SDE 及其数值解的遍历性

Zhizhou Liu (刘之洲)

August 31, 2023

Contents

欲研究 SDE 及其数值解的遍历性，我们需了解对于 Markov chain 或 process 的一般随机遍历性理论，再应用到我们关心的特定 Markov chain 或 process 上。在此之前，需要一些 Markov chain 理论的预备知识，熟悉刻画 Markov chain 的 transition kernel 和对应的 semigroup. 这一部分也可以参考本笔记的 Appendix A.

另外，Da Prato 所著的 [2, 3, 4] 是重要的参考文献。

Notations

Some uncommon notations are introduced here.

- $L(X,Y)$ denotes the space of bounded linear operator from X to Y. If $X=Y$, then simply denoted by $L(X)$.
- $\mathbb{B}_b(H)$ denotes the space of bounded measurable functions from H to $\mathbb{R}$.

- 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.

1 随机遍历性的一般理论

随机遍历性的一般理论旨在给出对于给定 Markov chain 或 process 遍历性的常用充分条件 – 不变测度的存在唯一性。

In §1.1, we breifly introduce the meaning and equivalent characterizations of ergodicity. In §1.2, 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. §1.3 provides some sufficient conditions for the Markov semigroups that process invariant measures and §1.4 for which of processing a unique invariant measure.

1.1 Ergodicity

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

1.1.1 Invariant Measure of Markov Semigroup

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

Definition 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

$$\int_H P_t\phi\mathrm{d}\mu=\int_H\phi\mathrm{d}\mu \tag{1}$$

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

Remark 1.2. It is clear that the above definition is equivalent of saying

$$\mu P_t(A)=\mu(A) \tag{2}$$

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

$$P_t^*\mu=\mu \tag{3}$$

for all $t\in\mathbb{T}$ by Remark A.13.

1.1.2 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 $\phi \in L^2(H,\mu)$, $\lim_{t\to 0} P_t\phi = \phi$) semigroup of $L(L^2(H,\mu))$, [8, p. 381, Theorem 1]. 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 1.3. *Let E be a Hilbert space and T be a bounded linear operator on E. Let*

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

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 [3, Theorem 5.11].

Then apply the result to the average

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

for all $\phi \in L^2(H,\mu)$ and $T > 0$. We obtain the well-knwon Von Neumann's ergodic theorem , [3, Theorem 5.12].

Theorem 1.4 (Von Neumann). *$\lim_{T\to\infty} M(T)\phi$ exists in $L^2(H,\mu)$, denoted by $M_\infty\phi$. Moreover, it is a projection operator on Σ and also*

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

1.1.3 Characterizations of Ergodic Measures

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

Definition 1.5 (ergodic, strongly mixing). Let μ be an invariant measure for P_t. We say that

- μ is *ergodic* iff
$$\lim_{T\to\infty} \frac{1}{T}\int_0^T P_t\phi \mathrm{d}t = \bar{\phi}$$
in $L^2(H,\mu)$-sense for all $\phi \in L^2(H,\mu)$,
- μ is *strongly mixing* iff
$$\lim_{T\to\infty} P_t\phi = \bar{\phi}$$
in $L^2(H,\mu)$-sense for all $\phi \in L^2(H,\mu)$,

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

Remark 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 [4, Subsection 12.4.3].

Let Σ of be the sets of *stationary points*

$$\Sigma \overset{\text{def}}{=} \{\phi \in L^2(H,\mu) : P_t\phi = \phi\} \tag{4}$$

Definition 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 1.8. *Let μ be an invariant measure for P_t. Then following statements are equivalent:*

(i) μ is ergodic.

(ii) The dimension of the linear space Σ of stationary points in (4) is 1.

(iii) Any invariant set is trivial.

1.2 Structure of the Set of Invariant Measures

Let

$$\Lambda \overset{\text{def}}{=} \{\mu \in \mathbb{B}_b(H)^* : P_t^*\mu = \mu\}. \tag{5}$$

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

Theorem 1.9. *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

$$\mu_\Gamma(A) = \frac{1}{\mu(\Gamma)}\mu(A\cap\Gamma) \tag{6}$$

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. $\square$

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 1.10. *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 $\mu(\Gamma) = \nu(\Gamma)$. $\square$

Definition 1.11 (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 1.12. *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 (6). 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 ergodicity.

$\square$

Theorem 1.13. *If μ and ν are both ergodic, 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 $M \cap N = \emptyset$, i.e. μ and ν are mutually singular. $\square$

1.3 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 1.14 (Feller)**.** Let P_t be a Markov semigroup on H. We say P_t is *Feller* iff $P_t\phi \in C_b(H)$ for any $\phi \in C_b(H)$ and any $t \geq 0$.

Lemma 1.15. *Let $\mu, \nu \in \mathbb{M}_1(H)$ be such that*

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

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

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

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

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 1.16 (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 $\phi \in C_b(H)$, choose $\psi = P_s\phi \in C_b(H)$ by Feller property, then

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

By Lemma 1.15, μ is an invariant measure for P_t. □

1.4 Uniqueness of Invariant Measure

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

Definition 1.17 (strong Feller, irreducible, regular)**.** Let P_t be a Markov semigroup on H.

