Foundations of Multi-Scale Stochastic Systems

Topics in Stochastic Analysis

Xue-Mei Li
EPFL

December 17, 2024

Contents

Chapter 1

Introduction

1.1 Introduction

A multi-scale stochastic system models a set of particles or objects of interest that evolve at different scales, are subject to randomness, and interact with each other. Of particular interest are two-scale stochastic differential equations, where a slow variable evolves in a rapidly changing random environment. For example, consider the equation:

$$dx_t^\varepsilon = g(x_t^\varepsilon, \epsilon, y_t^\epsilon)dt + \sum_{k=1}^{m} f_k(x_t^\varepsilon, \epsilon, y_t^\epsilon)dW_t^k,$$

where W_t^k are independent Brownian motions, x-variable takes value in a space $\mathcal{X}$, and the y-variable takes values in a space $\mathcal{Y}$. The parameter ϵ serves as an indicator for the separation of the time scales. In this setting, the slow variable is x_t^ϵ evolving in its natural scale; while y_t^ϵ evolves on a faster time scale, specifically $\frac{1}{\epsilon}$. It is often the case $y_t^\epsilon = \tilde{y}_{t/\epsilon}^\epsilon$, while both $\tilde{y}_t^\epsilon$ and x_t^ϵ are two stochastic processes evolving on the same time scale. Slow/ fast stochastic dynamics, with two scales, holds significant potential for applications.

Multi-scale systems are often observed in classical mechanics, particularly in the case of small perturbations to Hamiltonian systems, or more generally, in perturbations to dynamical systems with conservation laws. A key goal is to study the effect of these perturbations on the evolution of the Hamiltonian or energy along the system's trajectories. Multi-scale behaviour often emerges in the form of an ϵ-expansion of the perturbation.

Multi-scale dynamics are also observed in neuron response dynamics. The effect of a spike on a neuron's membrane potential can be quantified by the difference between the interior of the cell and its surroundings. Depending on the sign of this change, the effect can be excitatory or inhibitory. After the spike arrives, the neuron's potential returns to its resting state.

In a generic two-dimensional neuron evolution model [5], the equations are as follows:

$$\dot{u}(t) = f(u, w) + I, \qquad \dot{w}(t) = \epsilon b(u, w),$$

where I is the current that does not directly affect the spike w, but a change in potential leads to a small movement in the spike. A spike induces a significant movement in neuron potential. The firing of a spike can seem random, and the second equation can be replaced with a diffusion model. This model can be further simplified to a single equation as the separation scale parameter ϵ tends to zero.

In nature, multi-time scale phenomena are widespread. Klaus Hasselmann, a German oceanographer and climate modeller, was awarded the Nobel Prize in Physics for his groundbreaking work on climate science. He proposed a dynamical system that describes the interaction between climate and weather, where climate evolves at a slower pace compared to the rapidly changing weather. Quoting from the Nobel Prize webpage:

"Our world is full of complex systems characterised by randomness and disorder. One complex system of vital importance to humankind is Earth's climate. In the 1970s, Klaus Hasselmann created a model that links together weather and climate, thus answering the question of why climate models can be reliable despite weather being changeable and chaotic."

Randomness, whether perceived or intrinsic, must be accounted for. Multiple time-scale stochastic ordinary differential equations and stochastic partial differential equations have also emerged as promising tools in biological and social sciences [12]. For example, in biomedical research, a patient's observables can be viewed as a stochastic system interacting with drug treatments that operate at the cellular and molecular levels.

The model of physical Brownian motion developed by Smoluchowski, Langevin, Ornstein, and Uhlenbeck can be transformed into a two-scale system. Consider the position of a particle of small mass ϵ, where the velocity field is governed by the Ornstein-Uhlenbeck process:

$$\begin{aligned} \dot{x}_t^\epsilon &= y_t^\epsilon \\ dy_t^\epsilon &= -\frac{1}{\epsilon} y_t^\epsilon + \frac{1}{\epsilon} dW_t. \end{aligned}$$

In a more complex scenario y_t^ϵ receives a feedback from the x-variable. For example, introducing friction into the model yields the equation [13]:

$$dy_t^\epsilon = f(x_t^\epsilon) - y_t^\epsilon + dW_t.$$

1.1.1 Some slow / fast models

We point out some slow /fast stochastis differential equations.

1. *Stochastic Averaging of SDEs and SPDEs.* Consider a stochastic averaging problem of SDEs:

   $$dx_t^\varepsilon = g(x_t^\varepsilon, y_{\frac{t}{\epsilon}}^\epsilon)dt + \sum_{k=1}^m f_k(x_t^\varepsilon, y_{\frac{t}{\epsilon}}^\epsilon)dW_t^k,$$

   where $\{W_t^k\}$ are independent Brownian motions.

   Assume for simplicity $m = 1$, and x_t^ϵ and y_t^ϵ are real valued stochastic processes for which the following equation

   $$dx_t^\varepsilon = f(x_t^\varepsilon, y_{\frac{t}{\epsilon}}^\epsilon)dW_t, \qquad x_t^\epsilon = x_0$$

   are satisfied. Then the quadratic variation of the solution satisfies:

   $$\langle x^\epsilon \rangle_t = \int_0^t f^2(x_s^\varepsilon, y_{\frac{s}{\epsilon}}^\epsilon)ds.$$

   This naiive computation indeed pointing a way for obtaining the effective dynamics, by martingale problem.

2. The averaging principle is a law of large numbers, from here we may study fluctuations, e.g.

   $$\frac{x_t^\epsilon - \bar{x}_t}{\sqrt{\epsilon}}.$$

   Consider for example

   $$\dot{x}_t^\epsilon = \frac{1}{\sqrt{\epsilon}} g(x_t^\epsilon, y_{\frac{t}{\epsilon}}).$$

3. *Stochastic Averaging for Non-Markovian system.*

   Take for example SDEs driven by fractional Brownian motions B_t^H:

   $$dx_t^\varepsilon = g(x_t^\varepsilon, y_{\frac{t}{\epsilon}}^\epsilon)dt + \sum_{k=1}^m f(x_t^\varepsilon, y_{\frac{t}{\epsilon}}^\epsilon)dB_t^H.$$

1.2 An example of an averaging principle – Week 1

In highly oscillatory dynamical systems, averaging techniques can be used to simplify the analysis. The resulting averaged system approximates the original dynamics when there is a large separation of time scales.

Let $\mathcal{Y}$ denote a metric space with distance function d, and let $g : \mathbf{R}^d \times \mathcal{Y} \to \mathbf{R}^d$ be a Borel measurable function. Consider a system with a constant of motion and a small perturbation:

$$\dot{u}_t^\epsilon = \epsilon g(u_t^\epsilon, y_t),$$

where ϵ is a small parameter indicating the magnitude of the perturbation.

By rescaling time as $t \mapsto t/\epsilon$, we obtain the following rescaled problem:

$$\dot{x}_t^\epsilon = g(x_t^\epsilon, y_{\frac{t}{\epsilon}}).$$

In this setting, ϵ controls the separation of time scales. As $\epsilon \to 0$, the term $y_{\frac{t}{\epsilon}}$ evolves on a much faster time scale compared to x_t^ϵ.

We typically assume that y_t is ergodic. This means that the time-average of the vector field $g(x, y_t)$, as y_t evolves with time, converges in probability to a deterministic vector field $\bar{g}$ on $\mathbf{R}^d$, for every x in $\mathbf{R}^d$. Tis ergodicity ensures that, over long time intervals, on the time scale $[0, \frac{1}{\varepsilon}]$, the fast y_t-dynamics average out, leaving a persistent influence on the slower x_t^ε-dynamics.

The following defines the weak ergodic condition.

Definition 1.2.1 We say that the pair, g and (y_t^ϵ), satisfies the **weak ergodic condition**, if for any x, there exists a point $\bar{g}(x) \in \mathbf{R}^d$ such that for any $\delta > 0$ and any $0 \leqslant s \leqslant t$,

$$\lim_{\epsilon \to 0} \mathbb{P}(\omega : |\frac{1}{t-s} \int_s^t g(x, y_{\frac{r}{\epsilon}}(\omega))dr - \bar{g}(x)| > \delta) = 0. \tag{1.1}$$

This condition essentially indicates that, over time, the process y_t loses memory of its initial state, and the dynamics governed by $g(x, y_t)$ become predictable in the long run, with fluctuations around $\bar{g}(x)$ vanishing as $T \to \infty$.

In fact, in place of (1.2), we may assume i.e. for any $\delta > 0$ and any $T > 0$, and *uniformly* in $t > 0$,

$$\lim_{T \to \infty} \mathbb{P}(\omega : |\frac{1}{T} \int_t^{t+T} g(x, y_s^\epsilon(\omega))ds - \bar{g}(x)| > \delta) = 0. \tag{1.2}$$

We further assume that g is Lipschitz continuous. Specifically, there exists a constant $|g|_{\mathrm{Lip}}$ such that for all $x, x' \in \mathbf{R}^d$, and $y, y' \in \mathcal{Y}$,

$$|g(x,y) - g(x' - y')| \leqslant |g|_{\mathrm{Lip}}(|x - x'| + d(y, y')). \tag{1.3}$$

Lemma 1.2.2 *Suppose that g is bounded and Lipschitz continuous, and $g(y_t^\epsilon)$ satisfies (1.2). Consider:*

$$\frac{d}{dt}\bar{x}_t = \bar{g}(\bar{x}_t), \qquad \bar{x}_0 = \bar{x}(0).$$

then

$$\sup_{s\in[0,t]} \left| \int_0^s (g(\bar{x}_r, y_{\frac{r}{\epsilon}}) - \bar{g}(\bar{x}_r))dr \right| \to 0$$

in probability.

Proof Observing that $\bar{g}$ is bounded, Lipschitz continuous, and $|\bar{x}_t - \bar{x}_s| \leqslant |g|_{\mathrm{Lip}}|t - s|$. Write $y_r^\epsilon = y_{\frac{r}{\epsilon}}$. Let $0 \leqslant s_1 < \dots, s_n$ denotes a uniform partition of $[0, t]$.

$$\int_0^s g(\bar{x}_r, y_r^\epsilon)) - \bar{g}(\bar{x}_r))dr = \sum_i \int_{s_i}^{s_{i+1}} (g(\bar{x}_r, y_r^\epsilon)) - \bar{g}(\bar{x}_r))dr$$
$$= \sum_i \int_{s_i}^{s_{i+1}} dr[(g(\bar{x}_r, y_r^\epsilon)) - g(\bar{x}_{s_i}, y_r^\epsilon))) + (g(\bar{x}_{s_i}, y_r^\epsilon)) - \bar{g}(\bar{x}_{s_i})) + (\bar{g}(\bar{x}_{s_i}) - \bar{g}(x_r))].$$

Note that

$$|\sum_i \int_{s_i}^{s_{i+1}} (g(\bar{x}_r, y_r^\epsilon)) - g(\bar{x}_{s_i}, y_r^\epsilon)))dr| \leqslant \sum_i \int_{s_i}^{s_{i+1}} K|\bar{x}_r - \bar{x}_{s_i}|dr \leqslant \sum_i (\Delta s_i)^2 \lesssim \frac{1}{n}.$$

which converges to zero as n is taken to infinity. A similar computation applies to the third term. Now fixing n, as $\epsilon \to 0$,

$$\sum_i \int_{s_i}^{s_{i+1}} (g(\bar{x}_{s_i}, y_r^\epsilon)) - \bar{g}(\bar{x}_{s_i}))dr \to 0$$

in probability. Consequently,

$$|\sum_i \int_{s_i}^{s_{i+1}} (g(\bar{x}_r, y_r^\epsilon)) - g(\bar{x}_{s_i}, y_r^\epsilon)))dr| \lesssim \frac{1}{n} + \sum_i \int_{s_i}^{s_{i+1}} (g(\bar{x}_{s_i}, y_{\frac{r}{\epsilon}}) - \bar{g}(\bar{x}_{s_i}))dr \to 0,$$

as we take $\epsilon \to 0$ then take $n \to \infty$. □

To work with the convergence of solutions, iterative method or maximal principle lead to a large collection of very useful inequality. They go by the name of Gronwall inequality, referring to Gronwall's work published in 1919 [6].

Lemma 1.2.3 (Gronwall Inequality) *Let u and β be non-negative, continuous functions on $[a, b]$, for which the inequality*

$$u(t) \leqslant C + \int_0^t \beta(s)u(s)ds, \qquad a \leqslant t \leqslant b$$

holds, where C is a non-negative constant, then

$$u(t) \leqslant C \exp(\int_a^t \beta(s)ds), \qquad a \leqslant t \leqslant b.$$

Lemma 1.2.4 *[9] Let u, α, and β be piecewise continuous functions on $[0, T]$, and β is non-negative on this interval. If*

$$u(t) \leqslant \alpha(t) + \int_0^t \beta(s)u(s)ds, \qquad 0 \leqslant t \leqslant T$$

then

$$u(t) \leqslant \alpha(t) + \int_0^t \alpha(s)\beta(s) \exp(\int_s^t \beta(r)dr)ds, \qquad 0 \leqslant t \leqslant T.$$

See also [1].

Proposition 1.2.5 *Assume that g is bounded and Lipschitz continuous, satisfying (1.2). Let $x(0, \epsilon)$ be $\mathbf{R}^d$ valued random variables converging to $\bar{x}(0)$ in probability. Consider the solution of the equation*

$$\dot{x}_t^\epsilon = g(x_t^\epsilon, y_t^\epsilon), \qquad x_0^\epsilon = x(0, \epsilon), \tag{1.4}$$

Then, for any $\delta > 0$,

$$\lim_{\epsilon \to 0} \mathbb{P}(\sup_{s \in [0,T]} |x_s^\epsilon - \bar{x}_s| > \delta) = 0,$$

where $\frac{d}{dt}\bar{x}_t = \bar{g}(\bar{x}_t)$ and $\bar{x}_0 = \bar{x}(0)$.

Proof This is simply a corollary of Lemma 1.2.2. Firstly,

$$x_t^\epsilon - \bar{x}_t = x(0, \epsilon) - \bar{x}(0) + \int_0^t (g(x_r^\epsilon, y_r^\epsilon) - g(\bar{x}_r, y_r^\epsilon))dr + \int_0^t (g(\bar{x}_r, y_r^\epsilon) - \bar{g}(\bar{x}_r)))dr$$

Then,

$$\sup_{s \leqslant t} |x_s^\epsilon - \bar{x}_s| \leqslant |x(0, \epsilon) - \bar{x}(0)| + K \int_0^t \sup_{s \leqslant r} |x_s^\epsilon - \bar{x}_s| dr + | \int_0^t (g(\bar{x}_r, y_r^\epsilon) - \bar{b}(\bar{x}_r)))dr|.$$

Set

$$\alpha_\epsilon(t) = |x(0, \epsilon) - \bar{x}(0)| + \sup_{s \in [0,t]} | \int_0^s (g(\bar{x}_r, y_r^\epsilon) - \bar{g}(\bar{x}_r)))dr|.$$

By Grownall's inequality,

$$\sup_{s\leqslant t}|x_s^\epsilon - \bar{x}_s| \leqslant e^{Kt}\Big(|x(0,\epsilon)-\bar{x}(0)| + \sup_{s\in[0,t]}|\int_0^s (g(\bar{x}_r, y_r^\epsilon) - \bar{g}(\bar{x}_r))dr\Big).$$

This complete the proof. □

1.3 Notations

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

Definition 1.3.1 Let $p \geqslant 1$.

1. A family of Borel measurable functions $\{f_\alpha\}$ on a measure space is L^p bounded if $\sup_\alpha \int |f_\alpha|^p < \infty$.
2. A stochastic process (X_t) is L^p integrable if $\mathbb{E}(|X_t|^p) < \infty$ for all t; it is L^p bounded if $\sup_t \mathbb{E}(|X_t|^p) < \infty$.

Chapter 2

Basics on Stochastic Processes

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, the state space for random variables are assumed to be a connected metric space, satisfying the complete and separable assumptions, and endowed with its Borel σ-algebra.

Let $\mathcal{X}$ be a metric space and $\mathcal{B}(\mathcal{X})$ its Borel σ-algebra. Elements of $\mathcal{B}(\mathcal{X})$ are referred as the Borel subsets of $\mathcal{X}$. A measure on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ are referred as a Borel measure. We consider only measures that assigns a finite number to every metric ball. We denote by $P(\mathcal{X})$ the space of probability measures on $\mathcal{X}$.

Then the following holds:

- **regular:** for each Borel subset A and for each $\epsilon > 0$ there exists an open set U and a closed set C such that $C \subset A \subset U$ and $\mu(U - C) < \varepsilon$.
- **tight:** If $\mu(\mathcal{X}) < \infty$, then for any $\epsilon > 0$ there exists a compact subset $K \subset \mathcal{X}$ such that $\mu(K) > 1 - \epsilon$

2.1 Distributions of stochastic processes

Let $\mathcal{X}$ be a metric space. A random variable on $\mathcal{X}$ is simply a Borel measurable function from Ω to $\mathcal{X}$; a stochastic processes (X_t) on $\mathcal{X}$ is a collection of random variables parametrized by an index set $I \subset \mathbf{R}$. As is customary, we denote the time-variable of stochastic processes with a subscript and we often omit the round bracket for notational simplicity.

If $\mathcal{X}_\alpha$, where $\alpha \in I$, are metric spaces, the product space $\Pi_{\alpha \in I} \mathcal{X}_\alpha$ tis equipped with the product σ-algebra such that each coordinate map: $\pi_\alpha : \Pi_\alpha \mathcal{X}_\alpha \to \mathcal{X}_\alpha$ is measurable.

It is generated sets of the form $\Pi_{\alpha \in I} A_i$ where $A_i \in \mathcal{B}(\mathcal{X}_\alpha)$, with only finitely many A_α not equal to the whole space. Such sets are referred to as cylindrical sets.

A stochastic process X_t is a measurable map from $\Omega \to \Pi_\alpha \mathcal{X}$, its probability distribution is the pushforward measure $(X_\cdot)_* \mathbb{P}$ on $\Pi_\alpha \mathcal{X}$. The distribution of a stochastic process are determined by their finite dimensional distributions.

Definition 2.1.1 Let $(X_t, t \in I)$ be a stochastic process. The finite-dimensional distribution of the process at time $t_1, \dots t_n$, where $t_1 < t_2 < \cdots < t_n$, $t_i \in I$, is the joint distribution of $(X_{t_1}, \dots, X_{t_n})$:

$$\mu_{t_1,\dots,t_n}(\Pi_{i=1}^n A_i) = \mathbb{P}(\{\omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_n}(\omega) \in A_n\},$$

where A_i are Borel measurable subsets of $\mathcal{X}$. The collection of such probability measures is called the finite dimensional distributions of (X_t).

We denote $(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n) := \{\omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_n}(\omega) \in A_n\}$.

Definition 2.1.2 For any $t_1 < \cdots < t_n$, there is a projection $\pi_{t_1,\dots t_n} \mathcal{X}^I \to \mathcal{X}^n$ defined by

$$\pi_{t_1,\dots t_n}(\sigma) = (\sigma(t_1), \dots, \sigma(t_n)).$$

The following theorem guarantees the existence of a stochastic process with given finite-dimensional distributions.

