Advanced Topics in Stochastic Processes

Xue-Mei Li

Version: December 28, 2021

Contents

0.1 Prologue

These are the lecture notes for the CDT core module "Advanced Topics in Stochastic Processes". I will assume basic knowledge of stochastic calculus and the theory of martingales, and aim to study some more advanced topics in depth.

Stochastic processes are used to model the evolution of physical quantities which have either intrinsic randomness or are subject to external random influences (e.g. from a random environment). The properties one is usually interested in are then of course dependent on the particular phenomenon to be modelled. We will take a more theoretical point of view and study path properties, large time asymptotics, ergodic properties and the connections with PDEs.

In these lectures, we will focus on continuous time and sample continuous stochastic processes on a complete, separable (i.e. having a countable dense subset), metric state space $\mathcal{X}$. We can also work with Polish spaces (separable completely metrizable topological spaces), for all practical purposes we can just as well endow it a metric. As usual, $\mathcal{B}(\mathcal{X})$ is the Borel σ-field on $\mathcal{X}$. We also assume an underlying filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t), P\big)$ with a complete and right-continuous filtration $(\mathcal{F}_t)$ (the 'usual' conditions). A stochastic processes (X_t) is a random function $X : \mathbb{R}_+ \times \Omega \to \mathcal{X}$, assumed to be adapted to the filtration unless stated otherwise.

List of Notations

- $\triangleright$ $\mathcal{B}(\mathcal{X})$ denotes the Borel σ-algebra on a metric or topological space $\mathcal{X}$.
- $\triangleright$ $\mathcal{P}(\mathcal{X})$ denotes the set of Borel probability measures on a metric space $\mathcal{X}$.

▷ $\mathcal{B}_b(\mathcal{X})$ is the space of bounded, measurable functions $\mathcal{X} \to \mathbb{R}$ equipped with the sup-norm.

▷ $\mathrm{BC}(\mathcal{X})$ is the space of bounded continuous functions from $\mathcal{X}$ to $\mathbb{R}$.

▷ $\mathcal{C}(\mathcal{X})$, the space of real continuous functions on $\mathcal{X}$.

▷ $\mathcal{C}_0(\mathcal{X})$ is the space of continuous functions vanishing at infinity equipped with the sup-norm (assuming $\mathcal{X}$ is locally compact). To be more precise $f \in \mathcal{C}_0(\mathcal{X})$ if, for any $\varepsilon > 0$, there is a compact set $K \subset \mathcal{X}$ such that $|f(x)| \leqslant \varepsilon$ for all $x \in \mathcal{X} \setminus K$. This is a Banach space, provided $\mathcal{X}$ is locally compact. In fact, you can check that in this case $\mathcal{C}_0(\mathcal{X})$ is the closure of $\mathcal{C}_c(\mathcal{X})$, the space of continuous functions with compact support.

Example 0.1.1.

(i) $\mathcal{X} = \mathbb{R}^n$ or any finite dimensional, complete, connected, smooth Riemannian manifolds with the relevant metrics. (To clarify, in the definition of manifolds, we assume Hausdorff and second countability).

The state space in most examples of Markov processes encountered in these lectures will be $\mathbb{R}^n$, although it is sometimes useful to treat a sample path of a stochastic process as a 'point' in the path space.

(ii) The space $\mathcal{C}([0,1];\mathbb{R})$ with the supremum norm is a separable Banach space (the set of polynomials with rational coefficients is dense).

In general $\mathcal{C}(\mathcal{X})$ is separable if $\mathcal{X}$ is a compact metric space.

(iii) The *canonical path space* $\mathcal{X} = \mathcal{C}(I, \mathcal{Y})$ for some complete separable metric space $\mathcal{Y}$ and $I = [0, T]$ is a complete separable metric space with the supremum norm. If $I = \mathbb{R}_+$, it will be equipped with the topology of (local) uniform convergence which is metrized by

$$d(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \sup_{t \in [0,n]} d_{\mathcal{Y}}\big(f(t), g(t)\big).$$

(iv) The space $D\big([0,1], \mathcal{Y}\big)$ of càdlàg (right-continuous with left limits) functions $[0,1] \to \mathcal{Y}$. Equipping this space with the uniform topology renders it inseparable, while still being Banach. The remedy was found by Skorohod who introduced a separable metric, which however turned out to be non-complete. Shortly thereafter, Kolmogorov found a complete metric which induces the same topology (remember that completeness is not a topological property!). Kolmogorov's metric is given by

$$d(f,g) = \inf_{\alpha \in H} \Big\{ \sup_{t \in [0,1]} d_{\mathcal{Y}}\big(f(t) - g \circ \alpha(t)\big) + |\mathrm{id} - \alpha|_\infty \Big\},$$

where H denotes the space of strictly increasing homeomorphisms on $[0,1]$. Under this topology, a sequence of functions converges if they converge in the supremum topology after stretching and squashing. The associated topology is now often called the J_1-topology and is closest to the uniform topology (on $\mathcal{C}\big([0,1], \mathcal{Y}\big)$ they coincide). Since sometimes there is need for coarser topologies, other topologies on $D\big([0,1], \mathcal{Y}\big)$ were introduced. The take home fact is that there exists a metric on D rendering it Polish and inducing the Skorohod topology.

(v) The space $\mathcal{B}_b([0,1])$, of bounded measurable functions on $[0,1]$, is not separable. (The collection $\{\mathbf{1}_{[0,a]}, a \leqslant 1\}$ is uncountable, and any two functions in it has distance 1.)

The space $BC(\mathbb{R}, \mathbb{R})$ with the supremum norm is not separable. (To see this, observe that the set of continuous functions taking either 0 or 1 at integers is uncountable, any two distinct functions from it has distance 1 from each other.)

(vi) Assume that $\mathcal{X}$ is in addition locally compact. Then $\mathcal{C}_0(\mathcal{X})$, the collection of real valued functions on $\mathcal{X}$ vanishing at infinity, is separable. Let K_n be a collection of compact sets with $\cup_n K_n = \mathcal{X}$, this property is called countable at inifnity, and let E_n be a countable dense set of continuous functions with compact support on K_n. Then $E = \cup_n E_n$ is a dense subset of $C_0(\mathcal{X})$.

[1]

[1]This is based on the 2020-2021 lecture notes which was typed by Julian, I would like to thank him for typing it up for for proof reading.

Chapter 1

Introduction

1.1 Lecture 1

Let I be the index set and let $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P}\big)$ be a filtered probability space. where $\mathcal{F}_t$ is a filtration satisfying the usual completeness and right continuity assumptions. This is also denoted by $\big(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}\big)$ for short. Any random variable under discussion shall be assumed to be a complete and separable metric space $\mathcal{X}$, e.g. $\mathcal{X}$ is $\mathbb{R}^n$ or a Riemannian manifold. In addition, we index a stochastic process by $t \in I$ for some interval $I \subset \mathbb{R}$ (usually $I = [a, b]$ or $I = \mathbb{R}_+$).

Let $\sigma(X)$ denote the σ-algebra generated by a random variable X with values in $\mathcal{X}$. Likewise let $\mathcal{F}_t^X$ denote the σ-algebra generated by the stochastic process $(X_t, t \in I)$ up to time t, this contains all information from the stochastic process up to time t. This is the natural filtration for $X_\cdot$. If for almost surely all ω, $t \mapsto X_t(\omega)$ is continuous, we may consider $(X_t, t \in I)$ as a random variable on the path space $C(I, \mathcal{X})$, the latter is a metric space with the supremum norm in case I is a compact interval or metrized with the distance function $d(f, g) = \sum_{n=1}^\infty \frac{1}{2^n} \sup_{t\in[0,n]} d_{\mathcal{Y}}\big(f(t), g(t)\big)$, in case $I = \mathbb{R}_+$. (In case the sample paths have left limit and right continuous, the space of continuous functions is replaced by $D\big([0, 1], \mathcal{Y}\big)$, the space of càdlàg (right-continuous with left limits) functions $[0, 1] \to \mathcal{Y}$. As mentioned earlier, we shall focus on stochastic processes with sample continuous paths.

1.1.1 Markov Processes

The probability distribution of a stochastic process $(X_t, t \in I)$ is determined by its finite dimensional distributions $\mu_{t_1,\dots,t_n}$ where $\mu_{t_1,\dots,t_n}(A_1\times\cdots\times A_n) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n)$ for any $A_i \in \mathcal{B}(\mathcal{X})$, and $t_1, \dots, t_n \in I$. In addition to the distribution of the process at each time, we must also know their correlations. The simplest correlation is that any finite collections of random variables $\{X_{t_1}, \dots, X_{t_n}\}$ are mutually independent. A stochastic process is stationary if for any $t \geq 0$, $X_{t+\cdot}$ and $X_\cdot$ have the same finite dimensional distributions. A simplest stochastic process is a stationary process such that for any finite collections of times $\{t_i\}$, $\{X_{t_i}\}$ are uncorrelated and thus $\mu_{t_1,\dots,t_n} = \otimes^n \mu$ where $\mu = \mathcal{L}(X_t)$. The next simplest stochastic process is a Gaussian process, for its probability distribution is determined by the expectation $\mathbb{E}(X_t)$ and its correlation function $\mathbb{E}(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)$. The next simplest is perhaps the Markov property: it allows to determine the finite dimensional distributions with one step transition probabilities.

A stochastic process is said to have the Markov property if its evolution from the current time s depends only on the random variable X_s, given its present it does not depend on the whole history.

If $Z : \Omega \to \mathbb{R}$ is another random variable, then there exists a Borel measurable function $f : \mathcal{X} \to \mathbb{R}$ such that $\mathbb{E}(Z|\sigma(X)) = f(X)$, we denote the function f by $\mathbb{E}(Z|X = x)$. If X is a stochastic process and $G : \Omega \to \mathbb{R}$ a $\mathcal{F}_t^X = \sigma(X_s, s \leqslant t)$ measurable function, there exists $F : C([0,t], \mathcal{X})$ Borel measurable such that $\mathbb{E}(G|\mathcal{F}_t^X) = F((X_s, s \leqslant t))$.

Definition 1.1.1. An $\mathcal{F}_t$ adapted, $\mathcal{X}$-valued, stochastic process X is said to have the *Markov property* if for all $f \in \mathcal{B}_b(\mathcal{X})$ and all $s \leqslant t$, $s, t \in I$,

$$\mathbb{E}\left[f(X_t) \,|\, \mathcal{F}_s\right] = \mathbb{E}\left[f(X_t) \,|\, X_s\right], \qquad a.e.. \tag{1.1.1}$$

In particular, the finite dimensional distributions of a Markov processes are determined by the two time point evolutions and are given by (*exercise*):

$$P(X_{t_n} \in A_n, \ldots, X_{t_1} \in A_1) = P(X_{t_n} \in A_n \,|\, X_{t_{n-1}} \in A_{n-1}) \cdots P(X_{t_1} \in A_1). \tag{1.1.2}$$

In Definition 1.4.1, the Markov process is relative to the background filtration $(\mathcal{F}_t)$. We have allowed to take a filtration finer than $\mathcal{F}_t^X$ to accommodate for other random elements in the system.

Remark 1.1.2. If (X_t) is a Markov process, it is clearly a Markov process with respect to its natural filtration$(\mathcal{F}_t^X)$.

Fix $s, t \geqslant 0$ and $A \in \mathcal{B}(\mathcal{X})$. Then there is a $\psi \in \mathcal{B}_b(\mathcal{X})$ such that

$$\mathbb{P}(X_{t+s} \in A \,|\, X_s) = \psi(X_s) \quad \text{a.s.}$$

The null set depends on the data in general. In practice, however, we can often choose ψ in a nice way.

1.1.2 Transition Functions

As we have seen in (1.1.2), the evolution of a Markov process is completely determined by its one step conditional probabilities. By Definition 1.4.1, for any $f \in \mathcal{B}_b(\mathcal{X})$, any $s < t$, we have a function $P_{s,t}f$ such that the right hand side of 1.1.1 equals to $P_{s,t}f(X_s)$ almost surely. We shall need to assume measurability conditions in s, t and x, which we usually easy to obtain for separable complete metric spaces, and use the regular versions of these, the 'transition probabilities'. Roughly speaking, these represent the probability of finding X_{t+s} in a neighbourghood of a point y knowing that $X_s = x$.

To avoid handling measurability issues, we introduce transition functions.

Definition 1.1.3. A real valued mapping $P : \mathbb{R}_+ \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ is time-homogeneous *transition function* if

(i) $P_t(x, \cdot) \in \mathcal{P}(\mathcal{X})$ for all $(t, x) \in \mathbb{R}_+ \times \mathcal{X}$.

(ii) $P_0(x, \cdot) = \delta_x$ for all $x \in \mathcal{X}$.

(iii) for any $A \in \mathcal{B}(\mathcal{X})$, $(t, x) \mapsto P_t(x, A)$ is measurable.

(iv) for any $t, s \geqslant 0$, $x \in \mathcal{X}$, and $A \in \mathcal{B}(\mathcal{X})$, the Chapman-Kolmogorov equation

$$P_{t+s}(x, A) = \int_{\mathcal{X}} P_t(y, A) P_s(x, dy). \tag{1.1.3}$$

holds.

It is standard to only requires $x \mapsto P_t(x, A)$ measurable for any $t \geqslant 0$. But the joint measurability is often easily obtainable.

Remark 1.1.4. Equation (1.1.3) implies that, for all $0 \leqslant s \leqslant t \leqslant u$ and $f \in \mathcal{B}_b(\mathcal{X})$, the *Chapman-Kolmogorov equation*

$$\int_{\mathbb{R}^d} f(y)\, P_{t+s}(x, dy) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y)\, P_t(z, dy)\, P_s(x, dz) \tag{1.1.4}$$

holds.

Definition 1.1.5. An adapted process $(X_t)_{t \geqslant 0}$ is a (time homogeneous) Markov process with transition function P if, for any $0 \leqslant s \leqslant t$ and any $A \in \mathcal{B}(\mathcal{X})$,

$$P(X_t \in A \,|\, \mathcal{F}_s) = P_{t-s}(X_s, A) \quad \text{a.s.} \tag{1.1.5}$$

Note that if (X_t) is a stochastic process satisfying (1.1.5), then it is necessarily a Markov process. In particular,

$$\mathbb{P}(X_t \in A|\mathcal{F}_0) = \mathbb{P}(X_t \in A|X_0) = P_t(X_0, A), \qquad \mathbb{P}(X_t \in A|X_0 = x) = P_t(x, A) \quad a.e..$$

Consequently, for any $f \in \mathcal{B}_b(\mathcal{X})$,

$$\mathbb{E}(f(X_t)|X_0 = x) = \int_{\mathcal{X}} f(y) P_t(x, dy).$$

If X_0 is distributed as μ_0 we see (by taking expectation of the above),

$$\mathbb{P}(X_t \in A) = \mathbb{E}(\mathbb{E}(f(X_t)|X_0)) = \int_{\mathcal{X}} P_t(y, A) \mu_0(dy). \tag{1.1.6}$$

By the Chapman-Kolmogorov equation one can show (exercise):

Theorem 1.1.6. *Let $(X_t)_{t \geqslant 0}$ be a Markov process with transition function P and initial distribution $X_0 \sim \mu_0$. Then, for any $A_0, \ldots, A_n \in \mathcal{B}(\mathcal{X})$ and $0 = t_0 < t_1 < \cdots < t_n$,*

$$\mathbb{P}(X_{t_0} \in A_0, \ldots, X_{t_n} \in A_n) = \int_{A_0} \cdots \int_{A_n} P_{t_n - t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(y_0, dy_1) \mu_0(dy_0). \tag{1.1.7}$$

Remark 1.1.7. A Markov process with a transition function can start from any point, $P(X_t \in A \,|\, X_0 = x) = P_{t-s}(x, A)$, and so we have a family of stochastic processes satisfying the Markov property.

From now on all Markov processes are Markov processes with transition functions unless otherwise stated.

1.2 Markov Semi-groups

Definition 1.2.1. For any $t \geqslant 0$, we define a linear map $T_t : \mathcal{B}_b(\mathcal{X}) \to \mathcal{B}_b(\mathcal{X})$,

$$T_t f(x) \triangleq \int_{\mathcal{X}} f(y) P_t(x, dy).$$

If X_t is a Markov process with transition function P_t, then

$$T_t f(x) = \mathbb{E}[f(X_{t+s}) \,|\, X_s = x].$$

Note that T_0 is the identical transformation and the Chapman-Kolmogorov equation leads to

$$T_{t+s}f(x) = \int_{\mathcal{X}} f(z)P_{t+s}(x, dz) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(z)P_s(y, dz)P_t(x, dy)$$
$$= \int_{\mathcal{X}} T_s f(y)P_t(x, dy) = T_t(T_s f)(x).$$

Thus T_t is a semigroup of linear operators on $E = \mathcal{B}_b(\mathcal{X})$: $T_0 = id$, and

$$T_t \circ T_s = T_{t+s}. \tag{1.2.1}$$

In addition, $T_t f \geq 0$ if $f \geq 0$, $T_t 1(x) = 1$, and $\|T_t f\|_E \leqslant \|f\|_E$.

Remark 1.2.2. The property (1.2.1) encapsulates the key analytical property of Markov processes since (T_t) consequently defines a 'semigroup' of linear operators on $\mathcal{B}_b(\mathcal{X})$. We can now pull our most favorite functional analysis book off the shelf and have a look what results of the analysts we can hijack for our analysis. The first resignation comes as we realize that some regularity of the mapping $t \mapsto T_t$ is required for a rich theory. Moreover, we expect some more specialized results since (T_t) has additional structure, e.g. $T_t 1 = 1$ and $T_t f \geqslant 0$ for $f \geqslant 0$. We also need some continuity of the operator T_t.

We will study these in more detail, but before that let us have a look into some examples of Markov processes.

1.3 Brownian Motion and diffusion processes

A stochastic process (X_t) is said to have independent increments, if for any $t_1 < t_2 < \cdots < t_n$, $(X_{t_n} - X_{t_{n-1}}, \dots, X_{t_2} - X_{t_1})$ are independent random variables.

Exercise 1.3.1. If a stochastic process has independent increments, it is a Markov process.

If B_t is a 3-dimensional Brownian motion, then

$$\mathbb{P}(B_t \in A | B_0 = x) = \int_{\mathbb{R}^d} \frac{1}{((2\pi)t)^{3/2}} e^{-\frac{|y-x|^2}{2t}} dy.$$

A Brownian motion is the archetypal Markov process, first used to model the motion of a macroscopic particle (e.g. pollen) in a liquid at rest is the result of its collisions with the microscopic water molecules which undergo a thermodynamic movement. The probability of finding the Brownian particle in a set A at time $t > 0$ is thought to be

$$\mathbb{P}(B_{t+s} \in A \,|\, B_s) = \int_A p_t(B_s, y)\, dy$$

where$p : (0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ denote the *heat kernel*

$$p_t(x, y) \triangleq (2\pi t)^{-\frac{d}{2}} e^{-\frac{|x-y|^2}{2t}},$$

the fundamental solution to the heat equation

$$\frac{\partial}{\partial t} u = \frac{1}{2} \triangle. \tag{1.3.1}$$

We set $P_t(x, A) \triangleq \int_A p_t(x, y)\, dy$, the family of measures $\{P_t(x, \cdot)\}_{x \in \mathbb{R}^d}$ is a transition function.

The equation (1.3.1) and its generalizations, which we discuss later, are called diffusion equations where the Laplacian Δ is replaced by a diffusion operator:

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{k=1}^{d} b_k(x)\frac{\partial}{\partial x_k}.$$

The name diffusion stems from a molecular diffusion model in which a large number of particles move according to Fick's law. We can use a random walk model for the discrete version. Let Δx denote the spatial step and Δt the time step, so we have $s_j = j\Delta x$ and $t_n = t_0 + n\Delta t$. Denote by c_j^n the number of particles at site s_j at time t_n. Suppose that a molecule moves to its left with probability $p = \frac{1}{2}$, to its right with $q = \frac{1}{2}$, then

$$c_j^{n+1} - c_j^n = \frac{1}{2}c_{j+1}^n + \frac{1}{2}c_{j-1}^n - c_j^n.$$

Let $\Delta c_j^n = c_j^{n+1} - c_j^n$. Then

$$\frac{\Delta c_j^n}{\Delta t} = \frac{1}{2}\frac{(\Delta x)^2}{\Delta t}\frac{\left(c_{j+1}^n + c_{j-1}^n - 2c_j^n\right)}{(\Delta x)^2}.$$

If we keep the ratio $D \triangleq \frac{(\Delta x)^2}{\Delta t}$ constant, this discrete equation is a finite difference approximation for the equation

$$\frac{\partial}{\partial t}c = \frac{1}{2}D\triangle C.$$

A Brownian motion on $\mathbb{R}^n$ is a Gaussian process and a Markov process. From the definition one sees that

$$\mathbb{E}(f(B_t)|B_s = x) = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} f(y+x)e^{-\frac{|y|^2}{2(t-s)}}\,dy.$$

Note that if f is bounded measurable, the function on the right hand side is continuous in s, t and in x. This smoothing property is typical os elliptic diffusions. You are asked to give a proof of the following theorem on Example sheet 1.

Theorem 1.3.2. *Let $f \in \mathrm{B}C(\mathbb{R}^d)$, then*

$$u(t,x) = \int_{\mathbb{R}^d} f(y)p_t(x,y)\,dy$$

solves the heat equation (1.3.1) *and* $\lim_{t\searrow 0} u(t,x) = f(x)$ *for each* $x \in \mathbb{R}^d$.

A formal definition:

Definition 1.3.3. An n-dimensional Brownian motion B_t is a sample continuous sample path, with independent stationary increments, and $B_t - B_s \sim N(0, (t-s)I)$ for any $0 \leqslant s < t$.

Example 1.3.4. An example of a Markov process is the stationary Ornstein-Uhlenbeck process which solves the linear equation:

$$\dot{x}_t = v_t, \qquad dv_t = -v_t dt + dB_t.$$

Here (v_t) is a Markov process on $\mathbb{R}$, (x_t, v_t) is a Markov process on $\mathbb{R}^2$, but (x_t) alone is not a Markov process.

Exercise 1.3.5. Compute their transition functions, show (x_t) is not a Markov process.

You may want to check conditions on a Gaussian process to be a Markov process.

Remark 1.3.6. A standard Brownian motion on $\mathbb{R}^n$ has many desirable properties:

(i) It is a Markov process.

(ii) It is a martingale.

(iii) It is a Gaussian process.

(iv) It has stationary increments.

(v) It has independent increments.

(vi) Its probability distribution at any time $t > 0$ has exponential tails.

(vii) It is an elliptic diffusion.

(viii) Its semigroup has the strong Feller property.

(ix) Almost all of its sample paths are continuous (they are in fact in $\mathcal{C}^{\frac{1}{2}-}$).

The concepts of a Brownian motion can be defined on a Riemannian manifold: It is a strong Markov process with generator $\frac{1}{2}\Delta$.

Example 1.3.7. Not every Gaussian process is a Markov process. Fractional Brownian motion with Hurst parameter $H \in (0,1) \setminus \{\frac{1}{2}\}$ is a centred non-Markovian Gaussian process with covariance:

$$\mathbb{E}(B_t B_s) = \frac{1}{2}(t^{2H} + S^{2H} - |t-s|^{2H}).$$

Example 1.3.8. If W_t is a Brownian motion from 0, $x + W_t$ is a Markov process with the heat transition function $P_t(x, dy) = \frac{1}{\sqrt{2\pi}} e^{-|y-x|^2/2t} dy$ and initial value x. It starts from x, after time s, it restarts from $x + B_s$.

1.4 Strong Markov Process

Definition 1.4.1. An adapted, $\mathcal{X}$-valued stochastic process X is a *strong Markov process* if for all $f \in \mathcal{B}_b(\mathcal{X})$ and and for any stopping time τ and any $t \geq 0$,

$$\mathbb{E}\left[f(X_{\tau+t}) \,|\, \mathcal{F}_\tau\right] = \mathbb{E}\left[f(X_{\tau+t}) \,|\, X_\tau\right]$$

on the event $\{\tau < \infty\}$ where $\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \leqslant t\} \in \mathcal{F}_t, \forall t\}$.

Example 1.4.2. Let

$$X_t \triangleq \begin{cases} x + W_t, & \text{if } X_0 = x \neq 0, \\ 0, & \text{if } X_0 = 0, \end{cases}$$

for a one-dimensional Brownian motion $(W_t)_{t \geqslant 0}$. This is a Markov process with transition function

$$Q_t(x, dy) = \begin{cases} P_t(x, dy), & \text{if } x \neq 0, \\ \delta_0(dy), & \text{if } x = 0, \end{cases}$$

where $P_t(x, dy) = p_t(x, y)dy$ where $p_t(x, y)$ is the heat kernel and dy the Lebesgue measure on $\mathbb{R}$. It is clear that

$$\mathbb{E}\left[f(X_{t+s}) \,|\, X_s\right] = \begin{cases} T_t f(X_s), & \text{if } X_s \neq 0, \\ f(0), & \text{if } X_s = 0, \end{cases}$$

where $T_t f(x) \triangleq \int_{\mathbb{R}} f(y)\, P_t(x, dy)$ is the heat semigroup. Note that the Chapman-Kolmogorov equation

$$Q_{t+s}(x, A) = \int_{\mathcal{X}} Q_t(y, A) Q_s(x, dy)$$

holds. Indeed, $P_{t+s}(x, A) = \int_{\mathcal{X}} P_t(y, A) P_s(x, dy)$. If $x \neq 0$, since $Q_t(y, A)$ and $P_t(y, A)$ only differ at $y = 0$ and the Lebesgue measure does not charge singleton sets, this does not change the integral. Also $Q_t(0, A) = \delta_0(A)$ for each $t \geqslant 0$ and so

$$\int_{\mathbb{R}} Q_t(y, A) Q_s(0, dy) = \int_{\mathbb{R}} Q_t(y, A) \delta_0(dy) = \delta_0(A) = Q_{t+s}(0, A).$$

The process X is however not strong Markov. Indeed, let $\tau \triangleq \inf\{t \geqslant 0 : X_t = 0\}$. Then, by strong Markov, it would restart at time τ from $0 = X_\tau$ and we would have $X_{t+\tau} = 0$ for all $t \geqslant 0$. Our Markov process X_t starting away from zero goes straight through 0 like a Brownian motion.

Definition 1.4.3. We say T_t has the Feller property if $T_t f$ is continuous whenever f is bounded continuous; it is said to have the strong Feller property if $T_t f$ is continuous whenever f is bounded measurable.

Exercise 1.4.4. Check whether T_t defined in the example above has the Feller property.

1.5 Gaussian processes with the Markov Property

For any $a \in \mathbb{R}^n$ and any $n \times n$ symmetric positive definite matrix C, there is a unique probability measure with the following Fourier transform:

$$\int_{\mathbb{R}^n} e^{i\langle \xi, x\rangle} \mu(dx) = e^{i\langle \xi, a\rangle - \frac{1}{2}\langle X\xi, \xi\rangle}, \qquad \xi \in \mathbb{R}^n,$$

it is called the Gaussian measure with mean a and covariance C. If C is non-degenerate, the Gaussian measure is of the form

$$\frac{1}{\sqrt{(2\pi)^n \det C}} e^{-\frac{1}{2}\langle x-a, C^{-1}(x-a)\rangle} dx.$$

Let X_t be a Gaussian process with mean $\mu_t = \mathbb{E}(X_t)$ correlation function (auto-correlation function):

$$R(s, t) = \mathbb{E}\langle X_t - \mathbb{E}X_t, X_s - \mathbb{E}X_s\rangle.$$

For the one dimensional Brownian motion this is $s \wedge t$, for the one dimensional fractional Brownian motion of parameter H this is

$$R(s, t) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}).$$

Exercise 1.5.1. Show that for $H \in (0, 1) \setminus \{\frac{1}{2}\}$, a fractional Brownian motion is not a Markov process.

It is easy to verify whether a Gaussian process is a Markov process, for the conditional expectation of X_t on X_s is a linear function.

Proposition 1.5.2. *If a centred Gaussian process on $\mathbb{R}^n$ with covariance $R(t,s)$ is a Markov process, then for any $s < t < u$,*

$$R(s,u) = \frac{R(s,t)R(t,u)}{R(t,t)}. \tag{1.5.1}$$

Conversely if the covariance of a Gaussian process satisfies the above identity for any $s \leqslant t \leqslant u$, it is a Markov process. We have assumed implicitly the variances do not vanish at any t.