- P_t is *strong Feller* iff $P_t\phi \in C_b(H)$ for any $\phi \in \mathbb{B}_b(H)$ 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 1.18 (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 1.19 (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 ergodic.

If there is another invariant ergodic measure ν. Then μ must equivalent to ν so that $\mu = \nu$ by Lemma 1.10. □

Remark 1.20. Under the conditions of Doob's Theorem, the conclusion of μ can be stronger than ergodicity. In fact, μ is strongly mixing. The proof [2, Theorem 4.2.1] is not that easy so that we only quote the result.

2 SDE 及其数值解的 Markov 性和齐时性

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

$$\begin{cases} \mathrm{d}X(t) = b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t) \\ X(s) = \eta, \end{cases} \tag{7}$$

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

注意这里需要 drift b 和 diffusion σ 都与时间无关。SDE (7) 的数值解之 Markov 性和齐时性一般是显然的，因此只需讨论 SDE 真实解的相关性质。

Definition 2.1. An $\mathbb{R}^d$-valued stochastic process $\{X_t, s \le t \le T\}$ is called a *solution* of (7) 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

$$X_t = x_0 + \int_s^t b(X_u)\mathrm{d}u + \int_s^t \sigma(X_u)\mathrm{d}W_u \tag{8}$$

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 2.2. 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 SDE (7), where s, x means the SDE 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.

In this subsection, we wish to prove that

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

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

Define

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

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

The following property is an immediate consequence of uniqueness.

Lemma 2.3. *Let $\zeta \in L^2(\Omega, \mathscr{F}_s)$. Then*

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

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

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 $x \in \mathbb{R}^d$.

Assumption 2.4. Assume 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

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

这个假设对于 global Lipschitz 系数的 SDE 和单调单边 Lipschitz 系数的 SDE 均满足。For a proof, see , [4, Proposition 8.6], 其证明可以推广到单调单边 Lipschitz 系数的 SDE.

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

Lemma 2.5. *For all $\phi \in \mathbb{B}_b(\mathbb{R}^d)$ and all $\eta \in L^2(\Omega, \mathscr{F}_s)$, we have*

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

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

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

Proof. [4]. 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 $x_1, \dots, x_n \in \mathbb{R}^d$ and $A_1, \dots, A_n$ 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

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

Assume $\phi \in C_b(\mathbb{R}^d)$. 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 2.4, we have

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

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

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

since their domains are disjoint, which implies

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

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

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

by the property of conditional expectation. In conclusion,

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

Theorem 2.6. *Let $0 \le s \le r \le t \le T$ and $\phi \in \mathbb{B}_b(\mathbb{R}^d)$. Then we have*

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

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

Proof. By Lemma 2.5, we have

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

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

Theorem 2.7 (Markov Property)**.** *Let $0 \le s < r < t \le T$ and let $\eta \in L^2(\Omega, \mathscr{F}_s)$. Then for all $\phi \in \mathbb{B}_b(\mathbb{R}^d)$ we have*

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

Proof. Set $\zeta = X(r,s,\eta)$. Then by Lemma 2.5, using Lemma 2.3,

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

and the conclusion follows. □

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

Theorem 2.8. *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 2.9. *P_t defines a Markov semigroup (not necessarily strongly continuous).*

Proof. We have shown that $P_{0,s+t}\phi = P_{0,s}P_{s,s+t}\phi$ in Theorem 2.6. By homogenity, $P_{s,s+t} = P_t$ and the conclusion follows. □

3 SDE 及其数值解的遍历性

我们已经知道了 SDE 的真实解是一个齐时 Markov 过程，因此可以利用一般的遍历性理论。依靠 Hasminskii 和 Doob 定理，只需证明不变测度的存在性，semigroup 的 strong Feller 和 irreducible 性即可说明其遍历性。

对于真实解，global Lipshitz 系数情形属于 well-known result，单调的单边系数情况在 [9] 中讨论了（也是我毕业论文的主要内容）；对于数值解，global Lipshitz 系数情形可以参看毛学荣的系列论文，非 global Lipschitz 的情况是研究的课题。

A Markov Kernel and Markov Semigroup

In this section, 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 [5, Chapter 1]. 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 *Markov kernel*. However, it should be remarked that similar results in this section hold for σ-finite kernel [1].

A.1 Markov Kernel and its Corresponding Operator

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

Definition A.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 A.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 in the original thesis.

Remark A.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 A.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

$$F_N f(x) \stackrel{\text{def}}{=} \int_{\mathbb{Y}} N(x, \mathrm{d}y) f(y). \tag{9}$$

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

By Remark A.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 A.5. *Let N be a Markov kernel on $\mathbb{X} \times \mathscr{Y}$. Then*

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

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

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

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

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

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

Notation A.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 $L(\mathbb{B}_b(\mathbb{Y}), \mathbb{B}_b(\mathbb{X}))$ 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. Thus the notation F_N would be abandoned.

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

Lemma A.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.

□

A.2 Compositions of Kernels, Markov Semigroup

Theorem A.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 $f \in \mathbb{B}_b(\mathbb{Z})$,*

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

[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.

