Introduction to Stochastic Analysis

Z. Qian and J. G. Ying
Exeter College, Oxford and Fudan University

June 4, 2008

Contents

Introduction

This book has been written based on the lecture notes which the first author has delivered to various classes in the last 15 years. These lecture notes consist of the material which both authors have taught to different level groups of students at East China Normal University, Fudan University, Imperial College London and University of Oxford. The material is particularly welcomed by those practitioners in stochastic analysis, stochastic modelling and mathematical finance who demand for a good understanding of Itô's calculus and some computation skills in stochastic calculus.

A stochastic differential equation is a differential equation perturbed by random noise (its intensity may depend on the time-parameter t and the position X_t at instance t), and therefore possesses a form as following

$$\frac{dX_t}{dt} = A(t, X_t) + \sigma(t, X_t)\dot{W}_t \ .$$

The central limit theorem in probability theory suggests that $\dot{W}_t$ should have normal distribution, and for the sack of simplicity $(\dot{W}_t)_{t\geq 0}$ should be independent at different instance. Such random noise may be modelled ideally by a Brownian motion $(W_t)_{t\geq 0}$, a mathematical model describing chaotic movements of pollen particles in a liquid, first observed and reported by R. Brown in 1827. A mathematical formulation of Brownian motion and the description of its distribution were derived by Albert Einstein in a short paper titled "On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat" published in 1905, in *Annalen der Physik* 17, 549-560. About the same time, in 1900, L. Bachelier submitted his Ph. D. thesis in which he used Brownian motion to model stock markets. His results were published in a paper titled "Théorie de la spéculation" in *Ann. Sci. Ecole Norm. sup.*, 17 (1900), 21-86, which is probably the first paper devoted to applications of Brownian motion to finance.

On the other hand, the first mathematical construction of Brownian motion was achieved in 1923, in that year N. Wiener published his article "Differential space", *J. Math. Phys.* 2, 132-174. Fruitful results and many unusual features of Brownian motions were revealed mainly by Paul Lévy in 30's - 40's. Among of them, P. Lévy showed that almost surely $t \to W_t$ is non-where differentiable, and therefore the time-derivative of Brownian motion, $\dot{W}_t$, does not exist in ordinary sense. It is thus necessary to rewrite the previous stochastic differential equation in differential form

$$\mathrm{d}X_t = A(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

which has to be interpreted as an integral equation

$$X_t - X_0 = \int_0^t A(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s \ .$$

It is thus required to define integrals like

$$\int \sigma(t, X_t)\mathrm{d}W_t$$

which does not exist in ordinary sense. It was K. Itô in 1940's who first established an integration theory for Brownian motion, and therefore the theory of stochastic differential equations. Among of the manifold applications and connections with partial differential equations, one of the most remarkable recent applications of Itô's theory is in the theory of finance. Although Itô's theory has not been recognized by the Fields medals committee until the award to Wendelin Werner in 2006 for the sophisticated applications to mathematical physics by his and his coauthors, it was brought worldwide recognition by awarding H. Markowitz, W. Sharpe and M. Miller the 1990 Nobel Prize, and Robert Merton, and M. Scholes the 1997 Nobel Prize, both in Economics, but both for recognizing their works involving the novel applications of Itô's calculus in the economics.

This book is a moderate introduction to the theory of stochastic analysis, and aims to present the core part of Itô's calculus and the theory of Markov processes under the unified martingale treatment. The material covered in the book provides a necessary background in stochastic analysis for those who are interested in stochastic modelings and their applications including the theory of finance, stochastic control and filtering etc. Students who are majoring in (pure and applied) analysis, differential geometry, functional analysis, harmonic analysis, mathematical physics and PDEs will find this book relevant to their interests.

Zhongmin Qian, Oxford, UK
Jiangang Ying[1], Shanghai, China

[1]The work is supported partly by the National Basic Research Program of China (973 Program: Grant No. 2007CB814904)

Chapter 1

Preliminaries

1.1 The monotone class theorem

A collection $\mathcal{L}$ of subsets of Ω is called a π-system, if it is closed under finite intersections. Throughout the book, by the monotone class theorem (resp. by a monotone class argument), we mean the following lemma or a version of its variations (resp. the use of the monotone class theorem).

Lemma 1.1.1 *Let $\mathcal{L}$ be a π-system, $\mathcal{F} = \sigma\{\mathcal{L}\}$ be the smallest σ-algebra which contains $\mathcal{L}$, and $\mathcal{H}$ be a linear space of real valued functions on Ω which satisfies two conditions:*

1)$1 \in \mathcal{H}$ and $1_A \in \mathcal{H}$ for every $A \in \mathcal{L}$.

2) If $f_n \in \mathcal{H}$, $f_n \geq 0$, f_n is increasing in n, and $\sup_n f_n < +\infty$, then $\sup_n f_n \in \mathcal{H}$.

Then $\mathcal{H}$ contains all bounded, real-valued and $\mathcal{F}$-measurable functions on Ω.

1.2 Probabilities and processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is σ-algebra on Ω, and P is a probability measure on $(\Omega, \mathcal{F})$.

A measurable *function* on $(\Omega, \mathcal{F})$ is called a random *variable* in the probability theory. Recall that $X : \Omega \to \mathbb{R}^d$ is measurable, if for every Borel subset B of $\mathbb{R}^d$, the pre-image $X^{-1}(B) = \{\omega : X(\omega) \in B\}$ belongs to $\mathcal{F}$. Therefore, a *random variable* is such a *function* X on Ω that we may be able to talk about, for example, what is the probability of the event that X lies in a Borel subset.

For $p \geq 1$. Let $L^p(\Omega, \mathcal{F}, P)$ (or simply L^p if no confusion may arise) be the Banach space of p-th integrable random variables, and $||X||_{L^p} \equiv \sqrt[p]{E|X|^p}$ for every $X \in L^p(\Omega, \mathcal{F}, P)$. Throughout this book, EX (or $E(X)$) denotes the expectation of random variable, if it exists. More precisely

$$E(X) \equiv \int_\Omega X(\omega)P(d\omega) \ , \quad \forall X \in L^1.$$

A simple application of Hölder's inequality shows that, if $p \geq q$, then $L^p \subset L^q$ and$||X||_{L^q} \leq ||X||_{L^p}$.

Stochastic processes are mathematical models to describe random phenomena evolving with time. In this book, we only study stochastic processes in continuous time, thus the set of time-parameter will be $[0, +\infty)$ or its subset, unless otherwise specified.

A *stochastic process* is a parameterized family $X = (X_t)_{t\geq 0}$ of random variables taking values in some topological space S, called the state space of the stochastic process X. A stochastic process $X = (X_t)_{t\geq 0}$ may be considered as a function from $[0, \infty) \times \Omega \to S$, which is the reason why a stochastic process is also called a random function. A stochastic process $X = (X_t)_{t\geq 0}$ is integrable (resp., square integrable) if each X_t is integrable (resp., square integrable).

For each sample point $\omega \in \Omega$, function $t \to X_t(\omega)$ from $[0, \infty)$ to S is called a *sample path* (or a trajectory, or a sample function). Naturally, a stochastic process $X = (X_t)_{t\geq 0}$ is continuous (resp. right-continuous, right-continuous with left-limits) if sample paths $t \to X_t(\omega)$ are continuous (resp. right-continuous, right-continuous with left-limits) for almost all $\omega \in \Omega$. Similarly a stochastic process is bounded if almost all sample paths are controlled by a same constant.

Example 1.2.1 *(The Poisson process) Let (ξ_n) be a sequence of independent identically distributed (i.i.d.) random variables with exponential distribution of parameter $\lambda > 0$. Let*

$$T_0 = 0; \quad T_n = \sum_{j=1}^{n} \xi_j$$

and, for every $t \geq 0$ define

$$X_t = n \quad \text{ if } \quad T_n \leq t < T_{n+1}.$$

Then for every sample point ω, $t \to X_t(\omega)$ is a step function, constant on each (random) interval $[T_n, T_{n+1})$, with jump 1 at (random time) T_n, and is right-continuous with left limit $n-1$ at T_n.

Let $X = (X_t)_{t\geq 0}$ be a stochastic process with state space S, and $0 \leq t_1 < t_2 < \cdots < t_n$. The joint distribution (also called law) of $(X_{t_1}, \cdots, X_{t_n})$ is a probability measure on $S \times \cdots \times S$ (with the product Borel σ-algebra), denoted by $\mu_{t_1,t_2,\cdots,t_n}$, is defined by

$$\mu_{t_1,t_2,\cdots,t_n}(dx_1, \cdots, dx_n) = P\left(X_{t_1} \in dx_1, \cdots, X_{t_n} \in dx_n\right).$$

$\mu_{t_1,t_2,\cdots,t_n}$ is called a finite-dimensional distribution of $X = (X_t)_{t\geq 0}$. More precisely, $\mu_{t_1,t_2,\cdots,t_n}$ is the unique measure on $S \times \cdots \times S$ such that

$$\mu_{t_1,t_2,\cdots,t_n}(B_1 \times \cdots \times B_n) = P\left(X_{t_1} \in B_1, \cdots, X_{t_n} \in B_n\right).$$

Therefore, if f is a Borel measurable function on $S \times \cdots \times S$, then

$$\int_{S\times\cdots\times S} f(x_1, \cdots, x_n)\mu_{t_1,t_2,\cdots,t_n}(dx_1, \cdots, dx_n) = E\left(f(X_{t_1}, \cdots, X_{t_n})\right)$$

provided that $f(X_{t_1}, \cdots, X_{t_n})$ is integrable.

If $S = \mathbb{R}$, i.e. $(X_t)_{t\geq 0}$ is a real stochastic process, then the finite dimensional distribution $\mu_{t_1,t_2,\cdots,t_n}$ is determined by its distribution function

$$F_{t_1,t_2,\cdots,t_n}(x_1, \cdots, x_n) = P\left(X_{t_1} \leq x_1, \cdots, X_{t_n} \leq x_n\right),$$

where

$$F_{t_1,t_2,\cdots,t_n}(x_1, \cdots, x_n) = \mu_{t_1,t_2,\cdots,t_n}((-\infty, x_1] \times \cdots \times (-\infty, x_n]).$$

For a stochastic process $(X_t)_{t\geq 0}$ we are mainly interested in the properties determined by the family of its finite dimensional distributions

$$\{\mu_{t_1,t_2,\cdots,t_n} : 0 \leq t_1 \leq \cdots \leq t_n\}.$$

We may therefore consider the family of finite dimensional distributions as the distribution of the process $(X_t)_{t\geq 0}$, although the law of $(X_t)_{t\geq 0}$ is indeed a probability measure on a path type space which is typically infinite dimensional. There is no advantage to introduce the concept of the law or distribution for a stochastic process at this stage, and we will avoid this concept as far as possible before we have developed Itô's calculus. However it seems quite essential to have a proper understanding of the concept of the law of a process in the theory of Markov processes.

Two stochastic processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ on the same state space S are equivalent if they have the same family of finite dimensional distributions. If $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ are on the same probability space and for every $t \geq 0$

$$P\{\omega : X_t(\omega) = Y_t(\omega)\} = 1$$

then we say $(Y_t)_{t\geq 0}$ is a version of $(X_t)_{t\geq 0}$ (or $(X_t)_{t\geq 0}$ is a version of $(Y_t)_{t\geq 0}$). In this case, two processes are equivalent.

There are technical difficulties when we deal with stochastic processes in continuous time. For example, a subset of Ω like

$$\{\omega \in \Omega : X_t(\omega) \in B \text{ for all } t \in [0,1]\}$$

may be not measurable, i.e. not even an event, so that

$$P\left(\omega \in \Omega : X_t(\omega) \in B \text{ for all } t \in [0,1]\right)$$

may not make sense, unless additional conditions on $(X_t)_{t\geq 0}$ are imposed. Similarly, a function like $\sup_{t\in K} X_t$ may be not measurable. Such situation will be very inconvenient. To avoid such technical difficulties, a common condition, which is good enough to include a large class of interesting stochastic processes, is that X is right-continuous almost surely, and $(\Omega, \mathcal{F}, P)$ is complete in the sense that any trivial subsets of probability null are events.

Exercise 1.1 *Let $(X_t)_{t\geq 0}$ be a stochastic process in $\mathbb{R}^d$ on $(\Omega, \mathcal{F}, P)$, and B be a Borel measurable subset. If F is a finite or countable subset of $[0, +\infty)$, then both*

$$\{\omega : X_t(\omega) \in B \textit{ for any } t \in F\}$$

and $\sup_{t\in F} |X_t|$ are measurable.

In practical situations, we are given a collection of *consistent* finite dimensional distributions $\mathcal{D} = \{\mu_{t_1,\cdots,t_n}, \text{ for } t_1 < \cdots < t_n,\ t_j \in [0,\infty)\}$, we would like to construct a stochastic process $(X_t)_{t\geq 0}$ on some probability space $(\Omega, \mathcal{F}, P)$ so that the family of finite dimensional distributions determined by $(X_t)_{t\geq 0}$ is the given family $\mathcal{D}$. In this case, $(X_t)_{t\geq 0}$ is called a *realization* of $\mathcal{D}$.

1.3 Conditional expectations

The main concepts in the probability theory, including independence, martingale property and Markov property, are stated in terms of conditional expectations (and conditional probability).

Let X be an integrable random variable on $(\Omega, \mathcal{F}, P)$, and $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then there is a unique integrable random variable, denoted by $E(X|\mathcal{G})$, called the *conditional expectation* of X given $\mathcal{G}$, such that

1) $E(X|\mathcal{G})$ is measurable with respect to $\mathcal{G}$, and

2) For any $A \in \mathcal{G}$

$$E\left[E(X|\mathcal{G})1_A\right] = E\left(X1_A\right).$$

The conditional expectation $E(X|\mathcal{G})$ is the best prediction (under the mean square distance) of the random variable X based on available information $\mathcal{G}$. According to 1), 2) and the monotone class theorem

$$E\left(E(X|\mathcal{G})Y\right) = E\left(XY\right)$$

provided that Y is $\mathcal{G}$-measurable and XY is integrable.

If X and Y are two random variables and X is integrable, then $E(X|Y)$ means $E(X|\sigma(Y))$, where $\sigma(Y)$ is the smallest σ-algebra for which Y is measurable. It can be shown that $E(X|Y)$ is a measurable function of Y, that is, there is a Borel function F such that $E(X|Y) = F(Y)$.

Exercise 1.2 *If ξ, η are two random variables and ξ is $\sigma(\eta)$-measurable, then $\xi = f(\eta)$ for some Borel function f.*

According to definition, if Y is $\mathcal{G}$-measurable, then $E(YX|\mathcal{G}) = YE(X|\mathcal{G})$. If X and $\mathcal{G}$ are independent, then $E(X|\mathcal{G}) = E(X)$. Indeed X is independent of σ-algebra $\mathcal{G}$ if and only if $E(f(X)|\mathcal{G}) = E(f(X))$ for any bounded Borel measurable function f. Another property used frequently is tower property. Let $\mathcal{G}_1 \subset \mathcal{G}_2$ be two sub σ-algebra of $\mathcal{F}$. Then

$$E(E(X|\mathcal{G}_1)|\mathcal{G}_2) = E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(X|\mathcal{G}_1).$$

The verification of above properties of conditional expectation is left to readers as exercises.

1.4 Uniform integrability

The *uniform integrability* of a family of integrable random variables has been formulated to handle the convergence of random variables in $L^1(\Omega, \mathcal{F}, P)$. In spirit, it is very close to concepts of uniform convergence and uniform continuity.

If ξ is integrable, i.e. $\xi \in L^1$, then $\lim_{N\to\infty} E(1_{\{|\xi|\geq N\}}|\xi|) = 0$. Let $\mathcal{A}$ be a family of integrable random variables on $(\Omega, \mathcal{F}, P)$. We say $\mathcal{A}$ is uniformly integrable if

$$\lim_{N\to\infty} \sup_{\xi\in\mathcal{A}} E(1_{\{|\xi|\geq N\}}|\xi|) = 0,$$

that is, $E\left\{1_{\{|\xi|\geq N\}}|\xi|\right\}$ tends to zero uniformly on $\mathcal{A}$ as $N \to \infty$.

In terms of ε-δ language, $\mathcal{A}$ is uniformly integrable, if for every $\varepsilon > 0$ there is an $N > 0$ depending only on ε such that

$$E(1_{\{|\xi|\geq N\}}|\xi|) < \varepsilon$$

whenever $\xi \in \mathcal{A}$.

According to the definition, we have the followings.

1) Any finite family of integrable random variables is uniformly integrable.

2) Let $\mathcal{A} \subset L^1$ be a family of integrable random variables. If there is an integrable random variable η such that $|\xi| \leq \eta$ for every $\xi \in \mathcal{A}$, then $\mathcal{A}$ is uniformly integrable. Indeed

$$\sup_{\xi\in\mathcal{A}} E(1_{\{|\xi|\geq N\}}|\xi|) \leq E(1_{\{\eta\geq N\}}\eta) \to 0 \quad \text{as } N \to \infty.$$

3) If $\mathcal{A} \subset L^p$ for some $p > 1$ and $\sup_{\xi\in\mathcal{A}} E|\xi|^p < \infty$ (that is, $\mathcal{A}$ is a bounded subset of L^p), then $\mathcal{A}$ is uniformly integrable. Indeed

$$\begin{aligned}
\sup_{\xi\in\mathcal{A}} \int_{\{|\xi|\geq N\}} |\xi| dP &\leq \sup_{\xi\in\mathcal{A}} \int_{\{|\xi|\geq N\}} \frac{1}{N^{p-1}} |\xi|^p dP \\
&\leq \frac{1}{N^{p-1}} \sup_{\xi\in\mathcal{A}} E|\xi|^p \to 0
\end{aligned}$$

as $N \to \infty$.

4) If $\xi \in L^1$ and $\{\mathcal{G}_\alpha\}_{\alpha\in A}$ is a collection of sub σ-algebras of $\mathcal{F}$, then $\{E(\xi|\mathcal{G}_\alpha) : \alpha \in A\}$ is uniformly integrable. Indeed, let $\xi_\alpha = E(\xi|\mathcal{G}_\alpha)$. Then $\{\xi_\alpha \geq N\}$ and $\{\xi_\alpha \leq -N\}$ are $\mathcal{G}_\alpha$-measurable, so that

$$\begin{aligned}
E(1_{\{|\xi_\alpha|\geq N\}}|\xi_\alpha|) &= E(1_{\{\xi_\alpha\geq N\}}\xi_\alpha) - E(1_{\{\xi_\alpha\leq -N\}}\xi_\alpha) \\
&= E(1_{\{\xi_\alpha\geq N\}}\xi) - E(1_{\{\xi_\alpha\leq -N\}}\xi) \\
&= E(1_{\{|\xi_\alpha|\geq N\}}|\xi|) \\
&\leq E(1_{\{|\xi|\geq N\}}|\xi|)
\end{aligned}$$

which implies the uniform integrability.

The following equivalent condition of uniform integrability is often used, but its proof is elementary and left to the reader as an exercise.

Theorem 1.4.1 *Let $\mathcal{A} \subset L^1$. Then $\mathcal{A}$ is uniformly integrable if and only if $\mathcal{A}$ is a bounded subset of L^1 [i.e. $\sup_{\xi \in \mathcal{A}} E|\xi| < \infty$] and for every $\varepsilon > 0$ there is a $\delta > 0$ such that $E(1_A|\xi|) \leq \varepsilon$ whenever $\xi \in \mathcal{A}$ and $A \in \mathcal{F}$ with $P(A) \leq \delta$.*

The following theorem demonstrates the importance of uniform integrability.

Theorem 1.4.2 *Let $\{X_n\}_{n\in\mathbb{Z}^+}$ be a sequence of integrable random variables on $(\Omega, \mathcal{F}, P)$. Then $X_n \to X$ in L^1 for some random variable X as $n \to \infty$:*

$$E|X_n - X| \to 0 \quad \textit{as} \quad n \to \infty \ ,$$

if and only if $\{X_n\}_{n\in\mathbb{Z}^+}$ is uniformly integrable and $X_n \to X$ in probability.

1.5 Borel-Cantelli's lemma

The Borel-Cantelli lemma may be one of the mathematical theorems which are very easy to prove but very useful.

Theorem 1.5.1 *If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $\limsup_n A_n = 0$, where*

$$\begin{aligned} \limsup_n A_n &= \{\omega \text{ belongs to infinitely many } A_n\} \\ &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n \ . \end{aligned}$$

If $\{A_n\}$ are independent, then $\sum_{n=1}^{\infty} P(A_n) = \infty$ if and only if $P\left(\limsup_n A_n\right) = 1$.

Chapter 2

Elements in the martingale theory

The martingale theory was established by J.L. Doob, and summarized in his classical book "Stochastic Processes" published in 1950. Here we give a summary about martingales in discrete-time, then establish several sample path properties for martingales in continuous time.

Let $\mathbf{T}$ be either $[0, +\infty)$ or $\{0, 1, 2 \cdots\}$, $(\Omega, \mathcal{F}, P)$ be a probability space, and $(\mathcal{F}_t)_{t \in \mathbf{T}}$ be an increasing family of sub σ-algebras of $\mathcal{F}$ (called a *filtration*). Let

$$\mathcal{F}_\infty = \sigma \{\mathcal{F}_t : \ t \in \mathbf{T}\}$$

(we often use the notation that $\mathcal{F}_\infty = \bigvee_{t \in \mathbf{T}} \mathcal{F}_t$) which is the smallest σ-algebra containing all $\mathcal{F}_t$ as sub σ-algebras.

A stochastic process $X = (X_t)_{t \in \mathbf{T}}$ is adapted (to $(\mathcal{F}_t)_{t \in \mathbf{T}}$) if X_t is $\mathcal{F}_t$-measurable for every $t \in \mathbf{T}$. Let $X = (X_t : t \in \mathbf{T})$ be an adapted process, such that $X_t \in L^1(\Omega, \mathcal{F}, P)$ for every t. Then X is a *martingale* if

$$E(X_t | \mathcal{F}_s) = X_s \quad \text{a.s.} \quad \forall s, t \in \mathbf{T}, \, s \leq t,$$

X is a *sub-martingale* if

$$E(X_t | \mathcal{F}_s) \geq X_s \quad \text{a.s.} \ \forall s, t \in \mathbf{T}, \, s \leq t,$$

and X is a *super-martingale* if $-X$ is a sub-martingale. Actually martingale property may be defined for any subset $\mathbf{T}$ of $\mathbb{R}$.

As a simple application of Jensen's inequality, if $(X_t : t \in \mathbf{T})$ is a martingale and if $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, then $\{\varphi(X_t) : t \in \mathbf{T}\}$ is a sub-martingale provided $\varphi(X_t)$ is integrable for every t. Similarly, if $(X_t : t \in \mathbf{T})$ is a sub-martingale and if $\varphi : \mathbb{R} \to \mathbb{R}$ is an *increasing* convex function, then $\{\varphi(X_t)\}$ is a sub-martingale, provided every $\varphi(X_t)$ is integrable.

2.1 Martingales in discrete time

We shall first consider discrete-time case $\mathbb{Z}^+ = \{0, 1, 2, \cdots\}$. We usually use n as discrete time parameter. A stochastic process $\{X_n : n = 0, 1, 2 \cdots\}$ is also called a random sequence. It is a martingale (resp. sub-martingale, super-martingale) if and only if

$$E(X_{n+1}|\mathcal{F}_n) = (\text{resp. } \geq, \leq) X_n \quad \text{a.s.} \quad \forall n \geq 0.$$

Exercise 2.1 *1) If $(X_n : n \in \mathbb{Z}^+)$ is a sub-martingale, so is (X_n^+). If $X_n \log^+ X_n$ is integrable for every n, then $\left(X_n \log^+ X_n : n \in \mathbb{Z}^+\right)$ is a sub-martingale.*

2) Let $p \geq 1$. If $(X_n : n \in \mathbb{Z}^+)$ is a p-th integrable martingale (or a non-negative sub-martingale), then $(|X_n|^p)$ is a sub-martingale.

In the exercise, $\log^+ x = \log x$ if $x > 1$ and $\log^+ x = 0$ otherwise. Functions $(t \ln t)\, 1_{(1,\infty)}(t)$, $t1_{(0,\infty)}$ and $|t|^p$ (for $p \geq 1$) are all convex functions.

2.1.1 Doob's optional sampling theorem

The most important result in the theory of martingales, without doubt, is Doob's optional sampling theorem, which says (sub-, super-)martingale property remains true at (certain type of) random times.

Let $\mathbb{Z}^+$ denote the set of all non-negative integers. Then a map $T : \Omega \to \mathbb{Z}^+ \cup \{+\infty\}$ is called a *stopping time* (with respect to the filtration $(\mathcal{F}_n)$) if $\{\omega : T(\omega) = n\} \in \mathcal{F}_n$ for every $n \in \mathbb{Z}^+$. If T is a stopping time, then

$$\{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n \ \forall n \in \mathbb{Z}^+\}$$

is a σ-algebra, denoted by $\mathcal{F}_T$, called the σ-algebra at (random time) T. It is easy to show that if $X = (X_n)_{n\geq 0}$ is adapted, and T is a stopping time, then $X_T 1_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable, where

$$X_T 1_{\{T<\infty\}}(\omega) = \begin{cases} X_{T(\omega)}(\omega) & \text{if } T(\omega) < \infty \ , \\ 0 & \text{if } T(\omega) = \infty \ . \end{cases}$$

Optional sampling theorem If (X_n) is a martingale (resp. super-martingale), and if $S \leq T$ are two *bounded* stopping times, then

$$E(X_T|\mathcal{F}_S) = X_S \quad \text{a.s.}$$

(resp.

$$E(X_T|\mathcal{F}_S) \leq X_S \quad \text{a.s. })$$

For a complete proof see additional topics. Most results in this chapter are stated for sub-martingales and the readers may easily get the corresponding statements for super-martingales.

Exercise 2.2 *Let (X_n) be a sub-martingale, and T be a stopping time. Show that*

$$E|X_{T\wedge n}| \leq 2E\left(X_n^+\right) - E\left(X_0\right).$$

for any $n \in \mathbb{Z}^+$. Therefore, if $\sup_n E|X_n| < \infty$, then

$$E(|X_T|1_{\{T<\infty\}}) \leq 3\sup_n E|X_n| \ .$$

Hint. Using the decomposition $|X_n| = 2X_n^+ - X_n$, show that (X_n^+) is a sub-martingale, then estimate integrals by considering the partition determined by $\{T = k\}$.

Exercise 2.3 *Let $(X_n : n \geq 0)$ be a non-negative super-martingale and $T = \inf\{n : X_n = 0\}$. Prove that for any n,*

$$P(X_n > 0, n \geq T) = 0,$$

that is, X stays at 0 after it hits 0.

Exercise 2.4 *(Gambler's ruin) Let $X_n = \sum_{k=1}^n \xi_k$ where $\{\xi_n : n \geq 1\}$ is an i.i.d. sequence and $P(\xi_n = 1) = p$, $P(\xi_n = -1) = 1 - p$. For positive integers $a < b$, define*

$$\tau_0 = \inf\{n : X_n + a = 0\}, \ \tau_b = \inf\{n : X_n + a = b\}.$$

Using Doob's optional sampling theorem to compute the probability $P(\tau_0 < \tau_b)$.

2.1.2 Doob's inequalities

Martingale inequalities are among the most powerful analytic tools in mathematics and follows from Doob's optional sampling theorem.

Doob's maximal inequality Let $(X_n)_{n\geq 0}$ be a *sub-martingale*, $N \in \mathbb{N}$ and $\lambda > 0$. Then

$$\lambda P\left\{\sup_{n\leq N} X_n \geq \lambda\right\} \leq E(1_{\{\sup_{n\leq N} X_n \geq \lambda\}} X_N) \leq E(X_N^+) \tag{2.1}$$

and

$$\lambda P\left\{\inf_{n\leq N} X_n \leq -\lambda\right\} \leq E(1_{\{\sup_{n\leq N} X_n > -\lambda\}} X_N) - EX_0. \tag{2.2}$$

Moreover

$$\lambda P\left\{\sup_{n\leq N} |X_n| \geq \lambda\right\} \leq 2E(X_N^+) - E(X_0).$$

In particular, if $(X_n)_{n\geq 0}$ is a non-negative *sub-martingale*, then

$$P\left\{\sup_{n\leq N} X_n \geq \lambda\right\} \leq \frac{1}{\lambda} E(X_N).$$

For example to prove (2.1) we set $T = \inf\{n : X_n \geq \lambda\}$. Then

$$EX_N \geq EX_{N\wedge T}$$
$$= E(X_{N\wedge T}; max_{n\leq N} X_n \geq \lambda) + E(X_N; \max X_n < \lambda)$$
$$\geq \lambda P(\max_{n\leq N} X_n \geq \lambda),$$

and thus (2.1) follows.

Exercise 2.5 *If $(X_n : n \geq 0)$ is a non-negative super-martingale, prove that*

$$P\left\{\sup_{n\leq N} X_n \geq \lambda\right\} \leq \frac{1}{\lambda} E\left(X_0\right).$$

Kolmogorov's inequality If $(X_n)_{n\geq 0}$ is a square-integrable martingale, by applying Doob's maximal inequality (2.1) to X_n^2, one has

$$P\left\{\sup_{n\leq N} |X_n| \geq \lambda\right\} \leq \frac{1}{\lambda^2} E\left(X_N^2\right)$$

for any $\lambda > 0$, which is called Kolmogorov's inequality. Kolmogorov first proved it for sums of independent random variables.

Doob's L^p-inequality This is the most useful form under the name of martingale inequalities. Let $(X_n)_{n\geq 0}$ be a non-negative sub-martingale. Then

1) For every $p > 1$

$$E\left(\max_{n\leq N} X_n\right)^p \leq \left(\frac{p}{p-1}\right)^p E\left(X_N^p\right).$$

2) In the case $p = 1$

$$E\left(\max_{n\leq N} X_n\right) \leq \frac{e}{e-1}\left\{1 + \max_{n\leq N} E(X_n \log^+ X_n)\right\}.$$

In particular, if (X_n) is a martingale, then for any $p > 1$

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

2.1.3 The martingale convergence theorem

The convergence theorem is another powerful tool in the theory of martingales and follows from the upcrossing inequality, another important result of Doob's.

Let $(X_n)_{n\geq 0}$ be an adapted process, and $a < b$ be two numbers. Define a sequence of stopping times:

$$\begin{aligned} T_0 = 0T_1 &= \inf\{n \geq 0 : X_n \leq a\} \ , \\ T_2 = \inf\{n > T_0 : X_n \geq b\} \ , & \\ &\cdots\cdots \\ T_{2j-1} = \inf\{n > T_{2j-1} : X_n \leq a\} \ , & \\ T_{2j} = \inf\{n > T_{2j} : X_n \geq b\} \ . & \end{aligned}$$

According to the definition of T_j, if $T_{2j}(\omega) < \infty$, then the sequence

$$X_1(\omega), \cdots, X_{T_{2j}}(\omega)$$

crosses the interval $[a, b]$ from below j times. Let $U_a^b(X; n)$ denote the number of up-crossings through $[a, b]$ by $(X_n)_{n\geq 0}$ up to time n. Then

$$\{U_a^b(X; n) = j\} = \{T_{2j} \leq n < T_{2j+2}\} \ .$$

In particular, $\{U_a^b(X; n) = j\} \in \mathcal{F}_n$. Obviously $X_{T_{2j-1}} \leq a$ on $\{T_{2j-1} < \infty\}$ and $X_{T_{2j}} \geq b$ on $\{T_{2j} < \infty\}$.

Theorem 2.1.1 (Doob's up-crossing inequality) *If $X = \{X_n\}$ is a sub-martingale, then*

$$P\left\{U_a^b(X; n) \geq k\right\} \leq \frac{1}{b-a} \int_{\{U_a^b(X;n)=k\}} (X_n - a)^+ dP$$

and

$$E[U_a^b(X; n)] \leq \frac{1}{b-a} E(X_n - a)^+ \ .$$

Theorem 2.1.2 (The martingale convergence theorem) *Let $(X_n)_{n\geq 0}$ be a sub-martingale. If $\sup_n E|X_n| < +\infty$, then*

$$X_n \to X_\infty$$

almost surely for some integrable random variable X_∞.

Exercise 2.6 *If $(X_n)_{n\geq 0}$ is a non-negative super-martingale, then*

$$E(X_\infty|\mathcal{F}_n) \leq X_n \qquad \forall n.$$

Corollary 2.1.3 *If $(X_n)_{n\geq 0}$ is a* uniformly integrable *sub-martingale, then $X_n \to X_\infty$ almost surely as well as in L^1 for some integrable random variable X_∞. Moreover*

$$E(X_\infty|\mathcal{F}_n) \geq X_n \quad \forall n \ .$$

The following theorem says that a negative integer indexed sub-martingale is uniformly integrable as long as its expectation sequence is bounded below.

Theorem 2.1.4 *Let $(X_n : n \le 0)$ be a sub-martingale with respect to $(\mathcal{F}_n : n \le 0)$, i.e.,*

$$E(X_{n+1}|\mathcal{F}_n) \ge X_n, \ \forall n < 0 \ .$$

If $\inf_{n\le 0} E(X_n) > -\infty$, then $(X_n : n \le 0)$ is uniformly integrable. Hence X_n converges to some $X_{-\infty} \in L^1$ almost surely and in L^1 as n goes to $-\infty$. Moreover

$$E(X_n|\mathcal{F}_{-\infty}) \ge X_{-\infty},$$

where $\mathcal{F}_{-\infty} = \bigcap_{n<0} \mathcal{F}_n$.

Proof. Let $n \le 0$. Since $EX_n \ge EX_{n-1}$, $\inf_n EX_n > -\infty$ implies that $x = \lim_{n\to-\infty} EX_n$ exists and is finite. For any given $\epsilon > 0$, take k such that $EX_k - x < \epsilon$. Then for $n \le k$, we have

$$\begin{aligned} E(|X_n|1_{\{|X_n|>\lambda\}}) &= E(X_n 1_{\{X_n>\lambda\}}) - E(X_n 1_{\{X_n<-\lambda\}}) \\ &= E(X_n 1_{\{X_n>\lambda\}}) + E(X_n 1_{\{X_n\ge-\lambda\}}) - EX_n \\ &\le E(X_k 1_{\{X_n>\lambda\}}) + E(X_k 1_{\{X_n\ge-\lambda\}}) - EX_k + \epsilon \\ &\le E(X_k 1_{\{X_n>\lambda\}}) + E(-X_k 1_{\{X_n<-\lambda\}}) + \epsilon \\ &\le E(|X_k|1_{\{|X_n|>\lambda\}}) + \epsilon. \end{aligned}$$

Moreover

$$\begin{aligned} P(|X_n| > \lambda) &\le \frac{1}{\lambda} E|X_n| = \frac{1}{\lambda} E(2X_n^+ - X_n) \\ &= \frac{1}{\lambda}(2EX_n^+ - EX_n) \le \frac{1}{\lambda}(2EX_0^+ - x) \end{aligned}$$

which implies the uniform integrability of X by Theorem 1.4.1. ∎

Exercise 2.7 *Let $\{\xi_n : n \ge 1\}$ be an integrable independent and identically distributed (i.i.d. in short) sequence and*

$$X_n = \frac{1}{n}\sum_{k=1}^{n} \xi_k .$$

Prove that $\{X_{-n} : n \le -1\}$ is a martingale and converges almost surely and in L^1. (How to show that strong law of large numbers holds?)

2.2 Martingales in continuous time

The concept of martingales (super- and sub-martingales) and Doob's inequalities in discrete-time may be extended to continuous time. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t\ge0}$ which is an increasing family of σ-algebras $\mathcal{F}_t \subset \mathcal{F}$ for all $t \ge 0$. Let $\mathcal{F}_\infty = \sigma\left(\bigcup_{t\ge0} \mathcal{F}_t\right)$. An $(\mathcal{F}_t)$-adapted real valued integrable process $(X_t)_{t\ge0}$ is called a martingale (resp. super-martingale; resp. sub-martingale), if for any $t \ge s$, $E\left(X_t|\mathcal{F}_s\right) = X_s$ (resp. $E\left(X_t|\mathcal{F}_s\right) \le X_s$, resp. $E(X_t|\mathcal{F}_s) \geq X_s$) almost surely. Similarly, the concept of stopping times can be stated in this setting as well, namely, a function $T : \Omega \to [0, +\infty]$ is a $(\mathcal{F}_t)$-stopping time if for every $t \geq 0$, $\{T \leq t\}$ belongs to $\mathcal{F}_t$. If T is a stopping time, then

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \{T \leq t\} \cap A \in \mathcal{F}_t \ \forall t \geq 0\}$$

is a σ-algebra representing the information available up to the random time T. The first thing to solve is always the measurability, which is not a problem in discrete-time case.

Theorem 2.2.1 *If $X = (X_t)_{t \geq 0}$ is a right-continuous stochastic process adapted to $(\mathcal{F}_t)_{t \geq 0}$ and if $T : \Omega \to [0, +\infty]$ is a stopping time, then random variables T and $X_T 1_{\{T<\infty\}}$ are measurable with respect to σ-algebra $\mathcal{F}_T$.*

Exercise 2.8 *The proof may be split into two parts. 1) A real process $(X_t : t > 0)$ is called progressively measurable if for any $t > 0$, $(s, \omega) \mapsto X_s(\omega)$ is measurable on $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$. Prove that any right continuous (or left continuous) adapted process is progressively measurable. 2) If X is progressively measurable, prove the conclusion of the theorem by viewing X_T as a combination of two mappings: (1) $\omega \mapsto (T(\omega) \wedge t, \omega)$ is measurable from $(\Omega, \mathcal{F}_t)$ to $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$. (2) $(s, \omega) \mapsto X_s(\omega)$ is measurable from $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

We shall first give an example to show a large class of stopping times, which are enough for this book.

Example 2.2.2 *If $X = (X_t)_{t \geq 0}$ is an adapted, continuous process on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and if $D \in \mathbb{R}^d$ is a bounded closed subset of $\mathbb{R}^d$, then*

$$T = \inf\{t \geq 0 : X_t \in D\}$$

is a stopping time. If $X_0 \in D^c$ and $T < \infty$, then

$$X_T \in \partial D \ .$$

In particular, if $d = 1$ and b is a real number, then

$$T_b = \inf\{t \geq 0 : X_t = b\}$$

is a stopping time. In this case $\sup_{t \in [0,N]} X_t$ is a random variable,

$$\left\{ \sup_{t \in [0,N]} X_t < b \right\} = \{T_b > N\}$$

and

$$\left\{ \sup_{t \in [0,N]} X_t \geq b \right\} = \{T_b \leq N\} \ .$$

The following lemma can be used to generalize many results about martingales in discrete-time to continuous-time setting.

Lemma 2.2.3 *Let $T : \Omega \to [0, +\infty]$ be a stopping time (with respect to $(\mathcal{F}_t)$). For $n \in \mathbb{N}$ define*

$$T^{(n)} = \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{\left\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\right\}} + (+\infty) 1_{\{T=+\infty\}}.$$

Then $T^{(n)}$ are stopping times, $T^{(n)} \geq T$, and $T^{(n)} \downarrow T$ as $n \to \infty$.

Proof. For any $t \geq 0$

$$\begin{aligned}
\left\{T^{(n)} \leq t\right\} &= \bigcup_{k=1}^{\infty} \left\{T^{(n)} \leq t\right\} \cap \left\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\right\} \\
&= \bigcup_{k/2^n \leq t} \left\{T^{(n)} \leq t\right\} \cap \left\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\right\} \\
&\in \bigvee_{k/2^n \leq t} \mathcal{F}_{\frac{k}{2^n}} \subset \mathcal{F}_t \ .
\end{aligned}$$

■

The main goal of this part is to study the regularity of martingales, which is not needed in the discrete time setting. For each $t \geq 0$ let $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$. Then $(\mathcal{F}_{t+})_{t\geq 0}$ is again a filtration and $\mathcal{F}_{t+} \supseteq \mathcal{F}_t$ for every t. If $T : \Omega \to [0, +\infty]$ is a $(\mathcal{F}_{t+})$-stopping time, then

$$\mathcal{F}_{T+} = \{A \in \mathcal{F}_\infty : (T \leq t) \cap A \in \mathcal{F}_{t+} \ \forall t \geq 0\} \ .$$

A filtration $(\mathcal{F}_t)$ is right-continuous if $\mathcal{F}_{t+} = \mathcal{F}_t$ for each $t \geq 0$. Clearly $(\mathcal{F}_{t+})$ is right-continuous. Similarly, for $t > 0$ we define $\mathcal{F}_{t-} = \sigma\{\mathcal{F}_s : s < t\}$ which is a left-continuous filtration.

Theorem 2.2.4 *If $(X_t)_{t\geq 0}$ is a martingale (resp. super-martingale, resp. sub-martingale) on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with almost all right-continuous sample paths, then $(X_t)_{t\geq 0}$ is a martingale (resp. super-martingale, resp. sub-martingale) on $(\Omega, \mathcal{F}, \mathcal{F}_{t+}, P)$.*

Proof. Let us prove the sub-martingale case. Since $(X_t)_{t\geq 0}$ is adapted to $(\mathcal{F}_{t+})_{t\geq 0}$ so we need to prove

$$E\left(X_t|\mathcal{F}_{s+}\right) \geq X_s \qquad P\text{-a.s.} \tag{2.3}$$

for any $t > s$. For any u between s and t

$$E\left(X_t|\mathcal{F}_u\right) \geq X_u \qquad P\text{-a.s.}$$

so that for any $A \in \mathcal{F}_{s+} \subset \mathcal{F}_u$

$$E\left(1_A X_t\right) \geq E\left(1_A X_u\right) \ .$$

Take (a sequence) $u \downarrow s$ in the inequality. By Theorem 2.1.4, $(X_u : t \geq u > s)$ is uniformly integrable and X_u converges to X_s in L^1. Hence we obtain

$$E\left(1_A X_t\right) \geq E\left(1_A X_s\right)$$

for any $A \in \mathcal{F}_{s+}$. ∎

Corollary 2.2.5 (Doob's optional sampling theorem) *Let $(X_t)_{t\geq 0}$ be a sub-martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with almost all right-continuous sample paths, and let S, T be bounded $(\mathcal{F}_t)$-stopping times with $S \leq T$. Then X_S, X_T are integrable and*

$$E(X_T|\mathcal{F}_{S+}) \geq X_S.$$

Proof. We may assume that $(\mathcal{F}_t)$ is right continuous by theorem above. Let $T^{(n)}$ and $S^{(n)}$ be defined as in Lemma 2.2.3. Then $(X_{T^{(-n)}} : n < 0)$ is a sub-martingale and $X_T = \lim_{n\to-\infty} X_{T^{(-n)}}$ by right continuity. Since $E(X_{T^{(n)}}) \geq E(X_0)$, it follows from Theorem 2.1.4 that $(X_{T^{(n)}} : n > 0)$ is uniformly integrable. Then X_T is integrable and $X_{T^{(n)}} \longrightarrow X_T$ in L^1. The similar results hold for S.

Now we have $E(X_{T^{(n)}} 1_A) \geq E(X_{S^{(n)}} 1_A)$ for any

$$A \in \mathcal{F}_S = \bigcap_{n>0} \mathcal{F}_{S^{(n)}}$$

by the Doob's optional sampling theorem for discrete time. The conclusion follows by taking n to infinity. ∎

Similar conclusions hold for martingales and super-martingales. One would ask when a martingale (super-martingale) has right-continuous sample paths almost surely. The question can be answered via Doob's convergence theorem for sub-martingales. Let $X = (X_t)_{t\geq 0}$ be a real valued stochastic process, and let $a < b$. If

$$F = \{0 \leq t_1 < t_2 < \cdots < t_N\}$$

is a finite subset of $[0, +\infty)$, then $U_a^b(X, F)$ denotes the number of upcrossings by $\{X_{t_1}, \cdots, X_{t_N}\}$, and if $D \subset [0, +\infty)$ then $U_a^b(X, D)$ denotes the supremum of $U_a^b(X, F)$ over finite subsets $F \subset D$. Obviously $D \to U_a^b(X, D)$ is increasing with respect to the inclusion $\subset$. In particular, if $X = (X_t)_{t\geq 0}$ is a $(\mathcal{F}_t)$-adapted process and if D is a countable subset of $[0, +\infty)$ then for every $t \geq 0$, $U_a^b(X, D \cap [0, t])$ is measurable with respect to $\mathcal{F}_t$. Applying Doob's upcrossing inequality to $(X_t)_{t\in F}$ where F is any finite subset, we establish

Theorem 2.2.6 (Doob's upcrossing inequality) *Let $X = (X_t)_{t\geq 0}$ be a sub-martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, $a < b$, and $t > 0$. Then*

$$E[U_a^b(X, D)] \leq \frac{1}{b-a} E\left(X_t - a\right)^+$$

for any countable subset $D \subset [0, t]$.

The martingale convergence theorem follows from the estimate. We are now in a position to prove the following fundamental theorem, called Föllmer's lemma.

Theorem 2.2.7 (Föllmer) *Let $(X_t)_{t\geq 0}$ be a sub-martingale (resp. martingale) on the filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and let D be a countable dense subset in $[0, +\infty)$. Then*

1) For almost all $\omega \in \Omega$the following right limit

$$Z_t(\omega) = \lim_{s\in D, s>t, s\downarrow t} X_s(\omega)$$

exists for all $t \geq 0$, and Z_t is $\mathcal{F}_{t+}$-measurable,

2) For almost all $\omega \in \Omega$ the following left limit

$$Z_{t-}(\omega) = \lim_{s<t, s\uparrow} Z_s(\omega) \ ,$$

exists for all $t > 0$. Therefore $(Z_t)_{t\geq 0}$ is a $(\mathcal{F}_{t+})$-adapted process with right-continuous sample paths and left limits.

3) For any $t \geq 0$

$$E\left(Z_t | \mathcal{F}_t\right) \geq X_t \ , \quad (\text{resp. } E\left(Z_t|\mathcal{F}_t\right) = X_t) \quad P\text{-a.s.}$$

4) $(Z_t)_{t\geq 0}$ is a sub-martingale (resp. martingale) on $(\Omega, \mathcal{F}, \mathcal{F}_{t+}, P)$.

Proof. The conclusion 1 and 2 follows directly from martingale convergence theorem. We only need to prove conclusions 3 and 4. Firstly we show the third statement. For any $r > t$, $r \in D$, we have

$$E\left(X_r|\mathcal{F}_t\right) \geq X_t$$

namely for any $A \in \mathcal{F}_t$

$$E\left(1_A X_r\right) \geq E(1_A X_t) \ .$$

Letting $r \downarrow t$ along D and using Theorem 2.1.4 we obtain

$$E\left(1_A Z_t\right) \geq E(1_A X_t) \ .$$

which is equivalent to the inequality in 3. Similarly, if $t > s$, $u > t > r > s$ and $u, r \in D$ then

$$E\left(X_u|\mathcal{F}_r\right) \geq X_r$$

In particular, for any $A \in \mathcal{F}_{s+} \subset \mathcal{F}_r$

$$E\left(1_A X_u\right) \geq E(1_A X_r) \ .$$

Letting $u \in D \downarrow t$ and $r \in D \downarrow s$, and using Theorem 2.1.4 again, we obtain that

$$E\left(1_A X_t\right) \geq E(1_A X_s)$$

for every $A \in \mathcal{F}_{s+}$, which implies statement 4. ■

We however can not in general conclude that $(Z_t)_{t\geq 0}$ is a version of $(X_t)_{t\geq 0}$. Two processes can be different.

Corollary 2.2.8 *Under the same assumptions and notations as in Theorem 2.2.7. In addition, assume that $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous. Then $(Z_t)_{t\geq 0}$ is a version of $(X_t)_{t\geq 0}$, that is, for each $t \geq 0$, $Z_t = X_t$ almost surely, if and only if $t \to E(X_t)$ is right-continuous.*

Proof. Since $Z_t \in \mathcal{F}_{t+} = \mathcal{F}_t$, we have, according to Theorem 2.2.7

$$Z_t = E\left(Z_t|\mathcal{F}_t\right) \geq X_t$$

and

$$E(Z_t) = \lim_{s\in D, s>t, s\downarrow t} E(X_s) \ .$$

Therefore, in order to have the equality $Z_t = X_t$, the necessary and sufficient condition is that

$$E(Z_t) = \lim_{s\in D, s>t, s\downarrow t} E(X_s) = E(X_t) \ .$$

However $s \to E(X_s)$ is increasing, and the above equality is equivalent to

$$\lim_{s>t, s\downarrow t} E(X_s) = E(X_t)$$

i.e. $t \to E(X_t)$ is right-continuous. ■

Corollary 2.2.9 *If $(\mathcal{F}_t)_{t\geq 0}$ is right continuous, and if $(X_t)_{t\geq 0}$ is a martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, then the process $(Z_t)_{t\geq 0}$ constructed in 2.2.7 is a version of $(X_t)_{t\geq 0}$.*

This is because for a martingale $(X_t)_{t\geq 0}$, $t \to E(X_t) = E(X_0)$ is a constant. Doob's inequalities in §2.1.2 still hold for right continuous sub-martingales or martingales. We give a proof for Kolmogorov inequality and other may be proved similarly.

Theorem 2.2.10 (Kolmogorov's inequality) *Let M be a right continuous square-integrable martingale. Then for any $\lambda > 0$*

$$P\left\{\sup_{0\leq t\leq T} |M_t| \geq \lambda\right\} \leq \frac{1}{\lambda^2} E\left(M_T^2\right) \ .$$

Proof. Since $(M_t)_{t\geq 0}$ is continuous

$$\sup_{0\leq t\leq T} |M_t| = \sup_{t\in D} |M_t|$$

for any countable dense subset D of $[0, T]$, $\sup_{0\leq t\leq T} |M_t|$ is a random variable. For each $n \in \mathbb{N}$, we may apply the Kolmogorov inequality to martingale in discrete-time $\{M_{Tk/2^n}; \mathcal{F}_{Tk/2^n}\}_{k\geq 0}$ to obtain

$$P\left\{\sup_{0\leq k\leq 2^n} |M_{Tk/2^n}| \geq \lambda\right\} \leq \frac{1}{\lambda^2} E\left(M_T^2\right) \ .$$

However, since $D = \{Tk/2^n : n, k \in \mathbb{N}\}$ is dense in $[0, T]$ so that

$$\sup_{0 \leq k \leq 2^n - 1} |M_{Tk/2^n}| \uparrow \sup_{0 \leq t \leq T} |M_t|$$

as $n \to \infty$. Therefore

$$\begin{aligned} P\left\{\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right\} &= \lim_{n \to \infty} P\left\{\sup_{0 \leq k \leq 2^n - 1} |M_{Tk/2^n}| \geq \lambda\right\} \\ &\leq \frac{1}{\lambda^2} E\left(M_T^2\right) . \end{aligned}$$

■

Exercise 2.9 *Prove Doob's inequalities listed in §2.1.2.*

There is a previsible version of the optional sampling theorem. Let $(\mathcal{F}_t)_{t \geq 0}$ be a right-continuous filtration. We say a $(\mathcal{F}_t)$-stopping time $T : \Omega \to [0, +\infty]$ is predictable if there is a sequence $\{T_n\}$ of $(\mathcal{F}_t)$-stopping times such that $T_n < T$ for every n and $\lim_{n \to \infty} T_n = T$. The filtration $(\mathcal{F}_t)_{t \geq 0}$ is quasi-left continuous if for every predictable stopping time T we have $\mathcal{F}_{T-} = \mathcal{F}_T$.

Theorem 2.2.11 (Doob's optional sampling theorem) *Let $(\mathcal{F}_t)_{t \geq 0}$ be a right-continuous filtration, and let $(X_t)_{t \in [0, \infty]}$ be a super-martingale whose sample paths are right-continuous and have left-limits. Let T and S be bounded, strictly positive predictable stopping times, and $S \leq T$. Then*

$$E\left(X_T | \mathcal{F}_{S-}\right) \leq X_{S-} \qquad a.s.$$

2.3 Local martingales

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space. According to the regularity theory for martingales, we will assume the filtration $(\mathcal{F}_t)_{t \geq 0}$ satisfies the following conditions to be satisfied, unless otherwise specified.

1) $(\Omega, \mathcal{F}, P)$ is a complete probability space.

2) The filtration $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous, that is, for each $t \geq 0$

$$\mathcal{F}_t = \mathcal{F}_{t+} \equiv \bigcap_{s > t} \mathcal{F}_s .$$

3) Each $\mathcal{F}_t$ contains all null sets in $\mathcal{F}$.

Those are called *the usual conditions*. The following theorem, which we will not prove in this book, provides us with a class of interesting stopping times.

Theorem 2.3.1 *Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space satisfying the usual conditions. Let $X = (X_t)_{t \geq 0}$ be an $\mathbb{R}^d$-valued, adapted stochastic process that is* right-continuous and has left-limits. *Then for any Borel subset $D \subset \mathbb{R}^d$ and $t_0 \geq 0$*

$$T = \inf\{t \geq t_0 : X_t \in D\}$$

is a stopping time, where $\inf \emptyset = +\infty$. *T is called the* hitting time *of D by the process X.*

The concept of stopping times provides us with a means of "localizing" quantities. Suppose $(X_t)_{t\geq 0}$ is a stochastic process, and T is a stopping time, then $X^T = (X_{t\wedge T})_{t\geq 0}$ is a stochastic process stopped at (random) time T, where

$$X_{t\wedge T}(\omega) = \begin{cases} X_t(\omega) & \text{if } t \leq T(\omega)\ ; \\ X_{T(\omega)}(\omega) & \text{if } t \geq T(\omega)\ . \end{cases}$$

Another interesting stopped process at random time T associated with X is the process $X1_{[0,T]}$ which is by definition

$$\begin{aligned} \left(X1_{[0,T]}\right)_t(\omega) &= X_t 1_{\{t\leq T\}}(\omega) \\ &= \begin{cases} X_t(\omega) & \text{if } t \leq T(\omega)\ ; \\ 0 & \text{if } t > T(\omega)\ . \end{cases} \end{aligned}$$

It is obvious that

$$X_t^T = X_t 1_{\{t\leq T\}} + X_T 1_{\{t>T\}}\ .$$

If $(X_t)_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, so are the process $(X_{t\wedge T})_{t\geq 0}$ stopped at stopping time T and $X_t 1_{\{t\leq T\}}$.

Definition 2.3.2 *An adapted stochastic process $X = (X_t)_{t\geq 0}$ on the filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is called a* local martingale *if there is an increasing family $\{T_n\}$ of finite stopping times such that*

$$T_n \uparrow +\infty \quad \text{as } n \to +\infty$$

(called a localizing sequence) and for each n, $(X_{t\wedge T_n})_{t\geq 0}$ is a martingale.

Remark 2.3.3 *In many books, the definition of local martingale is a little different, say M is a local martingale if a localizing sequence $\{T_n\}$ exists such that for each n, $(X_{t\wedge T_n} 1_{\{T_n>0\}})_{t\geq 0}$ is a martingale. This is more general at least including X equals identically a non-integrable $\mathcal{F}_0$-measurable random variable X_0 which our definition does not include.*

Exercise 2.10 *The definition of local martingale looks more natural after proving that if $(X_t : t \geq 0)$ is a right continuous $(\mathcal{F}_t)$-martingale, then the stopped process X^T is also an $(\mathcal{F}_t)$-martingale.*

Exercise 2.11 *Prove that a non-negative integrable continuous local martingale is a super-martingale. A bounded continuous local martingale is a martingale.*

Exercise 2.12 *Prove that any local martingale X has a localizing sequence $\{T_n\}$ such that X^{T_n} is a uniformly integrable martingale for any n, and prove that any continuous local martingale X has a localizing sequence $\{T_n\}$ such that X^{T_n} is a bounded martingale for any n.*

2.4 Additional topics

Discrete time stochastic integral and Doob's optional sampling theorem

Doob's optional sampling theorem is the origin of martingale theory. This very important theorem may be proved by a simple and intuitive idea. A random sequence $H = (H_n : n \geq 0)$ is said to be predictable if H_0 is $\mathcal{F}_0$-measurable and for $n \geq 1$, H_n is $\mathcal{F}_{n-1}$-measurable. Let X be adapted and H predictable. Define

$$\begin{aligned}(H.X)_0 &:= H_0 X_0, \\ (H.X)_n &:= (H.X)_{n-1} + H_n(X_n - X_{n-1}), \ n \geq 1.\end{aligned}$$

The process $H.X$ is called the stochastic integral of H with respect to X and it is the discrete form of usual stochastic integral.

Theorem 2.4.1 *Let X be adapted and H predictable such that $H.X$ is integrable. If X is a martingale, so is $H.X$. If X is a sub-martingale and $H \geq 0$, then $H.X$ is a sub-martingale.*

The proof is left as an exercise. Then we have for any random time τ,

$$\begin{aligned}X_{\tau\wedge n} &= X_n 1_{\{\tau\geq n\}} + X_{n-1} 1_{\{n>\tau\geq n-1\}} + \cdots + X_0 1_{\{\tau=0\}} \\ &= X_0 + \sum_{i=1}^{n} H_i(X_i - X_{i-1}),\end{aligned}$$

where $H_n = 1_{\{\tau\geq n\}}$. It is easily seen that the stopped process $(X_{\tau\wedge n} : n \geq 0)$ may be represented as a stochastic integral when τ is a stopping time, and it is hence a martingale. If τ, σ are bounded stopping times with $\tau \geq \sigma$, then $(X_{\tau\wedge n} - X_{\sigma\wedge n} : n \geq 0)$ is a martingale. Let n be large and we have

$$E(X_\tau) = E(X_\sigma)$$

which implies $E(X_\tau|\mathcal{F}_\sigma) = X_\sigma$.

Exercise 2.13 *Prove that $E(X_\tau) = E(X_\sigma)$ for all bounded stopping time τ, σ if and only if $E(X_\tau|\mathcal{F}_\sigma) = X_\sigma$ for all bounded stopping time τ, σ.*

Similarly if X is a sub-martingale and τ, σ are bounded stopping times with $\tau \geq \sigma$, we may construct a non-negative and predictable H such that $(H.X)_n = X_{\tau\wedge n} - X_{\sigma\wedge n}$, and hence it is a sub-martingale. Then we have

$$E(X_\tau) \geq E(X_\sigma),$$

which implies $E(X_\tau|\mathcal{F}_\sigma) \geq X_\sigma$.

Chapter 3

Brownian motion

In this chapter we are going to introduce the most useful stochastic process in probability theory. It has various nice properties, for example, it is continuous, a Gaussian process, a martingale and a Markov process. Though its sample paths do not have bounded variation, they do have bounded quadratic variation, a property which allows us to define stochastic integral and do Itô's calculus.

3.1 Construction of Brownian motion

Brownian motion is a mathematical model of random movements observed by botanist Robert Brown.

Definition 3.1.1 *A stochastic process $B = (B_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, P)$ taking values in $\mathbb{R}^d$ is called a Brownian motion (BM) in $\mathbb{R}^d$, if*

1) $(B_t)_{t\geq 0}$ possesses independent increments: for any $0 \leq t_0 < t_1 < \cdots < t_n$ random variables

$$B_{t_0},\ \ B_{t_1} - B_{t_0},\ \cdots\ ,\ \ B_{t_n} - B_{t_{n-1}}$$

are independent.

2) For any $t > s \geq 0$, random variable $B_t - B_s$ has a normal distribution $N(0, t-s)$, that is, $B_t - B_s$ has the probability density function

$$p(t-s, x) = \frac{1}{\left(2\pi(t-s)\right)^{d/2}} e^{-\frac{|x|^2}{2(t-s)}}\ ;\quad x \in \mathbb{R}^d\ .$$

In other words

$$P\{B_t - B_s \in dx\} = p(t-s, x)dx\ .$$

3) Almost all sample paths of $(B_t)_{t\geq 0}$ are continuous.

If, in addition, $P\{B_0 = x\} = 1$ where $x \in \mathbb{R}^d$, then we say $(B_t)_{t\geq 0}$ is a Brownian motion starting at x. In particular if $P\{B_0 = 0\} = 1$ where 0 is the origin of $\mathbb{R}^d$, then we say $(B_t)_{t\geq 0}$ is a standard Brownian motion.

We will see that the condition 3) is not trivial, which ensure that many interesting functionals of Brownian motion are indeed random variables.

Let $p(t,x,y) = p(t,x-y)$, and define for every $t > 0$

$$P_t f(x) = \int_{\mathbb{R}^d} f(y) p(t,x,y) \mathrm{d}y \qquad \forall f \in C_b(\mathbb{R}^d) \ .$$

Since

$$p(t+s,x,y) = \int_{\mathbb{R}^d} p(t,x,z)p(s,z,y)dz,$$

$(P_t)_{t\geq 0}$ is a semigroup on $C_b(\mathbb{R}^d)$. $(P_t)_{t\geq 0}$ is called the heat semigroup in $\mathbb{R}^d$ in the sense that for each $f \in C_b^2(\mathbb{R}^d)$, then $u(t,x) = (P_t f)(x)$ solves the heat equation

$$\left(\frac{1}{2}\Delta + \frac{\partial}{\partial t}\right) u(t,x) = 0 \ ; \quad u(0,\cdot) = f \ ,$$

where $\Delta = \sum_i \frac{\partial^2}{\partial x_i^2}$ is the Laplace operator.

The connection between Brownian motion and the Laplace operator Δ (hence the harmonic analysis) is demonstrated through the following identity:

$$\begin{aligned} (P_t f)(x) &= E\left(f(B_t + x)\right) \\ &= \frac{1}{(2\pi t)^{d/2}} \int_{\mathbb{R}^d} f(y) \mathrm{e}^{-\frac{|y-x|^2}{2t}} \mathrm{d}y \end{aligned}$$

where B_t is a standard Brownian motion.

Example 3.1.2 *If $B = (B_t)_{t\geq 0}$ is a BM in $\mathbb{R}$, then*

$$E|B_t - B_s|^p = c_p |t-s|^{p/2} \quad \text{for all } s,t \geq 0 \tag{3.1}$$

for $p \geq 0$, where c_p is a constant depending only on p. Indeed

$$E|B_t - B_s|^p = \frac{1}{\sqrt{2\pi|t-s|}} \int_{\mathbb{R}} |x|^p \exp\left(-\frac{|x|^2}{2|t-s|}\right) dx \ .$$

Making change of variable

$$\frac{x}{\sqrt{|t-s|}} = y \ ; \quad dx = \sqrt{|t-s|} dy$$

we thus have

$$\begin{aligned} E|B_t - B_s|^p &= \frac{(\sqrt{|t-s|})^p}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^p \exp\left(-\frac{|x|^2}{2}\right) dx \\ &= c_p |t-s|^{p/2} \end{aligned}$$

where

$$c_p = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^d} |x|^p \exp\left(-\frac{|x|^2}{2}\right) dx \ .$$

(3.1) remains true for BM in $\mathbb{R}^d$ with a constant c_p depending on p and d.

Remark 3.1.3 *Since $B_t - B_s \sim N(0, t-s)$, it is an easy exercise to show that for every $n \in \mathbb{Z}^+$*

$$E(B_t - B_s)^{2n} = \frac{(2n)!}{2^n n!}|t-s|^n$$

or equivalently for $\xi \in \mathbb{R}$,

$$E\left(e^{-\xi(B_t - B_s)}\right) = \exp\left(\frac{1}{2}|\xi|^2(t-s)\right).$$

Let $B = (B_t)_{t\geq 0}$ be a standard BM in $\mathbb{R}$. Then B is a centered Gaussian process with co-variance function $C(s,t) = s \wedge t$. Indeed, any finite-dimensional distribution of B is Gaussian (exercise), and hence B is a centered Gaussian process, and its co-variance function (if $s < t$)

$$\begin{aligned} E(B_t B_s) &= E((B_t - B_s)B_s + B_s^2) \\ &= E((B_t - B_s)B_s) + EB_s^2 \\ &= E(B_t - B_s)EB_s + EB_s^2 \\ &= s\ . \end{aligned}$$

Exercise 3.1 *A continuous centered Gaussian process $(X_t : t \geq 0)$ with covariance $E(X_s X_t) = t \wedge s$ for any $s, t \geq 0$ is a standard Brownian motion.*

Theorem 3.1.4 (N. Wiener) *There is a standard Brownian motion in $\mathbb{R}^d$.*

Proof. We may assume that $d = 1$, the proof in higher dimension is similar. Observe that a standard BM (B_t) must be a Gaussian process (i.e. a process whose finite-dimensional distributions are Gaussian distributions) with mean zero and variance function $E(B_t B_s) = s \wedge t$. Therefore we may first construct a Gaussian process (X_t) such that $EX_t = 0$ and $E(X_t X_s) = s \wedge t$ on some complete probability space $(\Omega, \mathcal{F}, P)$ by Kolmogorov's extension theorem. It can be verified that $(X_t)_{t\geq 0}$ satisfies all conditions in the definition of BM, except the continuity of its sample paths. The Gaussian process (X_t) may be not continuous, and we thus need to modify the construction of X_t to make it continuous. Let $D = \{\frac{j}{2^n} : j \in \mathbb{Z}^+,\ n \in \mathbb{N}\}$ the dyadic real numbers. The important fact is that D is dense in $\mathbb{R}^+$. Define

$$H = \bigcup_{N=1}^{\infty} \bigcap_{l=1}^{\infty} \bigcup_{n=l}^{\infty} \bigcup_{j=1}^{N2^n} \left(\left| X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}} \right| \geq \frac{1}{2^{n/8}} \right).$$

Let, for fixed N,

$$A_l = \bigcup_{n=l}^{\infty} \bigcup_{j=1}^{N2^n} \left(\left| X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}} \right| \geq \frac{1}{2^{n/8}} \right).$$

We are going to show that each $\bigcap_{l=1}^{\infty} A_l$ has probability zero, and therefore as a sum of countable many events with probability zero, $P(H) = 0$. Since

$$\begin{aligned} & P\left\{\bigcup_{j=1}^{N2^n}\left(\left|X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}}\right| \geq \frac{1}{2^{n/8}}\right)\right\} \\ \leq & \sum_{j=1}^{N2^n} P\left(\left|X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}}\right| \geq \frac{1}{2^{n/8}}\right) \\ = & N2^n P\left(\left|X_{\frac{1}{2^n}}\right| \geq \frac{1}{2^{n/8}}\right) \\ \leq & N2^n \left(2^{n/8}\right)^4 E\left|X_{\frac{1}{2^n}}\right|^4 \\ = & \left(2^{n/8}\right)^4 N2^n 3\left(\frac{1}{2^n}\right)^2 \\ = & 3N\frac{1}{2^{n/2}} \end{aligned}$$

we obtain the estimate

$$\begin{aligned} P(A_l) & \leq \sum_{n=l}^{\infty} P\left\{\bigcup_{j=1}^{N2^n}\left(\left|X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}}\right| \geq \frac{1}{2^{n/8}}\right)\right\} \\ & \leq 3N\sum_{n=l}^{\infty}\frac{1}{2^{n/2}} \\ & = \frac{3N\sqrt{2}}{\sqrt{2}-1}\frac{1}{\left(\sqrt{2}\right)^l} . \end{aligned}$$

Therefore

$$\begin{aligned} P\left(\bigcap_{l=1}^{\infty} A_l\right) & = \lim_{n\to\infty} P\{A_l\} \\ & \leq \frac{3N\sqrt{2}}{\sqrt{2}-1}\lim_{n\to\infty}\frac{1}{\left(\sqrt{2}\right)^l} \\ & = 0 . \end{aligned}$$

It follows that $P(H) = 0$, thus $P(H^c) = 1$. On the other hand, by De Morgan law

$$H^c = \bigcap_{N=1}^{\infty}\bigcup_{l=1}^{\infty}\bigcap_{n=l}^{\infty}\bigcap_{j=1}^{N2^n}\left\{\omega : \left|X_{\frac{j}{2^n}}(\omega) - X_{\frac{j-1}{2^n}}(\omega)\right| < \frac{1}{2^{n/8}}\right\}$$

and thus, if $\omega \in H^c$, then for any N, there is an l such that for any $n > l$ and for all $j = 1, \cdots, N2^n$ we have

$$\left|X_{\frac{j}{2^n}}(\omega) - X_{\frac{j-1}{2^n}}(\omega)\right| < \frac{1}{2^{n/8}} .$$

We may show as the exercise below that for any $\omega \in H^c$, $X_t(\omega)$ is uniformly continuous on any bounded interval, and then for any $t \geq 0$ the limit of $X_s(\omega)$ exists as $s \to t$ along the dyadic numbers, i.e. as $s \to t$ and $s \in D$. Moreover, D is dense in $[0, \infty)$, thus for any $t \in [0, \infty)$ we may define

$$B_t(\omega) = \lim_{s \in D \to t} X_s(\omega) \quad \text{if } \omega \in H^c$$

otherwise if $\omega \in H$ we set $B_t(\omega) = 0$. By definition, $(B_t)_{t \geq 0}$ is a continuous process which coincides with X_t on H^c when $t \in D$. It remains to verify that $(B_t)_{t \geq 0}$ is a version of X and hence a Brownian motion in $\mathbb{R}$ as an exercise. ■

We need the following exercise to prove that for $\omega \in H^c$, $(X_t(\omega) : t \in D \cap [0, n])$ is uniformly continuous for any n.

