分类号 ____________　　　　　　　　　　　　　　　　编号 ____________

U D C ____________　　　　　　　　　　　　　　　　密级 ____________

南方科技大学

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题　　目：单调型随机常微分方程的遍历性分析

姓　　名：刘之洲

学　　号：11910524

系　　别：数学系

专　　业：数学与应用数学

指导教师：刘智慧 副教授

2023 年 5 月 7 日

CLC ___________ Number___________

UDC___________ Available for reference ☐Yes ☐No

Southern University of Science and Technology

Undergraduate Thesis

Thesis Title: **Analysis on Ergodicity of Monotone SODEs**

Student Name: **Zhizhou Liu**

Student ID: **11910524**

Department: **Department of Mathematics**

Program: **Mathematics and Applied Mathematics**

Thesis Advisor: **Assoicate Professor Zhihui Liu**

Date: May 7, 2023

诚信承诺书

1. 本人郑重承诺所呈交的毕业设计（论文），是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。

2. 除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。

3. 本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文（设计）中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

作者签名:

________年____月____日

COMMITMENT OF HONESTY

1. I solemnly promise that the paper presented comes from my independent research work under my supervisor's supervision. All statistics and images are real and reliable.

2. Except for the annotated reference, the paper contents no other published work or achievement by person or group. All people making important contributions to the study of the paper have been indicated clearly in the paper.

3. I promise that I did not plagiarize other people's research achievement or forge related data in the process of designing topic and research content.

4. If there is violation of any intellectual property right, I will take legal responsibility myself.

Signature:

Date:

单调型随机常微分方程的遍历性分析

刘之洲

（数学系　指导教师：刘智慧）

[摘要]：本文的主要目的是研究在非 Lipshcitz 条件下随机常微分方程（SODE）的性质，特别是其遍历性。为此，本文第二节从概率论的基本概念 Markov 核和 Markov 半群出发，并简要介绍了随机微积分（Itô 积分）。另一方面，我们还需关于给定 Markov 半群下遍历测度的一般理论，这是本文第三节的主要内容。在第四部分中，我们对一类系数非 Lipshitz 条件的 SODE 进行了分析，证明了其解的存在唯一性、齐时性、Markov 性和半群性质。最后，我们通过分别证明强 Feller 性和不可约性，我们得到了不变测度的唯一性；利用第三节给出的 Krylov-Bogoliubov 定理的应用证明了不变测度的存在性；由 Doob 定理，这说明了解的遍历性。文中出现的所有结果均非原创。文章的价值在于其系统地研究了单调型随机常微分的方程的遍历性，为未来进一步研究该类型的方程提供了参考。

[关键词]：不变测度；单调随机微分方程；非 Lipshcitz 条件；半群性质；遍历性

[**ABSTRACT**]: The aim of this paper is to investigate the properties, especially ergodicity, of Stochastic Ordinary Differential Equations (SODEs) under non-Lipschitz conditions of coefficients. To achieve this, we start from the basic concept in probability theory such as Markov kernel and Markov semigroup, and briefly illustrate the ideas of stochastic calculus (Itô's integral). On the other hand, we also need a general theory for finding ergodic measures of a given Markov semigroup, which is the content of the third section. In Section 4, we analyze a class of non-Lipschitz SODEs and prove the existence, uniqueness, homogenity, Markov and semigroup properties of their solutions. Finally, we prove the uniqueness of the invariant measure through showing that it is strong Feller and irreducible; and prove the existence of invariant measure utilizing Krylov-Bogoliubov theorem. By Doob's theorem, the result follows. None of the results appeared is claimed for originality. The value of the thesis is on the systematic analysis of the ergodicity of monotone SODEs, which gives a reference for future studies on this type of equation.

[**Key words**]: Invariant Measure, Monotone SODE, Non-Lipschitz Condition, Semigroup Property, Ergodicity

Table of Content

1. Introduction and Outlines

Description of a system using probability would make it more precise, however, more complex, simultaneously. In standard ergodic theory, the dynamic systems are deterministic; that is, given an x in the phase space of the system, its position would be at $T_t x$ after time t. There is no possibility for x to go to other places, even a mistake of small ϵ. However, no matter how simple a system is, as long as it exists in real world, it will "make mistakes" by disturbance. Therefore we define the *Markov kernel* (to be studied in §2.1) $P_t(x, A)$, which is the *conditional probability* (to be studied in §2.4) of X goes to A after time t given it started at x.

Although we can describe the system abstractly by a Markov kernel, it is generally impossible to solve it implicitly; that is, obtaining a mathematical formula of $P_t(x, A)$. However, intuitively, for a "regular" system, if we observe it for a sufficiently long time, we should obtain all the information of it. Such hypothesis is called *ergodicity* (to be studied in §3.1) — time average equaling space average.

In this paper, we shall focus on a purticular class of systems which are generated by the solutions of a class of SODEs (under non-Lipschitz conditions). The Lipschitz case has already been well-studied in [Da Prato et al., 1996]. Our non-Lipshitz conditions, although had been studied as well, is a trendency in recent studies. We need some preparatory work before having a close look at it.

Generally speaking, Section 2 provides us both of tools for the study of Markov semigroup and SODEs and section 3 studies the methods of finding ergodic measures for a given semigroup. The main results are

- existence and uniqueness of *invariant measure*(to be studied in §3.1) imply ergodicity (Doob's Theorem 3.4.3);
- strong Feller property and irreducibility imply uniqueness of invariant measure (Hasminskii's Theorem 3.4.2).

Therefore, provided that the solution is indeed a Markov semigroup, we only need to show three properties to achieve our goal, namely the existence of invariant measure, strong Feller property and irreducibility.

2. Essentials in Probability Theory

For the readers' convenience, the mathematical preliminaries in probability theory are introduced in this section. The representation style of this section is well-designed: for those are highly related to the understanding of our main object but merely mentioned in standard textbooks, rigorous mathematical treatments are implemented; for the others, we will only provide a brief description.

Generally speaking, in §2.1-§2.3, we introduce Markov kernels, semigroups and process, which will be needed for the analysis of the problem, and §2.4-§2.6 provides the necessary tools to the setup of our problem (to define an SODE). In §2.1, a probabilistic transport is described in both kernel and semigorup languages. Tensor product theorem helps us to define a probability measure on a finite dimensional space with a given transport. In §2.2, we investigate on infinite dimensions and make clear the widely-accepted but ambiguous terminologies in stochastic process such as *information flow*. These are essential to help understand the mathematical languages in human words. Then we move on to §2.3 to extend the finite dimensional probability measure to infinite dimensions, which explains the existence of Brownian motion. We also remark that this can be generalized to the construction of any Markov process. In §2.4, the connection between two kinds of conditional expectation is illustrated clearly. Furthermore, we point out that if we use tensor product theorem to build the probability measure, it can indeed be understood as conditional probability. We give the definitions for both distrete- and continuous-time martingale in §2.5. And finally in §2.6, we breifly discuss the construction of Itô integral and state the well-known formula established by Itô, which in my opinion is the marrow in his theory and would be helpful to the estimation of solutions in the main part of the thesis.

2.1 Markov Kernel and Markov Semigroup

In this subsection, we shall introduce the idea of *transition* in both the language of kernel and semigroup, which is not included in some standard textbooks of probability theory. The materials could be found in Chapter 1 of [Douc et al., 2018]. The beautiful notation makes it easier for us to illustrate the ideas of in both Markov chain and Markov process.

Since we are in the universe of probability, we only care for *Markvo kernel*. However, it should be remarked that similar results in this section hold for σ-finite kernel[3].

2.1.1 Markov Kernel and its Corresponding Operator

There are two mathematical languages to describe a *probabilistic transport*: *kernel* language and *semigroup* language.

Definition 2.1.1 (Markov kernel)**.** Let $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$ be two measurable spaces. A *Markov kernel* N on $\mathbb{X} \times \mathscr{Y}$ is a mapping $N : \mathbb{X} \times \mathscr{Y} \to [0, 1]$ satisfying the following conditions:

(i) for every $x \in \mathbb{X}$, the mapping $N(x, \cdot) : A \mapsto N(x, A)$ is a probability measure on $\mathscr{Y}$;

(ii) for every $A \in \mathscr{Y}$, the mapping $N(\cdot, A) : x \mapsto N(x, A)$ is a measurable function from $(\mathbb{X}, \mathscr{X})$ to $([0, 1], \mathscr{B})$[1].

Remark 2.1.2. We can understand a Markov kernel $N(x, A)$ as the probability of x going to A with the help of N. For a reason, see Remark 2.4.8.

Remark 2.1.3 (Probability measure seen as Markov kernel). A probability measure ν on a space $(\mathbb{Y}, \mathscr{Y})$ can be seen as a Markov kernel on $\mathbb{X} \times \mathscr{Y}$ by defining $N(x, A) = \nu(A)$ for all $x \in \mathbb{X}$. In this case, our previous understanding does not make sense since all the probability of x goes to a fixed set A equal. We can understand it as the *initial measure* on $(\mathbb{Y}, \mathscr{Y})$; that is, a given probability measure before transportations happen.

Notation 2.1.4. Let N be a Markov kernel on $\mathbb{X} \times \mathscr{Y}$ and $f \in \mathbb{B}_b(\mathbb{Y})$ (the set of all real-valued bounded functions on $\mathbb{Y}$). A function $F_N f : \mathbb{X} \to \mathbb{R}$ is defined by

$$F_N f(x) \stackrel{\text{def}}{=} \int_{\mathbb{Y}} N(x, \mathrm{d}y) f(y). \tag{1}$$

Notice that $F_N \mathbb{1}_A(x) = N(x, A)$, for $A \in \mathscr{Y}$.

[1] $\mathscr{B}$ will always denote the Borel σ-algebra of the corresponding metric space. In this case, $\mathscr{B} = \mathscr{B}([0, 1])$.

By Remark 2.1.3, we can consequently define F_ν similarly,

$$F_\nu f(x) \equiv \int_{\mathbb{Y}} \nu(\mathrm{d}y) f(y),$$

for all $x \in \mathbb{X}$. Since the function $F_\nu f(x)$ is a constant, we denote it simply by $F_\nu f$. Note that this is equivalent to $\mathrm{E}_\nu(f)$.

The following lemma ensures the measurablity of Nf.

Lemma 2.1.5. *Let N be a Markov kernel on $\mathbb{X} \times \mathscr{Y}$. Then*

(i) for all $f \in \mathbb{B}_b(\mathbb{Y})$, $F_N f \in \mathbb{B}_b(\mathbb{X})$;

(ii) $|F_N f|_\infty \leq |f|_\infty$.

Proof. Write down the definition to check that $F_N f$ is $\mathscr{X}$-measurable when f is a simple function. Then for $f \in \mathbb{B}_b(\mathbb{Y})$, there exists a sequence of functions f_n converges pointwise to f by the approximation theorem. Then by the dominated convergence theorem, $F_N f(x) = \lim_n F_N f_n(x)$ for all $x \in \mathbb{X}$. Therefore $F_N f$ is $\mathscr{X}$-measurable as being the pointwise limit of a sequence of measurable functions. Finally, from

$$F_N f(x) = \int_{\mathbb{Y}} f(y) N(x, \mathrm{d}y) \leq |f|_\infty \int_{\mathbb{Y}} N(x, \mathrm{d}y) = |f|_\infty,$$

we obtain $|F_N f|_\infty \leq |f|_\infty$. □

Notation 2.1.6 (Indentify F_N with N). Thanks to the lemma, F_N becomes an bounded linear operator from $\mathbb{B}_b(\mathbb{Y})$ to $\mathbb{B}_b(\mathbb{X})$; in other words, every Markov kernel $N(x, A)$ has a natural embedding to $L(\mathbb{B}_b(\mathbb{Y}), \mathbb{B}_b(\mathbb{X}))$ ($L(X, Y)$ denotes the space of bounded linear operator from X to Y. If $X = Y$, then simply denoted by $L(X)$.) by $N \mapsto F_N$. Moreover, if the Markov kernel is just a probability measure ν, then F_ν can be viewed as a linear functional.

With a slight abuse of notation for the convenience of representation, we will use the same symbol for both the kernel and the operator [2] ; that is, we will identify F_N with N.

[2]Although it sounds unreasonable, we have met such abusion already in *Linear Algebra*, when we identify matrix A with the linear map induced by A.

Thus the notation F_N would be abandoned proceedingly.

The following lemma provides a useful tool to verify a construction of operator being a Markov kernel.

Lemma 2.1.7. *Let $M : \mathbb{B}_b(\mathbb{Y}) \to \mathbb{B}_b(\mathbb{X})$ be an additive ($M(f+g) = Mf + Mg$) and homogeneous ($M(\alpha f) = \alpha M f$) operator such that $\lim_n M(f_n) = M(\lim_n f_n)$ for every increasing sequence $\{f_n, n \in \mathbb{N}\}$ of functions in $\mathbb{B}_b(\mathbb{Y})$. Furthermore, $M(\mathbb{1}_{\mathbb{Y}}) = 1$. Then*

(i) the function defined on $X \times \mathscr{Y}$ by $N(x, A) = M(\mathbb{1}_A)(x)$ for $x \in \mathbb{X}$ and $A \in \mathscr{Y}$ is a Markov kernel;

(ii) $M(f) = Nf$ for all $f \in \mathbb{B}_b(\mathbb{Y})$.

Proof.

1. Since M is additive for each $x \in \mathbb{X}$, the function $A \to N(x, A)$ is additive. σ-additive then follows by the monotone convergence property. Write down the definition of $N(x, A)$ being a Markov kernel to finish the proof.

2. To show $M(f) = Nf$ for all $f \in \mathbb{B}_b(\mathbb{Y})$. Consider firstly f being simple functions and then apply dominated convergence theorem.

□

2.1.2 Compositions of Kernels, Markov Semigroup

Theorem 2.1.8 (Compositions of kernels)**.** *Let $(\mathbb{X}, \mathscr{X})$, $(\mathbb{Y}, \mathscr{Y})$ and $(\mathbb{Z}, \mathscr{Z})$ be three measurable spaces and let M, N be two kernels on $X \times \mathscr{Y}$ and $\mathbb{Y} \times \mathscr{Z}$ respectively. Then there exists a kernel on $\mathbb{X} \times \mathscr{Z}$, called the* composition *of M and N, denoted by MN, such that for all $x \in \mathscr{X}$, $A \in \mathscr{Z}$ and $f \in \mathbb{B}_b(\mathbb{Z})$,*

$$MN(x, A) = \int_{\mathbb{Y}} M(x, \mathrm{d}y) N(y, A).$$

Furthermore, $MNf(x) = M[Nf](x)$. Consequently, the compositions (when there are more than three kernels) of kernels are associative.

Proof. The kernels M and N define two additive and positively homogeneous operators on $\mathbb{B}_b(\mathbb{Y})$ and $\mathbb{B}_b(\mathbb{Z})$. Then it is easy to check that $M \circ N$ is additive and positively homogeneous, where $\circ$ denote the usual composition of operators. The monotone convergence property also holds for $M \circ N$. Therefore by Lemma 2.1.7, there exists a kernel, denoted by MN, such that $M \circ N(f) = (MN)(f)$ for all $f \in \mathbb{B}_b(\mathbb{Z})$. To conclude the proof, it remains to write down the relationship between the kernel and its relating operator. □

Remark 2.1.9. (i) As Remark 2.1.2, we can understand $MN(x, A)$ as the probability of x goes A with the help of N then M.

(ii) From Remark 2.1.3, as a corollary, if $\nu \in \mathbb{M}_1(\mathscr{X})$ (the set of all probability measures on $(\mathbb{X}, \mathscr{X})$), then there exists a probability measure $\nu N \in \mathbb{M}_1(\mathscr{X})$ such that

$$\nu M(A) = \int_{\mathbb{X}} \nu(\mathrm{d}x) M(x, A). \tag{2}$$

Similarly, νM can be understood as the result measure after transported by M with initial measure ν.

Remark 2.1.10. Given a Markov kernel N on $\mathbb{X} \times \mathscr{X}$, we may define the n-th power of this kernel as the n-th compositions. Note that the associativity of the compositions yields the Chapman-Kolmogorov equation:

$$N^{n+k} = N^n \circ N^k \tag{3}$$

or equivalently

$$N^{n+k}(x, A) = \int_{\mathbb{X}} N^n(x, \mathrm{d}y) N^k(y, A). \tag{4}$$

Equation (3) is called a *semigroup* structure. Formally, we have the follwing definition.