Proof. We first claim that for $s < t$,

$$\mathbb{E}(X_t \mid X_s) = \frac{R(s,t)}{R(s,s)} X_s, \qquad a.e. \tag{1.5.2}$$

Indeed, the identity holds precisely when $X_t - \frac{R(s,t)}{R(s,s)} X_s$ and X_s are uncorrelated and therefore independent, so

$$\mathbb{E}(X_t|X_s) = \mathbb{E}\Big(X_t - \frac{R(s,t)}{R(s,s)} X_s \mid X_s\Big) + \mathbb{E}\Big(\frac{R(s,t)}{R(s,s)} X_s \mid X_s\Big) = \frac{R(s,t)}{R(s,s)} X_s.$$

Suppose that X_t is a Markov process, then for $s < t < u$,

$$\begin{aligned}
\frac{R(s,u)}{R(s,s)} X_s &= \mathbb{E}(X_u \mid X_s) \\
&= \mathbb{E}\big(\mathbb{E}(X_u|F_t)|X_s\big) = \mathbb{E}\Big(\frac{R(t,u)}{R(t,t)} X_t \mid X_s\Big) \\
&= \frac{R(t,u)}{R(t,t)} \frac{R(s,t)}{R(s,s)} X_s, \qquad a.e.,
\end{aligned}$$

concluding the require covariance identity. Conversely suppose the identities Equation (1.5.1) hold, then for any $t < u$, $X_u - \frac{R(t,u)}{R(t,t)} X_t$ is not only independent of X_t, it is also independent of X_s for any $s \leqslant t$, as

$$\mathbb{E}\Big(\Big(X_u^i - \frac{R(t,u)}{R(t,t)} X_t^i\Big) X_s^j\Big) = R_{i,j}(s,u) - \frac{R_{i,j}(t,u)}{R_{i,j}(t,t)} R_{i,j}(s,t) = 0,$$

consequently $X_u - \frac{R(t,u)}{R(t,t)} X_t$ is independent of $\sigma(X_s, s \leqslant t)$ and

$$\begin{aligned}
\mathbb{E}(f(X_u) \mid \mathcal{F}_t) &= \mathbb{E}\Big(f\Big(X_u - \frac{R(t,u)}{R(t,t)} X_t + \frac{R(t,u)}{R(t,t)} X_t\Big) \mid \mathcal{F}_t\Big) \\
&= \int f\Big(z + \frac{R(t,u)}{R(t,t)} X_t\Big) \nu(dz),
\end{aligned}$$

where ν is the distribution of the Gaussian random variable $X_u - \frac{R(t,u)}{R(t,t)} X_t$, since the right hand side is a function of X_t this proves the Markov property. Observe also that its transition semi-group is then

$$P_{u-t} f(x) = \int f\Big(z + \frac{R(t,u)}{R(t,t)} x\Big) \nu(dz).$$

Note that ν is Gaussian with covariance $R(t,u) - \frac{R^2(t,u)}{R(t,t)}$. ☐

Exercise 1.5.3. Prove this when the Gaussian process is not centred. Show that the fractional Brownian motion does not satisfy Equation (1.5.1).

At this point we would explore an interesting property of a Markov process and from which to construct a Gaussian process indexed by the state space of the Markov process.

Definition 1.5.4. A real valued stochastic process $(X_t, t \in E)$ where E is a general index set, is a Gaussian process if for any mult-indices $t_1, \dots, t_n$ from E, the random variable $(X_{t_1}, \dots, X_{t_n})$ has Gaussian distribution. For any $x, y \in E$, we set

$$\rho(x,y) = \mathrm{cov}(X_x, X_y).$$

Example 1.5.5. A Gaussian white noise on $\mathcal{X}$ is a family of centred Gaussian random variables $\{W(t,A) : t \geq 0, A \in \mathcal{B}(\mathcal{X})\}$ with the covariance

$$\mathbb{E}(W(t,A)W(s,A')) = (s \wedge t)\mu(A \cap A')$$

for μ a finite Borel measure on $\mathcal{X}$. This allows to define for any $f : \mathcal{X} \to \mathbb{R}$ in $L^2(\mathcal{X}, \mu)$ the Gaussian random variable $W(f, A)$ by setting $W(\sum_{i=1}^n a_i \mathbf{1}_{A_i}, t) = \sum_{i=1}^n W(A_i, t)$, so $\{W(f,t) : f \in L^2(\mathcal{X}, \mu)\}$ is a Gaussian family of random variables with:

$$\mathbb{E}(W(f,t)W(f,s)) = (s \wedge t) \int fg d\mu.$$

Let $P_t(x, dy)$ be a transition function on a space $\mathcal{X}$ with the property they have densities with respect to one, and the same one, σ-finite measure μ. We denote by $p(t,x,y)$ the densities so

$$P_t(x, dy) = p(t,x,y)\mu(dy).$$

Set

$$\rho(x,y) = \int_0^\infty p(t,x,y)dt.$$

Proposition 1.5.6. *Suppose that $\rho(x,y) < \infty$ for all $x, y \in \mathcal{X}$, it is the covariance of a Gaussian process.*

Proof. By the proposition below, it is sufficient to show ρ is of positive type. Let $\xi = (\xi_1, \dots, x_n) \in \mathbb{R}^n$, $x \in \mathcal{X}$, then

$$\begin{aligned}
\sum_{i,j=1}^n \rho(x_i, x_j)\xi_i\xi_j &= \sum_{i,j=1}^n \int_0^\infty p(t, x_i, x_j)dt\ \xi_i\xi_j \\
&= \sum_{i,j=1}^n \int_0^\infty \int_{\mathcal{X}} p(t/2, x_i, z)p(t/2, x_j, z)\mu(dz)dt\xi_i\xi_j \\
&= \int_0^\infty \int_{\mathcal{X}} \Big(\sum_{i=1}^n p(t/2, x_i, z)\xi_i\Big)^2 \mu(dz)dt \geq 0,
\end{aligned}$$

completing the proof. □

The condition $\rho(x,y) = \int_0^\infty p(t,x,y)dt < \infty$ is related to the transient property of the process, it is more likely to be finite, if the process does not spent much time there, For a Brownian motion on $\mathbb{R}^n$,

$$g(x,y) = \int_0^\infty p(t,x,y)dy = \frac{1}{2}\frac{1}{\sqrt{\pi^n}}|x-y|^{2-n}\Gamma(\frac{n}{2} - 1)$$

where $p(t,x,y)$ is the heat kernel, is the Green kernel, $g(x,x) = \infty$.

Formally, we interpret $\int_0^\infty P_t dt$ as $(-\frac{1}{2}\Delta)^{-1}$: if $-\frac{1}{2}\Delta$ to the integral which itself applied to a suitable function f: $-\frac{1}{2}\Delta \int_0^\infty P_t f dt = -\int_0^\infty \frac{\partial}{\partial t} P_t f dt = f - \text{`}P_\infty f\text{'}$. This can be made rigorous when applied to the Laplacian on a compact manifold then there exists a unique invariant probability measure π and we expect that $\mathcal{L}(X_t)$ converges to π and for any f with $\int f d\pi = 0$ we expect that $\lim_{t\to\infty} P_t f = \int f \mu(dy) = 0$.

Proposition 1.5.7. *A function $\rho : E \times E \to \mathbb{R}$ is the covariance of a Gaussian process if and only if ρ is positive type, i.e. for any $y_1, \dots, y_n \in E$, the matrix $(\rho(y_i, y_j))$ is a non-negative / positive definite matrix.*

The necessity part of the proposition is trivial, the converse follows from Kolmogorov's extension theorem.

1.6 Invariant Measure and Stationary Markov Processes

Given a measure μ on $\mathcal{X}$, we define for any $A \in \mathcal{B}(\mathcal{X})$,

$$T_t^* \mu(A) = \int_{\mathcal{X}} P_t(x, A) \mu(dx),$$

this defines a transformation on measures on $\mathcal{X}$:

$$\mu \mapsto T_t^* \mu(\cdot) = \int_{\mathcal{X}} P_t(x, \cdot) \mu(dx).$$

Then for any $f \in \mathcal{B}_b(\mathcal{X})$,

$$\int_{\mathcal{X}} f(y)(T_t^* \mu)(dy) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(z) P_t(x, dz) \mu(dy) = \int_{\mathcal{X}} T_t f(y) \mu(dy).$$

Definition 1.6.1. A measure μ is an invariant measure for a transition function $P_t(x, dy)$ (for the Markov process with the transition probability) if $T_t^* \mu = \mu$ for any t, i.e.

$$\int_{\mathcal{X}} T_t f(y) \mu(dy) = \int_{\mathcal{X}} f(y) \mu(dy)$$

for any $f \in \mathcal{B}_b(\mathcal{X})$.

If μ is furthermore a probability measure and if X_0 with distribution μ, then $T_t^* \mu$ is the probability distribution of X_t, c.f. Equation (1.1.6), so the probability distribution of X_t does not change with time.

Naturally, an invariant probability measure of T_t is also called an invariant probability measure of the Markov process.

1.7 Exercises

Coming back to the Ornstein-Uhlenbeck process on $\mathbb{R}$, the solution to

$$dv_t = -\beta v_t dr + \alpha dB_t$$

is a Gaussian process, where β, α are real numbers. The solution can be explicitly written:

$$v_t = e^{-\beta t} v_0 + \alpha \int_0^t e^{-\beta(t-s)} dB_s.$$

Exercise 1.7.1. Show that the distribution of

$$\int_{-\infty}^{0} e^{-t+s} dB_s$$

is an invariant probability measure for the solution of

$$dv_t = v_t dr + dB_t$$

Exercise 1.7.2. Show that the solution v_t^β to

$$dv_t^\beta = -\beta v_t^\beta dr + \sqrt{\beta} dB_t$$

has the same distribution as $v_{t/\beta}$ where v_t solves

$$dv_t = -v_t^\beta dt + \sqrt{\beta}_t dB_t.$$

Exercise 1.7.3. Show that the solution $x_\cdot^\beta$ of the system of equations,

$$\dot{x}_t^\beta = v_t^\beta, \qquad dv_t^\beta = -\beta v_t^\beta dr + \beta dB_t,$$

converges weakly to that of a Brownian motion as $\beta \to \infty$. Compute the covariance $\mathbb{E}(v_s^\beta v_t^\beta)$.

1.8 The Canonical Picture

Given a transition function P and a probability measure μ, does there exist a Markov process with t.f. P and initial distribution μ? To answer this question we work with the canonical space.

Let $\mathcal{X}^I = \{\omega : I \to \mathcal{X}\}$ denote the collection of mappings from I to $\mathcal{X}$ with the product Borel σ-algebra $\bigotimes_{t\in I} \mathcal{B}(\mathcal{X}) = \sigma(\pi_t, t \in I)$. A stochastic process $(X_t, t \in I)$ is a random variable with values in $\mathcal{X}^I$ and induces a measure $\mu_X = \mathcal{L}(X_\cdot)$ on $\mathcal{X}^I$. This measure encodes all statistical information of the process.

Let us change the point of view, take $\Omega \triangleq \mathcal{X}^I = \{\omega : I \to \mathcal{X}\}$ to be our measurable space , this is the canonical space. We endowed with the measure induced by the stochastic process for $X_\cdot$.

Let $\pi_t : \mathcal{X}^I \to \mathcal{X}$,

$$\pi_t(\omega) = \omega(t)$$

be the canonical evaluation map at time $t \in I$ and let $\mathcal{F}_t$ be its natural σ-algebra.

Remark 1.8.1. If $\{\mathcal{X}_n\}$ is a family of separable metric spaces, the Borel σ-algebra of $\Pi_{n=1}^\infty \mathcal{X}_n$ agrees with the product σ-algebras $\otimes_{n=1}^\infty \mathcal{B}(X_n)$. Note that $\mathcal{B}(\mathcal{X})^{\otimes I} \subset \mathcal{B}(\mathcal{X}^I)$, the latter is the Borel σ-field on $\mathcal{X}^I$ equipped with the product topology, and the inclusion is strict. Indeed, it is clear that singletons are closed in the product topology but a set $A \in \mathcal{B}(\mathcal{X})^{\otimes I}$ can only depend on countably many times.

Recall that if $(X_t)_{t\geqslant 0}$ is a Markov process with transition function P and initial distribution $X_0 \sim \mu$, then for any $A \in \mathcal{B}(\mathcal{X})$, $\mathbb{P}(X_t \in A) = \mathbb{E}\left[P_t(X_0, A)\right] = \int_{\mathcal{X}} P_t(y, A)\,\mu(dy)$. and by Theorem 1.1.6, for any $A_0, \dots, A_n \in \mathcal{B}(\mathcal{X})$ and $0 = t_0 < t_1 < \cdots < t_n$,

$$\mathbb{P}(X_{t_0} \in A_0, \dots, X_{t_n} \in A_n) = \int_{A_0} \cdots \int_{A_n} P_{t_n - t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(y_0, dy_1)\mu(dy_0). \tag{1.8.1}$$

This inspires the following definition. Given P_t, μ, and $\Delta = \{t_1 < \cdots < t_n\} \subset I$ a finite collection of times, we define a measure μ_Δ on $\mathcal{X}^{n+1}$ by

$$\mu_\Delta(A_0 \times \cdots \times A_n) \triangleq \int_{A_0} \cdots \int_{A_n} P_{t_n - t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(y_0, dy_1)\mu(dy_0). \quad (1.8.2)$$

This collection of finite-dimensional distributions is consistent in the sense that, if $A_k = \mathcal{X}$,

$$\mu_\Delta(A_0 \times \cdots \times A_n) = \mu_{\Delta \setminus \{t_k\}}(A_0 \times \cdots A_{k-1} \times A_{k+1} \times \cdots \times A_n).$$

We leave it to the reader to check the consistency. Kolmogorov's extension theorem then establishes the following result:

Theorem 1.8.2 (Canonical picture)**.** *Let P be a transition function and $\mu \in \mathcal{P}(\mathcal{X})$. Then there exists a unique measure $\mathbb{P}_\mu$ on $\mathcal{X}^I$ such that, for any finite set of times $\Delta \subset I$, $\Delta = \{t_1, \ldots, t_n\}$,*

$$\pi_\Delta^* P_\mu = \mu_\Delta,$$

where $\pi_\Delta(\omega) = \big(\omega(t)\big)_{t \in \Delta} = (\omega(t_1), \ldots, \omega(t_n))$. Consequently, the coordinate map π_t is a Markov process on $\big(\mathcal{X}^{\mathbb{R}_+}, \bigotimes_{t \in I} \mathcal{B}(\mathcal{X}), \mathbb{P}_\mu\big)$ with transition function P and initial distribution μ.

Equation (1.8.2) precisely means that the finite dimensional distributions of π_t are $\pi_\Delta^* P_\mu$, it is therefore a Markov process.

Definition 1.8.3. If $\mu = \delta_x$ in Theorem 1.8.2, we denote $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

Recall that in the definition of a transition function we required that $(t, x) \mapsto P_t(x, A)$ is measurable for each $A \in \mathcal{B}(\mathcal{X})$. Hence,

$$x \mapsto P_x(\pi_{t_1} \in A_1, \ldots, \pi_{t_n} \in A_n) = \int_{A_1} \cdots \int_{A_n} P_{t_n - t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(x, dy_1)$$

is measurable and, by an easy monotone class argument, the same holds for $x \mapsto \mathbb{P}_x(A)$ for a general $A \in \bigotimes_{t \in I} \mathcal{B}(\mathcal{X})$. We can hence integrate $\mathbb{P}_x(A)$ and In particular $\pi_\Delta^* P_\mu = \int \pi_\Delta^* P_x \mu(dx)$, we have

$$\mathbb{P}_\mu(A) = \int_{\mathcal{X}} \mathbb{P}_x(A)\, \mu(dx).$$

Remark 1.8.4. The collection of probability measures $\mathbb{P}_x$ are Markovian measures (on the path space). If the Markov process is furthermore strong Markov with sample continuous sample paths, they are called diffusion measures.

Let us now examine how the Markov property looks in the canonical picture, taking $I = \mathbb{R}^+$. To this end, let $\theta_s : \mathcal{X}^{\mathbb{R}_+} \to \mathcal{X}^{\mathbb{R}_+}$, $\theta_s \omega(t) = \omega(s+t)$ be the shift operator. If $\Phi : \mathcal{X}^{\mathbb{R}_+} \to \mathbb{R}$ is a Borel measurable function, we introduce the notation:

$$\mathbb{E}_\mu[\Phi] = \int_{\mathcal{X}^{\mathbb{R}_+}} \Phi(\sigma)\, d\mathbb{P}_\mu(\sigma), \qquad \mathbb{E}_x[\Phi] = \int_{\mathcal{X}^{\mathbb{R}_+}} \Phi(\sigma)\, d\mathbb{P}_x(\sigma),$$

Using the canonical process X, on the probability space $(\mathcal{X}^{\mathbb{R}_+}, \mathcal{B}_b(\mathcal{X}^{\mathbb{R}_+}), \mathbb{P}_x)$, we have another notation: $\mathbb{E}_x[\Phi] = \mathbb{E}_x[\Phi(X)]$.

Theorem 1.8.5. *Let $(X_t)_{t \geqslant 0}$ denote the canonical Markov process with transition function P. Then, for any $\Phi \in \mathcal{B}_b(\mathcal{X}^{\mathbb{R}_+})$,*

$$\mathbb{E}_x\big[\Phi(\theta_s X) \,|\, \mathcal{F}_s\big] = \mathbb{E}_{X_s}[\Phi(X)] \quad \mathbb{P}_x - a.s. \quad (1.8.3)$$

for each $x \in \mathcal{X}$.

This can be written as

$$\mathbb{E}_x\big[\Phi \circ \theta_s \,|\, \mathcal{F}_s\big] = \mathbb{E}_{X_s}[\Phi] \quad \mathbb{P}_x - \text{a.s.}$$

Proof. It is enough to prove this for

$$\Phi(\omega) = \mathbf{1}_{\{\omega:\, \omega(t_1)\in A_1,\ldots,\omega(t_n)\in A_n\}}.$$

Then (1.8.3) becomes

$$\mathbb{P}_x(X_{t_1+s} \in A_1, \ldots, X_{t_n+s} \in A_n \,|\, \mathcal{F}_s) = \mathbb{P}_{X_s}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n).$$

By Lemma 1.8.6

$$\begin{aligned}
&\mathbb{P}_x(X_{t_1+s} \in A_1, \ldots, X_{t_n+s} \in A_n \,|\, \mathcal{F}_{t_1+s}) \\
&= \int_{\mathcal{X}} \cdots \int_{A_n} P_{t_n-t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(y_0, dy_1)\mu(dy_0) \\
&= \int_{A_1} \cdots \int_{A_n} P_{t_n-t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(X_s, dy_1),
\end{aligned}$$

where the second line follows from (1.8.1) with $\mu = \delta_{X_s}$, proving the required identity. $\square$

Lemma 1.8.6. *Let (X_t) be a Markov process with transition function $P_t(x, A)$, $t_1 \leqslant \ldots \leqslant t_n$, and $f_1, \ldots, f_n$ from $\mathcal{B}_b(\mathcal{X})$, then*

$$\mathbb{E}(\Pi_{i=1}^n f_i(X_{t_i+s} \,|\, \mathcal{F}_s) = \int_{\mathcal{X}} \ldots \int_{\mathcal{X}} \Pi_{i=1}^n f_i(z_i) P_{t_n-t_{n-1}}(z_{n-1}, dz_n) \ldots P_{t_1}(X_s, dz_1).$$

Proof. The proof for this is routine, it is sufficient to prove it for f_i the indicator functions. We show this for $n = 2$,

$$\begin{aligned}
\mathcal{P}(X_{t_1+s} \in A_1, X_{t_2+s} \in A_2 \mid \mathcal{F}_s) &= \mathbb{E}\Big(\mathbb{P}(X_{t_2+s} \in A_2 \mid \mathcal{F}_{t_1+s}) \mid \mathbf{1}_{X_{t_1+s}\in A_1} \mid \mathcal{F}_s\Big) \\
&= \mathbb{E}\Big(P_{t_2-t_1}(X_{t_1+s}, A_2) \mid \mathbf{1}_{X_{t_1+s}\in A_1} \mid \mathcal{F}_s\Big) \\
&= \int_{A_1} P_{t_2-t_1}(z, A_2) P_{t_1}(X_s, dz).
\end{aligned}$$

For $n \geq 2$, the analogous conclusion follows from induction. $\square$

We stress that the expectations in (1.8.3) have to be understood as integrals on the path space. To be utterly precise, (1.8.3) requires that

$$\int_A \Phi \circ \theta_{\cdot+s}(\omega)\, \mathbb{P}_x(d\omega) = \int_A \int_{\mathcal{X}^{\mathbb{R}_+}} \Phi(\omega')\, \mathbb{P}_{X_s(\omega)}(d\omega')\, \mathbb{P}_x(d\omega)$$

for all $A \in \mathcal{F}_s = \sigma(\pi_r, r \leqslant s)$.

Remark 1.8.7. If $(Y_t)_{t\geqslant 0}$ has càdlàg or continuous sample paths, we can use similar arguments as above to construct a measure on $D(\mathbb{R}_+, \mathcal{X})$ and $\mathcal{C}(\mathbb{R}_+, \mathcal{X})$, respectively. Since these spaces are however not in $\mathcal{B}(\mathcal{X})^{\otimes \mathbb{R}_+}$, this is not a simple corollary of our results and one has to work with the trace σ-fields instead.

1.9 A Markovian framework for fBm

Let us show how to enlarge the state space to render a fractional Brownian motion (fBm) Markovian. This is based on famous work of Hairer [Hai05].

We build a Markov process on the product $\mathbb{R}^d \times \mathcal{H}_H$ whose first marginal is an fBm. The space $\mathcal{H}_H$ is defined by the closure of

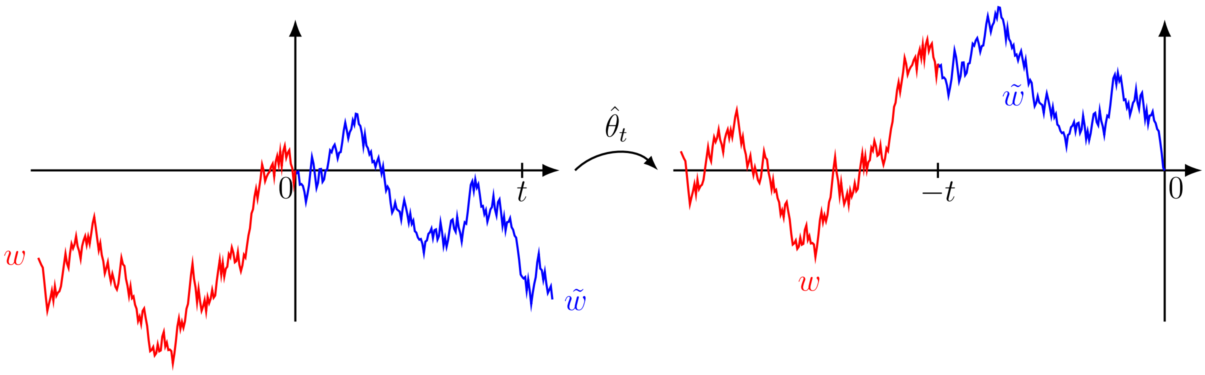
$$\mathcal{C}_0^\infty(\mathbb{R}_-, \mathbb{R}^d) = \{w : (-\infty, 0] \to \mathbb{R}^d : \, w(0) = 0, \, w \text{ is smooth and has compact support}\}$$

in the norm

$$\|w\|_H = \sup_{\substack{s \neq t \\ s,t \leqslant 0}} \frac{|w(t) - w(s)|}{|t-s|^{\frac{1-H}{2}} \sqrt{1 + |s| + |t|}}.$$

You are asked to show that $\mathcal{H}_H$ is indeed a separable Banach space on example sheet 1. Note that we fix the value of the path in $\mathcal{H}_H$ to be zero, so this norm is really a norm, not a semi-norm.

Then we define a transition function $P_t(w, \cdot)$ on $\mathcal{H}_H$ as follows: Take a Wiener path $\tilde{w}$ on $[0,t]$ and concatenate with the deterministic path w and shift it down such that the resulting function starts from the origin (see figure below). We call this operation $\hat{\theta}_t$.



Recall that fBm, centred to be 0 at 0, has the representation

$$B_t = \alpha_H \int_{-\infty}^{0} (-r)^{H-\frac{1}{2}} (dW_{t+r} - dW_r) \tag{1.9.1}$$

where α_H is a constant which we take to be 1 here. Conditioning on $(B_s, s \leqslant 0)$ is the same as conditioning on $(W_s, s \leqslant 0)$. The resulting evolution now involves a piece of Wiener process $(W_s)_{s\in[0,t]}$ independent of the history. Hence in more formal terms, the transition function is given by

$$P_t(w, \cdot) = \hat{\theta}_t^* \big(\delta_w \otimes \mathcal{L}\big((W_s)_{s\in[0,t]}\big)\big).$$

We are almost done with our construction; we just have to transform the Wiener to an fBm path. For this we make use of the representation (1.9.1). Define $\mathcal{D}_H : \mathcal{H}_H \to \mathcal{H}_{1-H}$,

$$\mathcal{D}_H(w)(t) = \int_{-\infty}^{t} (-r)^{H-\frac{1}{2}} \big(\dot{w}(r+t) - \dot{w}(r)\big)\, dr$$

initially for $w \in \mathcal{C}_0^\infty(\mathbb{R}_-, \mathbb{R}^d)$ and extended to $\mathcal{H}_H$. One checks that $\mathcal{D}_H$ is bounded and invertible, see [Hai05, Lemma 3.6]. We also want to know that the sample paths of a two sided Wiener processes belongs to $\mathcal{H}_H$ with probability 1, which follows from the following technical lemma:

Lemma 1.9.1 ([Hai05, Lemma 3.8]). *There exists a Gaussian measure* W *on* $\mathcal{H}_H$ *such that the coordinate process is a time-reversed Brownian motion.*

Now we just need to shift the path with $R_t : \mathcal{H}_{1-H} \to \mathcal{C}([0,t], \mathbb{R}^d)$, $R_t(w)(s) = w(s-t) - w(-t)$. In summary set

$$Q_t(x, w; A \times B) \triangleq \int_B \delta_{x+R_t\left(\mathcal{D}_H(w')\right)}(A)\, P_t(w, dw').$$

This defines a so-called Feller transition function, which we study in greater detail below. Moreover, the first marginal of the induced Markov process is the fBm we were after; the second marginal is the whole history of the backwards Wiener process.

1.10 Remarks

1.10.1 Treating Markov Processes with finite life time

A prominent class of Markov processes are solutions of stochastic differential equations of Markovian type. They may explode and have finite life time. Our setup excludes a Markov process with finite lifetime, to get around the problem we either ditch the requirement that $P_t(x, \mathcal{X}) = 1$ (it is customary to emphasize the condition $P_t(x, \mathcal{X}) = 1$ by referring to P as *conservative* Markov transition functions.) or enlarge the state space by adjoint an extra absorbing state Δ and define $d(x, \Delta) = 1$ for any $x \in \mathcal{X}$. Then $\hat{\mathcal{X}} = \mathcal{X} \cup \{\Delta\}$ is again a complete separable metric space. More precisely, if a stochastic process does explode (has a finite lifetime), we We define $X_t = \Delta$ for t greater or equal to its life time

$$\tau \triangleq \inf\{t \geqslant 0 : X_t = \Delta\}.$$

The Borel σ algebra on $\hat{\mathcal{X}}$ is that generated by $\{\Delta\}$ and $\mathcal{B}(\mathcal{X})$. If P_t is a family of transition measures with $P_t(x, \mathcal{X}) \leqslant 1$, we may define $\hat{P}_t$ on $\hat{\mathcal{X}}$ such that $\hat{P}_t(x, A) = P_t(x, A)$ for $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$, $\hat{P}_t(x, \{\Delta\}) = 1 - P_t(x, \mathcal{X})$ for $x \neq \Delta$ and $\hat{P}_t(\Delta, \{\Delta\})) = 1$. The canonical space contains paths $\omega : [0, \tau(\omega) \to \mathcal{X}$ where $\tau(\omega)$ is a positive number such that $\omega(t) = \Delta$ for any $t \geq \tau(\omega)$.

1.10.2 Non-time-homogeneous Markov processes

We could also define a non-time-homogeneous transition function $\{P_{s,t}(x, dy), 0 \leqslant s \leqslant t, x \in \mathcal{X}\}$, analogous to Definition 1.1.3. Then X_t is a Markov process with the transition function $P_{s,t}(x, dy)$ if $P(X_t \in A \,|\, \mathcal{F}_s) = P_{s,t}(X_s, A)$. The following self-evident claim shows that we can resort to this case in the sequel.