Exercise 3.2 *Let $\alpha > 0$ and f be a function on D such that for any N, there is an l such that for any $n > l$ and for all $j = 1, \cdots, N2^n$ we have*

$$\left| f\left(\frac{j}{2^n}\right) - f\left(\frac{j-1}{2^n}\right) \right| < \left(\frac{1}{2^n}\right)^{\alpha}.$$

Then for each $n > 0$, there exists a constant C_n such that for any $s, t \in D \cap [0, n]$,

$$|f(s) - f(t)| \leq C_n |s - t|^{\alpha}.$$

3.2 Scaling properties

Let $B = (B_t)_{t \geq 0}$ be a standard BM in $\mathbb{R}^d$. By definition, the distribution of increments of BM $B = (B_t)_{t \geq 0}$ is stationary, so that for any fixed time S, $\tilde{B}_t = B_{t+S} - B_S$ is again a standard Brownian motion. This statement is true indeed for any finite stopping time S, see additional topics.

Lemma 3.2.1 (Scaling invariance, self-similarity) *For any real number $\lambda \neq 0$*

$$M_t \equiv \lambda B_{t/\lambda^2}$$

is a standard BM in $\mathbb{R}^d$.

This statement follows directly from the definition of BM. In particular, $(-B_t)_{t \geq 0}$ is also a standard BM, and $(-B_t)_{t \geq 0}$ and $(B_t)_{t \geq 0}$ have the same distribution.

Lemma 3.2.2 *If U is an $d \times d$ orthonormal matrix, then $UB = (UB_t)_{t \geq 0}$ is a standard BM in $\mathbb{R}^d$. That is, BM is invariant under the action of orthogonal group of $\mathbb{R}^d$.*

This lemma is an easy corollary of the invariance property of Gaussian distributions under the orthogonal group action.

Lemma 3.2.3 *Let $B = (B_t)_{t\geq 0}$ be a standard BM in $\mathbb{R}$, and define*

$$M_0 = 0\ , \quad M_t = tB_{1/t} \quad \text{for } t > 0.$$

Then M is a standard BM in $\mathbb{R}$.

Proof. Obviously M_t is a centered Gaussian process and

$$\begin{aligned} E\left(M_t M_s\right) &= tsE\left(B_{1/t}B_{1/s}\right) \\ &= ts\left(\frac{1}{t} \wedge \frac{1}{s}\right) = s \wedge t. \end{aligned}$$

Moreover $t \to M_t$ is continuous for $t > 0$. To see the continuity of M_t at $t = 0$, we use the fact that

$$\lim_{t\to\infty} \frac{B_t}{t} = 0$$

which is the law of large numbers for BM. We will not prove this here, but refer readers to additional topics. ∎

3.3 Markov property and finite-dimensional distributions

Let $X = (X_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$. For every $t \geq 0$, set

$$\mathcal{F}_t^0 = \sigma\{X_s : s \leq t\}$$

which is the smallest σ-algebra such that every X_s (where $s \leq t$) is measurable. In particular, for each $t \geq 0$, $X_t \in \mathcal{F}_t^0$ and in this sense we say $(X_t)_{t\geq 0}$ is adapted to the filtration $\{\mathcal{F}_t^0\}$. The family $\{\mathcal{F}_t^0\}_{t\geq 0}$ is called the filtration generated by $X = (X_t)_{t\geq 0}$.

In this section, we use $(\mathcal{F}_t^0)_{t\geq 0}$ to denote the filtration generated by a standard Brownian motion $(B_t)_{t\geq 0}$, and let

$$\mathcal{F}_\infty^0 = \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t^0\right).$$

Lemma 3.3.1 *For any $t > s \geq 0$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s^0$.*

Recall that

$$p(t,x) = \frac{1}{(2\pi t)^{d/2}} \mathrm{e}^{-\frac{|x|^2}{2t}}$$

in $\mathbb{R}^d$, and $(P_t)_{t\geq 0}$ the heat semigroup

$$P_t f(x) = \int_{\mathbb{R}^d} f(y)p(t, y - x)\mathrm{d}y \tag{3.2}$$

for every $t > 0$.

Lemma 3.3.2 *If $t > s$, then the joint distribution of B_s and B_t is given by*

$$P\{B_s \in dx, \;\; B_t \in dy\} = p(s,x)p(t-s,y-x)\,dx\,dy \ .$$

Indeed, since B_s and $B_t - B_s$ are independent, $(B_s, B_t - B_s)$ has density function

$$p(s,x_1)p(t-s,x_2)$$

thus, for any bounded Borel measurable function f

$$\begin{aligned} Ef(B_s,B_t) &= Ef(B_s, B_t - B_s + B_s) \\ &= \iint f(x_1, x_2 + x_1)p(s,x_1)p(t-s,x_2)\mathrm{d}x_1\mathrm{d}x_2 \ . \end{aligned}$$

Making change of variables $x_1 = x$ and $x_2 + x_1 = y$ in the last double integral, the induced Jacobi is 1 so that $\mathrm{d}x_1\mathrm{d}x_2 = \mathrm{d}x\mathrm{d}y$ (as measures), and therefore

$$Ef(B_s,B_t) = \iint f(x,y)p(s,x)p(t-s,y-x)\mathrm{d}x\mathrm{d}y$$

which implies that the probability density function of (B_s, B_t) is $p(s,x)p(t-s,y-x)$.

Theorem 3.3.3 *Let $t > s$, and f be a bounded Borel measurable function. Then*

$$E\left\{f(B_t)|\mathcal{F}_s^0\right\} = P_{t-s}f(B_s) \qquad a.s. \tag{3.3}$$

where $(P_t)_{t>0}$ is the heat semigroup as defined in (3.2). In particular

$$E\left\{f(B_t)|\mathcal{F}_s^0\right\} = E\left\{f(B_t)|B_s\right\}$$

which is called Markov property.

Proof. First we show that

$$E\left\{f(B_t)|\mathcal{F}_s^0\right\} = E\left\{f(B_t)|B_s\right\}$$

which is called Markov property of $(B_t)_{t\geq 0}$.

Since B_s and $B_t - B_s$ are independent, we have for any Borel sets A_1, A_2,

$$\begin{aligned} E(1_{A_1}(B_t - B_s)1_{A_2}(B_s)|\mathcal{F}_s^0) &= 1_{A_2}(B_s)P(B_t - B_s \in A_1) \\ &= \int p(t-s,x)1_{A_1\times A_2}(x,B_s)dx. \end{aligned}$$

It follows from monotone class theorem that for any bounded Borel measurable function g on $\mathbb{R}^d \times \mathbb{R}^d$,

$$E[g(B_t - B_s, B_s)|\mathcal{F}_s^0] = \int_{\mathbb{R}^d} p(t-s,x)g(x,B_s)dx.$$

Hence we have

$$\begin{aligned} E[f(B_t)|\mathcal{F}_s^0] &= Ef(B_t - B_s + B_s) \\ &= \int p(t-s,x)f(x+B_s)dx \\ &= \int p(t-s,y-B_s)f(y)dy = P_{t-s}f(B_s). \end{aligned}$$

This approach leads to a general treatment which is Lemma 6.5.2. ■

The family of finite-dimensional distributions of BM can be calculated in terms of the Gaussian density function $p(t,x)$.

Proposition 3.3.4 *Let $B = (B_t)$ be a standard BM in $\mathbb{R}^d$. For any $0 < t_1 < t_2 < \cdots < t_n$, the ($\mathbb{R}^{n\times d}$-valued) random variable $(B_{t_1}, \cdots, B_{t_n})$ has density function*

$$p(t_1,x_1)p(t_2-t_1,x_2-x_1)\cdots p(t_n-t_{n-1},x_n-x_{n-1})$$

where

$$p(t,x) = \frac{1}{(2\pi t)^{d/2}} e^{-|x|^2/(2t)}$$

is a standard Gaussian probability density function on $\mathbb{R}^d$, namely, the joint distribution of $(B_{t_1}, \cdots, B_{t_n})$ is given by

$$\begin{aligned} & P\{B_{t_1} \in dx_1, \ \cdots, \ B_{t_n} \in dx_n\} \\ = \ & p(t_1,x_1)p(t_2-t_1,x_2-x_1)\cdots p(t_n-t_{n-1},x_n-x_{n-1})dx_1\cdots dx_n \quad (3.4) \end{aligned}$$

Proof. Let f be a bounded, continuous function. We want to calculate

$$E\left(f(B_{t_1}, \cdots, B_{t_n})\right) .$$

One can use the fact that $B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}$ are independent, and has the joint distribution with density function

$$p(t_1,z_1)p(t_2-t_1,z_2)\cdots p(t_n-t_{n-1},z_n) .$$

(3.4) follows after change of variables. Below we present an induction argument which uses only the Markov property. Indeed, by the Markov property

$$\begin{aligned} & E\left(f_1(B_{t_1})\cdots f_n(B_{t_n})\right) \\ = \ & E\left\{E\left(f_1(B_{t_1})\cdots f_n(B_{t_n})|\mathcal{F}^0_{t_{n-1}}\right)\right\} \\ = \ & E\left\{f_1(B_{t_1})\cdots f_{n-1}(B_{t_{n-1}})E\left(f_n(B_{t_n})|\mathcal{F}^0_{t_{n-1}}\right)\right\} \\ = \ & E\left\{f_1(B_{t_1})\cdots f_{n-1}(B_{t_{n-1}})\left(P_{t_n-t_{n-1}}f_n\right)(B_{t_{n-1}})\right\} \\ = \ & E\left\{f_1(B_{t_1})\cdots f_{n-2}(B_{t_{n-2}})\left(f_{n-1}P_{t_n-t_{n-1}}f_n\right)(B_{t_{n-1}})\right\} \end{aligned}$$

which reduces the number of times t_i to $n-1$, and then the conclusion follows from the induction immediately. ■

Corollary 3.3.5 *Let $B_t = (B_t^1, \cdots, B_t^d)$ be a d-dimensional standard Brownian motion. Then for each j, B_t^j is a standard BM in $\mathbb{R}$, and $(B_t^j)_{t\geq 0}$ $(j = 1, \cdots, d)$ are mutually independent.*

Therefore a d-dimensional BM is d independent copies of BM in $\mathbb{R}$.

3.4 The reflection principle

Brownian motion starts afresh at a stopping time, i.e. the Markov property for Brownian motion remains true at stopping times. Therefore Brownian motion possesses the *strong Markov property*, a very important property which had been used by Paul Lévy in the form of the *reflection principle*, long before the concept of strong Markov property had been properly defined. We will exhibit this principle by computing the distribution of the running maximum of a Brownian motion.

In many applications, especially in statistics, we would like to estimate distributions of running maxima of a stochastic process. For Brownian motion $B = (B_t)_{t\geq 0}$, the distribution of $\sup_{s\in[0,t]} B_s$ can be derived by means of the reflection principle.

Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ in $\mathbb{R}$. Let $b > 0$ and $b > a$, and let

$$T_b = \inf\{t > 0 : B_t = b\} .$$

Then T_b is a stopping time, and the Brownian motion starts afresh as a standard Brownian motion after hitting b, and therefore

$$\begin{aligned} P\left\{\sup_{s\in[0,t]} B_s \geq b, B_t \leq a\right\} &= P\left\{\sup_{s\in[0,t]} B_s \geq b, B_t \geq 2b - a\right\} \\ &= P\{B_t \geq 2b - a\} \end{aligned}$$

where the first equality follows from the "fact" that the Brownian motion starting at T_b (in position b): $B_{T_b} = b$, runs afresh like a Brownian motion, so that it moves with equal probability about the line $y = b$. For a rigorous proof, see additional topics. The second equality follows from $2b - a = b + (b - a) > b$.

The above equation may be written as

$$\begin{aligned} P\{T_b \leq t, B_t \leq a\} &= P\{T_b \leq t, B_t \geq 2b - a\} \\ &= P\{B_t \geq 2b - a\} \ , \end{aligned}$$

which can be justified by the *strong Markov property* of Brownian motion, a topic that will not pursue here. Therefore

$$P\left\{\sup_{s\in[0,t]} B_s \geq b, B_t \leq a\right\} = \frac{1}{\sqrt{2\pi t}} \int_{2b-a}^{+\infty} e^{-\frac{x^2}{2t}} \mathrm{d}t \ ,$$

which gives us the joint distribution of a Brownian motion and its maximum at a fixed time t. By differentiating in a and in b we conclude the following

Theorem 3.4.1 *Let $B = (B_t)_{t\geq 0}$ be a standard BM in $\mathbb{R}$, and let $t > 0$. Then the joint probability density function of random variables $(M_t = \sup_{s\in[0,t]} B_s, B_t)$ is given as*

$$P\left\{M_t \in db, B_t \in da\right\} = \frac{2(2b-a)}{\sqrt{2\pi t^3}} \exp\left\{-\frac{(2b-a)^2}{2t}\right\} da\, db$$

over the region $\{(b,a) : a \leq b, b > 0\}$ in $\mathbb{R}^2$.

In particular, for any $b > 0$,

$$\begin{aligned} P\left\{\sup_{s\in[0,t]} B_s \geq b\right\} &= P\left\{T_b \leq t\right\} \\ &= \frac{2}{\sqrt{2\pi t^3}} \int\int_{\{a\leq c, c\geq b\}} (2c-a) \exp\left\{-\frac{(2c-a)^2}{2t}\right\} \mathrm{d}a\mathrm{d}c \\ &= \frac{2}{\sqrt{2\pi t^3}} \int_b^{+\infty} \int_{-\infty}^{c} (2c-a) \exp\left\{-\frac{(2c-a)^2}{2t}\right\} \mathrm{d}a\mathrm{d}c \\ &= \frac{2}{\sqrt{2\pi t^3}} \int_b^{+\infty} \int_c^{+\infty} x \exp\left\{-\frac{x^2}{2t}\right\} \mathrm{d}x\mathrm{d}c \\ &= \frac{2}{\sqrt{2\pi t}} \int_b^{+\infty} \exp\left(-\frac{x^2}{2t}\right) \mathrm{d}x \end{aligned}$$

which gives the density function of $\sup_{s\in[0,t]} B_s$ (or the stopping time T_b) and leads to an exact formula for the tail probability of the Brownian motion.

Exercise 3.3 *The Laplace transform of T_b is*

$$E\left(e^{-sT_b}\right) = e^{-\sqrt{2s}b}.$$

This means that $b \mapsto T_b$ is a process of stationary and independent increments.

3.5 Martingale property

Let $B = (B_t^{(i)})_{t\geq 0}$ $(i = 1, \cdots, d)$ be a standard BM in $\mathbb{R}^d$, with its generated filtration $(\mathcal{F}_t^0)_{t\geq 0}$.

Proposition 3.5.1 *1) Each B_t is p-th integrable for every $p > 0$, and for $t > s$*

$$E(|B_t - B_s|^p) = c_{p,d}|t-s|^{p/2} \ . \tag{3.5}$$

2) $(B_t)_{t\geq 0}$ is a continuous, square-integrable martingale.

3) For each pair i, j, $M_t = B_t^{(i)} B_t^{(j)} - \delta_{ij} t$ is a continuous martingale.

Proof. The first part was proved before. Since $B_t - B_s$ is independent of $\mathcal{F}_s^0$ when $t > s$ we thus have

$$E(B_t - B_s|\mathcal{F}_s^0) = E(B_t - B_s) = 0.$$

Then

$$E(B_t|\mathcal{F}_s^0) = E(B_s|\mathcal{F}_s^0) = B_s$$

namely, $(B_t)_{t\geq 0}$ is a continuous martingale.

Obviously we only need to show 3) for BM in $\mathbb{R}$. In this case

$$\begin{aligned} E(B_t^2 - B_s^2|\mathcal{F}_s^0) &= E((B_t - B_s)^2 |\mathcal{F}_s^0) \\ &\quad + E(2B_s (B_t - B_s) |\mathcal{F}_s^0) \\ &= E((B_t - B_s)^2) + 2B_s E((B_t - B_s) |\mathcal{F}_s^0) \\ &= E (B_t - B_s)^2 \\ &= t - s \end{aligned}$$

so that

$$\begin{aligned} E(B_t^2 - t|\mathcal{F}_s^0) &= \mathbb{E}(B_s^2 - s|\mathcal{F}_s^0) \\ &= B_s^2 - s \end{aligned}$$

which shows that $B_t^2 - t$ is a martingale. ■

Theorem 3.5.2 *Let $B = (B_t)_{t\geq 0}$ be a continuous stochastic process in $\mathbb{R}$ such that $B_0 = 0$. Then $(B_t)_{t\geq 0}$ is a standard BM in $\mathbb{R}$, if and only if for any $\xi \in \mathbb{R}$ and $t > s$*

$$E\left\{\exp\left(\sqrt{-1}\langle \xi, B_t - B_s\rangle\right) |\mathcal{F}_s^0\right\} = e^{-(t-s)|\xi|^2/2}. \tag{3.6}$$

Proof. We observe that (3.6) implies $B_t - B_s$ is independent of $\mathcal{F}_s^0$ and has normal distribution with variance $t-s$. Conversely, if $(B_t)_{t\geq 0}$ is a standard BM in $\mathbb{R}$, then $B_t - B_s$ is independent of $\mathcal{F}_s^0$, and $B_t - B_s$ has a normal distribution of mean zero and variance $(t-s)$, so that

$$\begin{aligned} & E\left\{\exp\left(\sqrt{-1}\langle \xi, B_t - B_s\rangle\right) |\mathcal{F}_s^0\right\} \\ =\ & E\left\{\exp\left(\sqrt{-1}\langle \xi, B_t - B_s\rangle\right)\right\} \\ =\ & \frac{1}{\sqrt{2\pi(t-s)}} \int_{\mathbb{R}} e^{\sqrt{-1}\langle \xi, x\rangle - \frac{|x|^2}{2(t-s)}} \mathrm{d}x \\ =\ & \exp\left(-\frac{(t-s)|\xi|^2}{2}\right) . \end{aligned}$$

■

Corollary 3.5.3 *Let (B_t) be a standard BM in $\mathbb{R}^d$. If $\xi \in \mathbb{R}^d$, then*

$$M_t \equiv \exp\left(\sqrt{-1}\langle \xi, B_t\rangle + \frac{|\xi|^2}{2} t\right)$$

is a martingale.

Remark 3.5.4 *Note that both sides of (3.6) are analytic in ξ so that the identity continues to hold for any complex vector ξ. In particular, by replacing ξ by $-\sqrt{-1}\xi$ we obtain that*

$$E\left\{\exp\left(\langle \xi, B_t - B_s\rangle\right)|\mathcal{F}_s^0\right\} = e^{(t-s)|\xi|^2/2}$$

so that for any vector ξ

$$\exp\left(\langle \xi, B_t\rangle - \frac{|\xi|^2}{2}t\right)$$

is a continuous martingale. This statement will be extended to vector fields ξ in $\mathbb{R}$. The resulted identity is called Cameron-Martin formula.

Brownian motion is the basic example of Lévy processes: right continuous stochastic processes in $\mathbb{R}^d$ which possess stationary independent increments, and (3.6) is the Lévy-Khinchin formula for BM. In general if (X_t) is a Lévy process in $\mathbb{R}^d$, then

$$E\left\{\exp\left(\sqrt{-1}\langle \xi, X_t - X_s\rangle\right)|\mathcal{F}_s^0\right\} = e^{(t-s)\psi(\xi)}$$

for $t > s$ and $\xi \in \mathbb{R}^d$, where

$$\begin{aligned}\psi(\xi) &= -\frac{1}{2}\langle AA^T\xi, \xi\rangle + \sqrt{-1}\langle b, \xi\rangle \\ &\quad + \int_{\mathbb{R}^d\setminus\{0\}}\left(e^{\sqrt{-1}\langle \xi, x\rangle} - 1 - \sqrt{-1}1_{\{|x|<1\}}\langle \xi, x\rangle\right)\nu(dx)\end{aligned}$$

for some $d \times r$ matrix A, vector b and Lévy measure $\nu(dx)$ of (X_t) which is a σ-finite measure on $\mathbb{R}^d \setminus \{0\}$ satisfying the following integrable condition

$$\int_{\mathbb{R}^d\setminus\{0\}} \frac{|x|^2}{1+|x|^2}\nu(dx) < +\infty\ . \tag{3.7}$$

3.6 Quadratic variational processes

As we have seen that, both (one-dimensional) Brownian motion B_t and $M_t \equiv B_t^2 - t$ are martingales, thus

$$B_t^2 = M_t + A_t$$

with of course $A_t = t$. Therefore, the continuous sub-martingale B_t^2 is a *sum of a martingale and an adapted increasing process.* We will see in the next chapter that this decomposition for B_t^2 is the key to establish Itô's integration theory.

Lemma 3.6.1 *Let*

$$D = \{0 = t_0 < t_1 < \cdots < t_n = t\}$$

be a finite partition of the interval $[0, t]$, and let

$$V_D = \sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2$$

the quadratic variation of B over the partition D, which is a non-negative random variable. Then

$$EV_D = t$$

and the variance of V_D

$$E\left\{(V_D - EV_D)^2\right\} = 2\sum_{l=1}^{n}(t_l - t_{l-1})^2 \ .$$

Proof. Indeed

$$\begin{aligned} EV_D &= \sum_{l=1}^{n} E|B_{t_l} - B_{t_{l-1}}|^2 \\ &= \sum_{l=1}^{n}(t_l - t_{l-1}) \\ &= t \ . \end{aligned}$$

To prove the second formula we proceed as the following

$$\begin{aligned} & E\left\{(V_D - EV_D)^2\right\} \\ = \ & E\left\{\left(\sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2 - t\right)^2\right\} \\ = \ & E\left\{\left(\sum_{l=1}^{n} \left(|B_{t_l} - B_{t_{l-1}}|^2 - (t_l - t_{l-1})\right)\right)^2\right\} \\ = \ & \sum_{k,l=1}^{n} E\left\{\left(|B_{t_k} - B_{t_{k-1}}|^2 - (t_k - t_{k-1})\right)\left(|B_{t_l} - B_{t_{l-1}}|^2 - (t_l - t_{l-1})\right)\right\} \\ = \ & \sum_{l=1}^{n} E\left\{\left(|B_{t_l} - B_{t_{l-1}}|^2 - (t_l - t_{l-1})\right)^2\right\} \\ & + \sum_{k\neq l}^{n} E\left\{\left(|B_{t_k} - B_{t_{k-1}}|^2 - (t_k - t_{k-1})\right)\left(|B_{t_l} - B_{t_{l-1}}|^2 - (t_l - t_{l-1})\right)\right\} \ . \end{aligned}$$

Since the increments over different intervals are independent, the expectation of each product in the last sum on the right-hand side equals the product of their expectations, which gives contribution zero, and therefore

$$\begin{aligned}
& E\left\{(V_D - EV_D)^2\right\} \\
= & \sum_{l=1}^{n} E\left\{\left(|B_{t_l} - B_{t_{l-1}}|^2 - (t_l - t_{l-1})\right)^2\right\} \\
= & \sum_{l=1}^{n} E\left\{|B_{t_l} - B_{t_{l-1}}|^4 - 2(t_l - t_{l-1})|B_{t_l} - B_{t_{l-1}}|^2 + (t_l - t_{l-1})^2\right\} \\
= & \sum_{l=1}^{n} \left\{E|B_{t_l} - B_{t_{l-1}}|^4 - 2(t_l - t_{l-1})E|B_{t_l} - B_{t_{l-1}}|^2 + (t_l - t_{l-1})^2\right\} \\
= & 2\sum_{l=1}^{n}(t_l - t_{l-1})^2
\end{aligned}$$

where we have used the integral

$$E|B_{t_l} - B_{t_{l-1}}|^4 = 3(t_l - t_{l-1})^2 \ .$$

■

We are now in a position to prove the following

Theorem 3.6.2 *Let $B = (B_t)_{t\geq 0}$ be a standard BM in $\mathbb{R}$. Then*

$$\lim_{m(D)\to 0} \sum_l |B_{t_l} - B_{t_{l-1}}|^2 = t \qquad \text{in } L^2(\Omega, P)$$

for any t, where D runs over all finite partitions of interval $[0,t]$, and

$$m(D) = \max_l |t_l - t_{l-1}| \ .$$

Therefore

$$\lim_{m(D)\to 0} \sum_l |B_{t_l} - B_{t_{l-1}}|^2 = t \qquad \text{in probability.}$$

Proof. According to the previous lemma we have

$$\begin{aligned}
E\left|\sum_l |B_{t_l} - B_{t_{l-1}}|^2 - t\right|^2 &= E\left|V_D - E\left(V_D\right)\right|^2 \\
&= 2\sum_{l=1}^{n}(t_l - t_{l-1})^2 \\
&\leq 2m(D)\sum_{l=1}^{n}(t_l - t_{l-1}) \\
&= 2tm(D)
\end{aligned}$$

and therefore

$$\lim_{m(D)\to 0} E\left|\sum_l |B_{t_l} - B_{t_{l-1}}|^2 - t\right|^2 = 0 \ .$$

■

For good partitions the convergence in the above theorem takes place almost surely.

Proposition 3.6.3 *Let $(B_t)_{t\geq 0}$ be a standard BM in $\mathbb{R}$. Then for any $t > 0$ we have*

$$\sum_{j=1}^{2^n} \left| B_{\frac{j}{2^n}t} - B_{\frac{j-1}{2^n}t} \right|^2 \to t \quad a.s. \tag{3.8}$$

as $n \to \infty$.

Proof. Let D_n be the dyadic partition of $[0, t]$

$$D_n = \{0 = \frac{0}{2^n}t < \frac{1}{2^n}t < \cdots < \frac{2^n}{2^n}t = t\} \ .$$

and V_n denote V_{D_n}. Then, according to Lemma 3.6.1, $EV_n = t$ and

$$\begin{aligned} E\left|V_n - EV_n\right|^2 &= 2\sum_{l=1}^{2^n}\left(\frac{l}{2^n}t - \frac{l-1}{2^n}t\right)^2 \\ &= 22^n\left(\frac{1}{2^n}t\right)^2 \\ &= \frac{1}{2^{n-1}}t^2 \ . \end{aligned}$$

Therefore, by Markov's inequality,

$$\begin{aligned} P\left\{|V_n - EV_n| \geq \frac{1}{n}\right\} &\leq n^2 E\left|V_n - EV_n\right|^2 \\ &= \frac{n^2}{2^{n-1}}t^2 \end{aligned}$$

so that

$$\sum_{n=1}^{\infty} P\left\{|V_n - EV_n| \geq \frac{1}{n}\right\} = t^2\sum_{n=1}^{\infty}\frac{n^2}{2^{n-1}} < +\infty \ .$$

By the Borel-Cantelli lemma, it follows that $V_n \to t$ almost surely. ■

Remark 3.6.4 *Indeed the conclusion is true for monotone partitions. More precisely, for each n let*

$$D_n = \{0 = t_{n,0} < t_{1,n} < \cdots < t_{n_k,n} = t\}$$

be a finite partition of $[0, t]$. Suppose $D_{n+1} \supset D_n$ and

$$\lim_{n\to\infty} m(D_n) = \lim_{n\to\infty} \max|t_{n_i,n} - t_{n_{i-1},n}| = 0 \ .$$

Then

$$\sum_{i=1}^{n_k} \left| B_{t_{n_i,n}} - B_{t_{n_{i-1},n}} \right|^2 \to t \quad a.s. \tag{3.9}$$

as $n \to \infty$. *Indeed, in this case, if we denote by* M_n *the left-hand side of* (3.9), *then the reversed sequence*

$$\{\cdots\cdots, M_n, \cdots, M_2, M_1\} \tag{3.10}$$

is a non-negative martingale. Hence (3.9) *follows from the martingale convergence theorem. To prove* (3.10), *let* $D_1 = \{0, t\}$ *and* $D_2 = \{0, s, t\}$. *Set*

$$\tilde{B}_u = 1_{[0,s]}(u) B_u + 1_{[s,\infty)}(u)(B_s - (B_u - B_s)).$$

By the reflection principle $\tilde{B} = (\tilde{B}_u)$ *is also a standard BM. Define similarly* $\{\tilde{M}_n\}$ *for* $\tilde{B}$. *Then* $\tilde{M}_n = M_n$ *for* $n \geq 2$ *and*

$$\tilde{M}_1 = (2B_s - B_t)^2 = 2M_2 - M_1.$$

Clearly $E(M_1 | M_n, n \geq 2) = E(\tilde{M}_1 | \tilde{M}_n, n \geq 2)$ *and it follows then that*

$$\begin{aligned} E(M_1 | M_n, n \geq 2) &= \frac{1}{2} E(M_1 + \tilde{M}_1 | M_n, n \geq 2) \\ &= E(M_2 | M_n, n \geq 2) = M_2. \end{aligned}$$

It can be shown (not easy) that

$$\sup_D \sum_l |B_{t_l} - B_{t_{l-1}}|^p < \infty \quad \text{a.s.}$$

if $p > 2$, where sup is taken over all finite partitions of $[0, 1]$, and

$$\sup_D \sum_l |B_{t_l} - B_{t_{l-1}}|^2 = \infty \quad \text{a.s.}$$

That is to say, Brownian motion has finite p-variation for any $p > 2$. Indeed almost all Brownian motion sample paths are α-Hölder continuous for any $\alpha < 1/2$ but not for $\alpha = 1/2$. It follows that almost all Brownian motion paths are nowhere differentiable. We will not go into a deep study about the sample paths of BM, which are not needed in order to develop Itô's calculus for Brownian motion.

Definition 3.6.5 *Let* $p > 0$ *be a constant. A path* f *in* $\mathbb{R}^d$ *(a function on* $[0, T]$ *valued in* $\mathbb{R}^d$*) is said to have finite* p*-variation on* $[0, T]$, *if*

$$\sup_D \sum_l |f(t_i) - f(t_{i-1})|^p < +\infty$$

where D *runs over all finite partitions of* $[0, T]$. *It has finite (total) variation if it has finite* 1*-variation.*

A function with finite variation must be a difference of two increasing functions. In particular, it has at most countably many discontinuous points.

A stochastic process $V = (V_t)_{t\geq 0}$ is called a *variational process*, if for almost all $\omega \in \Omega$, the sample path $t \to V_t(\omega)$ possesses finite variation on any finite interval. A Brownian motion is not a variational process.

3.7 Additional topics

More about construction of Brownian motion

There are many approaches towards construction of Brownian motion. It is Wiener who first proved the continuity of sample paths of Brownian motion, but his proof is different. His original approach [22] is to construct a positive linear functional on $C([0,1])$ and prove that it induces a measure, which is later called Wiener measure. Here we would like to present another construction which Wiener provided later jointly with Paley. Let $\{g_n : n \geq 0\}$ be an i.i.d. sequence with standard Gaussian distribution on $L^2(\Omega, \mathcal{F}, P)$. It is known that a normal orthogonal base of $L^2([0,\pi]$ is

$$e_0(x) = \sqrt{1/\pi},\ e_n = \sqrt{2/\pi}\cos nx,\ n \geq 1.$$

Let H be a linear operator from $L^2([0,\pi])$ to $L^2(\Omega)$ such that $H(e_n) = g_n$ for each $n \geq 0$. Then H is isometric. For any $t \in [0,\pi]$, define $X_t = H(1_{[0,t]})$ and then

$$X_t = \frac{t}{\sqrt{\pi}} + \sqrt{\frac{2}{\pi}} \sum_{n\geq 1} \frac{\sin nt}{n} \cdot g_n. \tag{3.11}$$

It is not hard to verify that $(X_t : t \in [0,\pi])$ is a centered Gaussian process with co-variance $E(X_s X_t) = s \wedge t$. The magnificent thing is that Wiener shows that the sum in right hand side may be regrouped in some way,say

$$\frac{t}{\sqrt{\pi}}\xi_0 + \sqrt{\frac{2}{\pi}} \sum_{n\geq 0} \sum_{k=2^{n-1}}^{2^n - 1} \frac{\sin kt}{k} g_k$$

so that it converges uniformly on $[0,\pi]$ almost surely. The continuity follows then. See Itô and McKean [10] for details.

Law of large number for Brownian motion

To convince yourself why the law of large numbers for BM is true, we may look at a special way $t \to \infty$ through natural numbers, namely

$$\lim_{n\to\infty} \frac{B_n}{n} = \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n}$$

where $X_i = B_i - B_{i-1}$. Notice that (X_i) is a sequence of independent random variables with identical distribution $N(0,1)$, so that by the strong law of large numbers

$$\frac{X_1 + \cdots + X_n}{n} \to EX_1 = 0 \text{ almost surely.}$$

In order to handle the general case $t \geq 0$, we may write $t = [t] + r_t$ where $[t]$ is the integer part of t and $r_t \in [0, 1)$. Then

$$\frac{B_t}{t} = \frac{B_t - B_{[t]}}{t} + \frac{[t]}{t}\frac{B_{[t]}}{[t]}$$

the second term tends to 0 since as $t \to \infty$, $\frac{[t]}{t} \to 1$ and $\frac{B_{[t]}}{[t]} \to 0$. To see why

$$\frac{B_t - B_{[t]}}{t} \to 0$$

as $t \to \infty$, we need the following Gaussian tail estimate for BM (see §3.4 below)

$$\begin{aligned} P\left\{\omega : \sup_{t\in[0,T]} |B_t(\omega)| \geq R\right\} &= 2\sqrt{\frac{2}{\pi}} \int_{R/\sqrt{T}}^{\infty} e^{-x^2/2} dx \\ &\leq 2\exp\left(-\frac{R^2}{2T}\right) \qquad \text{for all } R > 0 \ . \end{aligned}$$

It follows that for any $\varepsilon > 0$

$$\sum_{n=0}^{\infty} P\left\{\omega : \sup_{t\in[n,n+1]} \frac{|B_t(\omega) - B_n(\omega)|}{n} \geq \varepsilon\right\} < \infty$$

and thus by the Borel-Cantelli lemma

$$\overline{\lim_{n\to\infty}} \sup_{t\in[n,n+1]} \left|\frac{B_t}{t} - \frac{B_n}{n}\right| = 0 \quad \text{almost surely.}$$

For more detail, see page 180-181 in [19].

A proof of reflection principle

The reflection principle follows from the fact that for any stopping time T, $(B_{t+T} - B_T : t \geq 0)$ is a standard Brownian motion independent of $\mathcal{F}_T$. Let us prove for a Borel set A,

$$P(B_{t+T} - B_T \in A|\mathcal{F}_T) = P(B_t \in A) \quad \text{ a.s. on } \{T < \infty\}. \tag{3.12}$$

Let $T^{(n)}$ be the discretization of T as in Lemma 2.2.3. Then

$$\begin{aligned} P(B_{t+T^{(n)}} - B_{T^{(n)}} \in A; T < \infty) &= \sum_{k\geq 1} P(B_{t+k/2^n} - B_{k/2^n} \in A; T^{(n)} = k/2^n) \\ &= \sum_{k\geq 1} P(B_{t+k/2^n} - B_{k/2^n} \in A)P(T^{(n)} = k/2^n) \\ &= \sum_{k\geq 1} P(B_t \in A)P(T^{(n)} = k/2^n) \\ &= P(B_t \in A)P(T < \infty) \ . \end{aligned}$$

The second equality follows from $(B_{t+s} - B_s : t \geq 0)$ is a standard BM independent of $\mathcal{F}_s$ for any $s \geq 0$. Taking limit as n goes to infinity, we get (3.12) by the continuity of B.

Let us back to reflection principle.

$$\begin{aligned} P(T_b \leq t, B_t \leq a) &= P(T_b \leq t, B_{(t-T_b)+T_b} - B_{T_b} \leq a - b) \\ &= P(T_b \leq t) P(B_{(t-T_b)+T_b} - B_{T_b} \leq a - b) \\ &= P(T_b \leq t) P(B_{(t-T_b)+T_b} - B_{T_b} \geq -(a - b)) \\ &= P(T_b \leq t, B_t \geq 2b - a). \end{aligned}$$

The third equality follows from the symmetry of standard BM.

Chapter 4

Itô's calculus

In this part we develop Itô's integration theory in a traditional way, that is, we first define stochastic integral $\int_0^t F_s dB_s$ for adapted simple processes $(F_t)_{t\geq 0}$, then extend the definition to a large class of integrands by exploiting the martingale characterization of Itô's integrals.

4.1 Introduction

Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}$ on a complete probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_t^0)_{t\geq 0}$ be the filtration generated by $(B_t)_{t\geq 0}$, called the Brownian filtration. That is, for each $t \geq 0$

$$\mathcal{F}_t^0 = \sigma\{B_s \text{ for } s \leq t\}$$

which represents the history of the Brownian motion $B = (B_t)_{t\geq 0}$ up to time t.

We are going to define Itô's integrals of the following form

$$\int_0^t F_s \mathrm{d}B_s \quad \text{for} \quad t \geq 0$$

as a continuous stochastic process, where the integrand $F = (F_t)_{t\geq 0}$ is a stochastic process satisfying certain conditions that will be described later. For example, we would like to define integrals like

$$\int_0^t f(B_s)\mathrm{d}B_s$$

for Borel measurable functions f.

Since, for almost all $\omega \in \Omega$, the sample path of Brownian motion $t \to B_t(\omega)$ is nowhere differentiable, the obvious definition via Riemann sums

$$\sum_i F_{t_i^*}(B_{t_i} - B_{t_{i-1}})$$

does not work: the limit of Riemann sums does not exist. The limit exists however in a probability sense, if for any finite partition we properly choose $t_i^* \in [t_{i-1}, t_i]$ and if the integrand process $(F_t)_{t\geq 0}$ is adapted to the Brownian filtration $(\mathcal{F}_t^0)_{t\geq 0}$. That is to say, for every $t \geq 0$, F_t is measurable with respect to $\mathcal{F}_t^0$. This approach works because both $(B_t)_{t\geq 0}$ and $(B_t^2 - t)_{t\geq 0}$ are continuous martingales.

In summary, Itô's integral $\int_0^t F_s \mathrm{d}B_s$ of *an adapted process* $F = (F_t)_{t\geq 0}$ (such that F is measurable with respect to the product σ-algebra $\mathcal{B}([0,t]) \otimes \mathcal{F}_t^0$ for any $t \geq 0$, a condition you are advised to forget at the first reading) with respect to the Brownian motion $B = (B_t)_{t\geq 0}$ may be simply defined to be the limit of special sort of Riemann sums:

$$\int_0^t F_s \mathrm{d}B_s = \lim_{m(D)\to 0} \sum_i F_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$$

where the limit takes place in L^2-sense (with respect to the product measure $P(\mathrm{d}\omega)\otimes \mathrm{d}t$: you are forgiven to ignore its meaning at this stage), over finite partitions

$$D = \{0 = t_0 < t_1 < \cdots < t_n = t\}$$

of $[0,t]$ so that $m(D) = \max_i (t_i - t_{i-1}) \to 0$. The reason to choose $F_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$ is the following: only with this choice

$$E\left(F_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})\right) = 0 \tag{4.1}$$

and

$$E\left(F_{t_{i-1}}^2(B_{t_i} - B_{t_{i-1}})^2 - F_{t_{i-1}}^2(t_i - t_{i-1})\right) = 0\ . \tag{4.2}$$

It will become clear that, it is these important features that this sort of Riemann sums converge to a martingale! (4.1, 4.2) imply that both the Itô integral

$$\int_0^t F_s \mathrm{d}B_s$$

and

$$\left(\int_0^t F_s \mathrm{d}B_s\right)^2 - \int_0^t F_s^2 \mathrm{d}s$$

are martingales.

Exercise 4.1 *Prove equations (4.1) and (4.2).*

Indeed, since $F_{t_{i-1}}$ is $\mathcal{F}_{t_{i-1}}^0$-measurable, so that

$$\begin{aligned} E\left(F_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})\right) &= E\left\{E\left(F_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}) \middle| \mathcal{F}_{t_{i-1}}^0\right)\right\} \\ &= E\left\{F_{t_{i-1}} E\left((B_{t_i} - B_{t_{i-1}}) \middle| \mathcal{F}_{t_{i-1}}^0\right)\right\} \\ &= E\left(F_{t_{i-1}} E\left(B_{t_i} - B_{t_{i-1}}\right)\right) \\ &= 0\ . \end{aligned}$$

Similarly, since $B_t^2 - t$ is a martingale

$$\begin{aligned}
& E\left((B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\middle|\, \mathcal{F}^0_{t_{i-1}}\right) \\
= \; & E\left(B^2_{t_i} - t_i\middle|\, \mathcal{F}^0_{t_{i-1}}\right) - 2E\left(B_{t_{i-1}}B_{t_i}\middle|\, \mathcal{F}^0_{t_{i-1}}\right) \\
& + B^2_{t_{i-1}} + t_{i-1} \\
= \; & 2B^2_{t_{i-1}} - 2B_{t_{i-1}}E\left(B_{t_i}\middle|\, \mathcal{F}^0_{t_{i-1}}\right) \\
= \; & 0
\end{aligned}$$

and therefore

$$\begin{aligned}
& E\left(F^2_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})^2 - F^2_{t_{i-1}}(t_i - t_{i-1})\right) \\
= \; & E\left\{E\left(F^2_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})^2 - F^2_{t_{i-1}}(t_i - t_{i-1})\middle|\, \mathcal{F}^0_{t_{i-1}}\right)\right\} \\
= \; & E\left\{F^2_{t_{i-1}}E\left((B_{t_i} - B_{t_{i-1}})^2 - F^2_{t_{i-1}}(t_i - t_{i-1})\middle|\, \mathcal{F}^0_{t_{i-1}}\right)\right\} \\
= \; & 0\,.
\end{aligned}$$

We have now so-called Itô isometry

$$E\left(\int_0^t F_s dB_s\right)^2 = E\int_0^t F_s^2 ds \tag{4.3}$$

which means that the mapping $F \mapsto \int_0^t F_s dB_s$ establishes an isometry from the space of such F with norm $||F|| = \sqrt{E\int_0^t F_s^2 ds}$ into the space of square integrable continuous martingale M with norm $||M|| = \sqrt{E(M_t^2)}$. Due to this isometry, we may define Itô integral for simple processes and extend it to the closure of simple processes by completeness of the space of square integrable martingales. However as we can see, the key for Itô isometry is the fact that $(B_t^2 - t)$ is a martingale, or (B_t) has quadratic variational process.

Itô's integration theory can be established for a continuous, square-integrable martingale by the same approach because such process has also quadratic variational process. In fact, if $M = (M_t)_{t\geq 0}$ is a continuous, square-integrable martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and $F = (F_t)_{t\geq 0}$ is an adapted stochastic process, then

$$\int_0^t F_s \mathrm{d}M_s = \lim_{m(D)\to 0} \sum_{l=1}^n F_{t_{l-1}}(M_{t_l} - M_{t_{l-1}})$$

exists under certain integrable conditions.

4.2 Quadratic variational processes

Let $T > 0$ be a fixed but arbitrary number, and $\mathcal{M}_0^2$ be the vector space of all continuous, square-integrable martingales up to time T on a probability space

$(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with *initial value zero*, endowed with the distance

$$d(M,N) = \sqrt{E|M_T - N_T|^2} \quad \text{for } M, N \in \mathcal{M}_0^2 .$$

By definition, a sequence of square-integrable martingales $(M(k)_t)_{t\geq 0}$ $(k = 1, \cdots,)$ converges to M in $\mathcal{M}_0^2$, if and only if

$$M(k)_T \to M_T \quad \text{in} \quad L^2(\Omega, \mathcal{F}, P)$$

as $k \to \infty$. The following maximal inequality, which is the "martingale version" of the Markov inequality, allows us to show that $(\mathcal{M}_0^2, d)$ indeed is complete.

Theorem 4.2.1 *$(\mathcal{M}_0^2, d)$ is a complete metric space.*

Proof. Let $M(k) \in \mathcal{M}_0^2$ ($k = 1, 2, \cdots$) be a Cauchy sequence in $\mathcal{M}_0^2$. Then

$$E|M(k)_T - M(l)_T|^2 \to 0 \ , \quad \text{as } k, l \to \infty \ .$$

According to Kolmogorov's inequality (2.2.10)

$$P\left\{ \sup_{0\leq t\leq T} |M(k)_t - M(l)_t| \geq \lambda \right\} \leq \frac{1}{\lambda^2} E|M(k)_T - M(l)_T|^2 \ ,$$

it follows that, $M(k)$ uniformly converges to a limit M on $[0,T]$ in probability. Therefore there exists a stochastic process $M \equiv (M_t)$ such that

$$\sup_{0\leq t\leq T} |M(k)_t - M_t| \to 0 \quad \text{in probability}$$

and converges almost surely at least along with a subsequence. It then follows that $(M_t)_{t\geq 0}$ is a continuous and square -integrable martingale (up to time T) as the uniform limit of a sequence of continuous martingales. ∎

If $M = (M_t)_{t\geq 0}$ is a continuous, square-integrable martingale, then $(M_t^2)_{t\geq 0}$ is no longer a martingale but a sub-martingale, except for the trivial case. As in the case of Brownian motion, the following limit

$$\langle M \rangle_t = \lim_{m(D)\to 0} \sum_l \left| M_{t_l} - M_{t_{l-1}} \right|^2$$

exists in probability, where the limit takes over all finite partitions D of the interval $[0,t]$. $\{\langle M \rangle_t\}_{t\geq 0}$ is called the (quadratic) variational process of $(M_t)_{t\geq 0}$, or simply *the bracket procees* of $(M_t)_{t\geq 0}$. The quadratic variational process $t \to \langle M \rangle_t$ is an adapted, continuous, increasing stochastic process [and therefore has finite variation] with initial zero. The following theorem demonstrates the importance of $\langle M \rangle_t$. For the proof, see additional topics.

Theorem 4.2.2 *Let $M = (M_t)_{t\geq 0}$ be a continuous, square-integrable martingale. Then $\langle M \rangle_t$ is the unique continuous, adapted and increasing process with initial zero, such that $M_t^2 - \langle M \rangle_t$ is a martingale.*

The uniqueness of $\langle M \rangle$ follows from the following theorem.

Theorem 4.2.3 *M has bounded variation if and only if it is identically M_0.*

Proof. Assume $M_0 = 0$.Let V be the total variation process of M. By localization we may assume that V and M are dominated by a constant K. For any $t \geq 0$ and a partition $D = \{t_i\}$ on $[0, t]$, Since $M_0 = 0$, we have

$$\begin{aligned} E(M_t^2) &= E \sum_i (M_{t_{i+1}}^2 - M_{t_i}^2) \\ &= E \sum_i (M_{t_{i+1}} - M_{t_i})^2 \\ &\leq K \cdot E \left(\sup_i |M_{t_{i+1}} - M_{t_i}| \right). \end{aligned}$$

By continuity of M, as $m(D) \to 0$, $\sup_i |M_{t_{i+1}} - M_{t_i}| \longrightarrow 0$ a.s. Applying the dominated convergence theorem, $E\left(\sup_i |M_{t_{i+1}} - M_{t_i}|\right) \longrightarrow 0$. It follows that $E(M_t^2) = 0$, i.e., M is identically zero. ∎ he process $\langle M \rangle$ is called the *quadratic variational process* associated with the martingale M. The theorem is a special case of the Doob-Meyer decomposition for sub-martingales: any sub-martingale can be decomposed into a sum of a martingale and a predictable, increasing process with initial value zero. The decomposition was conjectured by J.L. Doob, and proved by P. A. Meyer in the 60's, which opened the new era of stochastic calculus.

Remark 4.2.4 *Doob's decomposition in discrete-time case is almost nothing to prove. Consider a sub-martingale in discrete-time: $X = (X_n)_{n \in \mathbb{Z}^+}$ with respect to the filtration $(\mathcal{F}_n)_{n \in \mathbb{Z}^+}$. An increasing sequence $(A_n)_{n \in \mathbb{Z}^+}$ may be defined by*

$$\begin{aligned} A_0 &= 0\,; \\ A_n &= A_{n-1} + E(X_n - X_{n-1} | \mathcal{F}_{n-1}), \quad n = 1, 2, \cdots . \end{aligned}$$

Then

1. *$E(X_n - X_{n-1}|\mathcal{F}_{n-1}) = E(X_n|\mathcal{F}_{n-1}) - X_{n-1} \geq 0$, $(A_n)_{n \in \mathbf{Z}^+}$ is increasing.*
2. *$A_n \in \mathcal{F}_{n-1}$, so that $(A_n)_{n \in \mathbf{Z}^+}$ is* predictable*!*
3. *By definition*

$$E(X_n - A_n | \mathcal{F}_{n-1}) = X_{n-1} - A_{n-1}, \quad n = 1, 2, \cdots ,$$

and therefore $M_n = X_n - A_n$ is a martingale. However its generalization to continuous time case is by no means a trivial task.

Theorem 4.2.5 *Let $(M_t)_{t \geq 0}$ and $(N_t)_{t \geq 0}$ be two continuous, square-integrable martingales, and let*

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

called the bracket process of M and N. Then $\langle M, N\rangle_t$ is the unique adapted, continuous, variational process with initial zero, such that $M_t N_t - \langle M, N\rangle_t$ is a martingale. Moreover

$$\lim_{m(D)\to 0} \sum_{l=1}^{n} (M_{t_l} - M_{t_{l-1}})(N_{t_l} - N_{t_{l-1}}) = \langle M, N\rangle_t \ , \tag{4.4}$$

in probability.

4.3 Stochastic integrals for simple processes

Fix an arbitrary time $T > 0$. An adapted stochastic process $F = (F_t)_{t\geq 0}$ is called a simple process, if it has a representation

$$F_t(\omega) = f(\omega)1_{\{0\}}(t) + \sum_{i=0}^{\infty} f_i(\omega)1_{(t_i, t_{i+1}]}(t) \tag{4.5}$$

where $0 = t_0 < t_1 < \cdots < t_i \to \infty$, such that for any finite time $T \geq 0$, there are only finite many $t_i \in [0, T]$, each $f_i \in \mathcal{F}^0_{t_i}$ (i.e. f_i is measurable with respect to $\mathcal{F}^0_{t_i}$), $f_0 \in \mathcal{F}^0_0$, and F is a bounded process. The space of all simple (adapted) stochastic processes will be denoted by $\mathcal{L}_0$. If $F = (F_t)_{t\geq 0} \in \mathcal{L}_0$, then Itô's integral of F against Brownian motion $B = (B_t)_{t\geq 0}$ is defined as

$$I(F)_t \equiv \sum_{i=0}^{\infty} f_i(B_{t\wedge t_{i+1}} - B_{t\wedge t_i})$$

where the sum makes sense because only finite terms may not vanish. It is obvious that $I(F) = (I(F)_t)_{t\geq 0}$ is continuous, square-integrable, and adapted to $\left(\mathcal{F}^0_t\right)_{t\geq 0}$.

Lemma 4.3.1 *Let $M = (M_t)_{t\geq 0}$ be a continuous, square-integrable martingale, and $s < t \leq u < v$, $f \in \mathcal{F}^0_s$, $g \in \mathcal{F}^0_t$. Then*

$$E\left(g(M_v - M_u)(M_t - M_s)|\mathcal{F}^0_s\right) = 0$$

and

$$E\left(f(M_t - M_s)^2|\mathcal{F}^0_s\right) = E\left(f\left(\langle M\rangle_t - \langle M\rangle_s\right)|\mathcal{F}^0_s\right) \ .$$

Proof. By the tower property of conditional expectations

$$\begin{aligned}
& E\left(g(M_v - M_u)(M_t - M_s)|\mathcal{F}^0_s\right) \\
&= E\left\{E\left(g(M_v - M_u)(M_t - M_s)|\mathcal{F}^0_u\right)|\mathcal{F}^0_s\right\} \\
&= E\left\{g(M_t - M_s)E(M_v - M_u|\mathcal{F}^0_u)|\mathcal{F}^0_s\right\} \\
&= 0 \ .
\end{aligned}$$

The second equality is trivial as $f \in \mathcal{F}^0_s$ that can be moved out from the conditional expectation. ■

Lemma 4.3.2 $(I(F)_t)_{t\geq 0}$ *is a martingale*

$$E\left(I(F)_t - I(F)_s|\mathcal{F}_s^0\right) = 0\ , \quad \forall t > s\ .$$

Proof. Assume that $t_j < t \leq t_{j+1}$, $t_k < s \leq t_{k+1}$ for some $k, j \in \mathbb{N}$. Then $k \leq j$ and

$$\begin{aligned} I(F)_t &= \sum_{i=0}^{j-1} f_i(B_{t_{i+1}} - B_{t_i}) + f_j(B_t - B_{t_j})\ ; \\ I(F)_s &= \sum_{i=0}^{k-1} f_i(B_{t_{i+1}} - B_{t_i}) + f_k(B_s - B_{t_k})\ . \end{aligned}$$

If $k < j-1$, then

$$\begin{aligned} I(F)_t - I(F)_s &= \sum_{i=k+1}^{j-1} f_i(B_{t_{i+1}} - B_{t_i}) \\ &\quad + f_j(B_t - B_{t_j}) + f_k(B_{t_{k+1}} - B_s)\ . \end{aligned} \tag{4.6}$$

If $k+1 \leq i \leq j-1$, then $s \leq t_i$ and $\mathcal{F}_s^0 \subset \mathcal{F}_{t_i}^0$. Hence

$$\begin{aligned} &E\left(f_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_s^0\right) \\ &\quad = E\left\{E(\left\{f_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_{t_i}^0\right\}|\mathcal{F}_s^0\right\} \\ &\quad = E\left\{f_i E\left\{B_{t_{i+1}} - B_{t_i}|\mathcal{F}_{t_i}^0\right\}|\mathcal{F}_s^0\right\} \\ &\quad = 0\ . \end{aligned}$$

The first equality follows from $f_i \in \mathcal{F}_{t_i}^0$, and the second equality follows from that (B_t) is a martingale. Similarly

$$\begin{aligned} &E\left(f_j(B_t - B_{t_j})|\mathcal{F}_s^0\right) = 0, \quad t > t_j \geq s, f_j \in \mathcal{F}_{t_j}^0, \\ &E\left(f_k(B_{t_{k+1}} - B_s)|\mathcal{F}_s^0\right) = 0, \quad t_{k+1} \geq s > t_k, f_k \in \mathcal{F}_{t_k}^0 \subset \mathcal{F}_s^0\ . \end{aligned}$$

Putting these equations together we obtain

$$E\left(I(F)_t - I(F)_s|\mathcal{F}_s^0\right) = 0\ .$$

If $k = j-1$, then $t_{j-1} < s \leq t_j < t \leq t_{j+1}$ and

$$I(F)_t - I(F)_s = f_{j-1}(B_{t_j} - B_s) + f_j(B_t - B_{t_j}).$$

We thus again have

$$E\left(I(F)_t - I(F)_s|\mathcal{F}_s^0\right) = 0\ .$$

■

Lemma 4.3.3 $\left(I(F)_t^2 - \int_0^t F_s^2\, ds\right)_{t\geq 0}$ *is a martingale. Therefore* $I(F) \in \mathcal{M}_0^2$ *and*

$$\langle I(F)\rangle_t = \int_0^t F_s^2\, ds\ .$$

Proof. We want to prove that for any $t \geq s$

$$E\left(I(F)_t^2 - \int_0^t F_u^2 du \middle| \mathcal{F}_s^0 \right) = I(F)_s^2 - \int_0^s F_u^2 du \ .$$

In other words, we have to prove that

$$E\left(I(F)_t^2 - I(F)_s^2 - \int_s^t F_u^2 du \middle| \mathcal{F}_s^0 \right) = 0 \ .$$

Obviously

$$\begin{aligned} I(F)_t^2 - I(F)_s^2 &= (I(F)_t - I(F)_s)^2 - 2I(F)_t I(F)_s - 2I(F)_s^2 \\ &= (I(F)_t - I(F)_s)^2 - 2(I(F)_t - I(F)_s)I(F)_s. \end{aligned}$$

Since $(I(F)_t)_{t\geq 0}$ is a martingale,

$$E\left(I(F)_t - I(F)_s | \mathcal{F}_s^0\right) = 0 \ .$$

But $I(F)_s \in \mathcal{F}_s^0$ so that

$$\begin{aligned} &E\left\{I(F)_s \left(I(F)_t - I(F)_s\right) | \mathcal{F}_s^0\right\} \\ &= I(F)_s E\left\{I(F)_t - I(F)_s | \mathcal{F}_s^0\right\} = 0 \ . \end{aligned}$$

We therefore only need to show

$$E\left\{ (I(F)_t - I(F)_s)^2 - \int_s^t F_u^2 du \middle| \mathcal{F}_s^0 \right\} = 0 \ .$$

Now we use the same notations as in the proof of Lemma 4.3.2. It is clear from (4.6) that if $k < j-1$, then

$$\begin{aligned} (I(F)_t - I(F)_s)^2 &= \sum_{i,l=k+1}^{j-1} f_i f_l (B_{t_{i+1}} - B_{t_i})(B_{t_{l+1}} - B_{t_l}) \\ &+ \sum_{i=1}^{j-1} f_i f_j (B_{t_{i+1}} - B_{t_i})(B_t - B_{t_j}) \\ &+ \sum_{i=1}^{j-1} f_i f_k (B_{t_{i+1}} - B_{t_i})(B_{t_{k+1}} - B_s) \\ &+ f_j^2 (B_t - B_{t_j})^2 + f_k^2 (B_{t_{k+1}} - B_s)^2 \\ &+ f_k f_j (B_t - B_{t_i})(B_{t_{k+1}} - B_s) \ . \end{aligned}$$

Using Lemma 4.3.1 and the fact that both $(B_t)_{t\geq 0}$ and $(B_t^2 - t)_{t\geq 0}$ are martingales, we get

$$\begin{aligned} &E\left\{ (I(F)_t - I(F)_s)^2 \middle| \mathcal{F}_s^0 \right\} \\ &= E\left(\sum_{j=k+1}^{j-1} f_i^2 (t_{i+1} - t_i) + f_j^2 (t - t_j) + f_k^2 (t_{k+1} - s) \middle| \mathcal{F}_s^0 \right) \end{aligned}$$

and then

$$E\left\{(I(F)_t - I(F)_s)^2 | \mathcal{F}_s^0\right\} = E\left(\int_s^t F_u^2 du \middle| \mathcal{F}_s^0 \right) .$$

■

Lemma 4.3.4 *$F \to I(F)$ is linear, and for any $t \geq 0$*

$$E\left(I(F)_t^2\right) = E\left(\int_0^t F_s^2 \, ds\right) .$$

4.4 Stochastic integrals for adapted processes

In this section we extend the definition of Itô's integrals to integrands which are limits of simple processes. Obviously we only need to define Itô's integrals $I(F)_t$ for $t \leq T$ for arbitrary positive number T. Thus, throughout this section, we are given an arbitrary but fixed time $T > 0$.

4.4.1 Stochastic integrals as martingales

If F is a simple process, then the Itô integral $I(F)$ is a continuous, square-integrable martingale with initial zero, and its bracket process $\langle I(F)\rangle_t = \int_0^t F_s^2 \mathrm{d}s$. In particular we have the Itô isometry

$$E|I(F)_t|^2 = E\int_0^t F_s^2 ds, \ \forall \ t \leq T,$$

which allows us to extend the definition of Itô's integrals to a larger class of integrands.

If a process $F = (F_t)_{t\geq 0}$ is a limit of simple processes

$$E\int_0^T |F(n)_t - F_t|^2 dt \to 0$$

as $n \to \infty$ for a sequence of some simple processes $\{F(n) : n \in \mathbb{N}\}$, then we denote by $F \in \mathcal{L}^2$. Then the linearlity of Ito's integral together with Ito's isometry imply that

$$\begin{aligned} d(I(F(n)), I(F(m))) &= E|I(F(n))_T - I(F(m))_T|^2 \\ &= E\int_0^T |F(n)_t - F(m)_t|^2 dt \to 0 \end{aligned}$$

as $n, m \to \infty$, i.e. $\{I(F(n)\}$ is a Cauchy sequence in $(\mathcal{M}_0^2, d)$. Since $(\mathcal{M}_0^2, d)$ is complete, $\lim_{n\to\infty} I(F(n))$ exists in $(\mathcal{M}_0^2, d)$. We naturally define $I(F) = \lim_{n\to\infty} I(F(n))$, which is called Itô's integral of (F_t) against the Brownian motion B. We often write $I(F)_t$ as $\int_0^t F_s dB_s$ or $F.B_t$.

Remark 4.4.1 *1) A process $F = (F_t)_{t \leq T}$ in $\mathcal{L}^2$ is* adapted *and*

$$E \int_0^T F_t^2 \, dt < +\infty \ .$$

2) The map $F \to I(F)$ is a linear isometry from $\mathcal{L}^2$ to $\mathcal{M}_0^2$, where $\mathcal{L}^2$ is endowed with the norm

$$||F|| = \sqrt{E \int_0^T F_t^2 \, dt} \ .$$

$\mathcal{M}_0^2$ is a Hilbert space with norm $||M|| = \sqrt{E(M_T^2)}$.

3) If $F \in \mathcal{L}^2$, then $I(F)$ is a continuous, square-integrable martingale with initial value zero (up to time T), and $\langle I(F) \rangle_t = \int_0^t F_s^2 \, ds$.

The set of integrands $\mathcal{L}^2$ is a very big space which includes many interesting stochastic processes. For example

Lemma 4.4.2 *Let $F = (F_t)_{t \geq 0}$ be an* adapted, left-continuous *stochastic process, satisfying*

$$E \int_0^T F_s^2 \, ds < +\infty \ . \tag{4.7}$$

Then $F \in \mathcal{L}^2$ and

$$I(F)_t = \lim_{m(D) \to 0} \sum_l F_{t_{l-1}} \left(B_{t_l} - B_{t_{l-1}} \right) \quad \text{in probability}$$

where the limit takes over all finite partitions of $[0, t]$.

Proof. At first the set of bounded processes in $\mathcal{L}^2$ is dense in $\mathcal{L}^2$. Hence we may assume that F is bounded in addition. For $n > 0$, let

$$D_n \equiv \{0 = t_0^n < t_1^n < \cdots < t_{n_k}^n = T\}$$

be a sequence of finite partitions of $[0, T]$ such that

$$m(D_n) = \sup_j |t_j^n - t_{j-1}^n| \to 0 \quad \text{as } n \to \infty \ .$$

Let

$$F(n)_t = F_0 1_{\{0\}}(t) + \sum_{l=1}^{n_k} F_{t_{l-1}^n} 1_{(t_{l-1}^n, t_l^n]}(t) \ ; \quad \text{for} \ \ t \geq 0 \ . \tag{4.8}$$

Then each F_n is simple, and, since F is left-continuous $F(n)_t \to F_t$ for each t. Therefore

$$E \int_0^T |F(n)_s - F_s|^2 \mathrm{d}s \to 0 \ \text{ as } n \to \infty \ .$$

By definition, $F \in \mathcal{L}^2$. ∎

Remark 4.4.3 *The condition that $F = (F_t)_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by the Brownian motion, i.e. each F_t is measurable with respect to $\mathcal{F}_t$, is essential in the definition of Itô's integrals. On the other hand, left-continuity of $t \to F_t$ is a technical one, which can be replaced by some sort of Borel measurability (e.g.; right-continuous, continuous, measurable in (t,ω) etc.). Left-continuity becomes a correct condition if one attempts to define stochastic integrals of $F = (F_t)_{t\geq 0}$ against martingales which may have jumps. The reason is that the left-limit of F at time t "happens" before time t, and if $t \to F_t$ is left-continuous, then, for any time t, the value F_t can be "predicted" by the values taking place strictly before time t:*

$$F_t = \lim_{s\uparrow t} F_s \ .$$

Remark 4.4.4 *We should point out that some kind of measurability of random function $(t,\omega) \to F_t(\omega)$ is necessary in order to ensure (4.7) make sense. Note that (4.7) may be written as*

$$\int_\Omega \int_0^T F_s(\omega)^2 ds P(d\omega) < +\infty.$$

Thus the natural measurability condition should be that the function

$$F(s,\omega) \equiv F_s(\omega)$$

is measurable with respect to $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$ for any $t > 0$, where $\mathcal{B}([0,t])$ is the Borel σ-algebra generated by open subsets in $[0,t]$, and $\mathcal{B}([0,t])\otimes\mathcal{F}_t$ is the product σ-algebra on $[0,t]\times\Omega$. This is exactly what we call progressively measurable.

If $X = (X_t)_{t\geq 0}$ is a continuous stochastic process adapted to $(\mathcal{F}_t)_{t\geq 0}$, f is a Borel function, and

$$E\int_0^T f(X_t)^2 \mathrm{d}t < \infty$$

then the stochastic process $(f(X_t))_{t\geq 0}$ belongs to $\mathcal{L}^2$. In particular, for any Borel measurable function f such that

$$E\int_0^T f(B_t)^2 \mathrm{d}t < \infty \tag{4.9}$$

then $(f(B_t))_{t\geq 0}$ is in $\mathcal{L}^2$. What does condition (4.9) mean? While

$$\begin{aligned} E\int_0^T f(B_t)^2 \mathrm{d}t &= \int_0^T E f(B_t)^2 \mathrm{d}t \\ &= \int_0^T P_t(f^2)(0)\mathrm{d}t \end{aligned}$$

where

$$\begin{aligned} P_t(f^2)(0) &= \frac{1}{(2\pi t)^{d/2}} \int_{\mathbb{R}^d} f(x)^2 \mathrm{e}^{-|x|^2/2t} \mathrm{d}x \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(\sqrt{t}x)^2 \mathrm{e}^{-|x|^2/2} \mathrm{d}x \ . \end{aligned}$$

Therefore, if f is a polynomial, then $f(B_t)$ is in $\mathcal{L}^2$, and for any constant α the process $(\mathrm{e}^{\alpha B_t})_{t\geq 0}$ belongs to $\mathcal{L}^2$ as well. How about the stochastic process $F_t = \mathrm{e}^{\alpha B_t^2}$? In this case

$$E\int_0^T F_t^2 \mathrm{d}t = \frac{1}{(2\pi)^{d/2}}\int_0^T \int_{\mathbb{R}^d} \mathrm{e}^{2\alpha t x^2} \mathrm{e}^{-|x|^2/2}\mathrm{d}x$$

and therefore

$$E\int_0^T F_t^2 \mathrm{d}t < \infty \qquad \text{if } \alpha \leq 0 \ .$$

In the case $\alpha > 0$, then

$$E\int_0^T F_t^2 \mathrm{d}t < \infty \ \text{ if and only if } \quad T < \frac{1}{4\alpha} \ .$$

4.4.2 Summary of main properties

If $F = (F_t)_{t\geq 0} \in \mathcal{L}^2$, then both

$$\int_0^t F_s \mathrm{d}B_s \qquad \text{and} \qquad \left(\int_0^t F_s \mathrm{d}B_s\right)^2 - \int_0^t F_s^2 \mathrm{d}s$$

are continuous martingales with initial zero, and therefore $\langle F.B\rangle_t = \int_0^t F_s^2 \mathrm{d}s$. In general (by the use of the polarization $xy = \frac{1}{4}\left((x+y)^2 - (x-y)^2\right)$)

$$\langle F.B, G.B\rangle_t = \int_0^t F_s G_s \mathrm{d}s \qquad \forall F, G \in \mathcal{L}^2 \ .$$

In particular

$$E\left[\int_0^T F_s \mathrm{d}B_s\right]^2 = E\left(\int_0^T F_s^2 \mathrm{d}s\right) \ .$$

and for any $t \geq s$,

$$E\left\{\left(\int_s^t F_u \mathrm{d}B_u\right)^2 \middle| \mathcal{F}_s\right\} = E\left\{\int_s^t F_u^2 \mathrm{d}u \middle| \mathcal{F}_s\right\} \ .$$

4.5 Itô's integration for semi-martingales

We may apply the same procedure of defining Itô's integrals along Brownian motion to any continuous, square-integrable martingales. Indeed, if $M \in \mathcal{M}_0^2$ and if $F = (F_t)_{t\geq 0}$ is a bounded, adapted, simple process

$$F_t = f 1_{\{0\}}(t) + \sum_i f_i 1_{(t_i, t_{i+1}]}(t)$$

then define

$$I^M(F) = \sum_{i=0}^{\infty} f_i \cdot (M_{t\wedge t_{i+1}} - M_{t\wedge t_i}) \ .$$

As before, we have

1. $I^M(F) \in \mathcal{M}_0^2$.
2. The bracket process $\langle I^M(F)\rangle_t = \int_0^t F_s^2 \mathrm{d}\langle M\rangle_s$, i.e. $I^M(F)_t^2 - \int_0^t F_s \mathrm{d}\langle M\rangle_s$ is a martingale.
3. (Itô's isometry) For any $T > 0$, we have

$$E\left(\int_0^T F_t \mathrm{d}M_t\right)^2 = E\int_0^T F_t^2 \mathrm{d}\left\langle M\right\rangle_t \ .$$

Let $T > 0$ be a fixed time.