Definition 2.1.11. Let $\mathbb{T} = \mathbb{N}$ or $\mathbb{R}_+$. A *Markov semigroup* $\{P_t, t \in \mathbb{T}\}$ on $\mathbb{B}_b(\mathbb{Y})$ is a mapping $\mathbb{T} \to L(\mathbb{B}_b(\mathbb{Y}))$, $t \mapsto P_t$ such that

(i) $P_0 = \mathrm{Id}$, $P_{t+s} = P_t \circ P_s$ for all $t, s \in \mathbb{T}$.

(ii) For any $t \in \mathbb{T}$ and $x \in \mathbb{Y}$, there exists a probability measure $\pi_t(x, \cdot) \in \mathbb{M}_1(\mathbb{Y})$ such that

$$P_t\varphi(x) = \int_{\mathbb{Y}} \varphi(y)\pi_t(x, \mathrm{d}y)$$

for all $\varphi \in \mathbb{B}_b(H)$.

(iii) When $\mathbb{T} = \mathbb{R}_+$, for any $\varphi \in C_b(H)$ (the set of continuous and bounded functions on H) (resp. $\mathbb{B}_b(H)$) and $x \in H$, the function $t \mapsto P_t\varphi(x)$ is continuous (resp. Borel measurable).

It is easy to see $\pi_0(x, \cdot) = \delta_x$ for all $x \in \mathbb{Y}$; and $\pi_{t+s}(x, A) = \int_E \pi_t(x, \mathrm{d}y)\pi_s(y, A)$.

Very often, (iii) is not required in the definition of Markov semigroup P_t. In this case condition (iii) means that P_t is *stochastic continuous* (Definition 5.1, [Da Prato, 2006]).

Remark 2.1.12. When $\mathbb{T} = \mathbb{N}$, the semigroup can be constructed by only one Markov kernel. It is immediate, from (1) and (3), that $\{N^k, k \in \mathbb{N}\}$ is a Markov semigroup, provided that N is a Markov kernel.

However when $\mathbb{T} = \mathbb{R}_+$, the time index is continuous. We are required to have a sequence of Markov kernels satisfying $\pi_{t+s}(x, A) = \int_E \pi_t(x, \mathrm{d}y)\pi_s(y, A)$. Since we abuse the notation (Notation 2.1.6), $\pi_t(x, \cdot)$ would be written as $P_t(x, \cdot)$ for a semigroup induced by a Markov kernel.

Remark 2.1.13. Let $\mathbb{X}, \mathbb{Y}$ be metric space so that $\mathbb{B}_b(\mathbb{X}), \mathbb{B}_b(\mathbb{Y})$ would be Banach space (Theorem 4.9, [Robinson, 2020]). Now in the view point of semigroup, (2) is equivalent to

$$\nu M(f) = \int_{\mathbb{X}} Mf(x)\nu(\mathrm{d}x) = \nu(Mf).$$

Since $M \in L(\mathbb{B}_b(\mathbb{Y}), \mathbb{B}_b(\mathbb{X}))$ and $\nu \in \mathbb{B}_b(\mathbb{X})^*$ (here the star means the dual space), there is a adjoint operator $M^* \in L(\mathbb{B}_b(\mathbb{X})^*, \mathbb{B}_b(\mathbb{Y})^*)$ such that $M^*\nu(f) = \nu(Mf)$.

This remark emphasises that we could obtain similar expression as the composition in kernel language using only the language of semigroup. We will continue the discussion when the concept of invariant measure is introduced.

2.1.3 Tensor Products of Kernels

The compositions of kernels allow us to integrate on the middle steps of "transports" and care only on final effects the overall transports made, while the *tensor product* of kernels gives us the full information at each step.

We must deal with the measurablity[3]. E_y here means the section $\{z \in \mathbb{Z} : (y,z) \in E\}$.

Lemma 2.1.14. *Let $(\mathbb{Y}, \mathscr{Y})$ and $(\mathbb{Z}, \mathscr{Z})$ be two measurable spaces and N be a Markov kernel on $\mathbb{Y} \times \mathscr{Z}$. Suppose $\mathbb{1}_E, f \in \mathbb{B}_+(\mathscr{Y} \otimes \mathscr{Z})$ (recall that $\mathscr{Y} \otimes \mathscr{Z}$ means $\sigma(\mathscr{Y} \times \mathscr{Z})$).*

(i) $E_y \in \mathscr{Z}$ *for all* $y \in \mathbb{Y}$.

(ii) $N(y, E_y)$ *is* $\mathscr{Y}$*-measurable.*

(iii) $\int_{\mathbb{Z}} f(y,z) N(y, \mathrm{d}z)$ *is* $\mathscr{Y}$*-measurable.*

Proof. 1. Define

$$\mathscr{G}_1 \stackrel{\text{def}}{=} \{E \in \mathscr{Y} \otimes \mathscr{Z} : E_y \in \mathscr{Z}\}.$$

Then write down the definition to check $\mathscr{G}_1$ is a σ-algebra. On the other hand, if $A \in \mathscr{Y}, B \in \mathscr{Z}$, then $(A \times B)_y = B$ if $y \in A$ and $(A \times B)_y = \varnothing$ if $y \notin A$. Thus $A \times B \in \mathscr{G}_1$. As $\mathscr{Y} \otimes \mathscr{Z}$ is generated by such rectangles, we must have $\mathscr{G}_1 = \mathscr{Y} \otimes \mathscr{Z}$.

2. Define

$$\mathscr{G}_2 \stackrel{\text{def}}{=} \{E \in \mathscr{Y} \otimes \mathscr{Z} : N(y, E_y) \in \mathbb{B}_+(\mathbb{Y})\}.$$

Observe that $\mathscr{G}_2$ is a monotone class and contains the algebra of finite disjoint unions of measurable rectangles. $\mathscr{G}_2 = \mathscr{Y} \otimes \mathscr{Z}$ by the monotone class theorem.

3. Note that

$$\int_{\mathbb{Z}} \mathbb{1}_E(y,z) N(y, \mathrm{d}z) = \int_{\mathbb{Z}} \mathbb{1}_{E_y}(z) N(y, \mathrm{d}z) = N(y, E_y).$$

Therefore if f_n is non-negative simple functions, then $\int_{\mathbb{Z}} f_n(y,z) N(y, \mathrm{d}z)$ is measurable. The result then follows by the monotone convergence theorem.

[3] In [Douc et al., 2018], the author write (5) without checking the measurablity. We add Lemma 2.1.14 to make it rigorous. This step is also the key step when proving the classic Fubini's Theorem.

□

Theorem 2.1.15 (Tensor product)**.** *Let* $(\mathbb{X}, \mathscr{X})$, $(\mathbb{Y}, \mathscr{Y})$ *and* $(\mathbb{Z}, \mathscr{Z})$ *be three measurable spaces and let* M, N *be two Markov kernels on* $X \times \mathscr{Y}$ *and* $\mathbb{Y} \times \mathscr{Z}$ *respectively. Then there exists a Markov kernel on* $X \times (\mathscr{Y} \otimes \mathscr{Z})$*, called the* tensor product *of* M *and* N*, denoted by* $M \otimes N$*, such that for all* $f \in \mathbb{B}_b(Y \times Z, \mathscr{Y} \otimes \mathscr{Z})$ *its corresponding operator satisfies*

$$M \otimes Nf(x) = \int_{\mathbb{Y}} M(x, \mathrm{d}y) \int_{\mathbb{Z}} f(y, z) N(y, \mathrm{d}z). \tag{5}$$

Furthermore, if $(\mathbb{U}, \mathscr{U})$ *is a measurable space and* P *is a kernel on* $\mathbb{Z} \times \mathscr{U}$*, then* $(M \otimes N) \otimes P = M \otimes (N \otimes P)$*, i.e. the tensor product of kernels is associative.*

Proof. As Lemma 2.1.14 shows the integrand is measurable, we can define the mapping $I : \mathbb{B}_b(\mathbb{Y} \times \mathbb{Z}) \to \mathbb{B}_b(\mathbb{X})$ by

$$I(f) = \int_{\mathbb{Y}} M(x, \mathrm{d}y) \int_{\mathbb{Z}} f(y, z) N(y, \mathrm{d}z).$$

The mapping is additive and homogeneous. The monotone convergence property also holds. The Markov kernel $M \otimes N$ thus exists. Since we can explicitly write down the definition of tensor product, the associativity is also nature. □

Notation 2.1.16. For $n \geq 1$, the n-th tensor power $P^{\otimes n}$ of a kernel P on $\mathbb{X} \times \mathscr{X}$ is the kernel on $\mathbb{X} \times \mathscr{X}^{\otimes n}$ defined by $P \otimes \cdots \otimes P$, i.e.

$$P^{\otimes n} f(x) = \int_{\mathbb{X}^n} f(x_1, \ldots, x_n) P(x, \mathrm{d}x_1) P(x_1, \mathrm{d}x_2) \cdots P(x_{n-1}, \mathrm{d}x_n). \tag{6}$$

Remark 2.1.17. Different from compositions of kernels, tensor products $M \otimes N$ stored all the probabilistic information of the transport first N then M. For example, $M \otimes N(x, A \times B)$ for $A \in \mathscr{Y}, B \in \mathscr{Z}$ means the probability of x goes to A first with N then goes from A to B with M.

2.1.4 *Degression on Tonelli-Fubini Theorem

Let us leave the mainstream of the thesis for a while to introduce the classic Tonelli-Fubini Theorem (a well known result in measure theory) as a corollary (or some kind of remark) of Theorem 2.1.15. This may lead to a better understanding of tensor product.

Corollary 2.1.18 (Tonelli-Fubini)**.** *Let ν be a probability measure on $(\mathbb{Y}, \mathscr{Y})$ and N be a Markov kernel on $\mathbb{Y} \times \mathscr{Z}$. Then there exists a probability measure on $\mathscr{Y} \otimes \mathscr{Z}$, denoted $\nu \otimes N$, such that*

(i) for all $f \in \mathbb{B}_b(Y \times Z, \mathscr{Y} \otimes \mathscr{Z})$,

$$\nu \otimes Nf = \int_{\mathbb{Y}} \nu(\mathrm{d}y) \int_{\mathbb{Z}} f(y,z) N(y, \mathrm{d}z).$$

(ii) for all Borel measurable function f such that $\nu \otimes Nf$ exists (resp. is finite), then $\int_{\mathbb{Z}} f(y,z)N(y,\mathrm{d}z)$ exists (resp. is finite) for ν-almost every y, and defines a Borel measurable function of y if it is taken as 0 or as any Borel measurable function of y on the exceptional set. Also

$$\nu \otimes Nf = \int_{\mathbb{Y}} \nu(\mathrm{d}y) \int_{\mathbb{Z}} f(y,z) N(y, \mathrm{d}z).$$

Proof. For statement (i), just take $M = \nu$ in Theorem 2.1.15 in the sense of viewing measure as kernel (Remark 2.1.3). For (ii), suppose $\nu \otimes Nf^- < \infty$. By statement (i),

$$\int_{\mathbb{Y}} \nu(\mathrm{d}y) \int_{\mathbb{Z}} f^-(y,z) N(y, \mathrm{d}z) = \nu \otimes Nf^- < \infty$$

so that $\int_{\mathbb{Z}} f^-(y,z)N(y,\mathrm{d}z)$ is ν-integrable hence ν-a.e. finite. Therefore

$$\int_{\mathbb{Z}} f(y,z) N(y,\mathrm{d}z) = \int_{\mathbb{Z}} f^+(y,z) N(y,\mathrm{d}z) - \int_{\mathbb{Z}} f^-(y,z) N(y,\mathrm{d}z)$$

for ν-almost every y. The remaining part of proof is just discussing different cases for the existence (or finiteness) of $\nu \otimes Nf$. □

Corollary 2.1.19 (Classic Fubini). *Let ν_1, ν_2 be two probability measures on $(\mathbb{Y}, \mathscr{Y})$ and $(\mathbb{Z}, \mathscr{Z})$. Then there exists a probability measure on $\mathscr{Y} \otimes \mathscr{Z}$, denoted $\nu_1 \otimes \nu_2$, such that: if f is a Borel measurable function on $(Y \times Z, \mathscr{Y} \otimes \mathscr{Z})$ such that $\nu_1 \otimes \nu_2 f$ exists, then*

$$\nu_1 \otimes \nu_2 f = \int_{\mathbb{Y}} \nu_1(\mathrm{d}y) \int_{Z} f(y,z)\nu_2(\mathrm{d}z) = \int_{\mathbb{Z}} \nu_2(\mathrm{d}z) \int_{\mathbb{Y}} f(y,z)\nu_1(\mathrm{d}y).$$

Proof. Apply Tonelli-Fubini's Theorem (Corollary 2.1.18) with $N = \nu_2$. Then change the position of ν_1 and ν_2 to obtain the symmetric equality. □

2.2 Stochastic Process

The aim of this section is to help understand *stochastic process*. Traditionally, a stochastic process on a probability space $(\Omega, \mathscr{F}, \mathrm{P})$ is a family of random variables $\{X_t, t \in \mathbb{T}\}$, where $\mathbb{T}$ is the index set equals to $\mathbb{N}$ or $\mathbb{R}_+$. However, it can also be understood as a $\mathbb{R}^{\mathbb{T}}$-valued *random object*. From the view of the latter, the well-known understanding of *natrual filtration* as *information* would be mathematically reasonable.

2.2.1 Random Object

Definition 2.2.1 (random object). A *random object* X on a probability space $(\Omega, \mathscr{F}, \mathrm{P})$ is a measurable function from Ω to $(\mathbb{X}, \mathscr{X})$.

If $(\mathbb{X}, \mathscr{X}) = (\mathbb{R}^n, \mathscr{B})$, X is said to an *random vector* or $\mathbb{R}^n$*-valued random variable* or simply *random variable* if $n = 1$.

Definition 2.2.2 (induced measure). If X is a random object from $(\Omega, \mathscr{F}, \mathrm{P}) \to (\mathbb{X}, \mathscr{X})$, the *probability measure induced by* X is the probability measure P_X on $(\mathbb{X}, \mathscr{X})$ given by

$$\mathrm{P}_X(B) \stackrel{\text{def}}{=} \mathrm{P}\{X \in B\}^4$$

for $B \in \mathscr{X}$.

One can write down the definition to check P_X is indeed an probability measure.

The induced probability measure P_X is also called the *law* of X.

[4] There is a convention in probability theory that we will often omit the ω; that is, writing $\{\omega : X(\omega) \in B\}$ as $\{X \in B\}$.

Remark 2.2.3. (i) The probability measure P_X completely characterized the random object X in the sense that it provide the probabilities of all events involving X.

(ii) Very often, when we want to investigate a random object with its law P_X given, there is no reference to the underlying probability space $(\Omega, \mathscr{F}, \mathrm{P})$, and actually the nature of the underlying space is not important as long as we can define such random object on the space[3], i.e. $\{X \in B\} \in \mathscr{F}$ for all $B \in \mathscr{X}$. In fact, we can always supply the probability space in a canonical way; take $\Omega = \mathbb{X}, \mathscr{F} = \mathscr{X}$, $\mathrm{P} = \mathrm{P}_X$ and define X to be the identity map; that is, $X(\omega) = \omega$ for all $\omega \in \Omega$.

(iii) When we say "let X be a random object on a probability space $(\Omega, \mathscr{F}, \mathrm{P})$", it actually implicitly assumes that the space should be chosen in an appropriate way such that X could be defined [5] .

2.2.2 Induced Sigma-Algebra and Doob-Dykin Lemma

Definition 2.2.4 (induced σ-algebra)**.** Let $X : (\Omega, \mathscr{F}) \to (\mathbb{X}, \mathscr{X})$ be a random object. The σ-algebra induced by X is given by

$$\sigma(X) \stackrel{\text{def}}{=} X^{-1}(\mathscr{X}).$$

One can write down the definition to check $\sigma(X)$ is indeed an σ-algebra.

Element in $\sigma(X)$ is of the form $\{X \in A\}$ for some $A \in \mathscr{X}$.

The induced σ-algebra $\sigma(X)$ is also the smallest σ-algebra making X measurable (Theorem 5.4.2, [Ash, 2000]). The lemma below named after L. Doob and Dynkin is another key characterization.

Lemma 2.2.5 (Doob-Dynkin)**.** *Let X be an random object from $(\Omega, \mathscr{F}) \to (\mathbb{X}, \mathscr{X})$. If $Z : (\Omega, \sigma(X)) \to (\mathbb{R}, \mathscr{B})$ is a random variable, then $Z = f \circ X$ for some $f : (\mathbb{X}, \mathscr{X}) \to (\mathbb{R}, \mathscr{B})$.*

[5] In fact, this kind of abbreviation is commonly used. For example, when we say "let $x \in A$", we actually implicitly assumes that A is a non-empty set.