Theorem 2.1.3 (Kolmogorov's extension theorem) *Suppose that a family of probability measures $\{\mu_{t_1,\dots,t_n}\}$ are given, where n runs through all times points $t_1 < t_2 < \dots, t_n$, $t_i \in I$, satisfying the following consistency conditions. For any $n \in \mathbf{N}$, any $t_1 < \cdots < t_{n+1}$, $t_i \in I$, and for any $A_i \in \mathcal{B}(\mathcal{X})$, the following statements hold:*

(1)

$$\mu_{t_1,\dots,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathcal{X}) = \mu_{t_1,\dots,t_n}(\Pi_{i=1}^n A_i),$$

(2) for any permutation σ of $\{1, \dots, n\}$,

$$\mu_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = \mu_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}).$$

Then there exists a measure on $\mathcal{X}^I$ such that its finite dimensional projections agreeing with $\mu_{t_1,\dots,t_n}$. Consequently, there exists a stochastic process (X_t) on some probability space such that

$$\mu_{t_1,\dots,t_n}(\Pi_{i=1}^n A_i) = \mathbb{P}(\{\omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_n}(\omega) \in A_n\}).$$

A stochastic process is easy to analyze, if any family of random variables $\{X_{t_1}, \dots X_{t_n}\}$ are independent, in which case $\mu_{t_1,\dots,t_n} = \Pi_{i=1}^n \mu_{t_i}$. For stochastic integrals, the simplest are those whose increments are independent.

Definition 2.1.4 A stochastic process $(Y_t, t \geqslant 0)$ is stationary if its finite dimensional distributions are invariant under translations, i.e. for any $s > 0$,

$$(Y_{t_1+s}, \dots Y_{t_m+s}) \stackrel{law}{=} (Y_{t_1}, \dots Y_{t_m}).$$

Exercise 2.1.5 Construct an example of a stochastic process for which the probability distribution of its one time marginals are the same for all time, but it is not stationary.

2.2 Measure Separating sets

Definition 2.2.1 A collection E of measurable functions is said to be *measure separating* (or *measure determining*) if, for any two probability measures μ and ν on $\mathcal{X}$,

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

The set of functions $\mathcal{B}_b(\mathcal{X})$ is clearly measure determining, however it is often to be too large to be of any use. In fact, $\mathcal{B}_b(\mathcal{X})$ is not separable. It is sufficient to test these on the set of uniformly continuous functions.

If $f : \mathcal{X} \to \mathbf{R}$ we denote f^+ denotes its positive part: $f^+(x) = \max(f(x), 0)$.

Theorem 2.2.2 *[14, Thm. 5.9, pp39] Let $\mathcal{X}$ be a metric space, and let μ, ν be two probability measures on $\mathcal{X}$. If*

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

for every bounded and uniformly continuous function $f : \mathcal{X} \to \mathbf{R}$, then $\mu = \nu$.

Proof Since probability measures on a metric space are regular, they are determined by their values on closed sets. It is sufficient to show that $\mu(C) = \nu(C)$ for any closed set C. Let C be a closed set, define a sequence of functions $f_n : \mathcal{X} \to \mathbf{R}$ by

$$f_n(x) = (1 - n\, d(x, C))^+$$

where $d(x, C) = \inf_{y \in C} d(x, y)$. We have $f_n(x) = 0$ for $x \in C$. Since$d(x, C) > 0$ for $x \notin C$, $f_n(x) \to 0$ as $n \to \infty$. Therefore $f_n(x) \geqslant \mathbf{1}_C$ and each f_n is Lipschitz continuous:

$$|f_n(x) - f_n(z)| \leqslant n d(x, z).$$

For each n,

$$\int_{\mathcal{X}} f_n d\mu = \int_{\mathcal{X}} f_n d\nu.$$

Taking limit $n \to \infty$ and noting that $|f_n| \leqslant 1$, we can interchange the limit and the integrals (by the Dominated Convergence Theorem). Thus $\mu(C) = \int f d\mu = \int f d\nu = \nu(C)$ for all closed sets, concluding that the two measures are the same. □

Definition 2.2.3 Let $\mathcal{X}$ be a metric space. A set of functions is said to separate points if for any $x \neq y$ in M, there exists a function f such that $f(x) \neq f(y)$.

The collection of functions of the form $\{d(x, \cdot) \wedge 1 : x \in \mathcal{X}\}$ are bounded continuous functions, and separate points. Pointing separating is a fairly weak property, for example the set of linear functions on $\mathbf{R}^n$ separate point.

A subset M of $\mathbf{C}(\mathcal{X}; \mathbf{R})$ is an algebra if it is a vector space and $fg \in M$ whenever $f \in M$ and $g \in M$. The following theorem holds for any compact Hausdorff space.

Theorem 2.2.4 (The Stone-Weierstrass Theorem.) *Let $\mathcal{X}$ be a compact metric space and let $\mathcal{A}$ be a closed sub-algebra of $\mathbf{C}(\mathcal{X}, \mathbf{R})$ that contains the constant functions and separates points, then $M = \mathbf{C}(\mathcal{X}; \mathbf{R})$.*

With Theorem 2.2.2, we prove the following theorem, taken from Theorem 4.5 in Chapter 3 [3, pp.113].

Theorem 2.2.5 *Let $(\mathcal{X}, d)$ be a complete separable metric space. Let M be a sub-algebra of bounded continuous functions separating points, then M is measure separating.*

Proof Let μ and ν be two probability measures on $\mathcal{X}$ such that $\int g d\mu = \int g d\nu$ for all $g \in M$. Define $H = \{f + a : f \in M, a \in \mathbf{R}\}$. Then H is a sub-algebra of $\mathbf{C}_b(\mathcal{X})$ containing the constants and is point separating. Since $\mu(\mathcal{X}) = 1 = \nu(\mathcal{X})$, for any $h \in H$,

$$\int_{\mathcal{X}} h d\mu = \int_{\mathcal{X}} h d\nu.$$

If $f_n \in H$ converges to f in the uniform norm, then it follows that $\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f d\nu$. If $\mathcal{X}$ is compact, by the Stone-Weierstrass Theorem, the closure of H equals $\mathbf{C}(\mathcal{X}; \mathbf{R})$, and the theorem is proved.

Otherwise, we use Theorem 2.2.2 to show $\int f d\mu = \int f d\nu$ for any bounded uniformly continuous function f.

By tightness, for any $\epsilon > 0$, there exists a compact set $K \subset \mathcal{X}$such that $\mu(K_\varepsilon) > 1-\varepsilon$ and $\nu(K_\varepsilon) > 1-\varepsilon$. By the Stone-Wiererstrass theorem, there exists $g_n \in H$ such that g_n that approximates g uniformly on K_ε:

$$\lim_{n\to\infty} \sup_{x\in K_\varepsilon} |g_n(x) - g(x)| = 0. \tag{2.1}$$

Note that g_n depends on ϵ.

Consider the function $\varphi_\epsilon(x) = xe^{-\epsilon x^2}$. As $\epsilon \to 0$, we have $\varphi_\epsilon(x) \to x$. Since $|\varphi_\epsilon(g)|_\infty \leqslant |g|_\infty$ for any bounded function g, by the Dominated Convergence Theorem, we obtain:

$$|\int_{\mathcal{X}} g d\mu - \int_{\mathcal{X}} g d\nu| = \lim_{\epsilon\to 0} |\int_{\mathcal{X}} \varphi_\varepsilon(g) d\mu - \int_{\mathcal{X}} \varphi_\varepsilon(g) d\nu|.$$

On the other hand,

$$\begin{aligned}|\int_{\mathcal{X}} \varphi_\varepsilon(g)d\mu - \int_{\mathcal{X}} \varphi_\varepsilon(g)d\nu| \leqslant \Big|\int_{\mathcal{X}} (\varphi_\varepsilon(g) - \varphi_\varepsilon(g_n))\, d\mu\Big| + |\int_{\mathcal{X}} \varphi_\varepsilon(g_n)d\mu - \int_{\mathcal{X}} \varphi_\varepsilon(g_n)d\nu| \\ + \Big|\int_{\mathcal{X}} (\varphi_\varepsilon(g_n) - \varphi_\varepsilon(g))d\nu\Big|.\end{aligned}$$

Observe that $|\varphi_\varepsilon|_\infty \leqslant \frac{C}{\sqrt{\epsilon}}$ where $C = \sup_x xe^{-x^2}$. Since $\mu(\mathcal{X}\backslash K_\epsilon) \leqslant \varepsilon$, we can write

$$\begin{aligned}&\limsup_{n\to\infty}\Big|\int_{\mathcal{X}} (\varphi_\varepsilon(g) - \varphi_\varepsilon(g_n))\, d\mu\Big| \\ &\leqslant \limsup_{n\to\infty}\Big|\int_{K_\epsilon} (\varphi_\varepsilon(g) - \varphi_\epsilon(g_n))\, d\mu\Big| + \Big|\int_{\mathcal{X}\backslash K_\varepsilon} (\varphi_\varepsilon(g) - \varphi_\varepsilon(g_n))\, d\mu\Big|.\end{aligned}$$

This leads to

$$\limsup_{n\to\infty}\Big|\int_{\mathcal{X}} (\varphi_\varepsilon(g) - \varphi_\varepsilon(g_n))\, d\mu\Big| \leqslant \mu((K_\varepsilon)^c)\frac{2C}{\sqrt{\epsilon}} = 2C\sqrt{\epsilon}.$$

For the first term, we applied the result in (2.1) and used the Dominated Convergence Theorem. By applying the same argument to ν, we obtain

$$\limsup_{n\to\infty}\Big|\int_{\mathcal{X}} (\varphi_\varepsilon(g) - \varphi_\varepsilon(g_n))\, d\nu\Big| \leqslant 2C\sqrt{\epsilon}.$$

It remains to show that

$$\int_{\mathcal{X}} \varphi_\varepsilon(g_n)d\mu = \int_{\mathcal{X}} \varphi_\varepsilon(g_n)d\nu,$$

for which it is sufficient to demonstrate that $\varphi_\varepsilon(g_n)$ belongs to the closure of H.

Let $P_{m,\epsilon}$ denote the Taylor expansion of $\varphi_\epsilon : \mathbf{R} \to \mathbf{R}$, a smooth function, up to order m. Using the property that $g_n \in H$ and that H is an algebra, we conclude that

$P_{m,\varepsilon} \circ g_n \in H$. Taylor expansions of smooth functions converges uniformly on any bounded set K:

$$\lim_{m\to\infty} \sup_{x\in K} |P_{m,\varepsilon}(x) - \varphi_\varepsilon(x)| = 0.$$

Moreover, fixing ϵ, since g_n and g are uniformly bounded, we have $P_{m,\varepsilon} \circ g_n \to \varphi_\varepsilon \circ g_n$ uniformly on $\mathcal{X}$ as m approaches infinity. Therefore $\varphi_\varepsilon(g_n) \in \bar{H}$, the closure of H in the uniform topology.

From this, we deduce:

$$\int_{\mathcal{X}} \varphi_\varepsilon(g_n) d\mu = \int_{\mathcal{X}} \varphi_\varepsilon(g_n) d\nu,$$

allowing to conclude that $\int_{\mathcal{X}} g \; d\mu = \int_{\mathcal{X}} g d\mu$ for any $g \in \mathbf{C}_b(\mathcal{X})$, and hence $\mu = \nu$ by Theorem 2.2.2. □

Therefore on $\mathbf{R}^d$, the space of continuous functions with compact support is measure separating, and so are the space of smooth functions on compact supports. This can be verified with Urysohn's lemma and smooth Urysohn's lemma.

Let $\mathbf{C}_K(\mathcal{X})$ denote the space of continuous functions on $\mathcal{X}$ with compact support. Let $\mathbf{C}_0(\mathcal{X})$ denote the space of real-valued continuous functions on $\mathcal{X}$ that vanishes at infinity, which means that for any $\epsilon > 0$ there exists a compact set K such that $|f(x)| < \epsilon$.

A proof for the following statements can be found in [4, pp132,245]:

Proposition 2.2.6 *If $\mathcal{X}$ is a locally compact separable metric space, then $\mathbf{C}_0(\mathcal{X})$ is the closure of $\mathbf{C}_K(\mathcal{X})$ in the uniform metric.*

It follows that if $\mathcal{X}$ is a locally compact separable metric space, then $\mathbf{C}_0(\mathcal{X})$ is an algebra of $\mathbf{C}(\mathcal{X}; \mathbf{R})$ and is measure separating.

Proposition 2.2.7 *The space $\mathbf{C}_K^\infty$ is dense in $\mathbf{C}_0(R^n)$, and is therefore measure separating.*

If $f \in C_0(\mathbf{R}^n)$, then $\varphi_\epsilon \in \mathbf{C}_K^\infty$ for any φ positive, smooth with compact support with $\varphi_\epsilon(x) = \epsilon^{-n}\varphi(\frac{x}{\epsilon})$, normalised so that $\|\varphi\|_1 = 1$. Then if f is uniformly continuous and bounded, $f * \varphi_\epsilon \to \varphi$ uniformly.

For probability measures on $\mathbf{R}^n$ with bounded supports, the set of polynomials are measure separating. If the restrictions of two probability measures on $\mathbf{R}^n$ on balls agree, they are the same. These tao statement do not imply that polynomials are measure determining on $\mathbf{R}^n$. If the sequence of moments of a random variable grows too fast, they do not determine the distribution. Therefore the set of all polynomials are not measure separating on $\mathbf{R}^n$. Note that the integrals of polynomials of two measures agree do not imply that the integrals of polynomials of two measures restricted to balls equal.

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 2.2.8 *[3, Thm 4.6, pp115] Let $\mathcal{X}_i$ be complete separable metric spaces, and $E_k \subset \mathbf{C}_b(\mathcal{X}_k)$ is measure separating. Then*

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

is measure separating on $\Pi_{i=1}^{\infty}\mathcal{X}_i$.

2.3 Convergence determining set

Let $\mathbb{P}(\mathcal{X})$ denote the set of all probability measures on a metric space $\mathcal{X}$.

Definition 2.3.1 Let $\mu_n, \mu \in \mathbb{P}(\mathcal{X})$. We say that $\mu \to \mu$ weakly, if

$$\lim_{n\to\infty} f d\mu_n = \int f d\mu,$$

for every real-valued, bounded, and continuous function f on $\mathcal{X}$.

Let $x_n, x \in \mathcal{X}$. The sequence δ_{x_n} converges weakly if and only if $x_n \to x$.

Proposition 2.3.2 (Portmanteau Theorem) *The following statements are equivalent:*

(1) μ converges to μ weakly,

(2) $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$, for every real-valued, bounded and uniformly continuous function f on $\mathcal{X}$.

(3) $\limsup_{n\to\infty} \mu_n(F) \leqslant \mu(F)$ for all closed set F.

(4) $\liminf_{n\to\infty} \mu_n(G) \geqslant \mu(G)$ for all open set G.

Proposition 2.3.3 *A sequence $\mu_n \in \mathbb{P}(\mathcal{X})$ converges to $\mu \in \mathbb{P}(\mathcal{X})$ if and only if every subsequence of μ_n has a further subsequence that converges to μ weakly.*

Proof The only if part is clear. We show the converse. If μ_n does not converge to μ weakly, then there exists $f \in \mathbf{C}_b(\mathcal{X})$ such that $a_n := \int f d\mu_n$ does not converge to $\int f d\mu$. Consequently, for some $\epsilon > 0$, there exists a subsequence a_{n_k} with $|a_{n_k} - \int f d\mu| > \epsilon$, which contradicts the assumption. □

Definition 2.3.4 A family A of probability measures is said to be **relatively compact** if every sequence from A contains a weakly convergence subsequence.

We give a 'compactness' theorem that provides us with a very useful criteria to check whether a given sequence of probability measures has a convergent subsequence. In order to state this criteria, let us first introduce the notion of 'tightness'.

By tightness we mean that the measure is tightly packed into a small space, by 'small' we we mean the total mass can be almost packed into a compact set.

Definition 2.3.5 Let $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ be an arbitrary subset of the set of probability measures on some topological space $\mathcal{X}$. We say that $\mathcal{M}$ is (uniformly) **tight** if, for every $\varepsilon > 0$ there exists a compact set $K \subset \mathcal{X}$ such that $\mu(\mathcal{X} \backslash K) < \epsilon$ for every $\mu \in \mathcal{M}$.

By Lemma 9.2.5, every finite family of probability measures on a complete separable metric space is tight. One can show that: if $\{\mu_n\}$ is a tight sequence of probability measures on a complete separable metric space, then there exists a probability measure μ on $\mathcal{X}$ and a subsequence μ_{n_k} such that $\mu_{n_k} \to \mu$ weakly.

If $\{X_n\}$ is a a sequence of random variables with $\sup_n \mathbb{E}|X_n| < \infty$, then $\{|X_n|\}$ is tight. This is due to Markov-Chebychev inequality

$$\mathbb{P}(|X_n| > a) \leqslant \frac{1}{a}\mathbb{E}|X_n| \to 0$$

as $a \to \infty$.

Example 2.3.6 Let M be a subset of $\mathcal{P}(\mathbf{R})$. Suppose that there exists a non-decreasing function $\varphi : [0, \infty) \to [0, \infty)$ such that $\lim_{x\to\infty} \varphi(x) = \infty$ and $C = \sup_{\mu \in M} \int_{\mathcal{X}} \varphi(|x|)\mu(dx) < \infty$, then M is tight.

Proof Observe that

$$\begin{aligned} \mu(|x| \geqslant n) &= \int_{|x|\geqslant n} d\mu = \int_{|x|\geqslant n} \frac{\varphi(|x|)}{\varphi(|x|)}\, d\mu \leqslant \frac{1}{\varphi(n)} \int_{|x|\geqslant n} \varphi(|x|) d\mu \\ &\leqslant \frac{C}{\varphi(n)}. \end{aligned}$$

The quantity on the right hand side is the same for all $\mu \in M$, it converges to 0 uniform in $\mu \in M$, and tightness follows. □

Theorem 2.3.7 (Prohorov's Theorem) *If a subset of $\mathbb{P}(\mathcal{X})$ is tight, then it is relatively compact.*

Put everything together, to show that a sequence of probability measures converges, it is sufficient to demonstrate that the sequence is tight and that any accumulation point is the same.

There is a converse to Prohorov's theorem:

Theorem 2.3.8 (Prohorov's Theorem - the converse) *If $\mathcal{X}$ is complete and separable, and if $A \subset \mathbb{P}(\mathcal{X})$ is relatively compact, then it is tight.*

Once we have established that a sequence of probability measures is tight, it is necessary to identify its limit. We discuss what set of functions to work with for this purpose.

Definition 2.3.9 A collection E of measurable functions is said to be *convergence determining* if, for any sequence μ_n and μ in $\mathbb{P}(\mathcal{X})$, the following holds: whenever $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$ for all $f \in E$, it follows that $\mu_n \to \mu$ weakly.

Recall by convergence determining we refer only to probability measures.

Proposition 2.3.10 *Let $(\mathcal{X}, d)$ be a separable metric space. The space A of uniformly continuous functions with bounded support is convergence determining. Furthermore, if $\mathcal{X}$ is locally compact, then the space of uniformly continuous functions with* **compact** *support is also convergence determining.*

Proof Let $\{x_i\}$ be a dense subset of $\mathcal{X}$. Define

$$f_{i,n}(x) = 2(1 - n\, d(x, x_i))^+.$$

For any open set G, define

$$\varphi_k(x) = \sum_{\{i,n=1,\dots,k: B_{x_i}(\frac{1}{n}) \subset G\}} f_{i,n}(x) \wedge 1.$$

Then φ_k is uniformly continuous with support contained in the ball centred at x_1 with radius $\max_{i \leqslant m} d(x_i, x_1) + 1$.