Definition 4.5.1 *A stochastic process $F = (F_t)_{t\geq 0}$ is said to belong to $\mathcal{L}^2(M)$, if there is a sequence $\{F(n)\}$ of simple stochastic processes such that*

$$E\left\{\int_0^T F(n)_t^2 d\left\langle M\right\rangle_t\right\} < \infty$$

and

$$E\left\{\int_0^T |F(n)_t - F_t|^2 d\left\langle M\right\rangle_t\right\} \to 0 \text{ as } n\to\infty \ .$$

In other words, $\mathcal{L}^2(M)$ is the closure of all simple processes (up to time T) under the norm

$$||F|| = \left\{E\left(\int_0^T F_t^2 \mathrm{d}\langle M\rangle_t\right)\right\}^{\frac{1}{2}}$$

which of course depends on the running time T and the martingale $M \in \mathcal{M}_0^2$, and thus $\mathcal{L}^2(M)$ is a Banach space. Indeed, the above norm is induced by a scalar product, so that $\mathcal{L}^2(M)$ is a Hilbert space. If $F \in \mathcal{L}^2(M)$, and $||F - F(n)|| \to 0$ for a sequence of simple processes, thanks to Itô's isometry

$$\sqrt{E\left\{I^M(F)_T^2\right\}} = ||F|| \ ,$$

it follows that

$$I^M(F) \equiv \lim_{n\to\infty} I^M(F(n)) \ , \qquad \text{in } \mathcal{M}_0^2$$

exists. We use either $F.M$ or $\int_0^t F_s \mathrm{d}M_s$ to denote $I^M(F)$. According to definition, $F.M \in \mathcal{M}_0^2$ and $\langle F.M\rangle_t = \int_0^t F_s^2 \mathrm{d}\langle M\rangle_s$. By the use of the polarization identity, if M, $N \in \mathcal{M}_0^2$ and $F \in \mathcal{L}^2(M)$, $G \in \mathcal{L}^2(N)$, then

$$\left\langle F.M, G.N\right\rangle_t = \int_0^t F_s G_s \mathrm{d}\left\langle M, N\right\rangle_s$$

and $F.(G.M) = (FG).M$, that is,

$$\int_0^t F_s \mathrm{d}\left(\int_0^s G_u \mathrm{d}M_u\right)_s = \int_0^t F_s G_s \mathrm{d}M_s,$$

as far as these stochastic integrals make sense.

4.5.1 Extended to continuous local martingales

The Itô integration may be extended to local martingales. Let us briefly describe the idea. Suppose $M = (M_t)_{t\geq 0}$ is a continuous, local martingale with initial zero. Then we may choose a sequence $\{T_n\}$ of stopping times such that $T_n \uparrow \infty$ a.s. and for each n, $M^{T_n} = (M_{t\wedge T_n})_{t\geq 0}$ is a continuous, square-integrable (or bounded if necessary) martingale with initial zero. In this case we may define

$$\langle M\rangle_t = \langle M^{T_n}\rangle_t \quad \text{if } \; t \leq T_n$$

which is the unique adapted, continuous, increasing process with initial zero such that

$$M_t^2 - \langle M\rangle_t$$

is a continuous local martingale.

Exercise 4.2 *Prove that if T is a stopping time and M a bounded continuous martingale, then*

$$\langle M^T\rangle = \langle M\rangle^T.$$

Therefore $\langle M\rangle$ above is well-defined.

Let $F = (F_t)_{t\geq 0}$ be a left-continuous, adapted process with right limits and define

$$S_n = \inf\{t \geq 0 : |F_t| > n\},$$

which is a sequence of stopping times. It is easily seen that $|F_{S_n\wedge t}| \leq n$, i.e. F^{S_n} is bounded, due to the left continuity of F, and $S_n \uparrow \infty$. Let $\tilde{T}_n = T_n \wedge S_n$. Then $\tilde{T}_n \uparrow \infty$ almost surely, and for each n, $M^{\tilde{T}_n} \in \mathcal{M}_0^2$. Let

$$F(n)_t = F_t 1_{\{t\leq \tilde{T}_n\}} \; .$$

Then

$$\int_0^\infty F(n)_s^2 \mathrm{d}\langle M\rangle_s = \int_0^{\tilde{T}_n} F_s^2 \mathrm{d}\langle M\rangle_s \leq n$$

so that $F(n) \in \mathcal{L}^2(M^{\tilde{T}_n})$. We may define

$$(F.M)_t = \int_0^t F(n)_s \mathrm{d}\left(M^{\tilde{T}_n}\right)_s \qquad \text{if } \; t \leq \tilde{T}_n \uparrow \infty$$

for $n = 1, 2, 3, \cdots$, called the Itô integral of F with respect to local martingale M. It can be shown that $F.M$ does not depend on the choice of stopping times T_n. By definition, both $F.M$ and

$$(F.M)_t^2 - \int_0^t F_s^2 \mathrm{d}\langle M \rangle_s$$

are continuous, local martingales with initial zero.

Exercise 4.3 *A left continuous adapted process (with right limits) F has a localizing sequence $\{S_n\}$ such that F^{S_n} is bounded for each n. Does a right continuous adapted process have the same property?*

4.5.2 Extended to continuous semi-martingales

Finally let us extend the theory of stochastic integrals to the most useful class of (continuous) semimartingales. An adapted, continuous stochastic process $X = (X_t)_{t \geq 0}$ is a (continuous) semimartingale if X possesses a decomposition

$$X_t = M_t + V_t$$

where $(M_t)_{t \geq 0}$ is a continuous local martingale, and $(V_t)_{t \geq 0}$ is a continuous processes with finite variation on any finite interval. Such decomposition is unique by Theorem 4.2.3.

If $f(t)$ is a function on $[0, T]$ *having finite variation*:

$$\sup_D \sum_l |f(t_l) - f(t_{l-1})| < +\infty$$

where D runs over all finite partitions of $[0, t]$ (for any fixed t), then for Borel function g on $[0, \infty)$,

$$\int_0^t g(s) \mathrm{d}f(s)$$

is understood as the Lebesgue-Stieltjes integral. If in addition $s \to f(s)$ is continuous and g has bounded variation, then

$$\int_0^t g(s) \mathrm{d}f(s) = \lim_{m(D) \to 0} \sum_l g(t_{l-1})(f(t_l) - f(t_{l-1})) \ .$$

Therefore, if $V = (V_t)_{t \geq 0}$ is a continuous stochastic process with finite variation, then

$$\int_0^t F_s \mathrm{d}V_s$$

is a stochastic process defined path-wisely as the Lebesgue-Stieltjes integral

$$\begin{aligned} \left(\int_0^t F_s \mathrm{d}V_s \right)(\omega) &\equiv \int_0^t F_s(\omega) \mathrm{d}V_s(\omega) \\ &= \lim_{m(D) \to 0} \sum_l F_{t_{l-1}}(\omega)(V_{t_l}(\omega) - V_{t_{l-1}}(\omega)) \ . \end{aligned}$$

If $(F_t : t \geq 0)$ is a left-continuous adapted process, then the definition of stochastic integrals may be extended to any continuous *semi-martingale* in an obvious way, namely

$$\int_0^t F_s \mathrm{d}X_s = \int_0^t F_s \mathrm{d}M_s + \int_0^t F_s \mathrm{d}V_s$$

where, the first term on the right-hand side is the Itô's integral with respect to local martingale M defined in probability sense, which is again a local martingale, the second term is the usual Lebesgue-Stieltjes integral which is defined path-wisely. Moreover

$$\int_0^t F_s \mathrm{d}X_s = \lim_{m(D)\to 0} \sum_l F_{t_{l-1}} \left(X_{t_l} - X_{t_{l-1}} \right),$$

in probability.

Exercise 4.4 *Using localization to prove the convergence in probability above.*

Notice that the convergence in Itô integrals is not almost surely. This means that stochastic integration is not defined path-wisely.

4.6 Ito's formula

Ito's formula was established by K. Itô in 1944. Since Itô's stated it as a lemma in his seminal paper [ ], Itô's formula is also referred in literature as Itô's Lemma. Itô's Lemma is indeed the Fundamental Theorem in stochastic calculus.

We have used in many occasions the following elementary formula

$$X_{t_j}^2 - X_{t_{j-1}}^2 = \left(X_{t_j} - X_{t_{j-1}} \right)^2 + 2X_{t_{j-1}} \left(X_{t_j} - X_{t_{j-1}} \right) .$$

If in addition $(X_t)_{t\geq 0}$ is a continuous square-integrable martingale, then, by adding up the above identity over $j = 1, \cdots, n$, where $0 = t_0 < t_1 < \cdots < t_n = t$ is an arbitary finite partition, one obtains

$$X_t^2 - X_0^2 = 2\sum_{j=1}^n X_{t_{j-1}} \left(X_{t_j} - X_{t_{j-1}} \right) + \sum_{j=1}^n \left(X_{t_j} - X_{t_{j-1}} \right)^2 .$$

Letting $m(D) \to 0$, we obtain

$$X_t^2 - X_0^2 = 2\int_0^t X_s dX_s + \langle X \rangle_t .$$

which is the Itô formula for the martingale $(X_t)_{t\geq 0}$ applying to $f(x) = x^2$. By using polarization and localization, we establish the following integration by parts formula.

Lemma 4.6.1 (Integration by parts) *Let $X = M + A$ and $Y = N + B$ be continuous semi-martingales: M and N are continuous local martingales, and A, B are continuous, adapted processes with finite variations. Then*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle M, N \rangle_t \ .$$

The following is the fundamental theorem in stochastic calculus.

Theorem 4.6.2 (Itô's formula) *Let $X = (X_t^1, \cdots, X_t^d)$ be a continuous semi-martingale in $\mathbb{R}^d$ with decompositions $X_t^i = M_t^i + A_t^i$: $M_t^1, \cdots, M_t^d$ are continuous local martingales, and $A_t^1, \cdots, A_t^d$ are continuous, locally integrable, adapted processes with finite variations. Let $f \in C^2(\mathbb{R}^d, \mathbb{R})$. Then*

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d\langle M^i, M^j \rangle_s \ . \end{aligned} \tag{4.10}$$

The first term on the right-hand side of (4.10) can be splited into

$$\sum_{j=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dM_s^j + \sum_{j=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dA_s^j$$

so that $f(X_t) - f(X_0)$ is again a semi-martingale with its martingale part given by

$$M_t^f = \sum_{j=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dM_s^j \ .$$

It follows that

$$\langle M^f, M^g \rangle_t = \int_0^t \sum_{i,j=1}^d \frac{\partial f}{\partial x_i}(X_s) \frac{\partial g}{\partial x_j}(X_s) \mathrm{d}\langle M^i, M^j \rangle_s \ .$$

4.6.1 Itô's formula for BM

If $B = (B_t^1, \cdots, B_t^d)_{t \geq 0}$ is Brownian motion in $\mathbb{R}^d$, then, for $f \in C^2(\mathbb{R}^d, \mathbb{R})$

$$f(B_t) - f(B_0) = \int_0^t \nabla f(B_s).\mathrm{d}B_s + \int_0^t \frac{1}{2} \Delta f(B_s) \mathrm{d}s \ .$$

Let

$$M_t^f = f(B_t) - f(B_0) - \int_0^t \frac{1}{2} \Delta f(B_s) \mathrm{d}s \ .$$

Then M^f is a local martingale and

$$\langle M^f, M^g \rangle_t = \int_0^t \langle \nabla f, \nabla g \rangle (B_s) \mathrm{d}s \ .$$

4.6.2 Proof of Itô's formula.

Let us prove the Itô formula for one-dimensional case. By using localization technique, we only need to prove it for a continuous, square-integrable martingale $M = (M_t)_{t\geq 0}$. Thus we need to show

$$f(M_t) - f(M_0) = \int_0^t f'(M_s)\mathrm{d}M_s + \frac{1}{2}\int_0^t f''(M_s)\mathrm{d}\langle M\rangle_s \ . \tag{4.11}$$

The formula is true for $f(x) = x^2$ ($f'(x) = 2x$ and $f''(x) = 2$) as we have seen

$$M_t^2 - M_0^2 = 2\int_0^t M_s\mathrm{d}M_s + \langle M\rangle_t \ .$$

Suppose (4.11) is true for $f(x) = x^n$:

$$M_t^n - M_0^n = n\int_0^t M_s^{n-1}\mathrm{d}M_s + \frac{n(n-1)}{2}\int_0^t M_s^{n-2}\mathrm{d}\langle M\rangle_s \ ,$$

by applying integration by parts formula to M^n and M, one obtains

$$\begin{aligned} M_t^{n+1} - M_0^{n+1} &= \int_0^t M_s^n\mathrm{d}M_s + \int_0^t M_s\mathrm{d}M_s^n + \langle M, M^n\rangle_t \\ &= \int_0^t M_s^n\mathrm{d}M_s + \int_0^t M_s\mathrm{d}\left\{nM_s^{n-1}\mathrm{d}M_s + \frac{n(n-1)}{2}M_s^{n-2}\mathrm{d}\langle M\rangle_s\right\} \\ &\quad + \int_0^t nM_s^{n-1}\mathrm{d}\langle M\rangle_s \\ &= (n+1)\int_0^t M_s^n\mathrm{d}M_s + \frac{(n+1)n}{2}\int_0^t M_s^{n-1}\mathrm{d}\langle M\rangle_s \end{aligned}$$

which implies that (4.11) for power function x^{n+1}. Itô's formula holds thus for any polynomial, so does it for any C^2 function f due to Taylor's expansions.

4.7 Selected applications of Itô's formula

In this section, we present several applications of Itô's lemma.

4.7.1 Lévy's characterization of Brownian motion

Our first application is Lévy's martingale characterization of Brownian motion. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space satisfying the usual condition.

Theorem 4.7.1 (Lévy) *Let $M_t = (M_t^1, \cdots, M_t^d)$ be an adapted, continuous stochastic process on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ taking values in $\mathbb{R}^d$ with initial zero. Then $(M_t)_{t\geq 0}$ is a Brownian motion if and only if*

1) Each M_t^i is a continuous square-integrable martingale.

2) $M_t^i M_t^j - \delta_{ij}t$ is a martingale, that is, $\left\langle M^i, M^j\right\rangle_t = \delta_{ij}t$ for every pair (i, j).

Proof. We need only to prove the sufficiency part. Recall that, under the assumption, $(M_t)_{t\geq 0}$ is a Brownian motion if and only if

$$E\left(e^{\sqrt{-1}\langle \xi, M_t - M_s\rangle}\Big| \mathcal{F}_s\right) = \exp\left\{-\frac{|\xi|^2}{2}(t-s)\right\} \tag{4.12}$$

for any $t > s$ and $\xi = (\xi_i) \in \mathbb{R}^d$. We thus consider the adapted process

$$Z_t = \exp\left(\sqrt{-1}\sum_{i=1}^d \xi_i M_t^i + \frac{|\xi|^2}{2}t\right)$$

and we show it is a martingale. To this end, we apply Itô's formula to $f(x) = e^x$ (in this case $f' = f'' = f$) and semi-martingale

$$X_t = \sqrt{-1}\sum_{i=1}^d \xi_i M_t^i + \frac{|\xi|^2}{2}t \ ,$$

and obtain

$$\begin{aligned}
Z_t &= Z_0 + \int_0^t Z_s \mathrm{d}\left(\sqrt{-1}\sum_{i=1}^d \xi_i M_s^i + \frac{|\xi|^2}{2}s\right) \\
&\quad + \frac{1}{2}\int_0^t Z_s \mathrm{d}\langle \sqrt{-1}\sum_{i=1}^d \xi_i M^i\rangle_s \\
&= 1 + \sqrt{-1}\sum_{i=1}^d \xi_i \int_0^t Z_s \mathrm{d}M_s^i + \frac{|\xi|^2}{2}\int_0^t Z_s \mathrm{d}s \\
&\quad - \frac{1}{2}\int_0^t \sum_{i,j=1}^d \xi_i \xi_j Z_s \mathrm{d}\langle M^i, M^j\rangle_s \\
&= 1 + \sqrt{-1}\sum_{i=1}^d \xi_i \int_0^t Z_s \mathrm{d}M_s^i
\end{aligned}$$

where the last equality follows from

$$\frac{1}{2}\int_0^t \sum_{i,j=1}^d \xi_i \xi_j Z_s \mathrm{d}\langle M^i, M^j\rangle_s = \frac{1}{2}|\xi|^2 \int_0^t Z_s \mathrm{d}s \ .$$

due to the assumption that $\langle M^i, M^j\rangle_s = \delta_{ij}s$. Since $|Z_s| = e^{|\xi|^2 s/2}$, it is seen that for any $T > 0$

$$E\int_0^T |Z_s|^2 \mathrm{d}s = \int_0^T e^{|\xi|^2 s}\mathrm{d}s < +\infty$$

and therefore $(Z_t) \in \mathcal{L}^2(M^i)$ for $i = 1, \cdots, d$ as $\langle M^i\rangle_t = t$. It follows that

$$\int_0^t Z_s \mathrm{d}M_s^i \in \mathcal{M}_2^c \ .$$

That is to say, Z_s is a continuous, square-integrable martingale with initial value 1. (4.12) follows from the martingale property

$$E\left(e^{\sqrt{-1}\langle\xi,M_t\rangle+\frac{|\xi|^2}{2}t}\middle|\mathcal{F}_s\right)=e^{\sqrt{-1}\langle\xi,M_s\rangle+\frac{|\xi|^2}{2}s}$$

for $t>s$. ■

4.7.2 Time-changes of Brownian motion

A continuous local martingale is actually a time change of Brownian motion.

Theorem 4.7.2 (Dambis, Dubins, Schwarz) *Let $M=(M_t)_{t\geq 0}$ be a continuous, local martingale on $(\Omega,\mathcal{F},\mathcal{F}_t,P)$ with initial value zero satisfying $\langle M\rangle_\infty=\infty$, and let*

$$T_t=\inf\{s:\langle M\rangle_s>t\}\ .$$

Then T_t is a stopping time for each $t\geq 0$, $B_t=M_{T_t}$ is an $(\mathcal{F}_{T_t})$-Brownian motion, and $M_t=B_{\langle M\rangle_t}$.

Proof. The family $T=(T_t)_{t\geq 0}$ is called a time-change, because each T_t is a stopping time, and obviously $t\to T_t$ is increasing. Each T_t is finite $\mathbb{P}$-a.e. because $\langle M\rangle_\infty=\infty$ P-a.e. By continuity of $\langle M\rangle_t$

$$\langle M\rangle_{T_t}=t\quad P\text{-a.s.}$$

Applying Doob's optional sampling theorem for the square integrable martingale $(M_{s\wedge T_t})_{s\geq 0}$ and stopping times $T_t\geq T_s$ $(t\geq s)$, we obtain that

$$E\left(M_{T_t}|\mathcal{F}_{T_s}\right)=M_{T_s}$$

i.e. B_t is a $(\mathcal{F}_{T_t})$-local martingale. By the same argument but to the martingale $(M^2_{s\wedge T_t}-\langle M\rangle_{s\wedge T_t})_{s\geq 0}$ we have

$$E\left(M^2_{T_t}-\langle M\rangle_{T_t}|\mathcal{F}_{T_s}\right)=M^2_{T_s}-\langle M\rangle_{T_s}\ .$$

Hence (B_t^2-t) is an $(\mathcal{F}_{T_t})$-local martingale. We may verify that $t\to B_t$ is continuous, and then $B=(B_t)_{t\geq 0}$ is an $(\mathcal{F}_{T_t})$ Brownian motion. ■

Exercise 4.5 *Let f be a continuous increasing function on $[0,\infty)$ with $f(0)=0$ and $f(\infty)=\infty$. Define*

$$f^{-1}(x)=\inf\{y:f(y)>x\}.$$

Prove that $f(f^{-1}(x))=x$ for all $x\geq 0$.

4.7.3 Stochastic exponentials

In this section we consider a simple stochastic differential equation

$$dZ_t = Z_t dX_t \ , \quad Z_0 = 1 \tag{4.13}$$

where $X_t = M_t + A_t$ is a continuous semi-martingale. The solution of (4.13) is called the *stochastic exponential* of X. The equation (4.13) should be understood as an integral equation

$$Z_t = 1 + \int_0^t Z_s dX_s \tag{4.14}$$

where the integral is taken as Itô's integral. To find the solution to (4.14) we may try

$$Z_t = \exp(X_t + V_t)$$

where $(V_t)_{t\geq 0}$ is determined as a "correction" term (which has finite variation) due to the Itô's integration. Applying Itô's formula we obtain

$$Z_t = 1 + \int_0^t Z_s \mathrm{d}(X_s + V_s) + \frac{1}{2}\int_0^t Z_s \mathrm{d}\langle M\rangle_s$$

and therefore, in order to match the equation (4.14) we must choose $V_t = -\frac{1}{2}\langle M\rangle_t$.

Lemma 4.7.3 *Let $X_t = M_t + A_t$ (where M is a continuous local martingale, and A an adapted continuous process with finite total variation) with $X_0 = 0$. Then*

$$\mathcal{E}(X)_t = \exp\left(X_t - \frac{1}{2}\langle M\rangle_t\right)$$

is the solution to (4.14).

$\mathcal{E}(X)$ is called the *stochastic exponential* of $X = (X_t)_{t\geq 0}$.

Proposition 4.7.4 *Let $(M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$. Then the stochastic exponential $\mathcal{E}(M)$ is a continuous, non-negative local martingale.*

Remark 4.7.5 *According to definition of Itô's integrals, if $T > 0$ such that*

$$E\int_0^T e^{2M_t - \langle M\rangle_t} d\langle M\rangle_t < +\infty \tag{4.15}$$

then the stochastic exponential

$$\mathcal{E}(M)_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right)$$

is a non-negative, continuous martingale.

The remarkable fact is that, although $\mathcal{E}(M)$ may fail to be a martingale, it is nevertheless a super-martingale.

Lemma 4.7.6 *Let $X = (X_t)_{t \geq 0}$ be a* ***non-negative****, continuous local martingale. Then $X = (X_t)_{t \geq 0}$ is a super-martingale: $E(X_t|\mathcal{F}_s) \leq X_s$ for any $t < s$. In particular $t \to EX_t$ is decreasing, and therefore $EX_t \leq EX_0$ for any $t > 0$.*

Proof. Recall Fatou's lemma: if $\{f_n\}$ is a sequence of non-negative, integrable functions on a probability space $(\Omega, \mathcal{F}, P)$, such that

$$\liminf_{n \to \infty} E(f_n) < +\infty \ ,$$

then $\liminf_{n \to \infty} f_n$ is integrable and

$$E\left(\liminf_{n \to \infty} f_n | \mathcal{G}\right) \leq \liminf_{n \to \infty} E\left(f_n | \mathcal{G}\right)$$

for any sub σ-algebra $\mathcal{G}$ (as an exercise).

By definition, there is a sequence of finite stopping times $T_n \uparrow +\infty$ P-a.e. such that $X^{T_n} = (X_{t \wedge T_n})_{t \geq 0}$ is a martingale for each n. Hence

$$E\left(X_{t \wedge T_n} | \mathcal{F}_s\right) = X_{s \wedge T_n}, \quad \forall t \geq s, n = 1, 2, \cdots \ .$$

In particular

$$E\left(X_{t \wedge T_n}\right) = EX_0.$$

By Fatou's lemma, $X_t = \lim_{n \to \infty} X_{t \wedge T_n}$ is integrable. Applying Fatou's lemma to $X_{t \wedge T_n}$ and $\mathcal{G} = \mathcal{F}_s$ for $t > s$ we have

$$\begin{aligned} E(X_t|\mathcal{F}_s) &= E\left(\lim_{n \to \infty} X_{t \wedge T_n} | \mathcal{F}_s\right) \\ &\leq \liminf_{n \to \infty} E(X_{t \wedge T_n}|\mathcal{F}_s) \\ &= \liminf_{n \to \infty} X_{s \wedge T_n} \\ &= X_s \end{aligned}$$

According to definition, $X = (X_t)_{t \geq 0}$ is a super-martingale. ■

Corollary 4.7.7 *Let $M = (M_t)_{t \geq 0}$ be a continuous, local martingale with $M_0 = 0$. Then $\mathcal{E}(M)$ is a super-martingale. In particular,*

$$E \exp\left(M_t - \frac{1}{2}\langle M \rangle_t\right) \leq 1 \quad \textit{for all} \quad t \geq 0 \ .$$

Clearly, a continuous super-martingale $X = (X_t)_{t \geq 0}$ is a martingale if and only if its expectation $t \to E(X_t)$ is constant. Therefore

Corollary 4.7.8 *Let $M = (M_t)_{t \geq 0}$ be a continuous, local martingale with $M_0 = 0$. Then $\mathcal{E}(M)$ is a martingale up to time T, if and only if*

$$E \exp\left(M_T - \frac{1}{2}\langle M \rangle_T\right) = 1 \ . \tag{4.16}$$

Stochastic exponentials of local martingales play an important rôle in probability transformations. It is vital in many applications to know whether the stochastic exponential of a given martingale $M = (M_t)_{t\geq 0}$ is indeed a martingale. A sufficient condition to ensure (4.16) is the so-called Novikov's condition stated in below (see [13]).

Theorem 4.7.9 (Novikov) *Let $M = (M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$. If*

$$E\exp\left(\frac{1}{2}\langle M\rangle_T\right) < +\infty\ , \tag{4.17}$$

then $\mathcal{E}(M)$ is a martingale up to time T.

Proof. The following proof is due to J.A. Yan [23]. The idea is the following. Set

$$\mathcal{E}(\alpha M)_t \equiv \exp\left(\alpha M_t - \frac{1}{2}\alpha^2\langle M\rangle_t\right).$$

First show that, under the Novikov condition (4.17), for any $0<\alpha<1$

$$\{\mathcal{E}(\alpha M)_\tau :\ \tau \text{ is a stopping time and } \tau \leq T\}$$

is uniformly integrable and hence $\mathcal{E}(\alpha M)$ is a martingale.

Indeed, for any α, $\mathcal{E}(\alpha M)$ is the stochastic exponential of the local martingale αM, and $\mathcal{E}(\alpha M)$ is a non-negative, continuous local martingale, $E\left(\mathcal{E}(\alpha M)_t\right) \leq 1$. We also have the following scaling property

$$\begin{aligned}\mathcal{E}(\alpha M)_t &\equiv \exp\left\{\alpha\left(M_t - \frac{1}{2}\langle M\rangle_t\right) - \frac{1}{2}\alpha(\alpha-1)\langle M\rangle_t\right\}\\ &= \left(\mathcal{E}(M)_t\right)^\alpha \exp\left\{\frac{1}{2}\alpha(1-\alpha)\langle M\rangle_t\right\}\ .\end{aligned}$$

For any finite stopping time $\tau\leq T$ and for any $A\in\mathcal{F}_T$

$$E\left(1_A\mathcal{E}(\alpha M)_\tau\right) = E\left\{1_A\left(\mathcal{E}(M)_\tau\right)^\alpha\exp\left[\frac{1}{2}\alpha(1-\alpha)\langle M\rangle_\tau\right]\right\}\ . \tag{4.18}$$

Using Hölder's inequality with $\frac{1}{\alpha}>1$ and $\frac{1}{1-\alpha}$ in (4.18) one obtains

$$\begin{aligned}E\left\{1_A\mathcal{E}(\alpha M)_\tau\right\} &= E\left\{\left(\mathcal{E}(M)_\tau\right)^\alpha 1_A\exp\left[\frac{1}{2}\alpha(1-\alpha)\langle M\rangle_\tau\right]\right\}\\ &\leq \left\{E\left(\mathcal{E}(M)_\tau\right)\right\}^\alpha\left\{E\left[1_A\exp\left(\frac{1}{2}\alpha\langle M\rangle_\tau\right)\right]\right\}^{1-\alpha}\\ &\leq \left\{E\left[1_A\exp\left(\frac{1}{2}\alpha\langle M\rangle_T\right)\right]\right\}^{1-\alpha}\\ &\leq \left\{E\left[1_A\exp\left(\frac{1}{2}\langle M\rangle_T\right)\right]\right\}^{1-\alpha}\ .\end{aligned} \tag{4.19}$$

Hence it is easy to verify by Theorem 1.4.1 that

$$\{\mathcal{E}(\alpha M)_\tau : \text{any stopping times } \tau \le T\}$$

is uniformly integrable, and it follows that $\mathcal{E}(\alpha M)$ must be a martingale on $[0,T]$. Therefore

$$E\left(\mathcal{E}(\alpha M)_T\right) = E\left(\mathcal{E}(\alpha M)_0\right) = 1, \quad \forall \alpha \in (0,1).$$

Set $A = \Omega$ and $S = t \le T$ in (4.19), the first inequality of (4.19) becomes

$$\begin{aligned} 1 &= E\left(\mathcal{E}(\alpha M)_t\right) \\ &\le \left(E\left(\mathcal{E}(M)_t\right)\right)^\alpha \left\{E\left(\exp\left(\frac{1}{2}\langle M\rangle_T\right)\right)\right\}^{1-\alpha} \end{aligned}$$

for every $\alpha \in (0,1)$. Letting $\alpha \uparrow 1$ we thus obtain

$$E\left(\mathcal{E}(M)_t\right) \ge 1$$

namely $E\left(\mathcal{E}(M)_t\right) = 1$ for any $t \le T$. It follows thus that $\mathcal{E}(M)_t$ is a martingale up to T. ■

Consider a standard Brownian motion $B = (B_t)$, and $F = (F_t)_{t\ge 0} \in \mathcal{L}_2$. If

$$E\exp\left[\frac{1}{2}\int_0^T F_t^2 \mathrm{d}t\right] < \infty$$

then

$$X_t = \exp\left\{\int_0^t F_s \mathrm{d}B_s - \frac{1}{2}\int_0^t F_s^2 \mathrm{d}s\right\} \tag{4.20}$$

is a positive martingale on $[0,T]$. For example, for any bounded process $F = (F_t)_{t\ge 0} \in \mathcal{L}_2$: $|F_t(\omega)| \le C$ (for all $t \le T$ and $\omega \in \Omega$), where C is a constant, then

$$E\left\{\exp\left(\frac{1}{2}\int_0^T F_t^2 \mathrm{d}t\right)\right\} \le \exp\left(\frac{1}{2}C^2 T\right) < \infty$$

so that, in this case, $X = (X_t)$ defined by (4.20) is a martingale up to time T.

Novikov's condition is very nice, it is however not easy to verify in many interesting cases. For example, to see if the stochastic exponential of the martingale $\int_0^t B_s \mathrm{d}B_s$ is a martingale, the Novikov condition requires to estimate the integral

$$E\left\{\exp\left[\frac{1}{2}\int_0^T B_t^2 \mathrm{d}t\right]\right\}$$

which is already not an easy task.

4.7.4 Exponential inequality

We are going to present three significant applications of stochastic exponentials: a sharp improvement of Doob's maximal inequality for martingales, Girsanov's theorem, and the martingale representation theorem (in the next section). Additional applications will be discussed in the next chapter.

Recall that, according to Doob's maximal inequality, if $(X_t)_{t\geq 0}$ is a continuous super-martingale on $[0,T]$, then for any $\lambda > 0$

$$P\left\{\sup_{t\in[0,T]}|X_t| \geq \lambda\right\} \leq \frac{1}{\lambda}\left(E(X_0)+2E(X_T^-)\right).$$

In particular, if $(X_t)_{t\geq 0}$ is a *non-negative*, continuous super-martingale on $[0,T]$, then

$$P\left\{\sup_{t\in[0,T]}X_t \geq \lambda\right\} \leq \frac{1}{\lambda}E(X_0)\ . \tag{4.21}$$

This inequality has a significant improvement stated as follows.

Theorem 4.7.10 *Let $M=(M_t)_{t\geq 0}$ be a continuous square-integrable martingale with $M_0=0$. Suppose there is a (deterministic) continuous, increasing function $a=a(t)$ such that $a(0)=0$, $\langle M\rangle_t \leq a(t)$ for all $t\in[0,T]$. Then*

$$P\left\{\sup_{t\in[0,T]}M_t \geq \lambda a(T)\right\} \leq e^{-\frac{\lambda^2}{2}a(T)}\ . \tag{4.22}$$

Proof. For every $\alpha>0$ and $t\leq T$

$$\begin{aligned}\alpha M_t-\frac{\alpha^2}{2}\langle M\rangle_t &\geq \alpha M_t-\frac{\alpha^2}{2}\langle M\rangle_T\\ &\geq \alpha M_t-\frac{\alpha^2}{2}a(T)\end{aligned}$$

and hence

$$\mathcal{E}(\alpha M)_t \geq e^{\alpha M_t-\frac{\alpha^2}{2}a(T)} \qquad \text{for } \alpha>0\ .$$

Applying Doob's maximal inequality to the non-negative super-martingale $\mathcal{E}(\alpha M)$ we obtain

$$\begin{aligned}P\left\{\sup_{t\in[0,T]}M_t \geq \lambda a(T)\right\} &\leq P\left\{\sup_{t\in[0,T]}\mathcal{E}(\alpha M)_t \geq e^{\alpha\lambda a(T)-\frac{\alpha^2}{2}a(T)}\right\}\\ &\leq e^{-\alpha\lambda a(T)+\frac{\alpha^2}{2}a(T)}E\left\{\mathcal{E}(\alpha M)_0\right\}\\ &= e^{-\alpha\lambda a(T)+\frac{\alpha^2}{2}a(T)}\end{aligned}$$

for any $\alpha>0$. The exponential inequality follows by setting $\alpha=\lambda$. ■

In particular, by applying the exponential inequality to a standard Brownian motion $B=(B_t)_{t\geq 0}$,

$$P\left\{\sup_{t\in[0,T]}B_t \geq \lambda T\right\} \leq e^{-\frac{\lambda^2}{2}T}\ . \tag{4.23}$$

4.7.5 Girsanov's theorem

We are given a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Let $T > 0$, and Q be a probability measure on $(\Omega, \mathcal{F}_T)$ such that

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_T} = \xi$$

for some non-negative random variable $\xi \in L^1(\Omega, \mathcal{F}_T, P)$. By definition, for any bounded $\mathcal{F}_T$-measurable random variable X

$$\int_\Omega X(\omega) Q(d\omega) = \int_\Omega X(\omega)\xi(\omega) P(d\omega)$$

or simply written as

$$E^Q(X) = E^P(\xi X) .$$

If, however, X is $\mathcal{F}_t$-measurable, $t \leq T$, then

$$\begin{aligned} E^Q(X) &= E^P(E^P(\xi X | \mathcal{F}_t)) \\ &= E^P(E^P(\xi | \mathcal{F}_t) X) . \end{aligned}$$

That is, for every $t \leq T$

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = E^P(\xi | \mathcal{F}_t)$$

which is a non-negative martingale up to T under the probability P.

Conversely, if $T > 0$ and $Z = (Z_t)_{t \geq 0}$ is a continuous, positive martingale up to time T, with $Z_0 = 1$, on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. We define a measure Q on $(\Omega, \mathcal{F}_T)$ by

$$Q(A) = P(Z_T A) \quad \text{if } A \in \mathcal{F}_T \quad . \tag{4.24}$$

That is, $\left.\frac{dQ}{dP}\right|_{\mathcal{F}_T} = Z_T$. Q is a probability measure on $(\Omega, \mathcal{F}_T)$ as $E(Z_T) = 1$. Since $(Z_t)_{t \leq T}$ is a martingale up to time T, $\left.\frac{dQ}{dP}\right|_{\mathcal{F}t} = Z_t$ for all $t \leq T$.

Exercise 4.6 *An adapted process X is a Q-martingale if and only if XZ is a P-martingale. It is a Q-local martingale if XZ is a P-local martingale. The density Z is strictly positive Q-a.s.*

If $(Z_t)_{t \geq 0}$ is a positive martingale with $Z_0 = 1$, then there is a probability measure Q on $(\Omega, \mathcal{F}_\infty)$, where $\mathcal{F}_\infty \equiv \sigma\{\mathcal{F}_t : t \geq 0\}$, such that $\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = Z_t$ for all $t \geq 0$. We are now in a position to prove Girsanov's theorem.

Theorem 4.7.11 (Girsanov) *Let $(M_t)_{t \geq 0}$ be a continuous local martingale and $Z = (Z_t : t \geq 0)$ a continuous positive martingale with $Z_0 = 1$ on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ up to time T. Then*

$$X_t = M_t - \int_0^t \frac{1}{Z_s} d\langle M, Z\rangle_s$$

is a continuous local martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, Q)$ up to time T.

Proof. Using localization technique, we may assume that $M, Z, 1/Z$ are all bounded. In this case M, Z are bounded martingales. We want to prove that X is a martingale under the probability Q:

$$Q\{X_t|\mathcal{F}_s\} = X_s \quad \text{for all} \quad s < t \leq T ,$$

that is,

$$Q\{1_A(X_t - X_s)\} = 0 \quad \text{for all} \quad s < t \leq T , \; A \in \mathcal{F}_s .$$

By definition

$$Q\{1_A(X_t - X_s)\} = P\{(Z_t X_t - Z_s X_s)1_A\}$$

thus we only need to show that $(Z_t X_t)$ is a martingale up to time T under probability measure P. By use of integration by parts, we have

$$\begin{aligned} Z_t X_t &= Z_0 X_0 + \int_0^t Z_s \mathrm{d}X_s + \int_0^t X_s \mathrm{d}Z_s + \langle Z, X\rangle_t \\ &= Z_0 X_0 + \int_0^t Z_s \left(\mathrm{d}M_s - \frac{1}{Z_s} \mathrm{d}\langle M, Z\rangle_s \right) \\ &\quad + \int_0^t X_s \mathrm{d}Z_s + \langle Z, X\rangle_t \\ &= Z_0 X_0 + \int_0^t Z_s \mathrm{d}M_s + \int_0^t X_s \mathrm{d}Z_s \end{aligned}$$

which is a martingale. ∎

Since $Z_t > 0$ is a positive martingale up to time T, we may apply the Itô formula to $\log Z_t$, to obtain

$$\log Z_t - \log Z_0 = \int_0^t \frac{1}{Z_s} \mathrm{d}Z_s - \frac{1}{2} \int_0^t \frac{1}{Z_s^2} \mathrm{d}\langle Z\rangle_s ,$$

that is, $Z_t = \mathcal{E}(N)_t$ with

$$N_t = \int_0^t \frac{1}{Z_s} \mathrm{d}Z_s$$

is a continuous local martingale. Hence $Z_t = \mathcal{E}(N)_t$ solves the Itô integral equation

$$Z_t = 1 + \int_0^t Z_s \mathrm{d}N_s ,$$

and

$$\langle M, Z\rangle_t = \langle \int_0^t \mathrm{d}M_s, \int_0^t Z_s \mathrm{d}N_s \rangle = \int_0^t Z_s \mathrm{d}\langle N, M\rangle_s .$$

It follows thus that

$$\int_0^t \frac{1}{Z_s} \mathrm{d}\langle M, Z\rangle_s = \langle N, M\rangle_t .$$

Corollary 4.7.12 *Let N_t be a continuous local martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, $N_0 = 0$, such that its stochastic exponential $\mathcal{E}(N)_t$ is a continuous martingale up to T. Define a probability measure Q on the measurable space $(\Omega, \mathcal{F}_T)$ by*

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \mathcal{E}(N)_t \quad \textit{for all} \;\; t \le T.$$

If $M = (M_t)_{t\ge 0}$ is a continuous local martingale under the probability P, then

$$X_t = M_t - \langle N, M\rangle_t$$

is a continuous, local martingale under Q up to time T. (You should carefully define the concept of a local martingale up to time T).

4.7.6 The martingale representation theorem

The martingale representation theorem is a deep result about Brownian motion. There is a natural version for multi-dimensional Brownian motion, for simplicity of notations, we however concentrate on one-dimensional Brownian motion.

Let $B = (B_t)_{t\ge 0}$ be a standard Brownian motion in $\mathbb{R}$ on a complete probability space $(\Omega, \mathcal{F}, P)$, and $(\mathcal{F}_t^0)_{t\ge 0}$ (together with $\mathcal{F}_\infty^0 = \sigma\left(\bigcup_{t>0} \mathcal{F}_t^0\right)$) be the filtration generated by the Brownian motion $(B_t)_{t\ge 0}$. Let $\mathcal{F}_\infty$ be the completion of $\mathcal{F}_\infty^0$ and $\mathcal{F}_t$ the completion of $\mathcal{F}_t^0$ by adding in all null sets of $\mathcal{F}_\infty$.As a matter of fact, $(\mathcal{F}_t)_{t\ge 0}$ is continuous.

Theorem 4.7.13 *Let $M = (M_t)_{t\ge 0}$ be a square-integrable martingale on the filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Then there is a stochastic process $F = (F_t)_{t\ge 0}$ in $\mathcal{L}^2$, such that*

$$M_t = E(M_0) + \int_0^t F_s dB_s \qquad a.s.$$

for any $t \ge 0$. In particular, any martingale with respect to the Brownian filtration $(\mathcal{F}_t)_{t\ge 0}$ has a continuous version.

The proof of this theorem relies on the following several lemmata. Let $T > 0$ be any fixed time.

Lemma 4.7.14 *The following collection of random variables on $(\Omega, \mathcal{F}_T, P)$*

$$\left\{\phi(B_{t_1}, \cdots, B_{t_k}) : \forall k \in \mathbb{Z}^+, \;\; t_j \in [0,T] \;\textit{and}\; \phi \in C_0^\infty(\mathbb{R}^k)\right\}$$

is dense in $L^2(\Omega, \mathcal{F}_T, P)$.

Proof. If $X \in L^2(\Omega, \mathcal{F}_T, P)$, then, by definition, there is an $\mathcal{F}_T^0$-measurable function which equals X almost surely. Therefore, without losing generality, we may assume that $X \in L^2(\Omega, \mathcal{F}_T^0, P)$. According to definition, $\mathcal{F}_T^0 = \sigma\{B_t : t \le T\}$. Let $D = \mathbb{Q} \cap [0,T]$ the set of all rational numbers in the interval $[0,T]$.

Since D is dense in $[0,T]$, $\mathcal{F}_T^0 = \sigma\{B_t : t \in D\}$. Moreover D is countable, and we may write $D = \{t_1, \cdots, t_n, \cdots\}$. Let $D_n = \{t_1, \cdots, t_n\}$ for each n, and $\mathcal{G}_n = \sigma\{B_{t_1}, \cdots, B_{t_n}\}$. Then $\{\mathcal{G}_n\}$ is increasing, and $\mathcal{G}_n \uparrow \mathcal{F}_T^0$. Let $X_n = E(X|\mathcal{G}_n)$. Then $(X_n)_{n\geq 1}$ is square-integrable martingale, and thus, according to the martingale convergence theorem

$$X_n \to X \quad \text{almost surely.}$$

Moreover $X_n \to X$ in L^2. While, for each n, X_n is measurable with respect to $\mathcal{G}_n$, we have

$$X_n = f_n(B_{t_1}, \cdots, B_{t_n})$$

for some Borel measurable function $f_n : \mathbb{R}^n \to \mathbb{R}$. Since $X_n \in L^2$, $f_n \in L^2(\mathbb{R}^n, \mu)$ where μ is a Gaussian measure such that

$$EX_n^2 = \int_{\mathbb{R}^n} f(x)^2 \mu(dx) \ .$$

Since $C_0^\infty(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n, \mu)$, for each n, there is a sequence $\{\phi_{nk}\}$ in $C_0^\infty(\mathbb{R}^n)$ such that $\phi_{nk} \to f_n$ in $L^2(\mathbb{R}^n, \mu)$. It follows that

$$\phi_{nn}(B_{t_1}, \cdots, B_{t_n}) \to X$$

in L^2. ■

If $I \subset \mathbb{R}$ is an interval, then we use $L^2(I)$ to denote the Hilbert space of all functions h on I which are square-integrable.

Lemma 4.7.15 *Let $T > 0$. For any $h \in L^2([0,T])$, we associate with an exponential martingale up to time T:*

$$M(h)_t = \exp\left\{\int_0^t h(s)dB_s - \frac{1}{2}\int_0^t h(s)^2 ds\right\} \ ; \quad t \in [0,T]. \tag{4.25}$$

Then $\mathbb{L} = span\{M(h)_T : h \in L^2([0,T])\}$ is dense in $L^2(\Omega, \mathcal{F}_T, P)$.

Proof. It suffices to show that if $H \in L^2(\Omega, \mathcal{F}_T, P)$ such that

$$\int_\Omega H\Phi \mathrm{d}P = 0 \qquad \text{for all } \ \Phi \in \mathbb{L},$$

then $H = 0$.

For any $0 = t_0 < t_1 < \cdots < t_n = T$ and $c_i \in \mathbb{R}$, consider a step function $h(t) = c_i$ for $t \in (t_i, t_{i+1}]$. Then

$$M(h)_T = \exp\left\{\sum_i c_i(B_{t_{i+1}} - B_{t_i}) - \frac{1}{2}\sum_i c_i^2(t_{i+1} - t_i)\right\} \ .$$

Since $\int_\Omega H\Phi \mathrm{d}P = 0$ for any $\Phi \in \mathbb{L}$, so that

$$\int_\Omega H \exp\left\{\sum_i c_i(B_{t_{i+1}} - B_{t_i}) - \frac{1}{2}\sum_i c_i^2(t_{i+1} - t_i)\right\} \mathrm{d}P = 0 \ .$$

The deterministic, positive term $e^{-\frac{1}{2}\sum_i c_i^2(t_{i+1}-t_i)}$ can be removed from the integrand, and it follows therefore that

$$\int_\Omega H \exp\left\{\sum_i c_i(B_{t_{i+1}} - B_{t_i})\right\} \mathrm{d}P = 0 \ .$$

Since c_i are arbitrary numbers, it holds

$$\int_\Omega H \exp\left\{\sum_i c_i B_{t_i}\right\} \mathrm{d}P = 0$$

for any c_i and $t_i \in [0, T]$. Since the left-hand side is analytic in c_i, the equality remains true for any complex numbers c_i. If $\phi \in C_0^\infty(\mathbb{R}^n)$, then

$$\phi(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{\phi}(z) e^{\sqrt{-1}\langle z,x\rangle} \mathrm{d}z$$

where

$$\hat{\phi}(z) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \phi(x) e^{-\sqrt{-1}\langle z,x\rangle} \mathrm{d}x$$

is the Fourier transform of ϕ. Hence

$$\begin{aligned}
\int_\Omega H\phi(B_{t_1}, \cdots, B_{t_n})\mathrm{d}P &= \frac{1}{(2\pi)^{n/2}} \int_\Omega \left\{ H \int_{\mathbb{R}^n} \hat{\phi}(z) \exp\left(i \sum_j z_j B_{t_j} \right) \right\} \mathrm{d}z\mathrm{d}P \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left\{ \hat{\phi}(z) \int_\Omega H \exp\left(i \sum_i z_i B_{t_i} \right) \mathrm{d}P \right\} \mathrm{d}z \\
&= 0 \ .
\end{aligned}$$

Therefore, for any $\phi \in C_0^\infty(\mathbb{R}^n)$,

$$\int_\Omega H\phi(B_{t_1}, \cdots, B_{t_n})\mathrm{d}P = 0 \ . \tag{4.26}$$

By Lemma 4.7.14, the collection of all functions like $\phi(B_{t_1}, \cdots, B_{t_n})$ is dense in $L^2(\Omega, \mathcal{F}_T, P)$, and hence we have

$$\int_\Omega HG\mathrm{d}P = 0 \quad \text{for any} \quad G \in L^2(\Omega, \mathcal{F}_T, P) \ .$$

In particular, $E(H^2) = 0$ so that $H = 0$. ■

Theorem 4.7.16 (Itô) *Let $\xi \in L^2(\Omega, \mathcal{F}_T, P)$. Then there is a $F = (F_t)_{t\geq 0} \in \mathcal{L}^2$, such that*

$$\xi = E(\xi) + \int_0^T F_t dB_t \ .$$

Proof. By Lemma 4.7.15 we only need to show this lemma for $\xi = X(h)_T$ (where $h \in L^2([0,T])$) defined by (4.25). While, $X(h)_t$ is an exponential martingale so that it must satisfy the following integral equation

$$\begin{aligned} X(h)_T &= 1 + \int_0^T X(h)_t \mathrm{d}\left(\int_0^t h(s)\mathrm{d}B_s\right) \\ &= E(X(h)_T) + \int_0^T X(h)_t h(t)\mathrm{d}B_t \ . \end{aligned}$$

Therefore $F_t = X(h)_t h(t)$ will do. ∎

The martingale representation theorem now follows easily from the martingale property and Itô's representation theorem.

4.8 Additional topics

Quadratic variation of continuous square integrable martingale

As we said above, Theorem 4.2.2 is a direct consequence of Doob-Meyer decomposition. However the proof of Doob-Meyer's decomposition involves too much to be contained in this book. Here we introduce a direct proof taken from [14], which needs almost nothing new. In this topic, we assume that $M = (M_t : t \geq 0)$ is a bounded continuous martingale. The general case may be done by localization.

To prove the existence in Theorem 4.2.2 we prepare a lemma which is left as an exercise since its proof is just a careful verification. Define a process $T^D(X) = (T_t^D(X) : t \geq 0)$ for any process $(X_t : t \geq 0)$, by

$$T_t^D(X) := \sum_i (X_{t_{i+1}\wedge t} - X_{t_i \wedge t})^2, \ \forall t \geq 0,$$

where D is a partition of $[0, \infty)$ (D has finite many points on finite interval).

Exercise 4.7 *For any partition D, prove that $(T_t^D(M) : t \geq 0)$ is a continuous martingale.*

Actually the assertion in exercise is still true if assuming only that each M_t be bounded. In this case each $T_t^D(M)$ is bounded. Take two partitions D and D' on $[0, \infty)$, and denote by D'' the partition obtained by merging two. It follows from Exercise that the process

$$(T_t^D(M) - T_t^{D'}(M))^2 - T_t^{D''}(T^D(M) - T^{D'}(M)), \ t \geq 0$$

is a continuous martingale. Hence

$$\begin{aligned} E(T_t^D(M) - T_t^{D'}(M))^2 &= E\left(T_t^{D''}(T^D(M) - T^{D'}(M))\right) \\ &\leq 2T_t^{D''}(T^D(M)) + 2T_t^{D''}(T^{D'}(M)). \end{aligned}$$

We shall now verify that

$$\lim_{m(D)\to 0} E\left(T_t^{D''}(T^D(M))\right) \longrightarrow 0.$$

Indeed, let $D = \{t_i'\}$, $D'' = \{s_i'\}$ and let $t_i := t_i' \wedge t$, $s_i := s_i' \wedge t$. For any k, there exists a unique l such that $t_l \le s_k \le s_{k+1} \le t_{l+1}$. Hence we have

$$\begin{aligned} T^D_{s_{k+1}}(M) - T^D_{s_k}(M) &= (M_{s_{k+1}} - M_{t_l})^2 - (M_{s_k} - M_{t_l})^2 \\ &= (M_{s_{k+1}} - M_{s_k})(M_{s_{k+1}} + M_{s_k} - 2M_{t_l}), \end{aligned}$$

and

$$\begin{aligned} T_t^{D''}(T^D(M)) &= \sum_k (T^D_{s_{k+1}}(M) - T^D_{s_k}(M))^2 \\ &\le T_t^{D''}(M) \cdot \sup_k (M_{s_{k+1}} + M_{s_k} - 2M_{t_l})^2, \end{aligned}$$

By Cauchy-Schwarz's inequality, we have

$$[ET_t^{D''}(T^D(M))]^2 \le E[T_t^{D''}(M)]^2 \cdot E \sup_k (M_{s_{k+1}} + M_{s_k} - 2M_{t_l})^4.$$

When $m(D) \to 0$, the second expectation of right side goes to zero due to continuity and boundedness of M on $[0, t]$. It remains to check that $E(T_t^{D''}(M))^2$ is controlled by a constant independent of D''.

Exercise 4.8 *Prove that for any partition D,*

$$E(T_t^D(M))^2 \le 16 \sup_{s\le t, \omega\in\Omega} |M_s(\omega)|^4.$$

Therefore we have

$$\lim_{m(D), m(D')\to 0} E(T_t^D(M) - T_t^{D'}(M))^2 = 0.$$

Take now a sequence $\{D_n\}$ of partitions decreasing to zero and then $\{M^2 - T^{D_n}(M)\}$ is a Cauchy sequence in $\mathcal{M}_0^2$. By Theorem 4.2.1, it has a limit in $\mathcal{M}_0^2$. Therefore there exists a continuous process A, such that for any $t \ge 0$, $T_t^{D_n}(M)$ converges in L^2 and thus also in probability to A_t and $M^2 - A$ is a martingale. Since $T_0^{D_n}(M) = 0$ a.s., $A_0 = 0$ a.s. Moreover for any $t > s$ with $s, t \in \bigcup_n D_n$, there exists n large enough such that

$$T_t^{D_n}(M) \ge T_s^{D_n}(M), \text{ a.s.}$$

Hence $A_t \ge A_s$ a.s., and it follows from the continuity that A is increasing. By the uniqueness, A is independent of choice of D_n. Therefore $A = \langle M \rangle$.

Exercise 4.9 *Using localization to prove all corresponding conclusions for continuous local martingales.*

Strotonovich integral

The Stratonovich integral $\int_0^t F_s \circ \mathrm{d}M_s$, which was discovered later than Itô's, is defined on the other hand by

$$\int_0^t F_s \circ \mathrm{d}M_s = \lim_{m(D)\to 0} \sum_{l=1}^n \frac{F_{t_{l-1}} + F_{t_l}}{2}(M_{t_l} - M_{t_{l-1}})$$

which in general is different from the Itô integral $\int_0^t F_s \mathrm{d}M_s$.

According to definition,

$$\begin{aligned}
\int_0^t M_s \mathrm{d}M_s &= \lim_{m(D)\to 0} \sum_{l=1}^n M_{t_{l-1}}(M_{t_l} - M_{t_{l-1}}) \\
&= \lim_{m(D)\to 0} \sum_{l=1}^n \left\{ -\frac{1}{2}(M_{t_l} - M_{t_{l-1}})^2 + \frac{1}{2}\left(M_{t_l}^2 - M_{t_{l-1}}^2\right) \right\} \\
&= -\frac{1}{2} \lim_{m(D)\to 0} \sum_{l=1}^n (M_{t_l} - M_{t_{l-1}})^2 + \frac{1}{2} \lim_{m(D)\to 0} \sum_{l=1}^n \left(M_{t_l}^2 - M_{t_{l-1}}^2\right) \\
&= -\frac{1}{2}\langle M\rangle_t + \frac{1}{2}(M_t^2 - M_0^2) \ .
\end{aligned}$$

That is

$$M_t^2 - M_0^2 = 2\int_0^t M_s \mathrm{d}M_s + \langle M\rangle_t \ .$$

On the other hand

$$\begin{aligned}
\int_0^t M_s \circ \mathrm{d}M_s &= \lim_{m(D)\to 0} \sum_{l=1}^n \frac{M_{t_l} + M_{t_{l-1}}}{2}(M_{t_l} - M_{t_{l-1}}) \\
&= \frac{1}{2} \lim_{m(D)\to 0} \sum_{l=1}^n (M_{t_l} - M_{t_{l-1}})^2 + \lim_{m(D)\to 0} \sum_{l=1}^n M_{t_{l-1}}(M_{t_l} - M_{t_{l-1}}) \\
&= \frac{1}{2}\langle M\rangle_t + \int_0^t M_s \mathrm{d}M_s \\
&= \frac{1}{2}\left(M_t^2 - M_0^2\right)
\end{aligned}$$

so that

$$M_t^2 - M_0^2 = 2\int_0^t M_s \circ \mathrm{d}M_s$$

which looks very much like the fundamental theorem in Calculus. In general, we have

Lemma 4.8.1 *Let N, M be two continuous, square-integrable martingales, and $F_t = N_t + A_t$ where A_t is an adapted process with finite variation. Define Stratonovich's integral*

$$\int_0^t F_s \circ dM_s = \lim_{m(D)\to 0} \sum_l \frac{F_{t_l} + F_{t_{l-1}}}{2}(M_{t_l} - M_{t_{l-1}})$$

Then

$$\int_0^t F_s \circ dM_s = \int_0^t F_s \, dM_s + \frac{1}{2}\langle N, M\rangle_t$$

Indeed,

$$\begin{aligned}
\int_0^t F_s \circ \mathrm{d}M_s &= \lim_{m(D)\to 0} \sum_{l=1}^n \frac{N_{t_l} + N_{t_{l-1}}}{2}(M_{t_l} - M_{t_{l-1}}) \\
&\quad + \lim_{m(D)\to 0} \sum_{l=1}^n \frac{A_{t_l} + A_{t_{l-1}}}{2}(M_{t_l} - M_{t_{l-1}}) \\
&= \frac{1}{2} \lim_{m(D)\to 0} \sum_{l=1}^n \left(N_{t_l} - N_{t_{l-1}}\right)(M_{t_l} - M_{t_{l-1}}) \\
&\quad + \lim_{m(D)\to 0} \sum_{l=1}^n F_{t_{l-1}}(M_{t_l} - M_{t_{l-1}}) \\
&\quad + \frac{1}{2} \lim_{m(D)\to 0} \sum_{l=1}^n \left(A_{t_l} - A_{t_{l-1}}\right)(M_{t_l} - M_{t_{l-1}}) \\
&= \frac{1}{2}\langle M, N\rangle_t + \int_0^t F_s \mathrm{d}M_s \ .
\end{aligned}$$

In particular, if $F = (F_t)_{t\geq 0}$ is a process with finite variation, then

$$\int_0^t F_s \circ \mathrm{d}M_s = \int_0^t F_s \mathrm{d}M_s \ .$$

We will concentrate on Itô's integrals only.

Chapter 5

Stochastic differential equations

The main goal of the chapter is to establish the basic existence and uniqueness theorem for a class of stochastic differential equations which are most important in application.

5.1 Introduction

Stochastic differential equations (SDE) are ordinary differential equations perturbed by noises. We will consider a simple class of noises modelled by Brownian motion. Thus we consider the following type of equation

$$dX_t^j = \sum_{i=1}^n f_i^j(t, X_t)dB_t^i + f_0^j(t, X_t)dt \ , \quad j = 1, \cdots, N \tag{5.1}$$

where $B_t = (B_t^1, \cdots, B_t^n)_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}^n$ on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and

$$f_i^j : [0, +\infty) \times \mathbb{R}^N \to \mathbb{R}, \ 1 \leq i \leq n, 1 \leq j \leq N,$$

are Borel measurable functions. Of course, differential equation (5.1) should be interpreted as an integral equation in terms of Itô's integration. More precisely, an adapted, continuous, $\mathbb{R}^N$-valued stochastic process $X_t \equiv (X_t^1, \cdots, X_t^N)$ is a solution of (5.1), if

$$X_t^j = X_0^j + \sum_{k=1}^n \int_0^t f_k^j(s, X_s)dB_s^k + \int_0^t f_0^j(s, X_s)ds \tag{5.2}$$

for $j = 1, \cdots, N$. Since we are concerned only with the distribution determined by the solution $(X_t)_{t\geq 0}$ of (5.1), we therefore expect that any solution of SDE

(5.1) should have the same distribution for *any* Brownian motion $B = (B_t)_{t\geq 0}$. It thus leads to different concepts of solutions and uniqueness: strong solutions and weak solutions, path-wise uniqueness and uniqueness in law.

Definition 5.1.1 *1) An adapted, continuous, $\mathbb{R}^N$-valued stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is a (weak) solution of (5.1), if there is a Brownian motion $W = (W_t)_{t\geq 0}$ in $\mathbb{R}^n$, adapted to the filtration $(\mathcal{F}_t)$, such that*

$$X_t^j - X_0^j = \sum_{l=1}^n \int_0^t f_l^j(s, X_s)dW_s^l + \int_0^t f_0^j(s, X_s)ds, \quad j = 1, \cdots, N.$$

In this case we also call the pair (X, W) a (weak) solution of (5.1).

2) Given a standard Brownian motion $B = (B_t)_{t\geq 0}$ in $\mathbb{R}^n$ on $(\Omega, \mathcal{F}, P)$ with its natural filtration $(\mathcal{F}_t)_{t\geq 0}$, an adapted, continuous stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is a strong solution *of (5.1), if*

$$X_t^j - X_0^j = \sum_{i=1}^n \int_0^t f_i^j(s, X_s)dB_s^i + \int_0^t f_0^j(s, X_s)ds \ .$$

The strong solution is a functional of the given Brownian motion, or a stochastic process on Wiener space. We also have different concepts of uniqueness.

Definition 5.1.2 *Consider SDE (5.1).*

1) We say that the ***path-wise uniqueness*** *holds for (5.1), if whenever (X, B) and $(\widetilde{X}, B)$ are two solutions defined on the same filtered space and same Brownian motion B, and $X_0 = \widetilde{X}_0$, then $X = \widetilde{X}$.*

2) It is said that ***uniqueness in law*** *holds for (5.1), if (X, B) and $(\widetilde{X}, \widetilde{B})$ are two solutions (with possibly different Brownian motions B and $\widetilde{B}$, even can be on different probability spaces), and X_0 and $\widetilde{X}_0$ possess the same distribution, then X and $\widetilde{X}$ have same distribution.*

Theorem 5.1.3 (Yamada-Watanabe) *Path-wise uniqueness implies uniqueness in law.*

The following is a simple example of SDE for which has no strong solution, but possesses weak solutions and uniqueness in law holds.

Example 5.1.4 (Tanaka) *Consider 1-dimensional stochastic differential equation:*

$$X_t = \int_0^t \mathrm{sgn}(X_s)dB_s \ , \qquad 0 \leq t < \infty$$

where $\mathrm{sgn}(x) = 1$ *if $x \geq 0$, and equals -1 for negative value of x.*

1. Uniqueness in law holds, since X is a standard Brownian motion (by Lévy's Characterization Theorem).

2. There is a weak solution. Let W_t be a one-dimensional Brownian motion, and let $B_t = \int_0^t \mathrm{sgn}(W_s) dW_s$. Then B is a one-dimensional Brownian motion, and
$$W_t = \int_0^t \mathrm{sgn}(W_s) dB_s \ ,$$
so that (W, B) is a solution.

3. If (X, B) is a weak solution, then so is $(-X, B)$. Therefore path-wise uniqueness does not hold.

4. There is no strong solution.

5.2 Several examples

5.2.1 Linear-Gaussian diffusions

Linear stochastic differential equations can be solved explicitly. Consider
$$dX_t^j = \sum_{i=1}^n \sigma_i^j dB_t^i + \sum_{k=1}^N \beta_k^j X_t^k dt \tag{5.3}$$
$(j = 1, \cdots, N)$, where B is a Brownian motion in $\mathbb{R}^n$, $\sigma = (\sigma_i^j)$ a constant $N \times n$ matrix, and $\beta = (\beta_k^j)$ a constant $N \times N$ matrix. (5.3) may be written as
$$dX_t = \sigma dB_t + \beta X_t dt \ .$$

Let
$$e^{\beta t} = \sum_{k=0}^\infty \frac{t^k}{k!} \beta^k$$
be the exponential of the square matrix β. Using Itô's formula, we have
$$\begin{aligned} e^{-\beta t} X_t - X_0 &= \int_0^t e^{-\beta s} dX_s - \int_0^t e^{-\beta s} \beta X_s ds \\ &= \int_0^t e^{-\beta s} (dX_s - \beta X_s ds) \\ &= \int_0^t e^{-\beta s} \sigma dB_s \end{aligned}$$
and therefore
$$X_t = e^{\beta t} X_0 + \int_0^t e^{\beta(t-s)} \sigma \mathrm{d} B_s \ .$$
In particular, if $X_0 = x$, then X_t has a normal distribution with mean $e^{\beta t} x$. For example, if $n = N = 1$, then
$$X_t \sim N(e^{\beta t} x, \frac{\sigma^2}{2} \left(e^{2\beta t} - 1\right)) \ .$$

It can be shown that (X_t) is a diffusion process, and thus its distribution can be described by its transition probability $P_t(x,dz)$. Then

$$\begin{aligned}(P_tf)(x) &\equiv \int_{\mathbb{R}^N} f(z)P_t(x,dz) \\ &= E\left(f(X_t)|X_0=x\right) ,\end{aligned}$$

thus

$$\begin{aligned}(P_tf)(x) &= E\left(f(X_t)|X_0=x\right) \\ &= \int_{\mathbb{R}} f(z)\frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}\left(e^{2\beta t}-1\right)}}\exp\left(-\frac{|z-e^{\beta t}x|^2}{\frac{\sigma^2}{2}\left(e^{2\beta t}-1\right)}\right)dz \\ &= \int_{\mathbb{R}} f(e^{\beta t}x+\sqrt{\frac{\sigma^2}{2}\left(e^{2\beta t}-1\right)}z)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{|z|^2}{2}\right)dz \\ &= Ef(e^{\beta t}x+\sqrt{\frac{\sigma^2}{2}\left(e^{2\beta t}-1\right)}\xi)\end{aligned}$$

where ξ has the standard normal distribution $N(0,1)$. From the second line of the above formula, and compare to the definition of $P_t(x,dz)$, we may conclude that

$$P_t(x,dz)=p(t,x,z)dz$$

with

$$p(t,x,z)=\frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}\left(e^{2\beta t}-1\right)}}\exp\left(-\frac{|z-e^{\beta t}x|^2}{\frac{\sigma^2}{2}\left(e^{2\beta t}-1\right)}\right) .$$

$p(t,x,z)$ is called the transition density of the diffusion process $(X_t)_{t\geq 0}$.

Remark 5.2.1 *It is easy to see from the above representation that*

$$\frac{d}{dx}(P_tf)=e^{\beta t}P_t\left(\frac{d}{dx}f\right) .$$

Remark 5.2.2 *Diffusion processes are a kind of stochastic processes which inherit many characters from Brownian motion and will be carefully introduced in the later chapters. Its transition density, $p(t,x,\cdot)$, if exists, is understood as the density function of X_t conditioned on $X_0=x$ and it satisfies the Chapman-Kolmogorov equation*

$$\int_{\mathbb{R}^n} p(t,x,y)p(s,y,z)dy=p(s+t,x,z) \tag{5.4}$$

for any $s,t\geq 0$ and $x,z\in\mathbb{R}^n$. A diffusion may now be understood as a continuous process with transition density.