Conversely, if $Z = f \circ X$ and $f : (\mathbb{X}, \mathscr{X}) \to (\mathbb{R}, \mathscr{B})$, then $Z : (\Omega, \sigma(X)) \to (\mathbb{R}, \mathscr{B})$ is a random variable.

In other words, a real-valued function Z is $\sigma(X)$-measurable iff it can be written as some function of X.

We include its proof here since it is the cornerstone to understand $\sigma(X)$.

Proof. The converse is trivial as compositions of measurable functions is measurable. Now assume $Z : (\Omega, \sigma(X)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Consider first the case Z is an indicator function, then a simple function and finally the general case. Here we only consider the indicator function $Z = \mathbb{1}_C$ as the remaining procedure is standard. Since Z is $\sigma(X)$-measurable, $C \in \sigma(X) = \{X^{-1}(A) : A \in \mathscr{X}\}$, so that $C = X^{-1}(A)$ for some $A \in \mathscr{X}$. Let $f = \mathbb{1}_A$, then $f \circ X = \mathbb{1}_A \circ X = \mathbb{1}_{X^{-1}(A)} = \mathbb{1}_C = Z$. □

Remark 2.2.6 (The information of X). Intuitively, the *information* generated by X is all the things which can be completely determined by X; in other words, if Y is the information generated by X and X happens, then we should know Y happens or not. This is exactly the mathematical formulation $Y = f(X)$. Therefore, $\sigma(X)$ is said to contain all the information of X or simply said to be the information of X.

2.2.3 Applications to Stochastic Process

As said in the begining of the subsection, a stochastic process can be viewed as a $\mathbb{R}^{\mathbb{T}}$-valued random object. To illustrate this, we should first define $\mathbb{R}^{\mathbb{T}}$ and then define a σ-algebra on it.

Let $\mathbb{T}$ be an infinite index set.

Definition 2.2.7. Let $\mathbb{R}^{\mathbb{T}}$ denote the space of all real-valued functions ω on the interval $\mathbb{T}$. Let $\mathscr{R}$ be the σ-algebra generated by *cylinders*, i.e. sets of the form

$$\{\omega \in \mathbb{R}^{\mathbb{T}} : (\omega(t_1), \dots, \omega(t_n)) \in A\},$$

where $0 \le t_1 < t_2 < \cdots < t_n$, $t_i \in \mathbb{T}$ for all $i = 1, \ldots, n$ and $A \in \mathscr{B}(\mathbb{R}^n)$.

A *stochastic process* $X = \{X_t, t \in \mathbb{T}\}$ is a random object from $(\Omega, \mathscr{F})$ to $(\mathbb{R}^{\mathbb{T}}, \mathscr{R})$. It is easy to see that X is $\mathscr{R}$-measurable iff X_t is $\mathscr{B}(\mathbb{R})$-measurable for all $t \in \mathbb{T}$.

Definition 2.2.8 (filtration)**.** A *filtration* is an increasing sequence of σ-algebra indexed by $\mathbb{T}$, $\{\mathscr{F}_t, t \in \mathbb{T}\}$, i.e. $\mathscr{F}_t \subseteq \mathscr{F}_{t'}$ if $t \le t'$.

The *natural filtration* of a stochastic process $X = \{X_t, t \in \mathbb{T}\}$ is the filtration consists of the induced σ-algebras $\{\sigma(\{X_s, s \le t, s \in \mathbb{T}\}), t \in \mathbb{T}\}$. Here $\{X_s, s \le t, s \in \mathbb{T}\}$ is a *truncation process* of X.

Therefore, by Remark 2.2.6, the natural filtration of $X = \{X_t, t \in \mathbb{T}\}$ could be viewed as a sequence of information generated by the truncation process$\{X_s, s \le t, s \in \mathbb{T}\}$. For the same reason, a *filtration* is also called an *information flow*.

2.3 Brownian Motion

A botanist named R. Brown observed the erratic motion of grains of pollon suspended in a liquid. A. Einstein gave a mathematical formulation of the motion which can be summarized as the following.

Definition 2.3.1 (Brownian motion)**.** A real-valued *Brownian motion* (or named *Wiener process*) is a real-valued stochastic process with time index $\mathbb{T} = \mathbb{R}_+$, $W = \{W_t, t \in \mathbb{R}_+\}$, satisfying the following properties.

(i) $W_0 = x_0$ a.s.;

(ii) (independent increment) $W_{t_1} - W_{t_0}$, $W_{t_2} - W_{t_1}$, $\cdots$, $W_{t_n} - W_{t_{n-1}}$ are independent for all $n \ge 2$ and $0 \le t_0 < \cdots < t_n$;

(iii) (stationary Gaussian law) $W_t - W_s$ follows $N(\mu(t-s), \sigma^2(t-s))$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$ for all $0 \le s < t$; and finally

(iv) W has continuous sample paths, i.e. $t \mapsto W_t$ is a continuous function on $\mathbb{R}_+$ a.s.

If $x_0 = 0, \mu = 0$ and $\sigma = 1$, then W is called a *standard* Brownian motion.

Since R. Brown had “observed” such kind of process in the real world, such motion should also exist in the world of mathematics; that is, there is indeed a stochastic process satisfies Definition 2.3.1. In fact, there are at leat three ways to show its existence:

1. Wiener’s method ([Wiener, 1923]): first defines a pre-measure on the algebra of cylinders. Then use Carathéodory Extension Theorem to extend the measure on the σ-algebra generated by cylinders. Finally show that the continuous functions with such a measure is indeed a Brownian motion.

2. A method based on Kolmogorov extension theorem and continuity theorem. This method would be explained in detail later.

3. Lévy’s interpolation method ([Lévy, 1939]): define a sequence of stochastic processes iteratively and prove the limit of the process is indeed a Browian motion.

2.3.1 Kolmogorov Extension Theorem

The content of Kolmogorov extension theorem is the vadality to extend a class of measures on a finite dimensional spaces to a measure on an infinite dimensional space, provided that the class is *consistency*.

Definition 2.3.2 (consistency condition)**.** A family of probability measures $\mu_{t_1,t_2,\ldots,t_n}$ on $\mathbb{R}^n$ is said to satisfy the *consistency condition* if for all $0 \le t_1 < t_2 < \cdots < t_n$, $A_1 \in \mathscr{B}(\mathbb{R}^{i-1})$, $A_2 \in \mathscr{B}(\mathbb{R}^{n-i})$ with $i = 1, \ldots, n$,

$$\mu_{t_1,\ldots,t_{i-1},\widehat{t_i},t_{i+1},\ldots,t_n}(A_1 \times A_2) = \mu_{t_1,\ldots,t_n}(A_1 \times \mathbb{R} \times A_2), \tag{7}$$

where $\widehat{t_i}$ means that t_i is delated.

This condition ensures different measures in the family to have the same value for different representations of the same set.

Theorem 2.3.3 (Kolmogorov’s Extension Theorem)**.** *Suppose with each* $0 \le t_1 < t_2 < \cdots < t_n$, $n \ge 1$, *there is a probability measure* $\mu_{t_1,\ldots,t_n}$ *on* $\mathbb{R}^n$. *Assume the family satisfies* *the consistency condition. Then there exists a unique probability measure* P *on the space* $(\mathbb{R}^{[0,\infty)}, \mathscr{R})$ *such that*

$$\mathrm{P}\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), \ldots, \omega(t_n)) \in A\} = \mu_{t_1,\ldots,t_n}(A)$$

for all $0 \leq t_1 < t_2 < \cdots < t_n$*,* $n \geq 1$ *and* $A \in \mathscr{B}(\mathbb{R}^n)$*.*

For a proof, see (Theorem 2.7.5, [Ash, 2000]).

Remark 2.3.4 (Existence of Brownian Motion). For each $0 \leq t_1 < t_2 < \cdots < t_n$, $n \geq 1$, define a Markov kernel for each $i = 1, \ldots, n$ by a normal density,

$$P_{t_i - t_{i-1}}(x, A) \overset{\text{def}}{=} \int_A g(y, t_i - t_{i-1} \mid x)\mathrm{d}y, \tag{8}$$

where

$$g(y, t|x) = \frac{1}{\sqrt{2\pi t \sigma}} \exp\left[-\frac{(y - x - \mu t)^2}{2t\sigma^2}\right].$$

Then there is a probability measure

$$\delta_x \otimes P_{t_1 - t_0} \otimes \cdots \otimes P_{t_n - t_{n-1}}$$

on $\mathbb{R}^n$, which satisfies the consistency condition. Therefore, by Kolmogorov's extension theorem, there exists a unique probability measure P as an extension. Then the finite marginal distribution of $(\omega(t_0), \omega(t_1), \ldots, \omega(t_n))$ could be calculated. Use the standard transformation method, one can find the distribution law of $(\omega(t_0), \omega(t_1) - \omega(t_0), \ldots, \omega(t_n) - \omega(t_{n-1}))$. Then condition (i), (ii) and (iii) in Definition 2.3.1 are be checked.

Remark 2.3.5. In fact, the above procedure, which defines Brownian motion by a Markov semigroup, can be widely generalized to any *Markov* stochastic process. The (stochastic) continuity of the defined process can be inherited from such continuity of the Markov semigroup. For more details, see (Section 2.2, [Da Prato et al., 1996]).

It suffices to check condition (iv). However, the procedure is rather complicated. The tool we shall use is the Kolmogorov's continuity theorem.

2.3.2 Kolmogorov's Continuity Theorem

Definition 2.3.6. A stochastic process $\tilde{X}_t$ is called a *version* (or named *modification*) of X_t if $\mathrm{P}\{\tilde{X}_t = X_t\} = 1$ for each $t \in \mathbb{T}$.

Theorem 2.3.7 (Kolmogorov's Continuity Theorem)**.** *Let $\{X_t, 0 \le t \le 1\}$ be a stochastic process. Assume that there exists constant α, β satisfying the inequality*

$$\mathrm{E}\left|X_t - X_s\right|^{\alpha} \le K\left|t - s\right|^{1+\beta}$$

for all $0 \le t, s \le 1$. Then X_t has a continuous version [6].

For a proof, see (Theorem 3.3.8, [Kuo, 2006]) or (Appendix, [Evans, 2013]).

Therefore, if we take $x_0 = 0, \mu = 0$ and $\sigma = 1$ in Remark 2.3.4. Then it would satisfies

$$\mathrm{E}\left|\omega(t) - \omega(s)\right|^4 = 3\left|t - s\right|^2$$

since $\omega(t) - \omega(s)$ is normally distributed wiht mean 0 and variance $t - s$. By Kolmogorov's continuity theorem, it must possess a continuous version. Replace ω by its continuous version $\widehat{\omega}$ if necessary, then it becomes a Brownian motion.

So far, we have illustrated the existence of Brownian motion.

2.4 Conditional Expectation

The concept of *conditional expectation* is the highlight of advanced probability theory. It is an essential tool for the definition of *martingale* in §2.5. For this reason, many textbooks only illustrate conditional expectation given a σ-algebra. However, for many problems we concern in the thesis, a rigorous definition for $\mathrm{P}\{A \mid X = x\}$ is needed. The material of this subsection comes from (Chapter 5, [Ash, 2000]).

2.4.1 Classic Conditional Expectation

Commonly, there are two different ways to establish the concpet of conditional expectation:

[6] More actually, the sample path of the continuous version is γ-Hölder continuous, where $\gamma \in (0, \alpha/\beta)$.

1. via Radon-Nikodym theorem (Theorem 2.2.1, [Ash, 2000]); or
2. viewing as an image after projection in the Hilbert space $L^2(\Omega)$ and then generalizing the idea.

Here we follow the first way, which is less intuitive but much quicker.

Theorem 2.4.1 (Classic Conditional Expectation)**.** *Let Y be an extended random variable on $(\Omega, \mathscr{F}, \mathrm{P})$, $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. Assume that $\mathrm{E}(Y)$ exists. Then there is a function (random variable) h: $(\Omega, \mathscr{G}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ such that*

$$\int_C Y \mathrm{d}\,\mathrm{P} = \int_C h \mathrm{d}\,\mathrm{P}$$

for all $C \in \mathscr{G}$. Furthermore, if h' is another such function, then $h = h'$, P-a.s.

We define $\mathrm{E}(Y \mid \mathscr{G})$*, called* the conditional expectation of Y given $\mathscr{G}$*, as h.*

Proof. Let $\lambda(C) = \int_C Y \mathrm{d}\,\mathrm{P}$. Check that it is a signed measure and absolutely continuous w.r.t. P. Then the result follows from the Radon-Nikodym theorem. □

2.4.2 Conditional Expectation Given a Set

Theorem 2.4.2 (Conditional Expectation)**.** *Let Y be an extended random variable on $(\Omega, \mathscr{F}, \mathrm{P})$, and $X : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{F}')$, a random object. If $\mathrm{E}(Y)$ exists, there is a function g : $(\Omega', \mathscr{F}') \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ such that for each $A \in \mathscr{F}'$,*

$$\int_{\{X \in A\}} Y \mathrm{d}\,\mathrm{P} = \int_A g(x) \mathrm{d}\,\mathrm{P}_X(x). \tag{9}$$

Furthermore, if h is another such function, then $g = h$ P_X-a.s.

We define $\mathrm{E}(Y \mid X = x)$ as $g(x)$.

Proof. Let $\lambda(A) = \int_{\{X \in A\}} Y \mathrm{d}\,\mathrm{P}$. Check that λ is a signed measure and absolutely continuous w.r.t. P_X. Then the result follows from the Radon-Nikodym theorem. □

Conditional expectation includes conditional probability as a special case.

Definition 2.4.3 (conditional probability)**.** Let $A \in \mathscr{F}$ and $X : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{F}')$, a random object. Then we define $\mathrm{P}(A \mid X = x)$ by $\mathrm{E}(\mathbb{1}_A \mid X = x)$.

The next remark gives another characterization of the conditional expectation given a set.

Remark 2.4.4. Suppose we have $g(x) = \mathrm{E}(Y \mid X = x)$. If we define $h(\omega) = g(X(\omega))$, then $h = \mathrm{E}(Y \mid \sigma(X))$ since

$$\int_{\{X \in A\}} Y \mathrm{d}\,\mathrm{P} = \int_A g(x) \mathrm{d}\,\mathrm{P}_X(x) = \int_{\{X \in A\}} h(\omega) \mathrm{d}\,\mathrm{P}(\omega) \tag{10}$$

by changing of variable. In this case, we shall usually write $h = \mathrm{E}(Y \mid X)$ for convenience. We can understand $\mathrm{E}(Y \mid X)$ by either $g(X)$ or $\mathrm{E}(Y \mid \sigma(X))$ [7] .

The above remark tells us if we have the definition $\mathrm{E}(Y \mid X = x)$, then we can use it to define $\mathrm{E}(Y \mid X)$. And they essentially means the same thing. The next remark tells us the converse is also correct.

Remark 2.4.5. In fact, we can also define $\mathrm{E}(Y \mid X = x)$ using $\mathrm{E}(Y \mid \sigma(X))$ with the help of Doob-Dykin lemma. Since $\mathrm{E}(Y \mid \sigma(X))$ is $\sigma(X)$-measurable, it can be written as a function of X, say $g(X)$. Then $g(x)$ should be the same as $\mathrm{E}(Y \mid X = x)$ by (10).

A final remark is given, which ends the discussion of relationship between $\mathrm{E}(Y \mid X)$ and $\mathrm{E}(Y \mid X = x)$.

Remark 2.4.6. Any conditional expectation given a σ-algebra arises from a random object X in this way by taking X to be the identity map from $(\Omega, \mathscr{F}) \to (\Omega, \mathscr{G})$. Then $\sigma(X) = X^{-1}(\mathscr{G}) = \mathscr{G}$ so that $\mathrm{E}(Y \mid \mathscr{G}) = \mathrm{E}(Y \mid \sigma(X)) = \mathrm{E}(Y \mid X)$.

Another question is that: does the definition of conditional expectation (and conditional probability) given a set agrees with our intuition in simple cases?

[7] The former understanding is accepted in most elementary course of probability theory, while the latter is commonly accepted in advanced courses of probability.

Example 2.4.7. (i) if X takes discrete value, we should have

$$\mathrm{P}(A \mid X = x_i) = \frac{\mathrm{P}(A \cap \{X = x_i\})}{\mathrm{P}\{X = x_i\}};$$

(ii) if X is a continuous random variable with a density function, we should have [8]

$$\mathrm{P}(Y \in C \mid X = x) = \lim_{h \to 0} \frac{\mathrm{P}(\{Y \in C\} \cap \{x - h \le X < x + h\})}{\mathrm{P}\{x - h \le X < x + h\}} = \int_C \frac{f(x,y)}{f_X(x)} \mathrm{d}y.$$

The answers to the above two simple cases are of course "yes"es. The proof can be done by pluging in the r.h.s. of each above equality to (9) and then by the uniqueness of conditional expectation.