It is clear that $\mathbf{1}_G \geqslant g_m$ and $g_m \to \mathbf{1}_G$. Let $\mu_k, \mu \in \mathbb{P}(\mathcal{X})$ be such that $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$ for every $f \in A$. Then,

$$\liminf_{k\to\infty} \mu_k(G) \geqslant \liminf_{k\to\infty} \int g_m d\mu_k = \int g_m d\mu.$$

Taking the limit $k \to \infty$ allows us to conclude that $\mu_n \to \mu$. Note that the support of g_k is contained in a compact set if $\mathcal{X}$ is locally compact. □

Example 2.3.11 Let $\mathcal{X} = \mathbf{R}^n$. Denote by $\mathbf{C}_0^\infty(\mathbf{R}^n)$ the space of smooth function vanishing at infinity, and by $\mathbf{C}_K^\infty(\mathbf{R}^n)$ the set of smooth functions with compact support. Note that $\mathbf{C}_0(\mathbf{R}^n) = \overline{\mathbf{C}_K(\mathbf{R}^n)}$ (see proposition 4.3.5 in [4, pp.132]).

We have seen that $\mathbf{C}_K(\mathbf{R}^n)$ is convergence determining. In fact, $\mathbf{C}_K^\infty(\mathbf{R}^n)$ is dense in $\mathbf{C}_0(\mathbf{R}^n)$, and thus it is measure determining.

Definition 2.3.12 A family $E \subset \mathbf{C}_b(\mathcal{X})$ is said to **strongly separating points** if, for every $x \in \mathcal{X}$ and any $\delta > 0$, there exists a finite set $\{f_i : i = 1, \ldots, k\}$ from E such that

$$\inf_{y: d(y,x) \geqslant \delta} \max_{i=1,\ldots,k} |f_i(x) - f_i(y)| > 0.$$

For $\mathcal{X} = \mathbf{R}$, we can take a smooth function f with support in $B_x(\delta/2)$ such that $f(x) = 1$. Then $|f(x) - f(y)| = 1$ if $y \in B_x(\delta)$. Thus, the space $\mathbf{C}_K^\infty(\mathbf{R}^n)$ strongly separates points.

Theorem 2.3.13 *Let $(\mathcal{X}, d)$ be a complete separable metric space. IfE is an algebra of bounded continuous functions that strongly separates points, then E is convergence determining.*

2.4 Fourier transform of measures

Let $f \in L^1(\mathbf{R}^n; \mathbf{R})$, its Fourier transform

$$\hat{f}(\xi) = \int_{\mathbf{R}^n} e^{ix\cdot\lambda} f(x) dx, \qquad \lambda \in \mathbf{R}^n,$$

is uniformly continuous, and is bounded by the L_1 norm of f: $\|f|_\infty \leqslant \|f\|_1$. Furthermore $\lim_{|\lambda|\to\infty} |\hat{f}(\lambda)| = 0$ (Riemann-Lebesgue lemma). We can identify f with a finite measure $\mu(dx) = f dx$.

Likewise, if μ is a finite Borel measure on $\mathbf{R}^n$, we can define its Fourier transform.

Definition 2.4.1 The Fourier transform of a finite Borel measure μ on $\mathbf{R}^n$ is a complex-valued function given by the formula:

$$\hat{\mu}(\lambda) = \int_{\mathbf{R}^n} e^{i\langle \lambda, x\rangle} \mu(dx).$$

The Fourier transform of a finite measure is bounded by $\mu(\mathbf{R})$ and uniformly continuous.

Proposition 2.4.2 *If two finite Borel measures on $\mathbf{R}^n$ have the same Fourier transform, they are equal.*

Proof Firstly, if $f \in \mathcal{S}$, the space of Schwartz function of $\mathbf{C}^\infty$ functions of rapid decrease,

$$\int_{\mathbf{R}^n} \hat{f} d\mu = \int_{\mathbf{R}^n} f(x)\hat{\mu}(x)dx.$$

If $\hat{\mu}_1 = \hat{\mu}_2$, then $\int_{\mathbf{R}^n} \hat{f} d\mu_1 = \int_{\mathbf{R}^n} \hat{f} d\mu_2$. Since the Fourier transform is a bijection on $\mathcal{S}$, the above holds for all $f \in \mathcal{S}$, hence $\mu_1 = \mu_2$. □

Lemma 2.4.3 *If both f and $\hat{f}$ are in L^1, then the Fourier inversion formula holds:*

$$f(t) = \frac{1}{2\pi} \int_{\mathbf{R}^n} \hat{f}(\lambda) e^{-i\langle x, \lambda \rangle} d\lambda,$$

where $x \cdot \lambda$ denotes the scalar product on $\mathbf{R}^n$.

If $\hat{f}$ is not integrable, we cannot apply the Fourier inversion formula, it is useful to multiply by a rapidly decreasing function $\exp(-\epsilon x^2/2)$.

Example 2.4.4 If $p_\epsilon(x,t) = \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{x^2}{2\epsilon}}$, then $\hat{p}_\epsilon(x) = \exp(-\epsilon x^2/2)$.

Denote by γ and γ^n the standard Gaussian measure on $\mathbf{R}$ and on $\mathbf{R}^n$. The Fourier transform of $\gamma = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp^{-\frac{|x|^2}{2}} dx$ is

$$\int_{\mathbf{R}} e^{i\lambda x} \gamma(dx) = e^{-\frac{\lambda^2 \sigma^2}{2}}.$$

and $\gamma^n = \otimes \gamma$, so

$$\int_{\mathbf{R}} \sum_i e^{i\lambda_i x_i} \gamma(dx) = (\int_{\mathbf{R}} e^{i\lambda_i x_i} \gamma_i(dx)^n.$$

Theorem 2.4.5 (Lévy's continuity theorem) *Let $\mu_n, \mu \in \mathbb{P}(\mathbf{R}^d)$. Then if $\mu_n \to \mu$ weakly, then $\hat{\mu}_n \to \hat{\mu}$ pointwise. Conversely if $\hat{\mu}_n \to f$ pointwise and f os continuous at 0, there exists a probability measure μ such that $f = \hat{\mu}$ and $\mu_n \to \mu$ weakly.*

2.4.1 Gaussian Measures

Definition 2.4.6 A Borel measure μ on $\mathbf{R}^n$ is Gaussian if there exists a non-negative symmetric $n \times n$ matrix K and a vector $m \in \mathbf{R}^n$ such that

$$\int_{\mathbf{R}^n} e^{i\langle\lambda,x\rangle}\mu(dx) = e^{i\langle\lambda,m\rangle-\frac{1}{2}\langle K\lambda,\lambda\rangle}. \tag{2.2}$$

Denote by γ and γ^n the standard Gaussian measure on $\mathbf{R}$ and on $\mathbf{R}^n$. The Fourier transform of $\gamma = \frac{1}{(2\pi)^{\frac{n}{2}}}\exp^{-\frac{|x|^2}{2}} dx$ is

$$\hat\gamma(\lambda) = \int_{\mathbf{R}} e^{i\lambda x}\frac{1}{(2\pi)^{\frac{n}{2}}}\exp^{-\frac{|x|^2}{2}} dx = e^{-\frac{\lambda^2}{2}}.$$

and $\gamma^n = \otimes^N\gamma$, so

$$\int_{\mathbf{R}} e^{i\sum_j\lambda_j x_j}\gamma(dx) = (\int_{\mathbf{R}} e^{i\lambda_i x_i}\gamma(dx_i)^n = e^{-\frac{|\lambda|^2}{2}}.$$

Lemma 2.4.7 *Let C be an invertible positive definite symmetric matrix. For $\alpha \in \mathbf{C}$,*

$$\sqrt{\frac{\det C}{(2\pi)^n}}\int_{\mathbf{R}^n} e^{-\frac{\langle Cx,x\rangle}{2}}e^{\alpha\langle x,y\rangle}dx = e^{\frac{\alpha^2}{2}\langle C^{-1}y,y\rangle}.$$

Proof: First assume that $\alpha \in \mathbf{R}$, then the left hand side of the required equality equals

$$\begin{aligned}
&(2\pi)^{-\frac{n}{2}}\sqrt{\det C}\int_{\mathbf{R}^n} e^{-\frac{1}{2}\left\langle C(x-\alpha C^{-1}y),x-\alpha C^{-1}y\right\rangle}dx\; e^{\frac{\alpha^2}{2}\langle C^{-1}y,y\rangle}\\
=\;&(2\pi)^{-\frac{n}{2}}\sqrt{\det C}e^{\frac{\alpha^2}{2}\langle C^{-1}y,y\rangle}\int_{\mathbf{R}^n} e^{-\frac{1}{2}\langle Cx,x\rangle}dx\\
=\;&e^{\frac{\alpha^2}{2}\langle C^{-1}y,y\rangle}
\end{aligned}$$

by translation invariance of the Lebesgue measures. The result holds for $\alpha \in \mathbf{C}$ since both sides of the equality are analytic functions of α and they agree on $\mathbf{R}$. ■

If K is a symmetric non-degenerate matrix, and $m \in \mathbf{R}^n$, by translation invariance, we see that then

$$\mu := \frac{1}{\sqrt{(2\pi)^n\det(K)}}e^{-\frac{1}{2}\langle K^{-1}(x-m),\,x-m\rangle}dx$$

is a Gaussian measure with Fourier transform $\hat\mu(\lambda) = e^{\langle\lambda,m\rangle-\frac{1}{2}\langle K\lambda,\lambda\rangle}$.

Define the heat kernel on $\mathbf{R}^n$:

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{|x-y|^2}{2t}}.$$

Example 2.4.8 On the real line, a Gaussian measure has the Fourier transform $\hat{\mu}(\lambda) = e^{i\lambda m - \frac{\sigma^2\lambda^2}{2}}$. It is either a Dirac measure, δ_m, or it is absolutely continuous with respect to the Lebesgue measure with density $p_{\sigma^2}(m,x)$ where $m = \int_{\mathbf{R}} x\,\mu(dx)$ is the mean and $\sigma^2 = \int_{\mathbf{R}} x^2\mu(dx)$ is its variance.

Recall that

$$\int_{\mathcal{X}} \varphi \circ f \; d\mu = \int_{\mathcal{Y}} \varphi \; d(f_*\mu).$$

Observe that $\int_{\mathbf{R}^n} e^{i\langle \ell, x\rangle)}\mu(dx) = (\hat{\ell}_*\mu)(1)$ where the second ℓ denote the linear functional $x \mapsto \langle \ell, x\rangle)$. Hence (2.2) is equivalent to $(\hat{\ell}_*\mu)(1)e^{i\langle\lambda,m\rangle - \frac{1}{2}\langle K\lambda,\lambda\rangle}$.

Proposition 2.4.9 *A Borel measure μ on $\mathbf{R}^n$ is a Gaussian measure if and only if for every linear functional, $\ell : \mathbf{R}^n \to \mathbf{R}$, the push-forward measure $\ell_*\mu$ on $\mathbf{R}$ is Gaussian. The Gaussian measure μ has a density with respect to the Lebesgue measure if and only if K is non-degenerate in which case the density is*

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

Proof Let μ be a Gaussian measure on $\mathbf{R}^n$, with Fourier transform given by (2.2). Let $\ell : \mathbf{R}^n \to \mathbf{R}$ be linear, then $\ell(x) = \langle \ell^\sharp, x\rangle$ for some $\ell^\sharp \in \mathbf{R}^n$. For $\lambda \in \mathbf{R}$,

$$\begin{aligned}(\hat{\ell}_*\mu)(\lambda) &= \int_{\mathbf{R}} e^{i\lambda t}\ell_*\mu(dt) = \int_{\mathbf{R}^n} e^{i\lambda \ell(x)}\mu(dx) \\ &= \int_{\mathbf{R}^n} e^{i\langle \ell^\sharp, \lambda x\rangle}\mu(dx) = e^{i\lambda\langle \ell^\sharp, m\rangle - \frac{\lambda^2}{2}\langle K\ell^\sharp, \ell^\sharp\rangle},\end{aligned}$$

the last step follows from (2.2). Therefore $\ell_*\mu$ is a Gaussian measure with mean $\langle \ell^\sharp, m\rangle$ and variance $\langle K\ell^\sharp, \ell^\sharp\rangle$.

Suppose that for every linear functional ℓ on $\mathbf{R}^n$, $\ell_*\mu$ is Gaussian on $\mathbf{R}$. Below we identify the linear functional ℓ with $\ell^\#$. Denote their mean and variances, respectively, by $\tilde{m}(\ell)$ and $\sigma^2(\ell)$.

$$\tilde{m}(\ell) = \int x\ell_*\mu(dx) = \int t\ell_*\mu(dt) = \int_{\mathbf{R}} \ell(x)d\mu(x), \tag{2.3}$$

which is linear in ℓ so $\tilde{m}(\ell) = \langle \ell, m\rangle$ for some vector m. In addition,

$$\sigma^2(\ell) = \int (t - \langle \ell, m\rangle)^2 \ell_*\mu(dt) = \int_{\mathbf{R}} (\langle \ell, y\rangle - \langle \ell, m\rangle)^2\mu(dy) = \int_{\mathbf{R}^n} \langle \ell, y - m\rangle^2\mu(dy).$$

As a quadratic function, there exists K such that $\sigma^2(\ell) = \langle K\ell, \ell\rangle$. This means that

$$(\hat{\ell}_*\mu)(a) = e^{ia\langle \ell, m\rangle - \frac{a^2}{2}\langle K\ell, \ell\rangle},$$

which implies that
$$\int_{\mathbf{R}^n} e^{i\langle \ell, x\rangle)}\mu(dx) = (\hat{\ell}_*\mu)(1) = e^{i\langle \ell, m\rangle - \frac{1}{2}\langle K\ell, \ell\rangle},$$
and μ is Gaussian. □

Proposition 2.4.10 *If μ is a Gaussian measure determined by (2.2): $\int_{\mathbf{R}^n} e^{i\langle \lambda, x\rangle}\mu(dx) = e^{i\langle \lambda, m\rangle - \frac{1}{2}\langle K\lambda, \lambda\rangle}$, then the vector m and K are respectively the mean and the covariance matrix of the measure, which means that*
$$m = \int_{\mathbf{R}^n} x\,\mu(dx),$$
and
$$\langle Ku, v\rangle = \int_{\mathbf{R}^n} \langle x - m, u\rangle\langle x - m, v\rangle\mu(dx).$$

Proof Let $\ell(x) = x_i$, projection to its first component, $\{e_i\}$ the standard orthonormal basis of $\mathbf{R}^n$, then $\langle m, e_i\rangle = \int\langle e_i, x\rangle\mu(dx)$ by (2.3) , and
$$\int_{\mathbf{R}^n} \langle x - m, u\rangle\langle x - m, u\rangle\mu(dx) = \langle Ku, u\rangle,$$
By polarization we obtain the formula for $\langle Ku, v\rangle$. □

2.5 Gaussian Random Variables

A random variable with a Gaussian distribution is called a Gaussian random variable.

To rephrase Proposition 2.4.10 in terms of random variables, let $X = (X_1, \dots, X_n)$ a random variable with Gaussian distribution μ, where $\hat{\mu} = e^{i\langle \lambda, m\rangle - \frac{1}{2}\langle K\lambda, \lambda\rangle}$. Then $\mathbb{E}(X_i - m_i)(X_j - m_j) = K_{ij}$ and $\mathbb{E}X = m$. Consequently, any Gaussian measure on $\mathbf{R}^n$ is determined by its mean and its covariance operator. Furthermore, if K is diagonal, the measures decomposes into products of measures on $\mathbf{R}$ leading to:

Corollary 2.5.1 *A set of Gaussian processes $(X_1, \dots, X_n)$ is independent if and only if their covariances vanish.*

Lemma 2.5.2 *Assume that a sequence of Gaussian random variables on $\mathbf{R}^d$ converge, weakly, to a random variable X, then X is a Gaussian random variable.*

Proof For simplicity, assume that these variables are real-valued. Let $X_n \to X$ weakly. Let $m_n = \mathbb{E}X_n$, and K_n the covariance matrix of X_k. Then $\mathbb{E}[e^{i\langle \lambda, m_n\rangle - \frac{1}{2}\langle K\lambda, \lambda\rangle}] \to \mathbb{E}e^{itX}$, by (2.4.5), for every λ. The convergence of the left hand side implies that $m_n \to m$ and $\langle K_n\lambda, \lambda\rangle$ converges too. Thus the Fourier transform of X is $\hat{\mu}(\lambda) = e^{i\langle \lambda, m\rangle - \frac{1}{2}\langle K\lambda,\lambda\rangle}$, so the limit is Gaussian. □

Theorem 2.5.3 *If X is a Gaussian random variable on $\mathbf{R}^d$ with covariance operator K, and $A : \mathbf{R}^d \to \mathbf{R}^n$ a linear map, then AX is a Gaussian random variable with covariance AKA^T.*

Proof We only need to identify $\mathbb{E}\left[e^{i\langle \lambda, AX\rangle}\right]$ for any $\lambda \in \mathbf{R}^n$:

$$\begin{aligned}\mathbb{E}\left[e^{i\langle \lambda, AX\rangle}\right] &= \mathbb{E}\left[e^{i\langle A^T\lambda, X\rangle}\right]\\ &= e^{i\langle A^T\lambda, m\rangle - \frac{1}{2}\langle KA^T\lambda, A^T\lambda\rangle}\\ &= e^{i\langle \lambda, Am\rangle - \frac{1}{2}\langle AKA^T\lambda,\lambda\rangle}.\end{aligned}$$

This shows that X is a Gaussian random variable with mean Am and covariance AKA^T. □

The theorem holds also if X is Gaussian random variable on a Banach space E, and $A : E \to F$ is a bounded linear map from E to another Banach space.

Note that there exists a random variable $X = (X_1, X_2)$ with both marginals X_1 and X_2 Gaussian, but X is not Gaussian.

Let Z be a standard Gaussian variable on $\mathbf{R}^d$, $X = (X_1, \dots, X_n)$, K a matrix and C a vector, then $X = KZ + C$ has Gaussian distribution.

One of the nice properties of Gaussian random variables is the following. Let $(X_1, \dots, X_n)$ be Jointly Gaussian random variables. Then they are independent if and only if they are uncorrelated. Linear combinations of jointly Gaussian random variables are jointly Gaussian.

Exercise 2.5.4 If $\{X_1, \dots X_N\}$ are independent random variables with each X_i Gaussian on $\mathbf{R}^d$, and $a_i \in \mathbf{R}$, show that $\sum_{i=1}^N a_i X_i$ is a Gaussian random variable.

Proposition 2.5.5 *A random variable $(X_1, \dots, X_n)$ on $\mathbf{R}^n$ is Gaussian if and only if for any $a_i \in \mathbf{R}$, $\sum_{i=1}^n a_i X_i$ is a Gaussian random variable.*

Before closing this section we introduce a useful lemma. First recall that if $A(t)$ is a differentiable matrix-valued function with $A(0) = \mathrm{i}d$, then

$$\frac{d}{dt}\det(A(t))|_{t=0} = \mathrm{tr}(\frac{d}{dt}A(0)). \tag{2.4}$$

Lemma 2.5.6 *Suppose that C is a positive definite $n \times n$ matrix. Then the following statements hold.*

(i) If A is another symmetric $n \times n$ matrix, then

$$\mathrm{Trace} AC^{-1} = \frac{\sqrt{\det C}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbf{R}^n} \langle Ax, x \rangle e^{-\frac{\langle Cx,y \rangle}{2}} dx;$$

(ii)

$$(C^{-1})_{i,j} = \frac{\sqrt{\det C}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbf{R}^n} x_i x_j \, e^{-\frac{\langle Cx,y \rangle}{2}} dx;$$

Proof: For (i) take $h > 0$ small so $C + hA$ is positive, then by Lemma 2.4.7,

$$\begin{aligned}(2\pi)^{-\frac{n}{2}} \int_{\mathbf{R}^n} e^{-\frac{\langle (C+hA)x,x \rangle}{2}} dx &= \det(C + hA)^{-\frac{1}{2}} \\ &= (\det C)^{-\frac{1}{2}} \big(\det(I + hAC^{-1})\big)^{-\frac{1}{2}}.\end{aligned}$$

Differentiate for h at $h = 0$:

$$(2\pi)^{-\frac{n}{2}} \int_{\mathbf{R}^n} e^{-\frac{\langle Cx,x \rangle}{2}} \langle Ax, x \rangle dx = \frac{1}{2} (\det C)^{-\frac{1}{2}} \mathrm{trace} AC^{-1}.$$

For part (ii), fixing i, j, apply (i) with a matrix A whose elements $A_{p,q} = 0$, except $A_{i,j} = -1$ and $A_{j,i} = -1$, use symmetry of AC^{-1}.