The distribution of (X_t) is determined by the transition density $p(t,x,z)$. Indeed, for any $0 < t_1 < \cdots < t_k$, the joint distribution of $(X_{t_1}, \cdots, X_{t_k})$ is Gaussian, and indeed its pdf is

$$p(t_1, x, z_1)p(t_2 - t_1, z_1, z_2) \cdots p(t_k - t_{k-1}, z_{k-1}, z_k) \ .$$

If $B = (B_t^1, \cdots, B_t^n)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^n$, then the solution X_t of the SDE:

$$dX_t = dB_t - (AX_t)\, dt$$

is called the Ornstein-Uhlenbeck process, where $A \geq 0$ is a $d \times d$ matrix called the drift matrix. Hence we have

$$X_t = e^{-At}X_0 + \int_0^t e^{-(t-s)A} dB_s \ .$$

5.2.2 Geometric Brownian motion

Consider the Black-Scholes model

$$dS_t = S_t \left(\mu dt + \sigma dB_t\right) \ . \tag{5.5}$$

The solution of (5.5) is the stochastic exponential of

$$\int_0^t \mu ds + \int_0^t \sigma dB_s \ ,$$

and therefore

$$S_t = S_0 \exp\left\{ \int_0^t \sigma dB_s + \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) ds \right\} \ .$$

In the case σ and μ are constants,

$$S_t = S_0 \exp\left\{ \sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right) t \right\}$$

which is called the geometric Brownian motion. If $S_0 = x > 0$, then S_t remains positive, and

$$\log S_t = \log x + \sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right) t$$

has a normal distribution with mean $\log x + \left(\mu - \frac{1}{2}\sigma^2\right) t$ and variance σ^2. Again, as a solution to the stochastic differential equation (5.5), $(S_t)_{t\geq 0}$ is a diffusion process, its distribution is determined by its transition function $P_t(x, dz)$ (unfortunately we have to use the same notations as in the last sub-section), and according to definition

$$\begin{aligned}\int_{\mathbb{R}} f(z)P_t(x,dz) &= E\left(f(X_t)|X_0 = x\right) \\ &= E\left(f(xe^{\sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right)t})\right) \\ &= \int_{\mathbb{R}} f(xe^{\sigma z + \left(\mu - \frac{1}{2}\sigma^2\right)t}) \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2\pi t}} dz \\ &= \int_0^\infty f(y) \frac{1}{\sigma y\sqrt{2\pi t}} e^{-\frac{1}{2\pi t}\left(\frac{1}{\sigma}\log\frac{y}{x} - \left(\frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right)^2} dy\end{aligned}$$

where we assume that $\sigma > 0$ and have made the change of variable

$$xe^{\sigma z + \left(\mu - \frac{1}{2}\sigma^2\right)t} = y \ .$$

As usual, we define $(P_t f)(x) = \int_{\mathbb{R}} f(z)P_t(x,dz)$. By the third line of the previous formula

$$\begin{aligned}(P_t f)(x) &= \int_{\mathbb{R}} f(xe^{\sigma z + \left(\mu - \frac{1}{2}\sigma^2\right)t}) \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2\pi t}} dz \\ &= \int_{\mathbb{R}} f(xe^{\sigma\sqrt{t}y + \left(\mu - \frac{1}{2}\sigma^2\right)t}) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2\pi}} dy \\ &= E\left(f(xe^{\sigma\sqrt{t}\xi + \left(\mu - \frac{1}{2}\sigma^2\right)t})\right)\end{aligned}$$

(we have made a change variable z into $\sqrt{t}y$), where $\xi \sim N(0,1)$. Comparing with the definition of $P_t(x,dy)$ we have

$$P_t(x,dy) = \frac{1}{\sqrt{2\pi t}} \frac{1}{\sigma y} e^{-\frac{1}{2\pi t}\left(\frac{1}{\sigma}\log\frac{y}{x} - \left(\frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right)^2} dy \quad \text{on } (0,+\infty)$$

That is, (S_t) has the transition probability density

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} \frac{1}{\sigma y} e^{-\frac{1}{2\pi t}\left(\frac{1}{\sigma}\log\frac{y}{x} - \left(\frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right)^2} \quad \text{on } (0,+\infty) \ .$$

Therefore, for geometric Brownian motion

$$(P_t f)(x) = \int_0^\infty \frac{1}{\sqrt{2\pi t}} \frac{1}{\sigma y} e^{-\frac{1}{2\pi t}\left(\frac{1}{\sigma}\log\frac{y}{x} - \left(\frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right)^2} f(y)dy$$

for any $x > 0$.

5.2.3 Cameron-Martin's formula

Consider a simple stochastic differential equation

$$dX_t = dB_t + b(t,X_t)dt \tag{5.6}$$

where $b(t,x)$ is a bounded, Borel measurable function on $[0,+\infty)\times\mathbb{R}$. We may solve (5.6) by means of *change of probabilities*.

Let $(W_t)_{t\geq 0}$ be a standard Brownian motion on $(\Omega,\mathcal{F},\mathcal{F}_t,P)$, and define probability measure Q on $(\Omega,\mathcal{F}_\infty)$ by

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t}=\mathcal{E}(N)_t \qquad \text{for all} \quad t\geq 0$$

where $N_t=\int_0^t b(s,W_s)\mathrm{d}W_s$ is a martingale (under the probability P), with $\langle N\rangle_t=\int_0^t b(s,W_s)^2\mathrm{d}s$, which is bounded on any finite interval. Hence

$$\mathcal{E}(N)_t=\exp\left(\int_0^t b(s,W_s)\mathrm{d}W_s-\frac{1}{2}\int_0^t b(s,W_s)^2\mathrm{d}s\right)$$

is a martingale. According to Girsanov's theorem

$$B_t\equiv W_t-W_0-\langle W,N\rangle_t$$

is a martingale under the new probability Q, and $\langle B\rangle_t=\langle W\rangle_t=t$. By Lévy's martingale characterization of Brownian motion, $(B_t)_{t\geq 0}$ is a Brownian motion. Moreover

$$\begin{aligned}\langle W,N\rangle_t &= \langle\int_0^t \mathrm{d}W_s,\int_0^t b(s,W_s)\mathrm{d}W_s\rangle\\ &= \int_0^t b(s,W_s)\mathrm{d}s\end{aligned}$$

and therefore

$$W_t-W_0-\int_0^t b(s,W_s)\mathrm{d}s=B_t$$

is a standard Brownian motion on $(\Omega,\mathcal{F},Q)$. Thus

$$W_t=W_0+B_t+\int_0^t b(s,W_s)\mathrm{d}s \tag{5.7}$$

namely $(W_t)_{t\geq 0}$ on $(\Omega,\mathcal{F}_\infty,Q)$ is a solution of (5.6). The solution we have just constructed is a weak solution of SDE (5.6).

Theorem 5.2.3 (Cameron-Martin) *Let $b(t,x)=(b^1(t,x),\cdots,b^n(t,x))$ be bounded, Borel measurable functions on $[0,+\infty)\times\mathbb{R}^n$. Let $W_t=(W^1,\cdots,W_t^n)$ be a standard Brownian motion on a filtered probability space $(\Omega,\mathcal{F},\mathcal{F}_t,P)$, and let $\mathcal{F}_\infty=\sigma\{\mathcal{F}_t,t\geq 0\}$. Define probability measure Q on $(\Omega,\mathcal{F}_\infty)$ by*

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t}=\exp\left\{\sum_{k=1}^n\int_0^t b^k(s,W_s)dW_s^k-\frac{1}{2}\sum_{k=1}^n\int_0^t\left|b^k(s,W_s)\right|^2ds\right\}\quad \textit{for}\ \ t\geq 0\ .$$

Then $(W_t)_{t\geq 0}$ under the probability measure Q is a solution to

$$dX_t^j=dB_t^j+b^j(t,X_t)dt \tag{5.8}$$

for some Brownian motion $(B_t^1,\cdots,B_t^n)_{t\geq 0}$ under probability Q.

On the other hand, if (X_t) is a solution of SDE (5.8) on some probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and define $\tilde{P}$

$$\left.\frac{d\tilde{P}}{dP}\right|_{\mathcal{F}_t} = \exp\left\{-\sum_{k=1}^{n}\int_0^t b^k(s,X_s)\mathrm{d}B_s^k - \frac{1}{2}\sum_{k=1}^{n}\int_0^t \left|b^k(s,X_s)\right|^2 \mathrm{d}s\right\} \quad \text{for } \ t \geq 0$$

we may show that $(X_t)_{t\geq 0}$ under probability measure $\tilde{P}$ is a Brownian motion. Therefore solutions to SDE (5.8) is unique in law: all solutions have the same distribution.

5.3 Strong solutions: existence and uniqueness

In this section we present a fundamental result about the existence and uniqueness of strong solutions. By definition, any strong solution is a weak solution. We next prove a basic existence and uniqueness theorem for a stochastic differential equation under a global Lipschitz condition. Our proof will rely on two inequalities: The Gronwall inequality and Doob's L^p-inequality.

Lemma 5.3.1 (Gronwall inequality) *If a non-negative function g satisfies the integral equation*

$$g(t) \leq h(t) + \alpha\int_0^t g(s)ds \ , \qquad 0 \leq t \leq T$$

where $\alpha \geq 0$ is a constant and $h : [0,T] \to \mathbb{R}$ is an integrable function, then

$$g(t) \leq h(t) + \alpha\int_0^t e^{\alpha(t-s)}h(s)ds \ \ , \qquad 0 \leq t \leq T \ .$$

Proof. Let $F(t) = \int_0^t g(s)\mathrm{d}s$. Then $F(0) = 0$ and

$$F'(t) \leq h(t) + \alpha F(t)$$

so that

$$\left(e^{-\alpha t}F(t)\right)' \leq e^{-\alpha t}h(t) \ .$$

Integrating the differential inequality we obtain

$$\int_0^t \left(e^{-\alpha s}F(s)\right)' \mathrm{d}s \leq \int_0^t e^{-\alpha s}h(s)\mathrm{d}s$$

and therefore

$$F(t) \leq \int_0^t e^{\alpha(t-s)}h(s)\mathrm{d}s$$

which yields Gronwall's inequality. ∎

Consider the following stochastic differential equation

$$dX_t^j = \sum_{l=1}^n f_l^j(t, X_t)\mathrm{d}B_t^l + f_0^j(t, X_t)\mathrm{d}t \ ; \quad j = 1, \cdots, N \tag{5.9}$$

where $f_k^j(t,x)$ are Borel measurable functions on $\mathbb{R}^+ \times \mathbb{R}^N$, which are bounded on any compact subset in $\mathbb{R}^N$. We are going to show the existence and uniqueness by Picard's iteration. The main ingradient in the proof is a special case of Doob's L^p- inequality: if $(M_t)_{t\geq 0}$ is a square-integrable, continuous martingale with $M_0 = 0$, then for any $t > 0$

$$E\left\{\sup_{s\leq t}|M_s|^2\right\} \leq 4\sup_{s\leq t} E\left(|M_s|^2\right) = 4E\langle M\rangle_t \ . \tag{5.10}$$

Lemma 5.3.2 *Let $(B_t)_{t\geq 0}$ be a standard BM in $\mathbb{R}$ on $(\Omega, \mathcal{F}_t, \mathcal{F}, P)$, and $(Z_t)_{t\geq 0}$ and $(\tilde{Z}_t)_{t\geq 0}$ be two continuous, adapted processes. Let $f(t,x)$ be a Lipschitz function*

$$|f(t,x) - f(t,y)| \leq C|x-y| \ ; \quad \forall t \geq 0, \ \ x, y \in \mathbb{R}$$

for some constant C.

1) If

$$M_t = \int_0^t f(s, Z_s)dB_s - \int_0^t f(s, \tilde{Z}_s)dB_s \qquad \forall t \geq 0,$$

then

$$E\sup_{s\leq t}|M_s|^2 \leq 4C^2\int_0^t E\left|Z_s - \tilde{Z}_s\right|^2 ds, \ \forall t \geq 0.$$

2) If

$$N_t = \int_0^t f(s, Z_s)ds - \int_0^t f(s, \tilde{Z}_s)ds \qquad \forall t \geq 0$$

then

$$E\sup_{s\leq t}|N_s|^2 \leq C^2 t\int_0^t E\left|Z_s - \tilde{Z}_s\right|^2 ds \qquad \forall t \geq 0 \ .$$

Proof. To prove the first statement, we notice that

$$\sup_{s\leq t}|M_s|^2 = \sup_{s\leq t}\left|\int_0^s \left(f(u, Z_u) - f(u, \tilde{Z}_u)\right)\mathrm{d}B_u\right|^2 .$$

Then by Doob's L^2-inequality

$$\begin{aligned}
E\sup_{s\leq t}|M_s|^2 &= E\sup_{s\leq t}\left|\int_0^s \left(f(u, Z_u) - f(u, \tilde{Z}_u)\right)\mathrm{d}B_u\right|^2 \\
&\leq 4E\left|\int_0^t \left(f(s, Z_s) - f(s, \tilde{Z}_s)\right)\mathrm{d}B_s\right|^2 \\
&= 4E\int_0^t \left|f(s, Z_s) - f(s, \tilde{Z}_s)\right|^2 \mathrm{d}s \\
&\leq 4C^2\int_0^t E\left|Z_s - \tilde{Z}_s\right|^2 \mathrm{d}s \ .
\end{aligned}$$

Next we prove the second claim. Indeed

$$
\begin{aligned}
\sup_{s\le t}|N_s|^2 &= \sup_{s\le t}\left|\int_0^s \left(f(u,Z_u)-f(u,\tilde{Z}_u)\right)\mathrm{d}u\right|^2 \\
&\le \left(\int_0^t \left|f(s,Z_s)-f(s,\tilde{Z}_s)\right|\mathrm{d}s\right)^2 \\
&\le t\int_0^t \left|f(s,Z_s)-f(s,\tilde{Z}_s)\right|^2\mathrm{d}s \\
&\le C^2 t\int_0^t \left|Z_s-\tilde{Z}_s\right|^2\mathrm{d}s
\end{aligned}
$$

where the third line follows from the Schwartz inequality. ■

Theorem 5.3.3 *Consider SDE* (5.9). *Suppose that f_i^j satisfy the Lipschitz condition:*

$$
\left|f_i^j(t,x)-f_i^j(t,y)\right| \le C|x-y| \tag{5.11}
$$

and the linear-growth condition that

$$
\left|f_i^j(t,x)\right| \le C(1+|x|) \tag{5.12}
$$

for $t\in\mathbb{R}^+$ and $x,y\in\mathbb{R}^N$. Then for any $\eta\in L^2(\Omega,\mathcal{F}_0,P)$ and a standard Brownian motion $B_t=(B_t^i)$ in $\mathbb{R}^n$, there is a unique strong solution (X_t) of (5.9) with $X_0=\eta$.

Proof. For simplicity, let us prove a special case of this important theorem: the existence and uniqueness for one-dimensional stochastic differential equations

$$
dX_t = f(t,X_t)\mathrm{d}B_t\ ,\quad X_0=\eta\ ,
$$

and leave the details of the proof for the general case to the reader. Like for ODE, construct an approximation solutions via Picard's iteration:

$$
Y_0(t)=\eta
$$

and

$$
Y_{n+1}(t)=\eta+\int_0^t f(s,Y_n(s))\mathrm{d}B_s\ ,
$$

where $n=0,1,2,\cdots$. We are going to show that, for every $T>0$, the sequence $\{Y_n(t)\}$ converges to a solution $Y(t)$ uniformly on $[0,T]$ almost surely. Note that, every Y_n is a continuous square-integrable martingale. Then we have

$$
\begin{aligned}
E\sup_{0\le s\le t}|Y_1(s)-Y_0(s)|^2 &\le E\sup_{0\le s\le t}\left(\int_0^s |f(\tau,\eta)|\mathrm{d}B_\tau\right)^2 \\
&\le 4E\int_0^t f(\tau,\eta)^2\mathrm{d}s \\
&\le 8tC\left(1+E\eta^2\right)
\end{aligned}
$$

and, for any $t \leq T$, applying Doob's inequality and Lipschitz condition, we have

$$\begin{aligned} E\sup_{s\leq t}|Y_{n+1}(s)-Y_n(s)|^2 &= E\sup_{s\leq t}\left|\int_0^s (f(r,Y_n(r))-f(r,Y_{n-1}(r)))\,\mathrm{d}B_r\right|^2 \\ &\leq 4E\int_0^t (f(s,Y_n(s))-f(s,Y_{n-1}(s)))^2\,\mathrm{d}s \\ &\leq 4C^2E\int_0^t |Y_n(s)-Y_{n-1}(s)|^2\mathrm{d}s \\ &\leq 4C^2tE\sup_{s\leq t}|Y_n(s)-Y_{n-1}(s)|^2 \ . \end{aligned}$$

When $n=2$, we have

$$\begin{aligned} E\sup_{s\leq t}|Y_3(s)-Y_2(s)|^2 &\leq 4C^2\int_0^t E|Y_2(s)-Y_1(s)|^2 ds \\ &\leq 4C^2\cdot 4C^2\int_0^t sE\sup_{u\leq s}|Y_1(u)-Y_0(u)|^2 ds \\ &\leq (4C^2)^2\frac{t^2}{2}E\sup_{s\leq t}|Y_1(s)-Y_0(s)|^2. \end{aligned}$$

and therefore by induction

$$E\sup_{s\leq t}|Y_{n+1}(s)-Y_n(s)|^2 \leq \frac{\left(4C^2\right)^n t^n}{n!}E\sup_{s\leq t}|Y_1(s)-Y_0(s)|^2$$

for all $t \leq T$. In particular

$$E\sup_{t\leq T}|Y_{n+1}(t)-Y_n(t)|^2 \leq \frac{\left(4C^2\right)^n T^n}{n!}E\sup_{t\leq T}|Y_1(t)-Y_0(t)|^2$$

and it follows that

$$\begin{aligned} \sum_{n=0}^{\infty}E\sup_{t\leq T}|Y_{n+1}(t)-Y_n(t)|^2 &\leq \sum_{n=0}^{\infty}\frac{\left(4C^2\right)^n T^n}{n!}E\sup_{t\leq T}|Y_1(t)-Y_0(t)|^2 \\ &< \infty\ . \end{aligned}$$

Hence $\{Y_n : n \geq 1\}$ is a Cauchy sequence in $\mathcal{M}_0^2$, and

$$Y_n(t) \to X_t \quad \text{uniformly on } [0,T]\ , \quad P\text{-a.s.}$$

It is easy to see that (X_t) is a strong solution of the stochastic differential equation.

Next we prove the uniqueness. Let Y and Z be two solutions with same Brownian motion B. Then

$$Y_t = \eta + \int_0^t f(s,Y_s)\mathrm{d}B_s$$

and

$$Z_t = \eta + \int_0^t f(s, Z_s)\mathrm{d}B_s \ .$$

Then, as in the proof of the existence,

$$E\left(|Y_t - Z_t|^2\right) \leq 4C^2 \int_0^t E|Y_s - Z_s|^2\mathrm{d}s$$

The Gronwall inequality implies thus that

$$E\left(|Y_t - Z_t|^2\right) = 0 \ .$$

■

Remark 5.3.4 *The iteration Y_n constructed in the proof of Theorem 5.3.3 is a function of the Brownian motion B, and $Y_n(t)$ only depends on η and $(B_s : 0 \leq s \leq t)$.*

5.4 Martingales and weak solutions

Consider the time-homogenous stochastic differential equation

$$dX_t^i = \sum_{j=1}^m \sigma_j^i(X_t)\mathrm{d}B_t^j + b^i(X_t)\mathrm{d}t \tag{5.13}$$

where σ_j^i, $b^i \in C^\infty(\mathbb{R}^n)$ are smooth functions with bounded derivatives, and $B = (B_t)$ be an m-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Let $X = (X_t)_{\geq 0}$ be the unique strong solution with initial X_0 . If $f \in C_b^2(\mathbb{R}^n, \mathbb{R})$, then by Itô's formula

$$\begin{aligned} f(X_t) - f(X_0) &= \int_0^t \sum_{k=1}^n \frac{\partial f}{\partial x^k}(X_s)\mathrm{d}X_s^k \\ &+ \frac{1}{2}\int_0^t \sum_{k,l=1}^n \frac{\partial^2 f}{\partial x^k \partial x^l}(X_s)\mathrm{d}\langle X^k, X^l\rangle_s \ . \end{aligned}$$

According to (5.13)

$$\langle X^k, X^l\rangle_t = \int_0^t \sum_{j=1}^m \sigma_j^k(X_s)\sigma_j^l(X_s)ds$$

one thus obtains

$$\begin{aligned} f(X_t) - f(X_0) &= \int_0^t \sum_{j=1}^m \left(\sum_{k=1}^n \sigma_j^k \frac{\partial}{\partial x^k}\right) f(X_s)\mathrm{d}B_s^j \\ &+ \int_0^t \left(\frac{1}{2}\sum_{k,l=1}^n \sum_{j=1}^m \sigma_j^k \sigma_j^l \frac{\partial^2}{\partial x^k \partial x^l} + \sum_{k=1}^n b^k \frac{\partial}{\partial x^k}\right) f(X_s)ds \ . \end{aligned}$$

Define $a = (a^{kl})_{k,l\leq n}$ where

$$a^{kl} = \sum_{j=1}^{m} \sigma_j^k \sigma_j^l \ .$$

Then $(a^{kl})_{k,l\leq n}$ is symmetric and non-negative-definite. Let

$$L = \frac{1}{2} \sum_{k,l=1}^{n} a^{kl} \frac{\partial^2}{\partial x^k \partial x^l} + \sum_{k=1}^{n} b^k \frac{\partial}{\partial x^k} \tag{5.14}$$

which is an elliptic differential operator of second-order on $\mathbb{R}^n$. Then

$$f(X_t) - f(X_0) = \int_0^t \sum_{j=1}^{m} \left(\sum_{k=1}^{n} \sigma_j^k \frac{\partial}{\partial x^k} \right) f(X_s) \mathrm{d}B_s^j + \int_0^t (Lf)(X_s) \mathrm{d}s \ .$$

Define, for every f

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (Lf)(X_s) \mathrm{d}s \ .$$

Then, for every $f \in C_b^2(\mathbb{R})$

$$M_t^f = \int_0^t \sum_{j=1}^{m} \left(\sum_{k=1}^{n} \sigma_j^k \frac{\partial}{\partial x^k} \right) f(X_s) \mathrm{d}B_s^j$$

is a martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and

$$\begin{aligned} \langle M^f, M^g \rangle_t &= \int_0^t \sum_{j=1}^{m} \left(\sum_{k,l=1}^{n} \sigma_j^l \sigma_j^k \frac{\partial f}{\partial x^k} \frac{\partial g}{\partial x^l} \right) (X_s) \mathrm{d}s \\ &= \int_0^t \left(\sum_{k,l=1}^{n} a^{kl} \frac{\partial f}{\partial x^k} \frac{\partial g}{\partial x^l} \right) (X_s) \mathrm{d}s \ . \end{aligned}$$

Therefore we have proved

Proposition 5.4.1 *If $(X_t)_{t\geq 0}$ is a strong solution to SED (5.13) on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ (with a given Brownian motion), then for any $f \in C_b^2(\mathbb{R})$*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (Lf)(X_s) ds$$

is a martingale under the probability P, where L is defined by (5.14).

For example, if $\sigma_j^i = \delta_{ij}$ and $b^i = 0$ (in this case $L = \frac{1}{2}\Delta$), then $(B_t)_{t\geq 0}$ is a strong solution to

$$dX_t = dB_t$$

and

$$M_t^f = f(B_t) - f(B_0) - \frac{1}{2}\int_0^t (\Delta f)(B_s)\mathrm{d}s$$

is a martingale under P. On the other hand, Lévy's martingale characterization shows that the previous property that

$$f(B_t) - f(B_0) - \frac{1}{2}\int_0^t (\Delta f)(B_s)\mathrm{d}s$$

is martingale, in particular, implies that X_t^j and $X_i^j X_t^i - \delta_{ij}t$ are martingales, and completely characterizes Brownian motion. Therefore we may believe that the martingale property of M^f for all f should completely determine the distribution of a solution $(X_t)_{t\geq 0}$ to SDE (5.13), and hence those of weak solution of (5.13). Thus we give

Definition 5.4.2 *Let L be a linear operator on $C^\infty(\mathbb{R}^n)$. Let $(X_t)_{t\geq 0}$ be a continuous stochastic process on some $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Then we say that $(X_t)_{t\geq 0}$ together with the probability P is a solution to the L-martingale problem, if for every $f \in C_b^\infty(\mathbb{R}^n)$*

$$M_t^f \equiv f(X_t) - f(X_0) - \int_0^t Lf(X_s)ds$$

is a local martingale under the probability P.

Therefore a strong solution $(X_t)_{t\geq 0}$ of SDE (5.13) on $(\Omega, \mathcal{F}, P)$ is a solution to L-martingale problem, where L is given by (5.14) and

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Lf(X_s)\mathrm{d}s$$

is a martingale under P. Moreover, since

$$L(fg) - f\,(Lg) - g\,(Lf) = \sum_{k,l=1}^n a^{kl}\frac{\partial f}{\partial x^k}\frac{\partial g}{\partial x^l}$$

we thus have

$$\langle M^f, M^g\rangle_t = \int_0^t \{L(fg) - f\,(Lg) - g\,(Lf)\}\,(X_s)\mathrm{d}s\ .$$

Conversely, we can show that any solution to the L-martingale problem is a weak solution to SDE. Let us consider the one-dimensional case.

Theorem 5.4.3 *Let $b(\cdot)$ and $\sigma(\cdot)$ are Borel measurable on $\mathbb{R}$ which are bounded on any compact subset, and $\lambda^{-1} \leq \sigma(\cdot) \leq \lambda$ for some constant $\lambda > 0$. Let*

$$L = \frac{1}{2}\sigma(x)^2\frac{d^2}{dx^2} + b(x)\frac{d}{dx}.$$

If $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is a continuous process solving the L-martingale problem: for any $f \in C_b^2(\mathbb{R})$

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Lf(X_s)ds$$

is a continuous local martingale, then $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is a weak solution to SDE

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt \ . \tag{5.15}$$

Let us outline the proof only, a detailed proof will be given in additional topics in next chapter that handles the multi-dimensional case. To show that $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is a weak solution, we need to construct a Brownian motion $B = (B_t)_{t\geq 0}$ such that

$$X_t = X_0 + \int_0^t \sigma(X_s)\mathrm{d}B_s + \int_0^t b(X_s)\mathrm{d}s \ . \tag{5.16}$$

The key of the proof is to compute $\langle X\rangle_t$, and the result is

$$\begin{aligned} \left\langle M^f, M^g \right\rangle_t &= \int_0^t (L(fg) - fLg - gLf)(X_s)\mathrm{d}s \\ &= \int_0^t \left(\sigma^2 \frac{\partial f}{\partial x}\frac{\partial g}{\partial x}\right)(X_s)\mathrm{d}s \ . \end{aligned}$$

In particular, if we choose $f(x) = x$ the coordinate function (and write in this case M^f as M), then

$$\langle M\rangle_t = \int_0^t (\sigma(X_s))^2 \,\mathrm{d}s$$

so that

$$B_t = \int_0^t \frac{1}{\sigma(X_s)}\mathrm{d}M_s$$

is a Brownian motion (Lévy's martingale characterization for Brownian motion). It is then obvious that (X_t, B_t) satisfies the stochastic integral equation (5.16), and therefore $(X_t)_{t\geq 0}$ is a weak solution to (5.15).

5.5 Additional topics

Flows of diffeomorphisms

If the coefficients of a stochastic differential equation are bounded and smooth, then its strong solution defines a flow of diffeomorphisms. More precisely, let us consider the following stochastic differential equation on $\mathbb{R}^n$

$$dX_t^j = f_0^j(X_t)dt + \sum_{i=1}^n f_i^j(X_t)dW_t^i \tag{5.17}$$

where f_k^j ($j = 1, \cdots, n$ and $k = 0, 1, \cdots, n$), and $W = (W_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^n$ on a complete probability space $(\Omega, \mathcal{F}, P)$. Then there exists a mapping $X : [0, \infty) \times \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ satisfying the following conditions.

1) For $x \in \mathbb{R}^n$, $(X_t)_{t\geq 0}$, where $X_t(\omega) = X(t, x, \omega)$, is a strong solution of (5.17).

2) For $t \geq 0$, $\omega \in \Omega$, the mapping $x \to X(t, x, \omega)$ is a diffeomorphism on $\mathbb{R}^n$, and $x \to X(0, x, \omega)$ is the identity mapping.

Let $Z_k^l(t, x, \omega) = \frac{\partial}{\partial x^k} X^l(t, x, \omega)$. Then Z_k^j solves a stochastic differential equation which can be obtained by formly differentiating the previous equation (5.17). Namely, (X^j, Z_k^l) is the (unique) strong solution of the following stochastic differential equation

$$\begin{cases} dX(t,x,\cdot)^j = f_0^j(X(t,x,\cdot))dt + \sum_{i=1}^n f_i^j(X(t,x,\cdot))dW_t^i, \\ dZ_k^l(t,x,\cdot) = \frac{\partial f_0^l}{\partial x^\alpha}\Big|_{X(t,x,\cdot)} Z_k^a(t,x,\cdot)dt + \sum_{i=1}^n \frac{\partial f_i^l}{\partial x^\alpha}\Big|_{X(t,x,\cdot)} Z_k^a(t,x,\cdot)dW_t^i \end{cases} \tag{5.18}$$

with initial $X(0, x, \cdot) = x$ and $Z_k^l(0, x, \cdot) = \delta_k^l$. Clearly the second equation (5.18) is obtained by taking derivative in x^k both sides of (5.17). Moreover, this procedure can carry on as long as the stochastic differential equations obtained are well posed.

We will prove that the strong solution $(X(t, x, \cdot))_{t\geq 0}$ of (5.17) is a Markov process (see Theorem 6.5.4 below) with transition semigroup $(P_t)_{t\geq 0}$. Moreover

$$P_t u(x) = E\left(u(X(t, x, \cdot))\right).$$

Suppose u has bounded derivative, then, according to the Lebesgue convergence theorem,

$$\begin{aligned} \frac{\partial}{\partial x^k} P_t u(x) &= E\left\{\frac{\partial}{\partial x^k} u(X(t,x,\cdot))\right\} \\ &= \sum_{l=1}^n E\left\{\frac{\partial u}{\partial x^l}\Big|_{X(t,x,\cdot)} \frac{\partial X^l(t,x,\cdot)}{\partial x^k}\right\} \\ &= \sum_{l=1}^n E\left\{\frac{\partial u}{\partial x^l}\Big|_{X(t,x,\cdot)} Z_k^l(t,x,\cdot)\right\} \end{aligned}$$

which gives the probabilistic representation for $\frac{\partial}{\partial x^k} P_t$.

Theory of rough paths

In [], see also [ ], T. Lyons has developed a new approach which allows us to obtain strong solutions for all stochastic differential equations with one exceptional set with probability, and extends the research territory of stochastic analysis beyond the semi-martingale setting.

It is believed for quite long time that in order to solve the stochastic differential equation

$$dX_t = f_0^j(t, X_t)dt + \sum_{i=1}^n f_i^j(t, X_t) \circ dW_t^i \tag{5.19}$$

it is enough to know the sample path $t \to W_t$ and its Lévy area process

$$A^{k,l}_{s,t} = \int_s^t W^k_r dW^l_r - \int_s^t W^l_r dW^k_r \ ,$$

or roughly speaking, the strong solution to (5.19) is a "continuous" function of $(W_\cdot, A_\cdot)$, see [ ] for some results along this line. In [], T. Lyons introduced a new concept of rough paths, and proved that the strong solution of (5.19) is indeed, under certain conditions on the regularity about the coefficients f^j_k, a continuous function of the Brownian path together with its Lévy area with respect to the so-called p-variation distance. We must point out that the theory of rough paths, established in [ ], [], not only applies to stochastic differential equations driven by Brownian motion, but also to a large class of stochastic processes which are not semi-martingales.

Suppose $w : [0, T] \to \mathbb{R}^n$ is a continuous path with finite variation on $[0, T]$, then we can construct the iterated path integrals of any order by means of Riemann sums over any interval $[s, t] \subset [0, T]$. Namely, the first order iterated integral, denoted by $W^1_{s,t}$ is the increment of w over $[s, t]$, i.e. $W^1_{s,t} = w(t) - w(s)$, and the k-th order iterated integral, denoted by $W^k_{s,t}$, can be defined by induction

$$W^k_{s,t} = \lim_{m(D)\to 0} \sum_j W^{k-1}_{s,t_{j-1}} \otimes W^1_{t_{j-1},t_j}$$

where $D : s = t_0 < \cdots < t_m = t$ runs over finite partitions of $[s, t]$. In other words

$$\begin{aligned} W^k_{s,t} &= \int \cdots \int_{s<t_1<\cdots<t_k<t} dw(t_1) \otimes \cdots \otimes dw(t_k) \\ &= \left(\int \cdots \int_{s<t_1<\cdots<t_k<t} dw^{j_1}(t_1) \cdots dw^{j_k}(t_k) \right)_{j_1,\cdots,j_k \le n} . \end{aligned}$$

For example, $W^2_{s,t}$ is the matrix-valued process (index by two parameters s, t)

$$\left(\iint_{s<t_1<t_2<t} dw^{j_1}(t_1) dw^{j_2}(t_2) \right)_{j_1,j_2 \le n}$$

which can be decomposed into a symmetric matrix with entry

$$\begin{aligned} S^{j_1,j_2}_{s,t} &= \frac{1}{2} \iint_{s<t_1<t_2<t} dw^{j_1}(t_1) dw^{j_2}(t_k) + \frac{1}{2} \iint_{s<t_1<t_2<t} dw^{j_1}(t_1) dw^{j_2}(t_k) \\ &= \left(w^{j_2}(t) - w^{j_2}(s) \right) \left(w^{j_1}(t) - w^{j_1}(s) \right) \end{aligned}$$

and the skew-symmetric part with entry

$$A^{j_1,j_2}_{s,t} = \frac{1}{2} \iint_{s<t_1<t_2<t} dw^{j_1}(t_1) dw^{j_2}(t_k) - dw^{j_1}(t_1) dw^{j_2}(t_k)$$

which is called the Lévy area enclosed by the planer curve (w^{j_1}, w^{j_2}) and the line connecting the two end points $(w^{j_1}(s), w^{j_2}(s))$ and $(w^{j_1}(t), w^{j_2}(t))$.

Clearly, the symmetric part $S_{s,t}^{j_1,j_2}$ is a continuous functional of the path w (for example under the uniform norm on paths), but it is quite easy to construct a sequence paths which converges uniformly a continuous path but the corresponding sequence of Lévy areas diverges to infinity. That is, the Lévy area is not a continuous functional of the path. T. Lyons introduced the p-variation distances which take care of not only the path but also its iterated integrals. More precisely, for given $p \geq 1$, the p-variation distance between two variational paths w and x as the following

$$d_p(x,w) = \sup_{[0,T]} |x(t)-w(t)| + \max_{k\leq[p]} \sup_{0=t_0<\cdots<t_m=T} \left(\sum_{j=1}^{m} \left| X_{t_{j-1},t_j}^k - W_{t_{j-1},t_j}^k \right|^{p/k} \right)^{k/p} \tag{5.20}$$

where $[p]$ denotes the integer part of p, the second sup takes over all finite partitions of $[0,T]$. Now we can state the main result in the theory of rough paths. Consider the ordinary differential equation

$$dx(t) = f_0^j(t, x(t))dt + \sum_{i=1}^{n} f_i^j(t, x(t))dw^i(t) \tag{5.21}$$

with initial $x_0 = o \in \mathbb{R}^n$, where $w = (w(t))$ is a continuous path in $\mathbb{R}^n$ with finite variation on $[0,T]$, and f_i^j ($j \leq N$, $i = 0, \cdots, n$) are Lipschitz continuous. The unique solution $(x_t)_{t\geq 0}$ to (5.21) depends on the path w (and also the initial o which will be fixed), so we denote the solution by $F(w)$ which is thus a continuous path in $\mathbb{R}^N$ with finite variation $[0,T]$. F is called the Itô map defined by (5.21).

Theorem 5.5.1 *(T. Lyons [ ]) Under above notations and assumptions. Let $p \geq 1$. Suppose in addition that all f_i^j belong to C_b^γ where γ is an integer greater than p, then the Itô map $w \to F(w)$ defined by (5.21) is continuous with respect to the p-variation distance.*

Therefore, for every $p > 1$, one may extend the Itô map to the paths in the closure of the space of all variational paths inder the p-variation distance. Any element in this closure is called a rough path with roughness p. It turns out, for any p, as long as $p > 2$, almost all sample Brownian motion paths are rough paths, for more details the reader should consult [ ] etc.

Chapter 6

Markov processes

In this chapter we shall develop the fundamental theory of Markov processes. A stochastic process indexed by a subset of the real line has the Markov property if, roughly speaking, the past and the future of the process are independent conditioned on the present.

In many applications, we have a family of measures which depend on a random parameter. The typical example is the following. If $(X_t)_{t\geq 0}$ is a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$ in a state space M, then for fixed $t > 0$, the so-called *regular conditional probability*

$$K(x, dz) = P\left(X_t \in dz | X_0 = x\right)$$

is a transition kernel, if this family of measures is measurable in the parameter x.

6.1 Transition semigroups

6.1.1 Kernels and their associated operators

Let $(M, \mathcal{M})$ and $(N, \mathcal{N})$ be two measurable spaces. A function $K : N \times \mathcal{M} \to [0, \infty]$ is a kernel from $(M, \mathcal{M})$ to $(N, \mathcal{N})$ if

1) For every $x \in N$, $K(x, \cdot) : \mathcal{M} \to [0, \infty]$ is a σ-finite measure on $(M, \mathcal{M})$.

2) For any $A \in \mathcal{M}$, $x \to K(x, A)$ is a measurable function on $(M, \mathcal{M})$.

For every $x \in X$, $K(x, dz)$ denotes the σ-finite measure $K(x, \cdot) : \mathcal{M} \to [0, \infty]$. In the case that $(M, \mathcal{M}) = (N, \mathcal{N})$, then we say K is a kernel on $(M, \mathcal{M})$.

A kernel K from $(M, \mathcal{M})$ to $(N, \mathcal{N})$ is a *Markov kernel* (resp. *sub-Markov kernel*, resp. *bounded kernel*) if for every $x \in X$, $K(x, M) = 1$ (resp. $N(x, M) \leq 1$, resp. $\sup_{x\in X} N(x, M) < +\infty$).

Let K be a *bounded* kernel from $(M, \mathcal{M})$ to $(N, \mathcal{N})$. Then

$$(Kf)(x) = \int_M f(z)K(x, dz) \qquad \forall f \in b\mathcal{M} \text{ or } \mathcal{M}^+$$

define a linear operator from $b\mathcal{M}$ to $b\mathcal{N}$, again denoted by K, which is bounded under the maximal norms. Similarly, if μ is a σ-finite measure on $(N, \mathcal{N})$, then define

$$(\mu K)(B) = \int_N K(x, B)\mu(dx)$$

which sends a σ-finite measure μ on $(N, \mathcal{N})$ to a σ-finite measure μK on $(M, \mathcal{M})$.

Sub-Markov kernel

Any sub-Markov kernel K on $(M, \mathcal{M})$ may be extended to a Markov kernel, denoted by $\tilde{K}$, on the extended space $M_\partial = M \cup \{\partial\}$ by adding a point $\partial \notin M$ into M. The natural σ-algebra on M_∂, denoted by $\mathcal{M}_\partial$, is the smallest σ-algebra generated by $\mathcal{M}$ and $\{\partial\}$. $\tilde{K}$ is a kernel on $(M_\partial, \mathcal{M}_\partial)$ defined by

$$\tilde{K}(x, A) = K(x, A) \qquad \forall x \in M \text{ and } A \in \mathcal{M}$$

and

$$\tilde{K}(x, \{\partial\}) = 1 - K(x, M)\ ; \qquad \tilde{K}(\partial, \cdot) = \delta_\partial(\cdot).$$

Then obviously $\tilde{K}$ is a Markov kernel on the extended state space $(M_\partial, \mathcal{M}_\partial)$ that coincides with K on $(M, \mathcal{M})$.

6.1.2 Transition functions

The concept of transition functions is the generalization of transition matrices. A family $\{P_{s,t} : 0 \leq s < t\}$ of sub-Markov kernels on a measurable space $(M, \mathcal{M})$ is called a *transition function* on $(M, \mathcal{M})$, if $\{P_{s,t} : 0 \leq s < t\}$ satisfies the Chapman-Kolmogorov equation

$$P_{s,u}(x, A) = \int_M P_{s,t}(x, dz)P_{t,u}(z, A) \qquad \forall A \in \mathcal{M}$$

for every $x \in M$ and $0 \leq s < t < u$.

Sometimes a transition function $P_{s,t}(x,dy)$ is written as $P(s, x; t, dy)$ (for $x \in M$ and $0 \leq s < t$), which represents the probability that a Markov process $(X_t)_{t\geq 0}$ starting at x at time s enters dy at time t:

$$P(s, x; t, dy) = P\left(X_t \in dy | X_s = x\right)\ .$$

A transition function $\{P_{s,t} : 0 \leq s < t\}$ is homogenous if $P_{s,t}$ depends only on $t - s$ for any $t > s \geq 0$. In this case we may define $P_t \equiv P_{0,t}$, and then for $t > s$ we have $P_{s,t} = P_{t-s}$. We often write $P_t(x, dy)$ as $P(t, x, dy)$. The Chapman-Kolmogorov equation can be written as

$$P(s + t, x, A) = \int_M P(s, x, dz)P(t, z, A) \qquad \forall A \in \mathcal{M}.$$

Proposition 6.1.1 *Let $\{P_t : t > 0\}$ be a (homogenous) transition function on $(M, \mathcal{M})$. Then*

1) $(P_t)_{t>0}$ is a contraction semigroup on the Banach space $(b\mathcal{M}, ||\cdot||_\infty)$.

2) For every $t > 0$, P_t preserves positivity.

Proof. If $f \in b\mathcal{M}$ and $t > 0$ then

$$\begin{aligned} P_t(P_s f)(x) &= \int_M P_s f(z) P(t, x, dz) \\ &= \int_M \left\{ \int_M f(y) P(s, z, dy) \right\} P(t, x, dz) \\ &= \int_M \left\{ f(y) \int_M P(s, z, dy) P(t, x, dz) \right\} \\ &= \int_M f(y) P(s+t, x, dy) \\ &= P_{t+s} f(x) \ , \end{aligned}$$

and

$$\begin{aligned} |P_t f(x)| &= \left| \int_M f(z) P(t, x, dz) \right| \\ &\leq ||f||_\infty \int_M P(t, x, dz) \\ &= ||f||_\infty \ . \end{aligned}$$

Thus $(P_t)_{t>0}$ is a contraction semigroup. It is obvious that $P_t f \geq 0$ if f is non-negative. ∎

Therefore a homogenous transition probability function is also called a transition semigroup. Set $P_0 = I$ if necessary.

6.1.3 Feller semigroups

We begin with some notations. If M is a topological space, then $C(M)$ (resp. $C_b(M)$) denotes the vector space of (resp. bounded) continuous functions on M. $C_b(M)$ is a Banach space under the uniform convergence norm $||f||_\infty = \sup_{x \in M} f(x)$.

If M is a locally compact metric space, then it is often useful to consider its one-point compactification, called the Alexandroff compacification of M. That is, we choose a point $\partial \notin M$, and form an extended space $M_\partial = M \cup \{\partial\}$. Equip M_∂ with the topology consisting of all open sets in M together with all complements of compact sets in M, so that ∂ becomes the point at infinity. According to this topology, a function f defined on M has a limit l at infinity, written as $\lim_{x \to \partial} f(x) = l$, if for every $\varepsilon > 0$ there is a compact subset $K \subset M$ such that $|f(x) - l| < \varepsilon$ whenever $x \in M \backslash K$. Continuous functions on M_∂ can be identified with continuous functions on M which have limits at infinity. Let

$$C_0(M) = \left\{ f \in C(M) : \lim_{x \to \partial} f(x) = 0 \right\}$$

which is a closed subspace of $C_b(M)$. If M is a compact metric space, then $C_0(M)$ coincides with $C(M)$.

Definition 6.1.2 *A transition function* $\{P(t,x,dy) : t > 0\}$ *is called a* Feller semigroup *if*

1) For every $t > 0$, $C_0(M)$ *is invariant under* P_t, *that is* $P_t f \in C_0(M)$ *for every* $f \in C_0(M)$.

2) For any $f \in C_0(M)$ *and* $x \in M$, $\lim_{t \downarrow 0} P_t f(x) = f(x)$.

In this case we also say $(P_t)_{t \geq 0}$ *has the Feller property.*

The statement of the following theorem is important, but its proof uses a few results in the functional analysis.

Theorem 6.1.3 *Let* $\{P(t,x,dy) : t > 0\}$ *be a Feller semigroup on a* locally compact metric space M. *Then*

$$\lim_{t \downarrow 0} ||P_t f - f||_\infty = 0 \qquad \forall f \in C_0(M),$$

namely $(P_t)_{t \geq 0}$ *is a strongly continuous semigroup of contractions on the Banach space* $C_0(M)$.

Proof. We may assume that M is a compact metric space, otherwise consider M_∂ instead. If $f \in C(M)$, $P_s f \in C(M)$ for $s \geq 0$, and

$$\lim_{t \downarrow s} P_t f(x) = \lim_{r \downarrow 0} P_r(P_s f)(x) = P_s f(x) \qquad \forall x \in M \ , \ \ s \geq 0 \ .$$

That is $s \to P_s f(x)$ is right continuous. Let

$$R_\alpha f(x) = \int_0^\infty \mathrm{e}^{-\alpha s} P_s f(x) \mathrm{d}s$$

for $\alpha > 0$. Each αR_α is a contraction on $C(M)$, and, according to Lebesgue's dominated convergence theorem

$$\lim_{\alpha \to +\infty} \alpha R_\alpha f(x) = f(x) \quad \forall f \in C(M) \ .$$

We first show that the range of R_α is dense in $C(M)$. It is straightforward to verify the resolvent equation

$$R_\alpha - R_\beta = -(\alpha - \beta) R_\beta R_\alpha$$

which implies that the range of R_α does not depend on $\alpha > 0$. According to Hahn-Banach's theorem, it is sufficient to show that a continuous linear functional on $C(M)$ which is orthogonal to the range of R_α must be zero. Let

$\mu \in C(M)^*$ be a functional which vanishes on the range of R_α. Then μ is a finite signed measure on M, and it follows that

$$\begin{aligned} 0 &= \mu(\alpha R_\alpha f) = \int_M \alpha R_\alpha f(x)\mu(dx) \\ &\to \int_M f(x)\mu(dx) = \mu(f) \qquad\qquad \forall f \in C(M) \ . \end{aligned}$$

Therefore $\mu = 0$. Now for $\alpha > 0$ and $t > 0$ we have

$$\begin{aligned} P_t(R_\alpha f) &= \int_0^\infty e^{-\alpha s}(P_{t+s}f)ds \\ &= e^{\alpha t}\left(R_\alpha f - \int_0^t e^{-\alpha s}P_s f ds\right) \\ &\to R_\alpha f \end{aligned}$$

uniformly on M as $t \downarrow 0$. Since $\{R_\alpha f : f \in C(M)\}$ is dense in $C(M)$, $P_t f \to f$ uniformly on M as $t \downarrow 0$. ■

6.1.4 Examples

We now present several examples of Markov semigroups.

1. *Transition matrices.* Let M be a finite or countable set, and let $(p_{ij}(t))$ be a Markov matrix with state space M. Then

$$P(t, i, A) = \sum_{j \in A} p_{ij}(t)$$

is a Markov semigroup. If M is finite, then $\{P_t : t > 0\}$ is Feller if and only if

$$\lim_{t \downarrow 0} p_{ij}(t) = \delta_{ij} \ .$$

If M is a countable infinite set, then $\{P_t : t > 0\}$ is Feller if and only if for every $t > 0$ and $i \in M$

$$\lim_{j \to \infty} p_{ij}(t) = 0$$

and for every $i \in M$

$$\lim_{t \downarrow 0} p_{ij}(t) = \delta_{ij} \qquad \text{uniformly in } j \in M \ .$$

2. *Convolution semigroups.* Let $\{\mu_t : t > 0\}$ be a family of probability measures on the Euclidean space $\mathbb{R}^n$, which is called a convolution semigroup if it satisfies the following conditions

$$\mu_{t+s}(A) = \int_{\mathbb{R}^n} 1_A(x+y)\mu_s(\mathrm{d}x)\mu_t(\mathrm{d}y) \qquad \forall A \in \mathcal{B}(\mathbb{R}^n)$$

and $\mu_t \to \delta_0$ weakly as $t \downarrow 0$. Define

$$\begin{aligned} P(t,x,A) &= \int_{\mathbb{R}^n} 1_A(y+x)\mu_t(\mathrm{d}y) \\ &= \mu_t(A-x) \qquad\qquad \forall t>0 \text{ and } x \in \mathbb{R}^n \ . \end{aligned}$$

Then $\{P(t,x,A) : t > 0\}$ is a Markov semigroup, and P_t leaves $C_b(\mathbb{R}^n)$ invariant. Actually $(P_t)_{t\geq 0}$ is Feller semigroup. In later case $(P_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on $(C_b(\mathbb{R}^n), ||\cdot||_\infty)$.

Exercise 6.1 *Let μ be a probability measure on $\mathbb{R}^n$ and*

$$\mu_t = \sum_{k\geq 0} -\lambda t \frac{(\lambda t)^k}{k!}\mu^{*k},$$

where $*k$ denotes k-th convolution power. Prove that $\{\mu_t\}$ is a convolution semigroup which is called compound Poisson semigroup.

3. *Heat kernel on $\mathbb{R}^n$.* The heat kernel on $\mathbb{R}^n$ is the minimal fundamental solution $p(t,x)$ to the heat equation

$$\left(\Delta - \frac{\partial}{\partial t}\right) u(t,x) = 0 \ .$$

It is known that

$$p(t,x) = \frac{1}{(4\pi t)^{n/2}} \exp\left\{-\frac{|x|^2}{4t}\right\} \qquad \forall t > 0$$

and we however may regard this formula as the definition of the heat kernel on $\mathbb{R}^n$ Define $p(t,x,y) = p(t,y-x)$ for any $t \geq 0, x,y \in \mathbb{R}^n$ and

$$P(t,x,\mathrm{d}y) = p(t,x,y)\mathrm{d}y \ .$$

Then $\{P(t,x,\mathrm{d}y) : t > 0\}$ is a Feller semigroup on $\mathbb{R}^n$.

Exercise 6.2 *Let (P_t) be the heat kernel on $\mathbb{R}$. Let $E = \mathbb{R} \setminus \{0\}$ with its Borel σ-algebra $\mathcal{E}$. Define for any $x \in E$ and $B \in \mathcal{E}$,*

$$\hat{P}_t(x,B) = P_t(x,B).$$

Show that (P_t) is a transition semigroup but not Feller.

4. *The Ornstein-Uhlenbeck semigroup.* Let μ be the standard Gaussian measure on $\mathbb{R}^n$:

$$\mu(f) = \int_{\mathbb{R}^n} f(y)\frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{|y|^2}{2}\right\} \mathrm{d}y \ .$$

Define

$$P_t f(x) = \int_{\mathbb{R}^n} f\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}y\right)\mu(\mathrm{d}y) \ .$$

One can show that $(P_t)_{t\geq 0}$ is a Feller semigroup on $\mathbb{R}^n$.

6.2 Markov property

A stochastic process indexed by a subset of the real line has the Markov property if, roughly speaking, the past and the future of the process are independent conditioned on the present.

6.2.1 Simple Markov property

Let $(\Omega, \mathcal{G}, \mathcal{G}_t, P)$ be a filtered probability space, and let $(X_t)_{t\geq 0}$ be an adapted process with values in a measurable space $(M, \mathcal{M})$.

Definition 6.2.1 *The process* $(X_t)_{t\geq 0}$ *is said to have* Markov property *with respect to filtration* $(\mathcal{G}_t)_{t\geq 0}$ *if*

$$E\left(f(X_t)|\mathcal{G}_s\right) = E\left(f(X_t)|X_s\right) \qquad a.e.\text{-}P \tag{6.1}$$

for any $t > s$ *and for any bounded measurable function* f *on* M*.*

The equation (6.1) means that there is a measurable function g on M such that for every $A \in \mathcal{G}_s$

$$E\left(1_A \cdot f(X_t)\right) = E\left(1_A \cdot g(X_s)\right) . \tag{6.2}$$

For each $t \geq 0$ let

$$\mathcal{F}_t^0 = \sigma\left\{X_s : s \leq t\right\} \quad \text{and} \quad \mathcal{F}_t' = \sigma\left\{X_s : s \geq t\right\} .$$

Intuitively, an event in $\mathcal{F}_t^0$ is determined by the process up to time t, and those in $\mathcal{F}_t'$ after time t. Thus they represent respectively the past and the future of the process relative to the present instant t.

Theorem 6.2.2 *A process* $(X_t)_{t\geq 0}$ *has Markov property with respect to* $(\mathcal{G}_t)_{t\geq 0}$ *if and only if one of the following conditions is satisfied:*

1) For every $t \geq 0$*,* $A \in \mathcal{G}_t$ *and* $B \in \mathcal{F}_t'$ *we have*

$$P\left(A \cap B|X_t\right) = P\left(A|X_t\right) P\left(B|X_t\right) .$$

That is, the past and the future are independent conditioned on the present.

2) For every $t \geq 0$ *and* $B \in \mathcal{F}_t'$

$$P\left(B|\mathcal{G}_t\right) = P\left(B|X_t\right) .$$

3) For every $t \geq 0$ *and* $A \in \mathcal{G}_t$

$$P\left(A|\mathcal{F}_t'\right) = P\left(A|X_t\right) .$$

The second and the third claims indicate the Markov property is symmetric with respect to the past and the future.

Proof. It is evident that item 2) implies the Markov property (6.2). We prove that (6.2) implies 2). In fact, for $h \in \mathcal{F}_t'$ we show that

$$E\left(h|\mathcal{G}_t\right) = E\left(h|X_t\right) \ . \tag{6.3}$$

By the monotone class theorem, we only need to prove (6.3) for

$$h = f_1(X_{t_1}) \cdots f_n(X_{t_n})$$

where $t \leq t_1 < t_2 < \cdots < t_n$ and f_i are bounded measurable functions on M. The case that $n = 1$ and $t_1 = t$ is trivial. Let us use induction on n. If $n = 1$ and $t_1 > t$, then 2) reduces to the Markov property (6.1). Suppose now $n > 1$ and 2) is true for $n - 1$. Define $h' = E\left(h|\mathcal{G}_{t_{n-1}}\right)$. Then h' is a function of $X_{t_{n-1}}$ so that it may be represented as $g(X_{t_{n-1}})$ where h is a Borel measurable, bounded function, and thus

$$E(h|\mathcal{G}_t) = E\left(h'|\mathcal{G}_t\right) \ ; \quad E(h|X_t) = E(h'|X_t) \ .$$

On the other hand,

$$h' = f_1(X_{t_1}) \cdots (f_{n-1}g)(X_{t_{n-1}}).$$

By induction assumption we have

$$E\left(h'|\mathcal{G}_t\right) = E(h'|X_t)$$

and therefore

$$E(h|\mathcal{G}_t) = E(h|X_t) \ .$$

Hence (6.3) holds for any h.

1)$\Longrightarrow$2) and 3). For $A \in \mathcal{G}_t$ and $B \in \mathcal{F}_t'$ we have

$$\begin{aligned} E\left(1_A P(B|X_t)|X_t\right) &= P(B|X_t)E(1_A|X_t) \\ &= P(B|X_t)P(A|X_t) \\ &= P(A \cap B|X_t) \end{aligned}$$

and taking expectation

$$E\left(1_A P(B|X_t)\right) = P(A \cap B) = E\left(1_A 1_B\right)$$

which yields that

$$P(B|\mathcal{G}_t) = P(B|X_t) \ .$$

Similarly

$$E\left(1_B P(A|X_t)\right) = P(A \cap B)$$

and therefore

$$P(A|X_t) = P(A|\mathcal{F}_t') \ .$$

Conversely, let us for instance show that 2)$\Longrightarrow$1). If $A \in \mathcal{G}_t$ and $B \in \mathcal{F}'_t$, then by 2)

$$\begin{aligned} P(A \cap B|X_t) &= E\left(E\left(1_A 1_B|\mathcal{G}_t\right)|X_t\right) \\ &= E\left(1_A E\left(1_B|\mathcal{G}_t\right)|X_t\right) \\ &= E\left(1_A P(B|X_t)|X_t\right) \\ &= P(B|X_t)E\left(1_A|X_t\right) \\ &= P(B|X_t)P\left(A|X_t\right) \ . \end{aligned}$$

■

The item 2) of Theorem 6.2.2 is the most useful form of the Markov property, and is equivalent to the following: for any $t \geq 0$ and for every $Y \in b\mathcal{F}'_t$

$$E\left(Y|\mathcal{G}_t\right) = E\left(Y|X_t\right) \ . \tag{6.4}$$

It is also equivalent to

$$E\left(YZ|X_t\right) = E\left(Y|X_t\right)E\left(Z|X_t\right)$$

for every $Y \in \mathcal{G}_t$ and $Z \in \mathcal{F}'_t$.

In order to verify the Markov property with respect to the natural filtration $(\mathcal{F}^0_t)_{t\geq 0}$:

$$E\left(f(X_t)|\mathcal{F}^0_s\right) = E\left(f(X_t)|X_s\right) \qquad \text{a.e.}$$

it suffices to verify that

$$E\left(f(X_t)|X_{t_1}, \cdots, X_{t_n}, X_s\right) = E\left(f(X_t)|X_s\right)$$

for $0 = t_1 < t_2 < \cdots < t_n \leq s < t$ and for any bounded Borel measurable function f. This follows from an argument of the monotone class theorem. In particular, a stochastic process which has the same law as that of a process having Markov property has Markov property with respect to the natural filtration.

Lemma 6.2.3 *Let $(X_t)_{t\geq 0}$ be a Markov process with respect to $(\mathcal{G}_t)_{t\geq 0}$. Then $(X_t)_{t\geq 0}$ has Markov property with respect to its natural filtration $(\mathcal{F}^0_t)_{t\geq 0}$.*

This is due to the fact that $\mathcal{F}^0_t \subset \mathcal{G}_t$ for each $t \geq 0$. Therefore we may talk about Markov property without mentioning a filtration.

Lemma 6.2.4 *Let $(X_t)_{t\geq 0}$ be an adapted stochastic process on $(\Omega, \mathcal{G}, \mathcal{G}_t, P)$ with the state space $(M, \mathcal{M})$, and let $\{P_{s,t}(x, dy) : t > s \geq 0\}$ be a family of transition kernels on $(M, \mathcal{M})$, such that*

$$E\left(f(X_t)|\mathcal{G}_s\right) = P_{s,t}f(X_s) \qquad \forall t > s \geq 0 \ , \ f \in b\mathcal{M} \ . \tag{6.5}$$

Then $(X_t)_{t\geq 0}$ possesses Markov property. In this case we call $(X_t)_{t\geq 0}$ a Markov process with transition function $\{P_{s,t} : t > s \geq 0\}$.

Proof. Indeed

$$\begin{aligned} E\left(f(X_t)|X_s\right) &= E\left(E\left(f(X_t)|\mathcal{G}_s\right)|X_s\right) \\ &= E\left(P_{s,t}f(X_s)|X_s\right) \\ &= P_{s,t}f(X_s) \\ &= E\left(f(X_t)|\mathcal{G}_s\right) \ . \end{aligned}$$

■

If the transition semigroup is homogenous, i.e.

$$E\left(f(X_t)|\mathcal{G}_s\right) = P_{t-s}f(X_s) \quad \forall t > s \geq 0 \ , \ f \in b\mathcal{M} \tag{6.6}$$

for some transition function $(P_t)_{t>0}$, then $(X_t)_{t\geq 0}$ is called a homogenous Markov process with the transition function $(P_t)_{t>0}$. The Markov property implies that $(P_t)_{t>0}$ must satisfy the semigroup property

$$P_{t+s} = P_t P_s \ .$$

The Markov property (6.6) may be written as

$$E\left(f(X_{t+s})|\mathcal{G}_s\right) = P_t f(X_s) \quad \forall t, s \geq 0 \ , \ f \in b\mathcal{M} \tag{6.7}$$

which is called the *simple Markov property.*

6.2.2 Realizations of Markov semigroups

Let $(X_t)_{t\geq 0}$ be a Markov process on $(\Omega, \mathcal{G}, \mathcal{G}_t, P)$ with transition semigroup $(P_t)_{t>0}$ (see eqn. (6.7)) and let $\mu(dx)$ be the initial distribution of $(X_t)_{t\geq 0}$, i.e.

$$\mu(A) = P\left\{X_0 \in A\right\} \qquad \forall A \in \mathcal{M} \ .$$

Then the marginal distribution μ_t of X_t is given by $\mu_t(f) = \mu P_t$. Indeed

$$\begin{aligned} \mu_t(f) &= E(f(X_t)) = E(P_t f(X_0)) \\ &= \mu(P_t f) = \mu P_t(f) \ . \end{aligned}$$

The finite dimensional distribution $\mu_{t_0,\cdots,t_n}$ of X_{t_0}, X_{t_1}, $\cdots$, X_{t_n}:

$$\mu_{t_0,\cdots,t_n}(dx_0, \cdots, dx_n) = P\left\{X_{t_0} \in dx_0, \, X_{t_1} \in dx_1, \, \cdots, \, X_{t_n} \in dx_n\right\}$$

where $0 = t_0 < t_1 < \cdots < t_n$, can be determined in terms of the initial distribution μ and the transition semigroup $(P_t)_{t\geq 0}$. In order to write down a formula for the probability measure $\mu_{t_0,\cdots,t_n}$ on the product measurable space $(M^{n+1}, \mathcal{M}^{n+1})$, we compute

$$E(f(X_{t_0}, X_{t_1}, \cdots, X_{t_n})) \ .$$

By the monotone class theorem, we only need to do this functions in product forms

$$f(x_0, \cdots, x_n) = f_0(x_0) \cdots f_n(x_n)$$

where f_i are bounded measurable functions on $(M, \mathcal{M})$. In this case

$$\int_{M^{n+1}} f_0(x_0) \cdots f_n(x_n) \mu_{t_0,\cdots,t_n}(dx_0, \cdots, dx_n) = E(f_0(X_{t_0}) \cdots f_n(X_{t_n})).$$

By the Markov property at $t_n > t_{n-1}$ the right-hand side may be written as

$$\begin{aligned} E(f_0(X_{t_0}) \cdots f_n(X_{t_n})) &= E\left\{E\left(f_0(X_{t_0}) \cdots f_n(X_{t_n}) | \mathcal{G}_{t_{n-1}}\right)\right\} \\ &= E\left\{f_0(X_{t_0}) \cdots f_{n-1}(X_{t_{n-1}}) E\left(f_n(X_{t_n}) | \mathcal{G}_{t_{n-1}}\right)\right\} \\ &= E\left\{f_0(X_{t_0}) \cdots f_{n-1}(X_{t_{n-1}}) (P_{t_n - t_{n-1}} f_n)(X_{t_{n-1}})\right\} \\ &= E\left\{f_0(X_{t_0}) \cdots \tilde{f}_{n-1}(X_{t_{n-1}})\right\} \end{aligned}$$

where $\tilde{f}_{n-1} = f_{n-1}(P_{t_n - t_{n-1}} f_n)$. By repeating the same procedure we then arrive in the following

$$\begin{aligned} & E(f(X_{t_0}, X_{t_1}, \cdots, X_{t_n})) \\ = & \int_{M^{n+1}} f(x_0, \cdots, x_n) \mu(dx_0) P_{t_1}(x_0, dx_1) \\ & \qquad \times P_{t_2 - t_1}(x_1, dx_2) \cdots P_{t_n - t_{n-1}}(x_{n-1}, dx_n) \; . \end{aligned}$$

Therefore

$$\begin{aligned} & P\left\{X_{t_0} \in dx_0, \, X_{t_1} \in dx_1, \, \cdots, \, X_{t_n} \in dx_n\right\} \\ = & \mu(dx_0) P_{t_1}(x_0, dx_1) P_{t_2 - t_1}(x_1, dx_2) \cdots P_{t_n - t_{n-1}}(x_{n-1}, dx_n) \; . \end{aligned} \tag{6.8}$$

The family of finite-dimensional distributions

$$\{\mu_{t_0,\cdots,t_n} : 0 = t_0 < t_1 < \cdots < t_n\}$$

is called the finite dimensional distribution family of the Markov process $(X_t)_{t \geq 0}$, which is consistent in the sense that $\mu_{t_0,\cdots t_{j-1}, t_{j+1}, \cdots, t_n}$ (with t_j erased) coincides with the measure $\mu_{t_0,\cdots,t_n}$ by integrating the variable x_j over M.

Conversely, if we start with a Markov transition semigroup $(P_t)_{t>0}$ on $(M, \mathcal{M})$ and an initial distribution probability μ, then we may construct a finite dimensional distribution family via (6.8) which is consistent in the above sense. According to Kolmogorov's extension theorem, there is a stochastic process $(X_t)_{t \geq 0}$ on some probability space which has specified finite-dimensional distributions

$$\{\mu_{t_0,\cdots,t_n} : 0 = t_0 < t_1 < \cdots < t_n\} \; .$$

Clearly $(X_t)_{t \geq 0}$ is a Markov process with transition semigroup $(P_t)_{t>0}$.

More specifically, let Ω be the space of all paths $\omega : [0, +\infty) \to M$, let $(X_t)_{t \geq 0}$ be the coordinate process

$$X_t(\omega) = \omega(t) \qquad \forall \omega \in \Omega, \, t \geq 0,$$

let $\mathcal{F}^0$ and $(\mathcal{F}_t^0)_{t \geq 0}$ be generated by the coordinate process

$$\mathcal{F}^0 = \sigma\left\{X_t : t \geq 0\right\}, \;\; \mathcal{F}_t^0 = \sigma\left\{X_s : s \leq t\right\} \; .$$

Then, for any given initial distribution μ there exists a unique probability measure P^μ on $(\Omega, \mathcal{F}^0)$ such that $(X_t)_{t\geq 0}$ is a Markov process on $(\Omega, \mathcal{F}^0, \mathcal{F}^0_t, P^\mu)$ with transition semigroup $(P_t)_{t>0}$. P^μ is also called the distribution of the process $(X_t)_{t\geq 0}$ with initial distribution μ.

Any Markov process with transition semigroup $(P_t)_{t>0}$ is called a realization of the Markov semigroup $(P_t)_{t>0}$, and two Markov processes with the same state space are equivalent in law (i.e. they have the same family of finite dimensional distributions) if they have the same transition semigroup. Therefore any Markov semigroup has a realization on the space of paths Ω.

Given a transition semigroup $(P_t)_{t>0}$ on a state space $(M, \mathcal{M})$ and $x \in M$, we use P^x to denote P^{δ_x} for simplicity. The Markov process $(X_t)_{t\geq 0}$ with initial distribution $\mu = \delta_x$ is said to start at x. P^x is characterized by the following: for $0 < t_1 < \cdots < t_n$

$$\begin{aligned} & P^x\left(X_{t_1} \in dx_1, \cdots, X_{t_n} \in dx_n\right) \\ = \;& P_{t_1}(x, dx_1) \cdots P_{t_n - t_{n-1}}(x_{n-1}, dx_n) \; . \end{aligned} \tag{6.9}$$

The corresponding expectation will be denoted by E^x. Thus, if $Y \in b\mathcal{F}^0$, then

$$E^x(Y) = \int_\Omega Y(\omega) P^x(d\omega)$$

and

$$P_t(x, A) = E^x(1_A(X_t)) = P^x(X_t \in A).$$

It can be shown that if $Y \in \mathcal{F}^0$ then $x \to P^x(Y)$ is measurable on $(M, \mathcal{M})$. Indeed if $Y = X_t^{-1}(A)$ for some $A \in \mathcal{M}$, then the measurability of the mapping $x \to E^x(Y)$ follows from the defining property of transition kernels. The general case then follows by a monotone class argument.

In terms of notations we have established, the Markov property (6.7) can be written as

$$\begin{aligned} P\left(X_{t+s} \in A | \mathcal{F}_s\right) &= P^{X_s}\left(X_t \in A\right) \\ &= P_t(X_s, A) \qquad \forall t \geq 0,\, s > 0,\, A \in \mathcal{M} \; . \end{aligned} \tag{6.10}$$

If μ is the initial distribution of the Markov process $(X_t)_{t\geq 0}$, then

$$P^\mu(Y) = \int_M P^x(Y)\mu(dx) \qquad \forall Y \in b\mathcal{F}^0 \tag{6.11}$$

is the law of the Markov process $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}^0)$. Notice that (6.11) makes sense for any σ-finite measure μ on $(M, \mathcal{M})$. The right-hand side of (6.10) does not depend on the initial law of $(X_t)_{t\geq 0}$ and it may be stated as the following

$$P^\mu\left(X_{t+s} \in A | \mathcal{F}_s\right) = P_t(X_s, A) = P^{X_s}(X_t \in A) \qquad P^\mu\text{-a.s.} \tag{6.12}$$

for any probability measure μ on $(M, \mathcal{M})$, or

$$\begin{aligned} P^\mu\left(f(X_{t+s}) | \mathcal{F}_s\right) &= P^{X_s}(f(X_t)) \\ &= P_t f(X_s) \qquad P^\mu\text{-a.s.} \end{aligned}$$

For simplicity of notations, we often collect all items together and call

$$(\Omega, \mathcal{F}^0, \mathcal{F}_t^0, X_t, \theta_t, P^x)$$

a *canonical realization* of the Markov semigroup $(P_t)_{t>0}$, where $(\theta_t : t \geq 0)$ is shift operators on Ω, which satisfies

$$X_t \circ \theta_s = X_{t+s}, \ t, s \geq 0.$$

The shift operators, always exist at least on space of paths, make the Markov property look more simple and intuitive. For details of shift operators, see additional topics.