Next example illustrates the reason why we may think the Markov kernel $N(x, B)$ as the probability of x goes to A: $N(x, B) = \mu \otimes N(B \mid X = x)$.

Example 2.4.8. Let $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$ be given and N is a Markov kernel on $(\mathbb{X}, \mathscr{Y})$, μ is a probability measure on $(\mathbb{X}, \mathscr{X})$. Let X be the identity map on $\mathbb{X}$ so that $\mathrm{P}_X = \mu$. Then for $A \in \mathscr{X}, B \in \mathscr{Y}$, by the definition of tensor product,

$$\begin{aligned} \mu \otimes N(\{X \in A\} \times B) &= \int_{\mathbb{X}} \mathrm{d}\mu(x) \int_{\mathbb{Y}} \mathbb{1}_{A \times B}(x, y) N(x, \mathrm{d}y) \\ &= \int_A \mathrm{d}\mu(x) N(x, B). \end{aligned}$$

Therefore $N(x, B) = \mu \otimes N(B \mid X = x)$ by the definition conditional expectation.

2.5 Martingales

The importance of martingales and related topics can hardly be exaggerated[10]. However, in the thesis we only use it as an auxiliary tool. Thus the treatments in this subsection would be brief.

2.5.1 Discrete-Time Martingales

Definition 2.5.1 (martingale)**.** Let $\{X_k, k \in \mathbb{N}\}$ be a sequence of integrable random variables on $(\Omega, \mathscr{F}, \mathrm{P})$ and $\{\mathscr{F}_k, k \in \mathbb{N}\}$ be a filtration; X_k is assumed $\mathscr{F}_k$-measurable for each $k \in \mathbb{N}$

[8] In fact, the so-called *conditional density* in elementary courses of probability $h_{Y|X}(x \mid y)$ is defined as $\frac{f(x,y)}{f_X(x)}$.

(this is called *adapted*). Then sequence $\{X_k, k \in \mathbb{N}\}$ is said to be a *martingale* relative to $\mathscr{F}_n$ (alternatively, we say $\{X_n, \mathscr{F}_n\}$ is a martingale) iff for all $n \in \mathbb{N}$,

$$\mathrm{E}(X_{n+1} \mid \mathscr{F}_n) = X_n;$$ [9]

a *submartingale* (resp. *supermartingale*) iff $\mathrm{E}(X_{n+1} \mid \mathscr{F}_n) \geq X_n$ (resp. $\mathrm{E}(X_{n+1} \mid \mathscr{F}_n) \leq X_n$).

Definition 2.5.2 (stopping time)**.** A *stopping time* for a filtration $\{\mathscr{F}_t, t \in \mathbb{T}\}$ is a random variable T such that $\{T \leq t\} \in \mathscr{F}_t$ for each $t \in \mathbb{T}$.

Martingale convergence theorem, optimal sampling theorem and other related results can be found in (Sections 6.3-6.7, [Ash, 2000]); Doob's martingale inequalities can be found in (Chapter 26, [Jacod et al., 2003]).

2.5.2 Continuous-Time Martingales

Definition 2.5.3. A stochastic process $\{X_t, t \geq 0\}$ is a *(continuous-time) martingale* w.r.t. a filtration $\{\mathscr{F}_t, t \geq 0\}$ iff it is adapted to the filtration, integrable and satisfies

$$\mathrm{E}[X_t \mid \mathscr{F}] = X_s$$

when $0 \leq s < t$.

The notions of sub- and supermartingale can be similarly generalized.

The fact that most results in discrete-time martingale theory are also true in continuous-time is based on Doob's regularization theorem (Theorem 9.28, [Kallenberg, 2021]), which states that any martingale w.r.t. a right-continuous and complete filtration admits a right-continuous, left-hand limits (abbreviated as *rcll* or *càdlàg*) version. For the corresponding theorems we may need, see [Karatzas et al., 1991].

Lastly we need the concept of *local martingale* to describe the martingale-like process but without integrability.

[9] In statements involving conditional expectations, the "a.s." is always understood and will usually be omitted.

Definition 2.5.4 (local martingale). A stochastic process $\{X_k, k \in \mathbb{N}\}$ is a *local martingale* if there exists a nondecreasing sequence of stopping time $\{T_k, k \in \mathbb{N}\}$ such that $\lim_k T_k = \infty$ and each $X_{t \wedge T_k}$ is a martingale.

2.6 Itô Integral

Itô Integral has been well-studied in many textbooks, for example [Kuo, 2006], [Evans, 2013] and [Øksendal, 2003]. Therefore we will only provide a brief description.

2.6.1 Construction of Itô Integral

Fix a Brownian motion $\{W_t, t \geq 0\}$ and let a filtration $\{\mathscr{F}_t, t \geq 0\}$ be the natrual filtration of W_t.

Notation 2.6.1. We will use $\mathscr{M}^2(a, b)$ to denote the space of all stochastic process

$$f(t, \omega) : [a, b] \times \Omega \to \mathbb{R},$$

where $a \leq t \leq b$, $\omega \in \Omega$, satisfying the following:

(i) $(t, \omega) \mapsto f(t, \omega)$ is $\mathscr{B} \times \mathscr{F}$-measurable, where $\mathscr{B}$ denotes the Borel σ-algebra on $[a, b]$;

(ii) $f(t, \omega)$ is adapted to the filtration $\{\mathscr{F}_t\}$.

(iii) $\mathrm{E}[\int_a^b f(t, \omega)^2 \mathrm{d}t] < \infty$.

We need condition (i) to ensure that $\int_a^b f(t, \omega)^2 \mathrm{d}t$ is $\mathscr{F}$-measurable by Fubini's theorem so that condition (iii) makes sense. Suppse $X \in \mathscr{M}^2(a, b)$, if $\|X\| \overset{\text{def}}{=} \{\mathrm{E}[\int_a^b X(t, \omega)^2 \mathrm{d}t]\}^{1/2}$, then one can check the space is a Banach space.

The steps for the construction of Itô integral are:

1. Define the value of integral $I(\sigma)$ for *elementary process* $\sigma \in \mathscr{M}^2(0, T)$ as Riemann sum.

2. Observe the Itô isometry: $\mathrm{E}(|I(\sigma)|^2) = \mathrm{E} \int_0^T |\sigma(t)|^2 \mathrm{d}t$. The l.h.s. is the $L^2(\Omega)$-norm of $I(\sigma)$ on $L^2(\Omega)$ and the r.h.s. can be regarded as the $L^2((0, T) \times \Omega)$-norm of σ on $\mathscr{M}^2(0, T)$.

3. Prove that the elementary process is dense in $\mathscr{M}^2(0,T)$ with $L^2((0,T)\times\Omega)$-norm.

4. Prove that $\lim_n I(\sigma_n)$ converges in $L^2(\Omega)$-norm and define $I(f)$ by $\lim_n I(\sigma_n)$, where σ_n approximates f on $\mathscr{M}^2(0,T)$.

For more details, see [Itô, 1944], which is the original paper, or the textbooks listed in the begining of this subsection.

Now consider $I(t) \stackrel{\text{def}}{=} \int_0^t f(r)\mathrm{d}W(r)$ as a stochastic process with a little abuse of notation. The following theorem might be one of the most important non-trivial properties.

Theorem 2.6.2. *Suppose $f \in \mathscr{M}^2(0,T)$. Then the stochastic process $I(t)$ is a centered, squared integrable, continuous martingale.*

For a proof, see (Theorem 4.3.5, 4.6.1, 4.6.2, [Kuo, 2006]).

Previously, $\int_s^t f(r)\mathrm{d}W(r)$ makes sense only when $f(s) = f(s,\omega) \in \mathscr{M}^2(s,t)$. Now we extend the class of stochastic processes.

Notation 2.6.3. Denote $\mathscr{L}^2(a,b)$ (resp. $\mathscr{L}^1(a,b)$) the space of all stochastic processes

$$f(t,\omega) : [a,b]\times\Omega \to \mathbb{R}$$

where $a \le t \le b, \omega \in \Omega$, satisfying the following:

(i) $(t,\omega) \mapsto f(t,\omega)$ is $\mathscr{B}\times\mathscr{F}$-measurable, where $\mathscr{B}$ denotes the Borel σ-algebra on $[a,b]$;

(ii) $f(t,\omega)$ is non-anticipating w.r.t. $\mathbb{F}$;

(iii) $\int_a^b f(t,\omega)^2\mathrm{d}t < \infty$ (resp. $\int_a^b |f(t,\omega)|\,\mathrm{d}t < \infty$) a.s.

The difference between $\mathscr{L}^2(a,b)$ and $\mathscr{M}^2(a,b)$ is in condition (iii). For $f \in \mathscr{M}^2(a,b)$, we require $\mathrm{E}[\int_a^b f(t,\omega)^2\mathrm{d}t] < \infty$ thus $\int_a^b f(t,\omega)^2\mathrm{d}t < \infty$ a.s.; that is, $\mathscr{M}^2(a,b) \subseteq \mathscr{L}^2(a,b)$.

Remark 2.6.4. One can still define Itô integral for $f \in \mathscr{L}(a,b)$. However, we will lose

1. Itô isometry (but Burkholder-Davis-Gundy inequality is valid, see (Chapter 1, Theorem 7.3, [Mao, 2008])); and

2. the convergence in $L^2(\Omega)$ of $I(\sigma_n)$ to $I(f)$. Instead, we only have the convergence in probability.

3. $I(t)$ as a stochastic process would no longer be a martingale (because of the lack of integrability) but a *local martingale*.

4. $I(t)$ is not continuous, but it processes a continuous version (or stronger, a continuous *realization*).

2.6.2 Itô's Formula

Due to the fact of Brownian motion's non-zero quadratic variation, there will be an additional term for the chain rule of Itô integral[14].

Definition 2.6.5 (Itô process)**.** An *Itô process* is a stochastic process of the form

$$X_t = X_a + \int_a^t f_s \mathrm{d}W_s + \int_a^t g_s \mathrm{d}s \tag{11}$$

where $a \le t \le b$, X_a is $\mathscr{F}_a$-measurable, $f \in \mathscr{L}^2(a,b)$ and $g \in \mathscr{L}^1(a,b)$.

It is convenient (and widely accepted) to write (11) by its *symbolic* shorthand

$$\mathrm{d}X_t = f_t \mathrm{d}W_t + g_t \mathrm{d}t. \tag{12}$$

Theorem 2.6.6 (Itô's Formula)**.** *Let X_t be an Itô process given by* (12)*. Suppose $F(t,x)$ is a continuous function with continuous partial derivatives $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x}$ and $\frac{\partial^2 F}{\partial x^2}$.*

Then $F(t, X_t)$ is also an Itô process and

$$\mathrm{d}F(t,X_t) = \frac{\partial F}{\partial t}(t,X_t)\mathrm{d}t + \frac{\partial F}{\partial x}(t,X_t)\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,X_t)\mathrm{d}X_t \cdot \mathrm{d}X_t,$$

and we can calculate the symbols by $\mathrm{d}t \cdot \mathrm{d}t = 0, \mathrm{d}t \cdot \mathrm{d}W_t = 0$ *and* $\mathrm{d}W_t \cdot \mathrm{d}W_t = \mathrm{d}t$*; in other* *words, by substituting* (12) *into the symbolic shorthand,*

$$\begin{aligned} F(t, X_t) =& F(a, X_a) + \int_a^t \frac{\partial F}{\partial x}(s, X_s) f_s \mathrm{d}W_s \\ &+ \int_a^t \left[\frac{\partial F}{\partial t}(s, X_s) + \frac{\partial F}{\partial t}(s, X_s) g_s + \frac{1}{2} \frac{\partial^2 F}{\partial x^2}(s, X_s) f_s^2 \right] \mathrm{d}s. \end{aligned}$$

See (Theorem 18.18, [Kallenberg, 2021]) for a complete proof in a much more general case, which in fact includes the multidimensional case that we shall introduce proceedingly; and (Theorem 4.1.2, [Øksendal, 2003]) for a sketch of proof, which is enough to understand the idea of it.

The situations in multidimensions are similar. Let $W(t) = (W_1(t), \ldots, W_m(t))$ denote m-dimensional Brownian motion. If $f_i(t) \in \mathscr{L}^1(a, b)$ and $g_{ij}(t) \in \mathscr{L}^2(a, b)$ for each i, j, then we can form the following n Itô process

$$\begin{cases} \mathrm{d}X_1 = & f_1 \mathrm{d}t + g_{11} \mathrm{d}W_1 + \cdots + & g_{1m} \mathrm{d}W_m \\ \vdots & \vdots & \vdots \\ \mathrm{d}X_n = & f_n \mathrm{d}t + g_{n1} \mathrm{d}W_1 + \cdots + & g_{nm} \mathrm{d}W_m \end{cases} \tag{13}$$

Or, in matrix notation,

$$\mathrm{d}X(t) = f \mathrm{d}t + g \mathrm{d}W(t), \tag{14}$$

where

$$\mathrm{d}X(t) = \begin{bmatrix} \mathrm{d}X_1(t) \\ \vdots \\ \mathrm{d}X_n(t) \end{bmatrix}, f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}, g = \begin{bmatrix} g_{11} & \cdots & g_{1m} \\ \vdots & & \vdots \\ g_{n1} & \cdots & g_{nm} \end{bmatrix}, \mathrm{d}W(t) = \begin{bmatrix} \mathrm{d}W_1(t) \\ \vdots \\ \mathrm{d}W_m(t) \end{bmatrix}.$$

We can extend the Itô's formula to multidimensional case.

Theorem 2.6.7 (Multidimensional Itô's Formula)**.** *Suppose $F(t, x_1, \ldots, x_n)$ is a continuous function on $[a, b]$ and has continuous first-order and second-order partial derivatives $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x_i}$ and $\frac{\partial^2 F}{\partial x_i \partial x_j}$ for $i, j = 1, \ldots, n$.*

Then

$$\mathrm{d}F(t, X(t)) = \frac{\partial F}{\partial t} \mathrm{d}t + \sum_{i=1}^n \frac{\partial F}{\partial x_i} \mathrm{d}X_i(t) + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j} \mathrm{d}X_i(t) \cdot \mathrm{d}X_j(t), \tag{15}$$

where $\mathrm{d}t \cdot \mathrm{d}t = 0$, $\mathrm{d}B_i(t) \cdot \mathrm{d}t = \mathrm{d}t \cdot \mathrm{d}B_i(t) = 0$ *and* $\mathrm{d}B_i(t) \cdot \mathrm{d}B_j(t) = \delta_{ij}\mathrm{d}t$*; or in matrix notation,*

$$\begin{aligned}\mathrm{d}F(t, X(t)) &= \frac{\partial F}{\partial t}\mathrm{d}t + (\nabla_X F)^\top \mathrm{d}X(t) + \frac{1}{2}(\mathrm{d}X(t))^\top (H_X f)\mathrm{d}X_t \\ &= \left\{ \frac{\partial F}{\partial t} + ((\nabla_X F)^\top) f + \frac{1}{2}\,\mathrm{Tr}[g^\top (H_X F) g] \right\} \mathrm{d}t + (\nabla_X F)^\top g \mathrm{d}W(t),\end{aligned}$$

where $\nabla_X F$ *is the gradient of* F *w.r.t.* X *and* $H_X F$ *is the Hessian matrix of* F *w.r.t.* X *and* Tr *is the trace operator.*

3. General Thoery for Finding Ergodic Measures

The aim of this section is to provide some general tools for finding ergodic measures. Most of the preparatory results of showing ergodicity are provided with complete proofs.

In §3.1, we breifly introduce the meaning and equivalent characterizations of ergodicity. In §3.2, we investigate in details on the structure of the set of invariant measures. One of the key results is that the unique existence of invariant measure implies ergodicity. Therefore, we shall focus on those Markov semigroups which process exactly one invariant measure. §3.3 provides some sufficient conditions for the Markov semigroups that process invariant measures and §3.4 for which of processing a unique invariant measure.

3.1 Ergodicity

Ergodic measure is a special member in the family of invariant measures. In this subsection, we shall give definitions for both of them.

3.1.1 Invariant Measure of Markov Semigroup

Assume that H be a Hilbert space and $\mathbb{T} = \mathbb{R}_+$ or $\mathbb{N}$.

Definition 3.1.1. Let $(H, \mathscr{X})$ be a measurable space. A probability measure μ on it is said to be *invariant* w.r.t. a semigroup $P_t \in L(\mathbb{B}_b(H))$, $t \in \mathbb{T}$ iff

$$\int_H P_t\varphi \mathrm{d}\mu = \int_H \varphi \mathrm{d}\mu \tag{16}$$

for all $t \in \mathbb{T}$ and $\varphi \in \mathbb{B}_b(H)$.