Exercise 1.10.1. Let X be a Markov process on $\mathcal{X}$ with transition function $P_{s,t}$. We define a family of probabilities on $\mathbb{R}_+ \times \mathcal{X}$ as below. Letting $z = (s, x) \in \mathbb{R}_+ \times \mathcal{X}$ and $d\bar{z} \triangleq d(\bar{s}, \bar{x})$,

$$\hat{P}_h(z, d\bar{z}) = \delta_{h+s}(\bar{s}) P_{s,s+h}(x, d\bar{x}).$$

Show that $\hat{P}_h$ is indeed a time-homogeneous transition function and $\hat{X}_t \triangleq (Y_t, X_t)$, where $Y_t = Y_0 + t$, is a time-homogeneous Markov process with transition function $\hat{P}_h$.

We will focus on time homogeneous Markov processes and drop the prefix 'time-homogeneous' henceforth.

Chapter 2

Semi-groups, Generators, and Martingale Problems

Instead of specifying the transition probabilities, which is impossible most of the times, we specify the probability semigroup associated to it. To this end, we need a considerable amount of semigroup theory, which we recall in its general form in Section 2.1.2, and develop for so-called Markov semigroups in this chapter.

2.1 Preliminaries and Terminologies

Please familiarise yourself with material in this section ahead of the lecture.

2.1.1 Linear Operators

The set of all linear operators between two normed space E and F is denoted by $\mathcal{L}(E,F)$, on which we define the operator norm:

$$\|T\| \triangleq \sup_{|x|_E=1} |Tx|_F = \{|Tx|_F : |x|_E \leqslant 1\} < \infty.$$

An operator T is bounded if its operator norm is bounded.

Example 2.1.1. If E is finite dimensional space, then every linear map from $E \to F$ is bounded. Indeed, let e_i be an o.n.b. basis of E, then if $x = \sum x_i e_i$,

$$|Tx| \leqslant \max |x_i| \sum_{i=1}^{n} |Te_i|.$$

Since $\max_i |x_i|$ defines a norm on E and all norms on E are equivalent, then exists a constant C such that $\max |x_i| \leqslant C|x|$ for all $x \in E$, and $\|T\| \leqslant C \sum_{i=1}^n |Te_i|$.

Proposition 2.1.2. *The following are equivalent:*

(i) T is bounded,

(ii) T is continuous,

(iii) T is continuous at 0.

If $T_1 : D_1 \to F$ and $T_2 : D_2 \to F$, where $D_1 \subset D_2$ and $T_1 = T_2$ on D_1, then T_2 is an extension of T_2 and T_1 is the restriction of T_2 on D_1. Let E, F be Banach spaces. By a 'linear operator with domain $D \subset E$', we mean T is defined only on D.

Example 2.1.3. Let $T : \mathcal{C}^1([0,2\pi]) \to \mathcal{C}([0,2\pi])$ be the derivative operator $Tf = f'$. Then T is not bounded. Take $f_n(t) = \sin(nt)$ and use $|Tf|_\infty = |f'|_\infty$.

2.1.2 Semigroups of bounded linear operators

A transition function introduces a semigroup of linear operators on $\mathcal{B}_b$, often we must work with a smaller function space with a Banach space structure. Take for example the Laplacian Δ which is essentially self-adjoint on L^2 and generates a semi-group of linear operators $e^{t\Delta}$.

Definition 2.1.4. A one parameter family of bounded linear operators $T_t : E \to E$ on a Banach space E is said to be a semigroup if

$$T(t+s) = T(t)T(s), \qquad T(0) = I, \tag{2.1.1}$$

where I is the identity,

Example 2.1.5.

- ▷ Translation on the circle $S^1 = e^{is}$: $T_t(e^{is}) = e^{i(t+s)}$.
- ▷ Let $A \in \mathcal{M}_{n\times n}$, the set of $n \times n$ matrices. Define $e^{tA} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}$ and $T_t : \mathbb{R}^n \to \mathbb{R}^n$ by $T_t x = e^{tA}x$.
- ▷ $T_t f(x) = \int_{\mathcal{X}} f(y)\, P_t(x, dy)$, $f \in \mathcal{B}_b(\mathcal{X})$, for a transition function P.
- ▷ *Translation semi-group.* Let $T_t f(x) = f(x+t)$, then T_t is a semigroup of bounded linear operators on $\mathrm{B}C(\mathbb{R};\mathbb{R})$.
- ▷ *Conditioned shift.* Let E^0 denote the space of adapted L_1 $\mathcal{F}_t$-bounded processes. Set $\|X\| = \sup_t \mathbb{E}|X_t|$. Let E be the equivalent class of functions: $X = Y$ if $\|X - Y\| = 0$. Define $T_t f(s) = \mathbb{E}[f(t+s)|\mathcal{F}_s]$, then T_t is a semi-group on E.

We semigroup generated by a Markov process has also the following properties:

Definition 2.1.6. A linear operator T on E is said to have

(i) the positive preserving property if $T_t f \geq 0$ whenever $f \geq 0$;

(ii) the conservative property if $T1 = 1$

(iii) the contractive property if $\|T\| \leqslant 1$.

A semi-group of linear operators T_t on E is said to have these properties if for each t, T_t does.

A semi-group of linear operators on $\mathcal{B}_b(\mathcal{X})$ with positive preserving and conservative property introduces a family of probability measures satisfying the Chapman-Komogorov equation and $P_0(x,\cdot) = \delta_x$. In addition, $x \mapsto P_t(\cdot, \Gamma)$ is measurable. We do not yet have the joint measurability in (t,x) required for defining a transition function, it can be easily obtained from a suitable continuity in time assumption.

If a Markov process is stochastic continuous, then for each f bounded continuous, $T_t f(x) = \mathbb{E}(f(X_t) \mid X_0 = x) \to f(x)$. Since the time in the semigroup is taken from an uncountable space, we would impose some regularity on t. A natural concept of for a semigroup T_t on a Banach space E seems to be the norm continuity: $\|T_t - I\| \to 0$, however it is rare that a semi-group of interest is uniformly continuous. The continuity os the image $T_t f$ where $f \in E$ is more suitable.

Definition 2.1.7. A semigroup of bounded linear operators on a Banach space E is uniformly continuous if

$$\|T_t - I\| = \sup_{|x|=1} |T_t x - x| \to 0,$$

as $t \searrow 0$. It is called *strongly continuous* if

$$\lim_{t \searrow 0} |T_t x - x| = 0$$

for each $x \in E$.

If A is a $n \times n$ matrix, $|\exp(tA)x - x| = t|A \sum_{n=1}^{\infty} \frac{(tA)^n}{n!} x| \to 0$ uniformly in x.

Example 2.1.8. An example of a non-strongly continuous semigroup on $BC(\mathbb{R}; \mathbb{R})$ is: $T_0 = I$ and $T_t = 0$ for $t > 0$.

Exercise 2.1.9. Show that the translation semi-group is not strongly continuous on $BC(\mathbb{R}; \mathbb{R})$ either. Identify its generator and a space on which it is strongly continuous.

Definition 2.1.10. Let T be a strongly continuous semigroup. We define its generator by

$$\mathcal{L}x \triangleq \lim_{t \searrow 0} \frac{T_t x - x}{t} \tag{2.1.2}$$

if the limit exists. The domain of $\mathcal{L}$ is then defined by

$$\mathcal{D}(\mathcal{L}) \triangleq \{x \in E : \text{ the limit (2.1.2) exists}\}.$$

It is clear that on $\mathcal{D}(\mathcal{L})$ we necessarily have that $T_t x \to x$ as $t \to 0$. If T_t is a contractive semigroup, $\lim_{t \searrow 0} \frac{T_t x - x}{t}$ exists on a dense subset, Proposition 2.1.15 implies that T_t is strongly continuous.

2.1.3 Uniformly continuous semi-groups

Bounded linear operators resembles matrix operators. A semigroup of bounded linear operators on a Banach space is uniformly continuous if only if it is of the form $T_t = e^{tA}$ where A is a bounded linear operator on E. Good accounts of the semi-group theory can be found in [EN00, DS88, Paz83].

Proposition 2.1.11. *If $A : E \to E$ is a bounded linear operator, then e^{tA} is a uniformly continuous semigroup and*

$$\frac{d}{dt} T_t = A T_t, \qquad T_0 = I.$$

Conversely every uniformly continuous semi-group on a Banach space if of the form $T_t = e^{tA}$ for some bounded linear operator $A : E \to E$.

The first statement follows from that e^{tA} is norm convergent. Given T_t, we observe that $U(t) \triangleq \int_0^t T_s ds$ is a family bounded linear operator, and differentiable in t with $\dot{U}(t) = T_t$. Also, $\frac{1}{t} U(t)$ converges as $t \downarrow 0$ in norm to the identity, hence it is invertible on $[0, a]$ for some $a > 0$ and differentiable at 0. Then

$$T_t = U(t_0)^{-1} U(t_0) T_t = U(t_0)^{-1} \int_0^{t_0} T_{t+s} ds = U(t_0)^{-1} \int_t^{t+t_0} T_s ds.$$

Consequently,

$$T_t = U(t_0)^{-1} \left(U(t + t_0) - U(t) \right)$$

is differentiable. Set $A = \frac{d}{dt} T_t|_{t=0}$ It is then easy to verify that

$$\frac{d}{dt} T_t = \lim_{h \to 0} \frac{T_{t+h} - T_t}{h} = \lim_{h \to 0} \frac{T(h) - I}{h} T_t = \dot{T}(0) T_t = A T_t,$$

for all $t > 0$, $\frac{d}{dt} T_t = A T_t$.

2.1.4 Strongly continuous semi-groups

We don't expect a Markov semigroup uniformly continuous, however the generator of a strongly continuous M-dissipative semigroup is the limit of bounded operators. This is the content of Hille-Yosida theorem which we will discuss later.

Example 2.1.12. Let

$$T_t f(x) \triangleq \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(y) e^{-\frac{|y-x|^2}{2t}} \, dy.$$

(i) Then T is not strongly continuous on $\mathcal{B}_b(\mathbb{R})$. Indeed, let $f(y) = \mathbf{1}_{\{0\}}(y)$. Then $T_t f = 0$ for all $t > 0$ and $|T_t f - f|_\infty = 1$, concluding that T_t is not strongly continuous on $\mathcal{B}_b$. The same conclusion holds for any Markov process whose transition function $P_t(x, \cdot)$ is absolutely continuous with respect to the Lebesgue measure.

(ii) The heat semigroup will restricts to a semi-group on $BC(\mathbb{R}^n)$, it is however not strongly continuous. Let $f(x) = \sin(x^2)$.

$$T_t f(x) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} \sin(y^2) e^{-\frac{|y-x|^2}{2t}} \, dy \to 0,$$

for every $t > 0$, this converges to zero as $x \to \infty$, hence as $t \to 0$, $|T_t f - f|_\infty$ does not converge.

Proposition 2.1.13. *If T_t is a strongly continuous semigroup, there exist constants $M \geqslant 1, \omega \geqslant 0$ such that $\|T_t\| \leqslant M e^{\omega t}$.*

Proof. We first show that there exist $a > 0, M > 0$ with $\sup_{t \in [0,a]} \|T_t\| \leqslant M$. If not there exists a sequence $t_n \to 0$ with $\|T_{t_n}\| \to \infty$. But by the continuity $\sup_n |T_{t_n} x|$ is bounded for every x, this contradicts the Uniform Boundedness Principle. For any t suppose that $t = Na + \delta$. Then $\|T_t\| \leqslant \|T_a\|^N \|T_\delta\| = M \cdot M^N$. Set $\omega = \ln M$ to conclude. □

Exercise 2.1.14. Let $T_t : E \to E$ be a semigroup on a Banach space E. Show that if T_t is strongly continuous then for any $x \in E$, the map $t \mapsto T_t x \in E$ is continuous on $(0, \infty)$.

Proposition 2.1.15. *The following statements are equivalent for a semigroup T_t on a Banach space E.*

(i) T_t is strongly continuous.

(ii) There exists $\delta > 0$, and a number $M > 0$ such that

$$\sup_{t \in [0,\delta]} \|T_t\| \leqslant M$$

and there exists a dense subset D of E such that $\lim_{t \downarrow 0} T_t x = x$, for every $x \in D$.

Proof. Suppose the (ii) holds. Let $x \in E$. Let $\varepsilon > 0$. Let $y \in D$ with $|y - x| \leqslant \frac{\varepsilon}{3M+3}$. Choose $0 < \delta_0 < \delta$ such that for $t < \delta_0$, $|T_t y - y| < \varepsilon/2$. Then

$$|T_t x - x| \leqslant |T_t x - T_t y| + |T_t y - y| + |y - x| \leqslant M \frac{\varepsilon}{3M+3} + \varepsilon/3 + \frac{\varepsilon}{3}.$$

concluding the continuity of $T_t x$ at $t = 0$. □

Example 2.1.16. The heat semi-group is a strongly continuous semi-group on C_0. We first take $f \in C_K^\infty$. Since $T_t f(x) = \mathbb{E} f(x + B_t)$ and

$$f(x + B_t) = f(x) + \frac{1}{2} \int_0^t \Delta f(x_s) ds + \int_0^t df(x + B_s) dB_s.$$

Taking expectation we obtain

$$T_t f(x) = f(x) + \frac{1}{2}\int_0^t T_s f''(x) ds.$$

Suppose f has compact support K, $T_s f''$ converges uniformly on K. Hence T_t is a strongly continuous on C_K^∞, which is a dense subset of C_0. Since $\|T_t\| \leqslant 1$, the conclusion holds.

Remark 2.1.17. It is a very deep and surprising fact that a semigroup is strongly continuous iff it is weakly continuous in the sense that $\langle f^*, T_t f\rangle \to \langle f^*, f\rangle$ for each $f \in E$ and $f^* \in E^*$. See [EN00, pp40].

2.1.5 Equivalent formulations: semi-groups, generators, and resolvents

If $\mathcal{L} : \mathcal{D}(\mathcal{L}) \subset E \to \mathbb{E}$ is a closed linear operator on a Banach space E, we denote by $\rho(\mathcal{L})$ the resolvent set of $\mathcal{L}$, it is the set of complex numbers λ such that $\lambda I - \mathcal{L}$ is a bijection (therefore has a bounded inverse):

Definition 2.1.18. $\triangleright$ A number λ is said to be in the resolvent set $\rho(\mathcal{L})$ of $\mathcal{L}$, if $\lambda - \mathcal{L}$ is a bijection with bounded inverse.

The spectral of $\mathcal{L}$ is then defined to be $\mathcal{C}\backslash\rho(\mathcal{L})$. Given $\lambda \in \rho(\mathcal{L})$ the operator $(\lambda - \mathcal{L})^{-1}$ is called the resolvent of $\mathcal{L}$ at λ.

Remark 2.1.19. If $\mathcal{L}$ is a closed linear operator on a Banach space E and $\lambda - \mathcal{L}$ has an inverse, then its inverse is closed and bounded.

Proof. (1) An operator is closed $\Leftrightarrow$ its graph is a closed subset of $E \times E$. (2) Since A is closed, so is $(\lambda - A)^{-1}$ (A pair (f, g) is in the graph of $\lambda - A$ precisely when (g, f) is in the graph of $(\lambda - A)^{-1}$.) (3) Since $(\lambda - A)^{-1}$ is defined on the whole space E, it is bounded by the closed graph theorem (A linear map from a Banach space to another is bounded if and only if its graph is closed). $\square$

Example 2.1.20. Let A be a bounded linear operator then, $\sum_{n=0}^\infty \frac{A^n}{\lambda^{n+1}}$ is seen to be the inverse of $\lambda - \mathcal{L}$, agreeing with the formal expression $(\lambda - \mathcal{L})^{-1} = \frac{1}{\lambda}(1 - \frac{A}{\lambda})^{-1}$.

Definition 2.1.21. A family of bounded linear operators $\{R_\lambda, \lambda > 0\}$ of bounded linear operators on E is a strongly continuous contraction resolvents if the following holds

(i) $\lim_{\lambda\to\infty} \lambda R_\lambda x = x$ for any $x \in E$;

(ii) λR_λ is a contraction for every $\lambda > 0$.

(iii) $R_\lambda - R_\mu = (\lambda - \mu) R_\lambda R_\mu$ for any $\lambda, \mu > 0$.

Example 2.1.22. If T_t is a strongly continuous contraction semi-group, we define

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} T_t f(x) dt.$$

Then $R_\lambda f = (\lambda - \mathcal{L})^{-1} f$.

We can now state the connections between these concepts.

Theorem 2.1.23. *There is a one to one correspondence among the following objects:*

(i) A strongly continuous contraction semi-group on E;

(ii) a densely defined closed linear operator (as its generator) $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ such that $(0, \infty) \subset \rho(\mathcal{L})$ and such that $\mathcal{L}$ is maximaly dissipative.

(iii) A strong continuous contraction resolvent.

They are related by the formula:

$$\mathcal{L}x = \lim_{t \to 0} \frac{T_t x - x}{t}, \qquad R_\lambda f = \int_0^\infty e^{-\lambda t} T_t f dt, \qquad T_t = \lim_{\lambda \to \infty} e^{t\lambda(\lambda R_\lambda - 1)}.$$

Finally, if E is a Hilbert space these are in one to one correspondence with coersive closed bilinear form on a dense subset of H.

Le D be a linear subspace of a Hilbert space and $\mathcal{E} : D \times D \to \mathbb{R}$ a bilinear map. For any $a > 0$, define

$$\mathcal{E}_a(u, v) = \mathcal{E}(u, v) + a\langle u, v\rangle.$$

We say $\mathcal{E}$ is closed if its domain D is complete in E under the norm $\mathcal{E}_1$. It is positive if $\mathcal{E}(u, u) \geq 0$ for any $u \in D$.

Definition 2.1.24. We say $\mathcal{E}$ is a coercive symmetric closed form if D is dense in H and and satisfies the coercive condition: there exists a number K such that

$$|\mathcal{E}_1(u, v)| \leqslant K(|\mathcal{E}_1(u, u)|)^{\frac{1}{2}} K(|\mathcal{E}_1(, v)|)^{\frac{1}{2}}, \quad \forall u, v \in D. \tag{2.1.3}$$

We say it is a coercive closed form its symmetric part is a symmetric closed form and the coercive condition (2.1.3) holds for $\mathcal{E}_1$.

Theorem 2.1.25. *[MR91] Let $\mathcal{E}$ be a coercive closed form on $D \subset H$, then there exists two strongly continuous contraction resolvents G_α and $\hat{G}_\alpha$ with $G_\alpha(H) \subset D(\mathcal{E})$ and $\hat{G}_\alpha(H) \subset D(\mathcal{E})$ and such that*

$$\mathcal{E}_\alpha(G_\alpha f, g) = \langle f, g\rangle = \langle f, \hat{G}_\alpha g\rangle$$

for any $f \in H$ and $g \in \mathcal{D}$. Furthermore,

$$\langle G_\alpha f, g\rangle = \langle f, \hat{G}_\alpha g\rangle.$$

Also if L denotes the linear operator corresponds to G_α, then $Dom(\mathcal{L}) \subset D$ and

$$\mathcal{E}(f, g) = \langle Lf, g\rangle.$$

Definition 2.1.26. A coercive closed form is a Dirichlet form if for any $f \in D$, or any $f \in Dom(\mathcal{L})$ and $g \in D$ one has $f^+ \wedge 1 \in D$,

$$\mathcal{E}(f + f^+ \wedge 1, f - f^+ \wedge 1) \geq 0$$

and

$$\mathcal{E}(f - f^+ \wedge 1, f + f^+ \wedge 1) \geq 0.$$

If $\mathcal{E}$ is symmetric, the last two condition is equivalent to

$$\mathcal{E}(f^+ \wedge 1, f^+ \wedge 1) \leqslant \mathcal{E}(f, f).$$

We conclude this brief outline by nnoting that there is a Dirichlet form theory equivalent to the continuous contraction semi-group theory. The Dirichlet property of the coercive closed form corresponds to the sub-Markovian property; the coercive property corresponds to the contraction property and we expect $\mathcal{E}(fg) = \langle -\mathcal{L}f, g\rangle$ for $f \in Dom(\mathcal{L})$ and $g \in D$.

2.1.6 A mini Project

A possible project is to study Dirichlet forms and connect them to some Markov process models. I shall describe another one.

Suppose that $\int f d\mu = \int T_t f d\mu$ for any bounded continuous functions, so $\int f d\mu = \int f d(T_t)^* \mu$, then μ is an invariant measure for T_t.

Definition 2.1.27. Let μ be a probability measure. A bounded linear operator T on $L^2(\mu)$ is a Markov operator if

(i) $f \geq 0$ ae implies that $Tf \geq 0$ a.e.

(ii) $T1 = 1$

(iii) $T^*1 = 1$.

The last properties is the same as $\int f d\mu = \int T_t f d\mu$. We think of T as a generalisation of a doubly stochastic matrix.

Can T be an isometry? Isometry means that $\langle Tf, Tg \rangle_{L^2} = \langle f, g \rangle_{L^2}$ (It is unitary if it is furthermore surjective).

It turns out that a Markov operator on $L^2(\mathcal{X}, \mu)$ is an isometry if and only if the restriction of T on L^∞ is multiplicative (i.e. $T(fg) = TfTg$.) See [Arv86]– Markov operators and OS-positive processes.

An example of a multiplicative Markov operator is given by $Tf = f \circ S$ where $S : \mathcal{X} \to \mathcal{X}$ is a measure preserving transformation, $S^* \mu = \mu$, it is an isometry following from it being a measure preserving transformation. An example of a measure preserving transformation is $Se^{i\theta} = e^{2i\theta}$ on S^1 (This is the doubling map on $[0, 1]$, $S(t) = 2t$ mod 1.) Let $\mathcal{X} = \{0, 1\}$, $\mu(\{1\}) = 1$. Let S takes elements of $\mathcal{X}$ to 1. It is a measure preserving transformation. In this case $Tf(x) = f \circ S = f(1)$ for any $x \in \mathcal{X}$.

Any Markov process with $T_1 f = f \circ S$ where S is a measure preserving transformations is of the form $P_1(x, \cdot) = \delta_x$. An interesting concept 'totally non-deterministic process' is associated to aperiodicity. See [Ver05] – What does a generic Markov operator look like?

2.1.7 The Generator of a strongly continuous semi-group

An unbounded linear operator on a Banach space E is never defined on the whole space. It is useful to know the domain of the generator, which is however often tricky to identify. The domain can be thought of as 'smooth' functions. The semigroup T_t is thought of to smooth out a function, or at least not to rough it, for $t > 0$. Similarly, integration $\int_0^t$ is a smoothing operation. The integral $\int_0^t T_s x ds$ is defined by Riemann sum on E.

Theorem 2.1.28. *Let $T_t : E \to E$ be a strongly continuous semigroup on a Banach space E. Let $\big(\mathcal{L}, \mathcal{D}(\mathcal{L})\big)$ denote its generator. Then the following hold:*

(i) If $x \in E$ and $t > 0$, then

$$\int_0^t T_s x \, ds \in \mathcal{D}(\mathcal{L})$$

and

$$T_t x - x = \mathcal{L}\left(\int_0^t T_s x ds\right).$$

(ii) If $x \in \mathcal{D}(\mathcal{L})$, then $T_t x \in \mathcal{D}(\mathcal{L})$ for any $t > 0$ and

$$\frac{d}{dt} T_t x = T_t \mathcal{L} x = \mathcal{L}(T_t x).$$

(iii) $\mathcal{D}(\mathcal{L})$ is dense in E and $\mathcal{L}$ is closed.

Proof. (i) We have

$$\frac{1}{h}\left(T_h\int_0^t T_s x\,ds - \int_0^t T_s x\,ds\right) = \frac{1}{h}\left(\int_h^{t+h} T_s x\,ds - \int_0^t T_s x\,ds\right)$$
$$= \frac{1}{h}\int_t^{t+h} T_s x\,ds - \frac{1}{h}\int_0^h T_s x\,ds \to T_t x - x$$

as $h \searrow 0$ since $t \mapsto T_t x$ is continuous.

(ii) If $x \in \mathcal{D}(\mathcal{L})$ and $t > 0$, then

$$\frac{T_h T_t x - T_t x}{h} = T_t \frac{T_h x - x}{h} \to T_t \mathcal{L} x$$

by continuity of T_t. Hence, $T_t x \in \mathcal{D}(L)$ and $\mathcal{L} T_t x = T_t \mathcal{L} x$. Moreover,

$$\frac{d}{dt} T_t x = \lim_{h\to 0} \frac{T_{t+h}x - T_t x}{h} = T_t \mathcal{L} x = \mathcal{L} T_t x.$$

(iii) Since $\frac{1}{t}\int_0^t T_s x ds \in \mathcal{D}(\mathcal{L})$ for each $x \in E$ and $t > 0$ and

$$x = \lim_{h\to 0} \frac{1}{h}\int_0^h T_s x\,ds,$$

we see that $x \in \overline{\mathcal{D}(\mathcal{L})}$.

Finally we show that $\mathcal{L}$ is closed. Let $(x_n) \subset \mathcal{D}(\mathcal{L})$, $x_n \to x$, and suppose that $\mathcal{L} x_n \to y$. Then, by (ii),

$$T_t x_n - x_n = \int_0^t T_s \mathcal{L} x_n ds.$$

Taking $n \to \infty$, we see that $T_t x - x = \int_0^t T_s y ds$ and $\frac{T_t x - x}{t} \to y$. Thus, $x \in \mathcal{D}(\mathcal{L})$ and $Lx = y$. Consequently, $\mathcal{L}$ is closed. ☐

The following theorems shows that the generator of a strongly continuous semigroup determines it.

Theorem 2.1.29. *Let T_t and S_t be strongly continuous semigroups of bounded linear operators with the same generator, then $T_t = S_t$ for all $t \geq 0$.*

Proof. Let us denote the generator by L. Since $\mathcal{D}(L)$ is dense, and T_t, S_t are continuous linear operators, it is sufficient to show that $T_t = S_t$ on $\mathcal{D}(L)$. Note that $S_0 = T_0$. Take x in the domain, then, for each $r \geqslant 0$, $S_r x, T_r x \in \mathcal{D}(\mathcal{L})$. Hence,

$$\frac{d}{ds} S_{t-s} T_s x = -\mathcal{L} S_{t-s}(T_s x) + S_{t-s}(\mathcal{L} T_s x) = 0.$$

In the last line, we used part (ii) of Theorem 2.1.28 to commute L and its generator. This means that $s \mapsto S_{t-s} T_s x$ is a constant, concluding the proof. ☐

With stochastic differential equations of Markovian type, on a manifold without a boundary, it is easy to extract the formal generator, we hope knowing the generator on $\mathcal{C}_K^\infty$ is sufficient to identify the transition functions. Then if the martingale problem is well posed we are in good business.

Example 2.1.30. Let $T_t f(x) = \mathbb{E} f(x + B_t)$, where B_t is an n dimensional Brownian motion. Then for $f \in C^2$,

$$T_t f(x) = \frac{1}{\sqrt{2\pi t}^n} \int f(x+y) e^{-\frac{|y|^2}{2t}} dy = \frac{1}{\sqrt{2\pi}^n} \int f(x + \sqrt{t} y) e^{-\frac{|y|^2}{2}} dy.$$

Taylor expand around x gives, for some $s \in [0, 1]$,

$$T_t f(x) - f(x) = \frac{1}{\sqrt{2\pi}^n} \int \sqrt{t} \langle \nabla f(x), y \rangle + \frac{1}{2t} \langle \nabla^2 f(x + s\sqrt{t} y) y, y \rangle e^{-\frac{|y|^2}{2}} dy.$$

Using the mean zero property, for $f \in C_K^2$,

$$\frac{T_t f(x) - f(x)}{t} = \frac{1}{\sqrt{2\pi}^n} \int \frac{1}{2} \langle \nabla^2 f(x + s\sqrt{t} y) y, y \rangle e^{-\frac{|y|^2}{2}} dy \to \frac{1}{2} \operatorname{tr} \nabla^2 f(x) = \frac{1}{2} \Delta f(x).$$

Exercise 2.1.31. Check that T_t preserves the space $\mathcal{C}_0(\mathbb{R}^n)$.

We emphasise that for two operators to be equal, their domains must be equal. It is, as with solutions of a stochastic differential equation, sometimes easier to describe the formal generator of a Markov process, determining the domain is more subtle. The role played be the domain of a generator is significant. For example, let $\mathcal{X} = [0, 1]$ and let $L = \frac{1}{2} \frac{d^2}{dx^2}$, depending on their domains we may have a reflected Brownian motion or a killed Brownian motion.

Example 2.1.32. (BM on $\mathbb{R}_+$, Reflecting boundary) How do we keep a Brownian motion starting with $x > 0$ in $[0, \infty)$? One way is to reflect it back. The reflected Brownian motion behaves like a Brownian motion while away from 0, at 0, it moves only to the right. A Brownian motion on $\mathbb{R}$ with initial condition x reflected at 0 behaves like a Brownian motion from x before hitting 0, at 0 it reflects immediately, so it spent 0 time on the boundary ($\int_0^t \mathbf{1}_{\{0\}}(X_s) ds = 0$.) A realisation of the reflected Brownian motion is: $x + B_t|$.