Given a σ-finite measure μ on $(M, \mathcal{M})$, $\mathcal{F}^\mu$ denotes the completion of $\mathcal{F}^0$ under P^μ, and let $\mathcal{F}_t^\mu$ denote the smallest σ-algebra containing $\mathcal{F}_t^0$ and all P^μ-null sets. Let $\mathcal{M}_1$ denote the set of all possible initial distributions on $(M, \mathcal{M})$. Let

$$\mathcal{F} = \bigcap_{\mu \in \mathcal{M}_1} \mathcal{F}^\mu \qquad \text{and} \quad \mathcal{F}_t = \bigcap_{\mu \in \mathcal{M}_1} \mathcal{F}_t^\mu \ .$$

This procedure is called the augmentation of $(\mathcal{F}_t^0)$. From now on, we call $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration of Markov process $(X_t)_{t\geq 0}$.

6.2.3 Markov processes in topological spaces

Let M be either a locally compact separable metric space or a Polish space (i.e. a complete, separable metric space), together with its Borel σ-algebra $\mathcal{B}(M)$ (we shall however use the shorter notation $\mathcal{B}$ in place of $\mathcal{B}(M)$ unless clearness dictates otherwise). The Euclidean space $\mathbb{R}^n$ of dimension n is a typical example, and thus the readers may content themselves by considering M as a Euclidean space.

Lemma 6.2.5 *If M is a locally compact separable metric space or a Polish space, and if $(X_t)_{t\geq 0}$ is a stochastic process with values in M, adapted to $(\mathcal{G}_t)_{t\geq 0}$, i.e. for each $t \geq 0$, X_t is a measurable function from $(\Omega, \mathcal{G}_t)$ into $(M, \mathcal{B})$, then the Markov property (6.4) is equivalent to*

$$E\left(f(X_t)|\mathcal{G}_s\right) = E\left(f(X_t)|X_s\right) \qquad \forall t > s$$

for any $f \in C_b(M)$.

The lemma follows from the fact that the Borel σ-algebra $\mathcal{B}(M)$ is the generated by functions in $C_b(M)$, that is $\mathcal{B} = \sigma\left\{f : f \in C_b(M)\right\}$.

Universal measurability

Let M be a topological space. For any σ-finite measure μ, $\mathcal{B}^\mu$ denotes the completion of M with respect to μ and set

$$\mathcal{B}^u = \bigcap_{\mu(M)<+\infty} \mathcal{B}^\mu \ .$$

$\mathcal{B}^u$ is called the σ-algebra of universally (Borel) measurable sets. A function f on M is universally measurable if it is measurable with respect to the σ-algebra $\mathcal{B}^u$.

Let $(\Omega, \mathcal{F}^0, \mathcal{F}_t^0, X_t, \theta_t, P^x)$ be a canonical realization of a Borel semigroup $(P_t)_{t\geq 0}$ in a locally compact separable metric space or a Polish space M.

Proposition 6.2.6 *Let f be a universally Borel measurable function on M. Then for each $t \geq 0$, $f(X_t)$ is $\mathcal{F}_t$-measurable.*

Proof. We need to show that, for any probability μ on $(M, \mathcal{B})$, $f(X_t)$ is $\mathcal{F}_t^\mu$-measurable. By definition, we may choose two Borel measurable functions f_1, f_2 on $(M, \mathcal{B})$, such that $f_1 \leq f \leq f_2$ and

$$f_1 = f = f_2 \qquad \mu P_t\text{-a.s.}$$

By the Markov property, for any $g \in \mathcal{B}$

$$\begin{aligned} E^\mu(g(X_t)) &= E^\mu\left\{E^\mu(g(X_t)|X_0)\right\} = E^\mu(P_t g(X_0)) \\ &= \mu(P_t g) = \mu P_t(g). \end{aligned}$$

It follows that

$$f_1(X_t) \leq f(X_t) \leq f_2(X_t)$$

and

$$f_1(X_t) = f_2(X_t) \qquad P^\mu\text{-a.s.}$$

Therefore $f(X_t)$ is $\mathcal{F}^\mu$-measurable. ■

The argument used in the proof, which should be read twice, is standard regarding augmentation.

Theorem 6.2.7 *Let Y be an $\mathcal{F}$-measurable function on Ω, positive or bounded. Then*

1) The function $x \mapsto E^x(Y)$ is universally measurable.

2) For each $t \geq 0$, $Y \circ \theta_t$ is $\mathcal{F}$-measurable, and for any probability measure μ on $(M, \mathcal{B})$

$$E^\mu(Y \circ \theta_t|\mathcal{F}_t) = E^{X_t}(Y) \qquad P^\mu\text{-a.s.} \tag{6.13}$$

where $E^{X_t}(Y) = F(X_t)$ with $F(x) = E^x(Y)$.

Proof. The proof of 1) is left as an exercise. To prove 2), we first show that (6.13) for Y being $\mathcal{F}^0$-measurable. This follows from (6.12) and a monotone class argument. Now if

$$Y = f_1(X_{t_1}) \cdots f_n(X_{t_n})$$

for $0 \leq t_1 < \cdots < t_n$ and $f_i \in b\mathcal{B}$, then

$$Y \circ \theta_t = f_1(X_{t_1+t}) \cdots f_n(X_{t_n+t})$$

and therefore $Y \circ \theta_t \in \mathcal{F}^0$ and (6.13) holds.

If $Y \in b\mathcal{F}$, then for any $\mu \in \mathcal{B}_1$, we may choose two bounded random variables Y_1, $Y_2 \in b\mathcal{F}^0$ such that $Y_1 \leq Y \leq Y_2$ and $Y_1 = Y_2$ a.s. with respect to the probability measure $P^{\mu P_t}$. Let $F_i(x) = E^x(Y_i)$ which are $\mathcal{B}^u$-measurable. Then by (6.13)

$$E^\mu(Y_i \circ \theta_t|\mathcal{F}_t) = F_i(X_t) \qquad P^\mu\text{-a.s.}$$

so that

$$\begin{aligned} E^\mu(Y_1 \circ \theta_t) &= E^\mu\left(F_1(X_t)\right) = E^{\mu P_t}\left(Y_1\right) \\ &= E^{\mu P_t}\left(Y_2\right) = E^\mu(Y_2 \circ \theta_t) \ . \end{aligned}$$

It follows thus that

$$F_1(X_t) \leq F(X_t) \leq F_2(X_t)$$

and

$$F_1(X_t) = F_2(X_t) \qquad P^\mu\text{-a.s.}$$

Also it follows that

$$Y_1 \circ \theta_t = Y_2 \circ \theta_t \quad P^\mu\text{-a.s.}$$

and

$$Y_1 \circ \theta_t \leq Y \circ \theta_t \leq Y_2 \circ \theta_t \ .$$

Therefore $F(X_t) \in \mathcal{F}_t^\mu$ and $Y \circ \theta_t \in \mathcal{F}^\mu$ for any probability measure μ and (6.13) holds. In particular, if $t = 0$, then the previous argument yields that F is $\mathcal{B}^u$-measurable. ■

Theorem 6.2.8 *If $(X_t)_{t\geq 0}$ has the Markov property with respect to* the right-continuous filtration $(\mathcal{F}_{t+}^0)_{t\geq 0}$ *with transition semigroup $(P_t)_{t\geq 0}$, then*

1) for every initial distribution μ, the completion filtration $(\mathcal{F}_t^\mu)_{t\geq 0}$ is right-continuous, and

2) the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous.

Proof. Let us prove the second claim. The first statement is similar. By Markov property, for any initial distribution μ, $Y \in b\mathcal{F}^0$ and $t \geq 0$

$$E^\mu\left(Y \circ \theta_t|\mathcal{F}_{t+}^0\right) = E^{X_t}(Y) \qquad P^\mu\text{-a.s.} \tag{6.14}$$

Since the right-hand side is measurable with respect to $\mathcal{F}_t$,

$$E^\mu\left(Y \circ \theta_t|\mathcal{F}_{t+}^0\right) = E^\mu\left(Y \circ \theta_t|\mathcal{F}_t\right) \qquad P^\mu\text{-a.s.}$$

It follows that $E^\mu\left(Y|\mathcal{F}_{t+}^0\right) = E^\mu\left(Y|\mathcal{F}_t\right)$, P^μ-a.s. for every $Y \in b\mathcal{F}^0$ by a monotone class argument. Set in the identity $Y = 1_A$ for $A \in \mathcal{F}_{t+}^0$. Then $1_A = E^\mu\left(1_A|\mathcal{F}_t\right)$ for every μ, so that 1_A differs from an $\mathcal{F}_t$-measurable function on P^μ-set. Since μ is arbitrary, it is seen that $A \in \mathcal{F}_t$ and therefore $\mathcal{F}_{t+}^0 \subset \mathcal{F}_t$. On the other hand, it is easy to show that

$$\mathcal{F}_{t+} = \sigma\left\{\mathcal{F}_{t+}^0,\, P^\mu\text{-null sets for every } \mu\right\}.$$

Therefore $\mathcal{F}_{t+} = \mathcal{F}_t$. ■

6.2.4 Markov property and martingale property

Let M be a locally compact separable metric space or a Polish space, and let $\mathcal{B}$ be its Borel σ-algebra. Let $(P_t)_{t>0}$ be a Markov semigroup on $(M, \mathcal{B})$ which satisfies the following (minimal) measurability: $(t,x) \to P_t f(x)$ is Borel measurable on $[0,+\infty) \times M$. This measurability allows us to define the resolvent

$$(R_\alpha f)(x) = \int_0^\infty e^{-\alpha t} (P_t f)(x) dt$$

for bounded or non-negative Borel function f, and the potential kernels

$$R_\alpha(x, A) = \int_0^\infty e^{-\alpha t} P_t(x, A) dt \qquad \forall x \in M, \, A \in \mathcal{B}$$

where $\alpha > 0$.

Let $X = (\Omega, \mathcal{F}^0, \mathcal{F}_t^0, \theta_t, X_t, P^x)$ be the canonical realization of a Markov semigroup $(P_t)_{t>0}$ on $(M, \mathcal{B})$.

Definition 6.2.9 *Let $\alpha > 0$, and let $u \in \mathcal{B}^+$. Then u is an* α-excessive function *with respect to $(P_t)_{t \geq 0}$ if for each $x \in M$, $t \to e^{-\alpha t} P_t u(x)$ is decreasing and*

$$\lim_{t \downarrow 0} e^{-\alpha t} P_t u(x) = u(x) \ . \tag{6.15}$$

Example 6.2.10 *Brownian motion in $\mathbb{R}^n$. The transition semigroup for Brownian motion is*

$$P_t(x, dy) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{|y-x|^2}{2t}\right) dy \ .$$

For every $\alpha > 0$, $R_\alpha f$ is the convolution of f and

$$u_\alpha(z) = \frac{2}{(2\pi)^{n/2}} \left(\frac{|z|}{\sqrt{2\alpha}}\right)^{1-\frac{n}{2}} K_{\frac{n}{2}-1}\left(|z|\sqrt{2\alpha}\right)$$

where K_p is the Bessel functions. If $n = 1$, then

$$u_\alpha(z) = \frac{1}{\sqrt{2\alpha}} \exp\left(-|z|\sqrt{2\alpha}\right)$$

and if $n = 3$ then

$$\frac{1}{\sqrt{2\alpha}|z|} \exp\left(-|z|\sqrt{2\alpha}\right) \ .$$

The basic examples of α-excessive functions are potentials $u = R_\alpha f$, and those are indeed the only excessive function we will use in this book. In fact, if

$f \in \mathcal{B}^+$, by Fubini theorem

$$\begin{aligned}
P_t u(x) &= \int_M \int_0^{+\infty} e^{-\alpha s} P_s f(y) P_t(x, dy) ds \\
&= \int_M \int_0^{+\infty} \int_M e^{-\alpha s} f(z) P_s(y, dz) P_t(x, dy) ds \\
&= \int_0^{+\infty} e^{-\alpha s} \int_M \int_M f(z) P_s(y, dz) P_t(x, dy) ds \\
&= \int_0^{+\infty} e^{-\alpha s} \int_M f(z) P_{t+s}(x, dz)) ds \\
&= \int_0^{+\infty} e^{-\alpha s} P_{t+s} f(x) ds \\
&= e^{\alpha t} \int_t^{+\infty} e^{-\alpha s} \left(P_s f\right)(x) ds \ . \qquad (6.16)
\end{aligned}$$

Therefore, $t \to e^{-\alpha t} P_t u(x)$ is non-negative, decreasing, and

$$\lim_{t \downarrow 0} e^{-\alpha t} P_t u(x) = \lim_{t \downarrow 0} \int_t^{+\infty} e^{-\alpha s} \left(P_s f\right)(x) = u(x) \ .$$

It follows that $u = R_\alpha f$ is an α-excessive function. The following lemma is a connection between Markov processes and martingales.

Lemma 6.2.11 *Let $\alpha > 0$, let u be an α-excessive function with respect to $(P_t)_{t>0}$, and let μ be an initial distribution on $(M, \mathcal{B})$ such that*

$$\int_M u(x) \mu(dx) < +\infty \ .$$

Then $(e^{-\alpha t} u(X_t))_{t \geq 0}$ is a non-negative super-martingale on $(\Omega, \mathcal{F}^\mu, \mathcal{F}_t^\mu, P^\mu)$.

Proof. Since $P_t u \leq e^{\alpha t} u$, it implies that

$$E^\mu u(X_t) = \mu(P_t u) \leq e^{\alpha t} \mu(u) < +\infty$$

namely each $Z_t \equiv e^{-\alpha t} u(X_t)$ is integrable under the probability P^μ. For any $s > 0$ and $t \geq 0$, by the Markov property

$$\begin{aligned}
u(X_s) &\geq e^{-\alpha t} P_t u(X_s) \\
&= e^{-\alpha t} E^{X_s} u(X_t) \\
&= e^{-\alpha t} E^\mu \left(u(X_{t+s}) | \mathcal{F}_s^\mu\right) \ .
\end{aligned}$$

Dividing by $e^{\alpha s}$ both sides, we obtain

$$E^\mu \left(e^{-\alpha(t+s)} u(X_{t+s}) | \mathcal{F}_s^\mu\right) \leq e^{-\alpha s} u(X_s)$$

i.e., $(e^{-\alpha t} u(X_t))_{t \geq 0}$ is a P^μ-super-martingale. ∎

Lemma 6.2.12 *If there is a countable family $\mathcal{S}$ of continuous excessive functions with respect to $(P_t)_{t>0}$ which separates the points of M, then for any countable dense subset $D \subset [0, +\infty)$ the right and left limits along D of $(X_t)_{t\geq 0}$*

$$\lim_{s\in D, s>t, s\downarrow t} X_s \quad (\forall t \geq 0) \text{ and } \quad \lim_{s\in D, s<t, s\downarrow t} X_s \quad (\forall t > 0)$$

exist almost surely under P^μ for any initial distribution μ.

Proof. For every $u \in \mathcal{S}$ it follows that for some $\alpha > 0$, $(e^{-\alpha t}u(X_t))_{t\geq 0}$ is a P^μ-supermartingale. According to Föllmer's lemma, $e^{-\alpha t}u(X_t)$ has right and left limits along D almost surely. Since $\mathcal{S}$ is countable and separates the points of M, X_t has right and left limits along D almost surely. ■

We denote the right-limit process $\lim_{s\in D, s>t, s\downarrow t} X_s$ by X_{t+}, the left limit by X_{t-}. Both limits of course depend on the countable dense subset D.

If it happens (and indeed it is the case if $(P_t)_{t>0}$ is a Feller semigroup) that for each $t \geq 0$, $X_{t+} = X_t$ almost surely under P^μ for an initial distribution μ (which implies that X_{t+} is $\mathcal{F}_t^\mu$-measurable), then $(X_{t+})_{t\geq 0}$ is right-continuous, and has Markov property with respect to $(\mathcal{F}_t^\mu)_{t\geq 0}$ with transition semigroup $(P_t)_{t>0}$.

Theorem 6.2.13 *Let $(P_t)_{t>0}$ be a transition semigroup on $(M, \mathcal{B})$ such that $(t, x) \to P(t, x, A)$ Borel measurable for any $A \in \mathcal{B}$, and let $(X_t)_{t\geq 0}$ be a* right continuous *Markov process with transition semigroup $(P_t)_{t>0}$. For $\alpha > 0$, $f \in b\mathcal{B}$ and $u = R_\alpha f$, set*

$$M_t^f = \int_0^t e^{-\alpha s} f(X_s) ds + e^{-\alpha t} u(X_t) \qquad \text{for } t \geq 0 \ .$$

Then, for any initial distribution μ, $(M_t^f)_{t\geq 0}$ is a bounded martingale under P^μ and

$$M_t^f = E^\mu \left\{ \int_0^{+\infty} e^{-\alpha s} f(X_s) ds \middle| \mathcal{F}_t^\mu \right\} \qquad P^\mu\text{-a.s.} \tag{6.17}$$

Therefore the supermartingale $(e^{-\alpha t}u(X_t))_{t\geq 0}$ has a decomposition

$$e^{-\alpha t} u(X_t) = M_t^f - \int_0^t e^{-\alpha s} f(X_s) ds \quad \forall t \geq 0 \ . \tag{6.18}$$

Proof. Introduce for simplicity

$$A_t^f = \int_0^t e^{-\alpha s} f(X_s) ds$$

which is a variational process. Then for every $x \in M$

$$\begin{aligned} E^x(A_\infty^f) &= \int_0^\infty e^{-\alpha s} E^x f(X_s) ds \\ &= \int_0^\infty e^{-\alpha s} (P_s f)(x) ds \end{aligned}$$

and by the Markov property

$$\begin{aligned} E^\mu(A^f_\infty|\mathcal{F}^\mu_t) &= A^f_t + \int_t^\infty e^{-\alpha s} E^\mu\left\{f(X_s)|\mathcal{F}^\mu_t\right\} ds \\ &= A^f_t + \int_t^\infty e^{-\alpha s}(P_{s-t}f)(X_t)ds \\ &= A^f_t + e^{-\alpha t}\int_0^\infty e^{-\alpha s}(P_s f)(X_t)ds \\ &= A^f_t + e^{-\alpha t}(R_\alpha f)(X_t) \\ &= A^f_t + e^{-\alpha t} f(X_t) \\ &= M^f_t . \end{aligned}$$

Thus $(M^f_t)_{t\geq 0}$ is a bounded martingale under P^μ, and

$$e^{-\alpha t}u(X_t) = M^f_t - \int_0^t e^{-\alpha s} f(X_s)ds \ .$$

■

Notice that the decomposition for the stochastic process $e^{-\alpha t}u(X_t)$ (where $u = R_\alpha f$ is an α-potential of $f \in b\mathcal{B}$) is independent of the initial law μ. If $A^f_t = \int_0^t e^{-\alpha s} f(X_s)ds$ then the decomposition (6.18) may be written as

$$e^{-\alpha t}(R_\alpha f)(X_t) = E^\mu(A_\infty|\mathcal{F}^\mu_t) - A_t \qquad P^\mu\text{-a.s.} \tag{6.19}$$

6.3 Strong Markov property

It is known that Brownian motion has the strong Markov property, as the reflection principle shows. However an argument like reflection principle - applying the Markov property to a random time - needs a solid mathematical foundation.

In the last section, we have exhibited that the simple Markov property is more or less equivalent to the martingale property. We shall show that the strong Markov property is, indeed, a consequence of Doob's optional sampling theorem.

In order to discuss further sample properties of a Markov process, we require a set of minimal regularity properties listed as the following. Let $(X_t)_{t\geq 0}$ be a right continuous Markov process on $(\Omega, \mathcal{G}, \mathcal{G}_t, P^x)$ in a locally compact separable metric space or a Polish space M, with the transition semigroup $(P_t)_{t>0}$. Then $(t,x) \to P(t,x,A)$ is Borel measurable for every $A \in \mathcal{B}$.

Definition 6.3.1 (Strong Markov Property, SMP) *It is said that* $(X_t)_{t\geq 0}$ *has the* strong Markov property *(abbreviated as SMP), if for any stopping time* T *with respect to the (right-continuous) filtration* $(\mathcal{G}_{t+})$*,* $f \in b\mathcal{B}$*,* $t \geq 0$ *and for any initial distribution* μ*, it holds that*

$$E^\mu\left\{f(X_{t+T})1_{\{T<+\infty\}}|\mathcal{G}_{T+}\right\} = (P_t f)(X_T)1_{\{T<+\infty\}} \qquad P^\mu\text{-}a.s. \tag{6.20}$$

By setting $T = s$ a constant time, one can see that SMP implies the Markov property with respect to the right-continuous filtration $(\mathcal{G}_{t+})$. According to Theorem (6.2.8), for a strong Markov process, both its natural filtration $(\mathcal{F}_t)_{t\geq 0}$ and filtration $(\mathcal{F}_t^\mu)_{t\geq 0}$ satisfy the usual conditions for any initial distribution μ.

Proposition 6.3.2 *$(X_t)_{t\geq 0}$ has the strong Markov property if and only if for any $f \in C_b(M)$, $t \geq 0$ and any $(\mathcal{G}_{t+})$-stopping time T*

$$E^\mu\left(f(X_{t+T})1_{\{T<+\infty\}}\right) = E^\mu\left((P_t f)(X_T)1_{\{T<+\infty\}}\right) \ . \tag{6.21}$$

Proof. We only need to prove the sufficiency of (6.21). By Lebesgue's dominated convergence theorem, (6.21) remains true for any $f \in b\mathcal{B}$. If T is a $(\mathcal{G}_{t+})$-stopping time and $A \in \mathcal{G}_{T+}$, then $T_A = T1_A + (+\infty)1_{A^c}$ is again a stopping time with respect to the right-continuous filtration $(\mathcal{G}_{t+})$, and hence, by applying (6.21) to T_A,

$$E^\mu\left(f(X_{t+T_A})1_{\{T_A<+\infty\}}\right) = E^\mu\left((P_t f)(X_{T_A})1_{\{T_A<+\infty\}}\right) \ .$$

However $1_{\{T_A<+\infty\}} = 1_{\{T<+\infty\}}1_A$ and $T_A = T$ on $\{T_A < +\infty\}$. Therefore

$$E^\mu\left(f(X_{t+T})1_{\{T<+\infty\}}1_A\right) = E^\mu\left((P_t f)(X_T)1_{\{T<+\infty\}}1_A\right)$$

for every $A \in \mathcal{G}_{T+}$, and it follows that

$$E^\mu\left(f(X_{t+T})1_{\{T<+\infty\}}|\mathcal{G}_{T+}\right) = (P_t f)(X_T)1_{\{T<+\infty\}} \qquad P^\mu\text{-a.s.}$$

■

Proposition 6.3.3 *Let $(X_t)_{t\geq 0}$ be a strong Markov process. Then*

1) $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous, and $\mathcal{F}_{T+} = \mathcal{F}_T$ for any stopping time T on $(\Omega, \mathcal{F}, \mathcal{F}_t)$,

2) for any $Y \in b\mathcal{F}$, a stopping time T on $(\Omega, \mathcal{F}^\mu, \mathcal{F}_t^\mu)$ and an initial distribution μ, $Y \circ \theta_T 1_{\{T<+\infty\}} \in \mathcal{F}$ and

$$E^\mu\left\{Y \circ \theta_T 1_{\{T<+\infty\}}|\mathcal{F}_T^\mu\right\} = E^{X_T}(Y)1_{\{T<+\infty\}} \qquad P^\mu\text{-a.e.} \tag{6.22}$$

Proof. 1) follows from Theorem 6.2.8. We now prove 2). If $Y = f(X_t)$ for some $f \in b\mathcal{B}$, $t \geq 0$, then obviously

$$Y \circ \theta_T 1_{\{T<+\infty\}} = f(X_{t+T})1_{\{T<+\infty\}}$$

which is thus $\mathcal{F}$-measurable. It follows then from an application of the monotone class theorem that $Y \circ \theta_T 1_{\{T<+\infty\}} \in \mathcal{F}$ for any $Y \in b\mathcal{F}^0$. To prove the asserted equality for the case that $Y \in b\mathcal{F}^0$, it is sufficient, by the monotone class theorem, to prove the claim for

$$Y = f_0(X_{t_0}) \cdots f_n(X_{t_n})$$

where $0 = t_0 < t_1 < \cdots < t_n$ and $f_i \in b\mathcal{B}$. It is a routine argument.

For the general case $Y \in b\mathcal{F}$, for every μ we may choose Y_1, $Y_2 \in b\mathcal{F}^0$ such that

$$Y_1 \le Y \le Y_2 \text{ and } E^\nu(Y_2 - Y_1) = 0$$

where

$$\nu(g) = P^\mu(g(X_T)1_{\{T<+\infty\}}) \qquad \forall g \in b\mathcal{B} .$$

Then

$$\begin{aligned} & E^\mu \left\{ E^\mu \left(Y_2 - Y_1\right) \circ \theta_T 1_{\{T<+\infty\}} | \mathcal{F}_T) \right\} \\ & \quad = E^\mu \left\{ E^{X_T} \left(Y_2 - Y_1\right) 1_{\{T<+\infty\}}) \right\} \\ & \quad = E^\nu \left(Y_2 - Y_1\right) \\ & \quad = 0 \end{aligned}$$

and hence the identity (6.22) holds for $Y \in b\mathcal{F}$. ■

If in addition $(X_t)_{t\ge0}$ satisfies the following property: for any $\alpha > 0$ and for any α-excessive function u with respect to $(P_t)_{t\ge0}$, the sample paths $t \to u(X_t)$ are right-continuous almost surely under P^μ for any initial distribution μ, then

$$(\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, P^x)$$

is called a Borel right process. The concept of right processes was introduced by P. A. Meyer. The adjective "right" in the notion indicates that the class of right processes is a nice class of Markov processes for which a nice analysis (mainly for the propose of the potential theory) may be established.

Theorem 6.3.4 *Let $(X_t)_{t\ge0}$ be a Borel right process with transition semigroup $(P_t)_{t>0}$ and α-potential (R_α). Then $(X_t)_{t\ge0}$ has the strong Markov property.*

Proof. Let $f \in C_b(M)$. It is known from Theorem 6.2.13 and the definition of Borel right processes that

$$M_t^f = \int_0^t e^{-\alpha s} f(X_s) ds + e^{-\alpha t}(R_\alpha f)(X_t) .$$

is a right-continuous, bounded martingale under P^μ, and thus, by Doob's optional sampling theorem

$$E^\mu(M_\infty^f 1_{\{T<+\infty\}} | \mathcal{F}_{T+}^\mu) = M_T^f 1_{\{T<+\infty\}}.$$

It then follows that

$$E^\mu \left\{ 1_{\{T<+\infty\}} \int_T^{+\infty} e^{-\alpha t} f(X_t) dt \middle| \mathcal{F}_{T+}^\mu \right\} = 1_{\{T<+\infty\}} e^{-\alpha T} (R_\alpha f)(X_T) .$$

Multiplying both sides by $e^{\alpha T}$ and using the fact that $e^{\alpha T} \in \mathcal{F}_{T+}^\mu$ we obtain

$$E^\mu \left\{ 1_{\{T<+\infty\}} \int_T^{+\infty} e^{-\alpha(t-T)} f(X_t) dt \middle| \mathcal{F}_{T+}^\mu \right\} = 1_{\{T<+\infty\}} (R_\alpha f)(X_T) .$$

After making change of variable

$$E^{\mu}\left\{1_{\{T<+\infty\}}\int_0^{+\infty}e^{-\alpha t}f(X_{t+T})dt\middle|\mathcal{F}^{\mu}_{T+}\right\}=1_{\{T<+\infty\}}(R_\alpha f)(X_T),$$

for every $\alpha>0$. Since f is continuous, both $t\to f(X_{t+T})1_{\{T<+\infty\}}$ and $t\to(P_tf)(X_T)1_{\{T<+\infty\}}$ are right-continuous on $[0,+\infty)$. According to the uniqueness of Laplace transformations, we have

$$E^{\mu}\left\{1_{\{T<+\infty\}}f(X_{t+T})\middle|\mathcal{F}^{\mu}_{T+}\right\}=(P_tf)(X_T)1_{\{T<+\infty\}}$$

which proves the SMP. ■

Remark 6.3.5 *We need to write a remark for a result in §6.5. Carefully reading the proof of Proposition 6.3.2 and Theorem 6.3.4, we see that the condition in Theorem 6.3.4 may be weaker. Let* $\mathbb{L}$ *be a linear subspace of* $C_b(M)$ *such that for any open set* $G\subset M$*, there exists an increasing sequence* $\{f_n\}\subset\mathbb{L}$ *with* $f_n\uparrow 1G$*. The conclusion in Theorem 6.3.4 still holds true if* $t\mapsto R_\alpha f(X_t)$ *is right continuous almost surely under* P^μ *for any initial distribution* μ *and any* $f\in\mathbb{L}$*.*

6.4 Feller processes

In this section we construct a strong Markov process for a given Feller semigroup on a locally compact separable metric space M. If necessary, we may consider the one-point compactification of M, and thus without losing generality, we may assume throughout this section that M is a compact metric space. Let $\{P_t(x,dy)\colon t\geq 0\}$ be a Feller semigroup on M, and $P_t(x,M)=1$ for every $t\geq 0$. Each P_t preserves the space $C(M)$ of the continuous functions, and for every $x\in M$

$$\lim_{t\downarrow 0}P_tu(x)=u(x)$$

which implies the uniform convergence

$$\lim_{t\downarrow 0}||P_tu-u||_\infty=0\ .$$

For $\alpha>0$, let $R_\alpha(x,dy)$ denote the potential kernel

$$R_\alpha(x,A)=\int_0^\infty e^{-\alpha t}P_t(x,A)dt\qquad\forall x\in M,\ A\in\mathcal{B}\ .$$

For every $\alpha>0$, R_α preserves $C(M)$ and

$$\lim_{\alpha\to+\infty}||\alpha R_\alpha u-u||_\infty=0\ .$$

Proposition 6.4.1 *Let $(X_t)_{t\geq 0}$ be a Markov process on a complete filtred probability space $(\Omega, \mathcal{G}, \mathcal{G}_t, P)$ taking values in the compact metric space M with a Feller transition semigroup $(P_t)_{t\geq 0}$. Let D be a countable dense subset of $[0, +\infty)$. Then*

1) For almost all $\omega \in \Omega$, the right-limit along D at each $t \geq 0$

$$X_{t+}(\omega) = \lim_{s\in D, s>t, s\downarrow t} X_s(\omega)$$

and the left-limit along D at every $t > 0$

$$X_{t-}(\omega) = \lim_{s\in D, s<t, s\uparrow t} X_s(\omega)$$

exist.

2) For every $t \geq 0$, $X_{t+} = X_t$ almost surely, and for each $t > 0$, $X_{t-} = X_t$ almost surely. Therefore both $(X_t)_{t\geq 0}$ and $(X_{t+})_{t\geq 0}$ are stochastic continuous, and $(X_{t+})_{t\geq 0}$ is right-continuous and is a version of $(X_t)_{t\geq 0}$.

3) The right-continuous process $(X_{t+})_{t\geq 0}$ is a Markov process on $(\Omega, \mathcal{G}, \mathcal{G}_{t+}, P)$ with the same transition semigroup $(P_t)_{t\geq 0}$.

Proof. Let $\{f_n : n \geq 1\}$ be a countable family of non-negative continuous functions on M which separates points of the compact metric space M, and set $u_{mn} = mR_m f_n$. Since

$$\lim_{m\to\infty} u_{mn} = f_n$$

the family of continuous functions $\{u_{mn} : m, n \geq 1\}$ is countable and separates points of M as well. Thus according to Lemma 6.2.12, 1) follows.

Next we prove the second statement. To show for each $t \geq 0$, $X_t = X_{t+}$ we only need to prove that for any continuous function f on $M \times M$

$$E(f(X_t, X_{t+})) = E(f(X_t, X_t)) . \tag{6.23}$$

It is sufficient, by the monotone class theorem, to show (6.23) for $f(x, y) = f_1(x)f_2(y)$ where $f_i \in C(M)$. That is, we have to prove

$$E(f_1(X_t)f_2(X_{t+})) = E(f_1(X_t)f_2(X_t)) . \tag{6.24}$$

While

$$\begin{aligned} E(f_1(X_t)f_2(X_{t+})) &= \lim_{s\in D, s>t, s\downarrow t} E(f_1(X_t)f_2(X_s)) \\ &= \lim_{s\in D, s>t, s\downarrow t} E(f_1(X_t)P_{s-t}f_2(X_t)) \\ &= E(f_1(X_t)f_2(X_t)) \end{aligned}$$

where the first equality follows from the fact that $X_{t+} = \lim_{s\in D, s>t, s\downarrow t} X_s$ almost surely, the second follows from the Markov property

$$E(f_2(X_s)|\mathcal{G}_t) = P_{s-t}f_2(X_t) \quad P\text{-a.s.}$$

and the fact that $f_1(X_t)$ is $\mathcal{G}_t$-measurable, and the last equality follows from the Feller property that

$$\lim_{s\downarrow t} P_{s-t}f_2(x) = f_2(x)$$

uniformly.

Similarly, in order to show $X_t = X_{t-}$ almost surely for every $t > 0$, it is sufficient to prove

$$E(f_1(X_{t-})f_2(X_t)) = E(f_1(X_{t-})f_2(X_{t-})) \tag{6.25}$$

where $f_i \in C(M)$. Since

$$E(f_1(X_{t-})f_2(X_t)) = \lim_{s\in D, s<t, s\uparrow t} E(f_1(X_s)f_2(X_t))$$

and by the Markov property

$$\begin{aligned} E(f_1(X_s)f_2(X_t)|\mathcal{G}_s) &= f_1(X_s)E(f_2(X_t)|\mathcal{G}_s) \\ &= f_1(X_s)(P_{t-s}f_2)(X_s) \qquad P\text{-a.s.} \end{aligned}$$

Similarly $P_{t-s}f_2 \to f_2$ uniformly as $s \uparrow t$, and it follows thus that

$$\begin{aligned} \lim_{s\in D, s<t, s\uparrow t} E(f_1(X_s)f_2(X_t)) &= \lim_{s\in D, s<t, s\uparrow t} E(f_1(X_s)(P_{t-s}f_2)(X_s)) \\ &= E(f_1(X_{t-})f_2(X_{t-})) \ , \end{aligned}$$

which concludes the proof of the second part.

We are going to prove the third claim. Consider a continuous function f, $t > s$ and $A \in \mathcal{G}_{s+}$. For any $r \in (s,t) \cap D$, $A \subset \mathcal{G}_{s+} \subset \mathcal{G}_r$ and by the Markov property

$$E\left(f(X_t)|\mathcal{G}_r\right) = P_{t-r}f(X_r) \qquad P\text{-a.s.}$$

It implies that

$$E\left(1_A f(X_t)\right) = E(1_A P_{t-r}f(X_r)) \ .$$

By letting $r \in (s,t) \cap D \downarrow s$ we obtain

$$\begin{aligned} E\left(1_A f(X_t)\right) &= E(1_A P_{t-s}f(X_{s+})) \\ &= E(1_A P_{t-s}f(X_{s+})) \end{aligned}$$

for every $A \in \mathcal{G}_{s+}$. Since $X_t = X_{t+}$ almost surely for any fixed $t \geq 0$,

$$\begin{aligned} E\left(1_A f(X_t)\right) &= E\left(1_A f(X_{t+})\right) \\ &= E(1_A P_{t-s}f(X_{s+})) \\ &= E(1_A P_{t-s}f(X_s)) \qquad \forall A \in \mathcal{G}_{s+} \end{aligned}$$

which implies that both processes $(X_t)_{t\geq 0}$ and $(X_{t+})_{t\geq 0}$ are Markov processes on $(\Omega, \mathcal{G}, \mathcal{G}_{t+}, P)$ with the same transition semigroup $(P_t)_{t\geq 0}$. ∎

Thus, the realization $\left(\Omega, \mathcal{F}^0, \mathcal{F}^0_t, X_t, \theta_t, P^x\right)$ on the path space of a Feller transition semigroup $(P_t)_{t>0}$ has a version $(X_{t+})_{t\geq 0}$ (defined via limits along a dense set D) which is right-continuous with left limits, and moreover it has the Markov property relative to a right continuous filtration $(\mathcal{F}^0_{t+})$. Therefore there is a canonical realization of $(P_t)_{t>0}$

$$\left(\Omega, \mathcal{F}^0, \mathcal{F}^0_{t+}, X_t, \theta_t, P^x\right)$$

on the space Ω of right continuous paths with left limits in M, which is called Feller process.

Theorem 6.4.2 *Let $\left(\Omega, \mathcal{F}^0, \mathcal{F}^0_{t+}, X_t, \theta_t, P^x\right)$ be the Feller process with Feller semigroup $(P_t)_{t\geq 0}$ as above.*

1) For any initial distribution μ, the filtration $(\mathcal{F}^\mu_t)_{t\geq 0}$ is right continuous.

2) The natural filtration $(\mathcal{F}_t)_{t\geq 0}$ of the Markov process $(X_t)_{t\geq 0}$ is right continuous.

3) (Blumenthal 0-1 law) For any $x \in M$ and $A \in \mathcal{F}^0_{0+}$, $P^x(A) = 0$ or 1. The conclusion is true for any $A \in \mathcal{F}_0$.

4) $(X_t)_{t\geq 0}$ has strong Markov property.

Proof. The first and the second statements follow from the Markov property of $(X_t)_{t\geq 0}$ with respect to $(\mathcal{F}^0_{t+})_{t\geq 0}$, see Theorem 6.2.8. If $A \in \mathcal{F}^0_{0+}$, by Markov property

$$P^x(A) = E^{X_0}(1_A) = E^x(1_A|\mathcal{F}^0_{0+}) = 1_A \qquad P^x\text{-a.s.}$$

The last 4) follows from Theorem 6.3.4. ■

By applying Doob's optional sampling theorem to a predictable stopping time, we may prove the following

Theorem 6.4.3 *Let $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, P^x)$ be a Feller process with Feller semigroup $(P_t)_{t\geq 0}$. Then for any initial law μ, $f \in b\mathcal{B}$ and any predictable stopping time T on $(\Omega, \mathcal{F}^\mu, \mathcal{F}^\mu_t, P^\mu)$*

$$E^\mu\left(f(X_{s+T})1_{\{T<+\infty\}}|\mathcal{F}_{T-}\right) = (P_s f)(X_{T-})1_{\{T<+\infty\}} \qquad P^\mu\text{-a.s.} \tag{6.26}$$

(6.26) is called the moderate Markov property.

Proof. Apply Doob's optional sampling theorem (2.2.11) to the bounded martingale

$$M^f_t = \int_0^t e^{-\alpha s} f(X_s) ds + e^{-\alpha t}(R_\alpha f)(X_t) \ .$$

■

Corollary 6.4.4 *Under the same assumptions as in Theorem 6.4.3. Let T be a predictable stopping time with respect to the right-continuous filtration $(\mathcal{F}^\mu_t)_{t\geq 0}$. Then*

$$X_T = X_{T-} \qquad \text{on } \{T < +\infty\} \qquad P^\mu\text{-a.s.}$$

Proof. By replacing T by $T \wedge n$ we may assume T is a bounded predictable stopping time. By the moderate Markov property, for any f, $g \in C_0(M)$

$$E^\mu \left(f(X_T) | \mathcal{F}_{T-} \right) = f(X_{T-}) \qquad P^\mu\text{-a.s.}$$

so that, as $g(X_{T-}) \in \mathcal{F}_{T-}$,

$$E^\mu \left(f(X_T) g(X_{T-}) \right) = E^\mu \left(g(X_{T-}) f(X_{T-}) \right)$$

which implies that

$$E^\mu \left(f(X_T, X_{T-}) \right) = E^\mu \left(f(X_{T-}, X_{T-}) \right)$$

for any $f \in C(M \times M)$, and therefore $X_T = X_{T-}$ P^μ-a.s. ∎

Remark 6.4.5 *This property shows the jumps of a Feller process $(X_t)_{t \geq 0}$ are totally inaccessible. A stopping time τ is called totally inaccessible if $P(\tau = \sigma) = 0$ for any predictable stopping time σ.*

Corollary 6.4.6 (Quasi-left continuity) *Under the same assumptions as in Theorem 6.4.3. Let μ be an initial distribution.*

1) $(\mathcal{F}_t^\mu)_{t \geq 0}$ is quasi-left continuous in the sense that for any predictable $(\mathcal{F}_t^\mu)$-stopping time T, $\mathcal{F}_{T-}^\mu = \mathcal{F}_T^\mu$.

2) Let $\{T_n\}$ be an increasing family of $(\mathcal{F}_t^\mu)$-stopping times, and let $T = \lim_{n \to \infty} T_n$. Then

$$\lim_{n \to \infty} X_{T_n} = X_T \qquad \text{on } \{T < +\infty\} \qquad P^\mu\text{-a.s.} \tag{6.27}$$

Proof. Proof 1) ??? Now prove 2). On the part $\{T_n < T < +\infty \text{ for all } n\}$, (6.27) follows from Corollary 6.4.4, while on $\{T_n = T < +\infty \text{ for some } n\}$ (6.27) is trivial. ∎

In the remainder of this section, we assume that $(\Omega, \mathcal{F}, \mathcal{F}_t, \theta_t, X_t, P^x)$ is the canonical realization of a Feller semigroup $(P_t)_{t \geq 0}$ on the sample space of paths in M which are right-continuous with left limits. We next would like to look at the family of fundamental martingales

$$M_t^f = \int_0^t e^{-\alpha s} f(X_s) \mathrm{d}s + e^{-\alpha t} (R_\alpha f)(X_t)$$

where

$$M_t^f = E^\mu \left(A_\infty^f | \mathcal{F}_t^\mu \right), \quad A_t^f = \int_0^t e^{-\alpha s} f(X_s) \mathrm{d}s \ .$$

First let us compute the L^1 and L^2 norms of M_∞^f.

$$\begin{aligned}
E^\mu |M_\infty^f| &= E^\mu \left| \int_0^\infty e^{-\alpha s} f(X_s) \mathrm{d}s \right| \\
&\leq \int_0^\infty e^{-\alpha s} E^\mu \left| f(X_s) \right| \mathrm{d}s \\
&= \int_0^\infty e^{-\alpha s} \mu P_s(|f|) \mathrm{d}s \\
&= (\mu R_\alpha)(|f|) \ .
\end{aligned}$$

For the L^2-norm, we write

$$\begin{aligned} E^\mu|M^f_\infty|^2 &= E^\mu\left|\int_0^\infty e^{-\alpha s}f(X_s)\mathrm{d}s\right|^2 \\ &= \int_0^\infty\int_0^\infty e^{-\alpha(s+t)}E^\mu\left(f(X_s)f(X_t)\right)\mathrm{d}s\mathrm{d}t\ . \end{aligned} \tag{6.28}$$

The integral on the right-hand side can be computed explicitly, but here we deduce a useful estimate. Using the Hölder inequality we obtain

$$E^\mu\left(f(X_s)f(X_t)\right) \le \sqrt{E^\mu(f(X_s)^2)}\sqrt{E^\mu(f(X_t)^2)}$$

and

$$\begin{aligned} E^\mu|M^f_\infty|^2 &\le \int_0^\infty\int_0^\infty e^{-\alpha(s+t)}\sqrt{E^\mu(f(X_s)^2)}\sqrt{E^\mu(f(X_t)^2)}\mathrm{d}s\mathrm{d}t \\ &= \left(\int_0^\infty e^{-\alpha t}\sqrt{\mu P_t(f^2)}\mathrm{d}t\right)^2 \\ &\le \left(\int_0^\infty e^{-\alpha t}\mu P_t(f^2)\mathrm{d}t\right)\left(\int_0^\infty e^{-\alpha t}\mathrm{d}t\right) \\ &= \frac{1}{\alpha}(\mu U^\alpha)(f^2)\ . \end{aligned} \tag{6.29}$$

Lemma 6.4.7 *Given $\alpha > 0$ and an initial distribution μ, set $m(dx) \equiv \mu R_\alpha$ which is a finite measure. Then*

$$E^\mu|M^f_\infty - M^g_\infty| \le ||f-g||_{L^1(m)}$$

and

$$E^\mu|M^f_\infty - M^g_\infty|^2 \le \frac{1}{\alpha}||f-g||^2_{L^2(m)}\ . \tag{6.30}$$

The following lemma follows from a standard application of Doob's maximal inequality.

Lemma 6.4.8 *Let μ be an initial distribution. Let $\mathbb{M}^2_\mu$ denote the space of all square-integrable martingales $(W_t)_{t\in[0,+\infty]}$ which are right-continuous with left-limits on $(\Omega, \mathcal{F}^\mu, \mathcal{F}^\mu_t, P^\mu)$ endowed with norm*

$$||W|| = \sqrt{E^\mu|W_\infty|^2} \qquad \forall W \in \mathbb{M}^2_\mu\ .$$

Then $\mathbb{M}^2_\mu$ is a Hilbert space.

Theorem 6.4.9 *Let $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, P^x)$ be a Feller process with Feller transition semigroup $(P_t)_{t\ge 0}$ on a locally compact metric space M. Let $\alpha > 0$ and let μ be an initial distribution, and let $u \in L^2(M, \mu R_\alpha)$.*

1) The process

$$M^f_t = \int_0^t e^{-\alpha s}f(X_s)ds + e^{-\alpha t}(R_\alpha f)(X_t) \tag{6.31}$$

is a square-integrable martingale on $(\Omega, \mathcal{F}^\mu, \mathcal{F}_t^\mu, P^\mu)$ *which is right-continuous with left limits.*

2) The processes $((R_\alpha f)(X.)_{t-})_{t\geq 0}$ *and* $((R_\alpha f)(X_{t-}))_{t\geq 0}$ *are indistinguishable under* P^μ*.*

3) $(M_t^f)_{t\geq 0}$ *is a quasi-left continuous martingale on* $(\Omega, \mathcal{F}^\mu, \mathcal{F}_t^\mu, P^\mu)$*.*

Proof. The first two conclusions are obvious if f is continuous, Lemma 6.4.8 and (6.30) allow us to extended to any $f \in L^2(M, \mu R_\alpha)$. The quasi-left continuity of $(M_t^f)_{t\geq 0}$ follows from the second result and Corollary 6.4.4. ■

Corollary 6.4.10 *If* $f \in b\mathcal{B}$*, then* $(M_t^f)_{t\geq 0}$ *is a right-continuous and quasi-left continuous bounded martingale on* $\left(\Omega, \mathcal{F}^0, \mathcal{F}_t^0, P^\mu\right)$ *for any initial law* μ*.*

As we have indicated that $E^\mu |M_\infty^f|^2$ can be computed explicitly. Indeed, by the Markov property

$$\begin{aligned} E^\mu |M_\infty^f|^2 &= 2 \iint_{0\leq s<t<+\infty} e^{-\alpha(s+t)} E^\mu \left(f(X_s) f(X_t)\right) \mathrm{d}s\mathrm{d}t \\ &= 2 \iint_{0\leq s<t<+\infty} e^{-\alpha(s+t)} \mu(P_s(f P_{t-s} f) \mathrm{d}s\mathrm{d}t \\ &= 2\mu \left(\iint_{0\leq s<t<+\infty} e^{-\alpha(s+t)} P_s(f P_{t-s} f) \mathrm{d}s\mathrm{d}t \right) . \end{aligned}$$

Making change of limits in the last integral we thus obtain

$$\begin{aligned} & \iint_{0\leq s<t<+\infty} e^{-\alpha(s+t)} P_s(f P_{t-s} f) \mathrm{d}s\mathrm{d}t \\ = & \int_0^{+\infty} P_s \left(u \int_s^{+\infty} e^{-\alpha(s+t)} (P_{t-s} f) \mathrm{d}t \right) \mathrm{d}s \\ = & \int_0^{+\infty} e^{-2\alpha s} P_s \left(f(R_\alpha f)\right) \mathrm{d}s \\ = & R_{2\alpha} \left(f(R_\alpha f)\right) \end{aligned}$$

hence

$$E^\mu |M_\infty^f|^2 = 2(\mu R_{2\alpha}) \left(f(R_\alpha f)\right) .$$

6.5 Diffusion processes

A continuous Markov process with the strong Markov property is called a *diffusion process*. In this section, we prove the solutions to stochastic differential equations are diffusions.

6.5.1 The Wiener space

The heat semigroup $(P_t)_{t\geq 0}$ defined by $P(t,x,dy) = p(t,y-x)dy$, where

$$p(t,z) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|z|^2}{2t}\right)$$

the Gaussian kernel on $\mathbb{R}^d$, is a Feller semigroup.Therefore, there is a canonical realization on the space of continuous paths. More precisely, let $\mathbf{W}^d$ be the space of all continuous paths in $\mathbb{R}^d$, and, for simplicity of notations, let $(W_t)_{t\geq 0}$ be the coordinate process, i.e. $W_t(\omega) = \omega(t)$ for all $\omega \in \mathbf{W}^d$ and $t \geq 0$. Let $\mathcal{F}^0 = \sigma\{W_s : s \in [0,\infty)\}$ and $\mathcal{F}^0_t = \sigma\{W_s : s \in [0,t]\}$. Then, for any $x \in \mathbb{R}^d$, there is a unique probability measure P^x on $(\mathbf{W}^d, \mathcal{F}^0)$ such that $(W_t)_{t\geq 0}$ is a strong Markov process with transition semigroup $(P_t)_{t>0}$. Since $P^0\{W_0 = 0\} = 1$, the measure P^0 has support $\mathbf{W}^d_0$ the space of all continuous paths with initial zero. The measure $P = P^0$ on $\left(\mathbf{W}^d_0, \mathcal{F}^0\right)$ is called the Wiener measure. Let $\mathcal{F}$ be the completion of $\mathcal{F}^0$ under P and call $(\mathbf{W}^d_0, \mathcal{F}^0, P)$ the Wiener space. Then $(W_t)_{t\geq 0}$ is the standard Brownian motion in $\mathbb{R}^d$ or Wiener process on $(\mathbf{W}^d_0, \mathcal{F}, P)$.

Moreover, if $B = (B_t : t > 0)$ be a standard Brownian motion on $(\Omega, \mathcal{F}, \hat{P})$ and define for any $\omega \in \Omega$ such that $t \mapsto B_t(\omega)$ is continuous,

$$\Phi(\omega)(t) = B_t(\omega)$$

otherwise $\Phi(\omega)(t) = 0$. Then it is easy to check that Φ is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbf{W}^d_0, \mathcal{F}^0)$. The image measure $\hat{P} \circ \Phi^{-1}$ of $\hat{P}$ is Wiener measure P on $(\mathbf{W}^d_0, \mathcal{F}^0)$.

6.5.2 Diffusions as strong solutions of SDE's

Since strong solutions are considered, we may simply move the SDE on a generic probability space to the Wiener space. This doesn't make essential difference in the argument. Consider the stochastic differential equation on the Wiener space $(\mathbf{W}^d_0, \mathcal{F}^0, P)$

$$dX^j_t = b^j(X_t)dt + \sigma^j_i(X_t)dW^i_t, \quad X_0 = x \tag{6.32}$$

where $j = 1, \cdots, n$, b^j and σ^j_i are Lipschitz continuous with at most linear growth. According to the existence and uniqueness of SDE, Theorem 5.3.3, for each $x \in \mathbb{R}^n$ there is a unique strong solution to (6.32), denoted by $X(t,x,\omega)$. We shall first prove that the solution has nice continuity in initial value x as follows.

Theorem 6.5.1 *The function $x \to X(t,x,\cdot)$ is uniformly continuous almost surely for t in any finite interval $[0,T]$:*

$$\lim_{\delta\downarrow 0} \sup_{|x-y|<\delta} E\left\{\sup_{0\leq t\leq T} |X(t,x,\cdot) - X(t,y,\cdot)|^2\right\} = 0\ . \tag{6.33}$$

Proof. Let us write $X^x(t) = X(t,x,\cdot)$ simply in this proof. We shall only consider 1-dimensional case. Thus

$$X^x(t) = x + \int_0^t \sigma(s, X^x(s))\mathrm{d}W_s + \int_0^t b(s, X^x(s))\mathrm{d}s$$

and

$$X^y(t) = y + \int_0^t \sigma(s, X^y(s))\mathrm{d}W_s + \int_0^t b(s, X^y(s))\mathrm{d}s \ .$$

Therefore, by Doob's maximal inequality,

$$\begin{aligned}
& E\left\{\sup_{0\leq t\leq T} |X^x(t) - X^y(t)|^2\right\} \leq 3|x-y|^2 \\
& +3E\left\{\sup_{0\leq t\leq T}\left|\int_0^t (\sigma(s,X^x(s)) - \sigma(s,X^y(s)))\mathrm{d}W_s\right|^2\right\} \\
& +3E\left\{\sup_{0\leq t\leq T}\left|\int_0^t (b(s,X^x(s)) - b(s,X^y(s)))\mathrm{d}s\right|^2\right\} \\
\leq \ & 3|x-y|^2 + 12E\left\{\left|\int_0^T (\sigma(s,X^x(s)) - \sigma(s,X^y(s)))\,\mathrm{d}W_s\right|^2\right\} \\
& +3TE\left\{\int_0^T |b(X^x(s)) - b(X^y(s))|^2\,\mathrm{d}s\right\} \\
\leq \ & 3|x-y|^2 + 12E\left\{\int_0^T |\sigma(s,X^x(s)) - \sigma(s,X^y(s))|^2\,\mathrm{d}s\right\} \\
& +3TC^2E\left\{\int_0^T |X^x(s) - X^y(s)|^2\,ds\right\} \\
\leq \ & 3|x-y|^2 + 3C^2(4+T)\int_0^T E\left(|X^x(t) - X^y(t)|^2\right)\mathrm{d}t.
\end{aligned}$$

Setting

$$\Delta(t) = E\left\{\sup_{0\leq s\leq t} |X^x(s) - X^y(s)|^2\right\} \ ,$$

we have

$$\Delta(T) \leq 3|x-y|^2 + 3C^2(4+T)\int_0^T \Delta(t)dt$$

and therefore by Gronwall's inequality

$$\Delta(T) \leq 6|x-y|^2\exp(12C^2 + 3TC^2)$$

which yields (and is actually stronger than) (6.33). ■

Define $P_t f(x) = Ef(X(t,x,\cdot))$. Then according to Theorem 6.5.1, for any function $f \in C^1(\mathbb{R}^d)$, one has

$$\begin{aligned}|P_t f(x) - P_t f(y)| &\leq E\,|f(X(t,x,\cdot)) - f(X(t,y,\cdot))| \\ &\leq ||\nabla f||_\infty E\,|X(t,x,\cdot) - X(t,y,\cdot)|\,,\end{aligned}$$

and then

$$\sup_{t\leq T}\sup_{|x-y|<\delta}|P_t f(x) - P_t f(y)| \to 0 \quad \text{ as } \delta \downarrow 0\ .$$

It follows that $x \to P_t f(x)$ is continuous, so that $P(t,x,A) = P(X(t,x,\cdot) \in A)$ is measurable in x and thus is a kernel on $\mathbb{R}^n$.

For each $s \geq 0$, we define $\delta_s : \mathbf{W}_0^d \to \mathbf{W}_0^d$ by

$$\delta_s\omega(t) = \omega(t+s) - \omega(s)\ .$$

Then for each $s \geq 0$, δ_s is a measurable mapping, and δ_s is independent of $\mathcal{F}_s^0$ (with respect to the Wiener measure). Moreover $(W_t \circ \delta_s)_{t\geq 0}$ is again a standard Brownian motionon on $(\mathbf{W}_0^d, \mathcal{F}^0, P)$ which is independent of $\mathcal{F}_s^0$. Let $Z_t(\omega) = X(t+s,x,\omega)$. Since

$$\begin{aligned}Z_t - Z_0 &= \int_s^{s+t} b^j(X(r,x,\cdot))dr + \int_s^{s+t}\sigma_i^j(X(r,x,\cdot))dW_r \\ &= \int_0^t b^j(Z_r)dr + \int_0^t \sigma_i^j(Z_r)d(W_{r+s} - W_s)\ ,\end{aligned}$$

therefore, $(Z_t)_{t\geq 0}$ is a strong solution of the stochastic differential equation

$$dZ_t^j = b^j(Z_t)dt + \sigma_i^j(Z_t)dW_t^i \circ \delta_s,\ Z_0 = X(s,x,\cdot)$$

on $(\mathbf{W}_0^d, \mathcal{F}, P)$. By the uniqueness of strong solutions, it must hold that

$$X(t+s,x,\cdot) = X(t, X(s,x,\cdot), \delta_s(\cdot)) \quad P\text{-a.s.} \tag{6.34}$$

To prove the Markov property, we prepare a general lemma.

Lemma 6.5.2 *Let ξ, η be two random variable on some probability* $(\Omega, \mathcal{F}, P)$ *and $\mathcal{G}$ a sub σ-algebra of $\mathcal{F}$. If ξ is $\mathcal{G}$-measurable and η is independent of $\mathcal{G}$, then for any bounded measurable function f,*

$$E(f(\xi,\eta)|\mathcal{G}) = E\left(f(y,\eta)\right)\big|_{y=\xi} \ \ a.s.$$

This lemma can be easily verified for product type $f(x,y) = f_1(x)f_2(y)$ and it then follows from the monotone class theorem.

Proposition 6.5.3 *The strong solution $(X(t,x,\cdot) : t \geq 0)$ of (6.32) possesses Markov property.*

Proof. Let f be a bounded Borel measurable function on $\mathbb{R}^d$. We need to show

$$E\left(f(X(s+t,x,\cdot))|\mathcal{F}_s^0\right) = \left(P_t f\right)(X_s) \text{ a.s.} \tag{6.35}$$

where X_s is a shortcut for $X(s,x,\cdot)$. From (6.34), it follows that

$$E\left\{f(X(t+s,x,\cdot))|\mathcal{F}_s^0\right\} = E\left\{f(X(t,X_s,\delta_s(\cdot)))|\mathcal{F}_s^0\right\}.$$

Now X_s is $\mathcal{F}_s^0$-measurable, δ_s is independent of $\mathcal{F}_s^0$ and P is invariant under δ_s. Applying Lemma above, we obtain

$$\begin{aligned} E\left\{f(X(t,X_s,\delta_s(\cdot)))|\mathcal{F}_s^0\right\} &= E\left(f(X(t,y,\delta_s))\right)|_{y=X_s} \\ &= E(f(X(t,y,\cdot)))|_{y=X_s} \\ &= P_t f(X_s) \end{aligned}$$

which proves the claim. ■

Theorem 6.5.4 *The strong solution $X = (X(t,x,\cdot))_{t\geq 0}$ to the stochastic differential equation (6.32) is a diffusion process.*

Proof. X is a Markov process with continuous sample paths, and and clearly $t \to (R_\alpha f)(X(t,x,\cdot))$ is continuous for any $f \in C^1(\mathbb{R}^d)$, which may generate open sets of $\mathbb{R}^d$. Therefore, according to the remark below Theorem 6.3.4, X has the strong Markov property, or more than that, X is a continuous Borel right process. ■

Actually there is a similar result when the uniqueness in law holds but the proof, see the additional topics, is much harder.

6.6 Additional topics

Shift operators on path-type sample spaces

It is obvious that, if M is a topological space and $(X_t)_{t\geq 0}$ is a *continuous* Markov process with transition function $(P_t)_{t>0}$, then the semigroup $(P_t)_{t>0}$ has a canonical realization on the space Ω of all continuous paths in M. The same remark applies to processes with right-continuous sample paths, or right continuous paths which have left-limits. According to the regularity of the semigroup $(P_t)_{t\geq 0}$, the sample space Ω we will use falls into one of the following spaces of paths, called *path type sample spaces.*

If M is a topological space, then a path ω in M is continuous (resp. right-continuous, resp. right-continuous with left-limits), if ω is continuous on $[0,+\infty)$ (resp. right-continuous on $[0,\infty)$, resp. ω is right-continuous and for every $t > 0$ the left-limit $\lim_{s<t,s\uparrow t}\omega(s)$ exists, and in this case this left-limit is denoted by $\omega(t-)$).

1. Let $(M,\mathcal{M})$ be a measurable space, and $(P_t)_{t\geq 0}$ be a Markov semigroup on $(M,\mathcal{M})$. The sample space we may use is the space of all paths in M.

2. Let M be a topological space (in this book, for simplicity, M is either a locally compact separable metric space or a Polish space), and let $(P_t)_{t\geq 0}$ be a Markov semigroup on $(M, \mathcal{B})$ (such a Markov semigroup is called a *Borel semigroup*). In this case, we expect that a Markov semigroup $(X_t)_{t\geq 0}$ with transition semigroup $(P_t)_{t>0}$ possesses additional regularities, the sample space Ω will be the space of all continuous paths in M, or the space of all paths in M which are right-continuous with left-limits, or the space of all right-continuous paths in M.

Any sample space Ω listed in items 1 and 2 above is called a path type sample space on M.

On a path type sample space Ω on M, we introduce a family $(\theta_t)_{t\geq 0}$ of operators called *shift operators*. For each $t \geq 0$, $\theta_t : \Omega \to \Omega$ defined by $(\theta_t \omega)(s) = \omega(t+s)$ for any $\omega \in \Omega$ and $s \geq 0$. Let $(X_t)_{t\geq 0}$ be the coordinate process on Ω. Then $X_s(\theta_t \omega) = X_{s+t}(\omega)$ for any $s, t \geq 0$ and $\omega \in \Omega$. That is $X_s \circ \theta_t = X_{s+t}$ for any $s, t \geq 0$. Let $\mathcal{F}_t^0 = \sigma\{X_s : s \leq t\}$, $\mathcal{F}^0 = \sigma\{X_t : t \geq 0\}$ and $\mathcal{F}_t' = \sigma\{X_s : s \geq t\}$.

Exercise 6.3 *Prove that*

1. $\theta_t : (\Omega, \mathcal{F}^0) \to (\Omega, \mathcal{F}_t')$ *is a measurable transformation.*
2. *Let* $(X_t)_{t\geq 0}$ *be a Markov process on* $(\Omega, \mathcal{F}^0, \mathcal{F}_t^0, P^\mu)$. *Then* $\theta_t : (\Omega, \mathcal{F}^\mu) \to (\Omega, \mathcal{F}^\mu)$ *is a measurable transformation.*

Exercise 6.4 *If*

$$(\Omega, \mathcal{F}^0, \mathcal{F}_t^0, X_t, \theta_t, P^x)$$

is a canonical realization of a Markov semigroup (P_t), *then*

$$E^x(Y \circ \theta_t | \mathcal{F}_t^0) = E^{X_t}(Y), \ \forall Y \in b\mathcal{F}^0.$$

Infinitesimal generator and martingale property

Let us look at the decomposition in terms of the infinitesimal generator L of the semigroup $(P_t)_{t\geq 0}$ (for detail, see next chapter). Let $f \in b\mathcal{B}$ and $F(t,x) = e^{-\alpha t}u(x)$ where $u = R_\alpha f$ and $\alpha > 0$. From the proof of (6.16) we have seen that

$$P_h u(x) = e^{\alpha h} \int_h^{+\infty} e^{-\alpha s} \left(P_s f\right)(x) ds.$$

Therefore

$$\lim_{h\downarrow 0} \frac{P_h u(x) - u(x)}{h} = \alpha u(x) - f(x) \ .$$

We say that u belongs to the domain of the infinitesimal generator L of the semigroup $(P_t)_{t>0}$, and $Lu = \alpha u - f$. While it is obvious that $\frac{\partial}{\partial t} F = -\alpha F$ and

$$\left(\frac{\partial}{\partial t} + L\right) F = -e^{-\alpha t} f \ .$$

In terms of F, we may rewrite the decomposition (6.18) as

$$F(t,X_t)-F(0,X_0)=M_t^f-M_0^f+\int_0^t\left(\frac{\partial}{\partial t}+L\right)F(s,X_s)ds,$$

namely, if $F(t,x)=e^{-\alpha t}(R_\alpha f)(x)$, then

$$F(t,X_t)-F(0,X_0)-\int_0^t\left(\frac{\partial}{\partial t}+L\right)F(s,X_s)ds$$

is a bounded martingale under P^μ.

Applying the Itô formula (indeed integration by pats) to $e^{\alpha t}$ and $F(t,X_t)$ we have

$$\begin{aligned} u(X_t)-u(X_0) &= \int_0^t e^{\alpha s}dM_s^f-\int_0^t f(X_s)ds\ +\int_0^t \alpha u(X_s)ds \\ &= \int_0^t e^{\alpha s}dM_s^f\ +\int_0^t(\alpha u-f)(X_s)ds \\ &= \int_0^t e^{\alpha s}dM_s^f\ +\int_0^t(Lu)(X_s)ds\ . \end{aligned}$$

Therefore

$$u(X_t)-u(X_0)-\int_0^t(Lu)(X_s)ds\ . \tag{6.36}$$

is a bounded martingale under P^μ, where $Lu=\alpha u-f$.

This argument may not be so convincing since the integration by parts formula is used for M^f which may not be a continuous martingale. Let us give another proof which does not use Itô formula.

Exercise 6.5 *Let $u=R_\alpha f$. Then (6.36) is a bounded martingale under P^μ for any probability μ on M if and only if for each $t\geq 0$ and $x\in M$,*

$$P_tu(x)-u(x)=\int_0^t P_sLu(x)ds. \tag{6.37}$$

However (6.37) is equivalent to

$$P_tR_\alpha-R_\alpha=\int_0^t P_s(\alpha R_\alpha-I)ds. \tag{6.38}$$

Since both sides are right continuous in t, it suffices to prove both sides have the same Laplace transform. Multiplying $e^{-\beta t}$, integrating and applying Fubini theorem, we reach the resolvent equation

$$\beta R_\beta R_\alpha-R_\alpha=\alpha R_\beta R_\alpha-R_\beta.$$

That proves (6.37).

More about strong Markov property

If μ is a given initial distribution, if T is a stopping time with respect to $(\mathcal{F}_t^\mu)_{t\geq 0}$ and if $G : \Omega \times \mathbb{R}^+ \times \Omega \to \mathbb{R}$ is bounded and measurable with respect to the σ-algebra

$$\mathcal{F}_T^\mu \otimes \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{F}^\mu$$

then

$$E^\mu \left\{ G(\cdot, T(\cdot), \theta_{T(\cdot)}(\cdot)) 1_{\{T(\cdot)<+\infty\}} | \mathcal{F}_T^\mu \right\}(\omega) = \int_\Omega G(\omega, T(\omega), \omega') P^{X_{T(\omega)}}(d\omega')$$

on $\{T(\omega) < +\infty\}$, P^μ-a.s. To see the power of this formula, let us use it to prove the reflection principle of Brownian motion. Let

$$G(s, \cdot) = 1_{\{s\leq t, B_{t-s}\leq a\}}.$$

Due to symmetry, $P^b(B_t \leq a) = P^b(B_t \geq 2b - a)$. Hence we have

$$\begin{aligned} P^0(T_b \leq t, B_t \leq a) &= E^0 \left[G(T_b, \theta_{T_b}(\cdot)) \right] \\ &= E^0 \left(E^{B(T_b)}(G(s, \cdot))|_{s=T_b} \right) \\ &= E^0 \left(P^b(s \leq t, B_{t-s} \geq 2b - a)_{s=T_b} \right) \\ &= P^0(T_b \leq t, B_t \geq 2b - a). \end{aligned}$$

Trap of a Markov process

If $(P_t(x, dy) : t > 0)$ is a sub-Markov semigroup on a state space $(M, \mathcal{M})$, then we may construct a Markov process associated with its extension $\{\bar{P}_t(x, dy) : t > 0\}$ on the extended state space $(M_\partial, \mathcal{M}_\partial)$. Let $(X_t)_{t\geq 0}$ be such Markov process with state space M_∂ and semigroup $(\bar{P}_t)_{t>0}$. Recall that $\bar{P}_t(x, A) = P_t(x, A)$ if $x \in M$ and $A \in \mathcal{M}$, $\bar{P}_t(x, \{\partial\}) = 1 - P_t(x, M)$ if $x \neq \partial$, $\bar{P}_t(\partial, M) = 0$ and $\bar{P}_t(\partial, \{\partial\}) = 1$. We might hope that ∂ is an absorbing state (or called a "trap") which means the process will stay in ∂ forever as soon as the process hits it. In that case the time

$$\zeta(\omega) = \inf \{t \geq 0 : X_t(\omega) = \partial\} \tag{6.39}$$

is called the *life-time* of the Markov process $(X_t)_{t\geq 0}$.

However this is not true in general. Without further conditions, the process may come back from ∂, see Exercise 6.2. If $(P_t)_{t\geq 0}$ is a Feller semigroup of sub-Markov kernels on a locally compact metric space M, then $(P_t)_{t\geq 0}$ has a canonical realization on the space of right-continuous paths in M. More precisely, let M_∂ be the one-point compacification of M, so that $M_\partial = M \cup \{\partial\}$. Let $(X_t : t \geq 0)$ be the right continuous (with left limits) realization of $(P_t)_{t\geq 0}$. Then ∂ is a trap for $(X_t : t \geq 0)$. Before proving it, we need a result which says that a non-negative super-martingale will stay zero forever after it hits zero.