Remark 3.1.2. It is clear that the above definition is equivalent of saying

$$\mu P_t(A) = \mu(A) \tag{17}$$

for all $t \in \mathbb{T}$ by the classic method; or

$$P_t^*\mu = \mu \tag{18}$$

for all $t \in \mathbb{T}$ by Remark 2.1.13.

3.1.2 Ergodic Theorems

A basic fact for invariant measure w.r.t. a semigroup P_t is that we can extend P_t from an element in $L(\mathbb{B}_b(H))$ to a strongly continuous (for each $\varphi \in L^2(H, \mu)$, $\lim_{t\to 0} P_t\varphi = \varphi$) semigroup of $L(L^2(H, \mu))$ (p. 381, Theorem 1, [Yosida, 1995]). Then P_t could be view as a linear operator on a Hilbert space, so that we can use the following result in the operator theory on Hilbert space.

Theorem 3.1.3. *Let E be a Hilbert space and T be a bounded linear operator on E. Let*

$$M_n \stackrel{def}{=} \frac{1}{n}\sum_{k=0}^{n-1} T^k$$

on E. Assume that $\sup_{n\in\mathbb{N}} \|T^n\| < \infty$. Then $\lim_n M_n(x)$ exists for all $x \in E$, denoted the limiting value by $M_\infty(x)$. Moreover, $M_\infty \in L(E)$, $M_\infty^2 = M_\infty$ and $M_\infty(E) = \ker(I - T)$.

For a proof, see (Theorem 5.11, [Da Prato, 2006]).

Then apply the result to the average

$$M(T)\varphi \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T P_t\varphi \mathrm{d}t$$

for all $\varphi \in L^2(H, \mu)$ and $T > 0$. We obtain the well-knwon Von Neumann's ergodic theorem (Theorem 5.12, [Da Prato, 2006]).

Theorem 3.1.4 (Von Neumann). $\lim_{T\to\infty} M(T)\varphi$ *exists in $L^2(H, \mu)$, denoted by $M_\infty\varphi$. Moreover, it is a projection operator on Σ and also*

$$\int_H M_\infty\varphi \mathrm{d}\mu = \int_H \varphi \mathrm{d}\mu.$$

3.1.3 Characterizations of Ergodic Measures

Thanks to Von Neumann's Theorem, the following definition makes sense.

Definition 3.1.5 (ergodic, strongly mixing)**.** Let μ be an invariant measure for P_t. We say that

- μ is *ergodic* iff

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T P_t\varphi \mathrm{d}t = \bar{\varphi}$$

  in $L^2(H,\mu)$-sense for all $\varphi \in L^2(H,\mu)$,

- μ is *strongly mixing* iff

$$\lim_{T\to\infty} P_t\varphi = \bar{\varphi}$$

  in $L^2(H,\mu)$-sense for all $\varphi \in L^2(H,\mu)$,

where $\bar{\varphi} = \mu(\varphi)$ (the expected value of φ).

Remark 3.1.6. (i) Ergodicity is often interpreted by saying that the "time average" converges to the "space" average as T goes to infinity. If μ is strongly mixing, then it is erogdic by L' Hospital's theorem.

(ii) The main problems we focused in this thesis would be the existence and uniqueness of invariant measure for a *given* system. Therefore we define ergodicity for measures. However, for the problems that considering a fixed measure space and discuss the systems, one may say the ergodicity for semigroups or operators.

Ergodicity can also be characterized as the following. In fact, this is a standard result in ergodic theory. The discussion can be found in (Subsection 12.4.3, [Da Prato, 2014]).

Let Σ of be the sets of *stationary points*

$$\Sigma \overset{\text{def}}{=} \{\varphi \in L^2(H,\mu) : P_t\varphi = \varphi\} \tag{19}$$

Definition 3.1.7. Let μ be an invariant measure of P_t. A measurable set A is said to be invariant for P_t iff its characteristic function $\mathbb{1}_A$ belongs the stationary points Σ. If $\mu(A)$ equals 0 or 1, we say it is *trivial*.

Theorem 3.1.8. *Let μ be an invariant measure for P_t. Then following statements are equivalent:*

(i) *μ is erogdic.*

(ii) *The dimension of the linear space Σ of stationary points in* (19) *is* 1.

(iii) *Any invariant set is trivial.*

3.2 Structure of the Set of Invariant Measures

Let

$$\Lambda \stackrel{\text{def}}{=} \{\mu \in \mathbb{B}_b(H)^* : P_t^* \mu = \mu\}. \tag{20}$$

Then it is clear a convex susbet of $\mathbb{B}_b(H)^*$.

Theorem 3.2.1. *Assume that there is a unique invarinat measure μ for P_t. Then μ is ergodic.*

Proof. Assume by contradiction that μ is not ergodic. Then μ process a nontrivial invariant set Γ, i.e. $P_t \mathbb{1}_\Gamma = \mathbb{1}_\Gamma$. Let

$$\mu_\Gamma(A) = \frac{1}{\mu(\Gamma)} \mu(A \cap \Gamma) \tag{21}$$

for all $A \in \mathscr{B}(H)$. It is a probability measure and we are going to show it is another invariant measure, i.e.,

$$\mu_\Gamma(A) = \int_H P_t(x, A) \mu_\Gamma(\mathrm{d}x);$$

or equivalent (by classic method)

$$\mu(A \cap \Gamma) = \int_\Gamma P_t(x, A) \mu(\mathrm{d}x).$$

Since Γ is an invariant set,

$$\begin{aligned}\int_\Gamma P_t(x,A)\mu(\mathrm{d}x) &= \int_\Gamma P_t(x,A\cap\Gamma)\mu(\mathrm{d}x) + \int_\Gamma P_t(x,A\cap\Gamma^c)\mu(\mathrm{d}x)\\ &= \int_\Gamma P_t(x,A\cap\Gamma)\mu(\mathrm{d}x)\\ &= \int_\Gamma P_t(x,A\cap\Gamma)\mu(\mathrm{d}x) + \int_{\Gamma^c} P_t(x,A\cap\Gamma)\mu(\mathrm{d}x)\\ &= \int_H P_t(x,A\cap\Gamma)\mu(\mathrm{d}x) = \mu(A\cap\Gamma),\end{aligned}$$

by the invariance of μ in the last step. □

Now we would like to prove the set of extreme points of Λ is precisely the set of ergodic measures. We need the following lemma.

Lemma 3.2.2. *Let $\mu,\nu\in\Lambda$ with μ ergodic and ν absolutely continuous w.r.t. μ. Then $\mu=\nu$.*

Proof. By the definition of ergodicity,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T P_t\mathbb{1}_\Gamma \mathrm{d}t = \mu(\Gamma)$$

in $L^2(\mu)$. Therefore there exists a sequence $T_n\uparrow\infty$ such that

$$\lim_{n\to\infty}\frac{1}{T_n}\int_0^{T_n} P_t\mathbb{1}_\Gamma \mathrm{d}t = \mu(\Gamma)$$

μ-a.s. Since $\nu\ll\mu$, it holds ν-a.s. Then integrate w.r.t. ν, the l.h.s. equals $\nu(\Gamma)$ by the invariance of ν; the r.h.s. maintains the same since ν is a probability measure. Hence $\mu(\Gamma)=\nu(\Gamma)$. □

Definition 3.2.3 (extreme points)**.** Let C be a convex set. $x\in C$ is said to be an *extreme point* iff the existence of $\alpha\in(0,1)$ such that $x=\alpha y+(1-\alpha)z$ for $y,z\in C$ implies $x=y=z$.

Theorem 3.2.4. *The set of all invariant ergodic measures of P_t coincides with the set of all extreme points of Λ.*

Proof. 1. Assume μ is ergodic. If there exists $\alpha \in (0,1)$ such that $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ then clearly $\mu_1 \ll \mu, \mu_2 \ll \mu$. Hence $\mu_1 = \mu_2 = \mu$.

2. Assume μ is a extreme point. Let Γ be an invariant set. Define μ_Γ as (21). We know that μ_Γ is an invariant measure. Then one can easily check the following

$$\mu = \mu(\Gamma)\mu_\Gamma + (1-\mu(\Gamma))\mu_{\Gamma^c}.$$

Therefore $\mu(\Gamma)$ must equal to zero or one, which shows the erogdicity.

□

Theorem 3.2.5. *If μ and ν are both erogdic, then $\mu = \nu$ or $\mu \perp \nu$ (μ and ν are mutually singular).*

Proof. Assume $\mu \neq \nu$. Let $\Gamma \in \mathscr{B}(H)$ such that $\mu(\Gamma) \neq \mu(\Gamma)$. Then by the definition of ergodicity, there exists $T_n \uparrow \infty$ and M, N Borel sets such that $\mu(M) = \mu(N) = 1$ and

$$\lim_{n\to\infty} \frac{1}{T_n}\int_0^{T_n} P_t\mathbb{1}_\Gamma(x)\mathrm{d}t = \mu(\Gamma),$$

for all $x \in M$; and

$$\lim_{n\to\infty} \frac{1}{T_n}\int_0^{T_n} P_t\mathbb{1}_\Gamma(x)\mathrm{d}t = \nu(\Gamma),$$

for all $x \in N$. We can take the common sequence T_n by replacing it with subsequence if necessary. Then we must have $M \cap N = \varnothing$, i.e. μ and ν are mutually singular. □

3.3 Existence of Invariant Measure

In this subsection, we shall prove the famous Krylov-Bogoliubov Theorem and its consequences, which are important tools to show the existence of invariant measures.

Definition 3.3.1 (Feller)**.** Let P_t be a Markov semigroup on H. We say P_t is *Feller* iff $P_t\varphi \in C_b(H)$ for any $\varphi \in C_b(H)$ and any $t \geq 0$.

Lemma 3.3.2. *Let $\mu, \nu \in \mathbb{M}_1(H)$ be such that*

$$\int_H \varphi(x)\mu(\mathrm{d}x) = \int_H \varphi(x)\nu(\mathrm{d}x)$$

for all $\varphi \in C_b(H)$*. Then* $\mu = \nu$*.*

Proof. Note that $\varphi_n \in \mathbb{B}_b(H)$ defined by

$$\varphi_n(x) = \begin{cases} 1, & \text{if } x \in C \\ 1 - nd(x, C) & \text{if } d(x, C) \leq 1/n \\ 0 & \text{if } d(x, C) \geq 1/n \end{cases}$$

is uniformly bounded by 1 and converges to $\mathbb{1}_C$ when C is closed. Then the dominated convergence theorem implies $\mu(C) = \nu(C)$. As the collection of closed sets generates the Borel σ-algebra of H, $\mu = \nu$ as claimed. □

Theorem 3.3.3 (Krylov-Bogoliubov)**.** *If* P_t *is Feller and for some* x_0*, the sequence of measures*

$$\mu_T(x_0, G) = \frac{1}{T} \int_0^T P_t \mathbb{1}_G(x_0) \mathrm{d}t = \frac{1}{T} \int_0^T P_t(x_0, G) \mathrm{d}t$$

is tight, then there exists an invariant measure μ *for* P_t *on* H*.*

Proof. By the well-known Prokhorov theorem, tightness implies weak compactness. There exists $\{\mu_{T_k}\}_{k \in \mathbb{N}}$ weakly converge to μ. That is, for $\psi \in C_b(H)$,

$$\lim_k \int_H \psi \mathrm{d}\mu_{T_k} = \int_H \psi \mathrm{d}\mu.$$

From the definition of μ_T,

$$\int \mathbb{1}_G \mathrm{d}\mu_T = \mu_T(G) = \frac{1}{T} \int_0^T \left[\int \mathbb{1}_G(y) P_t(x_0, \mathrm{d}y) \right] \mathrm{d}t.$$

Therefore

$$\int \psi \mathrm{d}\mu_T = \frac{1}{T} \int_0^T \left[\int \psi(y) P_t(x_0, \mathrm{d}y) \right] \mathrm{d}t$$

for all $\psi \in C_b(H)$. Using this,

$$\lim_k \int_H \psi \mathrm{d}\mu_{T_k} = \lim_k \frac{1}{T_k} \int_0^T \left[\int \psi(y) P_t(x_0, \mathrm{d}y) \right] \mathrm{d}t = \lim_k \frac{1}{T_k} \int_0^T P_t \psi(x_0) \mathrm{d}t.$$

For any $\varphi \in C_b(H)$, choose $\psi = P_s\varphi \in C_b(H)$ by Feller property, then

$$\begin{aligned}\int_H P_s\varphi \mathrm{d}\mu &= \lim_k \frac{1}{T_k}\int_0^{T_k} P_{t+s}\varphi(x_0)\mathrm{d}t \\ &= \lim_k \frac{1}{T_k}\left[\int_0^{T_k} P_t\varphi(x_0)\mathrm{d}t + \int_{T_k}^{T_k+s} P_t\varphi(x_0)\mathrm{d}t - \int_0^s P_t\varphi(x_0)\mathrm{d}t\right] \\ &= \lim_k \int_H \varphi \mathrm{d}\mu_{T_k} = \int \varphi \mathrm{d}\mu.\end{aligned}$$

By Lemma 3.3.2, μ is an invariant measure for P_t. □

3.4 Uniqueness of Invariant Measure

The following definitions is crucial for the existence and uniqueness of the invariant measure, as we shall see later.

Definition 3.4.1 (strong Feller, irreducible, regular)**.** Let P_t be a Markov semigroup on H.

- P_t is *strong Feller* iff $P_t\varphi \in C_b(H)$ for any $\varphi \in \mathbb{B}_b(H)$ and any $t > 0$.
- P_t is *irreducible* iff $P_t\mathbb{1}_{B(x_0,r)}(x) > 0$ for all $x, x_0 \in H$, $r > 0$ and any $t > 0$.
- P_t is *regular* iff for fixed $t > 0$, all probability measures $\{\pi_t(x,\cdot)\colon\ x \in H\}$ are mutually equivalent (two measures are equivalent iff $\mu \ll \nu$ and $\nu \ll \mu$, i.e. $\mathscr{N}_\mu = \mathscr{N}_\nu$, where $\mathscr{N}_\mu$ denotes the collection of sets of measure zero by μ.).

Theorem 3.4.2 (Hasminskii)**.** *Assume that the Markov semigroup P_t is strong Feller and irreducible. then it is regular.*

Proof. To prove the regularity, it suffice to show that $P_t(x, A) > 0$ implies $P_t(y, A) > 0$ for all $x, y \in H$. Now assume $P_t(x, A) > 0$. Pick $h \in (0, t)$. We have

$$P_t(x, A) = \int_H P_h(x, \mathrm{d}z)P_{t-h}(z, A)$$

so that $P_{t-h}(z_0, A) > 0$. By strong Feller, there exists $B(z_0, r)$ such that $P_{t-h}(z, A) > 0$ for all $z \in B(z_0, r)$. Hence

$$\begin{aligned} P_t(y, A) &= \int_H P_h(y, \mathrm{d}z) P_{t-h}(z, A) \\ &\geq \int_{B(z_0,r)} P_h(y, \mathrm{d}z) P_{t-h}(z, A) > 0 \end{aligned}$$

by irreducibility. □

Theorem 3.4.3 (Doob)**.** *Assume that the Markov semigroup P_t is regular and processes an invariant measure μ. Then μ is equivalent to $P_t(x, \cdot)$ for any $t > 0$ and $x \in H$. Moreover, μ is the unique ergodic measure for P_t.*

Proof. Note that

$$\mu(A) = \int_H P_t(y, A) \mu(\mathrm{d}y).$$

Therefore the equivalence of μ and $P_t(x, \cdot)$ follows immediately by the definition of regularity.

Let Γ be the invariant set, with $\mu(\Gamma) > 0$, $P_t \mathbb{1}_\Gamma = \mathbb{1}_\Gamma$. Since $\mu(\Gamma) > 0$, we must have $P_t \mathbb{1}_\Gamma(x) = P_t(x, \Gamma) > 0$, for all $x \in \mathbb{R}^n$ by equivalence. Then we obtain $\mathbb{1}_\Gamma(x) > 0$ for all $x \in \mathbb{R}^n$ so that $\mathbb{1}_\Gamma = \mathbb{1}$. Hence μ is erogdic.

If there is another invariant erogdic measure ν. Then μ must equivalent to ν so that $\mu = \nu$ by Lemma 3.2.2. □

Remark 3.4.4. Under the conditions of Doob's Theorem, the conclusion of μ can be stronger than ergodicity. In fact, μ is strongly mixing. The proof (Theorem 4.2.1, [Da Prato et al., 1996]) is not that easy so that we only quote the result.