A Gaussian measure on a finite dimensional vector space W is the pushed forward measure $T_*(\gamma^n)$ for some linear map $T : \mathbf{R}^n \to W$.

2.5.1 Probability measures on Banach spaces

It is a fact that probability measures on a Banach space is determined by the set of its push-forward measures by linear functionals. We denote by E^* the dual of E, it is the set of bounded linear functionals on E.

Proposition 2.5.7 *Two probability Borel measures μ and ν on a separable Banach space are the same, if $\ell_* \mu = \ell_* \nu$ for all $\ell \in E^*$.*

Proposition 2.5.8 *A Borel measure μ on a Banach space E is said to be a Gaussian measure if for every bounded linear functional, $f : E \to \mathbf{R}$, the push-forward measure on E is Gaussian.*

A vector m is said to be the mean of a Gaussian measure μ if for all $\ell \in E^*$, $\int \ell(x)\mu(dx) = \ell(m)$. The linear operator $\mathcal{K} : E^* \to E$ defined by

$$\ell'(\mathcal{K}\ell) = \int_E \ell(x)\ell'(x)\mu(dx).$$

is its covariance operator. If X is a random variable on Ehas distribution μ, then $\mathbb{E}(\ell(X)) = m(X)$ and $= \mathbb{E}[\ell(X)\ell'(X)] = \ell(\mathcal{K}\ell')$.

Definition 2.5.9 Let μ be a probability measure on Banach space E, define its Fourier transform $\hat{\mu} : E \to \mathbf{C}$ by

$$\hat{\mu}(\ell) = \int_E e^{i\ell(x)} d\mu(x).$$

For probability measures on a Hilbert space H, by the Reisz representation theorem we can identify H^* with H by $\ell \mapsto \ell^\#$ for $\langle \ell^\#, x\rangle = \ell(x)$, all $x \in H$. Then we define $\hat{\mu} : H \to \mathbf{C}$ by

$$\hat{\mu}(y) := \int_H e^{i\langle x,y\rangle} d\mu(x).$$

Thus $\hat{\mu}(y) = \widehat{\ell_* \mu}$ where $\ell(y) = \langle y, \cdot\rangle$. The covariance operator $\mathcal{K} : H \to H$ is then defined by

$$\langle \mathcal{K}e, e'\rangle = \int_H \langle x, e\rangle \langle x, e'\rangle \mu(dx).$$

If H has an orthonormal basis $\{e_i\}$, $X = \sum_{i=1}^\infty X_i e_i$ is a random variable with distribution μ, then $K_{i,j} = \mathbb{E}[X_i X_j]$.

Remark 2.5.10 If μ is a probability measure on a Banach space E then $\hat{\mu}$ is of positive type with $\mu(0) = 1$ and is continuous on E^*. (i) The continuity: if $\ell_n \to \ell$ in E, then $e^{i\ell_n(x)} \to e^{i\ell(x)}$ for each x. The dominated convergence theorem applies.

(ii) If $\lambda_1, \dots, \lambda_N \in E^*$, and $\xi_1, \dots, \xi_N \in \mathbf{C}$, then

$$\sum_{i,j=1}^N \hat{\mu}(\lambda_i - \lambda_j)\xi_i \overline{\xi_j} = \int_E |\sum \xi_j \exp^{i\lambda_j(x)}|^2 d\mu(x) \geqslant 0.$$

Theorem 2.5.11 (Bochner's theorem) *[**?**] The set of Fourier transformations of probability measures on $\mathbf{R}^n$ is precisely the set of functions $\hat{\mu} : \mathbf{R}^n \to \mathbf{C}$ of positive type with $\hat{\mu}(0) = 1$. Moreover each such $\hat{\mu}$ corresponds to a unique μ.*

2.6 Gaussian Processes

Definition 2.6.1 A stochastic process on $\mathbf{R}^n$ is a Gaussian process if its finite dimensional distributions are Gaussian measures.

Gaussian processes are widely used to represent thermal noise in electronic devices, resulting from random movement of electrons due to thermal agitation.

Let $m(t) = \mathbb{E}X_t$ denote its mean and

$$K(s,t) = \mathbb{E}(X_t - m(t))(X_s - m(s))$$

its covariance function. The finite dimensional distributions of a Gaussian process at time $(t_1, \dots, t_n)$ is determined by its mean $(m(t_1), \dots m(t_n))$ and the covariance matrix $(K(t_i, t_j))$.

For simplicity consider a real valued Gaussian process X_t with $\int_0^1 \mathbb{E}(X_t)^2 dt < \infty$. We may consider $X.$ as a random variable on $\Omega \to L^2([0,1]; \mathbf{R})$ with distribution μ. Denote $H = L^2([0,1]; \mathbf{R})$. Then, $\hat{\mu} : H \to \mathbf{R}$ is given by

$$\hat{\mu}(\varphi) = \int_H e^{\langle f, \varphi \rangle_{L^2}} \mu(df).$$

As computed with $\mathbf{R}^n$ case, there exists $m \in H$ and $\mathcal{K} : L^2 \to L^2$ bounded linear such that if $\ell = \langle \varphi, \cdot \rangle_H$, then $\ell_* \mu(t)$ is a Gaussian measure with mean $\langle \varphi, m \rangle_{L^2}$ and covariance $\langle \mathcal{K}\varphi, \varphi \rangle_{L^2}$. In fact, $m(t) = \mathbb{E}X_t$ and $\mathcal{K}\varphi(s) = \int_0^1 K(s,t)\varphi(s)ds$.

Definition 2.6.2 A linear operator T on a Hilbert space H is positive if for any $x \in H$, $\langle Tx, x \rangle \geqslant 0$. It is symmetric if for all $x, y \in H$, $\langle Tx, y \rangle = \langle x, Ty \rangle$.

Lemma 2.6.3 *Let X_t be a Gaussian process with covariance $K(s,t)$. Suppose that $\int_0^1 \mathbb{E}(X_s)^2 ds < \infty$ and define*

$$\mathcal{K}f(s) = \int_0^1 K(s,t) f(t) dt. \tag{2.5}$$

Then $\int_0^1 \int_0^1 (K(s,t))^2 ds dt < \infty$, and $\mathcal{K}$ is a positive symmetric operator on $L^2([0,1]; \mathbf{R}^n)$.

Proof We first show that $K \in L^2([0,1] \times [0,1])$. Let $m(t) = \mathbb{E}(X_t)$. Then,

$$\begin{aligned}\int_0^1 \int_0^1 K^2(s,t) ds dt &= \int_0^1 \int_0^1 \mathbb{E}(X_t - m(t))(X_s - m(s))]^2 ds dt \\ &\leqslant \int_0^1 \int_0^1 \mathbb{E}(X_t - m(t))^2 \mathbb{E}(X_s - m(s))^2 ds dt = (\int_0^1 \mathbb{E}(X_t - m(t))^2 dt)^2.\end{aligned}$$

Let $f \in L^2([0,1]; \mathbf{R}^n)$, then

$$\begin{aligned}\int_0^1 (Kf(s))^2 ds &= \int_0^1 (\int_0^1 K(s,t) f(t) ds)^2 ds \leqslant \int_0^1 \int_0^1 (K(s,t))^2 dt \int_0^1 f^2(t) dt ds \\ &= \|f\|_{L^2([0,1];\mathbf{R}^n)} \int_0^1 \int_0^1 (K(s,t))^2 dt ds < \infty,\end{aligned}$$

and $\mathcal{K}f \in L^2([0,1];\mathbf{R}^n)$. Finally, since K is symmetric, $\mathcal{K}$ is symmetric,

$$\langle \mathcal{K}f, g\rangle = \int_0^1 \int_0^1 \mathbb{E}(X_t - m(t))(X_s - m(s))]^2 f(t)g(s)dsdt = \langle Kg, f\rangle,$$

and

$$\langle \mathcal{K}f, f\rangle = \int_0^1 \int_0^1 \mathbb{E}(X_t - m(t))(X_s - m(s))]^2 f(t)f(s)dsdt = \mathbb{E}\int_0^1 (X_t - m(t))f(t)dt)^2 \geqslant 0,$$

i.e. $\mathcal{K}$ is positive. □

Exercise 2.6.4 Prove the multi-dimensional version of the statement in the previous lemma.

Definition 2.6.5 A function $K : [0,1]^2 \to \mathbf{R}$ is said to be a symmetric positive kernel if the operator $\mathcal{K}$, defined by (2.5), is symmetric and positive definite.

Definition 2.6.6 Let H be a separable Hilbert space. Let $\{e_n\}$ be an orthonormal basis of H and $T : H \to H$ a positive symmetric linear operator. Its trace is defined by

$$\text{tr}(T) = \sum_n \langle Te_n, e_n\rangle,$$

which is independent of the choice of the basis. The operator T is said to be of trace class if $\text{tr}(T) < \infty$.

If T is a linear operator, not necessarily of positive type, we define $|T| = \sqrt{TT^*}$, then T is of trace class if $\text{tr}(|T|) < \infty$.

Lemma 2.6.7 *Suppose that K is a symmetric positive kernel, then*

$$\text{tr}(K) = \mathbb{E}\int_0^1 (X_s)^2 ds.$$

Proof Let $\{e_n\}$ be an orthonormal basis of $L^2([0,1])$. Since $\mathcal{K}$ is of positive type, then

$$\begin{aligned}\text{tr}(K) &= \sum_i \langle Ke_i e_i\rangle = \sum_i \int_0^1 \int_0^1 \mathbb{E}[(X_t - m(t)(X_s - m(s))]e_i(t)dte_i(s)ds \\ &= \mathbb{E}\sum_i (\int_0^1 (X_t - m(t))e_i(t)dt)^2 = \sum_i \mathbb{E}\langle X - m, e_i\rangle^2_{L^2([0,1])} = \mathbb{E}\|X\|^2_{L^2([0,1])},\end{aligned}$$

confirming the claim. □

Definition 2.6.8 A linear operator $T : E \to F$ is compact if it takes a bounded subset of E into a pre-compact set in F. This means precisely the following: for any bounded sequence $\{x_n\}$ in E there exists a convergence subsequence of Tx_n.

If T has finite dimensional range then it is clearly compact.

Proposition 2.6.9 *[?] A trace class operator T on a separable Hilbert space is compact. For any self-adjoint compact operator on H there exists a complete othonormal basis $\{\varphi_n\}$ such that $T\varphi_n = \lambda_n \varphi_n$ where φ_n are real numbers.*

A non-negative and compact linear operator has a countable many real eigenvalues, for which only 0 is a possible accumulation point, the multiplicity of any non-zero eigenvalues if finite. In particular,

Theorem 2.6.10 (Mercer's Theorem) *[?, pp243]Riesz-Nagy] Let $K : [0,1] \times [0,1] \to \mathbf{R}$ be a real valued symmetric, continuous kernel such that $\int_0^1 \int_0^1 K^2(s,t)dsdt < \infty$. Suppose that*

$$\mathcal{K}f(t) = \int_0^1 K(s,t)f(s)ds,$$

is non-negative. Then,

$$K(s,t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} e_k(s) e_k(t)$$

where $\{e_i\}$ is an orthonormal sequence of eigen-functions of $\mathcal{K}$ and $\mathcal{K}e_k = \lambda_k e_k$. Furthermore,

$$\sup_{s,t \in [0,1]} \left| \sum_{k=1}^{N} \lambda_k e_k(s) e_k(t) - K(s,t) \right| \to 0,$$

as $N \to \infty$.

Note that e_n are non-negative and continuous.

Proposition 2.6.11 *Let X_t be a mean zero Gaussian process with mean zero and covariance K. Assuming that K is continuous. Let $\{e_k\}$ be an orthonormal system such that $\mathcal{K}e_i = \lambda_i e_i$. Then,*

$$X_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \beta_k e_k,$$

where $\{\beta_k\}$ are independent standard Gaussian random variables, in the sense that

$$\lim_{N \to \infty} \mathbb{E} \sup_{t \leqslant 1} \left| \sum_{k=1}^{N} \sqrt{\lambda_k} \beta_k e_k - X_t \right|^2 = 0.$$

Proof Set $X_t^N = \sum_{k=1}^N \sqrt{\lambda_k}\beta_k e_k$. Then

$$\sup_{t\in[0,1]} \mathbb{E}\|X_t^N - X_t^M|^2 = \sup_{t\in[0,1]} \mathbb{E}|\sum_{k=M+1}^N \sqrt{\lambda_k}\beta_k e_k(t)|^2 = \sup_{t\in[0,1]} \sum_{k=M+1}^N \lambda_k e_k^2(t),$$

which converges to zero as $N, M \to \infty$. This is due to the fact that

$$\sum_{k=1}^N \lambda_k e_k(s) e_k(t) \to K(s,t)$$

uniformly on $[0,1]^2$.

We denote by X_t the limit of X_t^N. Since for any $0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant 1$, the random vector $(X_{t_1}^N, \dots X_{t_n}^N)$ converges in the mean square sense to $(X_{t_1}, \dots X_{t_n})$. The limit of Gaussian random variables are Gaussian random variables, Proposition 2.5.2, proving that the limit process (X_t) is a Gaussian process. We can identify its limit by identifying its mean and covariance. Firstly, $\mathbb{E}(X_t) = \lim_{N\to\infty} X_t^N$, which follows from the L^2 convergence. Similarly,

$$\mathbb{E}(X_t^N X_s^N) = \mathbb{E}(\sum_{k=1}^N \sqrt{\lambda_k}\beta_k e_k(s))(\sum_{k=1}^N \sqrt{\lambda_k}\beta_k e_k(t)) = \sum_{k=1}^N \lambda_k (e_k(t))(e_k(s)) \to K(s,t).$$

Thus the limit $X(t)$ is a mean zero Gaussian process with mean zero and covariance K.

□

Definition 2.6.12 The Cameron-Martin space of the Gaussian measure on the separable metric space is the range of $\sqrt{K}$.

Note that if $T : E \to F$ is a linear, we want to induced an inner product on T from that on E as follows: if $f = T\tilde{f}$ and $g = \tilde{g}$, then $\langle f, g\rangle = \langle \tilde{f}, \tilde{g}\rangle$.

Let $K(s,t) = \min(s,t)$. Taking $T = \sqrt{\mathcal{K}}$ in this procedure to the Gaussian distribution of a stochastic process. For example, take $\sqrt{K}f(s) = \int_0^s f(r)dr$. Then the range of $\sqrt{K}$ is the Sobolev space H of finite energy.

Although we do not define the Cameron-Martin space of a measure, this concept is indeed intrinsic to the measure, not depending on whether we view the process as in $L^2([0,1])$ on in the Wiener space. If considered on $\mathbf{C}([0,1])$ we view $\mathbf{C}([0,1])$ as a subset of $\mathbf{C}([0,1])^*$. The latter is the space of measures, so $f \in \mathbf{C}([0,1])$ shall be identified with the measure $f\,dx$. The covariance operator is then from $\mathbf{C}([0,1])^*$ to $\mathbf{C}([0,1])$:

$$\mathcal{K}\mu(s) = \int_0^1 K(s,t)\mu(dt).$$

2.7 Brownian motion

Definition 2.7.1 A stochastic process $(X_t : t \geqslant 0)$ is said to have independent increments if for any $n \in \mathbf{N}$ and for any $0 \leqslant t_0 < t_1 < \cdots < t_n$, $X_{t_1} - X_{t_0}, \dots, X_{t_n - t_{n-1}}$ are independent random variables.

Definition 2.7.2 A standard one dimensional Brownian motion W_t on $\mathbf{R}$ is a sample continuous stochastic process starting from zero and such that

1. $W_t - W_s$ is a mean zero Gaussian random variable with variance $t - s$,
2. W has independent increments.

In particular $W_t - W_s$ is independent of $\sigma(W_r : 0 \leqslant r \leqslant s)$.

Denote by $\mathcal{F}_s^W := (W_r : 0 \leqslant r \leqslant s)$ the completion of the natural filtration of $(W_t : t \geqslant 0)$, then by Blumenthal's $0-1$ law, $\mathcal{F}_s^W$ is right continuous. However, we often need to consider a Brownian motion in a larger information system, and the filtration we use could be generated by multiples stochastic processes on the same filtered probability space. For this reason we introduce yet another definition of a Brownian motion.

Definition 2.7.3 Consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. An $(\mathcal{F}_t)$-adapted stochastic process (W_t) with values in $\mathbf{R}^d$ is a $\mathcal{F}_t$-Brownian motion if $W_t - W_s$ is a mean zero Gaussian random variable with variance $t - s$, and if for every pair of numbers $0 \leqslant s, t$, $W_{t+s} - W_s$ is independent of $\mathcal{F}_s$.

Definition 2.7.4 An n-dimensional Brownian motion with zero initial condition is a vector valued stochastic process $B_t = (B_t^1, \dots, B_t^n)$, where $\{B_t^i : 1 \leqslant i \leqslant n)$ is family of independent standard one dimensional Brownian motions.

Their covariance is $\mathbb{E}(B_t B_s) = \min(s,t) I_{n \times n}$ where $I_{n \times n}$ denotes the $n \times n$-matrix.

By Kolmogorov's continuity theorem, Theorem 2.8.9 below, we can infer that a stochastic process with $|X_t - X_s|_p \leqslant C|t-s|^\gamma$ for some $p > 1$ and $\gamma > 0$, this applies to a Gaussian process with covariance $\mathbb{E}(W_t W_s) = \min(s,t)$, has a continuous version which is furthermore locally Hölder continuous of order $\alpha < \frac{1}{2}$.

Theorem 2.7.5 *A standard one Brownian motion on* $\mathbf{R}$*, which we denote by* W_t*, is a sample continuous stochastic process starting from zero and such that*

1. *It is a Gaussian process;*
2. *It has mean zero and covariance function* $\mathbb{E}(W_t W_s) = \min(s,t)$*.*

Proof If (X_t) is a Brownian motion, then $\mathbb{E}(X_t X_s) = \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X_s)^2 = \min(s,t)$ for $s < t$, and any increment process $(X_{t_2} - X_{t_1}, \dots, X_{t_{n+1}} - X_{t_n})$ is a Gaussian process. Suppose that the for any n times, the process $(X_{t_1}, \dots, X_{t_n})$ is Gaussian, then

$$(X_{t_1}, \dots, X_{t_{n+1}}) = (X_{t_1}, \dots, X_{t_{n+1}} - X_{t_n}) - (X_{t_1}, \dots, X_{t_n}, X_{t_n})$$

is Gaussian, as $\sum_{i=1}^n a_i X_{t_i} + a_{n+1} X_n$ is Gaussian.