Exercise 6.6 *Let Y be a non-negative right continuous super-martingale with left limits. and $\tau := \inf\{t : Y_t \cdot Y_{t-} = 0\}$. Then $Y_t = 0$ if $t > \tau$ almost surely.*

Hint: Set $\tau_n := \inf\{t \geq 0 : Y_t < \frac{1}{n}\}$. Using Doob's optional sampling theorem to prove that$E(Y_t; \tau \leq t) = 0$ for any fixed t and then for all t by the right continuity of Y.

Take $f \in C_0(M)$ which is strictly positive on M. Then $(e^{-\alpha t} R_\alpha f(X_t) : t \geq 0)$ is a right continuous super-martingale and $R_\alpha f \in C_0(M)$ is also strictly positive on M. It vanishes if and only if X_t hits ∂. The conclusion follows.

If $P(\zeta = +\infty) = 1$, then we say the Markov process $(X_t)_{t\geq 0}$ is *conservative.*

Introduction to Hunt processes

We discuss Markov processes which share the main sample path properties of the Feller processes. Let M be a locally compact metric space or a Polish space, and $\mathcal{B}$ the Borel σ-algebra on M. Let $(P_t)_{t\geq 0}$ be a Borel semigroup, i.e. $\{P_t(x,\mathrm{d}y) : t \geq 0\}$ be a family of sub-Markov kernels on $(M, \mathcal{B})$ satisfying the semigroup property

$$P_{t+s}(x, A) = \int_M P_t(z, A) P_s(x, \mathrm{d}z) \qquad \forall s, t \geq 0$$

and for any $f \in b\mathcal{B}$ function $(t, x) \to (P_t f)(x)$ is Borel measurable.

Let

$$(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, P^x)$$

be a homogenous Markov process taking values in M_∂ with transition function $(P_t)_{t\geq 0}$, where $\mathcal{F}$, $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of the Markov process $(X_t)_{t\geq 0}$. Let $\mathcal{F}_t^0 = \sigma\{X_s : s \leq t\}$. $\mathbf{X}$ is called a *Hunt process* if the following conditions are satisfied.

1. $(X_t)_{t\geq 0}$ is *right-continuous.*

2. $\mathbf{X}$ has the *strong Markov property*: for any $(\mathcal{F}_{t+}^0)$-stopping time T , any bounded Borel measurable function f and any initial distribution μ

$$E^\mu \left(f(X_t) \circ \theta_T 1_{\{T<+\infty\}} | \mathcal{F}_{T+}^0 \right) = (P_t f)(X_T) 1_{\{T<+\infty\}} \qquad P^\mu\text{-a.s.} \tag{6.40}$$

3. $(X_t)_{t\geq 0}$ is *quasi-left continuous*: for any initial distribution μ, if $\{T_n : n \geq 1\}$ is an increasing sequence of $(\mathcal{F}_t^\mu)$-stopping times which converges to T on $\{T < +\infty\}$ P^μ-a.s., then

$$\lim_{n\to\infty} X_{T_n} = X_T \quad \text{on } \{T < +\infty\} \qquad P^\mu\text{-a.s.}$$

In literature, if $\{T < +\infty\}$ is replaced by $\{T < \zeta\}$ (where ζ is the life-time) in condition 3 above, then $\mathbf{X}$ is called a *standard process.*

Condition 2 implies that $(X_t)_{t\geq 0}$ has Markov property with respect to the filtration $(\mathcal{F}_{t+}^0)_{t\geq 0}$. Prove that for a Hunt process, the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ and $(\mathcal{F}_t^\mu)_{t\geq 0}$ for any initial distribution μ are right-continuous, by a similar argument used to prove Theorem 6.2.8.

The quasi-left continuity is somewhat between left continuity and having left limits. If $\mathbf{X}$ is a Hunt process, then for any initial distribution μ, under P^μ almost all sample paths of $(X_t)_{t\geq 0}$ have left-limits on $(0,+\infty)$. As a consequence, Hunt processes retain all major sample properties we have established for Feller processes.

Diffusion processes and a continuity condition

A Markov process

$$\mathbf{X} = (\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, P^x)$$

with Borel semigroup $(P_t)_{t\geq 0}$ in the state space $(M, \mathcal{B})$ is a *diffusion process* if it has the strong Markov property and has continuous sample paths. Of course, a diffusion process is quasi-left continuous, so that a diffusion process must be a Hunt process. We now give a nice sufficient condition for sample path to be continuous in terms of transition function.

Exercise 6.7 *This is a nice analysis exercise. Let (M,d) be a complete metric space, and $h : [0,1] \to M$ is right continuous on $[0,1)$ and has left-limits on $(0,1]$. Then h is not continuous on $[0,1]$ if and only if there is a positive number ε and a natural number n_0 (depending on ε) such that*

$$\max_{0\leq k\leq n-1} d(h(\frac{k+1}{n}), h(\frac{k}{n})) \geq \varepsilon \ . \tag{6.41}$$

If h is continuous, then h is uniformly continuous, so we cannot have (6.41) for any $\varepsilon > 0$. For any $t \in (0,1]$, then for any n we may choose k_n such that $t \in (k_n/n, (k_n+1)/n]$. By the right continuity

$$\lim_{n\to\infty} h((k_n+1)/n) = h(t)$$

and by definition

$$\lim_{n\to\infty} h(k_n/n) = h(t-) \ .$$

Therefore

$$\lim_{n\to\infty} d(h(\frac{k_n+1}{n}), h(\frac{k_n}{n})) = d(h(t), h(t-))$$

and thus

$$\lim_{n\to\infty} \max_{0\leq k\leq n-1} d(h(\frac{k+1}{n}), h(\frac{k}{n})) \geq d(h(t), h(t-)) \ .$$

Hence, if there is a point $t_0 \in (0,1]$ at which $h(t_0) \neq h(t_0-)$, so that $\varepsilon = d(h(t_0), h(t_0-)) > 0$ and

$$\lim_{n\to\infty} \max_{0\leq k\leq n-1} d(h(\frac{k+1}{n}), h(\frac{k}{n})) \geq \varepsilon \ .$$

Theorem 6.6.1 *Let $(\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, P^x)$ be a Markov process in a complete metric space M, whose sample paths are right continuous, and let $(P_t)_{t\geq 0}$ be its transition semigroup. Suppose $(P_t)_{t\geq 0}$ satisfies the following condition: for any compact subset $K \subset M$ and any $\varepsilon > 0$*

$$\lim_{t\downarrow 0} \frac{1}{t} \sup_{x\in K} (1 - P_t(x, B_x(\varepsilon)) = 0 \tag{6.42}$$

where $B_x(\varepsilon)$ is the metric ball $\{y \in M : d(x, y) \leq \varepsilon\}$. Then almost all sample paths of $(X_t)_{t\geq 0}$ are continuous.

For a proof refer to K. L. Chung [3] and Blumenthal & Getoor [2].

Uniqueness and strong Markov property

We begin with the definition of a *regular conditional probability.*

Definition 6.6.2 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra. A regular conditional probability of P given $\mathcal{G}$ is a map $Q(\omega, A)$: $\Omega \times \mathcal{F} \to [0, 1]$ such that*

1) For any $\omega \in \Omega$, $Q(\omega, \cdot)$ is a probability measure on $(\Omega, \mathcal{F})$.

2) For any $A \in \mathcal{F}$, the map $\omega \to Q(\omega, A)$ is $\mathcal{G}$-measurable.

3) For any $f \in L^1(\Omega, \mathcal{F}, P)$,

$$E\,(f|\mathcal{G}) = \int_\Omega f(\omega) Q(\cdot, d\omega) \quad P\text{-a.s.}$$

In general a regular conditional probability needs not exist. We say regular conditional probability given $\mathcal{G}$ is unique, if Q_1 and Q_2 are two regular conditional probabilities given $\mathcal{G}$, then there is a null set $N \in \mathcal{F}$ such that

$$Q_1(\omega, \cdot) = Q_2(\omega, \cdot), \quad \forall \omega \notin N.$$

Theorem 6.6.3 *Suppose Ω is a Polish space (a Polish space is a complete and separable metric space), and $\mathcal{F} = \mathcal{B}(\Omega)$ is the Borel σ-field. Let P be a probability on $(\Omega, \mathcal{F})$ and $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra. Then there exists a unique regular conditional probability given $\mathcal{G}$.*

Proof. 1) There is a equivalent metric ρ on Ω, such that (Ω, ρ) is totally bounded, i.e. for any $\varepsilon > 0$, there exist finite points $x_1, \cdots, x_n \in \Omega$, such that

$$\cup_{i=1}^n B_{x_i}(\varepsilon) = \Omega,$$

where

$$B_{x_i}(\varepsilon) = \{x \colon \rho(x, x_i) < \varepsilon\},$$

so that the completion $\overline{\Omega}$ of Ω under the new metric ρ is compact.

2) Denote by $U_\rho(\Omega)$ the set of all uniformly continuous and bounded functions on Ω. The set $U_\rho(\Omega)$ is separable under the norm

$$\|f\| = \sup_{x\in\Omega} |f(x)|.$$

3) Any function f in $U_\rho(\Omega)$ has a unique continuous extension on $\overline{\Omega}$, so that

$$C(\overline{\Omega}) = U_\rho(\Omega).$$

4) It is clear that $\mathcal{B}(\Omega) = \sigma\{f \in U_\rho(\Omega)\}$.

5) Choosing a sequence of functions $\{f_i\}$ in $U_\rho(\Omega)$, such that

$$\begin{aligned}
&f_1 = 1,\\
&\{f_i\} \text{ linear independent}\\
&\text{span}\{f_i\colon i = 1, 2, \cdots,\} \text{ is dense in } U_\rho(\Omega)\ .
\end{aligned}$$

6) For any $i = 1, \cdots$, choosing a version of $E(f_i|\mathcal{G})$ denoted by g_i, $i \geq 2$, and $g_1 = 1$. Then

$$g_i \in \mathrm{b}\mathcal{G}, \quad g_1 = 1,$$

i.e. each g_i is bounded $\mathcal{G}$-measurable function.

7) For any n, denote by Q_n the set of all $(r_1, \cdots, r_n)$, such that $r_i \in Q$ (the set of all rational numbers), and

$$\sum_{i=1}^{n} r_i f_i(x) \geq 0, \quad \forall x \in \Omega,$$

so that Q_n is countable.

8) For any $(r_1, \cdots, r_n) \in Q_n$, we have

$$\sum_{i=1}^{n} r_i g_i(x) \geq 0, \quad P\text{-a.s. } x \in \Omega$$

so that

$$F_{(r_1,\cdots,r_n)} = \left\{x \in \Omega\colon \sum_{i=1}^{n} r_i g_i(x) < 0\right\}$$

is a P-zero set. Let

$$F = \cup_{n=1}^{\infty} \cup_{(r_1,\cdots,r_n)\in Q_n} F_{(r_1,\cdots,r_n)}\ .$$

Then $P(F) = 0$.

9) For any $x \in \Omega \setminus F$, let

$$L_x f = t_1 g_1(x) + \cdots + t_n g_n(x),$$

if $f = t_1 f_1 + \cdots + t_n f_n$, in the span$\{f_i\colon i = 1, \cdots\}$. We claim that $L_x\colon U_\rho(\Omega) \to \mathbb{R}$ is a positive functional. In fact, if $f \geq 0$, i.e.

$$t_1 f_1 + \cdots + t_n f_n \geq 0.$$

Since each f_i is uniformly continuous, so that for any $\varepsilon > 0$, there is a $(r_1, \cdots, r_n) \in Q_n$ such that

$$r_1 f_1 + \cdots + r_n f_n \geq -\varepsilon, \quad |t_i - r_i| < \varepsilon,$$

so that

$$(r_1+\varepsilon)f_1+\cdots+r_nf_n\geq 0.$$

Thus for any $\varepsilon>0$, $\varepsilon\in Q$,

$$(r_1+\varepsilon)g_1+\cdots+r_ng_n\geq 0.$$

Letting $\varepsilon\to 0$, we get

$$L_xf=t_1g_1(x)+\cdots+t_ng_n(x)\geq 0.$$

10) Since $U_\rho(\Omega)=C(\overline{\Omega})$, so that L_x is a positive functional on $C(\overline{\Omega})$ and $L_x(f_1)=1$. by Riesz representation theorem, for any $x\in\Omega\setminus F$, there is a probability measure on $\overline{\Omega}$, such that

$$\int_{\overline{\Omega}}f(y)Q_x(dy)=L_x(f),\quad\forall f\in C(\overline{\Omega}).$$

However

$$\int_{\overline{\Omega}}f_i(y)Q_x(dy)=g_i(x),\quad\forall x\in\Omega\setminus F$$

so that

$$x\to\int_{\overline{\Omega}}f(y)Q_x(dy)$$

is $\mathcal{G}$-measurable, for any $f\in U_\rho(\Omega)$.

11) By definition,

$$E^P\left\{\int f(y)Q.(dy),A\cap(X-F)\right\}=E^P\{f(\cdot),A\}$$

for any $A\in\mathcal{G}$ and $f\in U_\rho(\Omega)$. Therefore for any compact subset $K\in\Omega$, choosing $\psi_n\in U_\rho(\Omega)$, such that $\psi_n\to 1_K$, so that

$$E^P(Q.(K),\Omega-F)=P(K).$$

Now choosing compact subset $K_n\in\Omega$, so that

$$K_n\subset K_{n+1},\quad P(K_n)\geq 1-\frac{1}{n}.$$

Then $D=\cup K_n\in\mathcal{B}(\Omega)$ and

$$E^P(Q.(D),\Omega-F)=1\ .$$

Thus there is a zero set F', such that

$$Q_x(D)=1,\quad\forall x\in\Omega-(F\cup F')$$

that is $Q_x(dy)$ is a probability measure $(\Omega,\mathcal{B}(\Omega))$ for any $x\in\Omega\setminus(F\cup F')$.

12) For any $x \in F \cup F'$, we define $Q_x(dy) = P(dy)$. Then

$$x \to \int_\Omega f(y) Q_x(dy), \quad \forall f \in U_\rho(\Omega)$$

is $\mathcal{G}$-measurable, and since $P(F \cup F') = 0$, so that

$$E^P\left\{\int f(y) Q.(dy), A\right\} = E^P\left(f(\cdot), A\right)$$

for any $A \in \mathcal{G}$, $f \in U_\rho(\Omega)$, so that

$$E(f|\mathcal{G}) = \int f(y) Q.(dy) \qquad P\text{-a.s.}$$

■

Corollary 6.6.4 *Keep the same assumptions on $(\Omega, \mathcal{F}, P)$ in Theorem 6.6.3. Let*

$$X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$$

be a measurable, where $(S, \mathcal{S})$ is a measurable space. Denote by μ_X the law of X on $(S, ß)$ defined by

$$\mu_X(B) \widehat{=} P\left\{\omega \in \Omega : \ X(\omega) \in B\right\}, \quad \forall B \in \mathcal{S}.$$

Then there is a map $Q(x, A) : S \times \mathcal{F} \to [0, 1]$, called a regular conditional probability for $\mathcal{F}$ given X, such that

1) For any $x \in S$, $Q(x, \cdot)$ is a probability on $(\Omega, \mathcal{F})$.

2) For any $A \in \mathcal{F}$, $x \to Q(x, A)$ is $\mathcal{S}$-measurable.

3) For any $A \in \mathcal{F}$,

$$Q(x, A) = P\left\{A | X = x\right\}, \quad \mu_X\text{-a.s. } \ x \in S \ .$$

Moreover, such a regular conditional probability is unique in the following sense, if $\widetilde{Q}$ is another regular conditional probability, then there is $N \in \mathcal{S}$ such that $\mu_X(N) = 0$, and

$$\widetilde{Q}(x, \cdot) = Q(x, \cdot), \quad \forall x \notin N.$$

Now we are in a position to state the relation between uniqueness to martingale problem (i.e. uniqueness in law) and strong Markov property. As before, we let

$$L = \frac{1}{2} \sum_{i,j=1}^N a^{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^N b^i(x) \frac{\partial}{\partial x_i} \ .$$

Let $\Omega = W(\mathbb{R}^+, \mathbb{R}^N)$ the space of all continuous function on $[0, +\infty)$ with values in $\mathbb{R}^N$, and $x(t)$ be the coordinate process:

$$x(t)(\omega) = \omega(t), \quad \forall t \geq 0, \omega \in \Omega$$

and

$$\mathcal{F}_t = \sigma\{x(s) : s \leq t\}, \quad \mathcal{F} = \sigma\{x(t) : t \geq 0\}.$$

A probability P on $(\Omega, \mathcal{F})$ is called a solution of the L-martingale problem, if for any $f \in C_0^2(\mathbb{R}^N, \mathbb{R})$

$$M_t^f = f(x(t)) - f(x(0)) - \int_0^t Lf(x(s))\mathrm{d}s$$

is a continuous martingale under P. We say L-martingale problem is well-posed if for any $x \in \mathbb{R}^N$, there is a unique solution P^x of the L-martingale problem such that

$$P^x\{x(0) = x\} = 1.$$

Example 6.6.5 *If A_i, $i = 0, \cdots, n$ are $n+1$ vector fields satisfying global Lipschitz condition, then*

$$dX_t = \sum_{i=1}^n A_i(X_t)dB_t^i + A_0(X_t)dt, \quad X_0 = x$$

possesses a unique strong solution, so that the uniqueness in law holds. Hence L-martingale problem is well-posed, where

$$A_k(x) = \sum_{j=1}^N \sigma_k^j(x)\frac{\partial}{\partial x_j}; \quad A_0(x) = \sum_{j=1}^N b^j(x)\frac{\partial}{\partial x_j}$$

$$a^{ij} = \sum_{l=1}^n \sigma_l^i \sigma_l^j$$

and

$$L = \sum_{i,j=1}^N a^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{j=1}^N b^j(x)\frac{\partial}{\partial x_j} .$$

In terms of martingale problem we have

Theorem 6.6.6 *If L-martingale problem is well-posed, then $(\Omega, \mathcal{F}, \mathcal{F}_t, x(t), P^x)$ is a strong Markov process.*

Before the proof of Theorem 6.6.6, we need more notation. We use θ_t to denote the shift operator, i.e. $\theta_t : \Omega \to \Omega$, $(\theta_t\omega)(s) = \omega(t+s)$ for any $\omega \in \Omega$. Similarly we define θ_T for any finite stopping time T, i.e.

$$(\theta_T\omega)(s) = \omega(s + T(\omega)), \quad \forall \omega \in \Omega.$$

Same shift operator θ_T applies to measurable functions on Ω.

Definition 6.6.7 *We say $(\Omega, \mathcal{F}, \mathcal{F}_t, x_t, \theta_t, P^x)$ has strong Markov property, if*

$$E^x\left(f(x_{t+T})|\mathcal{F}_T\right) = E^{x_T} f(x_t) \qquad P^x\text{-}a.s. \tag{6.43}$$

for any $x \in \mathbb{R}^N$, any bounded stopping time T, and bounded Borel measurable function f, where

$$E^{X_T}\left\{f(x_t)\right\} = E^z\left\{f(x_t)\right\}|_{z=x_T} .$$

In the sequel we fix a x, and let $Q(\omega, d\omega')$ be the unique regular conditional probability given $\mathcal{F}_T$, where T is a fixed bounded stopping time, so that

$$\int F(\omega')Q(\omega, d\omega') = E^x(F|\mathcal{F}_T)(\omega) \quad P^x\text{-a.s.}$$

Proposition 6.6.8 *Under above notation, $(M_t^f)_{t\geq 0}$ is a continuous martingale under the probability $Q(\omega, d\omega') \circ \theta_T$ for P^x-a.e. $\omega \in \Omega$.*

Proof. We have to prove that

$$Q(\omega, d\omega') \circ \theta_T^{-1}(M_t^f 1_A) = Q(\omega, d\omega') \circ \theta_T^{-1}(M_s^f 1_A)$$

P^x-a.e. for any $t > s$, $A \in \mathcal{F}_s$, i.e.

$$P^x\left(M_{t+T}^f 1_A \circ \theta_T|\mathcal{F}_T\right) = P^x\left(M_{s+T}^f 1_A \circ \theta_T|\mathcal{F}_T\right).$$

However $1_A \circ \theta_T \in \mathcal{F}_{s+T}$, so that

$$\begin{aligned}
&P^x\left(M_{t+T}^f 1_A \circ \theta_T|\mathcal{F}_T\right) \\
&= P^x\left\{P^x\left(M_{t+T}^f 1_A \circ \theta_T|\mathcal{F}_{s+T}\right)|\mathcal{F}_T\right\} \\
&= P^x\left(M_{s+T}^f 1_A \circ \theta_T|\mathcal{F}_T\right),
\end{aligned}$$

where we have used Doob's stopping theorem. ∎

Lemma 6.6.9 *For any $\omega \in \Omega$, for $Q(\omega, d\omega') \circ \theta_T^{-1}$-a.e. ω',*

$$x_0(\omega') = x_{T(\omega)}(\omega).$$

Proof. We have to prove that for any f,

$$Q(\omega, d\omega') \circ \theta_T^{-1}(f(x_0)) = f(x_{T(\omega)}(\omega)) .$$

While

$$\begin{aligned}
Q(\cdot, d\omega') \circ \theta_T^{-1}(f(x_0)) &= E^x\left(f(x_0) \circ \theta_T|\mathcal{F}_T\right) \\
&= E^x\left(f(x_T)|\mathcal{F}_T\right) \\
&= f(x_T) .
\end{aligned}$$

∎

Corollary 6.6.10 *For P^x-a.e. $\omega \in \Omega$, $Q(\omega, d\omega') \circ \theta_T^{-1}$ is a solution of L-martingale problem with initial $x_{T(\omega)}(\omega)$.*

Thus if L-martingale is well posed, then for P^x-a.e. ω,

$$Q(\omega, d\omega') \circ \theta_T^{-1} = P^{x_{T(\omega)}(\omega)}(d\omega')$$

which implies that

$$E^x\left(f(x_{t+T})|\mathcal{F}_T\right) = P^{X_T}(f(x_t)) \quad P^x\text{-a.s.}$$

Hence we have completed the proof of Theorem 6.6.6.

Corollary 6.6.11 *If A_i, $i = 0, \cdots, d$ are $d+1$ vector fields satisfying Lipschitz condition, then for any $x \in \mathbb{R}^N$, the unique strong solution with given Brownian motion $B = (B_t)$ of the following stochastic differential equation,*

$$dX_t = \sum_{i=1}^{n} A_i(X_t)dB_t^i + A_0(X_t)dt, \quad X_0 = x$$

is a diffusion process: a continuous Markov process which possesses the strong Markov property.

Stroock-Varadhan's Theorem

The following is a solution of a martingale problem which was proved by Stroock and Varadhan.

Theorem 6.6.12 (Stroock & Varadhan, 1969) *Let*

$$L = \frac{1}{2}\sum_{i,j=1}^{N} a^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{N} b^i(x)\frac{\partial}{\partial x_i}$$

where a^{ij} are bounded, uniformly continuous functions on $\mathbb{R}^N$ satisfying the following condition:

$$\sum_{i,j=1}^{N} a^{ij}(x)\xi^i\xi^j \geq \lambda|\xi|^2, \quad \forall \xi = (\xi_i) \in \mathbb{R}^N$$

for some $\lambda > 0$, i.e. L is uniformly elliptic, and b^i are bounded Borel measurable on $\mathbb{R}^N$. Then the L-martingale problem is well-posed, so that there is a unique L-diffusion $(\Omega, \mathcal{F}, \mathcal{F}_t, x(t), \theta_t, \mathbb{P}^x)$.

Outline of the proof. Existence. Using **weak convergence method**. We briefly describe this method as following.

1. By the Cameron-Martin formula, we only need to solve the martingale problem associated to
$$L = \frac{1}{2}\sum_{i,j=1}^N a^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} \ .$$

2. Mollifying a^{ij}. Let ρ be a non-negative smooth function with compact support on $\mathbb{R}^N$ (e.g. $\mathrm{Supp}(\rho) \subset$ the curb $[-1,1]^N$) satisfying
$$\int_{\mathbb{R}^N} \rho(x)dx = 1 \ .$$
Set $\rho_n(x) = n^{-N}\rho(x/n)$, so that $\int \rho_n = 1$, and let
$$a_n^{ij} = a^{ij} * \rho_n,$$
i.e.
$$a_n^{ij}(x) = \int_{R^N} a^{ij}(y)\rho_n(x-y)dy \ .$$

3. Since a^{ij} is uniformly continuous, so that $a_n^{ij} \to a^{ij}$ uniformly on $\mathbb{R}^N$ as $n \to \infty$.

4. Each $a_n = (a_n^{ij})$ is uniformly elliptic, i.e.
$$\lambda|\xi|^2 \le \sum_{i,j=1}^N a_n^{ij}(x)\xi^i\xi^j \le \lambda^{-1}|\xi|^2, \quad \forall \xi \in \mathbb{R}^N \ ,$$
so that there is an $A_n = (A_{i,j}^n)$ such that
$$\sum_{i,j=1}^N A_{il}^n A_{lj}^n = a_n^{ij}$$
where A_{ij}^n are bounded and smooth.

5. Solve the following stochastic differential equation:
$$dX_t^j = \sum_{i=1}^N A_{ij}^n(X_t)\mathrm{d}B_t^i, \quad X_0 = x.$$
Denote by X_n the unique strong solution of the above stochastic differential equation.

6. Let P_n be the law of X_n on $C(\mathbb{R}^+, \mathbb{R}^N)$, i.e. $P_n = P \circ X_n^{-1}$.

7. By 3) and 4), we conclude that there is a subsequence of (P_n), such that
$$P_{n_k} \to P^x \ \ \text{weakly.}$$

8. By 3) and 4) again, we can check that P^x is a solution of the L-martingale problem.

Uniqueness. Hard part, using results coming from analysis!

Chapter 7

Analysis of Markov semigroups

In this part we study the analytic aspect of Markov semigroups and their associated Markov processes. We outline the beautiful theory of Markov semigroups, which is the natural product by combining Hille-Yosida's theory of one-parameter semigroups with the Markov property.

We begin with a short summary of Hille-Yosida's theory of semigroups on Banach spaces, which is necessary for the study of Markov semigroups. We then present a few special features about symmetric Markov semigroups and their associated Dirichlet spaces.

7.1 Contraction semigroups

Recall that a time-homogenous Markov chain $(X_t)_{t\geq 0}$ on a discrete state space M is described through its transition probability

$$p_{ij}(t) = P\left(X_t = j | X_0 = i\right).$$

The transition matrix $P(t) = (p_{ij}(t))$ satisfies the Chapman-Kolmogorov equation

$$p_{ij}(s+t) = \sum_{k\in M} p_{ik}(s)p_{kj}(t) \qquad \text{for any} \quad s,\, t > 0\ .$$

The transition matrix $(p_{ij}(t))$ allows us to define a linear operator P_t on the space $C_b(M)$ of bounded (continuous) functions on M by

$$\left(P_t f\right)(i) = \sum_{j\in M} f(j)p_{ij}(t) \qquad \forall i \in M\ .$$

If $C_b(M)$ is endowed with the supremum norm

$$||f|| = \sup_{i\in M} |f(i)| \qquad \forall f \in C_b(M)$$

then $C_b(M)$ is a Banach space, and each linear operator $P_t : C_b(M) \to C_b(M)$ is a contraction $||P_t f|| \leq ||f||$ for any $f \in C_b(M)$. The Chapman-Kolmogorov equation implies that $(P_t)_{t>0}$ is a *semigroup* on $C_b(M)$, i.e. $P_t(P_s f) = P_{t+s} f$. It is thus not surprising that the theory of 1-parameter semigroups plays a fundamental rôle in the theory of Markov processes.

7.1.1 Contraction semigroups on Banach spaces

Let B be a (real or complex) Banach space with a norm $||\cdot||$. Typical examples are L^p-spaces on σ-finite measure spaces. A linear operator $T : B \to B$ is bounded if there is a non-negative constant C such that $||T(x)|| \leq C||x||$ for any $x \in B$. In this case, the least $C \geq 0$ such that the previous statement is true is called the norm of T, denoted by $||T||$. A basic fact in functional analysis is that a linear operator is continuous if and only if it is bounded. For simplicity, if no confusion may arises, $T(x)$ will be simply written as Tx.

A linear operator $T : B \to B$ is called a contraction if $||T|| \leq 1$.

A one-parameter family $(P_t)_{t\geq 0}$ of bounded linear operators $P_t : B \to B$ is a semigroup (of linear operators) on B, if

1. $P_0 = I$ the identity operator on B,
2. $(P_t)_{t\geq 0}$ satisfies the semigroup property: $P_{t+s} = P_t P_s$ for every s, $t \geq 0$, where $P_t P_s$ is the composition of operators, that is, $P_t P_s(f) = P_t(P_s(f))$.

A semigroup $(P_t)_{t\geq 0}$ is *strongly continuous* if

$$\lim_{t\downarrow 0} P_t x = x \quad \text{for every } x \in B \ .$$

A semigroup $(P_t)_{t\geq 0}$ is a semigroup of contractions (or called *contraction semigroup*) if each P_t is a contraction on B, that is, $||P_t|| \leq 1$.

If $(P_t)_{t\geq 0}$ is a semigroup on B, then its infinitesimal generator (or simply generator) is the (unbounded) linear operator $(L, D(L))$ defined by

$$Lx = \lim_{t\downarrow 0} \frac{1}{t} (P_t x - x)$$

where $x \in D(L)$, and

$$D(L) = \left\{ x \in B : \lim_{t\downarrow 0} \frac{1}{t} (P_t x - x) \text{ exists} \right\} .$$

The following theorem summarizes the basic properties of the infinitesimal generator of a semigroup, the proofs leave for the reader as an exercise.

Theorem 7.1.1 *Let $(P_t)_{t\geq 0}$ be a strongly continuous semigroup on Banach space B, and let L be its infinitesimal generator with domain $D(L)$.*

1) The map $t \to P_t x$ is uniformly continuous for every $x \in B$, and

$$\lim_{h\downarrow 0} \frac{1}{h} \int_t^{t+h} (P_s x)\, ds = P_t x \quad \forall t \geq 0, \;\; x \in B. \tag{7.1}$$

Moreover, $\int_0^t (P_s x) ds \in D(L)$ *for* $t > 0$, $x \in B$, *and*

$$L \int_0^t (P_s x)\, ds = P_t x - x\ .$$

2) If $x \in D(L)$, *so is* $P_t x$,

$$\frac{d}{dt} P_t x = L(P_t x) = P_t(Lx) \tag{7.2}$$

and

$$\begin{aligned} P_t x - P_s x &= \int_s^t (P_u L x)\, du \\ &= \int_s^t (L P_u x)\, du = L\left(\int_s^t (P_u x)\, du\right)\ . \end{aligned} \tag{7.3}$$

As a consequence, $D(L)$ is dense in B. Therefore the infinitesimal generator L of a strongly continuous semigroup on a Banach space B is densely defined linear operator on B, each P_t leaves $D(L)$ invariant, L commutates with P_t, and for any $x \in B$ the integral $\int_0^t (P_s x) \mathrm{d}s$ (for $t > 0$) is an element in $D(L)$.

Generally, the infinitesimal generator L of a strongly continuous semigroup on a Banach space B is unbounded, and in general $D(L) \neq B$. To further investigate the properties of the densely defined linear operator L, we need the following

Definition 7.1.2 *The graph of a densely defined linear operator* $(T, D(T))$ *on a Banach space* B *is*

$$G(T) = \{(x, Tx) : x \in D(T)\}$$

which is a subset of the product space $B \times B$ *(endowed with norm* $||(x, y)|| = ||x|| + ||y||$*).* T *is called a* closed operator *if* $G(T)$ *is a closed subset of* $B \times B$ *(and thus* $G(T)$ *itself is a Banach space). In other words,* T *is closed if for every sequence* $\{x_n\}$ *of* $D(T)$ *such that* $x_n \to x$ *and* $Tx_n \to y$*, then* $x \in D(T)$ *and* $Tx = y$*.*

Proposition 7.1.3 *The infinitesimal generator* $(L, D(L))$ *of a strongly continuous contraction semigroup* $(P_t)_{t \geq 0}$ *is a closed operator.*

Proof. Suppose $x_n \in D(L) \to x$ and $Lx_n \to y$. We have to show that $x \in D(L)$ and $Lx = y$. Since $Lx_n \to y$, $\{Lx_n\}$ is bounded in B and

$$||P_s L(x_n)|| \leq ||L(x_n)|| \leq \sup ||L(x_n)|| \qquad \text{for all } s \geq 0.$$

A computation leads us to

$$\begin{aligned}
\frac{P_t x - x}{t} &= \frac{1}{t}\left(P_t(\lim_{n\to\infty} x_n) - \lim_{n\to\infty} x_n\right) \\
&= \lim_{n\to\infty} \frac{1}{t}\left(P_t(x_n) - x_n\right) && \text{(as } P_t \text{ is continuous)} \\
&= \lim_{n\to\infty} \frac{1}{t}\int_0^t P_s L(x_n) \mathrm{d}s \\
&= \frac{1}{t}\int_0^t \lim_{n\to\infty} P_s L(x_n) \mathrm{d}s && \text{(Dominated Convergence)} \\
&= \frac{1}{t}\int_0^t P_s y \mathrm{d}s
\end{aligned}$$

and

$$\lim_{t\downarrow 0} \frac{1}{t}\left(P_t x - x\right) = \lim_{t\downarrow 0} \frac{1}{t}\int_0^t P_s y \mathrm{d}s = y \ .$$

Therefore $x \in D(L)$ and $Lx = y$. ■

It is usually very difficult to determine the domain $D(L)$ of the infinitesimal generator L. Since the graph $G(L)$ of L obviously determines L uniquely, and since $G(L)$ is a closed subspace of $B\times B$, any dense subset of $G(L)$ will determine $G(L)$ and therefore the closed linear operator L uniquely. A subset C of $D(L)$ is a *core* for a closed linear operator $(L, D(L))$ if $\{(x, Lx) : x \in C\}$ is dense in $G(L)$. Precisely, for any $x \in D(L)$ there is a sequence $\{x_n\}$ in C such that $x_n \to x$ and $Lx_n \to y$ for some $y \in B$.

Another important concept associated with a strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ is the *resolvent* $\{R_\lambda : \lambda > 0\}$ which we have met in the previous chapter. By definition

$$R_\lambda = \int_0^\infty e^{-\lambda t} P_t \mathrm{d}t \ .$$

Since for every $x \in B$

$$\begin{aligned}
||R_\lambda x|| &= \left\|\int_0^\infty e^{-\lambda t}\left(P_t x\right) \mathrm{d}t\right\| \\
&\leq \frac{1}{\lambda}\|x\| ,
\end{aligned}$$

each R_λ (for $\lambda > 0$) is a bounded linear operator of B with $||R_\lambda|| \leq 1/\lambda$. $\{R_\lambda : \lambda > 0\}$ is a commutative family of bounded linear operators of B, and the semigroup property of $(P_t)_{t\geq 0}$ implies that $\{R_\lambda : \lambda > 0\}$ satisfies the resolvent equation:

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu \quad \forall \lambda, \mu > 0 \ .$$

In particular, the region $\{R_\lambda x : x \in B\}$ does not depend on $\lambda > 0$.

Proposition 7.1.4 *If $(L, D(L))$ is the infinitesimal generator of a strongly continuous semigroup $(P_t)_{t\geq 0}$ of contractions of B, then for any $\lambda > 0$, $\lambda - L$ (where λ means the multiplier λI) is invertible and $R_\lambda = (\lambda - L)^{-1}$. In particular, for every $\lambda > 0$ the region of R_λ*

$$\{R_\lambda x : x \in B\} \subset D(L) \ .$$

Proof. We only need to show that for every $x \in B$

$$(\lambda - L)(R_\lambda x) = x \ .$$

Firstly show that $R_\lambda x$. In fact

$$\begin{aligned} P_h\left(R_\lambda x\right) - R_\lambda x &= P_h \int_0^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t - \int_0^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t \\ &= \int_0^\infty \mathrm{e}^{-\lambda t}\left(P_{t+h} x\right) \mathrm{d}t - \int_0^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t \\ &= \mathrm{e}^{\lambda h} \int_h^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t - \int_0^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t \\ &= \left(\mathrm{e}^{\lambda h} - 1\right) \int_h^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t - \int_0^h \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t \end{aligned}$$

and we then have

$$\begin{aligned} \lim_{h\downarrow 0} \frac{1}{h}\left(P_h\left(R_\lambda x\right) - R_\lambda x\right) &= \lim_{h\downarrow 0} \frac{\mathrm{e}^{\lambda h} - 1}{h} \int_h^\infty \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t \\ &\quad - \lim_{h\downarrow 0} \frac{1}{h} \int_0^h \mathrm{e}^{-\lambda t}\left(P_t x\right) \mathrm{d}t \\ &= \lambda R_\lambda x - x \ . \end{aligned}$$

Therefore $R_\lambda x \in D(L)$ and

$$L\left(R_\lambda x\right) = \lambda R_\lambda x - x$$

which proves the claim. ■

Therefore any real $\lambda > 0$ belongs to the resolvent set of L. The resolvent set $\rho(L)$ of L is the set of all complex numbers λ for which $\lambda I - L$ is invertible, i.e. $(\lambda I - L)^{-1}$ is a bounded linear operator on B. The family

$$\{R_\lambda \equiv (\lambda I - L)^{-1} : \lambda \in \rho(L)\}$$

of bounded linear operators is also called the resolvent of L. The complement of the resolvent set of L is called the spectrum of L denoted by $\sigma(L)$.

The necessary and sufficient condition for a given densely defined linear operator L (with domain $D(L)$) to be the infinitesimal generator of some strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ is known as Hille-Yosida's theory of one-parameter semigroups.

Theorem 7.1.5 (Hille-Yosida) *A linear (unbounded) operator L is the infinitesimal generator of a strongly continuous semigroup of contractions on a Banach space B if and only if*

1) L is closed and the $D(L)$ is dense in B, i.e. L is a densely defined closed operator.

2) $(0,+\infty) \subset \rho(L)$ and $||R_\lambda|| \leq 1/\lambda$ for every $\lambda > 0$. In other words

$$||(\lambda - L)x|| \geq \lambda ||x||$$

for every $\lambda > 0$ and $x \in D(L)$.

We have proven the necessity of two conditions 1 and 2 (see Proposition 7.1.4) and

$$(\lambda I - L)^{-1} = \int_0^\infty e^{-\lambda t} P_t dt \qquad \forall \lambda > 0,$$

namely, the resolvent of L is the Laplace transform of its semigroup $(P_t)_{t\geq 0}$. For a complete proof of the Hille-Yosida theorem, see the additional topics.

7.1.2 Contraction semigroups on Hilbert spaces

In this sub-section we specialize our study to a class of strongly continuous contraction semigroups of symmetric linear operators on a Hilbert space H.

A bounded linear operator T on a (real) Hilbert space H is called a symmetric operator if

$$\langle Tx, y\rangle = \langle x, Ty\rangle$$

where $\langle x, y\rangle$ is the inner product of H. The adjoint operator L^* of a densely defined linear operator L (with domain $D(L)$) on H is defined as the following: $x \in D(L^*)$ if

$$|\langle Ly, x\rangle| \leq C_x ||y|| \qquad \text{for every } y \in D(L)$$

for some non-negative constant C_x, and L^*x is the unique element in H (F. Riesz's representation) such that

$$\langle Ly, x\rangle = \langle y, L^*x\rangle \qquad \text{for every } y \in D(L)\ .$$

L^* is a closed linear operator on H. If $L^* = L$, then L is called a self-adjoint operator.

The fundamental tool in the study of self-adjoint operators is the spectral decomposition theorem. To appreciate this theorem, let us investigate an example which is in turn to be the general (function) model for self-adjoint operators.

Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space, and let $H = L^2(\Omega, \mathcal{F}, \mu)$. If ϕ is a real-valued measurable function on Ω, then we use T_ϕ to denote the multiplier operator

$$T_\phi x = \phi x \qquad \text{for any } x \in H$$

and

$$D(T_\phi) = \{x \in H : \phi x \in L^2(\Omega, \mathcal{F}, \mu)\}\ .$$

Then T_ϕ is a self-adjoint operator on H, with $\sigma(T_\phi)$ the essential range of ϕ. We note that if ϕ is an indicator function of a measurable subset A, then $T_{1_A} : L^2(\Omega, \mathcal{F}, \mu) \to L^2(A, \mathcal{F}, \mu)$ is a projection.

Given a real-valued measurable function ϕ we associate it with a right-continuous, increasing family $\{E_\lambda : \lambda \in \mathbb{R}\}$ of projections on H defined by $E_\lambda = T_{1_{\{\phi<\lambda\}}}$. Obviously $\lambda \to \langle E_\lambda x, x\rangle$ is increasing so that

$$\lambda \to \langle E_\lambda x, y\rangle$$

is right-continuous and has finite variation. Moreover $\lim_{\lambda\to-\infty} E_\lambda = 0$ and $\lim_{\lambda\to\infty} E_\lambda = I$. If $x \in H$ such that $||x|| = 1$, then

$$\mathbf{P}_x(\mathrm{d}w) \equiv x(w)^2\mu(\mathrm{d}w)$$

is a probability measure, and

$$\begin{aligned}\langle E_\lambda x, x\rangle &= \int_X 1_{\{\phi<\lambda\}} x^2 \mathrm{d}\mu \\ &= \mathbf{P}_x\left\{\phi(w) < \lambda\right\}\end{aligned}$$

is the distribution function of ϕ under the probability measure $\mathbf{P}_x$. It is seen that ϕ is square-integrable with respect to $\mathbf{P}_x$ if and only if

$$\begin{aligned}\mathbf{E}_x|\phi|^2 &= \int_X |\phi|^2 \mathrm{d}\mathbf{P}_x = \int_X |\phi|^2 x^2 \mathrm{d}\mu \\ &= \int_{-\infty}^{\infty} |\lambda|^2 \mathrm{d}\langle E_\lambda x, x\rangle < +\infty \ ;\end{aligned}$$

which implies that $x \in D(T_\phi)$ if and only if

$$\int_{-\infty}^{\infty} \lambda^2 \mathrm{d}\langle E_\lambda x, x\rangle < +\infty \ .$$

Moreover

$$\begin{aligned}\int_{-\infty}^{\infty} \lambda \mathrm{d}\langle E_\lambda x, x\rangle = \mathbf{E}_x\left(\phi\right) &= \int_X \phi \mathrm{d}\mathbf{P}_x \\ &= \int_X \phi x^2 \mathrm{d}\mu\end{aligned}$$

and therefore

$$\langle T_\phi x, x\rangle = \int_{-\infty}^{\infty} \lambda \mathrm{d}\langle E_\lambda x, x\rangle \ .$$

By the polarization identity, we thus have

$$\langle T_\phi x, y\rangle = \int_{-\infty}^{\infty} \lambda \mathrm{d}\langle E_\lambda x, y\rangle \ .$$

The last equality shows that T_ϕ has a spectral decomposition

$$T_\phi = \int_{-\infty}^{\infty} \lambda \mathrm{d}E_\lambda = \int_{\sigma(T_\phi)} \lambda \mathrm{d}E_\lambda$$

as $\lambda \to E_\lambda$ increases only on the spectrum $\sigma(T_\phi)$, and

$$D(T_\phi) = \left\{ x : \int_{-\infty}^{\infty} \lambda^2 \mathrm{d}\langle E_\lambda x, x\rangle < +\infty \right\} .$$

One of the main achievement in Functional Analysis is the following spectral theorem, which claims that the above spectral decomposition holds for any self-adjoint operator.

Theorem 7.1.6 *Let L be a self-adjoint operator on a Hilbert space H.*

1) The spectrum $\sigma(L) \subset \mathbb{R}$.

2) There is a right-continuous and increasing family $\{E_\lambda : \lambda \in \mathbb{R}\}$ of projections in H such that

$$\lim_{\lambda\to-\infty} E_\lambda = 0 \quad \textit{and} \quad \lim_{\lambda\to\infty} E_\lambda = I \ ;$$

$\lambda \to E_\lambda$ increases only on $\sigma(L)$,

$$D(L) = \left\{ x \in H : \int_{-\infty}^{\infty} \lambda^2 d\langle E_\lambda x, x\rangle < +\infty \right\}$$

and

$$L = \int_{-\infty}^{+\infty} \lambda dE_\lambda = \int_{\sigma(L)} \lambda dE_\lambda$$

in the sense that

$$\langle Lx, y\rangle = \int_{-\infty}^{+\infty} \lambda d\langle E_\lambda x, y\rangle \qquad \forall x \in D(L) \textit{ and } y \in H,$$

where the right-hand side is the Riemann-Stieltjes integral.

A self-adjoint linear operator L is positive-definite if $\langle Lx, x\rangle \geq 0$ for every $x \in H$. Such a self-adjoint operator possesses a spectral decomposition

$$L = \int_0^{+\infty} \lambda \mathrm{d}E_\lambda \ .$$

Consider a self-adjoint linear operator L which is negative-definite, that is, $-L$ is positive-definite, and let $-L$ have the spectral decomposition

$$-L = \int_0^{+\infty} \lambda \mathrm{d}E_\lambda$$

or equivalently

$$L = \int_0^{+\infty} -\lambda \mathrm{d}E_\lambda \ .$$

Define

$$P_t = \int_0^{+\infty} \mathrm{e}^{-\lambda t} \mathrm{d}E_\lambda$$

which is a self-adjoint operator, and

$$\begin{aligned} |\langle P_t x, x\rangle| &\leq \int_0^{+\infty} \mathrm{e}^{-\lambda t} \mathrm{d}\langle E_\lambda x, x\rangle \\ &\leq \int_0^{+\infty} \mathrm{d}\langle E_\lambda x, x\rangle \\ &= ||x||^2 \ . \end{aligned}$$

Therefore each P_t is a contraction on H, hence $(P_t)_{t\geq 0}$ is a strongly continuous contraction semigroup of symmetric operators on H with infinitesimal generator L. Conversely, it is obvious that the infinitesimal generator of a strongly continuous contraction semigroup of symmetric linear operators on a Hilbert space H is a negative-definite self-adjoint operator.

Theorem 7.1.7 *L is the infinitesimal generator of a strongly continuous contraction semigroup of symmetric linear operators on a Hilbert space H if and only if L is a negative-definite self-adjoint operator. If*

$$L = \int_0^{+\infty} -\lambda dE_\lambda \tag{7.4}$$

is the spectral decomposition of $-L$, then

$$e^{tL} = \int_0^{+\infty} e^{-\lambda t} dE_\lambda$$

and for every $\mu > 0$

$$R_\mu = (\mu - L)^{-1} = \int_0^{+\infty} \frac{1}{\mu + \lambda} dE_\lambda \ .$$

In general, if $-L$ is a positive-definite self-adjoint linear operator on H with spectral decomposition (7.4) then for any continuous function f on $[0, +\infty)$, $f(L)$ is a self-adjoint operator

$$f(L) = \int_0^{+\infty} f(-\lambda) \mathrm{d}E_\lambda$$

with domain

$$D(f(L)) = \left\{ x \in H : \int_0^{+\infty} f(-\lambda)^2 \mathrm{d}\langle E_\lambda x, x\rangle < +\infty \right\} \ .$$

The most important for our propose is the square root of $-L$ which can be defined as

$$\sqrt{-L} = \int_0^{+\infty} \sqrt{\lambda} \mathrm{d}E_\lambda$$

with domain

$$D(\sqrt{-L}) = \left\{ x \in H : \int_0^{+\infty} \lambda \mathrm{d}\langle E_\lambda x, x\rangle < +\infty \right\} .$$

$\sqrt{-L}$ is a positive-definite, self-adjoint operator. Obviously $D(L) \subset D(\sqrt{-L})$ and

$$\langle -Lx, y\rangle = \langle \sqrt{-L}x, \sqrt{-L}y\rangle \quad , \qquad \forall x \in D(L);\, y \in D(\sqrt{-L})$$

Let $(P_t)_{t\geq 0}$ be the semigroup generated by L:

$$P_t = \int_0^{+\infty} \mathrm{e}^{-\lambda t} \mathrm{d}E_\lambda$$

and set

$$\begin{aligned}
\mathcal{E}_t(x,x) &\equiv \frac{1}{2t}\left(||x||^2 - ||P_t x||^2\right) \\
&= \frac{1}{2t}\langle x - P_{2t}x, x\rangle \\
&= \frac{1}{2t}\int_0^{+\infty} \left(1 - \mathrm{e}^{-2\lambda t}\right) \mathrm{d}\langle E_\lambda x, x\rangle \; .
\end{aligned}$$

Since

$$\begin{aligned}
\frac{d}{ds}\frac{1-\mathrm{e}^{-s}}{s} &= \frac{\mathrm{e}^{-s}s - 1 + \mathrm{e}^{-s}}{s^2} \\
&= -\frac{\mathrm{e}^{s} - 1 - s}{s^2\mathrm{e}^{s}} < 0 \quad \text{for } s > 0,
\end{aligned}$$

$t \to \mathcal{E}_t(x,x)$ is decreasing and therefore

$$\lim_{t\downarrow 0} \mathcal{E}_t(x,x) = \sup_{t>0} \mathcal{E}_t(x,x)$$

exists, which will be denoted by $\mathcal{E}(x,x)$ $(\leq +\infty)$.

Theorem 7.1.8 *Let L be a negative-definite self-adjoint operator on Hilbert space H, and let $P_t = e^{tL}$ be the semigroup generated by L. Then $x \in D(\sqrt{-L})$ if and only if $\mathcal{E}(x,x) < +\infty$. Moreover*

$$\begin{aligned}
||\sqrt{-L}x||^2 &= \mathcal{E}(x,x) \\
&= \sup_{t>0} \frac{1}{2t}\left(||x||^2 - ||P_t x||^2\right)
\end{aligned}$$

for every $x \in D(\sqrt{-L})$.

Proof. If $x \in D(\sqrt{-L})$ then

$$\int_0^{+\infty} \lambda \mathrm{d}\langle E_\lambda x, x\rangle < +\infty \ .$$

However

$$\frac{1-\mathrm{e}^{-s}}{s} \leq 1 \quad \text{for all} \quad s \in (0, +\infty)$$

so that by Lebesgue's dominated convergence

$$\begin{aligned}
\lim_{t\downarrow 0} \frac{1}{2t}\left(||x||^2 - ||P_t x||^2\right) &= \int_0^{+\infty} \lim_{t\downarrow 0} \frac{1-\mathrm{e}^{-2\lambda t}}{2t} \mathrm{d}\langle E_\lambda x, x\rangle \\
&= \int_0^{+\infty} \lambda \mathrm{d}\langle E_\lambda x, x\rangle \\
&= ||\sqrt{-L}x||^2 \ .
\end{aligned}$$

On the other hand, if

$$\begin{aligned}
\lim_{t\downarrow 0} \frac{1}{2t}\left(||x||^2 - ||P_t x||^2\right) &= \sup_{t>0} \frac{1}{2t} \int_0^{+\infty} \left(1-\mathrm{e}^{-2\lambda t}\right) \mathrm{d}\langle E_\lambda x, x\rangle \\
&< +\infty,
\end{aligned}$$

then by Fatou's lemma and the fact that $\frac{1-\mathrm{e}^{-2\lambda t}}{2t} > 0$, we have

$$\begin{aligned}
\int_0^{+\infty} \lambda \mathrm{d}\langle E_\lambda x, x\rangle &= \int_0^{+\infty} \lim_{t\downarrow 0} \frac{1-\mathrm{e}^{-2\lambda t}}{2t} \mathrm{d}\langle E_\lambda x, x\rangle \\
&\leq \lim_{t\downarrow 0} \int_0^{+\infty} \frac{1-\mathrm{e}^{-2\lambda t}}{2t} \mathrm{d}\langle E_\lambda x, x\rangle < +\infty
\end{aligned}$$

that is $x \in D(\sqrt{-L})$. ■

Corollary 7.1.9 *Let L be a negative-definite self-adjoint operator L on Hilbert space H, and let $P_t = e^{tL}$. Then $\mathcal{E}$ is lower semi-continuous on H, that is*

$$\mathcal{E}(x,x) \leq \underline{\lim}_{n\to\infty} \mathcal{E}(x_n, x_n)$$

if $x_n \to x$ in H.

The lower semi-continuity follows from that $\mathcal{E}$ is the supremum of a family of continuous functions

$$\mathcal{E}(x,x) \equiv \sup_{t>0} \frac{1}{2t}\left(||x||^2 - ||P_t x||^2\right) \ .$$

Definition 7.1.10 *The quadratic form $(\mathcal{E}, D(\mathcal{E}))$ associated with a negative-definite self-adjoint operator L is defined by*

$$\mathcal{E}(x,y) = \langle \sqrt{-L}x, \sqrt{-L}y\rangle \ , \quad x, y \in D(\mathcal{E})$$

where

$$D(\mathcal{E}) = D(\sqrt{-L}) \ .$$

The main advantage by considering the quadratic form $(\mathcal{E}, D(\mathcal{E}))$ instead of the (unbounded) self-adjoint operator L is that, as we have seen, $\mathcal{E}(x,x)$ is well-defined for every $x \in H$

$$\mathcal{E}(x,x) = \lim_{t \downarrow 0} \frac{1}{2t} \left(||x||^2 - ||P_t x||^2 \right)$$

and $\mathcal{E}(x,x)$ is finite if and only if $x \in D(\mathcal{E})$.

Proposition 7.1.11 *Let L be a negative-definite self-adjoint operator on Hilbert space H, and let $(\mathcal{E}, D(\mathcal{E}))$ be the quadratic form associated with L. Define*

$$\begin{aligned} ||x||_1^2 &= ||x||^2 + \mathcal{E}(x,x) \\ &= ||x||^2 + ||\sqrt{-L}x||^2 \qquad \textit{for } x \in D(\mathcal{E}) \ . \end{aligned}$$

Then $(D(\mathcal{E}), ||\cdot||_1)$ is a Hilbert space.

7.2 Symmetric Markov semigroups

In this section we study transition semigroups on a measure space.

7.2.1 Invariant and symmetric measures

Let $(M, \mathcal{M}, m)$ be a σ-finite measure space and let $p \geq 1$. Then $L^p(M, \mathcal{M}, m)$ (or simply by L^p) denotes the Banach space of p-th integral functions

$$L^p(M, \mathcal{M}, m) = \left\{ f \in \mathcal{M} : \int_M |f(x)|^p m(\mathrm{d}x) \right\}$$

which is endowed with the L^p-norm

$$||f||_{L^p} = \left(\int_M |f(x)|^p m(\mathrm{d}x) \right)^{1/p} .$$

If $p = 2$ then $L^2(M, \mathcal{M}, m)$ is a Hilbert space with the scalar product

$$\langle f, g \rangle_{L^2(m)} = \int_M g(x) \left(P_t f \right)(x) m(\mathrm{d}x) \ .$$

If no confusion may arise, $\langle f, g \rangle_{L^2(m)}$ will be denoted simply by $\langle f, g \rangle$.

Let $\{P(t, x, \mathrm{d}y) : t > 0\}$ be a transition function on the measurable space $(M, \mathcal{M})$. The transition semigroup $(P_t)_{t \geq 0}$ is defined by

$$P_t f(x) = \int_M f(y) P(t, x, \mathrm{d}y) \ , \qquad \forall f \in b\mathcal{M} \cup \mathcal{M}^+ \ .$$

Definition 7.2.1 *Let μ be a σ-finite measure on $(M, \mathcal{M})$. μ is called a sub-invariant measure of $(P_t)_{t\geq 0}$ if $\mu P_t \leq \mu$ for every $t > 0$, i.e.*

$$\int_M P_t f(x)\mu(dx) \leq \int_M f(x)\mu(dx) \qquad \forall f \geq 0$$

If the equality holds, μ is called an invariant measure.

Lemma 7.2.2 *If μ is a sub-invariant measure of $\{P(t,x,dy) : t > 0\}$, then for each $t > 0$, P_t may be extended to a contraction on $L^p(M, \mathcal{M}, \mu)$ (where $p \geq 1$).*

Proof. For any $t > 0$ and $x \in M$, $P(t, x, M) \leq 1$, thus, by Hölder's inequality, for every $p \geq 1$

$$\begin{aligned} |P_t f(x)|^p &= \left| \int_M f(z) P(t, x, \mathrm{d}z) \right|^p \\ &\leq \int_M |f(z)|^p P(t, x, \mathrm{d}z) \\ &= P_t\left(|f|^p\right)(x) \ . \end{aligned}$$

Therefore for any $f \in L^p \cap b\mathcal{M}$

$$\begin{aligned} \int_M |P_t f|^p \mathrm{d}\mu &\leq \int_M P_t\left(|f|^p\right) \mathrm{d}\mu \\ &\leq \int_M |f|^p \mathrm{d}\mu \ . \end{aligned}$$

■

Definition 7.2.3 *Let $(M, \mathcal{M}, m)$ be a σ-finite measure space. Then $(P_t)_{t\geq 0}$ is called symmetric with respect to m if*

$$\langle P_t f, g\rangle_{L^2(m)} = \langle f, P_t g\rangle_{L^2(m)} \qquad \forall f, g \geq 0 \ .$$

In this case, we say that $(P_t)_{t\geq 0}$ is a symmetric Markov semigroup on $(M, \mathcal{M}, m)$.

If $(P_t)_{t\geq 0}$ is symmetric with respect to a σ-finite measure m, then $mP_t(f) = m(fP_t 1) \leq m(f)$ and therefore m is sub-invariant. In particular, if $P_t 1 = 1$, then m is an invariant measure of $(P_t)_{t\geq 0}$.

Corollary 7.2.4 *If $(P_t)_{t\geq 0}$ is symmetric with respect to a σ-finite measure m, then, for every $p \geq 1$, $(P_t)_{t\geq 0}$ can be extended to a contraction on L^p, and is symmetric on $L^2(M, \mathcal{M}\ , m)$.*

Unless otherwise specified, we will use the same notation P_t to denote the transition function $P(t, x, \mathrm{d}y)$, its associated contraction $P_t : b\mathcal{M} \to b\mathcal{M}$ (and $P_t : \mathcal{M}^+ \to \mathcal{M}^+$), its continuous extension $P_t : L^p \to L^p$ (for every $p \geq 1$), and $P_t : C_0(M) \to C_0(M)$ (if the semigroup $(P_t)_{t\geq 0}$ is Feller).

Theorem 7.2.5 *Let M be a locally compact separable metric space, and let $\{P_t : t > 0\}$ be a Feller semigroup which is symmetric with respect to a σ-finite measure m on $(M, \mathcal{B}(M))$. Then $(P_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on $L^p(M, \mathcal{B}(M), m)$ for every $p \geq 1$.*

Proof. Since $P_t f \to f$ as $t \to 0$ for every $f \in C_0(M)$, by Lebesgue's dominated convergence, for any $f \in C_0(M)$

$$||P_t f - f||_p \to 0 \qquad \text{as } t \downarrow 0 .$$

The conclusion follows from the fact that $C_0(M)$ is dense in L^p for any $p \geq 1$. ■

Actually the transition semigroup of a symmetric Borel right process has the same property.

Theorem 7.2.6 *The semigroup of a symmetric Borel right process may also be extended into a strongly continuous symmetric contraction semigroup on L^2.*

Proof. Let (P_t) be the semigroup of an m-symmetric Borel right process X and also denote the symmetric operator on L^2. This proof is due to Fitzsimmons and Getoor [8]. For any $\alpha > 0$ and bounded integrable function $f \geq 0$ on M, set $h = R_\alpha f$. Then h is α-excessive, and $h(X_t)$ is right continuous. By the dominated convergence theorem, as t goes to 0,

$$||P_t h - h||^2 = \int (E^x(h(X_t)) - h(x))^2 m(dx) \longrightarrow 0.$$

We need to verify the space $K = \{R_\alpha f : f \in b\mathcal{B} \cap L^1(M, m)\}$ is dense in L^2. This amounts to show that any square integrable g orthogonal to K vanishes a.e. Take a strictly positive bounded integrable function ϕ on E. Set $f = U^1\phi$. By the right continuity of X, f is also bounded, integrable and strictly positive. For any bounded continuous function w, since $wf(X.)$ is right continuous, $\alpha R_{\alpha+1}(wf) \longrightarrow wf$. Applying the dominated convergence theorem again, g is orthogonal to wf, or $(g, w)_{f\cdot m} = 0$. Since the set of bounded continuous functions is dense $L^2(M, f \cdot m)$, $g = 0$ a.e.-$f \cdot m$. It follows that $g = 0$ a.e.-m. ■

7.2.2 Dirichlet (energy) forms

Let $(M, \mathcal{M}, m)$ be a σ-finite measure space, and let $(P_t)_{t\geq 0}$ be a Markov semigroup symmetric with respect to m. Then for every $p \geq 1$, $(P_t)_{t\geq 0}$ is a semigroup of contractions on $L^p(M, \mathcal{M}, m)$. In this section, we study the case that $p = 2$, i.e. consider $(P_t)_{t\geq 0}$ as a semigroup of contractions on $L^2(M, \mathcal{M}, m)$.

Throughout this section, we assume that $(P_t)_{t\geq 0}$ is strongly-continuous

$$\lim_{t\downarrow 0} ||P_t u - u||_{L^2(m)} = 0 \quad \text{for every } u \in L^2(M, \mathcal{M}, m). \tag{7.5}$$

Let $(L, D(L))$ be the infinitesimal generator of $(P_t)_{t\geq 0}$ (as strongly continuous contraction semigroup on L^2), which is negative-definite, self-adjoint operator on $L^2(M, \mathcal{M}, m)$. Let

$$L = \int_0^{+\infty} (-\lambda) \mathrm{d}E_\lambda$$

be the spectral decomposition of L. The quadratic form $(\mathcal{E}, D(\mathcal{E}))$ associated with L is defined as

$$\mathcal{E}(u,u) = ||\sqrt{-L}u||^2_{L^2(m)} \ ; \qquad D(\mathcal{E}) = D(\sqrt{-L}) \ ,$$

by the polarization identity

$$\mathcal{E}(u,v) = \langle \sqrt{-L}u, \sqrt{-L}v \rangle_{L^2(m)} \qquad \forall u,v \in D(\mathcal{E}) \ .$$

The form $(\mathcal{E}, D(\mathcal{E}))$ is called the Dirichlet form associated with the symmetric Markov semigroup $(P_t)_{t\geq 0}$. As we have seen, $\mathcal{E}(u,u)$ can be naturally extended to every $u \in L^2(M, \mathcal{M}, m)$ by means of

$$\begin{aligned} \mathcal{E}(u,u) &= \lim_{t\downarrow 0} \frac{1}{t} \left(||u||^2_{L^2(m)} - ||P_{t/2}u||^2_{L^2(m)} \right) \\ &\equiv \lim_{t\downarrow 0} \mathcal{E}_t(u,u) \end{aligned}$$

so that $u \to \mathcal{E}(u,u)$ is lower semi-continuous and takes values in $[0, +\infty]$. On the other hand

$$\begin{aligned} \mathcal{E}_t(u,u) &= \frac{1}{t} \langle u, u - P_t u \rangle_{L^2(m)} \\ &= \frac{1}{t} \int_M u(x) \left(u(x) - (P_t u)(x) \right) m(\mathrm{d}x) \\ &= \frac{1}{t} \int_M u(x) \left(u(x) - \int_M u(y) P(t,x,\mathrm{d}y) \right) m(\mathrm{d}x) \\ &= \frac{1}{t} \int_M u(x) \left(\int_M (u(x) - u(y)) P(t,x,\mathrm{d}y) \right) m(\mathrm{d}x) \\ &\quad + \frac{1}{t} \int_M u^2 \left(1 - P_t 1 \right) \mathrm{d}m \\ &= \frac{1}{t} \int_{M\times M} u(x) \left(u(x) - u(y) \right) P(t,x,\mathrm{d}y) m(\mathrm{d}x) \\ &\quad + \frac{1}{t} \int_M u^2 \left(1 - P_t 1 \right) \mathrm{d}m \end{aligned}$$

and therefore we introduce a family of measures $\{\mu_t : t > 0\}$ on $M \times M$ by

$$m_t(\mathrm{d}x, \mathrm{d}y) = \frac{1}{2t} P(t,x,\mathrm{d}y) m(\mathrm{d}x)$$

and

$$k_t(\mathrm{d}x) = \frac{1}{t} \left(1 - (P_t 1)(x) \right) m(\mathrm{d}x) \ .$$

Lemma 7.2.7 *For every $t > 0$, $m_t(dx,dy)$ is a symmetric measure on $M \times M$:*

$$\int_{M\times M} h(x,y)m_t(dx,dy) = \int_{M\times M} h(y,x)m_t(dx,dy) \ . \tag{7.6}$$

Proof. (7.6) reduces to the symmetric assumption of (P_t) with respect to m for functions $u(x)v(y)$. ■

By the previous lemma

$$\begin{aligned}\mathcal{E}_t(u,u) &= 2\int_{M\times M} u(y)\,(u(y)-u(x))\,m_t(\mathrm{d}x,\mathrm{d}y) \\ &\quad + \int_M u(x)^2 k_t(\mathrm{d}x) \\ &= \int_{M\times M} u(y)\,(u(y)-u(x))\,m_t(\mathrm{d}x,\mathrm{d}y) \\ &\quad - \int_{M\times M} u(y)\,(u(x)-u(y))\,m_t(\mathrm{d}x,\mathrm{d}y) \\ &\quad + \int_M u(x)^2 k_t(\mathrm{d}x)\end{aligned}$$

and therefore

$$\begin{aligned}\mathcal{E}_t(u,u) &= \int_{M\times M} (u(x)-u(y))^2\, m_t(\mathrm{d}x,\mathrm{d}y) \\ &\quad + \int_M u(x)^2 k_t(\mathrm{d}x) \ .\end{aligned}$$

Hence we have

Theorem 7.2.8 *Let $(P_t)_{t\geq 0}$ be a symmetric Markov semigroup on $(M,\mathcal{M},m)$ satisfying (7.5). Then for every $u \in L^2(M,\mathcal{M},m)$*

$$\mathcal{E}(u,u) = \lim_{t\downarrow 0}\left(\int_{M\times M} (u(x)-u(y))^2\, m_t(dx,dy) + \int_M u(x)^2 k_t(dx)\right) \tag{7.7}$$

exists. $\mathcal{E}(u,u) < +\infty$ if and only if u belongs to the Dirichlet space $D(\mathcal{E})$.

Corollary 7.2.9 *Let $F : \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function*

$$|F(a)-F(b)| \leq C_F|a-b| \qquad \forall a,b \in \mathbb{R} \tag{7.8}$$

where $C_F \geq 0$ is a constant (called the Lipschitz constant) such that $F(0) = 0$, and let $u \in D(\mathcal{E})$. Then $F \circ u \in D(\mathcal{E})$ and

$$\mathcal{E}(F\circ u, F\circ u) \leq C_F^2 \mathcal{E}(u,u) \ . \tag{7.9}$$

In particular

$$\mathcal{E}(|u|,|u|) \le \mathcal{E}(u,u) \qquad\qquad \forall u \in L^2(M,\mathcal{M},m)\ .$$

Actually both limits

$$\lim_{t\downarrow 0}\int_{M\times M}(u(x)-u(y))^2\, m_t(\mathrm{d}x,\mathrm{d}y)$$

and

$$\lim_{t\downarrow 0}\int_M u(x)^2 k_t(\mathrm{d}x)$$

exist as long as $u \in D(\mathcal{E})$.

7.2.3 Dirichlet spaces

We have seen that, if L is the infinitesimal generator of a strongly continuous semigroup $(P_t)_{t\ge 0}$ of symmetric contractions on a Hilbert space H, then the domain $D(\sqrt{-L})$ can be described by means of the bi-linear form associated with the semigroup $(P_t)_{t\ge 0}$. If, in addition, $H = L^2(M,\mathcal{B},m)$ and P_t is a sub-Markov kernel, then $D(\sqrt{-L})$ possesses a few very nice properties.