Exercise. Show that $|x + B_t|$ is a Markov process and the semi-group is: for $x \geq 0$,

$$T_t f(x) = \frac{1}{\sqrt{2\pi t}} \int_0^\infty f(y) e^{-\frac{|y-x|^2}{2t}} + e^{-\frac{|y+x|^2}{2t}} dy.$$

Then, $\mathcal{L} f = \frac{1}{2} f''$ with domain:

$$\{f \in C_0(\mathbb{R}) : f' \in C_0(\mathbb{R}_+), f'' \in C_0(\mathbb{R}_+), f'(0) = 0\}.$$

Before closing this section, note that it is remarkable that a strongly continuous semi-group on E is automatically differentiable on a dense set of E and on which x solves the equation:

$$\frac{d}{dt} T_t x = \mathcal{L} T_t x.$$

As we will see later it is often easy to identify the form of the generator for the semi-group corresponds to a Markov process on the class C_K^∞, the space of smooth functions on the compact support, should the space has no boundary. Then for such functions $T_t f$ solves the Cauchy problem $\frac{d}{dt} u = u$ with $u(0, \cdot) = f$.

Definition 2.1.33. The Markov uniqueness problem concerns whether there exists a unique Markov process on the continuous path space over a complete Riemannian manifold such that its Markov generator is the infinite dimensional Laplacian. This is an open problem.

2.1.8 Preliminary to the next section

A linear operator (matrix) on $\mathbb{R}^n$ is invertible is equivalent to one of the following statements: (1) A is injective, (2) A is surjective, (3) there exists a left inverse, (4) there exists a right inverse.

Example 2.1.34. Consider the shift operator T on bounded sequences such that

$$A((a_0, a_1, \dots)) = (a_1, a_2, \dots).$$

Then it is surjective, not injective, and it has a right inverse $B : (a_0, a_1, \dots) \mapsto (0, a_0, a_1, \dots)$, but not a left inverse.

Proposition 2.1.35. *If $A : E \to E$ is a linear operator on a Banach space with $\|A\| < \lambda$ then $\lambda - A$ is invertible and*

$$(\lambda - A)^{-1} = \frac{1}{\lambda} \sum_{n=0}^{\infty} \frac{A^n}{\lambda^n}.$$

If $T : E \to F$ is bounded then its kernel is closed. Unlike in finite dimensions, the range of a bounded linear operator needs not be closed.

Example 2.1.36. Let $E = L^1(\mu)$ where $\mu = p(x)dx$ is a probability measure on $\mathbb{R}$ with second moment, but with its third moment ∞. We may assume that $p > 0$. Let $g(x) = \frac{1}{1+x^4}$. We define the multiplication operator $T : E \to E$ by $T = fg$. Smooth functions with compact supports are in $L^1(\mu)$, but the constant function 1 is not in the range of T.

2.1.9 Resolvent Operator

Having seen that a strongly continuous semi-group T_t on a Banach space E is determined by its generator $\mathcal{L}$ (which is always densely defined and closed), we define the resolvent operator $(R_\lambda, \lambda \geq 0)$ of the semi-group and show that it is the inverse to $\lambda - \mathcal{L}$.

Definition 2.1.37. For any $\lambda > 0$, we define $R_\lambda : E \to E$ by

$$R_\lambda x = \int_0^\infty e^{-\lambda s} T_s x ds, \qquad \forall x \in E.$$

This is an improper integral using the strong continuity of T_s and that

$$\int_0^\infty e^{-\lambda s} |T_s x| ds \leqslant |x| \int_0^\infty e^{-\lambda s} ds < \infty.$$

This also shows give the norm bound: $\|R_\lambda\| \leqslant \frac{1}{\lambda}$.

Proposition 2.1.38. *If T_t is a strongly continuous contraction semi-group E, then R_λ is a strongly continuous contraction resovlent on E.*

Proof. We have seen already $|\lambda R_\lambda\| \leqslant 1$, we next show the continuity:

$$|\lambda R_\lambda x - x| = \Big| \int_0^\infty \lambda e^{-\lambda s} T_s x ds - \int_0^\infty e^{-s} x ds \Big| = \int_0^\infty e^{-u} \big| T_{u/\lambda} x - x \big| du,$$

passing limit inside the integral by the contraction property of T_t and dominated convergence. Finally let τ^λ, τ^μ be independent exponentially distributed random variables on $\mathbb{R}$ with parameter $\lambda > 0, \mu > 0$ respectively. Then, $\mathbb{E} T_{\tau^\lambda} x = \int_0^\infty T_s x \lambda e^{-\lambda} ds = \lambda R_\lambda x$ and

$$\mathbb{E} T_{\tau^\lambda} T_{\tau^\mu} x = \lambda \mu R_\lambda \mathbb{R}_\mu.$$

Now $\tau_1 + \tau_2$ is distributed as

$$\frac{\lambda\mu}{\lambda - \mu}(e^{-\lambda s} - e^{-\mu s})ds.$$

Using the semigroup property,

$$\frac{\lambda\mu}{\lambda - \mu}(R_\lambda - R_\mu) = \lambda\mu R_\lambda \mathbb{R}_\mu$$

proving the resolvent equation $R_\lambda - R_\mu = (\lambda - \mu)R_\lambda R_\mu$. $\square$

Example 2.1.39. If $T_t f(x) = \int_{\mathcal{X}} f(y)P_t(x, dy)$ on $\mathcal{B}_b(\mathcal{X})$ be given by a transition function. Then

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} f(x)dt.$$

Observe that $0 \leqslant f \leqslant 1$ implies that $0 \leqslant R_\lambda f \leqslant 1$. Also the conservative property $T_t 1 = 1$ is equivalent to $R_\lambda 1 = \frac{1}{\lambda}$.

Proposition 2.1.40. *Let T_t a strongly continuous contraction semi-group E with generator $\mathcal{L}$, then the following statements hold for any $\lambda > 0$.*

(i) For any $x \in E$, $R_\lambda x \in \mathcal{D}(\mathcal{L})$;

(ii) For any $x \in \mathcal{D}(\mathcal{L})$, $\mathcal{L}R_\lambda x = R_\lambda \mathcal{L}x$.

(iii) Any number $\lambda > 0$ belongs to the resolvent set $\rho(\mathcal{L})$ and $R_\lambda = (\lambda - \mathcal{L})^{-1}$. Consequently,

$$\|(\lambda - \mathcal{L})^{-1}\| \leqslant \frac{1}{\lambda}.$$

Proof. (1) Let $\lambda > 0$, and $x \in$E, by the contractive property,

$$\|R_\lambda x\| = \left\| \int_0^\infty e^{-\lambda t} T_t x \, dt \right\| \leqslant \int_0^\infty e^{-\lambda t} \, dt \, \|x\| \leqslant \frac{1}{\lambda}\|x\|, \tag{2.1.4}$$

hence $R_\lambda x$ is well defined. For any $h > 0$,

$$\begin{aligned}
\frac{T_h - I}{h} R_\lambda x &= \frac{1}{h}\int_0^\infty e^{-\lambda t}(T_h T_t x - T_t x)dt \\
&= \frac{1}{h}\int_h^\infty e^{-\lambda(t-h)} T_t x dt - \frac{1}{h}\int_0^\infty e^{-\lambda t} T_t x dt \\
&= \frac{e^{\lambda h} - 1}{h}\int_0^\infty e^{-\lambda t} T_t x dt - \frac{1}{h}\int_0^h e^{-\lambda t} T_t x dt \\
&\overset{(h \to 0)}{\longrightarrow} \lambda R_\lambda \, x - x.
\end{aligned}$$

Hence $R_\lambda x \in \mathcal{D}(\mathcal{L})$ and

$$\mathcal{L}(R_\lambda x) = \lambda R_\lambda \, x - x, \tag{2.1.5}$$

proving

$$(\lambda - \mathcal{L})R_\lambda = I_E.$$

So, $\lambda - \mathcal{L}$ is injective on $\mathcal{D}(\mathcal{L})$, and R_λ is the right inverse.

For $x \in \mathcal{D}(\mathcal{L})$,

$$R_\lambda(\mathcal{L}x) \overset{\text{definition}}{=} \lim_{s\to\infty} \int_0^s e^{-\lambda t} T_t(\mathcal{L}x)\, dt$$

$$= \lim_{s\to\infty} \int_0^s \mathcal{L}(e^{-\lambda t} T_t x)\, dt = \lim_{s\to\infty} \mathcal{L}\Big(\overbrace{\int_0^s T_t(e^{-\lambda t} x)\, dt}^{R_\lambda^s} \Big).$$

We used part (i) of Theorem 2.1.28. Since

$$R_\lambda^s \to R_\lambda x, \qquad \mathcal{L}(R_\lambda^s) \to R_\lambda \mathcal{L} x,$$

and $\mathcal{L}$ is closed by Theorem 2.1.28, $\mathcal{L}(R_\lambda^s) \to \mathcal{L} R_\lambda$, concluding

$$R_\lambda \mathcal{L} x = \mathcal{L} R_\lambda x, \qquad R_\lambda(\lambda - \mathcal{L}) = I_{\mathcal{D}(\mathcal{L})},$$

the latter follows from (2.1.5). Thus, $\mathrm{Range}(\lambda - \mathcal{L}) = E$, and $(\lambda - \mathcal{L})^{-1} = R_\lambda$. □

If λ is a complex number with strictly positive real part, R_λ is well defined, which allows to conclude that $\rho(\mathcal{L})$ is contained in the open right half of the complex plane. Strictly speaking, for this we should complexify the Banach space and extend the operator to the complexification by $\tilde{\mathcal{L}}(x+iy) = \mathcal{L}x + i\mathcal{L}y$. Note that $\lambda - \mathcal{L}$ being injective, surjective, invertible, as well as its boundedness are the same for $\mathcal{L}$ and $\tilde{\mathcal{L}}$. With this set up, the proof above leads to:

Corollary 2.1.41. *Let $\mathcal{L}$ be the generator of a strongly continuous contraction semigroup on E. Then $\rho(\mathcal{L}) \supset \{\lambda : Re(\lambda) > 0\}$, for such λ,*

$$\|(\lambda - \mathcal{L})^{-1}\| \leqslant \frac{1}{Re(\lambda)}.$$

Example 2.1.42. Let $E = \{f : \mathbb{R}_+ \to \mathbb{R}_+ : \text{bounded and uniformly continuous}\}$, then $T_t f(x) = f(x+t)$ defines a strongly continuous contraction semi-group on E. If $\lambda = -a + bi$ with $a < 0$, then $f(t) = e^{\lambda t} \in E$ and in $\mathcal{D}(\mathcal{L})$. Now, $T_t f = e^{\lambda t} f$ and $\mathcal{L} f = \lambda f$, so the resolvent set $\rho(\mathcal{L})$ is the right half of the plane.

2.1.10 M-Dissipative operators

In this section we show the Hille-Yosida theorem: a closed and densely defined linear operator $\mathcal{L}$ on a Banach space E is the generator of a strongly continuous contraction semigroup on E if and only if it is M-dissipative.

If A is a symmetric matrix and λ is in its resolvent set, then one expects that

$$\|(\lambda - A)^{-1}\| \leqslant \frac{1}{d(\lambda, Spe(A))}.$$

For an unbounded operator, we do not expect this to hold. We make an assumption of this nature.

Definition 2.1.43. Consider a linear operator $A : \mathcal{D}(A) \subset E \to E$.

▷ A is said to be dissipative if

$$\|(\lambda - A)x\| \geq \lambda \|x\| \quad \forall x \in \mathcal{D}(A),\ \forall \lambda > 0.$$

▷ A is said to be M-dissipative (maximal dissipative) if for any $\lambda > 0$, $\lambda - A$ has an inverse and

$$\|(\lambda - A)^{-1}\| \leqslant \frac{1}{\lambda}. \tag{2.1.6}$$

If A is M-dissipative, it is clearly dissipative. Indeed,

$$\|(\lambda - A)^{-1}x\| \leqslant \frac{1}{\lambda}\|x\|, \quad \forall x \in E, \ \forall \lambda > 0,$$

For any $g \in \mathcal{D}(A)$, simply replace x in the M-dissipative inequality with $(\lambda - A)g$.

Exercise 2.1.44. Suppose that A is closed and $(\lambda - A)$ is invertible any $\lambda > 0$. Show that A is dissipative if and only if A is M-dissipative.

Let E be a Hilbert space, and $A : E \to E$ a densely defined linear operator. Its adjoint operator is defined on the set of x such that there exists an element of E which we denote by A^*x with

$$\langle A^*x, y\rangle = \langle x, Ay\rangle, \qquad \forall\, y \in \mathrm{Dom}(A).$$

We say A is self-adjoint if $A^* = A$. If A is a self-adjoint operator on a Hilbert space, being dissipative means $\langle Ax, x\rangle \leqslant 0$. This agrees with our intuition that A is sort of a generalisation of a symmetric negative definite matrix (A self-adjoint operator is called negative definite if $\langle x, Ax\rangle \leqslant 0$ for any $x \in \mathrm{Dom}(A)$). The following theorem holds, [Paz83]:

Theorem 2.1.45. *Let A be closed and densely defined. Suppose that both A and its dual A^* dissipative, then A is the generator of a strongly continuous semi-group.*

We recall that the generator of a strongly continuous semigroup is dense. Anyhow, if it is not dense we could think of getting ride of the superfluous parts.

Recall that $\rho(\mathcal{L}) = \{\lambda \in \mathcal{C} : (\lambda - \mathcal{L}) : E \to E \text{ is bijection}\}$. If $\lambda \in \rho(\mathcal{L})$ we denote $R_\lambda = (\lambda - \mathcal{L})^{-1}$ its inverse. Then $\mathcal{L}R_\lambda = \lambda R_\lambda - id$. M-dissipative means $\|R_\lambda\| \leqslant \frac{1}{\lambda}$.

Lemma 2.1.46. *Let $\mathcal{L} : E \to E$ be a M-dissipative, closed, and densely defined operator. Then,*

$$\lim_{\lambda\to\infty} \lambda R_\lambda\, x = x, \qquad \forall x \in E.$$

Consequently, for every $x \in \mathcal{D}(\mathcal{L})$,

$$\mathcal{L}x = \lim_{\lambda\to\infty} \lambda \mathcal{L} R_\lambda\, x.$$

Proof. Let $x \in \mathcal{D}(\mathcal{L})$ and denote $R_\lambda = (\lambda - \mathcal{L})^{-1}$. We have:

$$\|\lambda R_\lambda x - x\| = \|\lambda R_\lambda x - R_\lambda(\lambda - \mathcal{L})x\| = \|R_\lambda \mathcal{L}x\| \leqslant \frac{1}{\lambda}\|\mathcal{L}x\| \to 0,$$

we used the M-dissipative condition $\|\lambda R_\lambda\| \leqslant 1$. Since $\mathcal{D}(\mathcal{L})$ is dense, this holds for all $x \in E$. □

Let us write $R_\lambda = (\lambda - \mathcal{L})^{-1}$, then

$$\mathcal{L}_\lambda \triangleq \lambda \mathcal{L} R_\lambda = \lambda(\lambda R_\lambda - id) = \lambda^2 R_\lambda - \lambda.$$

Definition 2.1.47. $\mathcal{L}_\lambda$ is said to be the Yosida approximation for $\mathcal{L}$.

Lemma 2.1.48. *Let $\mathcal{L}$ be a densely defined closed M-dissipative operator. Then $\mathcal{L}_\lambda$ is the generator of a uniformly continuous semigroup of contractions which we denote by T_t^λ. Furthermore,*

$$\|T_t^\lambda x - T_t^\mu x\| \leqslant t\|\mathcal{L}_\lambda x - \mathcal{L}_\mu x\|, \qquad \forall \lambda, \mu \geq 0.$$

Proof. Since $\|\mathcal{L}_\lambda\| \leqslant 2\lambda$, $\mathcal{L}_\lambda$ is a bounded operator and $T_t = e^{t\mathcal{L}_\lambda}$ is a uniformly continuous semi-group. Furthermore,

$$\|e^{t\mathcal{L}_\lambda}\| = \|e^{t(\lambda^2 R_\lambda - \lambda)}\| = e^{-\lambda t}\ e^{t\lambda^2 |R_\lambda|} \leqslant 1.$$

Also,

$$\begin{aligned}
\|e^{t\mathcal{L}_\lambda}x - e^{t\mathcal{L}_\mu}x\| &= \left\|\int_0^1 \frac{d}{ds} e^{st\mathcal{L}_\lambda + (1-s)t\mathcal{L}_\mu} x ds\right\| \\
&= \left\|\int_0^1 t(\mathcal{L}_\lambda - \mathcal{L}_\mu) e^{st\mathcal{L}_\lambda + (1-s)t\mathcal{L}_\mu} x\ ds\right\| \\
&\leqslant t\|\mathcal{L}_\lambda x - \mathcal{L}_\mu x\|.
\end{aligned}$$

□

Theorem 2.1.49 (Hille-Yosida theorem)**.** *A linear operator $\mathcal{L}$ on a Banach space E is the generator of a strongly continuous contraction semigroup on E if and only if the following statements hold.*

(i) $\mathcal{L}$ is closed and densely defined.

(ii) $\mathcal{L}$ is M-dissipative.

Proof. $\Longrightarrow$ The only if part follows from Proposition 2.1.29 and Theorem 2.1.28.

$\Longleftarrow$ Suppose that $\mathcal{L}$ is closed, densely defined, and M-dissipative. Then ro $x \in \mathcal{D}(\mathcal{L})$,

$$\begin{aligned}
\|\mathcal{L}_\lambda x - \mathcal{L}_\mu x\| = \|\lambda \mathcal{L} R_\lambda x - \mu \mathcal{L} R_\mu x\| &\leqslant \|\lambda \mathcal{L} R_\lambda x - \mathcal{L}x\| + \|\mu \mathcal{L} R_\mu x - \mathcal{L}x\| \\
&\leqslant (\frac{1}{\lambda} + \frac{1}{\mu})\|\mathcal{L}x\|.
\end{aligned}$$

By Lemma 2.1.48,

$$\|T_t^\lambda x - T_t^\mu x\| \leqslant t(\frac{1}{\lambda} + \frac{1}{\mu})\|\mathcal{L}x\|, \qquad \forall \lambda, \mu \geq 0.$$

So $T_t^\lambda x$ converges as $\lambda \to \infty$ uniformly in t on finite intervals. Set,

$$T_t x = \lim_{\lambda \to \infty} T_t^\lambda x.$$

Then $t \mapsto T_t x$ is continuous as uniform limit. Similarly, $\|T_t\| \leqslant 1$, and $T_0 x = x$. $T_t(T_s x) = \lim_{\lambda \to \infty} e^{t\mathcal{L}_\lambda}(T_s x)$. Since $e^{t\mathcal{L}_\lambda}$ is a contradiction, we can approximate $T_s f$ by $e^{s\mathcal{L}_\lambda}$, which gives $\lim_{\lambda \to \infty} e^{t\mathcal{L}_\lambda}(e^{s\mathcal{L}_\lambda}x) = T_{t+s}x$.

Finally let $\mathcal{A}$ denote its generator. Let $x \in \mathcal{D}(\mathcal{L})$. Then,

$$\begin{aligned}
\frac{1}{t}(T_t x - x) &= \frac{1}{t}\lim_{\lambda \to \infty}(T_t^\lambda x - x) = \frac{1}{t}\lim_{\lambda \to \infty}\int_0^t \frac{d}{ds} T_s^\lambda x\ ds \\
&= \frac{1}{t}\lim_{\lambda \to \infty}\int_0^t \mathcal{L}_\lambda T_s^\lambda x\ ds = \frac{1}{t}\lim_{\lambda \to \infty}\int_0^t T_s^\lambda \mathcal{L}_\lambda x\ ds = \frac{1}{t}\int_0^t T_s \mathcal{L} x ds \to \mathcal{L}x.
\end{aligned}$$

Hence $x \in \mathcal{D}(\mathcal{A})$, on which $\mathcal{L} = \mathcal{A}$. Note that $\mathcal{D}(\mathcal{L}) \subset \mathcal{D}(\mathcal{A})$.

By Theorem 2.1.29, any positive number $\lambda \in \rho(\mathcal{A})$, $(\lambda - \mathcal{A})$ is a bijection, and

$$(\lambda - \mathcal{A})(\mathcal{D}(\mathcal{A})) = E.$$

By the M-dissipative property, so is $(\lambda - \mathcal{L})^{-1}$, $(\lambda - \mathcal{L})(\mathcal{D}(\mathcal{L})) = E$. In particular, since $\mathcal{L} = \mathcal{A}$ on $\mathcal{D}(\mathcal{L}) \subset \mathcal{D}(\mathcal{A})$,

$$(\lambda - \mathcal{A})(\mathcal{D}(\mathcal{L})) = E.$$

As $\lambda - \mathcal{A}$ is injective, the two domains have to be the same. $\square$

Corollary 2.1.50. *A closed densely defined linear operator $\mathcal{L}$ on E is the generator of a strongly continuous contraction semigroup on E if and only if it is M-dissipative.*

Reference: [Paz83, EN00].

Corollary 2.1.51. *If T_t is a symmetric strongly continuous contraction semigroup on E, then there exists a self-adjoint operator $\mathcal{A}$ bounded from below s.t. $T_t = e^{-t\mathcal{A}}$.*

Proof. The generator $\mathcal{L}$ of T_t is closed and densely defined, and the resolvent $\rho(\mathcal{L}) \supset (0, \infty)$. Also,

$$\langle \mathcal{L}f, g \rangle = \left\langle \lim_{t \to 0} \frac{T_t f - f}{t}, g \right\rangle = \lim_{t \to 0} \left\langle f, \frac{T_t g - g}{t} \right\rangle,$$

hence $\mathcal{L}$ is symmetric. The spectrum of a closed positive symmetric operator are: the upper half complex plane, the lower half, the whole space, or a subset of $\mathbb{R}$. Hence $\sigma(\mathcal{L}) \subset [0, \infty)$ which means the range of $(\mathcal{L} \pm i)$ is E which implies that $\mathcal{L}$ is self-adjoint. That $\sigma(\mathcal{L}) \subset (-\infty, 0]$, which implies $\mathcal{L}$ is bounded from above. The two semi-groups, with the same generator must agree: $T_t = e^{t\mathcal{L}}$. $\square$

2.1.11

Example 2.1.52. Brownian motion with a drift: Consider $f : \mathbb{R} \to \mathbb{R}$ and let $\mathcal{L} = \frac{1}{2} f'' + c f'$ with domain:

$$\{f \in C_0(\mathbb{R}) : f' \in C_0(\mathbb{R}), f'' \in C_0(\mathbb{R})\}.$$

Example 2.1.53. (BM on $\mathbb{R}_+$, sticky boundary) $\mathcal{L} = \frac{1}{2} f''$ with domain.

$$\{f \in C_0(\mathbb{R}) : f' \in C_0(\mathbb{R}_+), f'' \in C_0(\mathbb{R}_+), f'(0) - \alpha f''(0) = 0\}.$$

Hene $\alpha > 0$ is a constant.

Chapter 3

Markov Semi-groups

Let us now state the formal definition of the class of semigroups on a subset of measurable functions we are going to study throughout these notes.

Definition 3.0.1. Let $(T_t)_{t\geqslant 0}$ be a semigroup on $\mathcal{B}_b(\mathcal{X})$. If $\|T_t\| \leqslant 1$ for all $t \geqslant 0$, $(T_t)_{t\geqslant 0}$ is called a *contraction* semigroup. If $T_t f \geqslant 0$ for all $f \geqslant 0$, we say that (T_t) is *positivity preserving*. If $T_t 1 = 1$, then it is called *conservative*. A conservative, positivity preserving contraction semigroup is called a Markov transition semigroup.

Definition 3.0.2. If (X_t) is a Markov process on $\mathcal{X}$ and T_t is a semigroup of bounded linear operators on a closed subspace $E \subset \mathcal{B}_b(\mathcal{X})$, where E is measure separable, s.t.

$$T_t f(X_s) = \mathbb{E}(f(X_{t+s})|\mathcal{F}_s), \qquad \forall f \in E,$$

we say that X_t corresponds to T_t.

Recall that the dual space E^* to a Banach space E is the set of continuous linear functions from E to $\mathbb{R}$. Then E^* is also a Banach space with the operator norm $\|\ell\| = \sup_{f\neq 0} \frac{|\ell(f)|}{\|f\|}$. We remark that a positive linear functional $\ell : \mathcal{C}(\mathcal{X}) \to \mathbb{R}$ is automatically bounded and $\ell(f) \leqslant \ell(\|f\|) = \|f\|\ell(1)$. The Riesz representation theorem states that if $\mathcal{C}(\mathcal{X})$ is a compact metric the dual space $\mathcal{C}(\mathcal{X})^*$ is the space of finite signed Borel measures on $\mathcal{X}$, with the total variation norm. (It is customary to define the bilinear map: $\langle \ell, f\rangle = \ell(f)$.) The tale of caution to a positive answer to the question is that the dual of $\mathcal{B}_b(\mathcal{X})$ are not necessarily a subset of measures. To obtain some sort of measurability on the transition probabilities, it would be helpful if the semigroup has continuity property. The continuity in time of $T_t f$ is automatic if f is continuous and T_t comes from a Markov process that is continuous in probability. For continuous f, the spatial continuity of $T_t f$ comes from the Feller property, otherwise from the strong Feller property.

3.0.1 Measure Separating sets

Please familiarise yourself with material in this section ahead of the lecture.

The semi-group of linear operators associated to a Markov process with a transition function is defined on the space of bounded (real-valued) measurable functions $\mathcal{B}_b(\mathcal{X})$ on $\mathcal{X}$. Before we proceed to study the abstract theory of a semigroup of linear operators on a Banach space E, we think of E as a space of functions, we explain that the function space needs to be sufficiently large to be useful to us. ‘Being large’ we mean that for example E should be rich enough to determine probability measures. Naturally, especially since $\mathcal{B}_b(\mathcal{X})$ is not separable, it is convenient to use a smaller set.

We should first formulate conditions which ensure that the semigroup (T_t) determines the Markov process. For this we need the following terminology:

Definition 3.0.3. A family of probability measures μ_n on $\mathcal{X}$ is said to converge weakly if for all bounded continuous functions f,

$$\int_{\mathcal{X}} f d\mu_n \to \int_{\mathcal{X}} f d\mu.$$

In the sequel, $\Rightarrow$ indicates weak convergence.

See [Par67, Theorem 6.1, pp40] for equivalent statements.

Definition 3.0.4. A collection E of continuous functions is said to be *separating* (or *measure determining*) if, for any two probability Borel measures μ, ν on $\mathcal{X}$,

$$\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f d\nu \quad \forall f \in E \quad \Rightarrow \quad \mu = \nu.$$

Recall that if $\mathcal{X}$ is in addition locally compact, $\mathcal{C}_0(\mathcal{X})$, the collection of real valued functions on $\mathcal{X}$ vanishing at infinity, is separable. Let K_n be a collection of compact sets with $\cup_n K_n = \mathcal{X}$, this property is called countable at infinity, and let E_n be a countable dense set of continuous functions with compact support on K_n. Then $E = \cup_n E_n$ is a dense subset of $C_0(\mathcal{X})$.

Example 3.0.5. If $\mathcal{X}$ is a locally compact separable metric space then $\mathcal{C}_0(\mathcal{X})$ is measure separating.

Remark 3.0.6. It is not necessary to assume elements of E are bounded, e.g. we may want to use polynomial functions. For we it is convenient to have property that if E is a measure determining set, the set of products of functions from E is measure separating on all finite product spaces. This depends on the tail of the probability distribution of the stochastic processes under the consideration.

Theorem 3.0.7 ([Par67, Thm. 5.9, pp39])**.** *Let $\mathcal{X}$ be a complete metric space and μ, ν be two probability measures on $\mathcal{X}$. Then $\mu = \nu$, if*

$$\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f d\nu$$

for every bounded and uniformly continuous function $f : \mathcal{X} \to \mathbb{R}$.

This theorem depends Urysohn's lemma and the fact that every probability measure on a metric space is regular and hence it is sufficient to show μ and ν agree on closed sets.

Let us now turn to product spaces, this will be useful for studying the finite dimensional distributions of a stochastic process. Let $\mathcal{X}_i$, be metric spaces, then the product space $\Pi_{i=1}^{\infty} \mathcal{X}_i$ is metrisable. The product space inherits the completeness and separability properties.