4. Ergodicity of Monotone SODEs

We are here concerned with the study of the asymptotic behaviour of the *Stochastic Ordinary Differential Equation* (SODE)

$$\begin{cases} \mathrm{d}X(t) = b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t) \\ X(s) = \eta, \end{cases} \tag{22}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $X(t), W(t) \in \mathbb{R}^d$, $\eta \in L^2(\Omega, \mathscr{F}_s)$. Assume b, σ are both continuous maps.

First let us review some basic notions and inequalities in SODE theory. The outline of this section would be presented at the end of §4.2, after the problem has been setted up.

4.1 Basic Notions and Inequalities in SODE Theory

Definition 4.1.1. An $\mathbb{R}^d$-valued stochastic process $\{X_t, s \le t \le T\}$ is called a *solution* of (22) if it has the following properties:

(i) $\{X_t\}$ is continuous and $\mathscr{F}_t$-adapted.

(ii) $b(X_t) \in \mathscr{L}^1(s,T)$ and $\sigma(X_t) \in \mathscr{L}^2(s,T)$.

(iii) The following stochastic integral equation

$$X_t = x_0 + \int_s^t b(X_u)\mathrm{d}u + \int_s^t \sigma(X_u)\mathrm{d}W_u \tag{23}$$

holds a.s. for $t \in [s,T]$.

A solution $\{X_t\}$ is said to be *unique* if any other solution $\{\tilde{X}_t\}$ is *indistinguishable* from $\{X_t\}$, that is,

$$\mathrm{P}\{X_t = \tilde{X}_t, \forall t \in [s,T]\} = 1.$$

Notation 4.1.2. We shall use $X(t,s,x,\omega)$ (or $X_t^{s,x}(\omega)$ when there are to many parentheses) to denote the solution of SODE (22), where s, x means the SODE is initialized at s with value x and t means at time t. If $s = 0$, then we simply write $X(t,x,\omega)$ (or $X_t^x(\omega)$) instead of

$X(t, 0, x, \omega)$. Sometimes when there is no chance of ambiguity, we would only write $X_t(\omega)$. We often omit to write ω as the convention in probability theory.

The advantage of the notation $X(t, s, x, \omega)$ is that, when the initial value possesses randomness, i.e. $x = x(\omega)$ is a random variable, then there will be two different contributions to the randomness of $X(t, s, x(\omega), \omega)$. Using our notation, those two kinds of randomnesses are seperated clearly in mind.

In the following, we shall use η, ζ to denote a random initial value and x, y to denote a constant.

The following two Gronwall-type inequalities are our main tools when finding boundaries. Their proofs can be found in (Section 1.8, [Mao, 2008]).

Lemma 4.1.3 (Gronwall's Inequality). *Let $T > 0$ and $c \geq 0$. Let $u(\cdot)$ be a Borel measurable bounded non-negative function of $[0, T]$, and let $v(\cdot)$ be a non-negative integrable function on $[0, T]$. If*

$$u(t) \leq c + \int_0^t v(s)u(s)\mathrm{d}s$$

for all $0 \leq t \leq T$, then

$$u(t) \leq c \exp\left(\int_0^t v(s)\mathrm{d}s\right)$$

for all $0 \leq t \leq T$.

Lemma 4.1.4 (Bihari's Inequality). *Let $T > 0$ and $c > 0$. Let $K : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous non-decreasing function such that $K(t) > 0$ for all $t > 0$. Let $u(\cdot)$ be a Borel measurable bounded non-negative function on $[0, T]$, and let $v(\cdot)$ be a non-negative integrable function on $[0, T]$. If*

$$u(t) \leq c + \int_0^t v(s)K(u(s))\mathrm{d}s,$$

for all $0 \leq t \leq T$, then

$$u(t) \leq G^{-1}\left(G(c) + \int_0^t v(s)\mathrm{d}s\right)$$

holds for all such $s \le t \le T$ that satisfies

$$G(c) + \int_0^t v(s)\mathrm{d}s \in Dom(G^{-1}),$$

where

$$G(r) = \int_1^r \frac{\mathrm{d}s}{K(s)}$$

on $r > 0$, and G^{-1} is the inverse function of G.

4.2 Problem Setups and Outlines

It is well-known that if both b and σ satisfies the Lipschitz condition, then the SODE processes a unique solution. To be more generalized, we shall study (22) under the following hypothesis.

Assumption 4.2.1 (Monotonicity)**.** There exists $\lambda_0 \in \mathbb{R}$ such that for all $x, y \in \mathbb{R}^d$,

$$2\langle x - y, b(x) - b(y)\rangle + \|\sigma(x) - \sigma(y)\|^2 \le \lambda |x - y|^2 (1 \vee \log |x - y|^{-1}).$$

Assumption 4.2.2 (Non-degenerate of σ)**.** There exists $\lambda_2 \in \mathbb{R}_+$ such that

$$\sup_{x \in \mathbb{R}^d} \left\|\sigma^{-1}(x)\right\| \le \lambda_2.$$

We need the above assumption to prove the uniqueness of invariant measure and the assumption below to prove the existence of invariant measure.

Assumption 4.2.3 (One side growth of b)**.** There exists $p > 2$ and $\lambda_3, \lambda_4 \in \mathbb{R}_+$ such that

$$2\langle x, b(x)\rangle + \|\sigma(x)\|^2 \le -\lambda_3 |x|^p + \lambda_4.$$

We are going to prove several properties for the solution. Firstly in §4.3, we will prove the existence and uniqueness of the solution under Assumption 4.2.1 by *contraction principle*. The estimation is based on a specific type of Bihari's inequality so we shall prove that inequality at the first place. In §4.4, our goal is to prove that the $P_t\varphi$ generated by the solution is indeed a Markov semigroup. We would see that the semigroup property relys

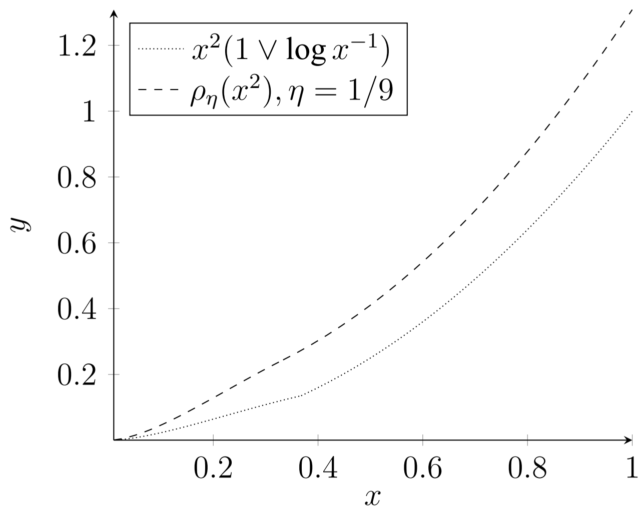

Figure 1 This figure illustrates why we can use $\rho_\eta(x^2)$ to bound $x^2(1 \vee \log x^{-1})$.

on both homogenity and Markov property. In §4.5, strong Feller property and irreducibility are proved under an additional assumption, Assumption 4.2.2. Hasminskii's theorem yields the overlap of these two properties implies the uniqueness of invariant measure. Finally, in §4.6, we prove the existence of invariant measure under another additional assumption, Assumption 4.2.3. Then by Doob's theorem, the unique ergodic measure exists.

4.3 Existence and Uniqueness of the Solution

We choose a perticular class of functions, ρ_η, for K in Bihari's inequality (Lemma 4.1.4). In order to treat the structure in Assumption 4.2.1, we define the following function.

For $0 < \eta < e^{-1}$, define the following concave and increasing function (see Fig. 1 for $\rho_\eta(x^2)$):

$$\rho_\eta(x) = \begin{cases} x \log x^{-1} & 0 < x \le \eta \\ \eta \log \eta^{-1} + (\log \eta^{-1} - 1)(x - \eta) & x > \eta. \end{cases} \tag{24}$$

Lemma 4.3.1 (Bihari's Inequality). *Let $g(s)$ be a strictly positive function on $\mathbb{R}_+$ satisfying for some $\delta > 0$,*

$$g(t) \le g(0) + \delta \int_0^t \rho_\eta(g(s)) \mathrm{d}s$$

for all $t \ge 0$.

Then for all $T > 0$, we have

(i) $g(t) \le g(0)^{\exp(-\delta T)}$ *if* $g(0) < \eta^{\exp(\delta T)}$;

(ii) $g(t) \le C(g(0)^{\exp(-\delta T)} + g(0))$*, for some* $C = C(T, \delta, \eta)$

for all $t \in [0, T]$.

Note that if $\delta \le 0$*, then trivially* $g(t) \le g(0)$ *for all* $t \in [0, T]$.

In the following, when we refer to Bihari's inequality, it means the above inequality instead of the original one.

Proof. For (i), we are going to use Bihari's inequality with $K = \rho_\eta \mathbb{1}_{(0,\eta]}$. Then

$$G(x) = \int_1^x \frac{\mathrm{d}s}{\rho_\eta(s)} = -\int_x^\eta \frac{\mathrm{d}s}{\rho_\eta(s)} \equiv \log\left(\frac{\log \eta}{\log x}\right).$$

Then $Dom(G^{-1}) = (-\infty, 0)$ and

$$G^{-1}(x) = \exp\left\{\log \eta \exp(-x)\right\}.$$

Direct calculation shows

$$G^{-1}(G(g(0)) + \delta t) = g(0)^{\exp(-\delta t)}.$$

Note that the condition $g(0) < \eta^{\exp(\delta T)}$ implies $G(g(0)) + \delta t < 0$. The result then follows by Bihari's inequality.

For (ii), it remains to consider $g(0) \ge \eta^{\exp(\delta T)}$. Then

$$\begin{aligned}\rho_\eta(x) &\le \eta \log \eta^{-1} + (\log \eta^{-1} - x)x \\ &\le g(0)^{\exp(-\delta T)} \log \eta^{-1} + (\log \eta^{-1} - x)x.\end{aligned}$$

So

$$g(t) \le g(0) + T\delta g(0)^{\exp(-\delta T)} \log \eta^{-1} + \delta(\log \eta^{-1} - x) \int_0^t g(s)\mathrm{d}s.$$

Gronwall's inequality yields the result. □

Note that if we apply Itô's formula to $|Y_t|^2$, we obtain

$$\mathrm{d}|Y_t|^2 = \left(\langle b(Y_s), Y_s\rangle + \|\sigma(Y_s)\|^2\right)\mathrm{d}s + 2\left\langle Y_s, \sigma(Y_s)\right\rangle \mathrm{d}W_s, \tag{25}$$

if symbolically $\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t$.

Theorem 4.3.2. *Let $\eta \in L^2(\Omega, \mathscr{F}_s)$. Under Assumption 4.2.1, SODE 22 with $X(s) = \eta$ processes an unique solution X_t. Moreover, $X_t \in \mathscr{M}^2(s,T)$.*

Proof. The idea of our proof is to use a fixed point argument in the space $\mathscr{M}^2([s,T])$. Define

$$\gamma(t,X) \stackrel{\text{def}}{=} \eta + \int_s^t b(X_u)\mathrm{d}u + \int_s^t \sigma(X_u)\mathrm{d}W_u \tag{26}$$

for $X \in \mathscr{M}^2([s,T])$, $t \in [s,T]$. Then it is a solution of (22) iff it is a fixed point of γ: $X = \gamma(X)$. Firstly we are going to show γ maps $\mathscr{M}^2(s,T)$ into itself, then that it is a 0-contraction. The result then follows by the contraction principle (Theorem D.2, [Da Prato, 2014]).

1. Similar to (25), Itô's formula yields that

$$|\gamma(t,X)|^2 = |\eta|^2 + \int_0^t \left(2\left\langle b(X_s), X_s\right\rangle + \|\sigma(X_s)\|^2\right)\mathrm{d}s + 2\int_0^t \left\langle X_s, \sigma(X_s)\mathrm{d}W_s\right\rangle.$$

   By Assumption 4.2.1,

$$|\gamma(t,X)|^2 \leq |\eta|^2 + \lambda_0 \int_0^t |X_s|^2(1 \vee \log|X_s|^{-1}) + 2\int_0^t \left\langle X_s, \sigma(X_s)\mathrm{d}W_s\right\rangle.$$

   From Figure 1, there exists $r^2(1 \vee \log r^{-1}) \leq \rho_\eta(r^2)$, so that

$$|\gamma(t,X)|^2 \leq |\eta|^2 + \lambda_0 \int_0^t \rho_\eta(|X_s|^2) + 2\int_0^t \left\langle X_s, \sigma(X_s)\mathrm{d}W_s\right\rangle.$$

   Now use the stopping time argument. Define

$$\tau_n \stackrel{\text{def}}{=} \{t \in [0,T] : |X_t| \geq n\}$$

   and replace t by $t \wedge \tau_n$. It is clear by the a.s. boundness of X_t on $[0,T]$ that $\tau_n \to T$ a.s. Then take expectation and apply Jensen's inequality with the notice of the concavity of ρ_η,

$$\mathrm{E}\,|\gamma(t\wedge\tau_n, X)|^2 \le \mathrm{E}\,|\eta|^2 + \lambda_0 \int_0^t \rho_\eta(\mathrm{E}\,|X_{s\wedge\tau_n}|^2)\mathrm{d}s.$$

Finally, the result follows by Bihari's inequality, letting $n\to\infty$ and the help of Fatou's lemma.

2. Arguing exactly the same as above except for replacing $\gamma(t,X)$ by $\gamma(t,X)-\gamma(Y)$, where Y is another element in $\mathscr{M}^2(s,T)$, we obtain

$$\mathrm{E}\,|\gamma(t\wedge\tau_n, X)-\gamma(t\wedge\tau_n, Y)|^2 \le \lambda_0 \int_0^t \rho_\eta(\mathrm{E}\,|X_{s\wedge\tau_n}-Y_{s\wedge\tau_n}|^2)\mathrm{d}s.$$

By Bihari's inequality, it follows that

$$\mathrm{E}\,|\gamma(t\wedge\tau_n, X)-\gamma(t\wedge\tau_n, Y)|^2 = 0.$$

Let $n\to\infty$, Fatou's lemma implies

$$\mathrm{E}\,|\gamma(t, X)-\gamma(t, Y)|^2 = 0.$$

Therefore by contraction principle, there exists a unique $X\in\mathscr{M}^2(s,T)$ such that $X(t)=\gamma(t,X(t))$. Moreover, $t\mapsto X(t)$ is continuous. Therefore $b(X_t)\in\mathscr{L}^1([s,T])$ and $\sigma(X_t)\in\mathscr{L}^2(s,T)$. The uniqueness follows by the standard method using similar argument (we have had shown the uniqueness over $\mathscr{M}^2(s,T)$ only). □

A similar argument yields the following, which I called the *continuity w.r.t. initial value in $L^2(\Omega)$ sense*.

Theorem 4.3.3. *Let $X(t,s,x)$ and $X(t,s,y)$ be the solution of corresponding SODE* (22)*. Then*

$$\mathrm{E}\,|X(t,s,x)-X(t,s,y)|^2 \le |x-y|^{\exp(-\lambda_0 T)}$$

provided that x,y are close enough.

4.4 Homogenity, Markov and Semigroup Property

In this subsection, we wish to prove that

$$P_t\varphi(x) \stackrel{\text{def}}{=} \mathrm{E}[\varphi(X_t^x)]$$

satisfies the semigroup property: $P_s \circ P_t(\varphi) = P_{s+t}(\varphi)$.

Define

$$P_{s,t}\varphi(x) \stackrel{\text{def}}{=} \mathrm{E}[\varphi(X_t^{s,x})].$$

Then $P_t = P_{0,t}$.

The following property is an immediate consequence of uniqueness.

Lemma 4.4.1. *Let $\zeta \in L^2(\Omega, \mathscr{F}_s)$. Then*

$$X(t, s, \zeta) = X(t, r, X(r, s, \zeta))$$

holds for $0 \leq s \leq r \leq t \leq T$.

Proof. Since $X(t, s, \zeta)$ is the solution,

$$\begin{aligned} X(t,s,\zeta) =& \zeta + \int_s^t b(X_u^{s,\zeta})\mathrm{d}u + \int_s^t \sigma(X_u^{s,\zeta})\mathrm{d}W_u \\ =& \zeta + \int_s^r + \int_r^t b(X_u^{s,\zeta})\mathrm{d}u + \int_s^r + \int_r^t \sigma(X_u^{s,\zeta})\mathrm{d}W_u \\ =& X(r,s,\zeta) + \int_r^t b(X_u^{s,\zeta})\mathrm{d}u + \int_r^t \sigma(X_u^{s,\zeta})\mathrm{d}W_u. \end{aligned}$$

From the uniqueness, $X(t, s, \zeta) = X(t, r, X(r, s, \zeta))$. □

A useful relationship between $X(t, s, \eta)$ and $X(t, s, x)$ is given below, where $\eta \in L^2(\Omega, \mathscr{F}_s)$ and $x \in \mathbb{R}^d$.