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 A.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 A.9. (i) As Remark A.2, we can understand $MN(x, A)$ as the probability of x goes A with the help of N then M.

(ii) From Remark A.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 $\nu N \in \mathbb{M}_1(\mathscr{Z})$ such that

$$\nu M(A) = \int_{\mathbb{X}} \nu(\mathrm{d}x) M(x, A). \tag{10}$$

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

Remark A.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:

$$N^{n+k} = N^n \circ N^k \tag{11}$$

or equivalently

$$N^{n+k}(x, A) = \int_{\mathbb{X}} N^n(x, \mathrm{d}y) N^k(y, A). \tag{12}$$

Equation (11) is called a *semigroup* structure. Formally, we have the follwing definition.

Definition A.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

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

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

(iii) When $\mathbb{T} = \mathbb{R}_+$, for any $\phi \in C_b(H)$ (the set of continuous and bounded functions on H) (resp. $\mathbb{B}_b(H)$) and $x \in H$, the function $t \mapsto P_t\phi(x)$ 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* [3, Definition 5.1].

Remark A.12. When $\mathbb{T} = \mathbb{N}$, the semigroup can be constructed by only one Markov kernel. It is immediate, from (9) and (11), 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 A.6), $\pi_t(x, \cdot)$ would be written as $P_t(x, \cdot)$ for a semigroup induced by a Markov kernel.

Remark A.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 [7, Theorem 4.9]. Now in the view point of semigroup, (10) 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 emphasizes 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.

A.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 A.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 $(A \times B)_y = \emptyset$ 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] In [5], the author write (13) without checking the measurablity. We add Lemma A.14 to make it rigorous. This step is also the key step when proving the classic Fubini's 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.

□

Theorem A.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*
$$M \otimes Nf(x) = \int_{\mathbb{Y}} M(x,\mathrm{d}y) \int_{\mathbb{Z}} f(y,z)N(y,\mathrm{d}z). \tag{13}$$
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 A.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 A.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.
$$P^{\otimes n} f(x) = \int_{\mathbb{X}^n} f(x_1,\ldots,x_n)P(x,\mathrm{d}x_1)P(x_1,\mathrm{d}x_2)\cdots P(x_{n-1},\mathrm{d}x_n). \tag{14}$$

Remark A.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.

References

[1] Robert B. Ash. *Probability and measure theory*. Second. With contributions by Catherine Doléans-Dade. Harcourt/Academic Press, Burlington, MA, 2000, pp. xii+516. ISBN: 0-12-065202-1.

[2] G. Da Prato and J. Zabczyk. *Ergodicity for infinite-dimensional systems*. Vol. 229. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 1996, pp. xii+339. ISBN: 0-521-57900-7. DOI: 10.1017/CBO9780511662829. URL: https://doi.org/10.1017/CBO9780511662829.

[3] Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Universitext. Revised and extended from the 2001 original by Da Prato. Springer-Verlag, Berlin, 2006, pp. x+209. ISBN: 978-3-540-29020-9; 3-540-29020-6. DOI: 10.1007/3-540-29021-4. URL: https://doi.org/10.1007/3-540-29021-4.

[4] Giuseppe Da Prato. *Introduction to stochastic analysis and Malliavin calculus*. Third. Vol. 13. Appunti. Scuola Normale Superiore di Pisa (Nuova Serie) [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)]. Edizioni della Normale, Pisa, 2014, pp. xviii+279. ISBN: 978-88-7642-497-7; 88-7642-497-7; 978-88-7642-499-1. DOI: 10.1007/978-88-7642-499-1. URL: https://doi.org/10.1007/978-88-7642-499-1.

[5] Randal Douc et al. *Markov chains*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018, pp. xviii+757. ISBN: 978-3-319-97703-4; 978-3-319-97704-1. DOI: 10.1007/978-3-319-97704-1. URL: https://doi.org/10.1007/978-3-319-97704-1.

[6] Bernt Øksendal. *Stochastic differential equations*. Sixth. Universitext. An introduction with applications. Springer-Verlag, Berlin, 2003, pp. xxiv+360. ISBN: 3-540-04758-1. DOI: 10.1007/978-3-642-14394-6. URL: https://doi.org/10.1007/978-3-642-14394-6.

[7] James C. Robinson. *An Introduction to Functional Analysis*. Cambridge University Press, 2020. DOI: 10.1017/9781139030267.

[8] Kosaku Yosida. *Functional analysis*. Classics in Mathematics. Reprint of the sixth (1980) edition. Springer-Verlag, Berlin, 1995, pp. xii+501. ISBN: 3-540-58654-7. DOI: 10.1007/978-3-642-61859-8. URL: https://doi.org/10.1007/978-3-642-61859-8.

[9] Xicheng Zhang. "Exponential ergodicity of non-Lipschitz stochastic differential equations". In: *Proc. Amer. Math. Soc.* 137.1 (2009), pp. 329–337. ISSN: 0002-9939. DOI: 10.1090/S0002-9939-08-09509-9. URL: https://doi.org/10.1090/S0002-9939-08-09509-9.