Proposition 3.0.8 ([EK86, Thm 4.6, pp115])**.** *Let $\mathcal{X}_i$ be complete separable metric spaces, and $E_k \subset \mathrm{BC}(\mathcal{X}_k)$ is measure separating on $\mathcal{X}_k$. Then*

$$L = \{\Pi_{i=1}^{n} f_i : f_i \in E_k \cup \{1\}, n \geq 1\}$$

is separating on $\Pi_{i=1}^{\infty} X_i$.

3.0.2 Markov transition / Probabilistic Semi-groups

Proposition 3.0.9 ([EK86, pp161]). *Let $\mathcal{X}$ be a separable metric space and let $E \subset \mathcal{B}_b(\mathcal{X})$ be closed, linear space which is measure determining. Let (X_t) be a Markov process on $\mathcal{X}$ with initial distribution μ and (T_t) be a semigroup of linear operators on E such that*

$$\mathbb{E}[f(X_{t+s}) \,|\, \mathcal{F}_s] = T_t f(X_s), \qquad \forall f \in E.$$

Then T_t and μ uniquely determine the finite dimensional distributions of (X_t).

Proof. Let $t > 0$. Since for every $f \in E$,

$$\mathbb{E}\left[f(X_t)\right] = \mathbb{E}\left[T_t f(X_0)\right] = \int_{\mathcal{X}} T_t f(x)\mu(dx) = \int_{\mathcal{X}} f(x)(T_t)^*\mu(dx),$$

and E is measure determining, the distribution of X_t equals $(T_t)^*\mu$. For the multi-dimensional distributions we use that

$$L = \{f(x) = \Pi_{i=1}^n f_i(x_i) : f_i \in E \cup \{1\}, n \geq 1\}$$

is measure separating on $\Pi_{i=1}^n \mathcal{X}$, see Proposition 3.0.8. We claim for any $n \geq 1$, $f_1, \dots, f_n \in E$, and $0 \leqslant t_1 < \cdots < t_n$,

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_{t_i})\right] = T_{t_1}\Big(f_1 \times \cdots \times T_{t_n - t_{n-1}} f_n)(X_{t_1})\Big).$$

For $n = 2$, this is

$$\mathbb{E}(f_1(X_{t_1}) f_2(X_{t_2})) = \mathbb{E}[(f_1 T_{t_2 - t_1} f_2)(X_{t_1})],$$

so the two point motion is determined. Assume this holds for $k \leqslant n-1$, we prove by induction and by the Markov property. For $t_1 < t_2 < \cdots < t_n$,

$$\begin{aligned}
\mathbb{E}[\Pi_{i=1}^n f_i(X_{t_i})] &= \mathbb{E}\big(f_1(X_{t_1})\mathbb{E}(\Pi_{i=2}^n f_i(X_{t_i})|\mathcal{F}_{t_1})\big) \\
&= \mathbb{E}\left(f_1(X_{t_1})\left(T_{t_2-t_1}\left(f_2 \times \cdots \times T_{t_n - t_{n-1}} f_n\right)(X_{t_1})\right)\right),
\end{aligned}$$

this concludes the proof. $\square$

3.1 The dual space of $\mathcal{C}_0(\mathcal{X})$

Let us now return to make connections with Markov processes (on a locally compact space). The reason that we can even hope to construct a Markov transition function from a semigroup of linear operators in the first place is Riesz's representation theorem which we recall below.

Definition 3.1.1. Let E be a vector space of functions with values in K (where $K = \mathbb{R}$ or $\mathcal{C}$). A linear functional ℓ on E is a linear map $\ell : E \to K$. A positive linear functional $\ell : E \to \mathbb{R}$ is a linear functional such that $\ell(f) \geq 0$ whenever $f \in E$ is a function with $f \geq 0$ pointwise.

Let E be a normed vector space, its *dual space* is the set of all bounded linear functionals on E and is denoted by E'. The dual space E' of a normed vector space with the operator norm is always a Banach space. The dual space contains linear functionals of the form $\ell(x_0) = \|x_0\|$ and $\|\ell\| = 1$ (use Hahn-Banach Theorem). Then $\|x\| = \sup\{\frac{|\ell(x)|}{\|\ell\|} : \ell \in E^*, \ell \neq 0\}$. The dual E' is large enough to separate points in E (for any $x \neq y$ in E, there exists $\ell \in E'$ with $\ell(x) \neq \ell(y)$).

Definition 3.1.2. (i) A sequence x_n in a normed space E is said to convergent (strongly convergent) if $\|x_n - x\| \to 0$ for some $x \in E$.

(ii) A sequence x_n in a normed space E is said to weakly convergent if there exists $x \in E$ such that $\ell(x_n) \to \ell(x)$ for every $\ell \in E'$.

Given a function space E, it is interesting to know what is its dual space. A desirable property for functions space E is that E' consists of measures on E. In case E' consists of measures then the weak convergence of $f_n \in E$ to f in E means:

$$\int_E f_n d\mu \to \int_E f d\mu,$$

for every $\mu \in E'$. This is a very useful concept. If $f_n \to f$ then $\|f_n\|$ is in fact bounded. Indeed, for every $\mu \in E'$, the convergent real sequence $\mu(f_n) \triangleq \int_E f_n d\mu$ is bounded. From this the boundedness of the norm follows from the uniform boundedness principle. A measure is said to have finite total variation if $|\mu|(E) = \sup_{j=1}^{\infty} \sum |\nu(E_j)|$ where $E = \cup_j E_j$ is a partition of E.

Theorem 3.1.3 (Riesz-Markov). Let $\mathcal{X}$ be a locally compact metric space. Then the dual of the $\mathcal{C}_0(\mathcal{X})$ is the space of signed Borel measures on $\mathcal{X}$ with finite total variation. In particular, if $\ell : \mathcal{C}_0(\mathcal{X}) \to \mathbb{R}$ is a positive linear functional, then there exists a unique Borel measure μ on $\mathcal{X}$ with finite total variation such that

$$\ell(f) = \int_{\mathcal{X}} f d\mu \qquad \forall f \in \mathcal{C}_0(\mathcal{X}).$$

This is originally obtained for $\mathcal{X}$ compact, the measure is constructed by:

$$\begin{aligned} \rho(O) &= \sup\{\ell(f) : f \in \mathcal{C}(X), 0 \leqslant f \leqslant 1, \operatorname{supp}(f) \subset O\}, \\ \mu_*(E) &= \inf\{\rho(O) : E \subset O, O \text{ is open}\}. \end{aligned}$$

See e.g. [SS05] for a proof in the compact case.

Note. References for this section are: [DS88, RS72, SS05]

3.2 The C_0-property

We are specially interested in a Markov process with a Markov transition function P, in which case

$$T_t f(x) = \int_{\mathcal{X}} f(x)\, P_t(x, dy) = \mathbb{E}_x[f(X_t)],$$

defines a *Markov transition semigroup* on $\mathcal{B}_b(\mathcal{X})$. There is, *a priori*, no regularity of the mapping $t \mapsto T_t$ (strong continuity). It turns out that most Markov transition semigroups are not strongly continuous on $\mathcal{B}_b(\mathcal{X})$. It is however this regularity which allows us to encode the semigroup in terms of a generator by means of the Hille-Yosida theorem. This can be remedied by restricting the semigroup to a smaller space. We therefore define

$$\mathcal{E} \triangleq \{f \in \mathcal{B}_b(\mathcal{X}) : \lim_{t \to 0} \|T_t f - f\|_\infty = 0\}.$$

This is the maximal subspace on which $(T_t)_{t \geqslant 0}$ is strongly continuous. An $\varepsilon/3$-argument shows that $\mathcal{E}$ is a Banach space and, clearly, $T_t(\mathcal{E}) \subset \mathcal{E}$.

One way to ensure the existence of the transition function is to exploit Corollary 3.2.2 and to develop a theory for Markov semigroups (T_t) which leave $\mathcal{C}_0(\mathcal{X})$ invariant. This leads to the so-called Feller-Dynkin processes. Another possible resolution of the dilemma is via L^p-space, provided we have a guess for the invariant measure and work on an L^2 space, see Section 3.6 below.

We state the following theorem without proof, which can be proved similarly to the proof that a super-martingale has a cádlág version. The interested reader may refer to [RY99, Thm 2.7, pp91], [LG16], [RW00, Section III.7].

Theorem 3.2.1. *If (X_t) is a Markov process with transition semigroup (T_t), which is strongly continuous on $\mathcal{C}_0(\mathcal{X})$, then there exists a càdlàg modification of (X_t), which is a $(\mathcal{F}_t^+)$-Markov process with the same transition semigroup.*

Corollary 3.2.2. *If $\mathcal{X}$ is locally compact and $T_t : \mathcal{C}_0(\mathcal{X}) \to \mathcal{C}_0(\mathcal{X})$, $t \geqslant 0$, is a positive preserving contraction semigroup and also defined on 1 with $T_t 1 = 1$, then there exists a transition function $P_t(x, dy)$ on $\mathcal{X}$ such that*

$$T_t f(x) = \int_{\mathcal{X}} f(y) P_t(x, dy) \qquad \forall f \in \mathcal{C}_0(\mathcal{X}). \tag{3.2.1}$$

Furthermore for any $A \in \mathcal{B}(\mathcal{X})$, $x \mapsto P_t(x, A)$ is measurable.

Proof. Then for each $x \in \mathcal{X}$ and $t > 0$, we have a probability measure $P_t(x, dy)$, which is dual to the bounded positive linear map $f \in \mathcal{C}(\mathcal{X}) \mapsto T_t f(x) \in \mathbb{R}$ we define a linear functional by $f \mapsto T_t f$. Note that $|T_t f(x)|_\infty \leqslant |f|_\infty$. The measurability of $x \mapsto P_t(x, A)$ for any $A \in \mathcal{B}(\mathcal{X})$ follows by a simple monotone class argument. By Theorem 3.2.1 the Markov process has a cádág version, hence $\mapsto P_t(x, A)$ is measurable and has at most a countable number of jumps. The joint measurability of $(t, x) \mapsto P_t(x, A)$ follows. $\square$

Exercise 3.2.3. Write down a Markov process for which $\mathcal{E} \neq \mathcal{B}_b(\mathcal{X})$.

We say that X defines a *Feller process* if $T_t\big(\mathrm{BC}(\mathcal{X})\big) \subset \mathrm{BC}(\mathcal{X})$ for all $t \geqslant 0$.

Exercise 3.2.4. Show that X is Feller if and only if, for all $t \geqslant 0$, $x \mapsto P_t(x, \cdot)$ is continuous as a map $\mathcal{X} \to \mathcal{P}(\mathcal{X})$ if the latter is equipped with the topology of weak convergence.

The terminology is not uniform across different textbooks. Sometimes authors call X Feller if $\mathcal{X}$ is locally compact and $T_t\big(\mathcal{C}_0(\mathcal{X})\big) \subset \mathcal{C}_0(\mathcal{X})$ where

$$\mathcal{C}_0(\mathcal{X}) \triangleq \{f \in \mathcal{C}(\mathcal{X}) : \forall \varepsilon > 0\, \exists K \subset \mathcal{X} \text{ compact} : |f(x)| \leqslant \varepsilon\, \forall x \in \mathcal{X} \setminus K\}.$$

For distinction we speak in this latter case of a *Feller-Dynkin process*. It is clear that this approach is problematic for infinite-dimensional $\mathcal{X}$. In fact, let $\mathcal{X}$ be an infinite-dimensional normed space, then $\mathcal{C}_0(\mathcal{X}) = \{0\}$. Nonetheless, we have the following result, see e.g. [RY99, Prop 2.4, pp89]:

Lemma 3.2.5. Let (T_t) be the Markov transition semigroup of a right continuous Markov process with $T_t\big(\mathcal{C}_0(\mathcal{X})\big) \subset \mathcal{C}_0(\mathcal{X})$. Then (T_t) is strongly continuous on $\mathcal{C}_0(\mathcal{X})$.

(It is sufficient to replace the right continuity of X_t by $\lim_{t \downarrow 0} P_t f(x) \to f(x)$ for any c and any $f \in \mathcal{C}_0(\mathcal{X})$.)

Proof of Lemma 3.2.5. For $\alpha > 0$, let $R_\alpha f \triangleq \int_0^t e^{-\alpha s} T_s g(x)\, ds$. Let $f = R_\alpha g$ for some $g \in \mathcal{C}_0(\mathcal{X})$. Then

$$T_t f(x) = e^{\alpha t} \int_t^\infty e^{-\alpha s} T_s g(x)\, ds = e^{\alpha t} f(x) - e^{\alpha t} \int_0^t e^{-\alpha s} T_s g(x)\, ds \quad \forall x \in \mathcal{X},$$

whence

$$\|T_t f - f\|_\infty \leqslant \big(e^{\alpha t} - 1\big)\|f\|_\infty + e^{\alpha t} \int_0^t \|T_s g\|_\infty\, ds \to 0$$

as $t \to 0$. Consequently, (T_t) is strongly continuous on $R_\alpha\big(\mathcal{C}_0(\mathcal{X})\big)$.

We then show that $R_\alpha(\mathcal{C}_0(\mathcal{X}))$ is dense in $\mathcal{C}_0(\mathcal{X})$. If not, since $\mathcal{C}_0(\mathcal{X})^*$ separate points and as a consequence of the Hahn-Banach and Riesz-Markov theorems, there is a finite, non-zero (signed) measure μ on $\mathcal{X}$ such that

$$\int_{\mathcal{X}} R_\alpha g\, d\mu = 0 \quad \forall g \in \mathcal{C}_0(\mathcal{X}).$$

It follows by the *(first) resolvent identity*

$$R_\beta = R_\alpha - (\beta - \alpha) R_\alpha R_\beta, \qquad \forall \alpha, \beta > 0, \tag{3.2.2}$$

we have

$$\int_{\mathcal{X}} R_\beta g\, d\mu = 0 \quad \forall g \in \mathcal{C}_0(\mathcal{X}), \beta > 0.$$

But this contradicts the fact that, since $T_t g(x) = \mathbb{E}_x[g(X_t)] \to g(x)$ by right-continuity of X_t, $\beta R_\beta g(x) \to g(x)$ for any $x \in \mathcal{X}$ as $\beta \to \infty$. In fact, then by dominated convergence

$$0 = \lim_{\beta \to \infty} \beta \int_{\mathcal{X}} R_\beta g\, d\mu = \int_{\mathcal{X}} g\, d\mu, \quad \forall g \in \mathcal{C}_0(\mathcal{X}),$$

i.e., $\mu \equiv 0$, contracting the assumption that $\mathbb{R}_\alpha(\mathcal{C}_0(\mathcal{X}))$ is not dense. □

3.3 Strong Markov Property

For some purposes the natural filtration of a Markov process may be too small, e.g., the hitting times of open sets by Brownian motion are no stopping times with respect to the natural filtration. For a given filtration $(\mathcal{F}_t)$, we let $\mathcal{F}_t^+ \triangleq \bigcap_{r>t} \mathcal{F}_r$ denote its right-continuous version.

Proposition 3.3.1. *Let (X_t) be a Markov process with right-continuous sample paths. If its transition semigroup (T_t) leaves $\mathrm{BC}(\mathcal{X})$ or $\mathcal{C}_0(\mathcal{X})$-invariant, then (X_t) is an $(\mathcal{F}_t^+)$-Markov process.*

Proof. Let $0 \leqslant s < t$ and $\varepsilon > 0$. For $f \in \mathrm{BC}(\mathcal{X})$, we have that

$$\mathbb{E}\left[f(X_{t+s+\varepsilon}) \,|\, \mathcal{F}_s^+\right] = \mathbb{E}\left[\mathbb{E}\left[f(X_{t+s+\varepsilon}) \,|\, \mathcal{F}_{s+\varepsilon}\right] \,|\, \mathcal{F}_s^+\right] = \mathbb{E}\left[T_t f(X_{s+\varepsilon}) \,|\, \mathcal{F}_s^+\right].$$

By right-continuity and bounded convergence, we can take $\varepsilon \to 0$ to conclude

$$\mathbb{E}\left[f(X_{t+s}) \,|\, \mathcal{F}_s^+\right] = \mathbb{E}\left[T_t f(X_s) \,|\, \mathcal{F}_s^+\right] = T_t f(X_s).$$

for bounded continuous test functions $f : \mathcal{X} \to \mathbb{R}$. To see that this in fact holds for any bounded measurable f, we fix $A \in \mathcal{F}_s^+$ and define the measures

$$\mu_A(B) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_B(X_{t+s}) \,|\, \mathcal{F}_s^+\right] \mathbf{1}_A\right], \quad \nu_A(B) = \mathbb{E}\left[T_t \mathbf{1}_B(X_s) \mathbf{1}_A\right].$$

Both have the same total finite mass, and

$$\int_{\mathcal{X}} f\, d\mu = \int_{\mathcal{X}} f\, d\nu \quad \forall f \in C_0(\mathcal{X}).$$

Since $\mathcal{C}_0(\mathcal{X})$ is measure-determining class, $\mu_A = \nu_A$, as required. $\square$

Let τ be a stopping time and recall that

$$\mathcal{F}_\tau \triangleq \{A \in \mathcal{F} :\ A \cap \{\tau \leqslant t\} \in \mathcal{F}_t \, \forall t \geqslant 0\}$$

defines a σ-field. The following two lemmas are standard:

Lemma 3.3.2. *Let*

$$\tau_n \triangleq \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{\left\{\frac{k}{2^n} \leqslant \tau < \frac{k+1}{2^n}\right\}} + \infty \mathbf{1}_{\{\tau=\infty\}}, \quad n \in \mathbb{N}.$$

Then τ_n is a stopping time for each $n \in \mathbb{N}$ and $\tau_n \downarrow \tau$ a.s.

With this one can show that

Lemma 3.3.3. *If (X_t) is adapted and right-continuous, then $X_\tau \mathbf{1}_{\tau<\infty} \in \mathcal{F}_\tau$.*

The next theorem shows that Feller processes are strong Markov:

Theorem 3.3.4. *Let (X_t) be a right-continuous Markov process whose transition function leaves either $\mathcal{C}_0(\mathcal{X})$ or $\mathrm{BC}(\mathcal{X})$ invariant. Then it is strong Markov. If (X_t) is cádlág (respectively continuous) , the Markov process in the canonical picture is:*

$$\mathbb{E}\left[\Phi \circ \theta_\tau \mathbf{1}_{\{\tau<\infty\}} \,|\, \mathcal{F}_\tau\right] = \mathbf{1}_{\{\tau<\infty\}} \mathbb{E}_{X_\tau}[\Phi], \tag{3.3.1}$$

where Φ is a bounded measurable function on $D\big([0,1], \mathcal{X}\big)$ (on the Wiener space).

Proof. Let us first suppose that τ takes only a countable number of values $\{t_k :\ k \in \mathbb{N}\}$ with $0 \leqslant t_1 < t_2 < \cdots < \ldots \leqslant \infty$. Then, using Theorem 1.8.5, we get for each $B \in \mathcal{F}_\tau$,

$$\begin{aligned}
\mathbb{E}\left[\Phi \circ \theta_\tau \mathbf{1}_{\{\tau<\infty\}} \mathbf{1}_B\right] &= \sum_{k=1}^{\infty} \mathbb{E}\left[\Phi \circ \theta_{t_k}) \mathbf{1}_{\{\tau=t_k\}} \mathbf{1}_B\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\mathbb{E}\left[\Phi \circ \theta_{t_k} \,|\, \mathcal{F}_{t_k}\right] \mathbf{1}_{\{\tau=t_k\}} \mathbf{1}_B\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\mathbb{E}_{X_{t_k}}[\Phi] \mathbf{1}_{\{\tau=t_k\}} \mathbf{1}_B\right] = \mathbb{E}\left[\mathbb{E}_{X_\tau}[\Phi] \mathbf{1}_{\{\tau<\infty\}} \mathbf{1}_B\right].
\end{aligned}$$

Here we used the fact that $B \cap \{\tau = t\} \in \mathcal{F}_t$ for each $B \in \mathcal{F}_\tau$ and $t \geqslant 0$.

If $f \in \mathcal{B}_b$ and $\Phi(X) = f(X_t)$, this is:

$$\mathbb{E}\left[f(X_{t+\tau}) \mathbf{1}_{\{\tau<\infty\}} | \mathcal{F}_\tau\right] = T_t f(X_\tau) \mathbf{1}_{\{\tau<\infty\}}. \tag{3.3.2}$$

Now assume a general τ, for the approximating sequence of Lemma 3.3.2,

$$\mathbb{E}\left[f(X_{t+\tau_n}) \mathbf{1}_{\{\tau_n<\infty\}} | \mathcal{F}_\tau\right] = T_t f(X_{\tau_n}) \mathbf{1}_{\{\tau_n<\infty\}}.$$

By the right-continuity of X and the Feller property of T_t, for any $f \in BC$ (or $f \in \mathcal{C}_0(\mathcal{X})$), (3.3.2) holds by bounded convergence, for any f continuous and bounded. By the standard method, this holds for bounded measurable f. It then remains to prove this for functions of the form $\Pi_{i=1}^n f_k(x_{t_k})$ and thus for all bounded measurable functions. For continuous paths, the analogous conclusion obviously holds. $\square$

The strong Markov property states that the process restarts at any stopping afresh.

Example 3.3.5. Let us return to Example 1.4.2, consider the transition function

$$Q_t(x, dy) = \begin{cases} P_t(x, dy), & \text{if } x \neq 0, \\ \delta_0(dy), & \text{if } x = 0, \end{cases}$$

where $P_t(x, dy) = p_t(x, y)$ where $p_t(x, y)$ is the heat/Gaussian kernel. If $x \neq 0$, we have a Brownian motion, e.g. $P(X_t \in A) = \int_A p_t(x, dy)$ for any $t > 0$. But when it hits zero (it does in finite time), it gets stuck at 0: from this stopping time, this is no longer a Brownian motion. However, the Markov property would require that $x_{t+\tau}$ to behave as a Brownian motion starting from 0. More precisely, let $\tau = \inf_{t>0}\{x_t = 0\}$, then $x_{\tau+t} = 0$ for all t.

Let us take a look from the definition of the strong Markov property. A realisation of the Markov process from x is:

$$X_t \triangleq \begin{cases} x + W_t, & \text{if } X_0 = x \neq 0, \\ 0, & \text{if } X_0 = 0, \end{cases}$$

for a one-dimensional Brownian motion $(W_t)_{t\geqslant 0}$. Take $\Phi(\sigma) = (\sigma(1))^2$. Suppose that $X(0) = 0$, then $\mathbb{E}_{X_\tau}(X(1))^2 = 0$, as $X(t) = 0$ for all time t when $X(0) = 0$. On the other hand,

$$\mathbb{E}((X_{1+\tau})^2 | \mathcal{F}_\tau) = \mathbb{E}((x + W_{1+\tau})^2 | \mathcal{F}_\tau) \neq 0.$$

This Markov process is not Feller!! Let f be a continuous and bounded function, then

$$P_t f(0) = f(0), \qquad P_t f(x) = \int_{\mathbb{R}} f(y) p_t(x, y)\, dy.$$

For $t > 0$, $\lim_{x\to 0} P_t f(x) \neq f(0)$ in general. Take for example $f(y) = y^2$.

3.4 Martingale Consideration

For the next result we require the measurability of the maps $r \to \mathcal{L}f(x_r)$ and $r \mapsto \mathcal{L}T_{s-r}f(X_r)$, which can be obtained be assuming X_r is progressively measurable and that $r \mapsto T_r f(x)$ is measurable which will follow if T_t is strongly continuous.

Proposition 3.4.1. *Let T_t be a strongly continuous semigroup on a Banach space $E \subset \mathcal{B}_b(\mathcal{X})$ with generator $\mathcal{L}$. Let (X_t) be a Cádlág Markov process corresponding to T_t. Then for every $f \in \mathcal{D}(\mathcal{L})$,*

$$M_t^f = f(X_t) - \int_0^t \mathcal{L}f(X_r) dr$$

is a martingale.

Proof. Let $s < t$. Since f and $\mathcal{L}f$ are bounded, for any $A \in \mathcal{F}_s$,

$$\mathbb{E}\left(\int_s^t \mathcal{L}f(X_r) dr \mathbf{1}_A\right) = \int_s^t \mathbb{E}\left(\mathcal{L}T_{r-s}f(X_s)\, \mathbf{1}_A\right) dr < \infty.$$

The cádlág property of X_r implies that $r \mapsto \mathcal{L}f(X_r)$ is measurable. It is then trivial to see that

$$
\begin{aligned}
\mathbb{E}(M_t^f - M_s^f | \mathcal{F}_s) &= \mathbb{E}\left[f(X_t) - f(X_s) \,\middle|\, \mathcal{F}_s\right] - \mathbb{E}\left[\int_s^t \mathcal{L}f(X_r)dr \,\middle|\, \mathcal{F}_s\right] \\
&= T_{t-s}f(X_s) - f(X_s) - \int_s^t T_{r-s}\mathcal{L}f(X_s)\, dr \\
&= T_{t-s}f(X_s) - f(X_s) - \int_s^r \mathcal{L}T_{r-s}f(X_s)\, dr = 0 \\
&= T_{t-s}f(X_s) - f(X_s) - \int_s^r \frac{d}{dr}T_{r-s}f(X_s)\, dr = 0,
\end{aligned}
$$

In the last step we used the fact that $\mathcal{L}T_tf = T_t\mathcal{L}f = \frac{d}{dt}T_tf$, for every $f \in \mathcal{D}(\mathcal{L})$. This completes the proof. $\square$

The converse holds if $E = \mathcal{C}_0(\mathcal{X})$ or $E = BC(\mathcal{X})$. Recall from Theorem 3.2.1 for the existence of a cádlag version of the Markov process.

Proposition 3.4.2. *Let T_t be a strongly continuous semigroup on $\mathcal{C}_0(\mathcal{X})$ with generator $\mathcal{L}$. Suppose that (X_t) is a Cádlág Markov process corresponding to T_t and with deterministic initial condition x. Suppose that $f, g \in \mathcal{C}_0(\mathcal{X})$ and*

$$N_t = f(X_t) - \int_0^t g(X_r)dr$$

is a martingale. Then $f \in \mathcal{D}(\mathcal{L})$ and $\mathcal{L}f = g$.

Proof. Again the regularity on X_t implies that the integral $\int_0^t g(X_r)dr$ is well defined. Since N_t is a martingale,

$$\mathbb{E}[f(X_t)] - \mathbb{E}\left[\int_0^t g(X_r)dr\right] = \mathbb{E}N_0 = f(X_0).$$

Since X_t is a Markov process corresponding to T_t with initial point x, $\mathbb{E}[f(X_t)] = T_tf(x)$. Hence

$$\frac{1}{t}\left[T_tf(x) - f(x)\right] = \frac{1}{t}\mathbb{E}\left[\int_0^t g(X_r)dr\right] = \frac{1}{t}\left[\int_0^t T_rg(x)dr\right],$$

Since T_rg is continuous, the right hand side converges to $g(x)$ and $\mathcal{L}f = g$. $\square$

Let $W_t = (W_t^1, \dots, W_t^m)$ be an m-dimensional Brownian motion. Let $X_k : \mathbb{R}^n \to \mathbb{R}^n$, $k = 1, \dots, m$, be vector fields on $\mathbb{R}^d$. We consider the stochastic differential equation

$$dx_t = \sum_{k=1}^m X_k(x_t)dW_t^k + X_0(x_t)dt. \tag{3.4.1}$$

Let $a_{i,j}(x) = \sum_{k=1}^m X_k^i(x)X_k^j(x)$. Set

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^n a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{j=1}^n X_0^j(x)\frac{\partial}{\partial x_j}. \tag{3.4.2}$$

Exercise 3.4.3. Show that if the vector fields X_i are Lipschitz continuous, then for any $f \in C_K^\infty$,

$$f(x_t) - f(x_0) - \int_0^t \mathcal{L}f(x_r)dr$$

is a martingale.

3.5 Martingale Problem

Definition 3.5.1. A right continuous $\mathcal{G}_t$-adapted process x_t is said to solve the martingale problem for a linear operator $\mathcal{L}$ on a subspace of $\mathcal{B}_b(\mathcal{X})$ with respect to the filtration $\mathcal{G}_t$ if for every $f \in \mathcal{D}(\mathcal{L})$,

$$M_t^f \triangleq f(x_t) - \int_0^t \mathcal{L}f(x_r)dr$$

is a $\mathcal{G}_t$ martingale.

The questions whether the martingale is well posed is a fundamental question, which leads essentially to the strong Markov property. The convention is to take $\mathcal{G}_t$ the natural filtration of X_t. It is also standard to assume that for $f \in C_K^\infty$, instead of $f \in \mathcal{D}(\mathcal{L})$, that M_t^f is a martingale.