The other way around, assume that the process if Gaussian with mean zero and covariance function $\mathbb{E}(W_t W_s) = \min(s,t)$, then $X(t) \sim N(0,t)$. Check it has independent increments by working out the covariance. □

Let $p(t,x,y) := (2\pi t)^{\frac{n}{2}} e^{-\frac{|x-y|^2}{2t}}$ denote the heat kernel.

Proposition 2.7.6 *A continuous stochastic process x_t with initial value x is a Brownian motion if its finite dimensional distributions are given by:*

$$\begin{aligned}&\mathbb{P}(x_{t_1} \in A_1, \dots, x_{t_n} \in A_n)\\ &= \int_{A_1} \cdots \int_{A_k} p(t_1, x, y_1) p(t_2 - t_1, y_1, y_2) \dots p(t_k - t_{k-1}, y_{k-1}, y_k) dy_k \dots dy_1.\end{aligned}$$

Besides the Brownian motion, we have the following classes of processes:

Example 2.7.7 The Ornstein-Uhlenbeck process is a Gaussian process with covariance $e^{-\beta|t-s|}$; and the Brownian bridge, starting from the origin at time 0 and ending at the origin at time 1, is a Gaussian process with covariance $\min(s,t) - st$.

It is easy to verify that $W_t - tW_1$ is a Gaussian process, and so is $e^{-t} W_{2t}$. What are their covariance? Equally, check that $\int_0^t g(s) W_s ds$ is a Wiener process if g_s is a continuous deterministic function. Hint : Proposition 2.5.2.

2.7.1 Self-similar stochastic Processes with stationary increments

Definition 2.7.8 A stochastic process $(X_t : t \geqslant 0)$ is said to have stationary increments if for any $n \in \mathbf{N}$ and for any $0 \leqslant t_0 < t_1 < \cdots < t_n$, the probability distribution of the stochastic process $X_{t_1+s}, \dots, X_{t_n+s}$ is independent of $s \geqslant 0$.

Definition 2.7.9 A stochastic process X_t is self-similar, with exponent $H \in (0,1]$, if $X_{at} = a^H X_t$ for all t.

A Brownian motion has stationary increments, and is self-similar with self-similarity exponent $H = \frac{1}{2}$.

Remark 2.7.10 Note that if $X_{at} = a^H X_t$ and $H > 0$ and $X(t)$ is continuous at zero, then $X(0) = 0$ a.e..

Proposition 2.7.11 *Suppose that X_t is real-valued such that $X_0 = 0$, H-self-similar , with stationary increments, and with finite second moment, then*

$$\mathbb{E}(X_t X_s) = \frac{1}{2}\sigma^2(t^{2H} + s^{2H} - |t-s|^{2H}),$$

where $\sigma^2 = \mathbb{E}(X_1)^2$.

Proof This follows from a simple computation:

$$\mathbb{E}(X_t X_s)^2 = -\frac{1}{2}(\mathbb{E}(X_t - X_s)^2 - \mathbb{E}(X_t)^2 - \mathbb{E}(X_s)^2) = \frac{1}{2}[t^{2H} + s^{2H} - (t-s)^{2H})]\mathbb{E}(X_1^2),$$

proving the claim. □

A generalisation to independent sequence of random variables are m-dependent sequence: $X_{n'}$ and X_n are independent if $|n' - n| > m$. If X_t is a stochastic process, we define:

$$\varrho(n) = X_{n+1} - X_n$$

to be the process of increment 1.

Definition 2.7.12 We say that X_t has long range dependence (rather, its increment has long range dependence) if there does not exist a number m such that $\varrho(n+k)$ and $\varrho(n)$ are independent for all $k > m$ and all n.

Proposition 2.7.13 *If X_t is a continuous self-similar process, with self-similar exponent $H \in (0,1)\backslash\{\frac{1}{2}\}$, has stationary increment and finite second moment, then X_t has long range dependence. In fact, if $R(n) := \mathbb{E}(\varrho_n \varrho_0)$ then as $n \to \infty$,*

$$R(n) \sim H(2H-1)n^{2H-2}\mathbb{E}(X_1^2).$$

Proof Since $X(0) = 0$,

$$R(n) = \mathbb{E}[X_1(X_{n+1} - X_n)] = \frac{1}{2}[((n+1)^{2H} - n^{2H}) - (n^{2H} - (n-1)^{2H})]\mathbb{E}[(X_1)^2],$$

concluding the proof by Taylor expansion. □

Remark 2.7.14 Observe that if $H \in (0, \frac{1}{2})$, then $R(n)$ decays sufficiently fast,

$$\sum |R(n)| = \sum n^{2H-2} < \infty.$$

If $H > \frac{1}{2}$, on the other hand, the decay in correlation is slow, and the series $\sum |R(n)|$ is not summable.

Definition 2.7.15 Let $H \in (0,1]$. If a real valued mean zero continuous Gaussian process, with $B_0^H = 0$, has covariance

$$\frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})\mathbb{E}[(B_1^H)^2],$$

it is called a fractional Brownian motion (fBM) with Hurst exponent H. It is a standard fBM if $\mathbb{E}[B_H(1)^2] = 1$.

Proposition 2.7.16 *A fractional Brownian motion is self-similar with exponent H and has stationary increments.*

Proof Firstly, $X_t^H := B_{at}^H$ is a Gaussian process. It is sufficient to identify its covariance:

$$\mathbb{E}[X_t^H X_s^H] = \frac{1}{2}a^{2H}(t^{2H} + s^{2H} - |t-s|^{2H})\mathbb{E}[(B_1^H)^2] = \mathbb{E}[(a^H B_t^H)(a^H B_s^H),$$

and thus $X^H = B^H$ in law.

Similarly, for any $a \geqslant 0$, the increment process $X_t := B_{t+a} - B_t$ is Gaussian, one compute:

$$\mathbb{E}[(B_{t+a}^H - B_t^H)(B_{s+a}^H - B_s^H)] = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})\mathbb{E}[(B_1^H)^2],$$

hence X^H and B^H are equal in law. □

2.7.2 Integral representation for fractional Brownian motion

The following theorem will need the basics on a Wiener integral. Wiener integral is of the form $\int_0^t f_s dW_s$ where $f : [0,T] \to \mathbf{R}$ is such that $\int_0^T (f_s)^2 ds < \infty$ and $t \leqslant T$ and W_s is a one dimensional Brownian motion. Let S denote the space of simple functions : $f \in S$ is of the form

$$f(t) = \sum_{k=1}^N a_k \mathbf{1}_{A_k}(t)$$

where $A_k \in \mathcal{B}(\mathbf{R})$. We can in fact choose $A_k = (a_k, b_k]$ and define

$$\int_0^1 f(t)dW_t = \sum_{k=1}^N a_k(B_{t_k} - B_{t_{k-1}}).$$

The integral $\int_0^1 f(t)dW_t$ is a Wiener process on can show that

$$\mathbb{E}(\int_0^t f_s dW_s)^2 = \|f\|_{L^2}.$$

If $f \in L^2$, we take simple functions $f_n \to f$ in L^2, then $\int_0^t f_n(s)dW_s$ converges in L^2 which we define to be $\int_0^1 f_s dW_s$. Itîsometry and Gaussian property remains to hold. Similarly, for any t

$$\mathbb{E}(\int_0^t f_s dW_s \int_0^t g_s dW_s) = \int_0^t f_s g_s ds.$$

Definition 2.7.17 If W^1 and W^t are two independent Brownian motions, we define $W_t = W_t^1$ for $t > 0$ and $W_t = W_{-t}^2$ for $t \leqslant 0$. Then W_t is called a two sided Brownian motion.

Theorem 2.7.18 *Let*

$$X_t = \int_{-\infty}^0 ((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}})dW_u + \int_0^t (t-u)^{H-\frac{1}{2}})^2 dW_u.$$

Then X_t is a fractional Brownian motion.

Proof For $u < t$, let

$$f(t,u) = \begin{cases} (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}, & t \geqslant 0 \\ (t-u)^{H-\frac{1}{2}}, & t < 0. \end{cases}$$

We first show that for any $t > 0$, $f \in L^2([-\infty, t)]$.

$$\begin{aligned} \int_{-\infty}^t (f(t,u))^2 du &= \int_{-\infty}^0 (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}})^2 du + \int_0^t (t-u)^{2H-1}\, du \\ &= t^{2H}(\int_{-\infty}^1 (1-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}})^2 du + \int_0^1 (1-s)^{2H-1}\, ds) \\ &= t^{2H} \int_{-\infty}^1 (f(1,u))^2 du < \infty. \end{aligned}$$

Note that X_t and Y_t are Gaussian processes. Observe that the integrals from $(-\infty, 0)$ and $[0, T)$ are independent, also $(\int_0^h [(t+h-u)^{H-\frac{1}{2}} - (h-u)^{H-\frac{1}{2}})dW_u$ and $\int_h^{t+h} (t+h-u)^{H-\frac{1}{2}} dW_u$ are independent. Thus,

$$\begin{aligned} \mathbb{E}[X_{t+h} - X_h)^2 &= \int_{-\infty}^0 ((t-u)^{H-\frac{1}{2}} - (h-u)^{H-\frac{1}{2}})^2 du + \mathbb{E}\big[(\int_0^{t+h} (t+h-u)^{H-\frac{1}{2}} dW_u - \int_0^h (h-u)^{H-\frac{1}{2}} dW_u)^2\big] \\ &= \int_{-\infty}^0 ((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}})^2 du \\ &\quad + \mathbb{E} \int_0^h ((t+h-u)^{H-\frac{1}{2}} - (h-u)^{H-\frac{1}{2}})^2 du + \int_h^{t+h} (t+h-u)^{2H-1} du \\ &= \int_{-\infty}^h ((t+h-u)^{H-\frac{1}{2}} - (h-u)^{H-\frac{1}{2}})^2 du + \int_h^{t+h} (t+h-u)^{2H-1} du \end{aligned}$$

$$= t^{2H} \int_{-\infty}^{1} [f(1,u)]^2 du.$$

By polarisation,

$$\mathbb{E}[X(t)X(s)] = \frac{1}{2}(t^{2H} + S^{2H} - |t-s|^{2H}) \int_{-\infty}^{1} [f(1,u)]^2 du.$$

Since $X(t)$ is a Gaussian process it is a fBM. □

2.8 Sample properties

A stochastic process Y_t is said to have a certain property (P), if for almost surely all ω, $t \mapsto Y_t(\omega)$ has the property. For example,

Definition 2.8.1 A stochastic process (X_t) is said to be continuous, if for almost surely all ω, $t \mapsto X_t(\omega)$ is continuous. Similarly, a stochastic process is Hölder continuous if $t \mapsto X_t(\omega)$ is everywhere Hölder continuous, almost surely.

2.8.1 Hölder spaces

Hölder continuity is a measurement for continuity. Given a function $f : \mathbf{R} \to \mathbf{R}$, one is interested whether there exists a function ω such that

$$|f(x) - f(y)| \leqslant \omega(|x-y|).$$

A useful way to strengthen the notion of continuity is to require its modulus of continuity proportional to $|x-y|^\alpha$ and $\alpha \in (0,1]$.

Definition 2.8.2 Let D be an open subset of $\mathbf{R}^d$. A function $f : D \to \mathbf{R}^n$ is locally Hölder continuous of exponent α if for every $u_0 \in D$ there exists a neigbourhood of D and a constant c such that for $u, v \in D$,

$$|f(u) - f(v)| \leqslant c|u-v|^\alpha,$$

here $| - |$ denotes the Euclidean norm. In our finite dimensional setting this is equivalent to require that for relatively compact subset D' of D

$$\sup_{u \neq v, u,v \in D'} \frac{|f(u) - f(v)|}{|u-v|^\alpha} < \infty.$$

The definition for Hölder continuity extends to functions between metric spaces.

Definition 2.8.3 Let D be an open subset of $\mathbf{R}^d$. A function $f : D \to \mathbf{R}^n$ is uniformly Hölder continuous of exponent α if

$$\sup_{u \neq v, u,v \in D} \frac{|f(u) - f(v)|}{|u - v|^\alpha} < \infty,$$

The quantity

$$|f|_\alpha := \sup_{u \neq v, u,v \in D} \frac{|f(u) - f(v)|}{|u - v|^\alpha}$$

is the Hölder semi-norm of f on D.

Definition 2.8.4 A function $x : \mathbf{R} \to \mathbf{R}^d$ is locally Hölder continuous of exponent α if

$$\sup_{0 \leqslant u \neq v, u,v \in K} \frac{|x_u - x_v|}{|u - v|^\alpha} < \infty,$$

for any compact subset K of $\mathbf{R}$.

We denote by $\mathbf{C}^\alpha$ the space of locally Hölder continuous functions. For example, the space of paths $\mathbf{C}^\alpha = \mathbf{C}^\alpha([0,T], \mathbf{R}^d)$ consists of paths with finite Hölder semi-norm.

$$\mathbf{C}^\alpha([0,T], \mathbf{R}^n) = \{x : [0,T] \to \mathbf{R}^n : |x|_\alpha < \infty.\}$$

Recall that a function $f : D \to \mathbf{R}$, where $D \subset \mathbf{R}^n$ is concave if for any $x, y \in D$ and any $t \in [0,1]$,

$$f(tx + (1-t)y) \geqslant tf(x) + (1-t)f(y).$$

Lemma 2.8.5 *If $f : \mathbf{R}_+ \to \mathbf{R}$ is concave and $f(0) = 0$ then $f(x) + f(y) \geqslant f(x+y)$ for any $x > 0, y > 0$. In particular, for $x > y$,*

$$f(x) - f(y) \leqslant f(x - y).$$

Proof Indeed, $x = \frac{y}{x+y} 0 + \frac{x}{x+y}(x+y)$, by concavity,

$$f(x) \geqslant \frac{y}{x+y} f(0) + \frac{x}{x+y} f(x+y),$$

Similarly,

$$f(y) \geqslant \frac{x}{x+y} f(0) + \frac{y}{x+y} f(x+y).$$

Adding up the two to conclude. □

Example 2.8.6 Let $\alpha \in (0,1)$. Consider $f : \mathbf{R}_+ \to \mathbf{R}$, $f(x) = |x|^\alpha$. On any interval $[\epsilon^2, \infty)$, $f'' = \alpha(\alpha-1)x^{\alpha-2} < 0$, so f is concave on $(0,\infty)$, and on $\mathbf{R}_+$. Thus, for $x > y$, $|x|^\alpha - |y|^\alpha \leqslant |x-y|^\alpha$. This shows that $|x|^\alpha$ is uniformly Hölder continuous.

Introducing a norm:

$$\|f|_{\mathbf{C}^{0,\alpha}} = |f|_\infty + \|f\|_\alpha.$$

The space $\mathbf{C}^{0,\alpha}(0,T]; \mathbf{R}^n)$ is the completion of the space of $\mathbf{C}^\infty$ smooth functions, from $[0,T]$ to $\mathbf{R}^n$, under the $\mathbf{C}^{0,\alpha}$ norm.

If Ω is a bounded subset of $\mathbf{R}^n$, there is a continuous compact inclusion map: for $0 < \alpha < \beta \leqslant 1$, $\mathbf{C}^{0,\beta} \subset \mathbf{C}^{0,\alpha}$:

$$\begin{aligned}\|f\|_\alpha = \frac{|f(u)-f(v)|}{|u-v|^\alpha} &= \Big(\frac{|f(u)-f(v)|}{|u-v|^\beta}\Big)^{\alpha/\beta} |f(u)-f(v)|^{1-\frac{\alpha}{\beta}}\\ &\leqslant (\|f\|_\beta)^{\frac{\alpha}{\beta}} (2|f|_\infty)^{1-\frac{\alpha}{\beta}}.\end{aligned}$$

If the stochastic processes has continuous sample paths, by its distribution we mean the measure it induces on the Wiener space $\mathbf{C}([0,T];\mathcal{X})$.

2.8.2 Kolmogorov's continuity theorem

Given a collection of random variables index by $[0,T]$, apriori we know nothing of its measurability in time or its continuity. We can often choose a continuous version

Definition 2.8.7
1. Two stochastic processes X_t and Y_t on the same probability space are *modifications* of each other if for each t, $P(X_t = Y_t) = 1$. The exceptional set $\{\omega : X_t(\omega) \neq Y_t(\omega)\}$ may depend on t. Such processes are said to be versions of each other.

2. Two stochastic processes X_t and Y_t on the same probability space are *indistinguishable* of each other if $P(X_t = Y_t, \forall \text{t}) = 1$.

By a Kolmogorov's continuity theorem, also referred as Kolmogorov-Chentsov Theorem, we mean the following type of result: Given appropriate moment bounds on the distance $d(X_t, X_s)$, there is a continuous and Hölder continuous modification. Furthermore, there is a bound on the L_p norm of the Hölder norm of the modification. In $\mathbf{R}^1$, Kolmogorov theorem infers continuous modification of a stochastic process from bounds of the form

$$\mathbb{E}|X(t) - X(s)|^p \leqslant C|t-s|^{1+\beta}.$$

We use the notation $\|X_t - X_s\|_p \equiv \mathbb{E}(|X(t) - X(s)|^p)^{\frac{1}{p}}$.

Definition 2.8.8 A stochastic process X_t is continuous in probability (also known as stochastic continuity) if for any s, and any $\delta > 0$,

$$\lim_{t\to s} \mathbb{P}(|X_t - X_s| > \delta) = 0.$$

Markov-Chebechev inequality implies that $\mathbb{P}(|X_t - X_s| > \delta) \leqslant C\delta^{-p}\mathbb{E}(|X_t - X_s|^p) \leqslant C|t-s|^{1+\beta}$, should the moment estimate given above holds. For example, a Poisson process N_t satisfies that $\|N_t - N_s\|_p = C_p|t-s|^{\frac{1}{p}}$, it is stochastic continuous, however does not have a continuous version.

Theorem 2.8.9 (Kolmogorov's theorem in dimension 1) *Let $(X_t, 0 \leqslant t \leqslant T)$ be a stochastic process with values in a Banach space E , such that for a number $p > 1$ and some $\delta > \frac{1}{p}$ and some constant C_p.*

$$\|X_t - X_s\|_p \leqslant C_p|t-s|^{\delta}.$$

Then there exists a continuous modification, $\tilde{X}_t$, which is furthermore Hölder continuous. For every $\gamma < \delta - \frac{1}{p}$, $X \in \mathbf{C}^\gamma([0,T];\mathbf{R})$ almost surely and

$$\||\tilde{X}|_\gamma\|_{L_p} \leqslant C_p \sum_{n\in\mathbf{N}_0} \frac{1}{2^{n(\delta-\frac{1}{p}-\gamma)}} < \infty.$$

Example 2.8.10 Consider the probability space $([0,1], \mathcal{B}([0,1]))$ with Lebesgue measure, and the stochastic process:

$$X_t(\omega) = \begin{cases} 1, & t = \omega \\ 0, & t \neq \omega \end{cases}$$

For any ω, X_t is not a continuous function. The conditions of the Kolmogorov Theorem holds, we can take $X_t(\omega) \equiv 0$.