Let M be a locally compact metric space or a Polish space, and let m be a positive, σ-finite Borel measure on $(M,\mathcal{B})$ with support M. Consider the real Hilbert space $L^2(M,\mathcal{B},m)$ which is endowed with the scalar product

$$\langle u,v\rangle = \int_M u(x)v(x)m(\mathrm{d}x)\ .$$

By a symmetric form $\mathcal{E}$ on $L^2(M,\mathcal{B},m)$ with domain $D(\mathcal{E})$ we mean a (real) bi-linear form

$$\mathcal{E} : D(\mathcal{E}) \times D(\mathcal{E}) \to \mathbb{R}$$

where $D(\mathcal{E})$ is a linear subspace of $L^2(M,\mathcal{B},m)$, which is symmetric in the sense that

$$\mathcal{E}(u,v) = \mathcal{E}(v,u) \qquad \forall u,v \in D(\mathcal{E})\ .$$

In this section we are only interested in non-negative symmetric form: $\mathcal{E}(u,u) \ge 0$ for every $u \in D(\mathcal{E})$. For such symmetric form, we use $\mathcal{E}(u)$ to denote $\mathcal{E}(u,u)$ for simplicity, and we introduce a new scalar product on $D(\mathcal{E})$

$$\langle u,v\rangle_{\mathcal{E}_\alpha} = \alpha\langle u,v\rangle + \mathcal{E}(u,v) \qquad \forall u,v \in D(\mathcal{E})$$

where $\alpha > 0$ is a constant. The corresponding norm is denoted by $||\cdot||_{\mathcal{E}_\alpha}$, that is

$$||u||_{\mathcal{E}_\alpha} = \sqrt{\alpha||u||^2_{L^2(m)} + \mathcal{E}(u)}\ .$$

Definition 7.2.10 *A non-negative symmetric form $(\mathcal{E}, D(\mathcal{E}))$ on $L^2(M,\mathcal{B},m)$ is* closable, *if $\{u_n\}$ is a Cauchy sequence in the norm $||\cdot||_{\mathcal{E}_1}$ such that $u_n \to 0$ in $L^2(M,\mathcal{B},m)$, then $u_n \to 0$ in $||\cdot||_{\mathcal{E}_1}$. In other words, the completion of $D(\mathcal{E})$ under the norm $||\cdot||_{\mathcal{E}_1}$ is a subspace of $L^2(M,\mathcal{B},m)$. $(\mathcal{E}, D(\mathcal{E}))$ is closed if $(D(\mathcal{E}), ||\cdot||_{\mathcal{E}_1})$ (hence $(D(\mathcal{E}), ||\cdot||_{\mathcal{E}\alpha})$ for any $\alpha > 0$) is a Hilbert space.*

The closability condition means the consistence between the symmetric form $\mathcal{E}$ and the Hilbert inner product $\langle\cdot,\cdot\rangle$.

Theorem 7.2.11 *Let $(\mathcal{E}, D(\mathcal{E}))$ is a non-negative, closed and symmetric form on $L^2(M,\mathcal{B},m)$. Then the following three conditions are equivalent.*

1) (contraction property) *For any $\varepsilon > 0$, there is a real function $\phi_\varepsilon : \mathbb{R} \to [-\varepsilon, 1+\varepsilon]$, $\phi_\varepsilon(t) = t$ for any $t \in [0,1]$ and*

$$0 \le \phi_\varepsilon(t) - \phi_\varepsilon(s) \le t - s \quad \text{for any } t < s$$

such that $\phi_\varepsilon(u) \in D(\mathcal{E})$ whenever $u \in D(\mathcal{E})$ and $\mathcal{E}(\phi_\varepsilon(u)) \le \mathcal{E}(u)$.

2) (unit contraction) *If $u \in D(\mathcal{E})$ and $v = (0 \vee u) \wedge 1$, then $v \in D(\mathcal{E})$ and $\mathcal{E}(v) \le \mathcal{E}(u)$.*

3) (normal contraction) *If $u \in D(\mathcal{E})$ and $v \in L^2(M,\mathcal{B},m)$ such that*

$$|v(x) - v(y)| \le |u(x) - u(y)|$$

and $|v(x)| \le |u(x)|$ for all $x, y \in M$, then $v \in D(\mathcal{E})$ and $\mathcal{E}(v) \le \mathcal{E}(u)$.

A symmetric form $(\mathcal{E}, D(\mathcal{E}))$ satisfying one of the equivalent conditions 1-3 is said to have the Markov property*.*

Proof. It is obvious that 3 implies 2 and 2 implies 1. We prove that 1 $\Longrightarrow$ 2. For each natural number n, we choose $\phi_{1/n}$ so that 1 is satisfied. If $u \in D(\mathcal{E})$, then $\mathcal{E}(\phi_{1/n}(u)) \le \mathcal{E}(u)$ for every n. By definition, $\phi_{1/n}(u) \to v \equiv \min\{1, \max(0,u)\}$ as $n \to \infty$ and therefore, by Lebesgue's dominated convergence theorem $\phi_{1/n}(u) \to v$ in $L^2(M,\mathcal{B},m)$. On the other hand $(D(\mathcal{E}), ||\cdot||_{\mathcal{E}^1})$ is a Hilbert space, thus, there is a sub-sequence of $\phi_{1/n}(u)$, its Cesaro mean convergent in $(D(\mathcal{E}), ||\cdot||_{\mathcal{E}^1})$. We therefore, without losing generality,

$$\frac{1}{n}\sum_{k=1}^{n}\phi_{1/k}(u)$$

converges in $||\cdot||_{\mathcal{E}^1}$. Therefore

$$\begin{aligned}
\sqrt{\mathcal{E}(v,v)} &= \sqrt{\mathcal{E}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\phi_{1/k}(u)\right)} \\
&= \lim_{n\to\infty}\frac{1}{n}\sqrt{\mathcal{E}\left(\sum_{k=1}^{n}\phi_{1/k}(u)\right)} \\
&\le \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sqrt{\mathcal{E}\left(\phi_{1/k}(u)\right)} \\
&\le \mathcal{E}(u)\ .
\end{aligned}$$

For the proof of 2 $\Longrightarrow$ 3, refer to [9]. ∎

A non-negative, closed and symmetric form $(\mathcal{E}, D(\mathcal{E}))$ on $L^2(M,\mathcal{B},m)$ which has the Markov property is called a Dirichlet form.

Definition 7.2.12 *A bounded linear operator P on $L^2(M,\mathcal{B},m)$ is Markovian if $0 \le Pu \le 1$ almost surely with respect to m whenever $u \in L^2(M,\mathcal{B},m)$ and $0 \le u \le 1$ m-a.e.*

A (unbounded) self-adjoint operator A with domain $D(A)$ on $L^2(M,\mathcal{B},m)$ is called a Dirichlet operator if

$$\langle Au,(u-1)^+\rangle \le 0 \qquad \forall u \in D(A) \ .$$

The following theorem is anticipated. For a proof, see [9].

Theorem 7.2.13 *Let $(T_t)_{t\ge 0}$ be a strongly continuous semigroup of symmetric contractions on $L^2(M,\mathcal{B},m)$, with the quadratic form $(\mathcal{E},D(\mathcal{E}))$, the resolvent (G_α), and the infinitesimal generator $(A,D(A))$. Then the following four conditions are equivalent to each other.*

1) For each $t \ge 0$, T_t is Markovian.

2) For each $\alpha > 0$, αG_α is Markovian.

3) $(\mathcal{E},D(\mathcal{E}))$ possesses the Markov property.

4) A is a Dirichlet operator.

In particular, if $(\mathcal{E},D(\mathcal{E}))$ is the Dirichlet form associated with a symmetric Markov semigroup $(P_t)_{t\ge 0}$, then all conclusions in the above theorem apply to $(P_t)_{t\ge 0}$.

A Dirichlet form shares the familiar properties as those of Dirichlet integrals. More precisely, if $(\mathcal{E},D(\mathcal{E}))$ a Dirichlet form on $L^2(M,\mathcal{B},m)$, then

1. $D(\mathcal{E})$ is closed under the lattice operations. More precisely, if $u,v \in D(\mathcal{E})$, so are $u \vee v$, $u \wedge v$, and $u \wedge 1$.

2. $D(\mathcal{E}) \cap L^\infty(M,B,m)$ is an algebra, that is, if $u,v \in D(\mathcal{E}) \cap L^\infty(M,B,m)$, then $uv \in D(\mathcal{E})$ and

$$\sqrt{\mathcal{E}(uv)} \le ||u||_\infty \sqrt{\mathcal{E}(v)} + ||v||_\infty \sqrt{\mathcal{E}(u)} \ .$$

3. If $u \in D(\mathcal{E})$ then $u_n \to u$ in $||\cdot||_{\mathcal{E}_1}$ as $n \to \infty$, where $u_n = ((-n)\vee u)\wedge n$. Similarly, $u^{(\varepsilon)} \to u$ in $||\cdot||_{\mathcal{E}_1}$ as $\varepsilon \downarrow 0$, where $u^{(\varepsilon)} = u - ((-\varepsilon)\vee u)\wedge \varepsilon$.

4. If $u_n \to u$ in $||\cdot||_{\mathcal{E}_1}$ (where u_n, $u \in D(\mathcal{E})$) then

$$\phi(u_n) \to \phi(u) \quad \text{weakly in } (D(\mathcal{E}), ||\cdot||_{\mathcal{E}_1})$$

as $n \to \infty$, for any Lipschitz function ϕ with $\phi(0) = 0$.

7.3 Symmetric Markov processes

In this section we examine a few special features about Markov processes with a symmetric transition semigroup. For simplicity we only consider conservative symmetric Markov processes.

Let M be a locally compact separable metric space or a Polish space. We assume that $(P_t)_{t\geq 0}$ is the transition function of a Borel right process with the canonical realization

$$\mathbf{X} = (\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, P^x)$$

on the space Ω of all right-continuous paths with left-limits in M, where $(X_t)_{t\geq 0}$ is the coordinate process on Ω. We further assume that X is conservative, i.e., $P^x(X_t \in M) = 1$ for any $t \geq 0$ and $x \in M$.

If μ is a σ-finite measure then

$$P^\mu(Y) = \int_M P^x(Y)\mu(\mathrm{d}x) \qquad \forall Y \in \mathcal{F} \ ;$$

if Y is an $\mathcal{F}$-measurable function, then $E^\mu(Y)$ denotes the expectation of Y with respect to measure P^μ.

Suppose $(P_t)_{t\geq 0}$ is symmetric with respect to a σ-finite measure m on $(M, \mathcal{B})$:

$$\int_M f(P_t h)\mathrm{d}m = \int_M h(P_t f)\mathrm{d}m$$

which will be written as $m(f(P_t h)) = m(h(P_t f))$ for each $t > 0$. We have shown that m is invariant under P_t so that for every $t > 0$ and $\alpha > 0$ we have

$$mP_t = m \quad \text{and} \quad mR_\alpha = \frac{1}{\alpha} m \ .$$

The goal of this section is to explore the connections between the energy form or the Dirichlet form associated with the symmetric semigroup $(P_t)_{t\geq 0}$ and the Markov process $\mathbf{X}$.

Let $(L, D(L))$ be the infinitesimal generator of $(P_t)_{t\geq 0}$ (as the strongly continuous contraction semigroup on $L^2(M, m)$), and let $(\mathcal{E}, D(\mathcal{E}))$ be the associated Dirichlet form.

Let $f \in L^2(M, m)$ which is Borel measurable, and let $u = R_\alpha f$ for $\alpha > 0$. Then $u \in D(L)$ and $Lu = \alpha u - f$.

By the Markov property, for any $t > s$

$$\begin{aligned} E^x\left(f(X_t)h(X_s)\right) &= E^x\left(E^x\left(f(X_t)h(X_s)|\mathcal{F}_s\right)\right) \\ &= E^x\left(h(X_s)(P_{t-s}f)(X_s)\right) \\ &= (P_s(h(P_{t-s}f)))(x) \end{aligned}$$

which is indeed a special case of finite dimensional distributions. Therefore

Lemma 7.3.1 *For any Borel measurable functions h, f and $t > s$ we have*

$$E^\mu\left(f(X_t)h(X_s)\right) = \mu(P_s(h(P_{t-s}f))) \ . \tag{7.10}$$

In particular, if μ is an invariant measure of $(P_t)_{t\geq 0}$ then for $t > s$

$$E^\mu\left(f(X_t)h(X_s)\right) = \mu(h(P_{t-s}f)) \ . \tag{7.11}$$

7.3.1 Energy integrals

If $u = R_\alpha f$ for some bounded Borel measurable function f on M and $\alpha > 0$, then

$$M_t^u = u(X_t) - u(X_0) - \int_0^t (Lu)(X_s)\mathrm{d}s \tag{7.12}$$

is a bounded martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^x)$ for any $x \in M$, where $Lu = \alpha u - f$ and L is the infinitesimal generator of $(P_t)_{t\geq 0}$. Since $(X_t)_{t\geq 0}$ is quasi-left continuous, so is $(M_t^u)_{t\geq 0}$ and therefore $(\langle M^u\rangle_t)_{t\geq 0}$ must be continuous, where $(\langle M^u\rangle_t)_{t\geq 0}$ is the unique increasing, adapted process with initial zero such that

$$(M_t^u)^2 - \langle M^u\rangle_t$$

is a martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^x)$. Let $\mathcal{R}$ denote the range of of R_α $(\alpha > 0)$:

$$\mathcal{R} = \{R_\alpha f : \alpha > 0 \text{ and } f \in b\mathcal{B}\}$$

which is independent of α.

Lemma 7.3.2 *If $u \in \mathcal{R}$ and $u = R_\alpha f$ for some $\alpha > 0$ and $f \in L^1(M, m) \cap b\mathcal{B}$, $(M_t^u)_{t\geq 0}$ is a martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^m)$, in the sense that for any $t > s$ and any $B \in \mathcal{F}_s$*

$$E^m(M_t^u 1_B) = E^m(M_s^u 1_B) \qquad \forall t > s\ ,\ B \in \mathcal{F}_s,$$

namely

$$E^m\left(M_t^u | \mathcal{F}_s\right) = M_s^u \qquad P^m\text{-a.s.}$$

for any $t > s$.

Proof. The potential $u = R_\alpha f$ is bounded and belongs to $L^1(M, m)$ as well, and thus

$$M_t^u = u(X_t) - u(X_0) - \int_0^t (\alpha u - f)(X_s)\mathrm{d}s$$

is bounded martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^x)$, and

$$\begin{aligned} E^m|M_t^u| &\leq E^m|u(X_t)| + E^m|u(X_0)| + \int_0^t E^m\left\{(\alpha|u| + |f|)(X_s)\right\}\mathrm{d}s \\ &\leq (2 + \alpha t)\, m(|u|) + tm(|f|)\ . \end{aligned}$$

■

Lemma 7.3.3 *If both u and u^2 belong to $\mathcal{R}$, then*

$$\langle M^u\rangle_t = \int_0^t \frac{1}{2}(L(u^2) - 2u(Lu))(X_s)ds\ .$$

Define for such u

$$\Gamma(u) = \frac{1}{2}(L(u^2) - 2u(Lu))\ .$$

Γ is called the square field operator on $\mathcal{R}$.

Proof. Since both u and u^2 belong to $\mathcal{R}$, we have

$$M_t^u = u(X_t) - u(X_0) - \int_0^t (Lu)(X_s)\mathrm{d}s$$

and

$$M_t^{u^2} = u^2(X_t) - u^2(X_0) - \int_0^t (Lu^2)(X_s)\mathrm{d}s \tag{7.13}$$

are martingales on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^x)$. On the other hand, by integration by parts formula

$$\begin{aligned} u^2(X_t) - u^2(X_0) &= 2\int_0^t u(X_s)\mathrm{d}M_s^u + 2\int_0^t u(X_s)(Lu)(X_s)\mathrm{d}s \\ &\quad +2\langle M^u\rangle_t \end{aligned} \tag{7.14}$$

comparing the martingale parts and the variational process parts of equation we thus must have

$$2\langle M^u\rangle_t + 2\int_0^t u(X_s)(Lu)(X_s)\mathrm{d}s = \int_0^t (Lu^2)(X_s)\mathrm{d}s$$

which implies

$$\langle M^u\rangle_t = \frac{1}{2}\int_0^t (Lu^2)(X_s)\mathrm{d}s - \int_0^t u(X_s)(Lu)(X_s)\mathrm{d}s \ .$$

■

We next would like to calculate $E^m\langle M^u\rangle_t$ for $u \in \mathcal{R}$, which is called the energy integral. To this end we need the following

Lemma 7.3.4 *If $u \in D(L)$, then*

$$\begin{aligned} \frac{d}{dt}\langle u, P_t u\rangle_{L^2(m)} &= \langle Lu, P_t u\rangle_{L^2(m)} \\ &= \langle u, LP_t u\rangle_{L^2(m)} \ . \end{aligned}$$

This follows from the definition of infinitesimal generators.

Theorem 7.3.5 *Let $u \in \mathcal{R}$ (i.e. $u = R_\alpha f$ for some $\alpha > 0$ and $f \in b\mathcal{B}$) and let*

$$A_t^u = \int_0^t (Lu)(X_s)ds = \int_0^t (\alpha u - f)(X_s)ds \qquad \forall t \geq 0 \ .$$

1) For $u \in \mathcal{R}$ we have

$$E^m\langle M^u\rangle_t = E^m(u(X_t) - u(X_0))^2 + E^m\left(A_t^u\right)^2 \ . \tag{7.15}$$

2) For any $u \in L^2(M, m)$ we have

$$E^m(u(X_t) - u(X_0))^2 = 2\langle u, u - P_t u\rangle_{L^2(m)} \ . \tag{7.16}$$

3) If $u \in \mathcal{R}$ then $u \in D(L)$,

$$E^m \langle M^u \rangle_t = 2t\mathcal{E}(u,u) \qquad \forall t \geq 0 \tag{7.17}$$

and

$$E^m (A_t^u)^2 = 2t\mathcal{E}(u,u) - 2\langle u, u - P_t u\rangle$$

Proof. Since $(M_t^u)_{t\geq 0}$ is a martingale under P^x for every x, $E^m\langle M^u\rangle_t = E^m\,(M_t^u)^2$. Squaring both sides of equation (7.12) and taking expectations with respect to P^m to obtain

$$\begin{aligned} E^m\,(M_t^u)^2 &= E^m(u(X_t)-u(X_0))^2 + E^m\left(\int_0^t (Lu)(X_s)\mathrm{d}s\right)^2 \\ &\quad -2E^m\left\{(u(X_t)-u(X_0))\int_0^t (Lu)(X_s)\mathrm{d}s\right\} \\ &= I_1 + I_2 + I_3\ . \end{aligned} \tag{7.18}$$

The first integral may be evaluated easily, indeed

$$\begin{aligned} I_1 &= E^m\left(u^2(X_t)\right) - 2E^m\left(u(X_t)u(X_0)\right) + E^m\left(u(X_0)^2\right) \\ &= ||u||^2_{L^2(m)} - 2m(u(P_tu)) + ||u||^2_{L^2(m)} \\ &= 2\langle u, u-P_tu\rangle_{L^2(m)} - \langle u^2, 1-P_t1\rangle_{L^2(m)} \\ &= 2\langle u, u-P_tu\rangle_{L^2(m)}\ . \end{aligned}$$

The computation of I_2 needs a little bit effort. First write I_2 as a triple integral

$$\begin{aligned} I_2 &= E^m\int_0^t\int_0^t (Lu)(X_s)(Lu)(X_r)\mathrm{d}s\mathrm{d}r \\ &= 2\int_0^t\left(\int_0^r E^m\left((Lu)(X_s)(Lu)(X_r)\right)\mathrm{d}s\right)\mathrm{d}r \\ &= 2\int_0^t\left(\int_0^r m\left(P_s\left((Lu)P_{r-s}(Lu)\right)\right)\mathrm{d}s\right)\mathrm{d}r \\ &= 2\int_0^t\left(\int_0^r m\left((Lu)(LP_{r-s}u)\right)\mathrm{d}s\right)\mathrm{d}r \\ &= 2\int_0^t\left(\int_0^r m\left((Lu)(LP_su)\right)\mathrm{d}s\right)\mathrm{d}r \\ &= 2\int_0^t\left(\int_0^r \frac{\mathrm{d}}{\mathrm{d}s}m\left((Lu)(P_su)\right)\mathrm{d}s\right)\mathrm{d}r \\ &= 2\int_0^t\left(m\left((Lu)(P_ru)\right) - m\left(u(Lu)\right)\right)\mathrm{d}r \\ &= 2\int_0^t \frac{\mathrm{d}}{\mathrm{d}r}m\left(u(P_ru)\right)\mathrm{d}r - 2t\langle u, Lu\rangle_{L^2(m)} \\ &= 2t\mathcal{E}(u,u) - 2\langle u, u-P_tu\rangle \end{aligned}$$

and finally we may treat the last integral I_3 similarly. In fact

$$\begin{aligned}
I_3 &= -2\int_0^t E^m\left(u(X_t)(Lu)(X_s)\right)\mathrm{d}s + 2\int_0^t E^m\left(u(X_0)(Lu)(X_s)\right)\mathrm{d}s \\
&= -2\int_0^t m((Lu)P_{t-s}u)\mathrm{d}s + 2\int_0^t m(u(P_sLu))\mathrm{d}s \\
&= 0 \ .
\end{aligned}$$

■

Recall that if $u = R_\alpha f$, then we have a family (indexed by $\alpha > 0$) of martingales

$$M_t^{\alpha,f} = \int_0^t e^{-\alpha s} f(X_s)\mathrm{d}s + e^{-\alpha t}u(X_t).$$

On the other hand, by means of integration by parts,

$$\begin{aligned}
e^{-\alpha t}u(X_t) - u(X_0) &= \int_0^t e^{-\alpha s}\mathrm{d}M_s^u \\
&\quad + \int_0^t e^{-\alpha s}(Lu)(X_s)\mathrm{d}s - \alpha\int_0^t e^{-\alpha s}u(X_s)ds
\end{aligned}$$

and hence we have

$$M_t^{\alpha,f} = \int_0^t e^{-\alpha s}\mathrm{d}M_s^u + u(X_0) \ .$$

Therefore

$$\begin{aligned}
E^m\langle M^{\alpha,f}\rangle_t &= E^m\left|\int_0^t e^{-\alpha s}\mathrm{d}M_s^u\right|^2 \\
&= E^m\left(\int_0^t e^{-2\alpha s}\mathrm{d}\langle M^u\rangle_s\right) \ .
\end{aligned}$$

According to integration by parts

$$\int_0^t e^{-2\alpha s}\mathrm{d}\langle M^u\rangle_s = e^{-2\alpha t}\langle M^u\rangle_t + 2\alpha\int_0^t e^{-2\alpha s}\langle M^u\rangle_s \mathrm{d}s$$

so that by (7.17)

$$\begin{aligned}
E^m\langle M^{\alpha,f}\rangle_t &= e^{-2\alpha t}E^m\langle M^u\rangle_t + 2\alpha\int_0^t e^{-2\alpha s}E^m\langle M_s^u\rangle \mathrm{d}s \\
&= \frac{1}{\alpha}\mathcal{E}(u,u) - \frac{1}{\alpha}e^{-2\alpha t}\mathcal{E}(u,u) \ .
\end{aligned}$$

Corollary 7.3.6 *If $u = R_\alpha f$ where $\alpha > 0$ and $f \in b\mathcal{B}$, then*

$$E^m\left(\langle M^{\alpha,f}\rangle_t\right) = \frac{1}{\alpha}\mathcal{E}(u,u) - \frac{1}{\alpha}e^{-2\alpha t}\mathcal{E}(u,u) \tag{7.19}$$

and

$$E^m \left| M_t^{\alpha,f} - M_0^{\alpha,f} \right|^2 = \frac{1-e^{-2\alpha t}}{\alpha} \mathcal{E}(u,u) \ ,$$

where

$$M_t^{\alpha,f} = \int_0^t e^{-\alpha s} f(X_s) ds + e^{-\alpha t} u(X_t) \ .$$

Lemma 7.3.7 *Let $\mathbb{M}^2$ denote the space of all square-integrable, right continuous martingales on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^m)$ endowed with norm*

$$||Z||_{\mathbb{M}^2} = \sum_{n=1}^{\infty} \frac{1}{2^n} \sqrt{1 \wedge E^m \left(|Z_n|^2 \right)} \ .$$

Then $(\mathbb{M}^2, || \cdot ||_{\mathbb{M}^2})$ is a complete metric space.

7.3.2 Lyons-Zheng's decomposition

We use the same notations established in the last sub-section. In particular the Markov process we are concerned is conservative. Given $T > 0$ a fixed constant time, let $\gamma_T : \Omega \to \Omega$ denote the time-inverse at time T:

$$(\omega \circ \gamma_T)(t) = \omega(T-t) \qquad \forall t \le T, \, \omega \in \Omega \ .$$

Then γ is a measurable transformation of $(\Omega, \mathcal{F}_T)$. The following lemma says that P^m on $\mathcal{F}_T$ is invariant under γ_T.

Lemma 7.3.8 *For any $Y \in L^1(\Omega, P^m)$ we have*

$$E^m (Y \circ \gamma_T) = E^m (Y) \ . \tag{7.20}$$

Next we suppose $u \in D(L)$ then

$$M_t^u = u(X_t) - u(X_0) - \int_0^t (Lu)(X_s) \mathrm{d}s$$

is a square-integrable (right-continuous) martingale under P^m. In particular

$$E^m \left(1_A M_t^u \right) = E^m \left(1_A M_s^u \right) \qquad \forall s < t \le T$$

for any $A \in \mathcal{F}_s$. By Lemma 7.3.8, it implies that

$$E^m \left((1_A \circ \gamma_T) M_t^u \circ \gamma_T \right) = E^m \left((1_A \circ \gamma_T) M_s^u \circ \gamma_T \right) \tag{7.21}$$

for any $A \in \mathcal{F}_s$ and $s < t \le T$. While

$$M_t^u \circ \gamma_T = u(X_{T-t}) - u(X_T) - \int_0^t (Lu)(X_{T-s}) \mathrm{d}s$$

and we thus define

$$\begin{aligned} N_t^u &= M_t^u \circ \gamma_T \\ &= u(X_{T-t}) - u(X_T) - \int_0^t (Lu)(X_{T-s})\mathrm{d}s \\ &= u(X_{T-t}) - u(X_T) - \int_{T-t}^T (Lu)(X_s)\mathrm{d}s \ . \end{aligned}$$

Let $\mathcal{G}_t = \sigma\{X_s:T-t \le s \le T\}$. Then $(N_t^u)_{t\le T}$ is $(\mathcal{G}_t)$-adapted, and equation (7.21) may be written as

$$E^m((1_A \circ \gamma_T)N_t) = E^m((1_A \circ \gamma_T)N_s)$$

for any $s \le t \le T$ and $A \in \mathcal{F}_s$. However, it is clear by definition that

$$\mathcal{G}_s = \left\{1_A \circ \gamma_T : A \in \mathcal{F}_{T-s}^0\right\}$$

and therefore

$$E^m(1_A N_s) = E^m(1_A N_t) \quad \forall s \le t \le T \text{ and } A \in \mathcal{G}_s \ .$$

That is, $(N_t^u)_{t\le T}$ is a martingale on $(\Omega, \mathcal{G}, \mathcal{G}_s, P^m)$, where $\mathcal{G} = \mathcal{F}_T^0$.

Lemma 7.3.9 *Let $u \in D(L)$. For $t \le T$ let $N_t^u = M_t^u \circ \gamma_T$. Then*

1) (N_t^u) is a martingale on $(\Omega, \mathcal{G}, \mathcal{G}_s, P^m)$ up to time T, and

$$N_T^u - N_{T-t}^u = u(X_0) - u(X_t) - \int_0^t (Lu)(X_s)ds \qquad \forall t \le T \ .$$

2) For any $t \le T$ the energy integral

$$E^m\langle N^u\rangle_t = 2t\mathcal{E}(u,u) \ .$$

3) If in addition $u^2 \in D(L)$ then

$$\langle N^u\rangle_t = \int_0^t \Gamma(u)(X_{T-s})ds = \int_{T-t}^T \Gamma(u)(X_s)ds \qquad P^m\text{-a.s.}$$

where $\Gamma(u) = \frac{1}{2}L(u^2) - u(Lu)$.

Proof. While

$$N_T^u = u(X_0) - u(X_T) - \int_0^T (Lu)(X_s)\mathrm{d}s$$

and

$$N_{T-t}^u = u(X_t) - u(X_T) - \int_t^T (Lu)(X_s)\mathrm{d}s$$

claim 1) follows immediately. Finally by Lemma 7.3.8

$$\begin{aligned} E^m\langle N^u\rangle_t &= E^m(N_t^u)^2 = E^m(N_t^u\circ\gamma_T)^2 \\ &= E^m\,(M_t^u)^2 \\ &= 2t\mathcal{E}(u,u) \end{aligned}$$

■

The following theorem follows directly from 2).

Theorem 7.3.10 (Lyons,Zheng) *Let $u\in D(L)$. For $t\le T$ let $N_t^u = M_t^u\circ\gamma_T$. We have the decomposition*

$$u(X_t)-u(X_0)=\frac{1}{2}M_t^u+\frac{1}{2}(N_{T-t}^u-N_T^u)\qquad P^m\text{-}a.s.$$

7.3.3 Fukushima's decomposition

It is shown that for any Borel measurable function $u\in D(\mathcal{E})$ and for any $T>0$, we have the Lyons-Zheng decomposition

$$u(X_t)-u(X_0)=\frac{1}{2}M_t^u+\frac{1}{2}\left(N_{T-t}^u-N_T^u\right)\quad P^m\text{-a.s.}$$

The above decomposition can be effectively to derive a maximal inequality for $u(X_t)$, from which we can see that the Lyons-Zheng's decomposition holds for $u\in D(\mathcal{E})$.

Lemma 7.3.11 *For any $u\in D(\mathcal{E})$ and $T>0$ we have*

$$P^m\left\{\sup_{t\le T}|u(X_t)-u(X_0)|\ge\lambda\right\}\le\frac{10}{\lambda^2}T\mathcal{E}\left(u,u\right)$$

for any $\lambda>0$.

Proof. According to Lyons-Zheng's decomposition, it holds that

$$\sup_{t\le T}|u(X_t)-u(X_0)|\le\frac{1}{2}\sup_{t\le T}|M_t^u|+\sup_{t\le T}|N_t^u|$$

and we then have

$$\begin{aligned} & P^m\left\{\sup_{t\le T}|u(X_t)-u(X_0)|\ge\lambda\right\} \\ \le\; & P^m\left\{\sup_{t\le T}|M_t^u|\ge\lambda\;\text{ or }\;\sup_{t\le T}|N_t^u|\ge\frac{\lambda}{2}\right\} \\ \le\; & P^m\left\{\sup_{t\le T}|M_t^u|\ge\lambda\right\}+P^m\left\{\sup_{t\le T}|N_t^u|\ge\frac{\lambda}{2}\right\} \\ \le\; & \frac{1}{\lambda^2}E^m(M_T^u)^2+\frac{4}{\lambda^2}E^m(N_T^u)^2 \\ =\; & \frac{10}{\lambda^2}T\mathcal{E}\left(u,u\right)\;. \end{aligned}$$

■

Theorem 7.3.12 (Fukushima) *For any $u \in D(\mathcal{E})$ (we choose a Borel measurable version) we have the following decomposition*

$$u(X_t) - u(X_0) = M_t^u + A_t^u \qquad P^m\text{-a.s.}$$

for any $t \geq 0$, where M^u is square-integrable (right continuous) martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^m)$, and $(A_t^u)_{t\geq 0}$ is continuous, $(\mathcal{F}_t)$-adapted process with zero energy in the sense that

$$\lim_{t\downarrow 0} \frac{1}{t} E^m (A_t^u)^2 = 0 \ .$$

Moreover for every $t > 0$

$$E^m (A_t^u)^2 = 2t\mathcal{E}(u,u) - 2\langle u, u - P_t u\rangle \ .$$

Proof. All conclusions hold for $u \in D(L)$. For a general $u \in D(\mathcal{E})$ we may choose a sequence of $u_n \in D(L)$ such that

$$\mathcal{E}(u_n - u, u_n - u) + ||u_n - u||^2_{L^2(m)} \leq \frac{1}{2^n}$$

for every n. Then the maximal inequality in Lemma 7.3.11 and Doob's maximal inequality for martingales imply that, for every $T > 0$, M^{u_n} converges to a square-integrable (right continuous) martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P^m)$ uniformly in $t \in [0, T]$, and $u_n(X_t) - u_n(X_0)$ tends to $u(X_t) - u(X_0)$ uniformly in $t \in [0, T]$, P^m-almost surely. Therefore A^{u_n} must go to a limit (A^u) uniformly on $[0, T]$, and A^u thus is continuous and

$$E^m (A_t^u)^2 = 2t\mathcal{E}(u,u) - 2\langle u, u - P_t u\rangle \ .$$

All other conclusions are easy. ■

7.4 Pinned diffusion processes

In this section we deduce several fundamental estimates about transition probability functions for some basic diffusion processes, by using elementary tools like Cameron-Martin's formula, pinned diffusion processes etc..

Let (X_t, P^x) be a time homogenous diffusion process in $\mathbb{R}^n$ with its natural filtration $(\mathcal{F}_t)_{t\geq 0}$. Suppose its transition probability function $P_t(x, dy)$ possesses a positive, continuous density function $h(x, t, y)$ for all $t > 0$ with respect to a σ-finite measure μ. The joint distribution of X_{t_0}, X_{t_1}, $\cdots$, X_{t_k} where $0 = t_0 < t_1 < \cdots < t_k$ then is given by

$$\begin{aligned} & P^x \left\{X_{t_0} \in dx_0, X_{t_1} \in dx_1, \cdots, X_{t_k} \in dx_k\right\} \\ = \quad & h(x_0, t_1, x_1) h(x_1, t_2 - t_1, x_2) \cdots h(x_{k-1}, t_2 - t_1, x_k) \mu(dx_1) \cdots \mu(dx_k) \ . \end{aligned}$$

7.4.1 Conditional diffusions

For a fixed $T > 0$ and a fixed point $y \in \mathbb{R}^n$, define a non-homogenous transition density function

$$H_{T,y}(s,z;t,w) = \frac{h(z,t-s,w)h(w,T-t,y)}{h(z,T-s,y)} \tag{7.22}$$

for all $0 \le s < t < T$, and a transition probability function

$$Q_{s,t}^{T,y} f(z) \triangleq \int_M f(w) H_{T,y}(s,z;t,w)\mu(dw) \quad \text{for } 0 \le s < t < T \ . \tag{7.23}$$

We will omit indices T, y if no confusion may arise. In most part of this section, both $T > 0$ and $y \in M$ are fixed.

Lemma 7.4.1 *Under above notations, we have*

$$P^x \left\{ Q_{s,t}^{T,X_T} f(X_s) g(X_T) | \mathcal{F}_s \right\} = P^x \left\{ f(X_t) g(X_T) | \mathcal{F}_s \right\} \tag{7.24}$$

for any $0 \le s < t < T$, *functions* f *and* g.

Proof. We need to show that

$$P^x \left\{ A Q_{s,t}^{T,X_T} f(X_s) g(X_T) \right\} = P^x \left\{ A f(X_t) g(X_T) \right\} \tag{7.25}$$

where

$$A = F(X_{t_1}, \cdots, X_{t_k}) \in \mathcal{F}_s$$

with $0 < t_1 < t_2 < \cdots < t_k \equiv s < t < T$. It is easily seen that

$$\begin{aligned}
P^x \left\{ A f(X_t) g(X_T) \right\} &= \int h(x,t_1,x_1) h(x_1,t_2-t_1,x_2) \cdots h(x_{k-1},t_k-t_{k-1},x_k) \\
&\quad \cdot h(x_k,t-s,x_{k+1}) h(x_{k+1},T-t,x_{k+2}) F(x_1,\cdots,x_k) f(x_{k+1}) g(x_{k+2}) \\
&= \int h(x,t_1,x_1) h(x_1,t_2-t_1,x_2) \cdots h(x_{k-1},t_k-t_{k-1},x_k) \\
&\quad \cdot H^{T,x_{k+2}}(s,x_k;t-s,x_{k+1}) f(x_{k+1}) F(x_1,\cdots,x_k) \\
&\quad \cdot g(x_{k+2}) h(x_k,T-s,x_{k+2}) \ .
\end{aligned}$$

Taking integration with respect to x_{k+1} we then obtain

$$\begin{aligned}
P^x \left\{ A f(X_t) g(X_T) \right\} &= \int h(x,t_1,x_1) h(x_1,t_2-t_1,x_2) \cdots h(x_{k-1},t_k-t_{k-1},x_k) \\
&\quad \cdot h(x_k,T-s,x_{k+2}) \left\{ Q_{s,t}^{T,x_{k+2}} f(x_k) \right\} F(x_1,\cdots,x_k) g(x_{k+2}) \\
&= P^x \left\{ A Q_{s,t}^{T,X_T} f(X_s) g(X_T) \right\}
\end{aligned}$$

which proves the claim. ■

Hence one may formally write (by taking $g=\delta_y$)

$$P^x\left\{Q_{s,t}f(X_s)\delta_y(X_T)|\mathcal{F}_s\right\}=P^x\left\{f(X_t)\delta_y(X_T)|\mathcal{F}_s\right\} \tag{7.26}$$

for every $0\leq s<t<T$, or equivalently

$$P^x\left\{\left(f(X_t)|\mathcal{F}_s\right)|X_T=y\right\}=P^x\left\{Q_{s,t}f(X_s)|X_T=y\right\} ,$$

which suggests that the process $(X_t)_{t<T}$ possesses the Markov property with transition function $Q_{s,t}$ under "the conditional probability" $P^x\left\{\cdot|X_T=y\right\}$.

Lemma 7.4.2 *Fix $T>0$ and a pair of points x and y in $\mathbb{R}^n$. Define*

$$M_t=\frac{h(X_t,T-t,y)}{h(x,T,y)} \qquad \forall t<T. \tag{7.27}$$

Then $(M_t)_{t<T}$ is a non-negative martingale under the probability P^x.

Proof. We need to show that

$$P^x\left(M_t|\mathcal{F}_s\right)=M_s \quad \text{for all} \quad 0\leq s<t<T .$$

Let

$$A=F(X_{t_1},\cdots,X_{t_k})\in\mathcal{F}_s$$

for $0<t_1<\cdots<t_k=s<t$. Then

$$\begin{aligned} P^x\left(M_tA\right) &= \int h(x,t_1,x_1)h(x_1,t_2-t_1,x_2)\cdots h(x_{k-1},t_k-t_{k-1},x_k) \\ &\quad \cdot h(x_k,t-s,x_{k+1})\frac{h(x_{k+1},T-t,y)}{h(x,T,y)}F(x_1,\cdots,x_k) , \end{aligned}$$

by integrating the variable x_{k+1} first

$$\int h(x_k,t-s,x_{k+1})h(x_{k+1},T-t,y)\mu(dx_{k+1})=h(x_k,T-s,y)$$

we obtain

$$\begin{aligned} P^x\left(M_tA\right) &= \int h(x,t_1,x_1)h(x_1,t_2-t_1,x_2)\cdots h(x_{k-1},t_k-t_{k-1},x_k) \\ &\quad \cdot\frac{h(x_k,T-s,y)}{h(x,T,y)}F(x_1,\cdots,x_k) \\ &= P^x\left(M_sA\right) , \end{aligned}$$

so that the conclusion follows. ■

We may therefore define a probability $P_T^{x,y}$ on the σ-algebra $\sigma\left\{\mathcal{F}_t:t<T\right\}$ by

$$\left.\frac{dP_T^{x,y}}{dP^x}\right|_{\mathcal{F}_t}=\frac{h(X_t,T-t,y)}{h(x,T,y)} \qquad \text{for all } t<T \tag{7.28}$$

called the conditional probability of (X_t) such that $X_0=x$ and $X_T=y$. Since (X_t) is continuous, $\sigma\left\{\mathcal{F}_t:t<T\right\}$ equals $\mathcal{F}_T$. Note that $P_T^{x,y}$ may be not absolutely continuous with respect to P^x on $\mathcal{F}_T$. In the literature, $P_T^{x,y}$ is denoted by $P^x(\cdot|X_T=y)$ or $P(\cdot|X_0=x,X_T=y)$.

Theorem 7.4.3 *The continuous process $(X_t, t \le T)$ under the probability $P_T^{x,y}$ is a Markov process with non-homogeneous transition density function $H(s,z;t,w)$.*

Proof. For any $0 \le s < t < T$ we need to show that

$$P_T^{x,y}\{f(X_t)|\mathcal{F}_s\} = Q_{s,t}f(X_s) \ ,$$

which implies that $(X_t, t \le T)$ is Markovian with transition function $Q_{s,t}$. Again, it is sufficient to show

$$P_T^{x,y}\{f(X_t)A\} = P_T^{x,y}\{AQ_{s,t}f(X_s)\}$$

for $A = F(X_{t_1}, \cdots, X_{t_k}) \in \mathcal{F}_s$ with

$$0 < t_1 < \cdots < t_k = s \ .$$

While as $s < t < T$, $f(X_t)A \in \mathcal{F}_t$, by definition,

$$\begin{aligned}
P_T^{x,y}\{f(X_t)A\} &= P^x\left\{\frac{h(X_t, T-t, y)}{h(x,T,y)} f(X_t)A\right\} \\
&= \int h(x,t_1,x_1)h(x_1,t_2-t_1,x_2)\cdots h(x_{k-1},t_k-t_{k-1},x_k) \\
&\quad \cdot h(x_k,t-s,x_{k+1})\frac{h(x_{k+1},T-t,y)}{h(x,T,y)}F(x_1,\cdots,x_k)f(x_{k+1}) \ .
\end{aligned}$$

Integrating the variable x_{k+1} gives that

$$\int \frac{h(x_k,t-s,x_{k+1})h(x_{k+1},T-t,y)}{h(x_k,T-s,y)} f(x_{k+1})\mu(dx_{k+1}) = Q_{s,t}f(x_k)$$

we therefore obtain

$$\begin{aligned}
P_T^{x,y}\{f(X_t)A\} &= \int h(x,t_1,x_1)h(x_1,t_2-t_1,x_2)\cdots h(x_{k-1},t_k-t_{k-1},x_k) \\
&\quad \cdot\frac{h(x_k,T-s,y)}{h(x,T,y)}F(x_1,\cdots,x_k)Q_{s,t}f(x_k) \\
&= P^x\left\{\frac{h(X_s,T-s,y)}{h(x,T,y)}AQ_{s,t}f(X_s)\right\} \\
&= P_T^{x,y}\{AQ_{s,t}f(X_s)\} \ .
\end{aligned}$$

■

The continuous Markov process $(X_t, t \le T; P_T^{x,y})$ is called a pinned diffusion, or a conditional diffusion given $X_0 = x$ and $X_T = y$, or a diffusion bridge from x to y with running time T.

Proposition 7.4.4 *If (X_t, P^x) is symmetric with respect to μ, i.e.,*

$$h(x,t,y) = h(y,t,x) \ ,$$

then $(X_t, t \le T; P_T^{x,y})$ possesses the same distribution as $(X_{T-t}, t \le T; P_T^{y,x})$.

Proof. It is easy to see that both processes have the same finite distributions.
■

7.4.2 Cameron-Martin's formula for pinned diffusions

Consider an L-diffusion process (X_t, P^x), where L is an elliptic differential operator of second order. L may be in non-divergence form

$$L = \frac{1}{2}\sum_{i,j} g^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_i b^i(x)\frac{\partial}{\partial x_i} \ , \tag{7.29}$$

in this case the symmetric matrix (g^{ij}) is uniformly continuous, or L may be in divergence form

$$L = \frac{1}{2}\frac{1}{q(x)}\sum_{i,j}\frac{\partial}{\partial x_i}\left(q(x)g^{ij}(x)\frac{\partial}{\partial x_j}\right) + \sum_i b^i(x)\frac{\partial}{\partial x_i} \ , \tag{7.30}$$

where coefficients (g^{ij}), positive function $q(x)$ and $(b^i(x))$ are Borel measurable. In both case (g^{ij}) is supposed to be positive definite. Let $(M_t)_{t\geq 0}$ be the martingale part of (X_t). Then

$$\langle M^i, M^j\rangle_t = \int_0^t g^{ij}(X_s)ds \ .$$

Theorem 7.4.5 *The conditional diffusion process $(X_t, t \leq T, P_T^{x,y})$ possesses infinitesimal generator*

$$\begin{aligned} L_{x,y} &= L + \nabla_z \log h(z, T-t, y).\nabla \\ &= L + \sum_{i,j} g^{ij}(z)\frac{\partial}{\partial z_i}\log h(z, T-t, y)\frac{\partial}{\partial z_j} \ . \end{aligned}$$

Proof. By Ito's formula

$$\begin{aligned} &\frac{h(X_t, T-t, y)}{h(x, T, y)} \\ &= \exp\{\log h(X_t, T-t, y) - \log h(x, T, y)\} \\ &= \exp\left\{\int_0^t \nabla_z \log h(X_s, T-s, y)dM_s - \frac{1}{2}\int_0^t |\nabla_z \log h(X_s, T-s, y)|^2 d\langle M\rangle_s\right\} \end{aligned}$$

where (M_t) is the martingale part of (X_t). The conclusion follows then easily. ■

Let

$$c(x) = \sum_i c^i(x)\frac{\partial}{\partial x_i}$$

be a measurable vector field on $\mathbb{R}^n$ which is of at most linear growth, such that

$$P^x \exp\left(\frac{1}{2}\int_0^T |c|_g^2(X_s)ds\right) < \infty \ .$$

Define a probability measure Q^x by

$$\left.\frac{dQ^x}{dP^x}\right|_{\mathcal{F}_t} = \exp\left[\int_0^t \langle c(X_s), dM_s\rangle_g - \frac{1}{2}\int_0^t |c|_g^2(X_s)ds\right] .$$

The lower index g indicates that both the norm and the inner product are computed in term of the metric (g_{ij}) (which is the inverse of (g^{ij})). Thus

$$\langle c(X_s), dM_s\rangle_g = \sum_{i,j} g_{ij}(X_s)c^i(X_s)dM_s^j$$

and

$$|c|_g^2(X_s) = \sum_{i,j} g_{ij}(X_s)c^i(X_s)c^j(X_s) .$$

By the Cameron-Martin formula, $(X_t, \mathcal{F}_t, Q^x)$ is an $L+c$-diffusion process. Let $p(x,t,y)$ denote its transition density.

Lemma 7.4.6 *Suppose $h(x,t,y)$ and $p(x,t,y)$ are continuous, and suppose the function*

$$y \to E_T^{x,y}\left(\exp\left\{\int_0^T \langle c(X_t), dM_t\rangle - \frac{1}{2}\int_0^T |c|^2(X_t)dt\right\}\right)$$

is continuous, then

$$\frac{p(x,T,y)}{h(x,T,y)} = P_T^{x,y}\exp\left\{\int_0^T \langle c(X_t), dM_t\rangle_g - \frac{1}{2}\int_0^T |c|_g^2(X_t)dt\right\} . \tag{7.31}$$

Proof. For every $0<\varepsilon<T$ and any bounded, continuous function φ we have

$$\begin{aligned} & P_T^{x,y}\left(\varphi(X_{T-\varepsilon})\mathrm{e}^{\int_0^{T-\varepsilon}\langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^{T-\varepsilon}|c|^2(X_t)dt}\right) \\ = \; & \frac{1}{h(x,T,y)}P^x\left(h(X_{T-\varepsilon},\varepsilon,y)\varphi(X_{T-\varepsilon})\mathrm{e}^{\int_0^{T-\varepsilon}\langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^{T-\varepsilon}|c|^2(X_t)dt}\right) \end{aligned}$$

then multiplying by $h(x,T,y)$ both sides and integrating in y over $\mathbb{R}^n$ we obtain

$$\begin{aligned} & \int_{\mathbb{R}^d} h(x,T,y)P_T^{x,y}\left(\varphi(X_{T-\varepsilon})\mathrm{e}^{\int_0^{T-\varepsilon}\langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^{T-\varepsilon}|c|^2(X_t)dt}\right)dy \\ = \; & P^x\left(\varphi(X_{T-\varepsilon})\mathrm{e}^{\int_0^{T-\varepsilon}\langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^{T-\varepsilon}|c|^2(X_t)dt}\right) . \end{aligned}$$

Letting $\varepsilon \to 0$ we thus obtain

$$\begin{aligned} & \int_{\mathbb{R}^d} h(x,T,y)P_T^{x,y}\left(\varphi(X_T)\mathrm{e}^{\int_0^T\langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^T|c|^2(X_t)dt}\right)dy \\ = \; & P^x\left(\varphi(X_T)\mathrm{e}^{\int_0^T\langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^T|c|^2(X_t)dt}\right) , \end{aligned}$$

where the exchange of limits with integrals is justified under our conditions. On the other hand

$$\begin{aligned}
&\int_{\mathbb{R}^d} p(x,T,y)\varphi(y)dy \\
&\quad = Q^x\left(\varphi(X_T)\right) \\
&\quad = P^x\left(\varphi(X_T)e^{\int_0^T \langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^T |c|^2(X_t)dt}\right) \\
&\quad = \int_{\mathbb{R}^d} h(x,T,y)P_T^{x,y}\left(\varphi(X_T)\mathrm{e}^{\int_0^T \langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^T |c|^2(X_t)dt}\right)dy \\
&\quad = \int_{\mathbb{R}^d} \varphi(y)h(x,T,y)P_T^{x,y}\left(\mathrm{e}^{\int_0^T \langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^T |c|^2(X_t)dt}\right)dy\ .
\end{aligned}$$

The conclusion follows from the fact that

$$y \to P_T^{x,y}\left(\mathrm{e}^{\int_0^T \langle c(X_t),dM_t\rangle - \frac{1}{2}\int_0^T |c|^2(X_t)dt}\right)$$

is continuous. ■

Theorem 7.4.7 *Same conditions as in Lemma 7.4.6. Let*

$$U_t = \exp\left(\int_0^t \langle c(X_s), dM_s\rangle_g - \frac{1}{2}\int_0^t |c(X_s)|_g^2 ds\right)$$

which is a martingale up to time T. Then

$$p(x,T,y) = h(x,T,y) + \int_0^T P^x\left\{U_t\langle c(X_t), \nabla_x h(X_t, T-t, y)\rangle_g\right\}dt\ . \tag{7.32}$$

Proof. By Lemma 7.4.6

$$\frac{p(x,T,y)}{h(x,T,y)} = P_T^{x,y}\left(U_T\right)$$

On the other hand,

$$\left.\frac{dP_T^{x,y}}{dP^x}\right|_{\mathcal{F}_s} = \frac{h(X_t, T-t, y)}{h(x,T,y)}\ , \qquad \forall t < T\ ,$$

of which the right-hand side will be denoted by N_t for $t < T$. By Girsanov's theorem

$$\tilde{U}_t = U_t - \langle U, N\rangle_t$$

is a martingale under $P_T^{x,y}$ for $t < T$. While as (U_t) is the exponential martingale of $\int_0^t \langle c(X_s), dM_s\rangle_g$ so that

$$U_t = 1 + \int_0^t U_s\langle c(X_s), dM_s\rangle\ .$$

Therefore

$$\langle U, N\rangle_t = \int_0^t U_s \langle c(X_s), \nabla_x \log h(X_s, T-s, y)\rangle_g ds \ ,$$

so that

$$\begin{aligned}
\frac{p(x,T,y)}{h(x,T,y)} &= P_T^{x,y}(U_T) = P_T^{x,y}\left(\tilde{U}_T + \langle U, N\rangle_T\right) \\
&= 1 + P_T^{x,y}\left(\int_0^T U_t \langle c(X_t), \nabla_x \log h(X_t, T-t, y)\rangle dt\right) \\
&= 1 + \int_0^T P_T^{x,y}\left\{U_t \langle c(X_t), \nabla_x \log h(X_t, T-t, y)\rangle\right\} dt \ . \qquad (7.33)
\end{aligned}$$

On the other hand, since U_t is $\mathcal{F}_t$-measurable so that

$$\begin{aligned}
& P_T^{x,y}\left\{U_t \langle c(X_t), \nabla_x \log h(X_t, T-t, y)\rangle\right\} \\
= \ & P^x\left(\frac{h(X_t, T-t, y)}{h(x,T,y)} U_t \langle c(X_t), \nabla_x \log h(X_t, T-t, y)\rangle\right) \\
= \ & \frac{1}{h(x,T,y)} P^x \left\{U_t \langle c(X_t), \nabla_x h(X_t, T-t, y)\rangle\right\}
\end{aligned}$$

and therefore; together with eqn (7.33);

$$\frac{p(x,T,y)}{h(x,T,y)} = 1 + \frac{1}{h(x,T,y)} \int_0^T P^x\left\{U_t \langle c(X_t), \nabla_x h(X_t, T-t, y)\rangle\right\} dt \qquad (7.34)$$

which in turn yields the claim of the theorem. ■

7.4.3 Brownian bridges

Consider a Brownian motion (B_t, P^x) in $\mathbb{R}^n$. For every $T > 0$ and a pair of points x and y, the conditional diffusion process $(B_t, t \le T, P_T^{x,y})$ defined as in the previous section is called a Brownian bridge (from x to y with running time T). Denote by

$$G(x,t,y) = \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{|x-y|^2}{2t}\right)$$

the Gaussian kernel. By definition

$$\begin{aligned}
\left.\frac{dP_T^{x,y}}{dP^x}\right|_{\mathcal{F}_t} &= \frac{G(X_t, T-t, y)}{G(x,T,y)} \\
&= \left(\frac{T}{T-t}\right)^{n/2} \exp\left(\frac{|x-y|^2}{2T} - \frac{|X_t - y|^2}{2(T-t)}\right) \\
&= \exp\left(-\int_0^t \frac{B_s - y}{T-s} dB_s - \frac{1}{2}\int_0^t \frac{|B_s - y|^2}{(T-s)^2} ds\right) \ .
\end{aligned}$$

The transition density function for a Brownian bridge is thus given by

$$\begin{aligned}&\frac{G(z,t-s,w)G(w,T-t,y)}{G(z,T-s,y)}\\ =\ &\frac{1}{(2\pi)^{n/2}}\left(\frac{T-s}{(T-t)(t-s)}\right)^{n/2}\exp\left(\frac{|z-y|^2}{2(T-s)}-\frac{|w-y|^2}{2(T-t)}-\frac{|z-w|^2}{2(t-s)}\right)\ .\end{aligned}$$

It follows that $(B_t, t \le T,\ P_T^{x,y})$ is a Gaussian process with covariance function

$$P_T^{x,y}(B_tB_s)=\left(\frac{T-t}{T}x+\frac{t}{T}y\right)\left(\frac{T-s}{T}x+\frac{s}{T}y\right)+s\wedge t-\frac{st}{T}\ . \tag{7.35}$$

On the other hand we may define, by using the linear structure of $\mathbb{R}^n$, a process

$$\xi_t = x + B_t + \frac{t}{T}(-B_T + y - x) \quad \text{for } 0 \le t \le T\ . \tag{7.36}$$

Proposition 7.4.8 *The Brownian bridge $(B_t, t \le T,\ P_T^{x,y})$ has the same distribution as the process $(\xi_t, t \le T,\ P^0)$. (Under P^0, (B_t) is a standard Brownian motion).*

Proof. Both processes are Gaussian with the same covariance function, so they have the same distribution. ∎

Proposition 7.4.9 *The diffusion process $(\xi_t)_{t<T}$ solves the following stochastic differential equation*

$$dX_t = dW_t + \frac{X_t - y}{T-t}dt\ , \quad X_0 = x\ , \tag{7.37}$$

where (W_t) is a standard Brownian motion in $\mathbb{R}^n$.

Proof. It follows directly from the equation (7.36). ∎

We next give some applications in the estimation of transition density function for Brownian motion with drifts. Given a Borel measurable vector field

$$b(x) = \sum_{i=1}^n b^i(x)\frac{\partial}{\partial x_i}\ ,$$

let $p_b(x,t,y)$ denote the transition density of the diffusion process

$$dX_t = dB_t + b(X_t)dt\ ,$$

where (B_t, P^x) is a Brownian motion in $\mathbb{R}^n$. That is, $p_b(x,t,y)$ is the transition density of the $L = \frac{1}{2}\Delta + b$ diffusion process with respect to the Lebesgue measure. By Lemma 7.4.6

$$\frac{p_b(x,T,y)}{G(x,T,y)} = P_T^{x,y}\exp\left\{\int_0^T b(B_t).dB_t - \frac{1}{2}\int_0^T |b|^2(B_t)dt\right\}\ , \tag{7.38}$$

and by Proposition 7.4.8 we may rewrite the equation (7.38) in terms of the process (ξ_t) defined by (7.36)

$$\frac{p_b(x,T,y)}{G(x,T,y)} = P\exp\left\{\int_0^T b(\xi_t).d\xi_t - \frac{1}{2}\int_0^T |b|^2(\xi_t)dt\right\} \tag{7.39}$$

where by definition

$$\xi_t = x + B_t + \frac{t}{T}(-B_T + y - x)$$

and (B_t, P) is a standard Brownian motion in $\mathbb{R}^n$. However we should be careful with the Itô integral

$$\int_0^T b(\xi_t).d\xi_t \ .$$

We cannot indeed replace $d\xi_t$ simply by

$$d\xi_t = dB_t + \frac{(-B_T + y - x)}{T}dt \ ,$$

rather we need to use Proposition 7.4.9 and substitute the Itô differential $d\xi_t$ by

$$d\xi_t = dW_t + \frac{\xi_t - y}{T-t}dt$$

where (W_t) is, under the probability P, a standard Brownian motion in $\mathbb{R}^n$. If we do make this change of variable, we then obtain

$$\begin{aligned}\frac{p_b(x,T,y)}{G(x,T,y)} &= P\exp\left\{\int_0^T b(\xi_t).dW_t - \frac{1}{2}\int_0^T |b|^2(\xi_t)dt\right. \\ &\qquad \left. + \int_0^T \frac{\xi_t - y}{T-t}.b(\xi_t)dt\right\} \ .\end{aligned} \tag{7.40}$$

Theorem 7.4.10 *Let $G(x,t,y)$ be the Gaussian kernel, and let b be a bounded vector field. Then for any $p > 1$ we have*

$$\frac{p_b(x,t,y)}{G(x,t,y)} \geq \exp\left\{-\frac{||b||^2}{2}t - ||b||\left(\left(\frac{2}{3}a_{n,1} + 2\right)\sqrt{t} + |x-y|\right)\right\}$$

where

$$a_{n,p} = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} |z|^p e^{-|z|^2/2}dz \ .$$

Proof. We may use the Jensen inequality in (7.40) to obtain

$$\begin{aligned}\frac{p_b(x,T,y)}{G(x,T,y)} &\geq \exp\left\{-\frac{1}{2}\int_0^T P\left(|b|^2(\xi_t)\right)dt + \int_0^T \frac{1}{T-t}P\left((\xi_t - y).b(\xi_t)\right)dt\right\} \\ &\geq \exp\left\{-\frac{1}{2}||b||^2 T - ||b||\int_0^T \frac{1}{T-t}P\left(|\xi_t - y|\right)dt\right\} \ .\end{aligned}$$

While, using the representation of the Brownian bridge,

$$\begin{aligned}
P\left(|\xi_t - y|\right) &= P\left(|\frac{T-t}{T}(x-y+B_t) - \frac{t}{T}(B_T - B_t)|\right) \\
&\leq \frac{T-t}{T}P|x-y+B_t| + \frac{t}{T}P|B_T - B_t| \\
&\leq \frac{T-t}{T}\frac{1}{(2\pi t)^{n/2}}\int_{\mathbb{R}^n}|z+x-y|\mathrm{e}^{-|z|^2/(2t)}dz + \frac{t}{T}\sqrt{P|B_T-B_t|^2} \\
&\leq \frac{T-t}{T}\left(a_{n,1}\sqrt{t} + |x-y|\right) + \frac{t}{T}\sqrt{T-t}\ ,
\end{aligned}$$

so that

$$\begin{aligned}
&\int_0^T \frac{1}{T-t}P\left(|\xi_t - y|\right)dt \\
\leq\ & \frac{1}{T}\int_0^T\left(a_{n,1}\sqrt{t} + |x-y|\right) + \int_0^T \frac{1}{\sqrt{T-t}}\frac{t}{T}dt \\
\leq\ & \left(\frac{2}{3}a_{n,1} + 2\right)\sqrt{T} + |x-y|\ .
\end{aligned}$$

Therefore

$$\frac{p_b(x,T,y)}{G(x,T,y)} \geq \exp\left\{-\frac{||b||^2}{2}T - ||b||\left(\left(\frac{2}{3}a_{n,1}+2\right)\sqrt{T} + |x-y|\right)\right\}\ .$$

■

In order to deduce a useful upper bound for $p_b(x,t,y)$ we may use Theorem 7.4.7.

Lemma 7.4.11 *Let (B_t, P^x) be a Brownian motion in $\mathbb{R}^n$. Then for every $q \geq 1$ and all $T > t$ we have*

$$\begin{aligned}
P^x\left(|B_t - y|^q e^{-\frac{q|B_t-y|^2}{2(T-t)}}\right) &\leq (T-t)^{(n+q)/2}\left(\frac{1}{T}\right)^{n/2} e^{-\frac{|x-y|^2}{2T}} \\
&\quad \cdot\left(\sqrt{\frac{1}{q^q}}a_{n,q} + \left(\frac{|x-y|}{\sqrt{T}}\right)^q\right)\ . \qquad (7.41)
\end{aligned}$$

7.5 Additional topics

Proof of Hille-Yosida theorem

Due to the importance of Hille-Yosida, we would like to include a proof here.

Lemma 7.5.1 *Let L satisfy the conditions 1 and 2 of Theorem 7.1.5, and set $R_\lambda = (\lambda I - L)^{-1}$. Then*

$$\lim_{\lambda\to\infty}\lambda R_\lambda x = x \qquad \forall x \in M\ .$$

Indeed, consider first those $x \in D(L)$. Then

$$\lambda R_\lambda x - x = LR_\lambda x = R_\lambda Lx$$

so that

$$\begin{aligned} ||\lambda R_\lambda x - x|| &= ||R_\lambda Lx|| \\ &\leq \frac{1}{\lambda}||Lx|| \end{aligned}$$

as $\lambda \to \infty$. However $D(L)$ is dense in B and $||\lambda R_\lambda x|| \leq 1$, therefore $\lambda R_\lambda x \to x$ as $\lambda \to \infty$ for every $x \in B$. That proves Claim 1.

For every $\lambda > 0$, the Yosida approximation of L is defined as

$$L_\lambda = \lambda L R_\lambda = \lambda^2 R_\lambda - \lambda I \ .$$

Note L_λ is a bounded linear operator for each $\lambda > 0$, and moreover, if $x \in D(L)$ then

$$LR_\lambda x = R_\lambda Lx$$

and we have

Corollary 7.5.2 *Let L satisfy the conditions 1 and 2 of Theorem 7.1.5. Then*

$$\lim_{\lambda\to\infty} L_\lambda x = Lx \qquad \forall x \in D(L) \ .$$

If T is a bounded linear operator on B, then its exponential e^T is given by the formula

$$\mathrm{e}^T(x) = \sum_{k=0}^{\infty} \frac{1}{k!}(T^k x)$$

which is again a bounded linear operator on B. Indeed

$$\begin{aligned} ||\mathrm{e}^T|| &\leq \sum_{k=0}^{\infty} \frac{1}{k!}||T^k|| \\ &\leq \sum_{k=0}^{\infty} \frac{1}{k!}||T||^k = \mathrm{e}^{||T||} \ . \end{aligned}$$

If S and T are two bounded linear operator and if T and S commute, then

$$\mathrm{e}^{T+S} = \mathrm{e}^T \mathrm{e}^S \ .$$

In particular for a bounded linear operator T on a Banach space, then $(\mathrm{e}^{tT})_{t\in\mathbf{C}}$ is a commutative family of bounded linear operators, and

$$\begin{aligned} ||\mathrm{e}^{tT}(x) - x|| &= \left\| \sum_{k=1}^{\infty} \frac{t^k}{k!}(T^k x) \right\| \\ &\leq \sum_{k=1}^{\infty} \frac{t^k}{k!}||T||^k ||x|| \\ &= \left(\mathrm{e}^{t||T||} - 1\right) ||x|| \end{aligned}$$

so that

$$||\mathrm{e}^{tT} - I|| \leq \mathrm{e}^{t||T||} - 1 \to 0 \qquad \text{as } t \to 0 \ .$$

Moreover

$$\begin{aligned}
\left\| \frac{\mathrm{e}^{tT}(x) - x}{t} - Tx \right\| &= \left\| \sum_{k=2}^{\infty} \frac{t^{k-1}}{k!}(T^k x) \right\| \\
&= \left\| \sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+2)!}(T^{k+2} x) \right\| \\
&\leq t||T||^2 \left\| \sum_{k=0}^{\infty} \frac{t^k}{k!}(T^k x) \right\| \\
&\leq t||T||^2 \mathrm{e}^{t||T||}||x||
\end{aligned}$$

Therefore $(\mathrm{e}^{tT})_{t\geq 0}$ is a strongly semigroup of bounded linear operators with infinitesimal generator T.

Since L_λ is bounded for any $\lambda > 0$, it is the infinitesimal generator of the strongly continuous semigroup

$$\mathrm{e}^{tL_\lambda} = \sum_{n=0}^{\infty} \frac{t^n}{n!} L_\lambda^n \ .$$

Moreover, since λI and R_λ commute, so that $\mathrm{e}^{tL_\lambda} = \mathrm{e}^{-\lambda t}\mathrm{e}^{t\lambda^2 R_\lambda}$ and therefore for any $t > 0$

$$\begin{aligned}
||\mathrm{e}^{tL_\lambda}|| &\leq \mathrm{e}^{-\lambda t} \sum_{n=0}^{\infty} \frac{t^{2n}}{n!} \lambda^{2n} ||R_\lambda||^n \\
&= \mathrm{e}^{-\lambda t}\mathrm{e}^{\lambda^2 t||R_\lambda||} \\
&\leq 1.
\end{aligned}$$

Hence for each $\lambda > 0$, $(\mathrm{e}^{tL_\lambda})_{t\geq 0}$ is a semigroup of contractions on B with infinitesimal generator L_λ. The next lemma shows that e^{tL_λ} converges as $\lambda \to \infty$. The limit shall be the strongly continuous semigroup of contractions with infinitesimal generator L.

Lemma 7.5.3 *Under the same assumption as in Lemma 7.5.1. For any $\lambda, \mu > 0$*

$$\left\| e^{tL_\lambda}x - e^{tL_\mu}x \right\| \leq t \left\| L_\lambda x - L_\mu x \right\| \ .$$

Therefore, for every $x \in D(L)$, $e^{tL_\lambda}x$ converges as $\lambda \to \infty$ uniformly in t in any bounded interval.

Let us prove this claim. Since all bounded linear operators e^{tL_λ}, e^{tL_μ}, L_λ and L_μ commute with each other, consequently

$$\begin{aligned}
\mathrm{e}^{tL_\lambda}x - \mathrm{e}^{tL_\mu}x &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{e}^{stL_\lambda}\mathrm{e}^{t(1-s)L_\mu}x\right)\mathrm{d}s \\
&= t\int_0^t \mathrm{e}^{stL_\lambda}\mathrm{e}^{t(1-s)L_\mu}\left(L_\lambda x - L_\mu x\right)\mathrm{d}s \ ,
\end{aligned}$$

together with the fact that $||e^{stL_\lambda}|| \leq 1$ and $||e^{t(1-s)L_\mu}|| \leq 1$, it follows thus that

$$\begin{aligned}\left\|e^{tL_\lambda}x - e^{tL_\mu}x\right\| &\leq t\int_0^t ||e^{stL_\lambda}e^{t(1-s)L_\mu}\left(L_\lambda x - L_\mu x\right)||\mathrm{d}s \\ &\leq t||L_\lambda x - L_\mu x|| \ .\end{aligned}$$

Let $x \in D(L)$. Then

$$\begin{aligned}\left\|e^{tL_\lambda}x - e^{tL_\mu}x\right\| &\leq t||L_\lambda x - L_\mu x|| \\ &= t||L_\lambda x - Lx|| + t||L_\mu x - Lx|| \ ,\end{aligned}$$

by Corollary 7.5.2, it thus follows that $e^{tL_\lambda}x$ converges as $\lambda \to \infty$ uniformly in t in any bounded interval. That proves the conclusion.