Definition 3.5.2. A measure $\mathbb{P}_\mu$ on the canonical space Ω is said to solve the martingale problem for $\mathcal{L}$ with initial condition μ if for every $f \in \mathcal{D}(\mathcal{L})$ and for π_t the canonical process,

$$M_t^f \triangleq f(\pi_t) - \int_0^t \mathcal{L}f(\pi_r)dr$$

is a martingale with respect to the measure $\mathbb{P}_\mu$ and $(P_0)^*\mathbb{P}_\mu = \mu$ almost surely.

We do not have time to work with the martingale problem in great depth, will simply go over the important results for stochastic differential equations on $\mathbb{R}^n$ for which the following notion of local martingale problem is equivalent to the existence of a weak solution.

Definition 3.5.3. A continuous process x_t in $\mathbb{R}^n$ or its probability measure on the Wiener space is said to solve the local martingale problem for

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^n a_{i,j}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i \frac{\partial}{\partial x_i},$$

if for any $f \in C_K^\infty$,

$$M_t^f \triangleq f(x_t) - f(x_0) - \int_0^t \mathcal{L}f(x_r)dr$$

is a local martingale.

Let us consider the SDE (4.7.1) and its (formal) generator (3.4.2). Suppose that X_i are Borel measurable. Then the SDE has a weak solution with distribution P if and only if P is a solution to the local martingale problem for $\mathcal{L}$. Suppose that for every $x \in \mathbb{R}^n$, the local martingale problem for $\mathcal{L}$ with initial distribution δ_x has a unique solution $\mathbb{P}_x$, then strong Markov property holds for the family $\{\mathbb{P}^x\}$.

3.6 L^p-Semigroups and Invariant Measure

We introduce two examples of strongly continuous semi-groups on L^p.

Lemma 3.6.1. ▷ *(Minkowski's integral inequality) Let $f : \mathbb{R}^m \times \mathbb{R}^n$ be measurable. Then, for $1 \leqslant p < \infty$,*

$$\left(\int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^m} f(x,y)dy\right|^p dx\right)^{\frac{1}{p}} \leqslant \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^m}|f(x,y)|^p\,dx\right)^{\frac{1}{p}} dy.$$

In other words,

$$\left\| \int_{\mathbb{R}^m} f(\cdot, y) dy \right\|_p \leqslant \int_{\mathbb{R}^n} \|f(x,\cdot)\|_p \, dx.$$

▷ *(Young Inequality) Let $f, K : \mathbb{R}^n \to \mathbb{R}$ be measurable, $f \in L^p$ and $K \in L^1$. Then the convolution $f * K$ is in L^p for any $1 \leqslant p \leqslant \infty$:*

$$\|f * K\|_p \leqslant \|f\|_p \|K\|_1.$$

Example 3.6.2. For the heat semi-group, we already have a transition semi-group, we are only concerned with a space on which T_t is a strongly continuous semi-group. Indeed, on $L^p \cap L_\infty$, $\|P_t f\|_p \leqslant \|f\|_p$, by the Young inequality. The heat semigroup thus extends to a semi-group on L_p by the contraction property and the fact that C_K^∞ is dense in L^p.

To show the semi-group on L_p is strongly continuous, let f be smooth with compact support. For any $\varepsilon > 0$ choose $\delta > 0$ so $|f(x) - f(y)| < \varepsilon/2$ for $|x - y| < \delta$ and let $K_t(x) = P_t(0, x)$.

$$\left| \int_{\mathbb{R}^n} K_t(y)(f(x+y) - f(x)) dy \right|_\infty \leqslant \frac{\varepsilon}{2} + 2|f|_\infty \left| \int_{|y| \geq \delta} \frac{1}{\sqrt{2\pi t}^{n/2}} e^{-\frac{|y|^2}{2t}} dy \right|_\infty < \varepsilon$$

for t sufficiently small, $|P_t f(x) - f(x)| \to 0$ for such f. Since $|P_t f - f|$ is uniformly bounded in L^p for any p, then the convergence is in L^p. For $f \in L^p$, choose $f_n \to f$ in L^p and f_n smooth with compact supports, then

$$\|P_t f - f\|_p \leqslant \|P_t f - P_t f_n\|_p + \|P_t f_n - f_n\|_p + |f_n - f\|_p \to 0.$$

Definition 3.6.3. Let X be a Markov process on $\mathcal{X}$ with transition semi-group T_t on $\mathcal{B}_b(\mathcal{X})$. A measure π on $\mathcal{X}$ is called *invariant* for X if

$$\int_{\mathcal{X}} T_t f(x) \, \pi(dx) = \int_{\mathcal{X}} f(x) \, \pi(dx)$$

for all $t \geqslant 0$ and $f \in \mathcal{B}_b(\mathcal{X})$.

Lemma 3.6.4. *Let $\pi \in \mathcal{P}(\mathcal{X})$ be an invariant measure for a* right-continuous *sample paths Markov process X. Then (T_t) extends to a Markov transition semigroup on $L^p(\mathcal{X}, \pi)$ for any $p \geqslant 1$. Furthermore T_t is a positive preserving strongly continuous contraction on L^p.*

Proof. Let $f \in L^p(\mathcal{X}, \pi) \cap L_\infty$. Then $|T_t f|^p = \left| \int f(y) P_t(x, dy) \right|^p \leqslant T_t |f|^p$ by Jensen's inequality, whence

$$(\|T_t f\|_{L^p})^p = \int |T_t f|^p \pi(dx) \leqslant \int T_t(|f|^p) \pi(dx) = \|f\|_{L^p},$$

since π is invariant. The set of continuous compactly supported functions is dense in L^p, so T_t extends to a contraction semigroup on $L^p(\mathcal{X}, \pi)$. By the right-continuity of the process, $T_t f(x) \to f(x)$ as $t \to 0$ for any $f \in \mathrm{BC}(\mathcal{X}) \cap L_p$, $|T_t f - f|_{L^p} \to 0$ by the dominated convergence, and this holds for any $f \in L^p(\mathcal{X}, \pi)$ since $BC(\mathcal{X})$ is a dense subspace of $L^p(\mathcal{X}, \pi)$. The semigroup on L^p inherits the positive preserving property. □

3.7 Characterisation of Invariant Measures

Let $\mathcal{X}$ be a separable complete metric space. Let E be a closed subspace of $\mathcal{B}_b(\mathcal{X})$.

Suppose that $\mathcal{L}$ generates a strongly continuous contraction semi-group T_t on E and E is separating, then any solution X_t to the martingale problem for $\mathcal{L}$ with initial distribution μ is a Markov process for T_t and

$$\mathbb{E}[f(X_{t+s}) \mid \mathcal{F}_s] = T_t f(X_s) \tag{3.7.1}$$

for any $f \in E$. See Theorem 4.1 in [EK86, pp182]. Furthermore uniqueness holds for the martingale problem for $\mathcal{L}$ with the initial distribution μ.

The following theorem unifies several notions of invariant measures, see [EK86, pp239].

Theorem 3.7.1. *Suppose that $\mathcal{L}$ generates a strongly continuous contraction semi-group T_t on E and E is measure determining, and the martingale problem for $\mathcal{L}$ is well posed. Let X_t (right continuous) be the solution for the martingale problem for $\mathcal{L}$ with the initial condition μ. Then the following is equivalent for a probability measure π.*

(i) The distribution of X_t is μ for all time $t \geq 0$.

(ii) $\theta_t X$ and X have the same finite dimensional distributions.

(iii) $\int_{\mathcal{X}} T_t f \, d\pi = \int_{\mathcal{X}} f \, d\pi$, for every $f \in E, t \geq 0$.

(iv) $\int_{\mathcal{X}} \mathcal{L} f d\pi = 0$ for any $f \in \mathrm{Dom}(\mathcal{L})$.

Proof. ▷ (ii) obviously implies (i).

- ▷ (i) $\Longrightarrow$ (ii) If $\mathcal{L}(X_t) = \mu$ for some $t > 0$, then $\theta_t X$ is a solution of the martingale problem with the initial value μ also. By the uniqueness to the martingale problem, the process $\theta_t X$ and X have the same probability distributions.

- ▷ (ii) $\Longrightarrow$ (iii) Let $f \in E$, according to (3.7.1),

$$\int T_t f(x) \mu(dx) = \mathbb{E}[f(X_t)] = \mathbb{E} f(\theta_s X_t) = \int T_{t+s} f(x) \mu(dx).$$

- ▷ (iii) $\Longrightarrow$ (i) The above shows that $\mathbb{E}[f(X_t)] = \mathbb{E} f(X_{t+s}))$ for any $f \in E$. Since E is measure determining, $\mathcal{L}(X_t) = \mathcal{L}(X_{t+s})$.

- ▷ (iii) $\Longrightarrow$ (iv) is immediate from the definition of the generator.

- ▷ (iv) $\Longrightarrow$ (iii), for $f \in \mathcal{D}(\mathcal{L})$,

$$\int_{\mathcal{X}} (T_t f - f) d\mu = \int_{\mathcal{X}} \int_0^t \frac{\partial}{\partial s} T_s f \, ds \, d\mu = \int_{\mathcal{X}} \int_0^t \mathcal{L} T_s f \, ds \, d\mu,$$

the right hand side equals

$$\int_{\mathcal{X}} \mathcal{L} \left(\int_0^t T_s f \, ds \right) d\mu = 0.$$

By density of $\mathcal{D}(\mathcal{L})$ in E, $\int_{\mathcal{X}} (T_t f - f) d\mu = 0$ for every $f \in E$ and every $t > 0$.

This completes the proof. □

Chapter 4

Diffusion processes and diffusion operators

4.1 Diffusion operators

Let $x = \mathbb{R}^n$ and

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^n a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_l^n b_l(x)\frac{\partial}{\partial x_l}.$$

where, for any x, $(a_{i,j}(x))$ is a non-negative symmetric matrix. The operator $\mathcal{L}$ is elliptic if for any x and any $\xi \in \mathbb{R}^n$,

$$\sum_{i,j=1}^n a_{i,j}(x)\xi_i\xi_j > 0.$$

It is strictly elliptic if there exists $c > 0$ for any x and any $\xi \in \mathbb{R}^n$, such that

$$\sum_{i,j=1}^n a_{i,j}(x)\xi_i\xi_j > c|\xi|^2.$$

For some authors, strictly ellipticity includes also an upper bound. Observe that π being an invariant measure is equivalent to $\mathcal{L}^*\pi = 0$ in the distributional sense. If $\pi \ll dx$, then $\pi = gdx$ and $\int_{\mathbb{R}^n} \mathcal{L} f g dx = 0$ for some Borel measurable function g for $f \in \mathrm{Dom}(\mathcal{L})$. It is natural to work with $L^2(dx)$, in terms of the L^2 adjoint operator

$$\int f\mathcal{L}^* g dx = 0.$$

For elliptic operators, π has a (smooth) density with respect to dx. An operator with smooth coefficients and satisfying Hörmander's bracket conditions has a smooth density.

Note that

$$\mathcal{L}^* g = \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}(a_{i,j}g) - \sum_l \frac{\partial}{\partial x_l}(b_l g)$$

is the sum of a diffusion operator and a zero order term Vg where

$$V = \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 a_{i,j}}{\partial x_i \partial x_j} - \sum_{l=1}^n \frac{\partial b_l}{\partial x_l}.$$

Example 4.1.1. The Brownian motion on $\mathbb{R}^n$ has no finite invariant probability measure. Its only invariant measure is dx. It has no non-constant harmonic functions.

Example 4.1.2. The Ornstein-Uhlenbeck process has a unique invariant probability measure.

We write $L_b\, g = \sum_l \frac{\partial}{\partial x_l}(b_l g)$, the Lie derivative of g in the direction of b.

Exercise 4.1.3. Let $b : \mathbb{R}^n \to \mathbb{R}^n$ and $V : \mathbb{R}^n \to \mathbb{R}$ are smooth.

$$\mathcal{L} = \frac{1}{2}\Delta + L_b + L_{\nabla V}.$$

Suppose that $\operatorname{div}\big(b\, e^{-2V}\big) = 0$. Show that $e^{-2V}dx$ is an invariant measure.

Definition 4.1.4. A diffusion process is a continuous strong Markov process.

4.1.1 Stochastic processes defined up to a random time

The stochastic process $X_t(\omega) := \frac{1}{2-B_t(\omega)}$ is defined up to the first time $B_t(\omega)$ reaches 2. We denote this time by τ. For any given time t, no matter how small it is, there are a set of path of positive probability (measured with respect to the Wiener measure on $C([0,t];\mathbb{R}^d)$) which will have reached 2 by time t:

$$P(\tau \leqslant t) = \mathbb{P}(\sup_{s\leqslant t} B_s \geq 2) = 2\mathbb{P}(B_t \geq 2) = \sqrt{\frac{2}{\pi}} \int_{\frac{2}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} dy.$$

This probability converges to zero as $t \to 0$. We say that X_t is defined up to τ and τ is called its life time or explosion time.

Let $\mathbb{R}^d \cup \{\Delta\}$ be the one point compactification of $\mathbb{R}^d$, which is a topological space whose open sets are open sets of $\mathbb{R}^d$ plus set of the form $(\mathbb{R}^d \setminus K) \cup \{\Delta\}$ where K denotes a compact set. The one point compactification allows for the explosion of the solutions Given a process $(x_t, t < \tau)$ on $\mathbb{R}^d$ we define a process $(\hat{x}_t, t \geq 0)$ on $\mathbb{R}^d \cup \{\Delta\}$:

$$\hat{x}_t(\omega) = \left\{ \begin{array}{ll} x_t(\omega), & \text{if } t < \tau(\omega) \\ \Delta, & \text{if } t \geq \tau(\omega). \end{array} \right\}.$$

If X_t is a continuous process on $\mathbb{R}^d$ then $\hat{X}_t$ is a continuous process on $\mathbb{R}^d \cup \Delta$. Define

$\hat{W}(\mathbb{R}^d) \equiv \{Y \in C([0,\infty); \mathbb{R}^d \cup \Delta)$ with the property $Y_t(\omega) = \Delta$ if $Y_s = \Delta$ for some $s \leqslant t\}$

The last condition means that once a process enters the coffin state it does not return.

4.2 Stochastic Flows

Consider an Itô SDE of Markovian type on $\mathbb{R}^n$:

$$dx_t = \sum_{k=1}^{m} X_k(x_t)\, dB_t^k + \sigma_0(x_t)\, dt, \tag{4.2.1}$$

where $X_k : \mathbb{R}^n \to \mathbb{R}^n$, $j = 0, \dots, m$, and $\{B^k\}_{k=1}^m$ are independent one-dimensional Brownian motions. Solution to stochastic differential equations are Markov process, the Markov property follows from the well-posedness of solutions.

Definition 4.2.1. A solution to the SDE is a pair of adapted stochastic processes (X_t, B_t) on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ where (B_t) is a standard Brownian motion in $\mathbb{R}^m$ such that the integral equation

$$x_t = x_0 + \int_0^t X_k(x_s) dB_s + \int_0^t X_0(x_s) ds$$

holds almost surely.

More generally, we assume X_t takes its value in $\hat{W}(\mathbb{R}^d)$ then there is a well defined explosion time τ after which $x_t(\omega) = \Delta$. For the course we may assume non-explosion.

Definition 4.2.2. We say that *pathwise uniqueness* for the SDE (4.2.1) holds if for any two solutions X and $\tilde{X}$ on the same filtered probability space and driven by the same Brownian motion the following holds: If $x_0 = \tilde{x}_0$ a.s., then $x_t = \tilde{x}_t$ for every $t \geqslant 0$ a.s. We say Uniqueness in law holds if any two solutions with the same initial laws are the same in distribution.

Definition 4.2.3. A stochastic process (x_t) together with a Brownian motion (B_t) on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ is a strong solution if x_t is adapted to the filtration generated by the Brownian motion. Otherwise it is a weak solution.

A weak solution is tied with the well-posedness of the martingale problem.

Theorem 4.2.4. *If for every x there exists one and only one solution $\mathbb{P}_x$ to the martingale problem on the space C_K^∞ and if $\mathbb{P}_x(x_t \in A)$ is measurable then the canonical process x_t is a Markov process with transition function $\mathbb{P}_x(x_t \in A)$.*

Example 4.2.5 (Tanaka's example). Consider

$$dx_t = \text{sign}(x_t) dB_t$$

where $sign(x) = -1$ if $x \leqslant 0$ and 1 otherwise. The solution $x_t = x_0 + \int_0^t \text{sign}(x_s) dB_s$ is a martingale with quadratic variation t. By Lévy characterisation theorem, the distribution of x_t is $N(x, t)$. Uniqueness in law holds. On any probability space, if x_t is a solution so is $-x_t$. Hence pathwise uniqueness fails.

To construct a weak solution with initial value x, take any probability space and any Brownian motion B_t. Define

$$W_t = \int_0^t \text{sign}(B_s) dB_s.$$

Then

$$\int_0^t \text{sign}(B_s) dW_s = \int_0^t dB_s = B_t - B_0.$$

This means that the pair (B_t, W_t) is a solution to the SDE, with W_t the driving noise,

$$dx_t = \text{sign}(x_t) dW_t.$$

By construction the driving Brownian motion W_t is adapted to the filtration of the solution B_t and has in fact strictly smaller σ-algebra.

Proposition 4.2.6 (Tanaka's formula). *Let x_t be a continuous semi-martingale. There exists a continuous increasing process l_t such that*

$$|x_t| = |x_0| + \int_0^t \operatorname{sign}(x_s) dx_s + l_t.$$

The process l_t is the local time of $|x_t|$ at 0. (Recall we defined the local time of a semi-martignale to be the finite bounded variation part of $|x_t|$. The two definitions agree.) The process l_t is non zero only when $x_t = 0$: $\int_0^t \mathbf{1}_{x_s \neq 0} dl_s = 0$. [To prove this approximate $|x|$ by a sequence $f_n(x) \to |x|$ uniformly and $f_n'(x) \to \operatorname{sign}(x)$.]

With Tanaka's formula, $|B_t| = \int_0^t \operatorname{sign}(B_s) dB_s - L_t$ where L_t is the local time at zero of the Brownian motion, $\sigma\{W_t\} = \sigma\{|B_t|\}$, and so the solution $x_t = B_t$ cannot be determined by W_t alone. We would need to enlarge the filtration of the Brownian motion W_t to obtain one adapted to W_t. This observation would be the key for the Yamada-Watanabe theorem below.

Pathwise uniqueness implies uniqueness in law (for which one does not assume that the solutions are defined on the same probability space). There is in fact a beautiful theorem which states that if pathwise uniqueness holds, then uniqueness in law holds and every solution is a strong solution. [RY99, Thm 1.7, Chapter IX, p 368]. The following version of the theorem is taken from [IW89, Theorem 1.1, Chapter IV]:

Theorem 4.2.7 (Yamada-Watanabe)**.** *Suppose that pathwise uniqueness holds for* (4.2.1) *and suppose that there exists a (possibly weak) solution for any initial condition. Then the SDE has a unique strong solution in the sense below. There is a measurable map $F : \mathbb{R}^n \times \mathcal{C}(\mathbb{R}_+, \mathbb{R}^m) \to \mathcal{C}(\mathbb{R}_+, \mathbb{R}^n)$, where the domain space is given a universally complete product sigma-algebra, such that*

- ▷ *for any initial condition $Y \in \mathcal{F}_0$ and any m-dimensional Brownian motion B, $x_t = F_t(Y, B)$ is a solution to* (4.2.1) *with $X_0 = Y$ a.s.,*
- ▷ *if x_t is a solution to* (4.2.1)*, then $x_t = F_t(X_0, B)$ a.s.*

We now let $F(x, B)$ be the function of Theorem 4.2.7, as a stochastic process, $F_t(x, B)$ is adapted to the (completed) filtration generated by B. For $0 \leqslant s \leqslant t$, there exists $F_{s,t}(x, B)$ such that $F_{s,s}(x, \omega) = x$ a.s. and it satisfdies

$$F_{s,t}(x, B) = x + \sum_{k=1}^m \int_s^t \sigma_k\big(F_{s,r}(x, B)\big)\, dB_r + \int_s^t \sigma_0\big(F_{s,r}(x, B)\big).\, dr$$

The claim can be verified by a change of variable in time, using approximation theorems for stochastic integrals. In particular, $F_{0,t} = F_0$. We call $\{F_{s,t}\}$ the solution flow to the SDE. We also write $F_t \triangleq F_{0,t}$. An SDE is said to have no explosion if from any initial point its (maximal) solution exists for all time.

Condition 4.2.8. [Linear Growth Condition] Suppose that V_j is locally Lipschitz for each $j = 0, \dots, m$ and there is a constant $C > 0$ such that, for every $x \in \mathbb{R}^n$,

$$|\sigma_k(x)| \leqslant C(1 + |x|), \quad k = 1, \dots, m,$$
$$\langle \sigma_0(x), x \rangle \leqslant C(1 + |x|^2).$$

Theorem 4.2.9. *Under Condition 4.2.8 pathwise uniqueness to the SDE* (4.2.1) *holds. Moreover, for any initial condition $X_0 \sim \mu$, there is a global strong solution.*

4.2.1 The Cocycle Property

Theorem 4.2.10. *Assume that* (4.2.1) *has a unique global strong solution. Then, for any* $0 \leqslant s \leqslant t$,

$$F_t(x,B) = F_{s,t}(F_s(x,B),B). \tag{4.2.2}$$

and

$$F_{t-s}\big(F_s(x,B),\theta_s B\big) = F_t(x,B) \quad \text{a.s.}, \tag{4.2.3}$$

where $\theta_s\sigma = \sigma(\cdot+s)-\sigma(s)$ *is the shift operator. This is called the flow (or cocycle) property.*

Proof. This follows from the fact that for every $t \geq s$,

$$\int_s^t H_r\,dW_r = \int_0^{t-s} H_{r+s}\,d(\theta_s W_r), \qquad a.s..$$

This is due to the following fact for adapted sample continuous processes:

$$\sum H_{t_i}(W_{t_{i+1}} - W_{t_i}) \to \int_s^t H_r\,dW_r$$

in probability along partitions of $[0,t]$. Observe that $F_t(x)$ solves

$$F_t(x,B) = F_s(x,B) + \sum_k \int_s^t X_k(F_r(x,B))dB_r^k + \int_s^t X_0(F_r(x,B))dr.$$

Pathwise uniqueness shows that $F_t(x,B) = F_{s,t}(F_s(x,B),B)$. Set $y_r = F_{r+s}(x,B)$, then

$$y_{t-s} = y_0 + \sum_k \int_s^t X_k(y_{r-s})dB_r^k + \int_s^t X_0(y_{r-s})dr.$$

Hene $F_t(x,B) = F_{0,t-s}(F_s(x,B),\theta_s B)$. $\square$

Problem: Try to work out a version with s replaced by a stopping time.

4.2.2 Measure preserving transformation, random dynamical system, and ergodicity

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. A transformation on $T:\Omega\to\Omega$ is measure preserving if $T^*\mathbb{P}=\mathbb{P}$ where $T^*\mathbb{P} = \mathbb{P}\circ T^{-1}$ is the pushed forward measure of $\mathbb{P}$ by T.

Definition 4.2.11. Let T be a measure preserving transformation on a measure space. A measurable set A is T- invariant if $\theta^{-1}(A) = A$. A measure preserving map T is said to be ergodic if every T-invariant set has full or null measure.

Definition 4.2.12. A probability space $(\Omega,\mathcal{F},\mathbb{P})$ together with a family of measure preserving maps $(\theta_t, t\in I)$ is called a (metric) dynamical system.

The map θ_t is measure preserving if $(\theta_t)^*(\mathbb{P}) = \mathbb{P}$. An example of a dynamical system is the Wiener space with $\theta_t\omega = \omega(t+\cdot)-\omega(\cdot)$, $t\in\mathbb{R}_+$.

Definition 4.2.13. A random/stochastic dynamical system on a measurable space $(\mathcal{X}, \mathcal{B})$ over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, (\theta_t)_{t\in I})$ is a measurable mapping

$$\phi : I \times \Omega \times \mathcal{X} \to \mathcal{X}$$

with the cocycle property:

$$\phi(0, \omega, x) = x, \qquad \phi(t+s, \omega) = \phi(t, \theta_s \omega) \circ \phi(s, \omega).$$

This holds almost surely for all $s, t \geq 0$,

See [Arn13] for detailed account of random dynamical systems.

Example 4.2.14. For each x, the solution flow $F_t(x, \omega)$ defined earlier is a random dynamical system.

Definition 4.2.15. If the cocycle identity holds for every $\{F_t(x, \omega), t \geq 0\}$ on the complement of a null set $\mathcal{N}$, we say it has the perfect cocycle property.

A stochastic flow has the perfect cocycle property precisely when the null set $\mathcal{N}$ can be taken independent of x.

Question. When does a stochastic flow has the perfect cocycle property?

4.2.3 Stationary noise, and stochastic dynamical systems

Definition 4.2.16. A stationary noise process is a quadruple consisting of a Polish space $\mathcal{W}$, a Feller transition Markov kernel (P_t) on $\mathcal{W}$ admitting a unique invariant probability measure $\mathbb{P}_{\mathcal{W}}$, and a semi-flow $\theta_t : \mathcal{W} \to \mathcal{W}$ of measurable maps with the properties that:

$$(\theta_t)^* P_t(w, \cdot) = \delta_w$$

for every $w \in \mathcal{W}$.

Note that the last identity is: $P_t(w, \theta_t w' \in A) = 1_A(w)$.

Definition 4.2.17. A stochastic dynamical system on a Polish space $\mathcal{X}$ over a stationary noise process $(\mathcal{W}, P_t, \mathbb{P}_{\mathcal{W}}, \theta_t)$ is a map $\phi : \mathbb{R}_+ \times \mathcal{X} \times \mathcal{W} \to \mathcal{X}$, with the following properties. Writing $\phi_t(x, w) = \phi(t, x, w)$.

(i) For every $s, t > 0$ and $x \in \mathcal{X}$ and all $w \in \mathcal{W}$,

$$\phi_0(x, w) = w,$$
$$\phi_{s+t}(x, w) = \phi_s(\phi(t, \theta_t w), w).$$

(ii) For every $T > 0$, $x \in \mathcal{X}$ and $w \in \mathcal{W}$, define $\Phi_T(x, w) : [0, T] \to \mathcal{X}$ by

$$\Phi_T(x, w) = \phi_t(x, \theta_{T-t} w).$$

(a) Then $\Phi_T \in \mathcal{C}([0, T], \mathcal{X})$.

(b) $(x, w) \mapsto \Phi_T(x, w)$ is continuous from $\mathcal{X} \times \mathcal{W}$ to $\mathcal{C}([0, T], \mathcal{X})$.

Example 4.2.18. Let ψ_t be the stochastic flow of a stochastic differential equation, let us define $\phi_t(x, w) = \psi_t(x, \theta_t^{-1} w)$.

4.2.4 Markov Property

It is now easy to prove the Markov property of the solution:

Theorem 4.2.19. *Assume that the SDE (4.2.1) has a unique global strong solution. Let $P_t(x,\cdot)$ denote the law of $F_t(x,B)$. Then, for any $A \in \mathcal{B}(\mathbb{R}^n)$ and any $0 \leqslant s < t$,*

$$\mathbb{P}(x_t \in A \,|\, \mathcal{F}_s) = P_{t-s}(x_s, A) \quad \text{a.s.}$$

Proof. For any $g \in \mathcal{B}_b(\mathbb{R}^n)$, we have that

$$\begin{aligned}\mathbb{E}\left[g(x_t) \,|\, \mathcal{F}_s\right] &= \mathbb{E}\left[g\Big(F_{t-s}(x_s, \theta_s B)\Big) \,|\, \mathcal{F}_s\right] = \mathbb{E}\left[g\big(F_{t-s}(x,B)\big)\right]\Big|_{x=X_s} \\ &= \int_{\mathbb{R}^n} g(y)\, P_{t-s}(X_s, dy).\end{aligned}$$

□

The map $t \mapsto P_t(x,A)$ inherits measurability from that of $t \mapsto F_t(x,\omega)$. How about measurability in x? This is easy to see if we have an explicit formula and if there exists a (global) smooth stochastic flows. Also we can use Corollary 3.2.2 for this when the semi-group has the C_0-property. See Section 4.6 for whether the solutions has the C_0-property.

4.3 Ergodic Theorems for Markov Processes

If a continuous time homogeneous Markov process (x_t) start from an invariant measure π, let $\mathbb{P}_\pi$ denote its distribution on the path space $\mathcal{C}(\mathcal{X}) = C(I,\mathcal{X})$ where $I \subset \mathbb{R}_+$ is an interval. The Markov property implies that given the transition function, π and $\mathbb{P}_\pi$ are mutually determined. Consider $(\mathcal{C}(\mathcal{X}), \mathcal{B}(\mathcal{C}(\mathcal{X})), \mathbb{P}_\pi)$ and let θ_t be the shift map on path space: if $\sigma : \mathbb{R}_+ \to \mathcal{X}$ then $\theta_t\sigma = \sigma(t+\cdot)$. Then $\theta_{t+s} = \theta_t \circ \theta_s$ and θ_0 the identity map. By Theorem 3.7.1, the Markov process is stationary, the measures $\mathbb{P}_\pi$ are shift invariant. In particular, $(\mathcal{C}(\mathcal{X}), \mathcal{B}(\mathcal{C}(\mathcal{X})), \mathbb{P}_\pi)$ together with $\theta_t\omega = \omega(t+\cdot)$ is a metric dynamical system.