Lemma 4.4.2. *Assume that Assumption 4.2.1 holds and that*

$$\eta = \sum_{k=1}^n x_k \mathbb{1}_{A_k},$$

where $x_1, \ldots, x_n \in \mathbb{R}^d$ *and* $A_1, \ldots, A_n$ *are mutually disjoint sets in* $\mathscr{F}_s$ *such that* $\Omega = \bigcup_k A_k$.

Then

$$X(t,s,\eta) = \sum_{k=1}^{n} X(t,s,x_k)\mathbb{1}_{A_k}.$$

For a proof, see (Proposition 8.6, [Da Prato, 2014])[10].

We have the following preparation lemma for the proof of Markov property.

Lemma 4.4.3. *For all* $\varphi \in \mathbb{B}_b(\mathbb{R}^d)$ *and all* $\eta \in L^2(\Omega, \mathscr{F}_s)$*, we have*

$$\mathrm{E}[\varphi(X(t,s,\eta)) \mid \mathscr{F}_s] = P_{s,t}\varphi(\eta)$$

for $0 \le s < t \le T$*. Consequently,*

$$\mathrm{E}[\varphi(X(t,s,\eta))] = \mathrm{E}[P_{s,t}\varphi(\eta)].$$

Proof. [Da Prato, 2014]. Since the class of simple functions is dense in $L^2(\Omega, \mathscr{F}_s)$, $C_b(\mathbb{R}^d)$ is dense in $\mathbb{B}_b(\mathbb{R}^d)$, it is enough to take η of the form

$$\eta = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}$$

where $x_1, \ldots, x_n \in \mathbb{R}^d$ and $A_1, \ldots, A_n$ are mutually disjoint sets in $\mathscr{F}_s$ such that $\Omega = \bigcup_k A_k$. Once we have shown this, then we can find simple functions $\eta_n \to \eta$ for all ω satisfying

$$\mathrm{E}[\varphi(X(t,s,\eta_n)) \mid \mathscr{F}_s] = P_{s,t}\varphi(\eta_n).$$

Assume $\varphi \in C_b(\mathbb{R}^d)$. As we have shown the continuity of $X(t,s,x)$ w.r.t. x in L^2 sense, there exists a subsequence $\{n_k\}$ such that $X(t,s,\eta_n)$ converges to $X(t,s,\eta)$ a.s. Let $k \to \infty$, the result follows by bounded convergence theorem.

Now consider such case. By Lemma 4.4.2, we have

$$X(t,s,\eta) = \sum_{k=1}^{n} X(t,s,x_k)\mathbb{1}_{A_k}$$

[10] Although we have different hypotheses to the coefficients of SODE, the map γ defined in (26) are both contractions. Therefore the lemma holds in our situation.

for $0 \leq s \leq t \leq T$. Consequently,

$$\varphi(X(t,s,\eta)) = \sum_{k=1}^{n} \varphi(X(t,s,x_k))\mathbb{1}_{A_k}$$

since their domains are disjoint, which implies

$$\mathrm{E}[\varphi(X(t,s,\eta)) \mid \mathscr{F}_s] = \sum_{k=1}^{n} \mathrm{E}[\varphi(X(t,s,x_k))\mathbb{1}_{A_k} \mid \mathscr{F}_s].$$

Since $\mathbb{1}_{A_k}$ is $\mathscr{F}$-measurable and $\varphi(X(t,s,x_k))$ is independent of $\mathscr{F}_s$, we have

$$\mathrm{E}[\varphi(X(t,s,x_k))\mathbb{1}_{A_k} \mid \mathscr{F}_s] = P_{s,t}\varphi(x_k)\mathbb{1}_{A_k}$$

by the property of conditional expectation. In conclusion,

$$\mathrm{E}[\varphi(X(t,s,\eta)) \mid \mathscr{F}_s] = P_{s,t}\varphi(\eta). \qquad \square$$

Theorem 4.4.4. *Let* $0 \leq s \leq r \leq t \leq T$ *and* $\varphi \in \mathbb{B}_b(\mathbb{R}^d)$. *Then we have*

$$P_{s,t}\varphi(x) = \mathrm{E}[P_{r,t}\varphi(X(r,s,x))].$$

In other words, $P_{s,t}\varphi = P_{s,r}P_{r,t}\varphi$.

Proof. By Lemma 4.4.3, we have

$$\mathrm{E}[P_{r,t}\varphi(X(r,s,x))] = \mathrm{E}[\varphi(X(t,r,X(r,s,x)))] = \mathrm{E}[\varphi(X(t,s,x))] = P_{s,t}\varphi(x).$$

Since $\mathrm{E}[P_{r,t}\varphi(X(r,s,x))] = P_{s,r}[P_{r,t}\varphi(x)]$, the result follows. $\square$

Theorem 4.4.5 (Markov Property)**.** *Let* $0 \leq s < r < t \leq T$ *and let* $\eta \in L^2(\Omega, \mathscr{F}_s)$. *Then for all* $\varphi \in \mathbb{B}_b(\mathbb{R}^d)$ *we have*

$$\mathrm{E}[\varphi(X(t,s,\eta)) \mid \mathscr{F}_r] = P_{r,t}\varphi(X(r,s,\eta)).$$

Proof. Set $\zeta = X(r, s, \eta)$. Then by Lemma 4.4.3, using Lemma 4.4.1,

$$\begin{aligned}\mathrm{E}[\varphi(X(t, s, \eta)) \mid \mathscr{F}_r] =& \mathrm{E}[\varphi(X(t, s, X(r, s, \eta))) \mid \mathscr{F}_r] \\ =& \mathrm{E}[\varphi(X(t, r, \zeta)) \mid \mathscr{F}_r] = P_{t,r}\varphi(\zeta)\end{aligned}$$

and the conclusion follows. □

The solution is *time-homogeneous* in the following sense.

Theorem 4.4.6. *The solution $X_t^{s,x}$ is time-homogeneous, i.e. $\{X_{s+h}^{s,x}\}$ and $\{X_h^{0,x}\}$ have the same distribution. In other words, $P_{s,s+h} = P_{0,h} = P_h$.*

Proof. [Øksendal, 2003]. On one hand,

$$\begin{aligned}X_{s+h}^{s,x} =& x + \int_s^{s+h} b(X_u^{s,x})\mathrm{d}u + \int_s^{s+h} \sigma(X_u^{s,x})\mathrm{d}W_u \\ & \text{Let } v = u - s \text{ or } u = v + s \\ =& x + \int_0^h b(X_{v+s}^{s,x})\mathrm{d}v + \int_0^h \sigma(X_{v+s}^{s,x})\mathrm{d}W_{v+s} \\ & \text{Let } \tilde{W}_v = W_{v+s} - W_s. \text{ Check that } \Delta_k \tilde{W}_v = \Delta_k W_{v+s} \\ =& x + \int_0^h b(X_{v+s}^{s,x})\mathrm{d}v + \int_0^h \sigma(X_{v+s}^{s,x})\mathrm{d}\tilde{W}_v.\end{aligned}$$

Here $\tilde{W}_v$ is a Brownian motion started at 0 a.s. On the other hand,

$$X_h^{0,x} = x + \int_0^h b(X_v^{0,x})\mathrm{d}v + \int_0^h \sigma(X_v^{s,x})\mathrm{d}W_v.$$

As W_v and $\tilde{W}_v$ have the same distribution, $\{X_{s+h}^{s,x}\}$ and $\{X_h^{0,x}\}$ also have the same distribution by the uniqueness of the solution. □

Theorem 4.4.7. *P_t defines a Markov semigroup (not necessarily strongly continuous).*

Proof. We have shown that $P_{0,s+t}\varphi = P_{0,s}P_{s,s+t}\varphi$ in Theorem 4.4.4. By homogenity, $P_{s,s+t} = P_t$ and the conclusion follows. □

4.5 Uniqueness of Invariant Measure

The proofs in this subsection follow [Zhang, 2009].

4.5.1 Strong Feller Property

For convenience, we denote $z/|z|$ by $\overline{z}$ for $z \neq 0$.

Proof of Strong Feller Property. The proof of strong Feller property consists of three steps. In Step 1, we prove that the coupling equation

$$\begin{cases} \mathrm{d}Y(t) = b(X(t))\mathrm{d}t + a(X(t) - Y(t)) \cdot \mathbb{1}_{t<\tau}\mathrm{d}t + \sigma(Y(t))\mathrm{d}W(t) \\ Y(0) = y_0, y_0 \in \mathbb{R}^d, \end{cases} \tag{27}$$

where

$$a(z) \overset{\text{def}}{=} |x_0 - y_0|^{\alpha} \cdot \mathbb{1}_{z\neq 0} \cdot \overline{z}$$

called the *coupling function* and

$$\tau \overset{\text{def}}{=} \inf\{t > 0 : |X(t) - Y(t)| = 0\}$$

called the *coupling time*, is solvable. In Step 2, we use Itô's formula and Lemma to estimate the coupling time. In the last step, we use Girsanov's theorem (Theorem 8.9.4, [Kuo, 2006]) to find the estimate of

$$|P_T\varphi(x_0) - P_T\varphi(y_0)|\,. \tag{28}$$

Now we start the proof.

1. Considering the following equation

$$\begin{cases} \mathrm{d}Y_t^{\epsilon} = b(Y_t^{\epsilon})\mathrm{d}t + a_{\epsilon}(X_t - Y_t^{\epsilon})\mathrm{d}t + \sigma(Y_t^{\epsilon})\mathrm{d}W_t \\ Y_0^{\epsilon} = y_0, y_0 \in \mathbb{R}^d, \end{cases} \tag{29}$$

   where

$$a_{\epsilon}(z) = |x_0 - y_0|^{\alpha} \cdot f_{\epsilon}(|z|) \cdot \overline{z},$$

   $f_{\epsilon} : \mathbb{R}_+ \to [0, 1]$ is smooth and equals 1 when $r > \epsilon$; equals 0 when $r \in [0, \epsilon/2]$. Then the SODE (29) possesses a unique solution since

$$|a_{\epsilon}(z) - a_{\epsilon}(z')| \leq C_{\epsilon}|z - z'|.$$

The reason is that $z \mapsto \overline{z}$ is $4/\epsilon$-Lipschitz when $z > \epsilon/2$:

$$\begin{aligned}
|\overline{z} - \overline{z'}| &= \left| \frac{z}{|z|} - \frac{z'}{|z'|} \right| \\
&= \left| \frac{z}{|z|} - \frac{z'}{|z|} + \frac{z'}{|z|} - \frac{z'}{|z'|} \right| \\
&\leq \frac{1}{|z|}|z - z'| + |z'| \frac{||z'| - |z||}{|z||z'|} \\
&\leq \frac{4}{\epsilon}|z - z'|.
\end{aligned}$$

Therefore we have the solution Y_t^ϵ. Define

$$\tau_\epsilon \stackrel{\text{def}}{=} \int \{t > 0 : |X_t - Y_t^\epsilon| \leq \epsilon\}.$$

Then for any $\epsilon' < \epsilon$, we have $Y_t^{\epsilon'} = Y_t^\epsilon$ when $t < \tau_\epsilon$ by uniqueness. Comparing (27) and (29), we have $Y_t = Y_t^\epsilon$ when $t < \tau_\epsilon$. Hence $\tau = \lim_{\epsilon \downarrow 0} \tau_\epsilon$. Then Y_t is well-defined on $t < \tau$. When $t \in [\tau, T]$, let $Y_t = X_t$. Then it is clear that Y_t solves (27).

2. Let $Z_t = X_t - Y_t$. Apply Itô formula to the function $r \mapsto \sqrt{|r|^2 + \epsilon}$ and let $\epsilon \to 0$. Then

$$\begin{aligned}
&|Z_{t\wedge\tau}| - |x_0 - y_0| - \int_0^{t\wedge\tau} \left\langle \overline{Z}_s, (\sigma(X_s) - \sigma(Y_s))\mathrm{d}W_s \right\rangle \\
=& \int_0^{t\wedge\tau} (2|Z_s|)^{-1} \cdot \left(2\left\langle Z_s, b(X_s) - b(Y_s)\right\rangle + \|\sigma(X_s) - \sigma(Y_s)\|^2\right) \mathrm{d}s \\
&- \int_0^{t\wedge\tau} \left\langle \overline{Z}_s, a(Z_s) \right\rangle \mathrm{d}s - \int_0^{t\wedge\tau} (2|Z_s|)^{-1} \cdot |[\sigma(X_s) - \sigma(Y_s)]^*(\overline{Z}_s)|^2 \mathrm{d}s \\
\leq& \frac{\lambda_0}{2} \int_0^{t\wedge\tau} |Z_s|(1 \vee \log |Z_s|^{-1})\mathrm{d}s - |x_0 - y_0|^\alpha (t \wedge \tau).
\end{aligned}$$

Note that there exists an $0 < \eta < e^{-1}$ such that

$$r(1 \vee \log r^{-1}) \leq \rho_\eta(r)$$

for all $r > 0$. Taking expectations yields that

$$\begin{aligned}\mathrm{E}\,|Z_{t\wedge\tau}| \le & |x_0 - y_0| - |x_0 - y_0|^{\alpha} \cdot \mathrm{E}(t \wedge \tau) + \frac{\lambda_0}{2}\,\mathrm{E}\int_0^{t\wedge\tau} \rho_\eta(|Z_s|)\mathrm{d}s \\ \le & |x_0 - y_0| - |x_0 - y_0|^{\alpha} \cdot \mathrm{E}(t \wedge \tau) + \frac{\lambda_0}{2}\int_0^{t\wedge\tau} \rho_\eta(\mathrm{E}\,|Z_{s\wedge\tau}|)\mathrm{d}s,\end{aligned}$$

where the second step is due to Jensen's inequality.

Using Bihari inequality, we get that for any $t > 0$ and $|x_0 - y_0| < \eta^{\lambda_0 T/2} \wedge \eta$,

$$\mathrm{E}\,|Z_{t\wedge\tau}| \le |x_0 - y_0|^{\exp(-\lambda_0 t/2)},$$

where we also use the fact that ρ_η is increasing. Then

$$\mathrm{E}(t \wedge \tau) \le |x_0 - y_0|^{1-\alpha} + \frac{\lambda_0 t}{2}\rho_\eta(|x_0 - y_0|^{\exp(-\lambda_0 t/2)}) \cdot |x_0 - y_0|^{-\alpha}. \tag{30}$$

3. Let

$$R_T = \exp\left[\int_0^{T\wedge\tau} H(X_s, Y_s)\mathrm{d}W_s - \frac{1}{2}\int_0^{T\wedge\tau} |H(X_s, Y_s)|^2\mathrm{d}s\right]$$

and

$$\tilde{W}_t = W_t + \int_0^{t\wedge\tau} H(X_s, Y_s)\mathrm{d}s,$$

where $H(x, y) = |x_0 - y_0|^{\alpha} \cdot [\sigma(y)]^{-1}\overline{x - y}$. Then

$$|H(x, y)|^2 \le |x_0 - y_0|^{2\alpha} \|\sigma(y)\|^{-2} \le |x_0 - y_0|^{2\alpha} \cdot \lambda_3^2.$$

By Novikov condition (Remark 8.7.4, [Kuo, 2006]), $\mathrm{E}\,R_T = 1$ and

$$\mathrm{E}\,R_T^2 \le \exp(T\lambda_3^2\,|x_0 - y_0|^{2\alpha}).$$

Then

$$\begin{aligned}&|P_T\varphi(x_0) - P_T\varphi(y_0)| \\ =&|\mathrm{E}[\varphi(X_T^{x_0})] - \mathrm{E}[\varphi(X_T^{y_0})]| = |\mathrm{E}[\varphi(X_T^{x_0})] - \mathrm{E}[\varphi(Y_T^{y_0})]| \\ =&\mathrm{E}\,|[\varphi(X_T^{x_0}) - R_T\varphi(Y_T^{y_0})] \cdot \mathbb{1}_{T\ge\tau}| + \mathrm{E}\,|[\varphi(X_T^{x_0}) - R_T\varphi(Y_T^{y_0})] \cdot \mathbb{1}_{T<\tau}|\,.\end{aligned}$$

We have

$$\begin{aligned}\mathrm{E}\,|[\varphi(X_T^{x_0}) - R_T\varphi(Y_T^{y_0})] \cdot \mathbb{1}_{T\geq\tau}| =& \mathrm{E}\,|[\varphi(X_T^{x_0}) - R_T\varphi(X_T^{y_0})] \cdot \mathbb{1}_{T\geq\tau}| \\ \leq& \|\varphi\|_0 \cdot \mathrm{E}\,|1 - R_T|.\end{aligned}$$

Since

$$\begin{aligned}(\mathrm{E}\,|1 - R_T|)^2 =& \mathrm{E}\,R_T^2 - 1 \\ \leq& \exp(T\lambda_3^2|x_0 - y_0|^{2\alpha}) - 1 \\ \leq& T\lambda^2|x_0 - y_0|^{2\alpha}\exp(T\lambda|x_0 - y_0|^{2\alpha}) \\ =& C_{T,\lambda,\eta'} \cdot |x_0 - y_0|^{2\alpha}\end{aligned}$$

for $|x_0 - y_0| < \eta'$ (as α will be chosen w.r.t. λ_0 and T, we omit it from the subscript of C), we obtain the estimate for the first term. For the second term,

$$\begin{aligned}(\mathrm{E}[(1 + R_T) \cdot \mathbb{1}_{\tau\geq T}])^2 \leq& (\mathrm{E}\,|1 + R_T|^2) \cdot \mathrm{P}(\tau \geq T) \\ =& (3 + \mathrm{E}\,R_T^2)\,\mathrm{P}((2T \wedge \tau) \geq T) \\ =& C_{T,\lambda_0,\eta''}\,\mathrm{E}(2T \wedge \tau).\end{aligned}$$

By L' Hospital's thoerem,

$$\mathrm{E}(2T \wedge \tau) \leq C|x_0 - y_0|^{\exp(-\lambda_0 T/2)/2}.$$

Combining two estimation, we obtain

$$|P_T\varphi(x_0) - P_T\varphi(y_0)| \leq C_{T,\lambda,\eta} \cdot |x_0 - y_0|^{\exp(-\lambda_0 T/2)/4}.$$

Thus P_T is strong Feller.