Example 2.8.11 Let $f_n : [0,T] \times \mathbf{R}^d \to \mathbf{R}$ and $\bar{f} : \mathbf{R}^d \to \mathbf{R}$ be measurable functions. Suppose that we have that for $p > 1$, as $n \to \infty$,

$$\Big\| \int_s^t (f_n(r,x) - \bar{f}(x))\, dr \Big\|_p \lesssim K\, n^{-\frac{\delta}{p}} |t-s|^{1-\frac{\delta}{p}} .$$

Apply Kolmogorv theorem, we see that for every $\gamma < 1 - \frac{\delta}{p} - \frac{1}{p}$,

$$\Big\| \sup_{s\neq t, s,t\in[0,T]} \frac{|\int_s^t (f_n(r,x) - \bar{f}(x))\, dr|}{|t-s|^\gamma} \Big\|_p \lesssim CK\, n^{-\frac{\delta}{p}} ,$$

i.e. the Hölder norm of the map $t \mapsto \int_0^t (f_n(r,x) - \bar{f}(x))\, dr$ is in L_p.

There is a version of Kolmogorov's theorem that applies to a stochastic process with values in a general metric spaces and indexed by a parameter in subset of a n-dimensional Euclidean space. We shall keep it simple, taking $(X_t)_{t\in D}$ where X_t are real valued and $D \subset \mathbf{R}^n$.

Lemma 2.8.12 *Consider the dyadic partition of* $[0,1]$ *to* 2^m *points, this divides* $[0,1]^d$ *into sub-cubic of side length* $\frac{1}{2^m}$*. Let* S_m *denote the collection of vertices at level* m *on these sub-cubes and set* $S = \cup_m S_m$*. Let* $(X_t)_{t\in S}$ *be a family of random variables indexed by* S *with the property that there exists* $\delta > \frac{d}{p}$ *such that*

$$\sup_{s<t, s,t\in S} \mathbb{E}(|X_s - X_t|^p)^{\frac{1}{p}} \leqslant C|t-s|^{\delta}.$$

Let $\tilde{S}_m$ *denote the nearest neighbour pairs in* S_m*, then for every* $\gamma < \delta - \frac{d}{p}$*, for almost surely all* ω*, the following holds for some constant* $\tilde{C}$*:*

$$\mathbb{E}\Big[\Big(\sup_{(s,t)\in\cup_m \tilde{S}_m} \frac{\Big(|X_s - X_t|}{|s-t|^\gamma}\Big)^p\Big]^{\frac{1}{p}} < C\tilde{C}.$$

In particular, $\sup_m \sup_{(s,t)\in\tilde{S}_m} \frac{|X_s - X_t|}{|s-t|^\gamma}$ *is finite almost surely.*

Proof Note that S_m contains points whose coordinates taking values in $\{\frac{k}{2^m}, k = 0, 1, \dots, 2^m\}$. The cardinality of S_m is at most $C2^{md}$ where $C = 2^d$. We care about the exponential rate md, not the factor C, as m is taken to infinity while d is held fixed. For this reason, $\#(S_m)$ is said to be of order 2^{md}. Since $\tilde{S}_m$ denotes the nearest neighbour points in S_m, the cardinality of $\tilde{S}_m$ is also of order 2^{md}. Hence,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sup_m \sup_{(s,t)\in\tilde{S}_m} \Big(\frac{|X_s - X_t|}{|s-t|^\gamma}\Big)^p\Big] \leqslant \sum_m \mathbb{E}\Big[\big(\sup_{(s,t)\in\tilde{S}_m} \frac{(|X_s - X_t|}{2^{-m\gamma}}\big)^p\Big]\\
&\leqslant \sum_m \sum_{(s,t)\in\tilde{S}_m} \mathbb{E}\Big[\big(\frac{|X_s - X_t|}{2^{-m\gamma}}\big)^p\Big]\\
&\leqslant C^p \sum_m \sum_{(s,t)\in\tilde{S}_m} 2^{-m(\delta p - \gamma p)} \leqslant C^p \sum_m 2^{md} 2^{m(\gamma p - \delta p)}
\end{aligned}$$

This is finite if $\gamma < \delta - \frac{d}{p}$, concluding the proof. □

For continuity statement, we consider the open cubic $(0, 1^d)$ and let $D_m = \{(\frac{k_1}{2^m}, \dots, (\frac{k_j}{2^m}) : 0 \leqslant k_j \leqslant 2^m\} \cap (0,1)^d$, it is the set of points whose coordinates are at level m.

Lemma 2.8.13 *Let* $\tilde{D}_m$ *denote the nearest neighbour pairs in* D_m*. If* $f(t)$ *is a real valued function defined on* $\cup_m D_m \cap (0,1)^d$ *such that there exists* $\gamma > 0$ *and* $C > 0$ *so that*

$$\sup_m \sup_{s,t\in D_m : |s-t| = 2^{-m}} \frac{|f(t) - f(s)|}{|t-s|^\gamma} = C < \infty,$$

for any nearest neigbours s, t on D_m, then f extends to a continuous function on the cube $[0,1]^d$, and the extension is Hölder continuous of order γ, with Hölder constant $C\tilde{C}$ where $\tilde{C}$ is a constant independent of f.

Proof For every point $s \in [0,1]^d$, on the grid D_m there is a closest point, denote it by $\pi_m(s)$, in particular

$$|\pi_m(s) - s| \leqslant 2^{d/2} 2^{-m}.$$

Note that $|\pi_m(s) - \pi_{m+1}(s)| \leqslant 22^{d/2}\, 2^{-m}$, and the sequence $\{\pi_m(s)\}$ is a Cauchy sequence, converging to s. Furthermore, there is a chain of neigbouring pairs at level m connecting $\pi_m(s)$ to $\pi_{m-1}(s)$ (at most of length $2d$). By the nearest neighbour assumption on the Hölder norm,

$$|f(\pi_m(s)) - f(\pi_{m+1}(s))| \leqslant 2^{\gamma(1+\frac{d}{2})} C\, 2^{-\gamma m}.$$

Since $|\sum_m f(\pi_m(s)) - f(\pi_{m+1}(s))|$ is summable, $f(\pi_m(s))$ has a limit which we define to be $\tilde{f}(s)$:

$$\tilde{f}(s) = \lim_{m\to\infty} f(\pi_m(s)),$$

and

$$|\tilde{f}(s) - f(\pi_m(s))| \leqslant CC_2\, 2^{-m\gamma}.$$

On a point in $\cup_m D_m$, $\tilde{f}$ agrees with f, for the nearest dyadic point to it is itself. We have defined an extension of f to the cube, which we denote by $\tilde{f}$.

We proceed to show that $\tilde{f} : (0,1)^d \to \mathbf{R}$ is Hölder continuous. Take any $s \neq t$ from the cube, close to each other, then there exists m_0 with

$$\frac{1}{2^{m_0+1}} \leqslant |s - t| \leqslant \frac{1}{2^{m_0}}.$$

Then $|\pi_{m_0}(s) - \pi_{m_0}(t)| \leqslant C(d) 2^{-m_0}$ for a universal constant $C(d)$ depending only on d. There is a short chain of nearest points in D_{m_0} connecting $\pi_{m_0}(s)$ and $\pi_{m_0}(t)$, so

$$|f(\pi_{m_0}(s)) - f(\pi_{m_0}(t))| \leqslant C(d) C 2^{-m_0\gamma}.$$

Put these together,

$$|\tilde{f}(s) - \tilde{f}(t)| \leqslant |\tilde{f}(s) - f(\pi_{m_0}(s))| + ||\tilde{f}(s) - f(\pi_{m_0}(t))| \leqslant 2C(d)C2^{-m_0\gamma}(2^{-m_0})^{\gamma} \leqslant CC'|s-t|^{\gamma},$$

where C' depends only on d. This proved that the extension is Hölder continuous of order γ. □

Definition 2.8.14 If a family of random variables are indexed by a set $D \in \mathbf{R}^n$, it is often referred as a random field.

Definition 2.8.15 Let $p > 1$, $\alpha > 0$, we define the following function space of adapted stochastic processes.

$$\mathcal{B}_{\alpha,p} = \left\{ X : X_t \in \mathcal{F}_t, \sup_{s<t} \frac{\|X_t - X_s\|_p}{|t-s|^\alpha} < \infty \right\}.$$

Theorem 2.8.16 *Let D be a compact set of $\mathbf{R}^d$. If $(X_t)_{t\in D}$ is a random field in $\mathcal{B}_{\alpha,p}$, for some $\alpha > 0$, $p > 1$, and $C > 0$. And its norm is bounded as follows:*

$$\sup_{s \neq t, s,t \in D} \|X_s - X_t\|_p \leqslant C|t-s|^\alpha,$$

Then X_t has a continuous extension, denoted as $\tilde{X}_t$, which furthermore satisfies that for any $\gamma < \alpha - \frac{d}{p}$,

$$\Big\| \sup_{s \neq t, s,t \in D} \frac{|\tilde{X}_s - \tilde{X}_t|}{|s-t|^\gamma} \Big\|_p \leqslant C\tilde{C}$$

where $\tilde{C}$ is the constant appearing in Theorem 2.8.12. In particular, the space $\mathcal{B}_{\alpha,p} \subset L_p(\Omega, \mathbf{C}^\gamma)$ (up to modification) for $\gamma < \alpha - \frac{d}{p}$.

Proof By Lemma 2.8.12,

$$\mathbb{E}\Big[\Big(\sup_{(s,t)\in \cup_m \tilde{S}_m} \frac{|X_s - X_t|}{|s-t|^\gamma} \Big)^p \Big]^{\frac{1}{p}} < C\tilde{C}.$$

Then there exists a subset of Ω of full measure, for ω from this set, the Hölder norm of $X_\cdot(\omega)$ on the dyadic points is finite. Let $A(\omega)$ be the smallest number such that $\sup_{(s,t)\in\cup_m \tilde{S}_m} \frac{|X_s - X_t|}{|s-t|^\gamma} \leqslant A(\omega)$. For ω with $A(\omega) < \infty$, let $\tilde{X}_t(\omega)$ denote its continuous extension. Note that the process X_t is continuous in probability, so its values are determined by its values in a dense subset and $X_t(\omega) = \tilde{X}_t(\omega)$ almost surely. Note that Lemma 2.8.13, the Hölder norm for each sample path of the extension x is the same as that for x on nearest neighbours, consequently their L_p norm is finite with the same bound. □

Chapter 3

Markov Processes

Let the state space of the stochastic processes be denoted by $\mathcal{X}$, which is assumed to be a complete and separable metric space.

Definition 3.0.1 If X_t is a stochastic process, we denote by $\mathcal{F}_t^X$ its natural filtration:

$$\mathcal{F}_t^X = \sigma(X_s : s \leqslant t, s \in I),$$

which is the smallest σ-algebra generated X_s, for $s \in (-\infty, t] \cap I$.

Definition 3.0.2 An $\mathcal{F}_t$ adapted stochastic process (X_t) is said to be an $\mathcal{F}_t$-Markov process, if for any $s < t$ and for any $A \in \mathcal{B}(\mathcal{X})$, the following holds

$$\mathbb{P}[X_t \in A \,|\, \mathcal{F}_s] = \mathbb{P}[X_t \in A \,|\, X_s] \tag{3.1}$$

almost surely.

The notation $\mathbb{P}[- \,|\, x_s]$ denotes taking conditional expectation with respect to the random variable X_s. We shall drop $\mathcal{F}_t$ and simply call X_t a Markov process. A Markov process, with respect to any filtration, is also a Markov process with respect to its own filtration.

3.1 Markov Processes

Proposition 3.1.1 *If (X_t) is a Markov process, then for any $s \leqslant t_1 < \cdots < t_n$, and n, and bounded measurable functions $f_i : \mathcal{X} \to \mathbf{R}$, the following identity holds:*

$$\mathbb{E}\big[\Pi_{i=1}^n f_i(X_{t_i}) | \mathcal{F}_s\big] = \mathbb{E}\big[\Pi_{i=1}^n f_i(X_{t_i}) | X_s\big]. \tag{3.2}$$

The set of Borel measurable functions can be replaced by the union of a set of measure determining functions and the constants.

3.1.1 Transition Function

To address measurability issues, we introduce the notion of (Markov) transition functions and assume that a Markov process has such a transition function. A Markov transition function is used to specify the probability that a Markov process, starting from a point x, lands in a set A at time t.

Definition 3.1.2 A **time homogeneous transition kernel** is a family of probability measures, $P := \{P_t(x, \cdot) : x \in \mathcal{X}, 0 \leqslant t\}$ on $\mathcal{X}$, with the following properties:

(i) for each $x \in \mathcal{X}$, $P_t(x, \cdot)$ is a probability measure on $\mathcal{X}$;

(ii) for each $A \in \mathcal{B}(\mathcal{X})$, the function $x \mapsto P_t(x, A)$ is Borel measurable.

(iii) $P_0(x, \cdot) = \delta_x$, where δ_x is the Dirac measure at x.

(iv) for any $r \leqslant s \leqslant t$, $x \in X$ and $B \in \mathcal{B}(\mathcal{X})$,

$$P_{s+t}(x, B) = \int_{\mathcal{X}} P_t(y, B) P_s(x, dy). \tag{3.3}$$

This last equation is referred to as the Chapman-Kolmogorov equation. A family of probability measures satisfying the first two conditions are referred to as a Markov kernel.

Note that (3.3) is equivalent to:

$$\int_{\mathcal{X}} f(y) P_{t+s}(x, dy) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(z) P_t(y, dz) P_s(x, dy). \tag{3.4}$$

Definition 3.1.3 The transition function (P_t) is said to be the transition function for a (time homogeneous) Markov process (X_t) if, for any $A \in \mathcal{B}(\mathcal{X})$ and $s < t$,

$$\mathbb{P}[X_t \in A | \mathcal{F}_s] = P_{t-s}(X_s, A), \quad a.s. \tag{3.5}$$

The distribution of X_0 is called the initial distribution.

Remark 3.1.4 We shall assume that $P_t(x, \mathcal{X}) = 1$, which means that X_t does not explode.

Proposition 3.1.5 *If* $(X_t, t \geqslant 0)$ *is a Markov process with transition probabilities* P *and initial distribution* μ_0*, then for any* $f_i \in \mathcal{B}_b(\mathcal{X})$, $0 < t_1 < \cdots < t_n$,

$$\mathbb{E}(\Pi_{i=1}^n f_i(X_{t_i})) = \overbrace{\int_{\mathcal{X}} \cdots \int_{\mathcal{X}}}^{n+1} \Pi_{i=1}^n f_i(y_i) \Pi_{i=1}^n P(y_{i-1}, dy_i) \mu(dy_0). \tag{3.6}$$

Proof The proof for this is routine. For $f \in \mathcal{B}_b(\mathcal{X})$, we have

$$\mathbb{E}[f(X_t)] = \mathbb{E}(\mathbb{E}[f(X_t)|X_0]) = \mathbb{E}(\int f(y) P_t(X_0, dy)) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(y) P_t(y_0, dy) \mu_0(dy).$$

Let us assume that this holds for N time points, where $N \geqslant 1$. Then

$$\begin{aligned}\mathbb{E}\big(\Pi_{i=1}^N f_i(X_{t_i})\big) &\overset{tower}{=} \mathbb{E}\big(\mathbb{E}\big(\Pi_{i=1}^N f_i(X_{t_i})|\mathcal{F}_{N-1}\big)\big)\\ &= \mathbb{E}\Big(\Pi_{i=1}^{N-1} f_i(X_{t_i}) \mathbb{E}(f_N(X_{t_N})|\mathcal{F}_{N-1})\Big)\\ &\overset{Markov}{=} \mathbb{E}\Big(\Pi_{i=1}^{N-1} f_i(X_{t_i}) \mathbb{E}\big(f_N(X_{t_N})|\ X_{t_{N-1}}\big)\Big)\\ &= \mathbb{E}\left(\Pi_{i=1}^{N-1} f_i(X_{t_i}) \int_{\mathcal{X}} f_N(y_N) P(X_{t_{N-1}}, dy_N)\right)\end{aligned}$$

The last function involves only $t_0, t_1, \ldots, t_{N-1}$, so we can apply the induction hypothesis:

$$\begin{aligned}RHS &= \overbrace{\int_{\mathcal{X}} \cdots \int_{\mathcal{X}}}^{N} \left(\Pi_{i=1}^{N-1} f_i(y_i) \int_{\mathcal{X}} f_N(y_N) P(y_{N-1}, dy_N)\right) \Pi_{i=0}^{N-1} P(y_i, dy_{i+1}) \mu(dy_0)\\ &= \overbrace{\int_{\mathcal{X}} \cdots \int_{\mathcal{X}}}^{N+1} \Pi_{i=1}^N f_i(y_i)\, \Pi_{i=1}^N P(y_{i-1}, dy_i) \mu(dy_0).\end{aligned}$$

The last line follows after bring $\Pi_{i=1}^{N-1} f_i(y_i)$ inside the inner most integral. □

A similar proof shows that following:

Exercise 3.1.6 Let (X_t) be a Markov process with transition function $P_t(x, A)$, $t_1 \leqslant \cdots \leqslant t_n$, and $f_1, \ldots, f_n$ from $\mathcal{B}_b(\mathcal{X})$, then

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

It is sufficient to prove it for f_i the indicator functions. We show this for $n = 2$,

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

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

3.2 Semi-groups and Generators

The set of all linear operators between two normed vector spaces $(E, |\cdot|_E)$ and $(F, |\cdot|_F)$ shall be denoted by $\mathcal{L}(E, F)$, on which we define the operator norm:

$$\|T\| := \sup_{|x|_E=1} |Tx|_F.$$

A linear operator T is said to be bounded if its operator norm is bounded.

Example 3.2.1 Suppose that a $(E, |\cdot|)$ is a finite dimensional normed vector space. Denote by n its dimension. Then every linear map from $E \to F$ is bounded. Indeed, let $\{e_i\}_{i=1}^n$ 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 there 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 3.2.2 *Let $T \in \mathcal{L}(E, F)$. The following statements are equivalent:*

1. *T is bounded,*
2. *T is continuous,*
3. *T is continuous at 0.*

Example 3.2.3 Let $T : \mathbf{C}^1([0, 2\pi]) \to \mathbf{C}([0, 2\pi])$ be the derivative operator $Tf = f'$. Then T is not bounded. Take $f_n(t) = \sin(nt)$ and use $|Tf|_\infty = |f'|_\infty$.

Let us consider an index set Λ, usually taken to be an interval $[0, T], \mathbf{R}_+$, or the set of natural numbers $\mathbf{N}$ and $\mathbf{N}_0 = \mathbf{N} \cup \{0\}$.

Definition 3.2.4 A one parameter family of bounded linear operators $T(t) : E \to E$ on a Banach space E, $t \in \Lambda$, is said to be a semigroup if

$$T(t+s) = T(t)T(s), \qquad T(0) = I, \tag{3.7}$$

where I denotes the identity amp,

Example 3.2.5 The following defines a semi-group of bounded linear operators on sespective spaces:

- $A \in \mathcal{M}_{n\times n}$, the set of $n \times n$ matrices. Define $T_t : \mathbf{R}^{\to}\mathbf{R}^n$ by $T_t = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}$.
- Translation on the circle $S^1 = e^{is}$: define $T_t(e^{is}) = e^{i(t+s)}$.
- *Translation Semi-group.* Define $T_t : BC(\mathbf{R};\mathbf{R}) \to BC(\mathbf{R};\mathbf{R})$ by $T_t f(x) = f(x+t)$.
- *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 $T_t : E \to E$ by $T_t f(s) = \mathbb{E}[f(t+s)|\mathcal{F}_s]$.

Example 3.2.6 A time homogeneous Markov process induces a semi-group of linear operators on $\mathcal{B}_b(\mathcal{X})$ by the formula:

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

Firstly, $T_0 f(x) = \int f(y)\delta_x(dy) = f(x)$. Then,

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

in the last step we have used (3.4), the Chapmann-Kolmogorov equation.