Definition 4.3.1. Let T be a measure preserving transformation on a measure space. A measurable set A is T- invariant if $\theta^{-1}(A) = A$. A measure preserving map T is said to be ergodic if every T-invariant set has full or null measure.

Definition 4.3.2. We say an invariant measure π is ergodic if $\mathbb{P}_\pi$ is ergodic w.r.t. every θ_t.

Remark 4.3.3. For discrete time Markov processes, we have the concept of the smallest time unit, we only need to consider the case of $t = 1$.

For Markov processes with continuous time, we introduce the following definition.

Definition 4.3.4.

- ▷ A set A is invariant if $\theta_t^{-1}(A) = A$ for every t.
- ▷ A measure π is an ergodic invariant measure for a Markov process if for every invariant set A, $\mathbb{P}_\pi(A) \in \{0,1\}$.

A useful theorem is the following:

Theorem 4.3.5. *If a time homogeneous Markov process has a unique invariant probability measure, then $\mathbb{P}_\pi$ (and π) is ergodic.*

The set of invariant probability measures for a Markov process is a convex set, It turns out that the set of ergodic invariant probability measures are precisely the extremal points from the set.

Theorem 4.3.6. *Let π be an invariant probability measure of time homogeneous Markov process, then π is ergodic if and only if it is an extremal of the convex set of probability measures.*

The support of a measure is the intersection of all closed sets of full measure. A point x is in the support of the measure if and only if every open set containing x has positive measure.

Theorem 4.3.7. *Suppose that the transition semi-group T_t is strong Feller (or asymptotically strong Feller). If μ_1 and μ_2 are two distinct invariant probability measures for T_t, then* $\mathrm{supp}(\mu_1) \cap \mathrm{supp}(\mu_2) = \emptyset$.

Corollary 4.3.8. *If the support of every invariant probability measure of a strong Feller transition function contains a common point, then it is ergodic.*

4.3.1 Birkhoff's ergodic theorem

Let T be a measure preserving transformation on a probability space. Let $\mathcal{I}$ denotes the set of T-invariant sets, it is called the invariant σ-algebra.

Theorem 4.3.9. *If T is a measure preserving transformation and $f \in L^1$, then*

$$\frac{1}{n}\sum_{k=1}^{n} f(T^k x) \to \mathbb{E}(f|\mathcal{I})$$

almost surely as $n \to \infty$.

If the measure is ergodic the invariant σ-algebra is trivial, then $\mathbb{E}(f|\mathcal{I}) = \mathbb{E}[f]$.

Theorem 4.3.10. *Let y_t be a stationary ergodic Markov process with invariant measure π then for any $f \in L^1(\pi)$, then as $t \to \infty$,*

$$\frac{1}{t}\int_0^t f(y_s)ds \to \int f d\pi$$

almost surely.

An application is the following stochastic averaging theorem. Suppose that f is Lipschitz continuous and y_t an ergodic Markov process with exponential rate of convergence (this exponential condition is over kill, but makes the proof very easy), then the solutions to the following equations

$$dx_t^\varepsilon = f(x_t^\varepsilon, y_{\frac{t}{\varepsilon}})$$

converges to that of the ODE $\dot{x}_t = \bar{f}(x_t)$ where $\bar{f}(x) = \int f(x,y)dy$.

4.4 Global smooth flows/ strong completeness

A solution $F_t(x,\omega)$ to an SDE is defined for a set of full measure, and this set of full measure may depend on x.

Definition 4.4.1. Suppose that the SDE has a unique strong solution and is conservative. we say the SDE is complete.

Definition 4.4.2. An SDE is said to have a global smooth solution flow (or to be strongly complete) if it is complete and there exists a version of F such that $(t, x) \mapsto F_t(x, \omega)$ is continuous on $[0, T] \times \mathbb{R}^d$ almost surely for any $T > 0$.

Remark 4.4.3. If an ODE $\dot{y}_t = V(y_t)$ is well posed and has no explosion, we say the vector field is complete. If it is complete and if V is C^1, then $(t, x) \mapsto F_t(x)$ is continuous. —- This may fail for an SDE. Hence, if an SDE has a global solution flow, we say that it is strongly complete.

Let V be a C^1 vector field on $\mathbb{R}$, then $\dot{x}_t = V(x_t)dB_t$ is strongly complete if complete.

Example 4.4.4. [Elw78] Let $dx_t = dB_t$ on $\mathbb{R}^n \setminus \{0\}$ where $n \geq 2$. Then $F_t(x, \omega) = x + B_t(\omega)$ exists for all time for almost surely all ω (for a BM does not hit a point.) But the SDE is not strongly complete. Given any ω and any time t there exists x such that $x + B_t(\omega) = 0$.

The infinitesimal generator of an SDE does not determine the flow problem.

Example 4.4.5. Let $\mathcal{L} = r^4 \Delta$ where $r^2 = x^2 + y^2$. Then $\log r$ is a Lyapunov function. So the Markov process with this generator is conservative.

▷ The following equation is strongly complete.

$$\begin{aligned} dx_t &= \frac{x_t}{r_t} dB_t^1 - \frac{y_t}{r_t} dB_t^2 \\ dy_t &= \frac{y_t}{r_t} dB_t^1 + \frac{x_t}{r_t} dB_t^2 \end{aligned}$$

It can be written as $dz_t = \frac{z_t}{|z_t|} dB_t$ on $\mathcal{C}$.

▷ The following is not strongly complete

$$\begin{aligned} dx_t &= (y_t^2 - x_t^2) dB_t^1 + 2x_t y_t dB_t^2 \\ dy_t &= -2x_t y_t dB_t^1 + (y_t^2 - x_t^2) dB_t^2. \end{aligned}$$

This is Elworthy example on $\mathbb{R}^2 \setminus \{0\}$, turned into $\mathbb{R}^2$ by $x \mapsto \frac{1}{x}$.

In the last example, the vector fields are quadratic functions of x, might be asking whether the problem comes from the vector fields grows too fast? In fact you would find examples of SDEs with C^∞ smooth and bounded vector fields, for which there is no global smooth solution flows, see [LS11].

Theorem 4.4.6. *Suppose that the SDE is strongly complete. Then the family of probability measures $P_t(x, \cdot)$ defined by the solution flow is a transition function. Furthermore, the solution $F_t(x)$ is a Markov process with transition function $P_t(x, \cdot)$ and the initial distribution δ_x.*

Proof. Since $(t, x) \mapsto F_t(x, \omega)$ is continuous, for every $A \in \mathcal{B}(\mathcal{X})$, $(t, x) \mapsto P_t(x, A)$ is measurable. Furthermore,

$$\begin{aligned} P_{t+s}(x, A) &= \mathbb{P}\Big(F_t(x, \omega) \in A\Big) = \mathbb{E}\mathbb{E}\Big(\mathbf{1}_A(F_{s,s+t}\Big(F_s(x, \omega), \theta_s(\omega)\Big) \mid \Big) \\ &= \mathbb{E}\mathbb{E}\Big(\mathbf{1}_A\Big(F_{s,s+t}\Big(F_s(x, \omega), \theta_s(\omega)\Big)\Big) \mid \mathcal{F}_s\Big) \\ &= \mathbb{E}\Big((P_t \mathbf{1}_A)(F_s(x, \omega))\Big) = \int_{\mathcal{X}} P_t \mathbf{1}_A(y) P_s(x, dy) \\ &= \int_{\mathcal{X}} P_t(y, A) P_s(x, dy). \end{aligned}$$

Also, since $F_0(x) = x$, $P_0(x, A) = \delta_x$. □

Remark 4.4.7. This continuous dependence property is essential for numerical simulation of solutions. Note that stopping times $\inf\{t : |F_t(x)| \geq \mathbb{R}\}$ does not depend on x continuously, in general.

4.5 The derivative flow

Let us consider $x \mapsto F_t(x, \omega)$. Let us denote by $v_t = TF_t(v_0)$ its derivative in the direction v_0, whenever it exists.

Let us change notation and let X_k, A be vector fields on a complete connected Riemannian manifolds (e.g. $\mathbb{R}^n$, the spheres, and the tori). On $M = \mathbb{R}^n$, $X_k : \mathbb{R}^n \to \mathbb{R}^n$ and $A_0 : \mathbb{R}^n \to \mathbb{R}^n$. Consider

$$dx_t = \sum_{k=1}^{m} X_k(x_t) \circ dB_t^k + A(x_t)dt. \tag{4.5.1}$$

We have to use Stratonovich integral on a manifold. On $\mathbb{R}^n$ this translates to

$$dx_t = \sum_{k=1}^{m} X_k(x_t)dB_t^k + Z(x_t)dt,$$

where $Z = A + \frac{1}{2}\sum_{k=1}^{m} DX_k X_k$. The difference between the two is that the Itô form equation required one less degree of regularity. We shall not be worried about the regularity and assume comfortable X_k in C^2 (or in C^3 to make it even simpler) and and $X_0 \in C^1$.

Then there exists a unique solution $F_t(x, \omega)$ up to an explosion time τ. We shall mainly assume $\tau(x, \omega) = \infty$ for every x, although this is not necessary. Then

$$\lim_{\varepsilon \to 0} \frac{F_t(x + \varepsilon v, \omega) - F_t(x, \omega)}{\varepsilon}$$

exists in probability and solves the following equation (the derivative equation)

$$dv_t = \sum_{k=1}^{m} (DX_k)_{x_t}(v_t) \circ dB_t^k + (DA)_{x_t}(v_t)dt,$$

where $(DX_k)_x(v)$ and $(DA)_x(v)$ denote the directional derivatives of X_k and A in the direction of v at the point x.

On a manifold, we use for example the Levi-Civita connection to differentiate the vector fields

$$Dv_t = \sum_{k=1}^{m} (\nabla X_k)_{x_t}(v_t)dB_t^k + (\nabla A)_{x_t}(v_t)dt.$$

Lemma 4.5.1. *Suppose that X_i, A are in C^2 with bounded derivatives. Let x, v be fixed. Set*

$$G_t^\varepsilon \triangleq \frac{1}{\varepsilon}(F_t(x + \varepsilon v) - F_t(x)).$$

Then $\sup_{\varepsilon \in (0,1]} \sup_{t \leqslant T} \mathbb{E}|G_t^\varepsilon|^p < \infty$ *for any* $p \geq 1$.

Proof. (exercise) □

Lemma 4.5.2. *Suppose that X_i, A are in C^2 with bounded derivatives. Fix x, v again. Then,*

$$\mathbb{E}|G_t^\varepsilon - G_s^{\varepsilon'}|^{2p} \lesssim |s - t|^p + |\varepsilon - \varepsilon'|^p$$

Proof. We will take $p = 2$ and $s = t$ in the proof, it is not difficult to extend the proof below to the general case. Ee denote $x_t^\varepsilon = F_t(x + \varepsilon v)$ and $x_t = F_t(x)$. Then,

$$dG_t^\varepsilon = \frac{1}{\varepsilon}\Big(\sigma(x_t^\varepsilon) - \sigma(x_t)\Big)dB_t + \frac{1}{\varepsilon}\Big(A(x_t^\varepsilon) - A(x_t)\Big)dt.$$

By taking a stopping time if necessary, we could assume the local martingale part in the formula below, which we denote by N_t, to be a martingale.

$$\begin{aligned}\frac{d}{dt}|G_t^\varepsilon - G_t^{\varepsilon'}|^2 &= N_t + 2\left\langle G_t^\varepsilon - G_t^{\varepsilon'}, \frac{1}{\varepsilon}\Big(A(x_t^\varepsilon) - A(x_t)\Big) - \frac{1}{\varepsilon'}\Big(A(x_t^{\varepsilon'}) - A(x_t)\Big)\right\rangle dt\\ &+ \sum_{k=1}^m \left|\frac{1}{\varepsilon}\Big(X_k(x_t^\varepsilon) - X_k(x_t)\Big) - \frac{1}{\varepsilon'}\Big(X_k(x_t^{\varepsilon'}) - X_k(x_t)\Big)\right|^2 dt.\end{aligned}$$

By Taylor's expansion,

$$\frac{1}{\varepsilon}\left(A(x_t^\varepsilon) - A(x_t)\right) = \int_0^1 (DA)_{x_t + r(x_t^\varepsilon - x_t)}(G_t^\varepsilon)dr,$$

Hence,

$$\begin{aligned}&\frac{1}{\varepsilon}\Big(A(x_t^\varepsilon) - A(x_t)\Big) - \frac{1}{\varepsilon'}\Big(A(x_t^{\varepsilon'}) - A(x_t)\Big)\\ &= \int_0^1 (DA)_{x_t + r(x_t^\varepsilon - x_t)}(G_t^\varepsilon) - (DA)_{x_t + r(x_t^{\varepsilon'} - x_t)}(G_t^{\varepsilon'})dr\\ &= \int_0^1 r(D^2A)_{x_t + r(x_t^\varepsilon - x_t)}(G_t^\varepsilon, (x_t^\varepsilon - x_t^{\varepsilon'}))dr\\ &+ \int_0^1 (DA)_{x_t + r(x_t^{\varepsilon'} - x_t)}(G_t^\varepsilon - G_t^{\varepsilon'})dr.\end{aligned}$$

Putting everything together, N_t, to be a martingale.

$$\mathbb{E}|G_t^\varepsilon - G_t^{\varepsilon'}|^2 \leqslant \sup_{s\leqslant t}\mathbb{E}|G_s^\varepsilon|^2 \sup_{s\leqslant t}\mathbb{E}|x_t^\varepsilon - x_t^{\varepsilon'}|^2 + \int_0^t \mathbb{E}|G_s^\varepsilon - G_s^{\varepsilon'}|^2 dds.$$

Since

$$\mathbb{E}|x_t^\varepsilon - x_t^{\varepsilon'}|^2 \leqslant |\varepsilon - \varepsilon'|^2,$$

the required estimates for $p = 1$ follows. The estimates for $2p$ can be obtained analogously. □

Theorem 4.5.3. *[Li94c, Theorem 3.1] Suppose that X_i are C^2 and $A \in C^1$. Suppose that $\zeta(x_0) = \infty$ a.s. for some x_0. Suppose that*

$$\sup_{x\in K} E\left(\sup_{s\leqslant t}\|T_xF_s\mathbf{1}_{s<\zeta(x)}\|^n\right) < \infty, \tag{4.5.2}$$

Then the SDE is strongly complete.

Let

$$\begin{aligned}H(v,v) &= 2\langle DA(v), v\rangle + \sum_k |DX_k(v)|^2 + \sum_k \langle\nabla^2 X_k(X_k, v), v\rangle\\ &+ \sum_k \langle\nabla X_k(\nabla X_k(v)), v\rangle + \sum_k |\nabla X_k(v)|^2\\ &+ (n-2)\sum_{k=1}^m \frac{1}{|v|^2}\langle DX_k(v), v\rangle^2.\end{aligned}$$

Theorem 4.5.4. *[Li94c, Thm 5.1+section 6] Suppose that X_i are C^2 and there is no explosion. Assume there is a function $f : \mathbb{R}^n \to \mathbb{R}_+$ such that:*

(i) for all $t > 0$, K compact.

$$\sup_{x\in K} E\left(e^{6n^2\int_0^t f(F_s(x))ds}\right) < \infty,$$

(ii) $\sum_k |DX_k(x)|^2 \leqslant f(x)$.

(iii) $H(x)(v,v) \leqslant 6pf(x)|v|^2$

Then

$$E\left(\sup_{s\leqslant t} |T_xF_s|^p\right) < cE\left(\exp^{6n^2\int_0^t f(F_s(x))ds}\right).$$

In particular the SDE is strongly complete and the random dynamical system $F_t(x,\omega)$ has the perfect cocycle property.

4.5.1 Examples

We will give two examples for which the strong completeness holds. Consider

$$dx_t = \sum_{k=1}^m X_k(x_t)dB_t^k + A(x_t)dt.$$

Then

$$H(v,v) = 2\langle DA(v), v\rangle + \sum_k |DX_k(v)|^2 + (n-2)\sum_1^m \frac{1}{|v|^2}\langle DX_k(v), v\rangle^2.$$

Lemma 4.5.5. *Assume non-explosion for simplicity. For any C^2 function g,*

$$E(e^{cg(x_t)}) \leqslant e^{c(g(x_0)+kt}$$

where k is a constant, provided that $\frac{1}{2}\sum_i |Dg(X_i)|^2 + \frac{1}{2}\sum D^g(X_i, X_i) + Dg(A)$ is bounded above.

A straightforward application of Theorem 4.5.4 gives two new results:

Theorem 4.5.6. *[Li94c, section 6] If either Condition 4.5.7 or Condition 4.5.8 holds, the SDE is strongly complete.*

Condition 4.5.7. Suppose that each X_k, A are C^2 and C^1 respectively and there is a constant $C > 0$ such that, for every $x, v \in \mathbb{R}^n$,

$$\begin{aligned} |X_k(x)| &\lesssim 1 + |x|, \quad k = 1, \dots, m, \\ \langle A(x), x\rangle &\lesssim 1 + |x|^2, \\ |DX_k(x)| &\lesssim 1 + \ln(1 + |x|^2), \quad k = 1, \dots, m, \\ \langle DA(x)(v), v\rangle &\lesssim (1 + \ln(1 + |x|^2))|v|^2. \end{aligned}$$

Condition 4.5.8.

$$\begin{aligned} |X_i(x)| &\lesssim c(1 + |x|^2)^{\frac{1}{2}-\varepsilon} \\ \langle x, A(x)\rangle &\lesssim c(1 + |x|^2)^{1-\varepsilon} \\ |DX_i(x)|^2 &\lesssim c(1 + |x|^2)^{\varepsilon} \\ \langle DA(x)(v), v\rangle &\lesssim c(1 + |x|^2)^{\varepsilon}|v|^2. \end{aligned}$$

If X_i are globally Lipschitz continuous, this can be proved using fixed point theorem.

4.5.2 Local smooth flow

We will now introduce the local flow theorem, a theorem of Carverhill-Elworthy and H. Kunita [Kun90]

Theorem 4.5.9. *Suppose that X_k are C^2. Then there exists a unique solution $(F_{s,t}(x,\omega))$ and life time $\zeta(x,\omega))$ such that there exists a null set outside of which the following holds:*

(i) For each x, $(F_{0,t}(x,\omega), t < \zeta(x,\omega))$ is the maxima solution with initial value x.

(ii) Let

$$M_t(x,\omega) = \{x : t < \zeta(x,\omega\}.$$

On $M_t(x,\omega)$, for $0 \leqslant s \leqslant t$,

$$F_{s,t} \circ F_{0,s}(x,\omega) = F_{0,t}(x,\omega).$$

4.6 The C_0-property

Theorem 4.6.1. *Assume that X_i are locally Lipschitz and grow at most linearly (The drift only need to have the bound $\langle X_0(x), x\rangle \leqslant C(1+|x|^2)$). Let $\mathcal{L}$ be the infinitesimal generator.*

(i) If f is continuous, so is $P_t f$ (The Feller property).

(ii) If $f \in C_0$, so is $P_t f$ (The C_0-property).

(iii) P_t is a strongly continuous semi-group on $\mathcal{C}_0$.

(iv) If $f \in \mathcal{C}_0 \cap \mathcal{C}_b^2$, then $P_t f$ solves the Cauchy problem $\frac{\partial u}{\partial t} = \mathcal{L}u$ with $u(0) = f$.

Proof. (1) The Feller property follows from the bounded and continuity of the map $x \mapsto f(F_t(x))$.

(2) Let $f \in \mathcal{C}_0(\mathbb{R}^n)$. For any $\varepsilon > 0$, choose a compact set $K = \overline{B(x, R/2)}$ such that $|f(x)| \leqslant \frac{\varepsilon}{2}$ outside of K.

Let $|x| > R$. On $\{|F_t(x) - x| \leqslant R/2\}$, we have $|F_t(x)| \geq |x| - |F_t(x) - x| > R/2$. Then,

$$\begin{aligned}
\left|P_t f(x)\right| &\leqslant \left|\mathbb{E}\Big[f(F_t(x))\mathbf{1}_{\{|F_t(x)-x|>R/2\}}\Big]\right| + \left|\mathbb{E}\Big[f(F_t(x))\mathbf{1}_{\{|F_t(x)-x|\leqslant R/2\}}\Big]\right| \\
&\leqslant |f|_\infty \mathbb{P}(|F_t(x)| \geq R/2) + \frac{\varepsilon}{2}.
\end{aligned}$$

Since $\mathbb{E}|F_t(x) - x|^2 \leqslant C(t^2+t)$, $\mathbb{P}(|F_t(x)| \geq R/2) \leqslant \frac{1}{R} C(t^2+t) \to 0$, as $x \to \infty$. This shows that $P_t f \in C_0$.

(3) If $f \in \mathcal{C}_0$, then

$$\begin{aligned}
\left|P_t f(x) - f(x)\right|_\infty &\leqslant \left|\Big[\mathbb{E}[f(F_t(x)) - f(x)]\mathbf{1}_{\{|F_t(x)-x|\leqslant\varepsilon\}}\Big]\right| + \left|\Big[(P_t f(x) - f(x))\mathbf{1}_{\{|F_t(x)-x|\geq\varepsilon\}}\Big]\right| \\
&\leqslant \sup_{y\in B_\varepsilon(x)} |f(x) - f(y)| + 2\|f\|_\infty \mathbb{P}(\{|F_t(x) - x| \geq \varepsilon\}) \to 0,
\end{aligned}$$

as $\varepsilon \to 0$. This implies that P_t is a strongly continuous semigroup on $\mathcal{C}_0$.

By Itô's formula, if $f \in C_K^2$,

$$f(x_t) = f(x_0) + \int_0^t df(X_k(x_s))dB_s^k + \int_0^t \mathcal{L}f(x_s)ds.$$

he local martingale part is a square integrable martingale:

$$\mathbb{E}\left(\int_0^t df(X_k(x_s))dB_s^k\right)^2 = \int_0^t \mathbb{E}|df(x_s)|^2 ds \leqslant |df|_\infty \int_0^t \mathbb{E}(|x_s|^2)ds < \infty.$$

Since $f \in \mathcal{C}_0$, $\mathcal{L}f \in \mathcal{C}_0$, by Proposition 3.4.2, f is in the the domain of the generator $\mathcal{A}$ of the semi-group P_t and $\mathcal{A}f = \mathcal{L}f$.

□

Alternatively, we could prove the last statement as follows:

$$P_t f(x_0) = \mathbb{E}f(x_t) = f(x_0) + \mathbb{E}\int_0^t \mathcal{L}f(x_s)ds.$$

Since

$$\mathcal{L}f(x) = \frac{1}{2}\sum_{i,j=1}^n \sum_k (X_i^k \sigma_j^k)(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + Af(x).$$

Then $\mathcal{L}f \leqslant c(1+|x|^2)$. Since $\mathbb{E}\sup_{s\leqslant t}|x_s|^2 < \infty$,

$$P_t f(x_0) = f(x_0) + \int_0^t \mathbb{E}\mathcal{L}f(x_s)ds.$$

If $f \in C_K^2$, then $\mathcal{L}f$ is bounded, hence $s \mapsto \mathbb{E}\mathcal{L}f(x_s)$ is continuous,

$$\lim_{t\to 0}\frac{P_t f(x_0) - f(x_0)}{t} = \mathcal{L}f(x_0).$$

An easy way to conclude that f is in the domain of the generator is to see $f(X_t) - \int_0^t \mathcal{L}f(x_r)dr$ is a martingale, and use Theorem 3.4.2.

See [Li94a] for a a study of a generalised notion, the C^*-property, which also connect to a PDE problem.

4.7 Exercises

Exercise 4.7.1. Let $A \in \mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and $B_t = (B_t^1, \dots, B_t^m)$ an $\mathbb{R}^m$-valued standard Brownian motion. Let

$$X \in \mathcal{C}_b^\infty(\mathbb{R}^d; \mathbb{L}(\mathbb{R}^m, \mathbb{R}^d)).$$

We may set $X_k = X(x)(e_k)$ where e_k is an o.n.b. of $\mathbb{R}^m$. Then the distribution of the solutions to

$$dx_t = \sum_{k=1}^m X_k(x_t)dB_t^k + Z(x_t)dt,$$

in independent of the choice of the basis.

Consider the SDE on $\mathbb{R}^d$:

$$dx_t = \sum_{k=1}^m X_k(x_t)dB_t^k + Z(x_t)dt, \tag{4.7.1}$$

$$dv_t = \sum_{k=1}^m (DX_k)_{x_t}(v_t)dB_t^k + (DA)_{x_t}(v_t)dt. \tag{4.7.2}$$

(i) Show that for any initial condition $(x_0, v_0) \in \mathbb{R}^d$ (4.7.1)-(4.7.2) has a unique global strong solution, and the system of equations is strongly Complete.

(ii) Show that for each $t \geqslant 0$,

$$\left|\frac{F_t(x+\varepsilon v) - F_t(x)}{\varepsilon} - v_t\right|$$

converges in L_2 as $\varepsilon \to 0$ and $(F_t(x), v_t)$ solve the SDE (4.7.2) with the initial condition (x, v).

The solution to (4.7.2) is linear in v. This will be called the derivative flow and denoted by $J_t(x, v)$.

(iii) Consider the system of equation (4.7.1) together with the following equation

$$\begin{cases} dK_t = (D^2 A)_{x_t}\big(J_t(x,v), J_t(x,w)\big)dt + (DA)_{x_t}\big(K_t\big)dt \\ \qquad\qquad + (D^2\sigma)_{x_t}\big)\big(J_t(x,v), J_t(x,w)\big)\, dB_t + (DA)_{x_t}\big(K_t\big)\, dB_t, \\ K_0 = 0. \end{cases}$$

Show that for each $x_0 \in \mathbb{R}^d$, $v, w \in \mathbb{R}^d$, the L^2 derivative of $x \mapsto F_t(x)$ at x_0 in the direction (v, w) exists, and its derivative if K_t.

Exercise 4.7.2. Suppose that X_i, A are in C^2 with bounded derivatives. Let $F_t(x)$ denote its solution flow, then

$$\mathbb{E}|F_t(x) - F_s(y)|^{2p} \lesssim |x-y|^{2p} + |s-t|^p.$$

4.8 The differentiation formula

We return to SDEs

$$dx_t = \sum_{k=1}^m X_k(x_t)\, dB_t^k + b(x_t)\, dt. \tag{4.8.1}$$

We assume that the coefficients are locally Lipschitz continuous and that the equation is well posed with a unique global solution $F_t(x)$ for every initial condition. Recall the solution correspond to the semigroup

$$P_t f(x) = \mathbb{E}[f(F_t(x))].$$

We denote by $\mathcal{L}$ its infinitesimal generator:

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^n a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{l=1}^n b_l(x)\frac{\partial}{\partial x_l}.$$

Let v_t denote its derivative flow, the solution to the linearised SDE with the initial value v_0, along $x_t = F_t(x_0)$,

$$dv_t = \sum_{k=1}^m (DX_k)_{x_t}(v_t) dB_t^k + (Db)_{x_t}(v_t)dt. \tag{4.8.2}$$

Under restrictions on the growth of $|DX_i|$ and $|Db|$, solution of the above equation exists also for all time, see [Li94c, Thm 5.1+section 6]. Let $(F_t(x), D_xF_t(v))$ denote the solution to the above system of equations with the initial value (x, v), then the semi-group corresponds to it is:

$$Q_t\phi(x_0, v_0) = \mathbb{E}\big[\phi(F_t(x), D_xF_t(v))\big].$$

If f is a differentiable function, we treat df as a real valued function on $\mathbb{R}^n \times \mathbb{R}^n$, it is linear in the second variable (so df is a differentiable 1-form).

Lemma 4.8.1. *Assume that X_i are C^2, suppose that the SDE (4.8.1) is strongly complete, and $\sup_{x\in K}\mathbb{E}|T_xF_t|^2<\infty$ for any bounded set K. Then for any $f\in BC^1$,*

$$Q_t df(x_0,v_0)=(dP_tf)_{x_0}(v_0). \tag{4.8.3}$$

See also [Li94b]. For the next theorem, we introduce some notation, We define $X(x):\mathbb{R}^m\to\mathbb{R}^n$ as follows. Taking $\{e_i\}$ an o.n.b. of $\mathbb{R}^m$,

$$X(x)(e)=\sum_{i=1}^m X_i(x)\langle e_i,e\rangle.$$

Suppose $\mathcal{L}$ is elliptic, then $X(x)$ is a surjective, denote by $Y(x)$ the right inverse of $X(x)$. The following result is taken from [Li92, EL94].