□

4.5.2 Irreducibility

For proving the irreducibility of P_t, it means to prove that for any $x_0 \in \mathbb{R}^d$, $T > 0$ and $y_0 \in \mathbb{R}^d$, $a > 0$,

$$P_T(x_0, B(y_0, a)) = \mathrm{P}(|X_T(x_0) - y_0| \leq a) > 0.$$

Proof of Irreducibility. Let $t_1 \in (0, T)$, whose value will be determined below. Let $\epsilon > 0$. Set

$$X^\epsilon_{t_1} \stackrel{\text{def}}{=} X_{t_1} \cdot \mathbb{1}_{|X_{t_1}| \le \epsilon^{-1}}.$$

Then

$$\lim_{\epsilon \downarrow 0} \mathrm{E}\, |X^\epsilon_{t_1} - X_{t_1}|^2 = 0.$$

Define Y_s for $s \in [t_1, T]$ as the following:

$$Y^\epsilon_s = \frac{T-s}{T-t_1} X^\epsilon_{t_1} + \frac{s-t_1}{T-t_1} y_0$$

satisfies $Y^\epsilon_{t_1} = X^\epsilon_{t_1}$, $Y^\epsilon_T = y_0$ and the following relation:

$$Y^\epsilon_t = X^\epsilon_{t_1} + \int_{t_1}^t b(Y^\epsilon_s) \mathrm{d}s + \int_{t_1}^t h^\epsilon_s \mathrm{d}s$$

for $t \in [t_1, T]$, where

$$h^\epsilon_s \stackrel{\text{def}}{=} \frac{y_0 - X^\epsilon_{t_1}}{T - t_1} - b(Y^\epsilon_s).$$

Consider the following SODE on $[t_1, T]$:

$$X^\epsilon_t = X_{t_1} + \int_{t_1}^t b(X^\epsilon_s) \mathrm{d}s + \int_{t_1}^t h^\epsilon_s \mathrm{d}s + \int_{t_1}^t \sigma(X^\epsilon_s) \mathrm{d}W_s.$$

If we define

$$X^\epsilon_t = X_t$$

for $t \in [0, t_1]$, then for any $t \in [0, T]$,

$$X^\epsilon_t = x_0 + \int_0^t b(X^\epsilon_s) \mathrm{d}s + \int_0^t h^\epsilon_s \mathbb{1}_{s > t_1} \mathrm{d}s + \int_0^t \sigma(X^\epsilon_s) \mathrm{d}W_s.$$

Now define

$$\tilde{W}^\epsilon_t = W_t + \int_0^t H^\epsilon_s \mathrm{d}s$$

and

$$R^\epsilon_T = \exp\left[\int_0^T \langle \mathrm{d}W_s, H^\epsilon_s \rangle - \frac{1}{2} \int_0^T |H^\epsilon_s|^2 \mathrm{d}s \right],$$

where

$$H_s^\epsilon \overset{\text{def}}{=} \mathbb{1}_{s>t_1}[\sigma(X_s^\epsilon)]^{-1}h_s^\epsilon.$$

Note that by Assumption 4.2.2 and the continuity of b,

$$|H_s^\epsilon| \le \lambda_2 |h_s^\epsilon| \le C_{\lambda_2,\epsilon,t_1}.$$

By Noviki condition, $\mathrm{E}\, R_T^\epsilon = 1$, $\mathrm{P}(R_T^\epsilon > 0) = 1$ and $\tilde{W}_t^\epsilon$ is a d-dimensional Brownian motion under $R_T^\epsilon \cdot \mathrm{P}$. Thus $X_T^\epsilon(x_0)$ has the same law as $X_T(x_0)$ in different probability measure. Due to the equivalence, it suffices to show

$$\mathrm{P}(|X_T^\epsilon(x_0) - y_0| > a) < 1.$$

Set $Z_t^\epsilon = X_t^\epsilon - Y_t^\epsilon$. By Itô's formula, we have

$$\begin{aligned}\mathrm{E}\,|Z_t^\epsilon|^2 &= \mathrm{E}\,|X_{t_1} - X_{t_1}^\epsilon|^2 + \int_{t_1}^t \mathrm{E}\left(2\left\langle Z_s^\epsilon, b(X_s^\epsilon) - b(X_s^\epsilon)\right\rangle + \|\sigma(X_s^\epsilon)\|^2\right) \mathrm{d}s \\ &= \mathrm{E}\,|X_{t_1} - X_{t_1}^\epsilon|^2 + C_a \int_{t_1}^t \mathrm{E}(|Y_s^\epsilon|^2 + 1)\mathrm{d}s + C_{\lambda_0,a}\int_{t_1}^t \mathrm{E}(|Z_s^\epsilon|^2(1 \vee \log|Z_s^\epsilon|^{-1}))\mathrm{d}s.\end{aligned}$$

We have,

$$\begin{aligned}\int_{t_1}^t \mathrm{E}\,|Y_s^\epsilon|^2 \mathrm{d}s \le& 2(T-t_1)(\mathrm{E}\,|X_{t_1}^\epsilon|^2 + |y_0|^2) \\ \le& 2(T-t_1)(C_{T,x_0,\lambda_0,\lambda_1} + |y_0|^2).\end{aligned}$$

By Bihari's inequality,

$$\mathrm{E}\,|X_T^\epsilon - y_0|^2 \le \left[\mathrm{E}\,|X_{t_1} - X_{t_1}^\epsilon|^2 + C(T-t_1)\right]^{\exp(-C_{\lambda_0,a}T)}.$$

Hence

$$\begin{aligned}\mathrm{P}(|X_T^\epsilon(x_0) - y_0| > a) &\le \frac{1}{a^2}\,\mathrm{E}\,|X_T^\epsilon(x_0) - y_0|^2 \\ &\le \frac{1}{a^2}\left[\mathrm{E}\,|X_{t_1} - X_{t_1}^\epsilon|^2 + C(T-t_1)\right]^{\exp(-C_{\lambda_0,a}T)}.\end{aligned}$$

Let t_1 close to T and choose ϵ to be sufficiently small, the result follows. □

4.6 Existence of Invariant Measure

The following theorem is a consequence of Krylov-Bogoliubov theorem, which is frequently used to find invariant measures.

Theorem 4.6.1. *Let H be a Hilbert space. Assume there exists some $x_0 \in H$ and a constant $C = C(x_0) > 0$ such that*

$$\frac{1}{t}\int_0^t \mathrm{E}[V(X_{x_0}(t))] \leq C(x_0)$$

for all $t \geq 0$, where $V : H \to [1, \infty]$ is a Borel function whose level sets

$$K_\alpha \overset{def}{=} \{x : V(x) \leq \alpha\}$$

are compact for any $\alpha > 0$. Then there exists an invariant measure for X.

Proof. Recall the definition of μ_T in Theorem 3.3.3. Given $\epsilon > 0$, let $a(\epsilon) = C(x_0)/\epsilon$, then the level set $K_{a(\epsilon)}$ satisfies

$$\begin{aligned}
\mu_T(x_0, K_{a(\epsilon)^c}) =& \frac{1}{T}\int_0^T \int_{V(y)>a(\epsilon)} P_t(x_0, \mathrm{d}y)\mathrm{d}t \\
\leq& \int P_t(x_0, \mathrm{d}y)\frac{1}{T}\int_0^T \left[\frac{V(y)}{a(\epsilon)}\right] \mathrm{d}t \\
=& \int P_t(x_0, \mathrm{d}y)\frac{1}{a(\epsilon)}\frac{1}{T}\int_0^T \mathrm{E}[V(X_{x_0}(t))]\mathrm{d}t \leq \epsilon.
\end{aligned}$$

Hence $\{\mu_t(x_0, \cdot)\}$ is tight, which ensures the existence of invariant measures. □

Remark 4.6.2. When $H = \mathbb{R}^d$, the condition $\lim_{x\to\infty} V(x) = \infty$ could replace the compactness of level sets in Theorem 4.6.1. The reason is that $\lim_{x\to\infty} V(x) = \infty$ means that for every $M > 0$, there exists $R > 0$ such that for $|x| > R$ we have $V(x) > M$, so that we can always find $\overline{B_R(x)}^c \subseteq \{V(x) > M\}$. Consequently, we can choose $\overline{B_R(x)}$ instead of $\{V(x) \leq M\}$ in the proof of each statement.

Proof of Uniqueness. Using Itô's formula, we have by Assumption 4.2.3 and Hölder's inequality

$$\begin{aligned}\frac{\mathrm{d}\,\mathrm{E}\,|X_t|^2}{\mathrm{d}t} &= \mathrm{E}(2\,\langle X_t, b(X_t)\rangle + \|\sigma(X_t)\|^2) \\ &\leq -\lambda_3\,\mathrm{E}\,|X_t|^p + \lambda_4 \\ &\leq -\lambda_3\left(\mathrm{E}\,|X_t|^2\right)^{p/2} + \lambda_4.\end{aligned}$$

Hence for all $t > 0$

$$\frac{1}{t}\int_0^t \mathrm{E}\,|X_s|^2 \leq \lambda_4.$$

The result follows by Theorem 4.6.1. □

We summarize our results as the following theorem using Doob's theorem.

Theorem 4.6.3. *Assume Assumption 4.2.1-4.2.2 holds. Then the semigroup of the solution of SODE* (22) *is strong Feller and irreducible. If in addition, Assumption 4.2.3 holds, then the solution is strongly mixing thus erogdic.*

（This is the end of the thesis, 本文完）

参考文献

[1] Da Prato G, Zabczyk J. Ergodicity for infinite-dimensional systems[M/OL]. Cambridge University Press, Cambridge, 1996: xii+339. https://doi.org/10.1017/CBO9780511662829. DOI: 10.1017/CBO9780511662829.

[2] Douc R, Moulines E, Priouret P, et al. Markov chains[M/OL]. Springer, Cham, 2018: xviii+757. https://doi.org/10.1007/978-3-319-97704-1. DOI: 10.1007/978-3-319-97704-1.

[3] Ash R B. Probability and measure theory[M]. Second. Harcourt/Academic Press, Burlington, MA, 2000: xii+516.

[4] Da Prato G. An introduction to infinite-dimensional analysis[M/OL]. Springer-Verlag, Berlin, 2006: x+209. https://doi.org/10.1007/3-540-29021-4. DOI: 10.1007/3-540-29021-4.

[5] Robinson J C. An Introduction to Functional Analysis[M]. Cambridge University Press, 2020. DOI: 10.1017/9781139030267.

[6] Wiener N. Differential-space[J]. J. Math. and Phys., 1923, 2: 131-174.

[7] Lévy P. Sur certains processus stochastiques homogènes[J/OL]. Compositio Math., 1939, 7: 283-339. http://www.numdam.org/item?id=CM_1940__7__283_0.

[8] Kuo H H. Introduction to stochastic integration[M]. Springer, New York, 2006: xiv+278.

[9] Evans L C. An introduction to stochastic differential equations[M/OL]. American Mathematical Society, Providence, RI, 2013: viii+151. https://doi.org/10.1090/mbk/082. DOI: 10.1090/mbk/082.

[10] Kallenberg O. Foundations of modern probability[M/OL]. Springer, Cham, 2021: xii+946. https://doi.org/10.1007/978-3-030-61871-1. DOI: 10.1007/978-3-030-61871-1.

[11] Jacod J, Protter P. Probability essentials[M/OL]. Second. Springer-Verlag, Berlin, 2003: x+254. https://doi.org/10.1007/978-3-642-55682-1. DOI: 10.1007/978-3-642-55682-1.

[12] Karatzas I, Shreve S E. Brownian motion and stochastic calculus[M/OL]. Second. Springer-Verlag, New York, 1991: xxiv+470. https://doi.org/10.1007/978-1-4612-0949-2. DOI: 10.1007/978-1-4612-0949-2.

[13] Øksendal B. Stochastic differential equations[M/OL]. Sixth. Springer-Verlag, Berlin, 2003: xxiv+360. https://doi.org/10.1007/978-3-642-14394-6. DOI: 10.1007/978-3-642-14394-6.

[14] Itô K. Stochastic integral[J/OL]. Proc. Imp. Acad. Tokyo, 1944, 20: 519-524. http://projecteuclid.org/euclid.pja/1195572786.

[15] Mao X. Stochastic differential equations and applications[M/OL]. Second. Horwood Publishing Limited, Chichester, 2008: xviii+422. https://doi.org/10.1533/9780857099402. DOI: 10.1533/9780857099402.

[16] Yosida K. Functional analysis[M/OL]. Springer-Verlag, Berlin, 1995: xii+501. https://doi.org/10.1007/978-3-642-61859-8. DOI: 10.1007/978-3-642-61859-8.

[17] Da Prato G. Introduction to stochastic analysis and Malliavin calculus[M/OL]. Third. Edizioni della Normale, Pisa, 2014: xviii+279. https://doi.org/10.1007/978-88-7642-499-1. DOI: 10.1007/978-88-7642-499-1.

[18] Zhang X. Exponential ergodicity of non-Lipschitz stochastic differential equations[J/OL]. Proc. Amer. Math. Soc., 2009, 137(1): 329-337. https://doi.org/10.1090/S0002-9939-08-09509-9. DOI: 10.1090/S0002-9939-08-09509-9.

致谢

本篇论文的部分内容是我在刘智慧老师讨论班做讨论时的讲稿，因此首先感谢毕舒琪同学、郑函同学的聆听，同时也感谢他们所做的一些讲解。熊捷教授和 Jana Hertz 教授分别教授的*现代概率论*和*动力系统*两门课程（熊老师还教过我*应用随机过程*，其中的 Markov 链与本文中的 Markov 过程也紧密相连），对于我理解概率论、遍历论有着不小的帮助，尽管在 Jana 教授的课中我差点挂科，也在此表示感谢。另外，还应感谢于跃同学（我一般称他为老师）。我们讨论的*贝叶斯统计*和*最优运输*都与本论文探讨的问题有着一定的联系，虽然这个联系不是显式的。

除去我本人之外，对这篇论文影响最大，也是最应该感谢的，是我的导师刘智慧教授。从选题开始贯穿讨论班再到最后的论文结构，无不渗透着他的帮助。如果要追溯的话，那么甚至可以从刘老师在我大二时讲授的*初等概率论*开始，包括给担任*概率论与数理统计*助教的机会，选课指导等等十分久远的事情。当然，最直接的帮助是他在本学期教授的*随机分析及其在金融学中的应用*这门课。这篇论文的核心证明中使用的 Girsanov 定理，就是我在刘老师的这门课中习得的；Itô 积分等基本理论就更不用说了。

最后，感谢数学。它对于我学习能力、逻辑思维乃至为人处事上都有根本性的影响。不论我今后是否能够继续学习这门学科，学习数学和撰写此篇毕业论文的经历，都必将会是我人生道路上的宝贵财富。