Definition 3.2.7 A linear operator T on E is said to have

1. the positive preserving property if $Tf \geqslant 0$ whenever $f \geqslant 0$;
2. the conservative property if $T1 = 1$
3. 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, on a locally conpact space, 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 of the image $T_t f$ where $f \in E$ is more suitable.

Definition 3.2.8 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 3.2.9 An example of a non-strongly continuous semigroup on $BC(\mathbf{R}; \mathbf{R})$ is: $T_0 = I$ and $T_t = 0$ for $t > 0$.

3.2.1 Generators

Definition 3.2.10 Let T_t be a strongly continuous semigroup of bounded linear operators on a Banach space E. We define its generator by: for $f \in E$,

$$\mathcal{L} f := \lim_{t \searrow 0} \frac{T_t f - f}{t} \tag{3.8}$$

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

$$\mathbb{D}(\mathcal{L}) := \{f \in \mathcal{B}_b(\mathcal{X}) : \text{ the limit (3.8) exists}\}.$$

Exercise 3.2.11 Show that the translation semi-group is not strongly continuous on $BC(\mathbf{R}; \mathbf{R})$ either. Identify its generator and a space on which it is strongly continuous.

3.3 Feller and Strong Feller

Recall that the dual space E^* to a Banach space E is the set of continuous linear functions from E to $\mathbf{R}$. 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 $\ell : \mathbf{C}(\mathcal{X}) \to \mathbf{R}$ is automatically bounded and $\ell(f) \leqslant \ell(\|f\|) = \|f\|\ell(1)$. The Riesz representation theorem states that if $\mathbf{C}(\mathcal{X})$ is a compact metric the dual space $\mathbf{C}(\mathcal{X})^*$ 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.

Definition 3.3.1 A semigroup T_t on $\mathcal{B}_b(\mathcal{X})$ is said

1. to be Feller if it restricts to a semigroup on $\mathbf{C}(\mathcal{X})$.
2. to be strong Feller if $T_t f$ is continuous for any $f \in \mathcal{B}_b(\mathcal{X})$ and any $t > 0$.

Remark 3.3.2 If T_t is associated with a transition function, then the Feller property is equivalent to that $x \mapsto P_t(x, \cdot)$ is continuous from $\mathcal{X}$ to $\mathbb{P}(\mathcal{X})$ in the weak topology, which precisely means for any f bounded and continuous, whenever $x_n \to x$,

$$\lim_{n\to\infty} \int f(y)P(x_n, dy) = \int f(y)P(x, dy).$$

Example 3.3.3 Let $x_0 \in \mathcal{X}$, set $P(x, dy) = \delta_{x-x_0}$. Then $Tf(x) = \int_{\mathcal{X}} f(y)P(x, dy) = f(x-x_0)$ is Feller.

Example 3.3.4 (Not Feller) Let $P(x, A)$ be a family of transition probabilities on $\mathbf{R}$ given below

$$P(x, \cdot) = \begin{cases} \delta_1, & \text{if } x > 0 \\ \delta_0, & \text{if } x \leqslant 0. \end{cases}$$

Then

$$Tf(x) = \int_{\mathbf{R}} f(y)P(x, dy) = \begin{cases} f(1) & \text{if } x > 0 \\ f(0), & \text{if } x \leqslant 0, \end{cases}$$

and Tf fails to be continuous at 0 for continuous functions f with $f(1) \neq f(0)$.

3.4 Stopping times

A stopping time is, roughly speaking, the time that an event has arrived. This time is ∞ if the event does not arrive. For a nice account of stopping times see Kallenberg [**?**], here we only state the basic properties of stoppoing times. Let $I \subset \mathbf{R}_+$.

Definition 3.4.1 A function $T : \Omega \to I \cup \{\infty\}$ is a $(\mathcal{F}_t)$ stopping time if for every $t \in I$, the event $\{T \leqslant t\}$ belongs to $\mathcal{F}_t$.

Note that A constant time is a stopping time, and so is $T(\omega) \equiv \infty$.

Let (X_t) be a stochastic process on S. For $B \in \mathcal{B}(S)$ let

$$T_B(\omega) = \inf\{t > 0 : X_t(\omega) \in B\},$$

T_B is referred as the hitting time of B by (X_t). By convention, $\inf(\phi) = +\infty$.

For discrete index $I = \mathbf{N}$, f function $T : \Omega \to \mathbf{N}$ is a $\{\mathcal{F}_n\}$ stopping time if and only if $\{T(\omega) = n\} \in \mathcal{F}_n$ for all n. Indeed, if T is a stopping time, $\{T = n\} = \{T \leqslant n\} \cap \{T \leqslant n-1\}^c \in \mathcal{F}_n$. Conversely, $\{T \leqslant n\} = \cup_{i=1}\{T = i\} \in \mathcal{F}_n$ if $\{T(\omega) = n\} \in \mathcal{F}_n$ for all n.

Example 3.4.2 Suppose that $(X_n)_{n\in\mathbf{N}}$ is $(\mathcal{F}_n)$ adapted. Let B be a measurable set. Then T_B is an $\mathcal{F}_n$ stopping time:

$$\{T_B \leqslant n\} = \cup_{k\leqslant n}\{\omega : X_k(\omega) \in B\} \in \mathcal{F}_n.$$

If (X_t) is an right continuous $(\mathcal{F}_t)$-adapted stochastic process, the hitting time of an open set is an $\mathcal{F}_t^+$-stopping time. Recall one of the usual assumptions: $\mathcal{F}_t = \mathcal{F}_t^+$. The first hitting time of closed set by a continuous $(\mathcal{F}_t)$-adapted stochastic process is an $\mathcal{F}_t$- stopping time.

Proposition 3.4.3 *Let S, T, T_n be stopping times.*

(1) Then $S \vee T = \max(S,T)$, $S \wedge T = \min(S,T)$ are stopping times.

(2) $\limsup_{n\to\infty} T_n$ and $\liminf_{n\to\infty} T_n$ are stopping times.

Proof Part (1) follows from the following observations:

$$\{\omega : \max(S,T) \leqslant t\} = \{S \leqslant t\} \cap \{T \leqslant t\}, \quad \{\omega : \min(S,T) \leqslant t\} = \{S \leqslant t\} \cup \{T \leqslant t\}.$$

Since

$$\limsup_{n\to\infty} T_n = \inf_{n\geqslant 1} \sup_{k\geqslant n} T_n, \qquad \liminf_{n\to\infty} T_n = \sup_{n\geqslant 1} \inf_{k\geqslant n} T_n$$

we only proof that if T_n is an increasing sequence, $\sup_n T_n$ is a stopping time; and if S_n is a decreasing sequence of stopping times with limit S, $\inf_n S_n$ is a stopping time. These follows from $\{\sup_n T_n \leqslant t\} = \cap_n\{T_n \leqslant t\}$, $\{\inf S \leqslant t\} = \cup_n\{S_n \leqslant t\}$. □

Given a process (X_t) and a stopping time S, we define the stopped process X^S by : $(X^S)_t = X_{S\wedge t}$. For simplicity we remove the bracket and denote $X_t^S := (X^S)_t$.

Definition 3.4.4 Let T be a stopping time. Define

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leqslant t\} \in \mathcal{F}_t, \forall t \geqslant 0\}.$$

If $T = t$ is a constant time, $\mathcal{F}_T$ agrees with $\mathcal{F}_t$. For T takes values in **N**,

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T = n\} \in \mathcal{F}_n, \forall n \in \mathbf{N}\}.$$

Definition 3.4.5 A stochastic process $X : I \times \Omega \to E$ is progressively measurable if

(1) $X : I \times \Omega \to E$ is measurable

(2) for each $t > 0$, $X : [0, t] \times \Omega \to E$ is a measurable map with respect to the product σ-algebra $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$.

Theorem 3.4.6 *If T is a stopping time and $(X_t, t \in I)$ a stochastic process. Then X_T is $\mathcal{F}_T$-measurable if $I = [a, b]$ and (X_t) is progressively measurable.*

Proof By the definition X_T is measurable w.r.t. $\mathcal{F}_T$ if and only if for any $B \in \mathcal{B}(\mathbf{R})$ and $t \geqslant 0$. $\{X_T \in B\} \cap \{T \leqslant t\} \in \mathcal{F}_t$. Observe that

$$\{X_T \in B\} \cap \{T \leqslant t\} = \{X_{T\wedge t} \in B\} \cap \{T \leqslant t\}.$$

It is sufficient to show that for any stopping time $\tau \leqslant t$, X_τ is $\mathcal{F}_t$ -measurable. We define a random variable with values in $[0, t]$:

$$\psi : (\Omega, \mathcal{F}_t) \to \Big(\Omega \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t]\Big)$$
$$\psi(\omega) = (\omega, \tau(\omega)))$$

Then ψ is measurable. In fact, if $A \in \mathcal{F}_t$, $s_1 < s_2 \leqslant t$,

$$\{\omega : \psi(\omega) \in A \times (s_1, s_2)\} = A \cap \{\omega : \tau(\omega) \in (s_1, s_2)\} \in \mathcal{F}_t.$$

Let

$$Y : \Big(\Omega \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t]\Big) \to (\mathbf{R}, \mathcal{B}(\mathbf{R}))$$

be given by

$$Y(\omega, t) = X_t(\omega).$$

Since (X_t) is progressively measurable, Y is measurable. This concludes that $X_\tau = Y \circ \psi$ is $\mathcal{F}_t$-measurable. □

Theorem 3.4.7 *If T is finite stopping time, then $\mathcal{F}_T = \sigma\{X_T : X$ is càdłàg$\}$.*

For a proof see Revuz-Yor [16] and Protter [**?**].

Proposition 3.4.8 *Let S, T be stopping times.*

(1) If $S \leqslant T$ then $\mathcal{F}_S \subset \mathcal{F}_T$.

(2) Let $S \leqslant T$ and $A \in \mathcal{F}_S$. Then $S\mathbf{1}_A + T\mathbf{1}_{A^c}$ is a stopping time.

(3) S is $\mathcal{F}_S$ measurable.

(4) $\mathcal{F}_S \cap \{S \leqslant T\} \subset \mathcal{F}_{S\wedge T}$.

Proof (1) If $A \in \mathcal{F}_S$,

$$A \cap \{T \leqslant t\} = (A \cap \{S \leqslant t\}) \cap \{T \leqslant t\} \in \mathcal{F}_t$$

and hence $A \in \mathcal{F}_T$. (2) Since $\mathcal{F}_S \subset \mathcal{F}_T$,

$$\{S\mathbf{1}_A + T\mathbf{1}_{A^c} \leqslant t\} = (\{S \leqslant t\} \cap A) \cup (\{T \leqslant t\} \cap A^c) \in \mathcal{F}_t.$$

(3) Let $r, t \in \mathbf{R}$, $\{S \leqslant r\} \cap \{S \leqslant t\} = \{S \leqslant \min(r,t)\} \in \mathcal{F}_t$. Hence $\{S \leqslant r\} \in \mathcal{F}_r$. (4) Take $A \in \mathcal{F}_S$ and $t \geqslant 0$. Then

$$A \cap \{S \leqslant T\} \cap \{S \wedge T \leqslant t\} = (A \cap \{T \leqslant t\}) \cap \{S \leqslant t\} \cap \{S \wedge t \leqslant T \wedge t\} \in \mathcal{F}_t.$$

which follows as $S \wedge t$ and $T \wedge t$ are $\mathcal{F}_t$-measurable, and $A \cap \{S \leqslant T\} \in \mathcal{F}_{S\wedge T}$. □

Every stopping time can be approximated by stopping times taking only a countable number of values.

Lemma 3.4.9 *Let S be a stopping time. Define:*

$$S_n = \frac{1}{2^n}[2^n S + 1].$$

Then each S_n is a stopping time, S_n decreases with n, and $|S_n - S| \leqslant \frac{1}{2^n}$.

Indeed,

$$S_n(\omega) = \frac{j+1}{2^n}, \text{ if } S(\omega) \in \left[\frac{j}{2^n}, \frac{j+1}{2^n}\right), \qquad j = 0, 1, 2 \ldots.$$

If $t \in [\frac{j}{2^n}, \frac{j+1}{2^n})$, since $S_n(\omega)$ takes an integer value,

$$\{S_n(\omega) \leqslant t\} = \{S_n(\omega) \leqslant \frac{j}{2^n}\} = \{S(\omega) < \frac{j}{2^n}\} \in \mathcal{F}_{\frac{j}{2^n}} \subset \mathcal{F}_t.$$

So S_n are stopping times.

3.4.1 The Canonical Picture

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 $\mu_X = \mathcal{L}(X.)$ on $\mathcal{X}^I$. This measure encodes all statistical information of the process.

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

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 3.4.10 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 $A \in \mathcal{B}(\mathcal{X})$, $\mathbb{P}(X_t \in A) = \mathbb{E}[P_t(X_0, A)] = \int_{\mathcal{X}} P_t(y, A)\,\mu(dy)$. and by **??**, for any $A_0, \dots, A_n \in \mathcal{B}(\mathcal{X})$ and $0 = t_0 < t_1 < \cdots < t_n$,

$$\mathbb{P}(X_{t_0} \in A_0, \dots, X_{t_n} \in A_n) = \int_{A_0} \cdots \int_{A_n} P_{t_n - t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(y_0, dy_1)\mu(dy_0). \tag{3.9}$$

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

$$\mu_\Delta(A_0 \times \cdots \times A_n) := \int_{A_0} \cdots \int_{A_n} P_{t_n - t_{n-1}}(y_{n-1}, dy_n) \cdots P_{t_1}(y_0, dy_1)\mu(dy_0). \tag{3.10}$$

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

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

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

Equation (3.10) precisely means that the finite dimensional distributions of π_t are $\pi_\Delta^* P_\mu$, it is therefore a Markov process.

Definition 3.4.12 If $\mu = \delta_x$ in theorem 3.4.11, 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,

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

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 3.4.13 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 $I = \mathbf{R}^+$. To this end, let $\theta_s : \mathcal{X}^{\mathbf{R}_+} \to \mathcal{X}^{\mathbf{R}_+}$, $\theta_s\omega(t) = \omega(s+t)$ be the shift operator. If $\Phi : \mathcal{X}^{\mathbf{R}_+} \to \mathbf{R}$ is a Borel measurable function, we introduce the notation:

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

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

Theorem 3.4.14 *Let* $(X_t)_{t\geqslant 0}$ *denote the canonical Markov process with transition function* P*. Then, for any* $\Phi \in \mathcal{B}_b(\mathcal{X}^{\mathbf{R}_+})$*,*

$$\mathbb{E}_x\big[\Phi(\theta_s X) \,|\, \mathcal{F}_s\big] = \mathbb{E}_{X_s}[\Phi(X)] \quad \mathbb{P}_x - a.s. \tag{3.11}$$

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

Remark 3.4.15 This can be written as

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

We stress that the expectations in (3.11) have to be understood as integrals on the path space. To be utterly precise, (3.11) requires that

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

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

Since $\Pi_{i=1}^n P(y_{i-1}, dy_i)\mu(dy_0)$ is a consistent family of finite dimensional distributions, by Kolmogorov's extension theorem, one obtains the following:

Proposition 3.4.16 *If $P_t(x,\cdot)$ is a time homogeneous Markov transition function, then for any initial distribution μ_0, there exists is Markov process on $\mathcal{X}$ with initial distribution μ_0 and such that P_t is its Markov transition function.*

Proof It is enough to prove this for

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

Then (3.11) becomes

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

By Theorem 3.1.6

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

where the second line follows from (3.9) with $\mu = \delta_{X_s}$, proving the required identity. □

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 [17], [10], [18, Section III.7]. Recall that $\mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$. However note that we must add the condition that $\mathcal{X}$ is locally compact.

Theorem 3.4.17 *If (X_t) is a Markov process with transition semigroup (T_t), which is strongly continuous on $\mathbf{C}_0(\mathcal{X})$ where $\mathcal{X}$ is a locally compact space, then there exists a càdlàg modification of (X_t), which is an $(\mathcal{F}_t^+)$-Markov process with the same transition semigroup.*

Remark 3.4.18 If $(Y_t)_{t\geqslant 0}$ has càdlàg or continuous sample paths, we can use similar arguments as above to construct a measure on $D(\mathbf{R}_+, \mathcal{X})$ and $\mathbf{C}(\mathbf{R}_+, \mathcal{X})$, respectively. Since these spaces are however not in $\mathcal{B}(\mathcal{X})^{\otimes \mathbf{R}_+}$, this is not a simple corollary of our results and one has to work with the trace σ-fields instead.

3.5 Strong Markov Property

Given a stochastic [process $X.$, we define

$$(\theta_T X.)_n = X_{T+n}$$

This means for $\omega \in \Omega$ and $t \geqslant 0$, $(\theta_T X.)_t$ is a random variable given by $(\theta_T X.)_t(\omega) = X_{T(\omega)+t}(\omega)$. The shift Markov process starts from x_T. Observe that x_{T+s} is measurable with respect to $\mathcal{F}_{T+s}$.

Definition 3.5.1 A time-homogeneous Markov process (X_t) with transition probabilities P is said to have the strong Markov property if for every finite stopping time T and for every bounded measurable function $\Phi : \mathcal{X}^{\mathbf{N}} \to \mathbf{R}$, the following holds:

$$\mathbb{E}\big(\Phi(\theta_T X.) \,|\, \mathcal{F}_T\big) = \mathbb{E}\big(\Phi(\theta_T X.) \,|\, X_T\big) \qquad a.s. \tag{3.12}$$

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 $\mathcal{F}_t^+ := \bigcap_{r>t} \mathcal{F}_r$ denote its right-continuous version.

Proposition 3.5.2 *Let (X_t) be a Markov process with right-continuous sample paths. If its transition semigroup (T_t) leaves $\mathrm{BC}(\mathcal{X})$ or $\mathbf{C}_0(\mathcal{X})$-invariant, then (X_t) is an $(\mathcal{F}_t^+)$-Markov process.*

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

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

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

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

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

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

Both measures 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 $\mathbf{C}_0(\mathcal{X})$ is measure-determining class, $\mu = \nu$, as required. □

Let τ be a stopping time and recall that the stopped σ-field is defined by

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

The following two lemmas are standard:

Lemma 3.5.3 *Let*

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

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

With this one can show that

Lemma 3.5.4 *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 3.5.5 *Let (X_t) be a right-continuous Markov process whose transition function leaves either $\mathbf{C}_0(\mathcal{X})$ or $\mathrm{B}C(\mathcal{X})$ invariant. Then it has the strong Markov property. If (X_t) is cádlág, the Markov property in the canonical picture is as follows: if Φ is a real valued bounded measurable function on $D\big([0,1], \mathcal{X}\big)$,*

$$\mathbb{E}\big[\Phi \circ \theta_\tau \mathbf{1}_{\{\tau<\infty\}} \,|\, \mathcal{F}_\tau\big] = \mathbf{1}_{\{\tau<\infty\}} \mathbb{E}_{X_\tau}[\Phi]. \tag{3.13}$$

Proof This holds if the Markov process is indexed by only a countable number of times. Let us first suppose that τ takes only a countable number of values $\{t_k :\ k \in \mathbf{N}\}$ with $0 \leqslant t_1 < t_2 < \cdots < \cdots \leqslant \infty$. Then, using Theorem 3.4.14, we get for each $B \in \mathcal{F}_\tau$,

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

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:

$$\mathbb{E}\big[f(X_{t+\tau})\mathbf{1}_{\{\tau<\infty\}}|\mathcal{F}_\tau\big] = T_t f(X_\tau)\mathbf{1}_{\{\tau<\infty\}}. \tag{3.14}$$

Now assume a general τ, for the approximating sequence of theorem 3.5.3,

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

By the right-continuity of X and the Feller property of T_t, for any $f \in BC$ (or $f \in \mathbf{C}_0(\mathcal{X})$), (4.17) 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.