Theorem 4.8.2. *[Differentiation / BEL Formula] Suppose that the SDE (4.8.1) is strongly complete, $\sup_{t\leqslant T}\mathbb{E}|T_xF_t|^2<\infty$ and the conclusion of the lemma. Suppose also that Y is bounded. Suppose that P_tf is $C^{1,2}$ for any $f\in BC^2$. Then for any $f\in\mathcal{B}_b(\mathbb{R}^n,\mathbb{R})$,*

$$(dP_tf)_{x_0}(v_0)=\frac{1}{t}\mathbb{E}\left[f(x_t)\int_0^t\langle Y(x_s)v_s,dB_s\rangle\right]. \tag{4.8.4}$$

In particular, the Markov process has the strong Feller property.

Proof. Let $f\in BC^2$, Itô's formula,

$$f(x_t)=P_Tf(x_0)+\int_0^T\big(dP_{T-s}f(x_s)\big)X(x_s)dB_s,$$

By $dP_{T-s}f(x_s)\big)X(x_s)dB_s$ we mean $\sum_{i=1}^m dP_{T-s}f(x_s)\big)X_i(x_s)dB_s^i$. And

$$\begin{aligned}
&\mathbb{E}\left[f(x_t)\int_0^T\langle Y(x_s)v_s,dB_s\rangle\right]\\
&=\mathbb{E}\left[\int_0^T dP_{T-s}f(x_s)(XdB_s)\int_0^t\langle Y(x_s)v_s,dB_s\rangle\right]\Big]\\
&=\mathbb{E}\left[\int_0^T dP_{T-s}f(x_s)(XY)(x_s)v_sds\right]\\
&=\mathbb{E}\left[\int_0^T Q_{T-s}(df(x_s,v_s)ds\right]\\
&=T\,Q_Tdf(x_0,v_0).
\end{aligned}$$

Next let $f\in BC$, take $f_n\to f$ uniformly. Then, $\mathbb{E}\left[f_n(x_t)\int_0^T\langle Y(x_s)v_s,dB_s\rangle\right]$ converges locally uniformly as $n\to\infty$. This means the required formula holds and P_Tf is C^1 and furthermore, for any $T>0$,

$$|P_Tf(x)-P_Tf(y)|\leqslant C|x-y|$$

for any bounded continuous f, i.e. the total variational distance of the probability measures are bounded:

$$\|P_T(x,\cdot)-P_T(y,\cdot)\|_{TV}\leqslant C|x-y|,$$

which means for any $f\in\mathcal{B}_b$, P_Tf is in BC^1. We now use the fact the formula holds for bounded continuous functions to complete the proof. $\square$

From the proof of the theorem we see that $\|P_T(x,\cdot)-P_T(y,\cdot)\|_{TV}\leqslant C|x-y|$. A partial converse holds.

Remark 4.8.3. Suppose that Equation (4.8.4) holds fro any bounded measurable function, and $\sup_{s\leqslant t}\mathbb{E}|TF_s(v)|^2 ds<\infty$ for any v, then

$$\|P_T(x,\cdot)-P_T(y,\cdot)\|_{TV}\leqslant C|x-y|,$$

To see this it is sufficient to show for any f bounded measurable, $|P_tf(x)-P_tf(y)|\leqslant C|f|_\infty|x-y|$.

4.8.1 Application to uniqueness of invariant measures

It is relatively easy to show that a Markov process X_t has an invariant measure. If P_t is Feller and there exists x_0 with $\{P^n(x_0,\cdot)\}_{n\geq 0}$ tight (relatively compact), then

$$\nu_N(f)\triangleq\frac{1}{N}\int_0^N f(X_s)ds$$

has an accumulation point, which can be shown with the Feller property to be invariant for T_t. Recall that $\int_{\mathcal{X}}T_tf d\mu=\int f d\pi$ for every $t\geq 0$ and for every f bounded measurable. It is much harder to prove uniqueness.

Let (X_t) be a Markov process with initial distribution an invariant measure π, we denote by $\mathbb{P}_\pi$ the measure it induces on the path space (i.e. the law of the process). Recall from Theorem 4.3.7 that if T_t has the strong Feller property, we can distinguish ergodic invariant probability measures by their supports. If $\mathcal{X}$ is an infinite dimensional space, the strong Feller property often fails.

Definition 4.8.4. A function $d:\mathcal{X}\times\mathcal{X}\to\mathbb{R}_+$ is a pseudo-metric if $d(x,x)=0$ for every x, and the triangle inequality holds.

A pseudo-metric may fail to separate points. The discrete metric ρ which assigns $\rho(x,y)=1$ if $x\neq y$ is on the other side of the spectrum.

Definition 4.8.5. A family of pseudo-metric d_n is said to be totally separating, if d_n is an increasing sequence and

$$\lim_{n\to\infty}d_n(x,y)=1$$

for any $x\neq y$.

We say $d_1\geq d_2$ if $d_1(x,y)\geq d_2(x,y)$ for any $x,y\in\mathcal{X}$.

Example 4.8.6. On $\mathcal{C}_0(\mathbb{R}^n)$, we define for a sequence of numbers $a_n>0$, increasing to infinity,

$$d_n(x,y)=1\wedge\Big(a_n\sup_{s\in[-n,n]}|x_s-y_s|\Big).$$

Then d_n is totally separating.

Recall that if μ_1 and μ_2 are two positive finite measures of the same mass on a metric space,

$$\|\mu_1-\mu_2\|_{TV}=\inf_{\nu\in\mathcal{C}(\mu_1,\mu_2)}\nu\big(\{(x,y):x\neq y\}\big)=\inf_{\nu\in\mathcal{C}(\mu_1,\mu_2)}\int_X\rho(x,y)d\nu.$$

where $\mathcal{C}$ is the set of coupling of μ_1 and μ_2 and ρ the metric. We generalise this notion to psudo-metrics.

Definition 4.8.7. Let d be a pseudo-metric. Let μ_1 and μ_2 be two positive finite measures with equal mass. Define,

$$\|\mu_1 - \mu_2\|_d = \inf_{\nu \in \mathcal{C}(\mu_1, \mu_2)} \int d(x, y) d\nu,$$

where $\mathcal{C}$ is the set of positive measures with marginals μ_1 and μ_2.

Remark 4.8.8. Let d_n be a totally separating family of pseudo-metric, then

$$\lim_{n\to\infty} \|\mu_1 - \mu_2\|_{d_n} = \|\mu_1 - \mu_2\|_{TV}$$

for any two positive Borel measures of equal mass.

Definition 4.8.9. [HM06] A transition function is Asymptotically Strong Feller if for every $x \in \mathcal{X}$, there exist an increasing sequence of numbers $t_n > 0$ and a sequence of totally separating pseudo-metrics such that

$$\lim_{r\to 0} \limsup_{n\to\infty} \sup_{y \in B_x(r)} \Big\| P_{t_n}(x, \cdot) - P_{t_n}(y, \cdot) \Big\|_{d_n} = 0.$$

If the conclusion of Theorem 4.8.2, the BEL formula, holds, then the solutions of the SDE has the strong Feller property, its generalisation in finite dimensional space is the following: The following results are convenient for studying Stochastic Partial Differential Equations.

Theorem 4.8.10. *[DPZ96] Let T_t be a Markov semigroup on the set of real valued bounded measurable functions on a Hilbert space. It is strong Feller if there exists a function $C : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any bounded function $f : H \to \mathbb{R}$ with bounded first order derivative,*

$$|\nabla P_t f(x)| \leqslant C(|x|)(1 + \|f\|_\infty).$$

Theorem 4.8.11. *[HM06] Let T_t be a Markov semigroup on the set of real valued bounded measurable functions on a Hilbert space. It is asymptotically strong Feller if there exists a function $C : \mathbb{R}_+ \to \mathbb{R}_+$, a sequence of increasing times t_n and $\delta_n \to 0$ such that for any bounded function $f : H \to \mathbb{R}$ with bounded first order derivative,*

$$|\nabla P_{t_n} f(x)| \leqslant C(|x|)(1 + \|f\|_\infty + \delta_n \|\nabla f\|_\infty).$$

Example 4.8.12. Consider

$$dx_t = -x_t dt + dB_t, \qquad \dot{y}_t = -y_t.$$

Let $f(x, y) = \text{sign}(y)$. Then $P_t f(x, y) = \mathbb{E}[f(y_t)] = \text{sign}(y_t)$. This is not a continuous fucntion, the strong Feller property does not hold. However,

$$\nabla P_t f(x, y) = \nabla[f(x_t, y_t)] = |\mathbb{E}[df_{(x_t, y_t)}(e^{-t}, e^{-t})]| \leqslant \|df\|_\infty e^{-t} \to 0.$$

Hence the asymptotic strong Feller property holds. Of course this system trivially has a unique invariant measure (they are decoupled, the elliptic diffusion has a unique invariant measure, the y process has δ_0 as invariant measure, any invariant measure is a coupling of the two, hence the product measure.

Example 4.8.13. Consider

$$dx_t = -(x_t - x_t)^3 dt + dB_t, \qquad \dot{y}_t = -y_t.$$

The same conclusion holds, using Theorem 4.8.2 on the first eqution and the contraction on the second.

Example 4.8.14. The infinite dimensional Ornstein-Uhlenbeck process

$$u(x,t) = \sum_{k\in Z} \hat{u}(k,t)\exp(ikx),$$

where $x \in [-\pi, \pi]$, so we treat $u(\cdot, t)$ as a function in $L^2([-\pi,\pi])$. Its complex Fourier coefficients solve the SDEs

$$d\hat{u}(k,t) = -(1+|k|^2)\hat{u}(k,t)dt + \exp(-|k|^3)d\beta_k(t)$$

with β_k independent standard complex Brownian motions. Then, it fails the strong Feller property. Asymptotic strong Feller property holds.

We conclude the section with the following observation. If a Markov transition function $P(x,dy)$ is continuous in the total variation norm, then the transition semigroup is strong Feller. The former is stronger, because the convergence

$$\lim_{x\to x_0}\sup_A |T_t\mathbf{1}_A(x) - T_t\mathbf{1}_A(x_0))| = \lim_{x\to x_0}\sup_A |P(x,A) - P(x_0,A)| = 0,$$

is uniform in the set A.

There is a theorem which states that the composition of two strong Feller Markov kernels is continuous in the total variation norm. See the notes [Hai09] and [Sei01]. By the Chapman-Kolmogorov equations, a continuous time strong Feller Markov semigroup is continuous in total variation norm as soon as the time is positive. There are counter example of strong Feller Markov processes not continuous in the total variation norm.

4.9 Feynman-Kac Formla

Let $V : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. Let us now consider the Cauchy problem for

$$\frac{\partial}{\partial t}u = (\frac{1}{2}\Delta - V)u, \qquad u(0,x) = f(x). \tag{4.9.1}$$

A function $u : (0,\infty)\times\mathbb{R}^n \to \mathbb{R}$ is said to be a classical solution if is in $\mathcal{C}^{1,2}$, i.e. continuously differentiable in t and twice in x, the equation is satisfied pointwise, and $\lim_{(t,x)\to(0,x_0)} u(t,x) = f(x_0)$ for any x_0.

Theorem 4.9.1. *Let V be continuous and bounded from below, f is bounded Borel measurable, and u a classical solution of (4.9.1). Then*

$$u(t,x) = \mathbb{E}\left[f(x+B_t)\int_0^t e^{-V(x+B_r)dr}dr\right],$$

where B_t is a standard Brownian motion on $\mathbb{R}^n$.

Proof. We observe that $K_t \triangleq \int_0^t e^{-V(x+B_r)}dr$ is differentiable and $\frac{\partial}{\partial dt}K_t = -V(x+B_t)K_t$. Let $T>0$ and $t<T$, then

$$\frac{\partial}{\partial t}u(T-t,x) + (\frac{1}{2}\Delta + V(T-t))u = 0.$$

Observe that the quadratic variation $\langle U, K\rangle_t$ vanishes. Hence by Itô's formula,

$$\begin{aligned}K_t\, u(T-t, x+B_t) =& u(T,x) - \int_0^t V(x+B_s)K_s\, u(T-s,x_s)ds \\ &+ \int_0^t K_s\left[\frac{\partial}{\partial s} + \frac{1}{2}\Delta\right]u(T-s,x_s)\,ds + M_t \\ =& u(T,x) + M_t.\end{aligned}$$

where M_t is a local martingale. Taking $t \to T$, by the continuity,

$$K_T\ f(x+B_T) = u(T) + M_t.$$

Since f is bounded, V is bounded from below, M_t is a martingale, we take expectation to conclude. $\square$

Could we relax the condition on the potential function? Yes, indeed. The class of function for which the Feyman-Kac semi-group can be defined extends to Kato-classes.

A non-negative function is in the Kato class if

$$\lim_{t\to 0} \sup_x \mathbb{E} \int_0^t V(x+B_s)ds < \infty.$$

Lemma 4.9.2. *Suppose that V is bounded from below and in L^1_{loc}. Show that fro any $t \geq 0$,*

$$\mathbb{P}\left(\int_0^t V(x+B_s)ds = \infty\right) = 0$$

for almost surely x.

Proof. We can assume that $V \geq 0$. Then for any $R > 0$,

$$\int_{\mathbb{R}^d}\int_0^t V(x+B_s)\mathbf{1}_{\{|x+B_s|\leqslant R\}}dsdx = \int_0^t \int_{\mathbb{R}^d} V(x+B_s)\mathbf{1}_{\{|x+B_s|\leqslant R\}}dxds < \infty.$$

For almost surely all x, $A_R \triangleq \int_0^t V(x+B_s)\mathbf{1}_{\{|x+B_s|\leqslant R\}}ds < \infty$. Let us fix an ω, for which $x+B_s$ is continuous, (set of measure 1), and set

$$R = \max_{s\leqslant t} |x+B_s(\omega)|.$$

Then, since $\sup_{s\leqslant t}|x+B_s| \leqslant R$,

$$\int_0^t V(x+B_s)ds = \int_0^t V(x+B_s)\mathbf{1}_{\{|(x+B_s)|\leqslant R\}}ds < \infty.$$

$\square$

Lemma 4.9.3 (Khamsinski's lemma)**.** *Let $V \geq 0$ be a Borel measurable function. Suppose that for some $t > 0$ and $A < 1$,*

$$\sup_{x\in\mathbb{R}^n} \mathbb{E}\int_0^t V(x+B_s)ds = A,$$

then

$$\sup_{x\in\mathbb{R}^n} \mathbb{E}\left[e^{\int_0^t V(x+B_s)ds}\right] \leqslant \frac{1}{1-A}.$$

Proof. Let

$$F(x,B) = \int_0^t V(x+B_s)ds.$$

Then,

$$\begin{aligned}
\sup_x \mathbb{E}\left[\frac{(\int_0^t V(x+B_s)ds)^n}{n!}\right] &= \frac{1}{n!}\sup_x \mathbb{E}\left[\int_0^t \cdots \int_0^t \Pi_{i=1}^n V(x+B_{s_i})ds_i\right] \\
&= \sup_x \mathbb{E}\left[\int_0^t \cdots \int_{s_1}^t \cdots \int_{s_{n-1}}^t \Pi_{i=1}^n V(x+B_{s_i})ds_n ds_{n-1}\ldots ds_1\right]
\end{aligned}$$

Let

$$F_k(-) = \int_0^t \cdots \int_{s_1}^t \cdots \int_{s_{n-k-1}}^t \Pi_{i=1}^n V(x + B_{s_i}) ds_{n-k} ds_{n-1} \ldots ds_1(-)$$

Them

$$\begin{aligned}\mathbb{E}[F_n] &= \mathbb{E}F_{n-1}\mathbb{E}\left[\int_{s_{n-k-1}}^t V(x+B_{s_n})ds_n|\mathcal{F}_{s_{n-1}}\right] \\ &\leqslant \sup_x \mathbb{E}\left|\int_{s_{n-k-1}}^t V(x+B_{s_n})ds_n\right| \mathbb{E}F_{k-1} \leqslant A^n.\end{aligned}$$

□

Lemma 4.9.4. *Let X_t be the canonical process on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^n)$, $\mathbb{P}_x$ the distribution of the BM from x. Let f, g be Borel measurable such that $\int |f| d\mathbb{P}_x < \infty$ and $A \triangleq \sup_x \int |g| d\mathbb{P}_x < \infty$. Let f be $\mathcal{F}_t$ measurable, then*

$$\mathbb{E}_x|fg \circ \theta_t| \leqslant A\mathbb{E}_x|f|.$$

Proof.

$$\begin{aligned}\mathbb{E}_x[|fg \circ \theta_t|] &= \mathbb{E}_x\mathbb{E}_x[|fg \circ \theta_t| \mid \mathcal{F}_t] \\ &\leqslant E_x\Big[|f|\,\mathbb{E}_x\big[|g \circ \theta_t| \mid \mathcal{F}_t\big]\Big] \\ &= E_x\left[|f|\mathbb{E}_{X_{\theta_t}}\left[|g| \mid \mathcal{F}_t\right]\right] \leqslant A\,\mathbb{E}_x|f|.\end{aligned}$$

□

Theorem 4.9.5. *Suppose that $V = V_+ - V_-$ where $V_+ \in L^1_{loc}(\mathbb{R}^n)$ and V_- is in the Kato class. Then the map*

$$f \mapsto \mathbb{E}\left[f(x + B_t)\int_0^t e^{-V(x+B_r)dr} dr\right]$$

is a bounded operator on L^∞, and a semi-group.

4.10 Essay Topics

(i) Given a diffusion process and its Markov operator (called diffusion operator), a very important question is to determine its invariant measure and the convergence to equilibrium. An important application is the ergodic theorem. There are some literatures for you to explore. You can write an essay on aspects of this theory or/and read digest one of the articles. The emphasize would be to give (as many as possible) concrete examples and work out these examples in detail.

Ergodicity of Diffusion Processes. In the lecture notes, 'Ergodic Properties of Markov Processes by Luc Rey-Bellet' https://people.math.umass.edu/~lr7q/ps_files/EMP.pdf you found the foundational theory.

(a) Ergodicity of Markov processes for stochastic equations on Hilbert spaces can be found in 'On ergodicity of some Markov processes by By Tomasz Komorowski, Szymon Peszat, and Tomasz Szarek.'

Introduction to Ergodic Rates for Markov Chains and Processes by Alexei Kulik https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/7936/file/lpam02.pdf

(b) Ergodic properties of Markov Processes by Martin Hairer, http://www.hairer.org/notes/Markov.pdf

(c) Exponential ergodicity for Markov processes with random switching by Bertrand Cloez, Bernoulli 21(1): 505-536 (February 2015). DOI: 10.3150/13-BEJ577

(d) Exponential and Uniform Ergodicity of Markov Processes by D. Down, S. P. Meyn, R. L. Tweedie Ann. Probab. 23(4): 1671-1691 (October, 1995). DOI: 10.1214/aop/1176987798

(e) A quantitative ergodic theorem for diffusion processes satisfying Hörmander's conditions can be found in Perturbation of Conservation Laws and Averaging on Manifolds by Xue-Mei Li https://arxiv.org/pdf/1705.08857.pdf

(f) Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré Dominique Bakry, Patrick Cattiaux a,c,, Arnaud Guillin https://www.sciencedirect.com/science/article/pii/S0022123607004259?via%3Dihub

(ii) Feynman-Kac formula is a useful tool, it started with Kac, used in mathematical physics called Feynman path integration. A good book in Path Integration by Simon. You can read it also in a more modern book 'Feynman-Kac-Type Formulae and Gibbs Measures' by Lorinczi , Jzsef , Hiroshima, Fumio, and Betz. The following papers are for stochastic processes on manifolds. You can try to work out examples for $\mathbb{R}^n$ with a general diffusion operator. The derivatives of solutions of Feynman-Kac formula are interesting too. You can see the following articles. First order Feynman-Kac formula by Xue-Mei Li and James Thompson. Hessian formulas and estimates for parabolic Schrödinger operators by Xue-Mei Li, in the Journal of Stochastic Analysis, Vol:2, ISSN:2689-6931, https://digitalcommons.lsu.edu/cgi/viewcontent.cgi?article=1080&context=josa

(iii) Stochastic averaging and limit theorems are popular topics. There are multiple approaches. One article is the following. You can read about them and write up your findings.

Averaging for martingale problems and stochastic approximation by T. Kurtz https://www.researchgate.net/publication/226505555_Averaging_for_martingale_problems_and_stochastic_approximation

(iv) Consider a perturbation to the two dimensional Hamiltonian system $\dot{q} = p, \dot{p} = -q$ is studied, it corresponds to the Hamiltonian function $H(p,q) = \frac{1}{2}(p^2 + q^2)$ with skew gradient $J\nabla H = (\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q})$ where $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ gives the standard symplectic structure on $\mathbb{R}^{2n}$. Suppose M is a $2n$- dimensional manifold with a closed non-degenerate two form ω (it belongs to $\wedge^2 T^*M$, it is $dq \wedge dp$ on $\mathbb{R}^2$) Then for every function $H : M \to \mathbb{R}$ with corresponding differential dH, there corresponds to a vector field X_H with $\omega(X_H(x), v)) = dH(v)$ for every $x \in M$ and every tangent vector $v \in T_x M$. This is called the symplectic vector field corresponds to H.

In 'An averaging principle for a completely integrable stochastic Hamiltonian system' by Xue-Mei Li, a perturbation model is given of the form:

$$dx_t = \sum_{i=1}^{n} X_i(x_t) dB_t^i + \varepsilon K(x_t) dt$$

Where $\{H_i : 1 \leqslant i \leqslant n\}$ are Hamiltonian functions with the vector fields $X_i = X_{H_i}$ pairwise commute. Could you work out some examples? For example the structure in the following paper can be connected to the previously mentioned paper. Toward the Fourier law for a weakly interacting anharmonic crystal by Carlangelo Liverani and Stefano Olla. in J. Amer. Math. Soc., 25(2):555-583, 2012.

If $M = \mathbb{R}^{2n}$, with a typical point written as $x = (x_1, \dots x_{2n})$, we identify tangent spaces with $\mathbb{R}^{2n}$ and the cotangent spaces $T_x^* M$ with the dual spaces. Recall df is a differential, so $df(v) = \langle \nabla f(x), v \rangle$ for $x \in T_x \mathbb{R}^{2n} = \mathbb{R}^{2n}$. So at each x, $df \in (\mathbb{R}^{2n})^*$ is a real valued linear map. A differential two form is an antisymmetric map for each x fixed, a typical one if $dx^i \wedge dx^j = \frac{1}{2}(dx^i \otimes dx^j - dx^j \otimes dx^i)$. A differential 3-form is an anti-symmetric mulit-linear the three fold of $\mathbb{R}^{2n}$, e.g. $|x|^2 dx^1 \wedge dx^2 \wedge dx^3$. Note $dx^i \wedge dx^j \wedge dx^k = -dx^j \wedge dx^i \wedge dx^k$.

A differential two-form is of the following form: for each x, we have a linear map $\omega_x : \mathbb{R}^{2n} \otimes \mathbb{R}^{2n} \to \mathbb{R}$ which is anti-symmetric. The subscript x is often omitted. Identify $\mathbb{R}^{2n}$ with its tangent spaces, we write it as follows

$$\omega = \sum_{i,j} f_{ij} dx^i \wedge dx^j$$

with anti-symmetric property, where $f_{i,j}$ are real valued functions. Recall

$$dx^i \wedge dx^j (u, v) = \frac{1}{2} dx^i(u) dx^j(v) - \frac{1}{2} dx^i(v) dx^j(u),$$

in particular $d\omega(u, v) = \frac{1}{2} \sum_{i,j=1}^{2n} f_{ij}(x)(u_i v_j - v_i u_i)$. You want ω to have to property $d\omega(u, v) = 0$ for all v implies $u = 0$ (non-degenerate) and closed ($d\omega = 0$) where

$$d\omega = \sum_{i,j,k} \frac{\partial f_i}{\partial x_k} dx^k \wedge dx^i \wedge dx^j.$$

This imposes constraints on the functions $f_{i,j}$. Recall that $d\omega$ is anti-symmetric, whenever two coordinates are exchanged, the sign is exchanged, this imposes constraints on the functions $f_{i,j}$ as well.. Writing $e_i = \frac{\partial}{\partial x_i}$, You can then solve for

$$\omega(X_H, \cdot) = dH = \sum_i \frac{\partial H}{\partial x_i} dx^i$$

as follows. Write $X_H = \sum u_i e_i$, you want to solve

$$\omega(\sum u_i e_i, \sum v_i e_i) = \langle \nabla H, v \rangle.$$

You can also try to work out examples on compact surfaces.

(v) Markov processes in functions Spaces. Solutions of Stochastic Differential Equation of non-Markovian type. The article : Stochastic Delay Equations by Michael Scheutzow: http://page.math.tu-berlin.de/~scheutzow/Lecture_Notes_SDDEs.pdf studies solution on function space, existence of invariant measure, and convergence to invariant measure problem in this setting.

(vi) Another non-Markov setting turned into Markov setting in function space is as below, which we discussed briefly in the beginning of the lectures. Ergodicity of stochastic differential equations driven by fractional Brownian motion by Martin Hairer: https://www.jstor.org/stable/3481754

(vii) It is also possible to look at the strong completeness property, which implies the existence of a perfect cocycle. A theorem on this is given in [Li94c], section 6 of which is devoted to a treatment on $\mathbb{R}^6$. The results there were new for $\mathbb{R}^n$ at the time, it may still be possible to construct new examples from the main theorem there. Note the treatment there is very much different from the Lipschitz case.

Bibliography

[Arn13] L. Arnold. *Random Dynamical Systems*. Springer Monographs in Mathematics. Springer Berlin Heidelberg, 2013.

[Arv86] W. Arveson. Markov operators and os-positive oeprators. *J. Funct. Anal.*, 66, 1986.

[DPZ96] G. Da Prato and J. Zabczyk. *Ergodicity for infinite-dimensional systems*, volume 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[DS88] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[EL94] K. D. Elworthy and X.-M. Li. Formulae for the derivatives of heat semigroups. *J. Funct. Anal.*, 125(1):252–286, 1994.

[Elw78] K. D. Elworthy. Stochastic dynamical systems and their flows. In *Stochastic analysis (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1978)*, pages 79–95. Academic Press, New York-London, 1978.

[EN00] Klaus-Jochen Engel and Rainer Nagel. *One-parameter semigroups for linear evolution equations*, volume 194 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000. With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt.

[Hai05] Martin Hairer. Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Probab.*, 33(2):703–758, 2005.

[Hai09] Martin Hairer. Ergodic properties of a class of non-Markovian processes. In *Trends in stochastic analysis*, volume 353 of *London Math. Soc. Lecture Note Ser.*, pages 65–98. Cambridge Univ. Press, Cambridge, 2009.

[HM06] Martin Hairer and Jonathan C. Mattingly. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)*, 164(3):993–1032, 2006.

[IW89] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989.

[Kun90] Hiroshi Kunita. *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[LG16] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer, [Cham], french edition, 2016.

[Li92] Xue-Mei Li. Stochastic flows on non-compact manifolds. Ph. D. thesis, 1992.

[Li94a] Xue-Mei Li. Properties at infinity of diffusion semigroups and stochastic flows via weak uniform covers. *Potential Anal.*, 3(4):339–357, 1994.

[Li94b] Xue-Mei Li. Stochastic differential equations on noncompact manifolds: moment stability and its topological consequences. *Probab. Theory Related Fields*, 100(4):417–428, 1994.

[Li94c] Xue-Mei Li. Strong p-completeness of stochastic differential equations and the existence of smooth flows on noncompact manifolds. *Probab. Theory Related Fields*, 100(4):485–511, 1994.

[LS11] Xue-Mei Li and Michael Scheutzow. Lack of strong completeness for stochastic flows. *Ann. Probab.*, 39(4):1407–1421, 2011.

[MR91] Zhiming Ma and Michael Röckner. *Introduction to the theory of (non-symmetric) Dirichlet forms*. Springer, 1991.

[Par67] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press, Inc., New York-London, 1967.

[Paz83] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[RS72] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I. Functional analysis*. Academic Press, New York-London, 1972.

[RW00] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Foundations, Reprint of the second (1994) edition.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[Sei01] J. Seidler. A note on the strong feller property. Unpublished lecture notes, 2001.

[SS05] Elias M. Stein and Rami Shakarchi. *Real analysis*, volume 3 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2005. Measure theory, integration, and Hilbert spaces.

[Ver05] A. Vershic. What does a generic markov operator look like. arXiv:math/0510320, 2005.