We are now in a position to complete the proof of the Hille-Yosida theorem 7.1.5. Let

$$P_t x \equiv \lim_{\lambda\to\infty} e^{tL_\lambda}x \qquad \forall x \in D(L),\ t \geq 0\ . \tag{7.42}$$

Then $||P_t x|| \leq 1$. Since $D(L)$ is dense in B, so that $\lim_{\lambda\to\infty}e^{tL_\lambda}x$ exists for every $x \in B$. Moreover the convergence in (7.42) is uniform in t on any bounded interval, so that

$$\begin{aligned}P_{t+s}x &= \lim_{\lambda\to\infty} e^{(t+s)L_\lambda}x = \lim_{\lambda\to\infty} e^{tL_\lambda}\left(e^{sL_\lambda}x\right) \\ &= \lim_{\lambda\to\infty} e^{tL_\lambda}\left(\lim_{\lambda\to\infty} e^{sL_\lambda}x\right) \\ &= P_t(P_s x)\end{aligned}$$

and

$$\begin{aligned}\lim_{t\downarrow 0}||P_t x - x|| &= \lim_{t\downarrow 0}||\lim_{\lambda\to\infty}\left(e^{tL_\lambda}(x) - x\right)|| \\ &= ||\lim_{\lambda\to\infty}\lim_{t\downarrow 0}\left(e^{tL_\lambda}(x) - x\right)|| \\ &= 0\ .\end{aligned}$$

Therefore $(P_t)_{t\geq 0}$ is a strongly continuous semigroup of contractions on B. Next we prove that L is the infinitesimal generator of $(P_t)_{t\geq 0}$. Let $(A, D(A))$ be the infinitesimal generator of $(P_t)_{t\geq 0}$. If $x \in D(L)$ then

$$\begin{aligned}P_t x - x &= \lim_{\lambda\to\infty}\left(e^{tL_\lambda}(x) - x\right) = \lim_{\lambda\to\infty}\int_0^t e^{sL_\lambda}(L_\lambda x)\mathrm{d}s \\ &= \int_0^t \lim_{\lambda\to\infty} e^{sL_\lambda}(L_\lambda x)\mathrm{d}s = \int_0^t \lim_{\lambda\to\infty} e^{sL_\lambda}(\lim_{\lambda\to\infty} L_\lambda x)\mathrm{d}s \\ &= \int_0^t P_s(Lx)\mathrm{d}s\end{aligned}$$

and therefore $x \in D(A)$ and $Ax = Lx$. Since 1 belongs to the resolvent sets both of A and L we therefore have

$$(I - A)D(L) = (I - L)D(L) = B$$

so that

$$\begin{aligned} D(A) &= (I - A)^{-1}B = (I - L)^{-1}B \\ &= D(L) \ . \end{aligned}$$

Therefore $D(A) = D(L)$ and $L = A$.

The contraction semigroup $(P_t)_{t\geq 0}$ with infinitesimal generator L sometimes is denoted by e^{tL}, although e^{tL} is not necessarily given by power series.

Corollary 7.5.4 *Let L be the infinitesimal generator of a strongly continuous, contraction semigroup $(P_t)_{t\geq 0}$ on a Banach space B, and let $L_\lambda = \lambda^2 R_\lambda - \lambda I$ (where $R_\lambda = (\lambda I - L)^{-1}$) be the Yosida approximation of L. Then*

$$P_t x = \lim_{\lambda\to\infty} e^{tL_\lambda} x$$

uniformly in t on any bounded interval.

Corollary 7.5.5 *Under the same assumption in the previous corollary. The resolvent set $\rho(L) \supseteq \{\lambda : Re\lambda > 0\}$ and*

$$||(\lambda I - L)^{-1}|| \leq \frac{1}{Re\lambda}$$

for any λ such that such $Re\lambda > 0$.

Proof. If $\mathrm{Re}\lambda > 0$ then

$$R_\lambda \equiv \int_0^\infty \mathrm{e}^{-\lambda t}(P_t x)\mathrm{d}t$$

is well defined bounded linear operator, which is $(\lambda I - L)^{-1}$. ■

Exponential formula In real analysis, we have the following famous limit

$$\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = \mathrm{e}^x$$

which is indeed the definition of the real number e. The above formula still holds in the setting of semigroups.

If $(P_t)_{t\geq 0}$ is a strongly continuous semigroup of contractions in a Banach space B with infinitesimal generator L, then

$$\begin{aligned} P_t x &= \lim_{n\to\infty}\left(I - \frac{t}{n}L\right)^{-n} x \\ &= \lim_{n\to\infty}\left(\frac{t}{n}R_{n/t}\right)^n x \end{aligned}$$

uniformly in t on any bounded interval.

Chapter 8

Analysis of Dirichlet forms

In this chapter we present a few analytic aspects about symmetric Markov semigroups, and emphasize on the L^p-technique for Markov semigroups.

8.1 Heat semigroups

In this section we study Markov semigroups defined by metric structures on a smooth manifold.

8.1.1 Riemannian metrics

Let M be a smooth manifold of dimension n. A *metric* (or Riemannian metric) on M is a positive-definite symmetric $(0,2)$ type tensor field on M. In a local coordinate system $(x^1, \cdots, x^n)$, a metric may be written as

$$\mathrm{d}s^2 = \sum_{i,j=1}^{n} g_{ij}(x)\mathrm{d}x^i \mathrm{d}x^j$$

where (g_{ij}) is a smooth, symmetric and positive-definite $n \times n$ matrix-valued function (locally defined) on M. For simplicity we will write a metric on M as (g_{ij}).

Given a metric (g_{ij}) on M, (g^{ij}) is the inverse matrix of (g_{ij}) and g denotes the determinate $\det(g_{ij})$, both are locally defined quantities on M. The metric (g_{ij}) on M defines the following global quantities which are independent of the choice of a local coordinate system, although the formulas below are expressed in terms of local coordinates.

1. *Volume measure* μ_g: in a local coordinate system, $\mu_g(\mathrm{d}x) = \sqrt{\det(g_{ij}(x))}\mathrm{d}x$ where $\mathrm{d}x$ is the Lebesgue measure in $\mathbb{R}^n$.

2. *The scalar product* (also called metric) on the tangent space TM and co-tangent space T^*M: if $\xi = \sum_{i=1}^n \xi^i \frac{\partial}{\partial x_i}$ and $\eta = \sum_{i=1}^n \eta^i \frac{\partial}{\partial x_i}$ then

$$\langle \xi, \eta \rangle = \sum_{i,j=1}^n g_{ij} \xi^i \eta^j \ .$$

Similarly, if $\alpha = \sum_{i=1}^n \alpha_i \mathrm{d}x^i$ and $\beta = \sum_{i=1}^n \beta_i \mathrm{d}x^i$ then

$$\langle \alpha, \beta \rangle = \sum_{i,j=1}^n g^{ij} \alpha_i \beta_j \ .$$

We will use $|\xi|^2$ to denote $\langle \xi, \xi \rangle$ etc.

3. *The gradient* ∇u of a differentiable function u on M is a vector field on M defined by

$$\nabla u = \sum_{i=1}^n \left(g^{ij} \frac{\partial u}{\partial x^j} \right) \frac{\partial}{\partial x^i} \ .$$

Therefore

$$\langle \nabla u, \nabla v \rangle = \sum_{i,j=1}^n g^{ij} \frac{\partial u}{\partial x^i} \frac{\partial v}{\partial x^j} \ . \tag{8.1}$$

4. *The Laplace-Beltrami operator* Δ is an elliptic operator of second order on M, in a local coordinate system

$$\Delta = \frac{1}{\sqrt{g}} \sum_{i,j=1}^n \frac{\partial}{\partial x^j} \left\{ g^{ij} \sqrt{g} \frac{\partial}{\partial x^i} \right\} \ . \tag{8.2}$$

It can be also written as

$$\Delta = \sum_{i,j=1}^n g^{ij} \frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^n \left\{ \frac{1}{\sqrt{g}} \sum_{j=1}^n \frac{\partial}{\partial x^j} \left(g^{ij} \sqrt{g} \right) \right\} \frac{\partial}{\partial x^i} \ .$$

5. *Integration by parts*: for $u \in C_K^2(M)$ (the space of twice differentiable functions with compact supports), and $v \in C^2(M)$

$$\begin{aligned} \int_M v(\Delta u) \mathrm{d}\mu_g &= \int_M u(\Delta v) \mathrm{d}\mu_g \\ &= -\int_M \langle \nabla u, \nabla v \rangle \mathrm{d}\mu_g \ . \end{aligned} \tag{8.3}$$

The metric (g_{ij}) naturally introduces a distance function $d(x,y)$ on the manifold:

$$d(x,y) = \sup \left\{ u(x) - u(y) : u \in C^1(M) \text{ s.t. } ||\nabla u||_\infty \leq 1 \right\}$$

where

$$||\nabla u||_\infty = \sup\{|\nabla u|(x) : x \in M\} \ .$$

$d(x,y)$ is nothing but the geodesic distance between x and y.

A metric (g_{ij}) on a smooth manifold M is *complete*, if the metric space (M,d) (where d is the distance defined by the metric (g_{ij})) is complete. In this case (M,g_{ij}) is called a complete manifold.

We state several facts about complete manifolds, their proofs can be found in any standard text books on Riemannian geometry.

Lemma 8.1.1 *Let (g_{ij}) be a* complete metric *on smooth manifold M.*

1) If $x \in M$ and $r > 0$, the open ball

$$B(x,r) = \{y \in M : d(y,x) < r\}$$

is pre-compact.

2) There is an increasing sequence $\{D_n : n \geq 1\}$ of pre-compact subsets in M, such that $\bar{D}_n \subset D_{n+1}$ and $\cup_{n\geq 1} D_n = M$.

The sequence $\{D_n : n \geq 1\}$ in item 2) of the lemma is called an *exhausting sequence* of a complete manifold (M,g_{ij}).

Lemma 8.1.2 *Let (M,g_{ij}) be a complete manifold. Then there exists a sequence $(h_n)_{n\in\mathbf{N}}$ of smooth functions with compact supports such that $|\nabla h_n| \leq 1/n$, $0 \leq h_n \leq 1$ and $h_n \uparrow 1$ as $n \to \infty$. $(h_n)_{n\geq 1}$ is called an* exhausting sequence *of the complete manifold M.*

Indeed, if $\{D_n : n \geq 1\}$ is an exhausting sequence in Lemma 8.1.1, then $\{1_{D_n} : n \geq 1\}$ would do if 1_{D_n} were smooth. What one has to do is to modify the sequence $\{1_{D_n} : n \geq 1\}$ so that it satisfies the properties in Lemma 8.1.2.

Let $M = \mathbb{R}^n$ be the Euclidean space of dimension n, and let (g_{ij}) be a smooth function on $\mathbb{R}^n$ with values in the space of symmetric $n \times n$ square matrices, such that

$$\lambda|\xi|^2 \leq \sum_{i,j=1}^{n} g_{ij}(x)\xi^i\xi^j \leq \lambda^{-1}|\xi|^2 \qquad \forall \xi \in \mathbb{R}^n$$

for some constant $\lambda \in (0,1]$. Then the distance determined by the metric (g_{ij}) is equivalent to the Euclidean distance, and therefore (g_{ij}) is a complete metric on $\mathbb{R}^n$. The Laplace-Beltrami operator is an elliptic differential operator of second-order in divergence form

$$L = \frac{1}{\sqrt{\det(g_{ij})}} \sum_{i,j=1} \frac{\partial}{\partial x^i} g_{ij}(x)\sqrt{\det(g_{ij})}\frac{\partial}{\partial x^j}$$

where $\sqrt{\det(g_{ij})}$ is bounded away from 0 and above

$$\lambda^n \leq \sqrt{\det(g_{ij})} \leq \lambda^{-n} \ .$$

8.1.2 The heat kernel

Let M be a smooth manifold of dimension n, and let (g_{ij}) be a complete metric on M. The main result of this part is the following

Theorem 8.1.3 *There is a smooth function $h : (0,+\infty) \times M \times M \to (0,+\infty)$ satisfying the following.*

1) For every $y \in M$, $u(t,y) \equiv h(t,x,y)$ satisfies the heat equation

$$\left(\Delta - \frac{\partial}{\partial t}\right) u(t,y) = 0 \quad \text{on } (0,+\infty) \times M$$

such that

$$\lim_{t \downarrow 0} \int_M h(t,x,y)u(y)\mu_g(dy) = u(x) \qquad \forall x \in M \ .$$

2) $h(t,x,y)$ is symmetric in x and y:

$$h(t,x,y) = h(t,y,x)$$

and satisfies the Chapman-Kolmogorov equation

$$h(t+s,x,y) = \int_M h(s,x,z)h(t,z,y)\mu_g(dz) \ .$$

3) For any $t > 0$ and $x \in M$, $h(t,x,y)$ is a sub-probability density function

$$\int_M h(t,x,y)\mu_g(dy) \leq 1 \ .$$

We prove this theorem through a series of lemmas.

8.1.3 The heat semigroup

Let (M, g_{ij}) be a complete manifold, and let $h(t,x,y)$ be the heat kernel constructed in the previous sub-section. Define

$$P(t,x,\mathrm{d}y) = h(t,x,y)\mu_g(\mathrm{d}y) \ .$$

Then $(P_t)_{t\geq 0}$ is a (sub-)Markov semigroup on M, and therefore $(P_t)_{t\geq 0}$ can be considered as strongly continuous contraction semigroup on $C_b(M)$ or as strongly continuous contraction semigroup on $L^2(M,\mu_g)$ symmetric with respect to the volume measure μ_g. We are going to calculate the infinitesimal generator of $(P_t)_{t\geq 0}$ and identifies its Dirichlet space.

Lemma 8.1.4 *Let Δ be the Laplace-Beltrami operator on a complete manifold (M, g_{ij}). Then*

1) $(\Delta, C_K^\infty(M))$ is closable in the Banach space $(C_b(M), ||\cdot||_\infty)$,

2) for any $p \geq 1$, $(\Delta, C_K^\infty(M) \cap L^p(M,\mu_g))$ is closable on $L^p(M,\mu_g)$.

It is easy to verify via Lebesgue's dominated convergence theorem that $(P_t)_{t\geq 0}$ is *strongly Fellerian* in the sense that P_t maps any bounded measurable function to a bounded continuous function (for any $t > 0$) and $P_t u \to u$ in $C_b(M)$. In fact for every $t > 0$, P_t sends a bounded measurable function to a smooth function.

Proposition 8.1.5 *$C_K^\infty(M)$ is the core of the infinitesimal generator L of the heat semigroup $(P_t)_{t\geq 0}$ on $(C_b(M), ||\cdot||_\infty)$, and $L = \Delta$.*

Proof. Since

$$P_t u(x) = \int_M h(t,x,y) u(y) \mu_g(\mathrm{d}y)$$

for $u \in C_K^\infty(M)$, we have

$$\begin{aligned}
\frac{\partial}{\partial t} P_t u &= \int_M \left\{ \frac{\partial}{\partial t} h(t,x,y) \right\} u(y) \mu_g(\mathrm{d}y) \\
&= \int_M (\Delta h(t,x,y))\, u(y) \mu_g(\mathrm{d}y) \\
&= \int_M h(t,x,y) (\Delta u(y))\, \mu_g(\mathrm{d}y) \ .
\end{aligned}$$

Therefore

$$\left.\frac{\partial}{\partial t}\right|_{t=0} P_t u = \Delta u$$

which implies that $u \in D(L)$ and $Lu = \Delta u$. ■

Similarly, we may identify the infinitesimal generator of $(P_t)_{t\geq 0}$ as strongly continuous (sub-)Markov semigroup on L^p-spaces.

Proposition 8.1.6 *For any $p \in [1, +\infty]$, $(P_t)_{t\geq 0}$ is a strongly continuous (sub-)Markov semigroup on $L^p(M, \mu_g)$, its infinitesimal generator L possesses core $C_K^\infty(M)$ on which $L = \Delta$.*

Lemma 8.1.7 *The Laplace-Beltrami operator Δ is essentially self-adjoint on $L^2(M, \mu_g)$.*

Corollary 8.1.8 *(Integration by parts) If $u, v \in D(\Delta)$ the domain of the infinitesimal generator Δ of $(P_t)_{t\geq 0}$ on $L^2(M, \mu_g)$, then*

$$\begin{aligned}
\int_M v(\Delta u)\, d\mu_g &= \int_M u(\Delta v)\, d\mu_g \\
&= -\int_M \langle \nabla u, \nabla v \rangle\, d\mu_g \ .
\end{aligned} \tag{8.4}$$

Corollary 8.1.9 *Endow the vector space $C_K^\infty(M)$ with norm*

$$||u||_{\mathcal{E}_1}^2 = \int_M (|u|^2 + |\nabla u|^2)\, d\mu_g$$

and let $H_0^1(M)$ be the completion of the normed space $(C_K^\infty(M), ||\cdot||_{\mathcal{E}_1}^2)$. Then the Dirichlet space associated with the heat semigroup $(P_t)_{t\geq 0}$ on $L^2(M,\mu_g)$ is given as the following

$$\begin{aligned}
\mathcal{E}(u,v) &= \int_M \langle \nabla u, \nabla v\rangle d\mu_g \qquad \forall u,v\in D(\mathcal{E}),\\
D(\mathcal{E}) &= H_0^1(M)\ .
\end{aligned}$$

Lemma 8.1.10 *If (M, g_{ij}) is complete, then $H_0^1(M) = W^{1,2}(M)$ where*

$$W^{1,2}(M) = \{u \in L^2(M,\mu_g) \text{ such that } |\nabla u| \in L^2(M,\mu_g)\}$$

is the Sobolev space on (M, g_{ij}).

Proof. Need a proof or a reference???. ■

8.1.4 Curvature and dimension

As we have observed that, on a complete manifold, the Laplace-Beltrami operator Δ is the infinitesimal generator of the heat semigroup which can be considered as a strongly continuous contraction semigroup on L^p space (for $p\in[1,+\infty]$) or on the Banach space of bounded continuous functions on M. However

$$\Delta = \frac{1}{\sqrt{g}}\sum_{i,j=1}^d \frac{\partial}{\partial x^i}\sqrt{g}g^{ij}\frac{\partial}{\partial x^j}$$

is an elliptic differential operator of second-order, so we may consider the heat equation on a domain D

$$\left(\Delta - \frac{\partial}{\partial t}\right)u(t,x) = 0 \qquad \text{on } [0,T]\times D\ . \tag{8.5}$$

The following lemma can be verified by means of chain rule.

Lemma 8.1.11 *1) If $F\in C^1(\mathbb{R}^n;\mathbb{R})$ and $u=(u_1,\cdots,u_n)\in C^1(D)$ then*

$$\nabla F(u) = \sum_{j=1}^n \frac{\partial F}{\partial x^j}(u)\,(\nabla u_j) \tag{8.6}$$

and if $F\in C^2(\mathbb{R}^n;\mathbb{R})$ and $u=(_1,\cdots,u_n)\in C^2(D)$ then

$$\Delta(F(u)) = \sum_{i=1}^n \frac{\partial F}{\partial x^i}(u)\,(\Delta u_i) + \sum_{i,j=1}^n \frac{\partial^2 F}{\partial x^i \partial x^j}(u)\langle \nabla u_i, \nabla u_j\rangle\ . \tag{8.7}$$

2) If $u\in C^2(M)$ then

$$\frac{1}{2}\Delta u^2 - u(\Delta u) = |\nabla u|^2\ . \tag{8.8}$$

The last equation (8.8) allows us to introduce the concept of curvature and dimension in terms of the Laplace-Beltrami operator Δ. Obviously the method applies to a larger class of Markov operators rather then Laplacians on manifolds. We will follow the approach in Bakry-Emery, that is, we will take the Böchner identity as our starting point.

According to (8.8) we define

$$\Gamma(u) = \frac{1}{2}\Delta(u^2) - u(\Delta u) \qquad \forall u \in C^2(M)$$

so that $\Gamma(u) = |\nabla u|^2$. Clearly Γ satisfies the polarization identity so that it can be extended into a quadratic form, naturally

$$\Gamma(u,v) = \frac{1}{2}\left\{\Delta(uv) - v(\Delta u) - u(\Delta v)\right\} \quad \forall u, v \in C^2(M) \ .$$

Next iterating the previous procedure for Γ to define the curvature operator of Δ:

$$\Gamma_2(u) = \frac{1}{2}\Delta(\Gamma(u)) - \Gamma(\Delta u, u)$$

and extend Γ_2 through polarization into a quadratic form $\Gamma_2(u,v)$ such that $\Gamma_2(u) = \Gamma_2(u,u)$.

Lemma 8.1.12 *If* $F \in C^1(\mathbb{R}^n;\mathbb{R})$ *and* $u = (_1, \cdots, u_n), v \in C^1(D)$ *then*

$$\Gamma(F(u), v) = \sum_{j=1}^{n} \frac{\partial F}{\partial x^j}(u)\Gamma\left(u_j, v\right) \ . \tag{8.9}$$

The chain rule for Γ_2 is much more complicated, and we will demonstrate its applications later on.

The significance of these definitions of course lies in the fact synthesized as the Böchner identity

$$\Gamma_2(u) = |\nabla\nabla u|^2 + R_{ij}(\nabla^i u)(\nabla^j u) \tag{8.10}$$

where R_{ij} is the Ricci curvature tensor and ∇ is the Levi-Civita connection, and therefore $\nabla\nabla u$ is the hessian tensor of u.

It is said that the Ricci curvature (R_{ij}) is bounded below by K at a point $x \in M$, if at this point x, the eigenvalues of (R_{ij}) are all greater then K. Therefore, the fact that the Ricci curvature is bounded below by K can be stated as

$$\Gamma_2(u) \geq K\Gamma(u)$$

as $|\nabla\nabla u|^2 \geq 0$. This fact, together with the fact that Δu is the trace of the hessian tensor $\nabla\nabla u$, yields that

$$|\nabla\nabla u|^2 \geq \frac{1}{n}(\Delta u)^2 \ .$$

Therefore, if the Ricci curvature is bounded below by K (which may be function on M, for example K may be the lowest eigenvalue of the Ricci curvature tensor (R_{ij})), then we have the following fundamental inequality

$$\Gamma_2(u) \geq \frac{1}{n}(\Delta u)^2 + K\Gamma(u) \ . \tag{8.11}$$

This inequality is called CD(K, n), curvature / dimension inequality. We must point out, which is important in the following development, that the definitions of the operator $\Gamma(u)$, $\Gamma_2(u)$ and the curvature / dimension inequality depend only on the differential operator Δ itself. In other words, these definitions work for any reasonable operators, in particular diffusion operators.

Remark 8.1.13 *It can be shown (Exercise) that*

$$\Gamma(u) = \lim_{t\to 0} \frac{1}{2t}\left(P_t(u^2) - (P_t u)^2\right)$$

so it is always true that $\Gamma(u) \geq 0$, for any abstract diffusion semigroup

Example 8.1.14 *Consider the diffusion semigroup (P_t) associated to the Brownian motion*

$$P_t f(x) = \int_{\mathbf{R}^d} \frac{1}{(4\pi t)^{d/2}} e^{-|x-y|^2/(4t)} f(y)\, dy \ .$$

The generator $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$, and a simple computation shows that

$$\Gamma_2(u) = \sum_{i,j=1}^d \left(\frac{\partial^2 u}{\partial x_i \partial x_j}\right)^2 \ .$$

In short, $\Gamma_2(u) = |\nabla\nabla u|^2$, so we say that the curvature of (P_t) is zero. In this case

$$\nabla P_t f = P_t \nabla f; \qquad |\nabla P_t f| \leq P_t |\nabla f| \ .$$

Theorem 8.1.15 *Let (M, g_{ij}) be a complete manifold such that the curvature / dimension CD(K, n) holds. Then for any $u \in C_b^\infty(M) \cap D(L)$*

$$|\nabla P_t u| \leq e^{Kt} P_t |\nabla u|$$

Proof. We have shown that P_t preserves the space $C_b^\infty(M)$, where $(P_t)_{t\geq 0}$ is the heat semigroup

$$P_t u(x) = \int_M h(t, x, y) u(y) \mu_g(\mathrm{d}y) \ .$$

For $\varepsilon > 0$ consider the following function

$$v(s, \cdot) = \left(e^{2Ks}|\nabla P_{t-s} u|^2 + \varepsilon\right)^{1/2}$$

which is a smooth function on $[0,t)\times M$ and continuous on $[0,t]\times M$, satisfying the initial and terminal conditions

$$v(0,\cdot)=\left(|\nabla P_t u|^2+\varepsilon\right)^{1/2} \quad \text{and} \quad v(t,\cdot)=\left(e^{2Kt}|\nabla u|^2+\varepsilon\right)^{1/2}$$

A simple computation shows that

$$\left(\Delta+\frac{\partial}{\partial s}\right)v(s,x)\geq 0 \qquad \text{on } [0,t)\times M.$$

The maximum principle then implies that

$$\left(|\nabla P_t u|^2+\varepsilon\right)^{1/2}\leq\left(e^{2Kt}|\nabla u|^2+\varepsilon\right)^{1/2} .$$

The domination inequality follows by letting $\varepsilon\to 0$. ■

Theorem 8.1.16 *[S.-T. Yau] Let (M,g_{ij}) be a complete manifold of dimension n. Suppose that the Ricci curvature of (M,g_{ij}) is bounded below, that is, there is a constant K such that*

$$\Gamma_2(u)\geq\frac{1}{n}(\Delta u)^2+K\Gamma(u) \qquad \forall u\in C^2(M)\cap L^2(M,\mu_g)$$

then we have the following conclusions.

1) The heat semigroup $(P_t)_{t\geq 0}$ is conservative, that is,

$$\int_M h(t,x,y)\mu_g(dy)=1 \qquad \forall t>0 \text{ and } x\in M .$$

2) $(P_t)_{t\geq 0}$ is a Feller semigroup, i.e. $(P_t)_{t\geq 0}$ is a strongly continuous Markov semigroup on $C_0(M)$.

Proof. Let u_1 and u_2 be two arbitrary smooth functions on M having compact supports. Then

$$\begin{aligned}
\langle P_t u_1-u_1,u_2\rangle &= \int_0^t\langle\Delta P_s u_1,u_2\rangle ds\\
&= -\int_0^t\int_M(\nabla P_s u_1.\nabla u_2)\mathrm{d}\mu_g\mathrm{d}s\\
&\leq \int_0^t\int_M|\nabla P_s u_1||\nabla u_2|\mathrm{d}\mu_g\mathrm{d}s\\
&\leq \int_0^t\int_M \mathrm{e}^{Ks}P_s(|\nabla u_1|)|\nabla u_2|\mathrm{d}\mu_g\mathrm{d}s\\
&\leq ||\nabla u_1||_\infty||\nabla u_2||_1\int_0^t\mathrm{e}^{Ks}\mathrm{d}s
\end{aligned}$$

and it follows thus that

$$|\langle P_t u_1-u_1,u_2\rangle|\leq||\nabla u_1||_\infty||\nabla u_2||_1\int_0^t\mathrm{e}^{Ks}\mathrm{d}s .$$

Let h_n be a sequence constructed in Lemma 8.1.2, and apply the previous inequality to $u_1 = h_n$, and then let $n \to \infty$. We thus obtain

$$|\langle P_t 1 - 1, u_2 \rangle| = 0$$

for any $u_2 \in C_K^\infty(M)$. Therefore we must have

$$P_t 1(x) = \int_M h(t,x,y)\mu_g(\mathrm{d}y) = 1$$

for any $x \in M$. ■

8.1.5 Diffusion semigroups

Let $(M, \mathcal{M}, m)$ be a σ-finite measure space, and let $(P_t)_{t\geq 0}$ be a transition probability function which can be extended to be a strongly continuous (sub-)Markov semigroup on $L^2(M,m)$. Let $(L, D(L))$ be the infinitesimal generator of $(P_t)_{t\geq 0}$ (as strongly continuous contraction semigroup on $L^2(M,m)$), and let $\mathcal{A}$ be a *core* of $(L, D(L))$ which satisfies the following conditions:

1. $\mathcal{A}$ is a sub-algebra of $b\mathcal{M}$: $\mathcal{A}$ is a linear sub-space of $b\mathcal{M}$, and if $u, v \in \mathcal{A}$ then $uv \in \mathcal{A}$.

2. $\mathcal{A}$ is closed under L: $L(\mathcal{A}) \subset \mathcal{A}$.

3. $\mathcal{A}$ is closed under compositions with C^2 functions: if $u_1, \cdots, u_N \in \mathcal{A}$ and $F \in C^2(\mathbf{R}^N, \mathbf{R})$, then
$$F(u_1, \cdots, u_N) \in \mathcal{A} .$$

4. $\mathcal{A}$ is closed under P_t for every $t \geq 0$: $P_t(\mathcal{A}) \subset \mathcal{A}$.

Then we define $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ (metric operator) by

$$\Gamma(u,v) = \frac{1}{2}\{L(uv) - u(Lv) - v(Lu)\} \qquad \forall u, v \in \mathcal{A}$$

and $\Gamma_2 : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ (curvature operator) by

$$\Gamma_2(u,v) = \frac{1}{2}\{L(\Gamma(u,v)) - \Gamma(u, Lv) - \Gamma(v, Lu)\} \qquad \forall u, v \in \mathcal{A} .$$

Obviously both Γ and Γ_2 are bi-linear forms on $\mathcal{A}$ (and take values in $\mathcal{A}$ as well). For simplicity of notations, $\Gamma(u,u)$ (resp. $\Gamma_2(u,u)$) will be denoted by $\Gamma(u)$ (resp. $\Gamma_2(u)$).

Definition 8.1.17 *We say that $(P_t)_{t\geq 0}$ is a* diffusion semigroup *on M if Γ satisfies the chain rule (8.9): for every $F \in C^2(\mathbf{R}^N, \mathbf{R})$ and $u_1, \cdots, u_N, v \in \mathcal{A}$, we have $F(u) \in \mathcal{A}$ and*

$$\Gamma(F(u), v) = \sum_{j=1}^n \frac{\partial F}{\partial x^j}(u)\Gamma(u_j, v) \tag{8.12}$$

where $u = (u_1, \cdots, u_N)$ *and*

$$F(u) \equiv F \circ u = F(u_1, \cdots, u_N) \ .$$

In this case we also call the infinitesimal generator L a diffusion operator.

Remark 8.1.18 *The metric operator Γ, curvature operator Γ_2 and diffusion property (Definition 8.1.17) may be defined for a closable linear operator on a Banach space that possesses a core $\mathcal{A}$ satisfying conditions 1-3.*

In the remainder of this sub-section, we further assume that $(P_t)_{t\geq 0}$ is symmetric with respect to the σ-finite measure m, and we assume that L is a diffusion operator with a core $\mathcal{A}$ satisfying conditions 1-4.

Lemma 8.1.19 *If $u \in \mathcal{A}$ then*

$$\int_M (Lu)\,dm = 0 \ .$$

Proof. By the diffusion property (8.12), $1 \in \mathcal{A}$. Since L is self-adjoint, we have

$$\begin{aligned} \int_M (Lu)\mathrm{d}m &= \int_M u(L1)\mathrm{d}m \\ &= 0 \ . \end{aligned}$$

■

Lemma 8.1.20 *(Integration by parts) If $u, v \in \mathcal{A}$ then*

$$\begin{aligned} \int_M v(Lu)\,dm &= \int_M u(Lv)\,dm \\ &= -\int_M \Gamma(u,v)\,dm \ . \end{aligned} \tag{8.13}$$

The first equality holds for any $u, v \in D(L)$ and the second equality follows from the definition of Γ and the fact that $\int_M L(uv) = 0$.

It follows thus that the Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ associated with $(L, D(L))$ is the closure of the closable form

$$\mathcal{E}(u,v) = \int_M \Gamma(u,v)\mathrm{d}m \qquad \text{for } u, v \in \mathcal{A} \ . \tag{8.14}$$

Let $u \in \mathcal{A}$, and let $\phi(t,\cdot) = P_t u$. Then u solves the following heat equation

$$\left(L - \frac{\partial}{\partial t}\right)\phi = 0 \qquad \text{on } (0,+\infty) \times M \ .$$

Moreover

$$\left(L - \frac{\partial}{\partial t}\right)\phi^2 = 2\Gamma(\phi) \quad \text{on } (0,+\infty) \times M \tag{8.15}$$

and

$$\left(L-\frac{\partial}{\partial t}\right)\Gamma(\phi)=2\Gamma_2(\phi)\quad \text{on } (0,+\infty)\times M\ . \tag{8.16}$$

The integration by parts formula (8.13), basic equations (8.15, 8.16) explain why the metric operator Γ and the curvature operator Γ_2 are useful in the study of the diffusion semigroup $(P_t)_{t\geq 0}$.

Similarly, if we reverse the time at a fixed time $T>0$, and set $\psi(t,\cdot)=P_{T-t}u$, where $t\in[0,T]$, then

$$\left(L+\frac{\partial}{\partial t}\right)\psi^2=2\Gamma(\psi)$$

and

$$\left(L+\frac{\partial}{\partial t}\right)\Gamma(\psi)=2\Gamma_2(\psi)\ .$$

8.2 Contractivity of diffusion semigroups

From this section, we study Markov semigroups as strongly continuous contraction semigroups on L^p spaces. Unless otherwise stated, we consider a diffusion semigroup $(P_t)_{t\geq 0}$ on a σ-finite measure space $(M,\mathcal{M},m)$ which is symmetric with respect to m, with a core $\mathcal{A}$ satisfying conditions 1-4 in §8.1.5.

8.2.1 The integral maximum principle

Let $u\in D(L)$ and consider $f(t)=\int_M|P_tu|^2\mathrm{d}m$. Since

$$\begin{aligned} f'(t) &= 2\int_M (P_tu)(L(P_tu))\mathrm{d}m \\ &= -2\mathcal{E}(P_tu,P_tu)\leq 0\ , \end{aligned}$$

$t\to\int_M|P_tu|^2\mathrm{d}m$ is decreasing, a fact that $(P_t)_{t\geq 0}$ is a contraction semigroup on $L^2(M,m)$.

By the integral maximum principle we mean that the decreasing property is still true for a family of weighted norms.

Let us consider a function $\xi(x,t)$ and consider the following L^2-norm

$$f(t)=\int_M|P_tu|^2\mathrm{e}^{\xi(\cdot,t)}\mathrm{d}m\ .$$

Then, by integration by parts

$$\begin{aligned} f'(t) &= 2\int_M (L(P_tu))\,(P_tu)\mathrm{e}^{\xi}\mathrm{d}m+\int_M|P_tu|^2\frac{\partial\xi}{\partial t}\mathrm{e}^{\xi}\mathrm{d}m \\ &= -2\int_M\Gamma\left(P_tu,\mathrm{e}^{\xi}P_tu\right)\mathrm{d}m+\int_M|P_tu|^2\frac{\partial\xi}{\partial t}\mathrm{e}^{\xi}\mathrm{d}m\ . \end{aligned}\tag{8.17}$$

By diffusion property

$$\Gamma\left(\mathrm{e}^{\xi/2} w\right) = \Gamma(w)\,\mathrm{e}^{\xi} + \Gamma(w,\xi)\,w\mathrm{e}^{\xi} + \frac{1}{4}\Gamma(\xi)\,w^2\mathrm{e}^{\xi}$$

and

$$\Gamma\left(w, \mathrm{e}^{\xi} w\right) = \Gamma(w)\,\mathrm{e}^{\xi} + \Gamma(w,\xi)\,w\mathrm{e}^{\xi}.$$

We then have

$$\Gamma\left(w, \mathrm{e}^{\xi} w\right) = \Gamma\left(\mathrm{e}^{\xi/2} w\right) - \frac{1}{4}\Gamma(\xi)\,w^2\mathrm{e}^{\xi}\ .$$

Substituting the last equality into (8.17), we obtain

$$\begin{aligned} f'(t) &= -2\mathcal{E}\left(\mathrm{e}^{\xi/2}(P_t u), \mathrm{e}^{\xi/2}(P_t u)\right) \\ &\quad + \int_M \left\{\frac{\partial \xi}{\partial t} + \frac{1}{2}\Gamma(\xi)\right\}(P_t u)^2 \mathrm{e}^{\xi}\mathrm{d}m\ . \end{aligned} \tag{8.18}$$

We are in a position to show the following

Proposition 8.2.1 *Let $u \in L^2(M,m)$, and let $\{\xi(t) : 0 \le t < T\}$ be a family of functions on M such that $\mathrm{e}^{\xi/2}(P_t u) \in D(\mathcal{E})$ for every t and*

$$\frac{\partial \xi}{\partial t} + \frac{1}{2}\Gamma(\xi) \le -2\delta \tag{8.19}$$

where δ is a constant. Then

$$\frac{d}{dt} f(t) \le -2\delta f(t) \qquad \text{on} \quad [0,T)$$

where $f(t) = \int_M |P_t u|^2\ e^{\xi}\, dm$. In particular

$$\int_M |P_t u|^2 e^{\xi(t)}\, dm \le e^{-2\delta t} \int_M |u|^2 e^{\xi(0)}\, dm \qquad \forall t < T\ . \tag{8.20}$$

Proof. By (8.18)

$$\begin{aligned} f'(t) &\le \int_M \left\{\frac{\partial \xi}{\partial t} + \frac{1}{2}\Gamma(\xi)\right\}(P_t u)^2 \mathrm{e}^{\xi}\mathrm{d}m \\ &\le -2\delta \int_M (P_t u)^2 \mathrm{e}^{\xi}\mathrm{d}m \\ &= -2\delta f(t)\ . \end{aligned}$$

■

Observe in the above proof, we have only used the positivity of the first term on right-hand side of (8.18), and therefore there is a possibility of improving the contractivity of $(P_t)_{t\ge 0}$.

Definition 8.2.2 *(The Poincaré inequality) Let $(\mathcal{E}, D(\mathcal{E}))$ be a Dirichlet form. The spectral gap λ_1 for $(\mathcal{E}, D(\mathcal{E}))$ is the largest constant such that*

$$\lambda_1 ||u||^2_{L^2(m)} \leq \mathcal{E}(u,u) \qquad \forall u \in D(\mathcal{E}) \ . \tag{8.21}$$

λ_1 is also called the spectral gap of the infinitesimal generator L of the Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$. Obviously $\lambda_1 \geq 0$. If $\lambda_1 > 0$, then we say that there is a spectral gap for the self-adjoint operator L.

Let us return to (8.18). By definition of the spectral gap, we deduce that

$$f'(t) \leq -2\lambda_1 f(t) + \int_M \left\{ \frac{\partial \xi}{\partial t} + \frac{1}{2}\Gamma(\xi) \right\} (P_t u)^2 \mathrm{e}^{\xi} \mathrm{d}m \ . \tag{8.22}$$

Therefore

Theorem 8.2.3 *(The integral maximum principle) Let ξ be a function satisfying (8.19) for some constant δ. Then*

$$\int_M |P_t u|^2 e^{\xi(t)} dm \ \leq e^{-2(\lambda_1+\delta)t} \int_M |u|^2 e^{\xi(0)} dm \tag{8.23}$$

where $\lambda_1 \geq 0$ is the spectral gap of $-L$, L is the infinitesimal generator of $(P_t)_{t\geq 0}$.

Two basic examples

As far as authors, there are only two examples of ξ satisfying (8.19), which we will explain below.

Example1.

Let ρ be a Lipschitz function on M. That is, $\rho \in \mathcal{A}_{\mathrm{loc}}$ such that $\Gamma(\rho) \leq 1$. Let $T > 0$ be a fixed constant, and set

$$\xi(t) = -\frac{\rho^2}{2(T-t)} \qquad \forall t < T \ . \tag{8.24}$$

Then

$$\Gamma(\xi) = \frac{\rho^2}{(T-t)^2}\Gamma(\rho) \ ; \qquad \frac{\partial}{\partial t}\xi = -\frac{\rho^2}{2(T-t)^2}$$

so that

$$\frac{\partial \xi}{\partial t} + \frac{1}{2}\Gamma(\xi) \leq 0 \ .$$

Thus by (8.23)

Corollary 8.2.4 *Let $P_t = e^{tL}$ and let λ_1 be the spectral gap of $-L$. Let ρ be a Lipschitz function on M: $\Gamma(\rho) \leq 1$, and let $T > 0$. Then*

$$\int_M |P_t u|^2 e^{-\frac{\rho^2}{2(T-t)}} dm \ \leq e^{-2\lambda_1 t} \int_M |u|^2 e^{-\frac{\rho^2}{2T}} dm \qquad \forall t < T \ . \tag{8.25}$$

Example 2

Again, let ρ be a Lipschitz function: $\rho \in \mathcal{A}_{\rm loc}$ such that $\Gamma(\rho) \le 1$. Let $\alpha \in \mathbb{R}$ and consider

$$\xi(t) = \alpha\rho - \frac{\alpha^2}{2}t \ .$$

Then

$$\Gamma(\xi) = \alpha^2\Gamma(\rho) \ ; \quad \frac{\partial}{\partial t}\xi = -\frac{\alpha^2}{2}$$

hence

$$\frac{\partial\xi}{\partial t} + \frac{1}{2}\Gamma(\xi) \le 0 \ .$$

Corollary 8.2.5 *Let $P_t = e^{tL}$ and let λ_1 be the spectral gap of $-L$. Let ρ be a Lipschitz function on M: $\Gamma(\rho) \le 1$, $\alpha \in \mathbb{R}$ and define measure μ by $d\mu = e^{\alpha\rho}\,dm$. Then*

$$\int_M |P_t u|^2\,d\mu \ \le e^{\frac{\alpha^2}{2}t - 2\lambda_1 t}\int_M |u|^2\,d\mu \qquad \forall t \ge 0 \ . \tag{8.26}$$

Contractivity on L^p-spaces

There is a version of the integral maximum principle on $L^p(M,m)$.

Proposition 8.2.6 *Let $p > 1$ and set $k = \frac{p-1}{p}$. If ξ is a function satisfying the following inequality:*

$$\frac{\partial}{\partial t}\xi + \frac{1}{4k}\Gamma(\xi) \le 0 \qquad \forall t < T \tag{8.27}$$

then

$$\int_M |P_t u|^p\,e^{\xi(t)}\,dm \le \int_M |u|^p\,e^{\xi(0)}\,dm \qquad 0 \le t < T \ . \tag{8.28}$$

Proof. The proof is similar to the case that $p = 2$. We may assume that $u \ge 0$, and consider the function

$$f(t) = \int_M (P_t u)^p \mathrm{e}^{\xi(t)}\mathrm{d}m \ .$$

Then

$$\begin{aligned}
f'(t) &= \int_M p((P_t u)^{p-1}\mathrm{e}^{\xi})L(P_t u)\mathrm{d}m + \int_M \xi'(t)(P_t u)^p\mathrm{e}^{\xi}\mathrm{d}m \\
&= -p\int_M \Gamma((P_t u)^{p-1}\mathrm{e}^{\xi}, P_t u)\mathrm{d}m + \int_M (P_t u)^p\xi'(t)\mathrm{e}^{\xi}\mathrm{d}m \\
&= -p\int_M \Gamma((P_t u)^{p-1}, P_t u)\mathrm{e}^{\xi}\mathrm{d}m - p\int_M (P_t u)^{p-1}\Gamma(\mathrm{e}^{\xi}, P_t u)\mathrm{d}m \\
&\quad + \int (P_t u)^p\xi'(t)\mathrm{e}^{\xi}\mathrm{d}m \\
&= -p(p-1)\int_M e^{\xi}(P_t u)^{p-2}\Gamma(P_t u)\mathrm{e}^{\xi}\mathrm{d}m \\
&\quad -p\int_M (P_t u)^{p-1}\Gamma(\xi, P_t u)\mathrm{e}^{\xi}\mathrm{d}m + \int_M (P_t u)^p\xi'(t)\mathrm{e}^{\xi}\mathrm{d}m \ .
\end{aligned}$$

However,

$$-(P_t u)\Gamma(\xi, P_t u) \le (p-1)\Gamma(P_t u) + \frac{1}{4(p-1)}\Gamma(\xi)(P_t u)^2$$

and therefore by inserting this inequality into $f'(t)$, we get that

$$\begin{aligned} f'(t) &\le \frac{p}{4(p-1)}\int_M (P_t u)^p \Gamma(\xi) e^{\xi} dm + \int_M (P_t u)^p \xi'(t) e^{\xi} dm \\ &\le 0 . \end{aligned}$$

■

8.2.2 Universal Gaussian upper bound

In this sub-section we deduce a universal Gaussian upper bound for a symmetric Markov semigroups. Except those conditions we made in Section ??? we further assume that M is a metric space with metric d, and m is a Radon measure on $(M, \mathcal{B}(M))$. Let A and B be two Borel subsets of M. We make the following technical assumptions.

1. $\rho(x) = d(x, A)$ belongs to $\mathcal{A}_{\rm loc}$ and $\Gamma(\rho) \le 1$, where $d(x, A) = \inf\{d(x, z) : z \in A\}$.
2. $d(A, B) = \inf\{d(x, A) : x \in B\} < +\infty$.

Under these conditions, we may apply (8.26)

$$\int_M |P_t u|^2 e^{\alpha\rho} dm \le e^{\frac{\alpha^2}{2}t - 2\lambda_1 t}\int_M |u|^2 e^{\alpha\rho} dm \qquad \forall t \ge 0 \text{ and } \alpha \in \mathbb{R}$$

to $u = 1_A$ and ρ, and obtain

$$\begin{aligned} \int_M |P_t 1_A|^2 e^{\alpha\rho} dm &\le e^{\frac{\alpha^2}{2}t - 2\lambda_1 t}\int_A e^{\alpha\rho} dm \\ &= e^{\frac{\alpha^2}{2}t - 2\lambda_1 t} m(A) \end{aligned}$$

for any $\alpha \in \mathbb{R}$. In particular

$$\int_B |P_t 1_A|^2 e^{\alpha\rho} dm \le e^{\frac{\alpha^2}{2}t - 2\lambda_1 t} m(A) .$$

While $\rho \ge d(A, B)$ on B, we have

$$\int_B |P_t 1_A|^2 e^{\alpha d(A,B)} dm \le e^{\frac{\alpha^2}{2}t - 2\lambda_1 t} m(A)$$

that is

$$\int_B |P_t 1_A|^2 d\mu \le e^{\frac{\alpha^2}{2}t - \alpha d(A,B) - 2\lambda_1 t} m(A) . \tag{8.29}$$

By the Schwartz inequality

$$\begin{aligned}\langle P_t 1_A, 1_B\rangle &\leq \sqrt{\int_B |P_t 1_A|^2 \mathrm{d}m}\sqrt{\int_B \mathrm{d}m} \\ &\leq \sqrt{m(A)m(B)}\mathrm{e}^{\frac{t}{4}\alpha^2 - \frac{d(A,B)}{2}\alpha - \lambda_1 t}\end{aligned}$$

and by choosing $\alpha = d(A,B)/t$ we thus obtain the following inequality

$$\langle P_t 1_A, 1_B\rangle \leq \sqrt{m(A)m(B)}\mathrm{e}^{-\frac{d(A,B)^2}{4t} - \lambda_1 t} .$$

Theorem 8.2.7 (B. Davies, A. Grigor'yan) *The universal Gaussian bound holds for symmetric Markov semigroup $(P_t)_{t\geq 0}$:*

$$\langle P_t 1_A, 1_B\rangle \leq \sqrt{m(A)m(B)}e^{-\frac{d(A,B)^2}{4t} - \lambda_1 t}$$

for $A, B \in B(M)$ such that $d(A,B) < +\infty$.

If P_t possesses a transition density function $p(t,x,y)$, then the previous Gaussian bound may be written as

$$\int_A \int_B p(t,x,y)m(\mathrm{d}x)m(\mathrm{d}y) \leq \sqrt{m(A)m(B)}\mathrm{e}^{-\frac{d(A,B)^2}{4t} - \lambda_1 t} . \tag{8.30}$$

8.3 Hypercontractivity

We have seen that, a symmetric Markov semigroup $(P_t)_{t\geq 0}$ on $(M, \mathcal{M}, m)$ can be considered as a contraction semigroup on $L^p(M,m)$ for every $p \geq 1$. In this sub-section, we consider P_t as a linear operator from $L^p \to L^q$ for p, $q \geq 1$.

8.3.1 Logarithmic Sobolev inequality

Entropy

If μ is a probability measure on $(M, \mathcal{M})$, then the entropy Ent_μ associated with the probability measure μ is defined as

$$\mathrm{Ent}_\mu(u) = \int_M |u| \log |u| \mathrm{d}\mu - ||u||_{L^1(\mu)} \log ||u||_{L^1(\mu)} \qquad \forall u \in L^1(\mu) . \tag{8.31}$$

Let us state a few elementary properties of the entropy functional Ent_μ.

1. If ν is another probability measure that is absolutely continuous with respect to μ, then

$$\begin{aligned}\mathrm{Ent}_\mu(\frac{d\nu}{d\mu}) &= \int_M \frac{d\nu}{d\mu} \log\left(\frac{d\nu}{d\mu}\right) \mathrm{d}\mu \\ &= \int_M \log\left(\frac{d\nu}{d\mu}\right) \mathrm{d}\nu\end{aligned}$$

which is the relative entropy of ν with respect to μ.

2. Function $\Phi(x) = x \log x$ is continuous and convex on $[0, +\infty)$:

$$\lim_{x \downarrow 0} x \log x = 0, \quad \Phi'(x) = \log x + 1 \quad \text{and} \quad \Phi''(x) = \frac{1}{x} \ ;$$

   Φ achieves its minimal value 0 at e^{-1}. By Jensen's inequality

$$\Phi\left(\int_M |u| \mathrm{d}\mu\right) \leq \int_M \Phi\left(|u|\right) \mathrm{d}\mu \ .$$

   That is, $\mathrm{Ent}_\mu(u) \geq 0$.

3. $\mathrm{Ent}_\mu(u)$ is lower semi-continuous on $L^1(\mu)$, that is, if $u_n \to u$ in $L^1(\mu)$ then

$$\mathrm{Ent}_\mu(u) \leq \underline{\lim}_{n\to\infty} \mathrm{Ent}_\mu(u_n) \ .$$

Entropy type inequalities

Let $(P_t)_{t\geq 0}$ be a symmetric Markov semigroup on $(M, \mathcal{M}, m)$, where m is a *probability measure*. The Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ associated with $(P_t)_{t\geq 0}$ is given by

$$\begin{aligned}
\mathcal{E}(u,u) &= \lim_{t\downarrow 0} \frac{1}{t}\left(||u||^2_{L^2(m)} - ||P_t u||^2_{L^2(m)}\right) \\
&= \lim_{t\downarrow 0} \frac{1}{2t} \int_{M\times M} |u(x) - u(y)|^2 P_t(x, \mathrm{d}y) m(\mathrm{d}x) \\
&= \sup_{t>0} \frac{1}{2t} \int_{M\times M} (u(x) - u(y))^2 P_t(x, \mathrm{d}y) m(\mathrm{d}x)
\end{aligned}$$

and

$$D(\mathcal{E}) = \left\{u \in L^2(m) : \mathcal{E}(u,u) < +\infty\right\} \ .$$

L. Gross (Cornell University) has introduced an infinite dimensional version of the Sobolev inequality, called a logarithmic Sobolev inequality. A class of functional inequalities associated with a Dirichlet form (under a general name entropy type inequalities) have been playing an important role in the study of Markov semigroups and their applications in manifold fields. These inequalities can be in general written as

$$\mathrm{Ent}(u^2) \leq \Phi\left(\mathcal{E}(u,u)\right)$$

for some concave function. We however consider a very simple case, called the logarithmic Sobolev inequality:

$$\mathrm{Ent}(u^2) \leq a\mathcal{E}(u,u) + b||u||_2^2 \qquad \forall u \in \mathcal{F} \ . \tag{8.32}$$

8.3.2 Ornstein-Uhlenbeck semigroup

Let σ and α be two positive constants, and consider the the following stochastic differential equation

$$dX_t = \sigma \mathrm{d}W_t - \alpha X_t \mathrm{d}t$$

in $\mathbb{R}^n$, where $(W_t)_{t\geq 0}$ is a standard n-dimensional Brownian motion. Its solution $(X_t)_{t\geq 0}$ is called the Ornstein-Uhlenbeck process. By the use of Itô's formula we may show that its solution is given by

$$X_t = x\mathrm{e}^{-\alpha t} + \sigma \int_0^t \mathrm{e}^{-\alpha(t-s)} \mathrm{d}W_s \ .$$

Thus $(X_t)_{t\geq 0}$ is a Gaussian process and has a normal distribution with mean $\mathrm{e}^{-\alpha t}x$ and variance matrix

$$\frac{\sigma^2}{2} \frac{1-\mathrm{e}^{-2\alpha t}}{\alpha} I$$

therefore

$$\begin{aligned} (P_t f)(x) &= E^x f(X_t) = E^x f\left(x\mathrm{e}^{-\alpha t} + \sigma \int_0^t \mathrm{e}^{-\alpha(t-s)} \mathrm{d}W_s \right) \\ &= E\left\{ f\left(x\mathrm{e}^{-\alpha t} + \sqrt{\frac{\sigma^2}{2} \frac{1-\mathrm{e}^{-2\alpha t}}{\alpha}} \xi \right) \right\} \end{aligned}$$

where $\xi = (\xi^i)$ are independent Gaussian variables with common distribution $N(0,1)$.

By Itô's formula

$$\begin{aligned} f(X_t) - f(X_0) &= \sigma \int_0^t \nabla f(X_s) \mathrm{d}W_t \\ &\quad + \int_0^t \left\{ \frac{\sigma^2}{2} (\Delta f)(X_s) - \alpha X_s.(\nabla f)(X_s) \right\} \mathrm{d}s \end{aligned}$$

and therefore the infinitesimal generator of $(P_t)_{t\geq 0}$ is

$$\begin{aligned} L &= \frac{\sigma^2}{2} \Delta - \alpha x.\nabla \\ &= \frac{\sigma^2}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} - \alpha \sum_{i=1}^n x_i \frac{\partial}{\partial x_i} \ . \end{aligned}$$

L is called the Ornstein-Uhlenbeck operator and $P_t = e^{tL}$ the Ornstein-Uhlenbeck semigroup. The square field operator associated with L is given by

$$\Gamma(u) = \frac{1}{2} L u^2 - uLu = \frac{\sigma^2}{2} |\nabla u|^2 \ .$$

$(P_t)_{t\geq 0}$ is symmetric with respect to the Gaussian measure

$$m(\mathrm{d}x) = \frac{\alpha^{n/2}}{(\pi\sigma^2)^{n/2}} \exp\left(-\frac{\alpha |x|^2}{\sigma^2} \right)$$

and, by integration by parts,

$$\int_{\mathbf{R}^n} \Gamma(u)\mathrm{d}m = -\int_{\mathbf{R}^n} u(Lu)\mathrm{d}m \ .$$

Therefore the Dirichlet form associated with $(P_t)_{t\geq 0}$ is

$$\mathcal{E}(u,u) = \int_{\mathbf{R}^n} \Gamma(u)\mathrm{d}m = \frac{\sigma^2}{2}\int_{\mathbb{R}^d} |\nabla u|^2 \mathrm{d}m \ .$$

For bounded continuous function f

$$f\left(x\mathrm{e}^{-\alpha t} + \sqrt{\frac{\sigma^2}{2}\frac{1-\mathrm{e}^{-2\alpha t}}{\alpha}}\xi\right) \to E\left\{f\left(\sqrt{\frac{\sigma^2}{2\alpha}}\xi\right)\right\} \quad \text{as } t\to\infty \ .$$

Since $\sqrt{\frac{\sigma^2}{2\alpha}}\xi$ has a normal distribution $N(0,\frac{\sigma^2}{2\alpha})$, therefore

Lemma 8.3.1 *Let $(P_t)_{t\geq 0}$ be the Ornstein-Uhlenbeck semigroup with infinitesimal generator*

$$L = \frac{\sigma^2}{2}\Delta - \alpha x.\nabla \ .$$

Then

1) For any $x \in \mathbb{R}^n$, $t > 0$ and $f \in L^1(\mathbb{R}^n, m)$

$$(P_t f)(x) = E\left\{f\left(xe^{-\alpha t} + \sqrt{\frac{\sigma^2}{2}\frac{1-e^{-2\alpha t}}{\alpha}}\xi\right)\right\}$$

where $\xi \sim N(0, I)$.

2) If $f \in L^1(\mathbb{R}^n, m)$, then

$$P_t f \to \int_{\mathbb{R}^n} f(x)m(dx) \qquad \text{as } t \to \infty$$

almost surely.

3) We have

$$\frac{\partial}{\partial x^i}P_t f = e^{-\alpha t}P_t\left(\frac{\partial f}{\partial x^i}\right)$$

so that

$$\sqrt{\Gamma(P_t f)} \leq e^{-\alpha t}P_t\left(\sqrt{\Gamma(f)}\right)$$

for all $t > 0$.

Gross' log-Sobolev inequality

Let us first show the following local entropy-energy inequality.

Lemma 8.3.2 *Under the same assumptions as in Lemma 8.3.1. For any $t > 0$*

$$P_t(u^2 \log u^2) - P_t(u^2) \log P_t(u^2) \le \frac{2(1 - e^{-2\alpha t})}{\alpha} P_t(\Gamma(u)) \ . \tag{8.33}$$

Proof. Consider the following function

$$f(s) = P_s(\Phi(P_{t-s}v))$$

where $\Phi(x) = x \log x$, and v is a positive function. Therefore $f(0) = P_t(v) \log P_t(v)$ and $f(t) = P_t(v \log v)$. By the chain rule we have

$$\begin{aligned} f'(s) &= LP_s(\Phi(P_{t-s}v)) + P_s(\frac{d}{ds}\Phi(P_{t-s}v)) \\ &= P_s\left\{L(\Phi(P_{t-s}v)) + \frac{d}{ds}\Phi(P_{t-s}v)\right\} \\ &= P_s\left\{\Phi'(P_{t-s}v)\left(L + \frac{d}{ds}\right)P_{t-s}v + \Phi''(P_{t-s}v)\Gamma(P_{t-s}v)\right\} \\ &= P_s\left\{\Phi''(P_{t-s}v)\Gamma(P_{t-s}v)\right\} \ . \end{aligned}$$

while in our case $\Phi''(x) = \frac{1}{x}$ so that

$$f'(s) = P_s\left\{\frac{\Gamma(P_{t-s}v)}{P_{t-s}v}\right\} \ .$$

By Lemma 8.3.1

$$\Gamma(P_{t-s}v) \le \mathrm{e}^{-2\alpha(t-s)}\left(P_{t-s}(\sqrt{\Gamma(v)})\right)^2 ,$$

so that

$$f'(s) \le \mathrm{e}^{-2\alpha(t-s)} P_s\left\{\frac{\left(P_{t-s}(\sqrt{\Gamma(v)})\right)^2}{P_{t-s}v}\right\} \ . \tag{8.34}$$

Furthermore, since P_{t-s} is a probability measure, applying Hölder inequality we obtain

$$\begin{aligned} (P_{t-s}\sqrt{\Gamma(v)})^2 &= (P_{t-s}\frac{\sqrt{\Gamma(v)}}{\sqrt{v}}\sqrt{v})^2 \\ &\le P_{t-s}\left(\frac{\Gamma(v)}{v}\right)P_{t-s}v \end{aligned}$$

that is, by dividing both sides by $P_{t-s}v$,

$$\frac{(P_{t-s}\sqrt{\Gamma(v)})^2}{P_{t-s}v} \le P_{t-s}\left(\frac{\Gamma(v)}{v}\right) \ .$$

Applying this inequality to (**??**) to establish

$$\begin{aligned} f'(s) &\leq \mathrm{e}^{-2K(t-s)} P_s \left\{ P_{t-s} \left(\frac{\Gamma(v)}{v} \right) \right\} \\ &= \mathrm{e}^{-2K(t-s)} P_t \left\{ \frac{\Gamma(v)}{v} \right\} . \end{aligned}$$

It remains to integrate f on $[0,t]$ and to set $v = u^2$. ■

Theorem 8.3.3 (L. Gross, 1975) *Under the same assumptions as in Lemma 8.3.1, it holds that*

$$\int_{\mathbf{R}^n} u^2 (\log u^2)\, dm - ||u||^2 \log ||u||^2 \leq \frac{\sigma^2}{\alpha} \int_{\mathbf{R}^n} |\nabla u|^2\, dm \tag{8.35}$$

where $||u||^2 = \int_{\mathbf{R}^n} u^2\, dm$. *In particular, for standard gaussian measure* μ ($\frac{\sigma^2}{\alpha} = 2$) *we have*

$$\int_{\mathbf{R}^n} u^2 (\log u^2)\, d\mu - ||u||^2_{L^2(\mu)} \log ||u||^2_{L^2(\mu)} \leq 2 \int_{\mathbf{R}^n} |\nabla u|^2\, dm \ . \tag{8.36}$$

8.3.3 Gross' theorem on hypercontractivity

Let us consider the following function

$$F(t) = \mathrm{e}^{-m(t)} ||P_t u||_p = \mathrm{e}^{-m(t)} \left(\int |P_t u|^{p(t)} \right)^{1/p(t)}$$

where $m(t)$ and $p(t)$ are two functions satisfying $m(0) = 1$ and $p(t) > 1$ with $p(0) = p$ should be increasing. Both functions will be chosen so that $t \to F(t)$ is decreasing. If it is the case, then $F(t) \leq F(0)$, that is,

$$||P_t u||_{p(t)} \leq \mathrm{e}^{m(t)} ||u||_p \ . \tag{8.37}$$

That is, P_t is a bounded operator from $L^p(\mathbf{R}^d, \mu)$ to $L^q(\mathbf{R}^d, \mu)$ if $p, q = p(t)$ and t are properly related:

$$p(t) = 1 + (p-1)e^{4t/a}; \quad m(t) = b\left(p^{-1} - p(t)^{-1}\right) \tag{8.38}$$

where $a > 0$ and $b \geq 0$ are two constants. The property (8.37) is called hypercontractivity.

Gross' theorem

A straightforward computation shows that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\log F(t) &= -m'(t) + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{p(t)}\log\int |P_t u|^{p(t)}\right) \\
&= -m'(t) - \frac{p'(t)}{p(t)^2}\log\int |P_t u|^{p(t)} + \frac{1}{p(t)}\frac{1}{|P_t u|^{p(t)}}\int\left(\frac{\mathrm{d}}{\mathrm{d}t}|P_t u|^{p(t)}\right) \\
&= -m'(t) - \frac{p'(t)}{p(t)^2}\log\int |P_t u|^{p(t)} \\
&\quad + \frac{1}{p(t)}\frac{1}{|P_t u|^{p(t)}}\int\left(\frac{p(t)}{2}|P_t u|^{\frac{p(t)}{2}-1}\frac{\mathrm{d}}{\mathrm{d}t}|P_t u|^2 + p'(t)|P_t u|^{p(t)}\log|P_t u|\right) .
\end{aligned}$$

After simplification we then establish

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\log F(t) &= -m'(t) + \frac{p'(t)}{p(t)^2}\frac{1}{||P_t u||_{p(t)}^{p(t)}}\mathrm{Ent}_{p(t)}\left(|P_t u|^{p(t)}\right) \\
&\quad + \frac{1}{2||P_t u||_{p(t)}^{p(t)}}\int_M |P_t u|^{p(t)-2}\frac{\mathrm{d}}{\mathrm{d}t}|P_t u|^2 \, . \qquad (8.39)
\end{aligned}$$

Multiplying bote sides by $||P_t u||_{p(t)}^{p(t)}$ we then obtain

$$\frac{\frac{d}{dt}\log F(t)}{||P_t\varphi||_p^p} = \frac{p'(t)}{p(t)^2}\left\{\mathrm{Ent}_p\left(|P_t u|\right) + \frac{p(t)^2}{2p'(t)}\int_M |P_t u|^{p-2}\frac{d|P_t\varphi|^2}{dt} - \frac{m'(t)p(t)^2}{p'(t)}||P_t\varphi||_p^p\right\} \qquad (8.40)$$

where for simplicity we remove t from $p(t)$.

On the other hand, by integration by parts

$$\begin{aligned}
\int_M |P_t u|^{p-2}\frac{d|P_t\varphi|^2}{dt} &= 2\int_M |P_t u|^{p-2}(P_t u)L(P_t u) \\
&= -2\int_M \nabla\left\{|P_t u|^{p-2}(P_t u)\right\}.\nabla(P_t u) \\
&= -2\int_M \nabla(P_t u)^{p-1}.\nabla(P_t u) \\
&= -2\frac{4(p-1)}{p^2}\int_M |\nabla(P_t u)^{p/2}|^2
\end{aligned}$$

so that

$$\frac{\frac{d}{dt}\log F(t)}{||P_t u||_p^p} = \frac{p'(t)}{p(t)^2}\left\{\mathrm{Ent}_p\left(|P_t u|\right) - \frac{4(p(t)-1)}{p'(t)}\int_M |\nabla(P_t u)^{p/2}|^2 - \frac{m'(t)p(t)^2}{p'(t)}||P_t u||_p^p\right\} . \qquad (8.41)$$

Choose $p(t)$ and $m(t)$ by solving the following differential equations:

$$\frac{4(p(t)-1)}{p'(t)} = a \; , \quad p(0) = p$$

and

$$\frac{m'(t)p(t)^2}{p'(t)} = b \ , \qquad m(0) = 0 \ ,$$

where a, b and the initial value $p(0) = p$ are given. That is,

$$p(t) = 1 + (p-1)e^{\frac{4}{a}t} \ , \qquad m(t) = b\left(p^{-1} - p(t)^{-1}\right) \ . \tag{8.42}$$

We thus see, for these functions $p(t)$ and $m(t)$,

$$\frac{\frac{d}{dt}\log F(t)}{||P_t u||_p^p} = \frac{p'(t)}{p(t)^2}\left\{\mathrm{Ent}_p\left(|P_t u|\right) - a\int_M |\nabla(P_t u)^{p/2}|^2 - b||P_t u||_p^p\right\}$$

and therefore $F' \leq 0$ if and only if

$$\mathrm{Ent}_p\left(|P_t u|\right) \leq a\int_M |\nabla(P_t u)^{p/2}|^2 + b||P_t u||_p^p \ .$$

Hence we have proved the following

Theorem 8.3.4 *The hypercontractivity (???) of $(P_t)_{t\geq 0}$ is equivalent to the following log-Sobolev inequality*

$$Ent\left(|u|^2\right) \leq a\mathcal{E}(u,u) + b||u||_2^2 \ . \tag{8.43}$$

Spectral properties of Markov semigroups, convergence, ergodic theorem, Poincaré inequality, log-Sobolev inequality and hyper-contractivity, Nash inequality, and decay of semigroups

Bibliography

[1] D. Bakry, Un critère de non-explosion pour certaines diffusions sur une variété riemannienne complète, C. R. Acad. Sc. Paris, t. 303, Série I, n°1, 1986, pages 23-26.

[2] R. M. Blumenthal and R. K. Getoor, Markov Processes and Potential Theory, Academic Press, New York (1968).

[3] K. L. Chung, Lectures from Markov Processes to Brownian Motion. Springer-Verlag, New York (1982).

[4] E. B. Davies, One-parameter semigroups, Academic Press, 1980.

[5] C. Dellacherie and P. A. Meyer, Probabilités et potentiel, Chapitres V-VIII, Hermann (1980).

[6] C. Dellacherie and P. A. Meyer, Probabilités et potentiel, Chapitres XII à XVI, Hermann (1987).

[7] J. L. Doob, Stochastic Processes. Wiley, New-York (1953).

[8] P.J. Fitzsimmons and R.K. Getoor, On the potential theory of symmetric Markov processes, Math Ann 281(1988), no3, 495-512

[9] M. Fukushim, Dirichlet forms and Markov processes. North Holland, Kodansha (1980).

[10] K. Itô, and H.P. Mckean Jr., Diffusion Processes and Their Sample Paths, Sringer-Verlag, 1965

[11] T. Lyons and W. Zheng, A crossing estimate for the canonical process on a Dirichlet space and a tightness result, in “Colloque Paul Lévy sur les processus stochastiques”, Astérisque 157-158, pages 249-271 (1988).

[12] J. Neveu, Martingales à temps discret. Masson, Paris (1972).

[13] A. A. Novikov, On moment inequalities and identities for stochastic integrals, Proc. second Japan-USSR Symp. Prob. Theor., Lecture Notes in Math., 330, 333-339, Springer-Verlag, Berlin 1973

[14] Revuz, D., Yor, M., Continuous Martingales and Brownian Motion, Springer, 1991

[15] L. C. G. Rogers and D. Williams, Diffusions, Markov Processes and Martingales, Volume 1 Foundations, and Volume 2 Itô Calculus, Cambridge Mathematical Library. Cambridge University Press (2000).

[16] M. L. Silverstein, Symmetric Markov Processes, Lecture Notes in Mathematics 426, Springer-Verlag (1974).

[17] M. Sharpe, General Theory of Markov Processes, Academic Press, INC. (1988).

[18] D. W. Stroock, An Introduction to the Theory of Large Deviations, Springer-Verlag (1984).

[19] D. W. Stroock, Probability Theory: An Analytic View,

[20] D. Revuz and M. Yor, Continuous Martingales and Brownian Motion, Grundlehren der mathematischen Wissenschaften 293, Springer-Verlag (1991).

[21] D. Williams, Probability with Martingales, .

[22] N. Wiener, Differential space", J. Math. Phys. 2, 132-174

[23] J. A. Yan, Critères d'intégrabilité uniforme des martingales exponentielles, Acta. Math. Sinica 23, 311-318 (1980)

[24] K. Yosida, Functional Analysis, Springer-Verlag, 1978.