□

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

Example 3.5.6 Let us return to Example **??**, 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:

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

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

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

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

3.6 Remarks

3.6.1 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 define $X_t = \Delta$ for t greater or equal to its life time

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

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 $t \geqslant \tau(\omega)$.

Example 3.6.1 Let (B_t) be a real valued Brownian motion. The stochastic process $X_t(\omega) := \frac{1}{2 - B_t(\omega)}$ is defined up to the first time $B_t(\omega)$ reaches 2 which we denote by τ:

$$\tau(\omega) = \inf_{t\geqslant 0}\{B_t(\omega) \geqslant 2\}.$$

For any given time t, no matter how small it is, there is a set of path of positive probability (measured with respect to the Wiener measure on $C([0, t]; \mathbf{R}^d)$) which will have reached 2 by time t:

$$P(\tau \leqslant t) = P(\sup_{s\leqslant t} B_s \geqslant 2) = 2P(B_t \geqslant 2) = \sqrt{\frac{2}{\pi}} \int_{\frac{2}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} dy > 0.$$

This probability converges to 1 as $t \to 0$. We say that (X_t) is defined up to τ and τ is called its life time or explosion time.

3.6.2 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 **??**. 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 3.6.2 Let X be a Markov process on $\mathcal{X}$ with transition function $P_{s,t}$. We define a family of probabilities on $\mathbf{R}_+ \times \mathcal{X}$ as below. Letting $z = (s,x) \in \mathbf{R}_+ \times \mathcal{X}$ and $d\bar{z} := d(\bar{s},\bar{x})$,

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

Show that $\hat{P}_h$ is indeed a time-homogeneous transition function and $\hat{X}_t : (Y_t, X_t)$, 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.

3.7 Invariant probability measure

We may define a transition map on $\mathbb{P}(\mathcal{X})$, the space of probability measures :

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

Definition 3.7.1 A probability measure μ on $\mathcal{X}$ is said to be invariant for P_t if $P_t^*\mu = \mu$ for all t. If (X_t) is a Markov process with transition probabilities P_t, then μ is also referred as an invariant probability measure for (X_t).

Definition 3.7.2 Given a transition function (P_t) and $t_0 > 0$, we define a Markov chain with

$$\mathbb{P}(x_{n+1} \in A | \mathcal{F}_n) = P_{t_0}(x_n, A), \qquad \forall A \in \mathcal{B}(\mathcal{X}).$$

It is a Markov chain with transition probabilities $\tilde{P}^n$ constructed as follows/

1. $\tilde{P}^0(x,\cdot) = \delta_x$,
2. $\tilde{P}^1(x,\cdot) = P_{t_0}(x,\cdot)$,
3. For any $n \geqslant 1$ and $x \in \mathcal{X}$, $\tilde{P}^{n+1}(x,A) = \int_{\mathcal{X}} P_{t_0}(y,A)\, \tilde{P}^n(x,dy)$ for all $A \in \mathcal{B}(\mathcal{X})$.

Note that for all purpose, we can assume that $t_0 = 1$. A probability measure is an invariant measure for a discrete Markov chain with transition function (P^n) if and only if $P^*\mu = \mu$.

For the familiar example of a Markov chain (X_n) with discrete state space, for example $\mathcal{X} = N$, the one step Markov transition probabilities are the transition matrices for which an invariant probability measure is a vector such that $\mu P = \mu$ where the left hand side indicating matrix multiplication. On a finite state space, any Markov chain has an invariant probability measure; it has a unique invariant measure if it is irreducible and aperiodic.

From an invariant probability for an induced Markov chain with transition probability P_{t_0}, we can construct an invariant probability for P_t, the measures need not be the same. The following observation allow us to pass results on discrete time Markov processes to continuous time Markov processes.

Proposition 3.7.3 *Suppose that for some time t_0, there exists a probability measure μ with $P_{t_0}^*\mu = \mu$. Then there exists an invariant measure for (P_t). If there exists at most one invariant measure for P_{t_0}, then uniqueness holds for P_t.*

Proof Suppose that for some time t_0, $T_{t_0}^*\mu = \mu$ for some t_0, then $\tilde{\mu} = \frac{1}{t_0}\int_0^{t_0} T_s^*\mu ds$ is an invariant probability measure. It is sufficient to observe that for any $A \in \mathcal{B}(\mathcal{X})$, $s \mapsto T_s^*\mu(A)$ is t_0-periodic, its average is invariant under any shift.

Since every invariant probability measure for T_t is invariant for $T = T_{t_0}$, uniqueness of invariant measure holds fro P_t whenever it holds for any fixed time. □

Theorem 3.7.4 (Krylov-Bogoliubov) *Let $\mathcal{X}$ be a complete separable metric space. Suppose that T_t is Feller and suppose that there exists a $\mu \in \mathbb{P}(\mathcal{X})$ such that the family of measures $\{T_t^*\mu : t \geqslant 0\}$ is tight and that $t \mapsto \int_{\mathcal{X}} T_t \mathbf{1}_A d\mu$ is measurable for every measurable set A. Then there exists an invariant probability measure for T_t.*

Proof That $t \mapsto \int_{\mathcal{X}} T_t \mathbf{1}_A d\mu$ is measurable is equivalent to $t \mapsto T_t^*\mu(A)$ is meaurable. Set

$$\mu_n(A) = \frac{1}{n}\int_0^n T_s^*\mu(A)ds.$$

This $\{\mu_n\}$ is tight, we show that its accumulation points are invariant probability measures. To this end we may assume that $\mu_n \to \pi$ To check $T_t^*\pi = \pi$ for any $t > 0$, we only need to show that $\int_{\mathcal{X}} \varphi d(T_t^*\pi) = \int_{\mathcal{X}} \varphi d\pi$ for any $\varphi \in \mathcal{C}_b(\mathcal{X})$. Since T is Feller, $T\varphi$ is a continuous function, since it is also bounded, the dominated convergence theorem can be used:

$$\int_{\mathcal{X}} \varphi \, d(T_t^*\pi) = \int_{\mathcal{X}} T_t\varphi \, d\pi = \lim_{k\to\infty}\int_{\mathcal{X}} T_t\varphi \, d\mu_{n_k}$$

$$= \lim_{k\to\infty} \frac{1}{n_k} \int_0^{n_k} \int_{\mathcal{X}} T_t \varphi \, dT_s^*(\mu) = \lim_{k\to\infty} \frac{1}{n_k} \int_0^{n_k} \int_{\mathcal{X}} \varphi \, dT_{s+t}^*(\mu) ds$$
$$= \lim_{k\to\infty} \frac{1}{n_k} \int_t^{n_k+t} \int_{\mathcal{X}} \varphi \, dT_s^*(\mu) ds = \int \varphi d\pi,$$

Since φ was also arbitrary, this in turn implies that $T^*\pi = \pi$, concluding the proof. □

On a compact space, every family of probability measures is tight, hence the following Corollary.

Corollary 3.7.5 *If the space $\mathcal{X}$ is compact, then every strongly continuous Feller semigroup on $\mathcal{X}$ has an invariant probability measure.*

Given the link between invariant measures fro P_1 and for P_t, it is appropriate to present examples of Markov chains indexed by **N** and their invariant measures.

Example 3.7.6 Let $\Phi : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ be continuous and bounded. Define the Markov chain by $x_{n+1} = \Phi(x_n, \xi_{n+1})$, with $\mu \sim \xi_k$ iid random variables and $\{x_0, \xi_k, k \geqslant 1\}$ independent. Then if $f \in C_b(\mathcal{X})$,

$$Tf(x) = \mathbb{E}[f(\Phi(x, \xi_{n+1}))] = \int f \circ \Phi(x, y) \mu(dy),$$

then Tf is continuous. Hence (x_n) is Feller and has an invariant probability measure. An example is $x_{n+1} = \sin(x_n + \xi_{n+1})$.

Example 3.7.7 Consider (x_n) a Markov chain on $\mathbf{R}^n$ with initial position x_0. Assume P (equivalently T) is Feller, then there exists an invariant probability measure if any of the following holds:

1) $\sup_{n\geqslant 0} \mathbb{E}[|x_n|^p] < \infty$ for some $p > 0$.
2) $\sup_{n\geqslant 0} \mathbb{E}|\log(|x_n| + 1)| < \infty$.

Proof In these settings we have $P^n(x_0, \cdot) = \mathcal{L}(x_n)$, and tightness for 2) follows from below[1]

$$P^n(x_0, (B_M)^c) = \mathbb{P}(|x_n| > M) \leqslant \sup_{n\geqslant 0} \frac{\mathbb{E}\log(|x_n| + 1)}{\log(M + 1)} \to 0, \qquad \text{as } M \to \infty,$$

where B_M is the closed ball of radius M centred at 0. The proof for 1) is similar. □

[1] Using Markov's inequality with non-negative monotone function $u \mapsto \log(u + 1)$.

Example 3.7.8 (*Tightness*) Suppose $\{\xi_n\}$ are iid and independent of x_0, with $\mathbb{E}|x_0| < \infty$ and Markov chain $x_{n+1} = \frac{1}{2}x_n + \xi_{n+1}$. Assume also $\mathbb{E}|\xi_k| = a < 1$. The chain is Feller (check as in Example 3.7.7). The following arguments shows that the probability distribution of $\{x_n\}$ is tight. For all $n \geqslant 1$

$$\begin{aligned}
\mathbb{E}|x_{n+1}| \leqslant \frac{1}{2}\mathbb{E}|x_n| + \mathbb{E}|\xi_{n+1}| \leqslant \frac{1}{2}\left(\frac{1}{2}\mathbb{E}|x_{n-1}| + a\right) + a \\
= a + \frac{1}{2}a + \frac{1}{4}(|x_{n-2}| + a) \\
\leqslant a + \frac{1}{2}a + \frac{1}{4}a + \cdots + \frac{1}{2^{n+1}}a + \mathbb{E}|x_0| \\
\leqslant 2a + \mathbb{E}|x_0|.
\end{aligned}$$

Hence $\sup_{n\geqslant 0} \mathbb{E}|x_n| < \infty$ and the system has an invariant probability measure. We remark since we only need to show $\{P^n(x_0, \cdot)\}$ is tight for some x_0, we can simply start the chain from a fixed point.

The Lyapunov test function method allows us to use this reasoning for more general systems.

Chapter 4

Stochastic Differential Equations

4.1 Stochastic Integration

In this chapter we review Itô integration (stochastic integration), local martingales, total variations and quadratic variations. Throughout this chapter, we have a probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfying the usual assumptions. Denote by $B_t = (B_t^1, \dots, B_t^n$ an n-dimensional Brownian motion with respect to a filtration $\mathcal{F}_t$, which means that $\{B_t^i\}$ are independent one dimensional Brownian motion.

We begin by defining the elementary integral $\int_0^t K_s dM_s$ where K_t is an elementary process, and very quickly we specialise to the case of $M_t = B_t$, Brownian motion, We seek a class of stochastic processes (f_s) with the property that there exists a sequence of stochastic processes $K^n \in \mathcal{E}$ with K^n converges to f (in some sense), and $\int_0^t K_n(s) dM_s$ converges (in some sense) to a limit, the limit will be a candidate for the Itô integral $\int_0^t f_s dB_s$.

4.2 Elementary Integrals

An elementary stochastic process (with real values) is of the form:

$$K_t(\omega) = K_{-1}(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} K_i(\omega)\mathbf{1}_{(t_i, t_{i+1}]}(t)$$

where $0 = t_0 < \cdots < t_n < \dots$ with $\lim_{i\to\infty} t_i = \infty$, is any sequence of positive numbers increasing to infinity, $K_{-1} \in \mathcal{F}_0, K_i \in \mathcal{F}_{t_i}$, $\sup_i |K_i|$ is bounded. Let $\mathcal{E}$ denotes the collection of elementary processes.

Definition 4.2.1 Let $K \in \mathcal{E}$ and let (M_s) be a stochastic process. The elementary integration is defined by

$$\int_0^t K_s dM_s := \sum_{i=1}^{\infty} K_i (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}).$$

If $t \in [t_n, t_{n+1})$, the elementary integral expands as follows:

$$\int_0^t K_s dM_s = \sum_{i=1}^{n-1} K_i(\omega)(M_{t_{i+1}}(\omega) - M_{t_i}(\omega)) + K_n(\omega)(M_t(\omega) - M_{t_n}(\omega)).$$

Exercise 4.2.2 Given $K \in \mathcal{E}$, compute $\mathbb{E}(K_i K_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j}))$.

Definition 4.2.3 A stochastic process $(X_t, t \in I)$ is L^2 bounded if $\sup_{t\in I} \mathbb{E}[X_t]^2 < \infty$.

Note that a Brownian motion $(B_t, t \leqslant T)$ is an L^2 bounded martingale for on any finite time interval $[0, T]$.

Proposition 4.2.4 *Let B_t be a Brownian motion, and $K \in \mathcal{E}$. Then for any interval $[0, T]$, $(\int_0^t K_s dB_s,)$ is an L^2 bounded continuous martingale, and for any $t > 0$, we have that*

$$\mathbb{E}[(\int_0^t K_s dB_s)^2] = \mathbb{E}[\int_0^t (K_s)^2 ds] \qquad (\textit{Itô isometry}).$$

Proof We may assume that the summation is from 1 to N and $t_{N+1} = t$ and so

$$\int_0^t K_r dB_r = \sum_{i=1}^{N} K_i (B_{t_{i+1}} - B_{t_i}).$$

Note that $K_i \in \mathcal{F}_{t_i}$ and $B_{t_{i+1}\wedge t} - B_{t_i \wedge t}$ is independent fo $\mathcal{F}_{t_i \wedge t}$. Without loss of generality, assume that $t_i < t_j$, and $t_{i+1} \leqslant t_j$.

$$\mathbb{E}(K_i(B_{t_{i+1}} - B_{t_i})K_j(B_{t_{j+1}} - B_{t_j}) = \mathbb{E}\Big[K_i K_j (B_{t_{i+1}} - B_{t_i})\mathbb{E}\big(B_{t_{j+1}} - B_{t_j} \,\Big|\, \mathcal{F}_{t_j})\big)\Big] = 0.$$

So

$$\begin{aligned}\mathbb{E}(\int_0^t K_s dB_s)^2 &= \sum_{i=1}^{\infty} \mathbb{E}(K_i^2 (B_{t_{i+1}\wedge t} - B_{t_i \wedge t}))^2 \\ &= \sum_{i=1}^{\infty} \mathbb{E}(K_i^2 (t_{i+1} \wedge t - t_i \wedge t)) \\ &= \mathbb{E}\int_0^t (K_s)^2 ds.\end{aligned}$$

Let $s < t$, we compute

$$\mathbb{E}\left\{\sum_{i=1}^{N} K_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_s\right\}.$$

Let us analyze the ith interval $I_i = (t_i, t_{i+1}]$.

If $t_{i+1} \leqslant s$, then

$$\mathbb{E}\big\{K_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_s\big\} = K_i(B_{t_{i+1}} - B_{t_i}) = K_i(B_{t_{i+1}\wedge s} - B_{t_i\wedge s}).$$

If $s \leqslant t_i$, then

$$\mathbb{E}\big\{K_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_s\big\} = 0 = K_i(B_{t_{i+1}\wedge s} - B_{t_i\wedge s}).$$

If $t_i < s < t_{i+1}$,

$$\mathbb{E}\{K_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_s\} = K_i\mathbb{E}\{B_{t_{i+1}}|\mathcal{F}_s\} - K_iB_{t_i} = K_i(B_{t_{i+1}\wedge s} - B_{t_i\wedge s}).$$

Summing up the three cases to obtain

$$\mathbb{E}\{\int_0^t K_r dB_r|\mathcal{F}_s\} = \sum_i K_i(B_{t_{i+1}\wedge s} - B_{t_i\wedge s}).$$

and $K \cdot B$ is a martingale. □

Exercise 4.2.5 Show that if (M_t) is an L^2 bounded martingale, then for any elementary process K_t, $\int_0^t K_r dM_r$ is an L^2 bounded martingale.

4.3 Itô integration

Let (B_t) be a Brownian motion.

Definition 4.3.1 We define $L^2(B,T)$ to be the set of progressively measurable process (f_t) such that

$$\|f\|^2_{L^2(B,T)} := \mathbb{E}\int_0^T (f_s)^2 ds < \infty.$$

Proposition 4.3.2 *The set of elementary processes are dense in* $L_2(B,T)$.

Proof We prove the case when $f \in L^2(B)$ is left continuous. First assume f is bounded and let

$$f_n(s,\omega) = f_0(\omega)\mathbf{1}_{\{0\}}(s) + \sum_{j\geqslant 1} f_{\frac{j}{2^n}}(\omega)\mathbf{1}_{\{\frac{j}{2^n}\leqslant s<\frac{j+1}{2^n}\}}.$$

Note that $f_{\frac{j}{2^n}}$ is $\mathcal{F}_{j/2^n}$ measurable. Since f is left continuous and bounded, $|f_n|_\infty \leqslant |f|_\infty$, by the dominated convergence theorem, $f_n \to f$ in $L^2(B,T)$. If f is not bounded, let $f_n(s) = f_s \mathbf{1}_{|f_s|\leqslant n}$. Then

$$\|f_n - f\|^2_{L^2(M)} \leqslant \mathbb{E}\int_0^\infty f_s^2(\omega)\mathbf{1}_{\{(s,\omega):|f(s,\omega)|\geqslant n\}} ds \overset{(n\to\infty)}{\to} 0.$$

Let $g : [0,t] \times \Omega \to \mathbf{R}$ be progressively measurable and such that $\mathbb{E}\int_0^T g^2(s,\omega)ds) < \infty$, it can be approximated by continuous functions in L^2. In fact, setting $g_n(t) = \int_0^t ne^{-n(t-r)}g(r)dr$, then $g_n(\omega) \to g(\omega)$ in L^2. □

This result hold if B is replaced by an L^2-martingale.

Definition 4.3.3 Denote by H^2 the set of L^2 bounded martingales and H_0^2 the subspace of L^2 bounded martingales with initial value 0.

Proposition 4.3.4 *The elementary integration defines a linear map from $\mathcal{E}$ to H_0^2.*

$$I : K \mapsto \int_0^\cdot K_s dB_s$$

Furthermore it is isometric:

$$\left\|\int_0^\cdot K_s dB_s\right\|_{H^2} = \|K\|_{L^2(B,T)}.$$

The linear map $K \to I(K) = \int_0^\cdot K_s dB_s$ extends to $L^2(B,T)$, which is referred as Itô integral or stochastic integration with respect to the Brownian motion. In other words,

Definition 4.3.5 If $f \in L^2(B,T)$ and $\{f_n\} \subset \mathcal{E}$ is a sequence converging to f in $L^2(B,T)$. Then $\int_0^t f_n dB_s$ exists. We define the limit to be $\int_0^t f_s dB_s$. This limit is independent of the choices of the converging sequence.

Exercise 4.3.6 Let $T > 0$. Suppose that $f_n \to f$ in $L^2([0,T]\times\Omega$, show that $\{\int_0^\cdot f_n(s)dB_s\}$ is a Cauchy sequence in H_0^2.

4.4 Local martingales and martingale brackets

Definition 4.4.1 An $\mathcal{F}_t$-adapted stochastic process (X_t) is a local martingale, if there exists a non-decreasing sequence of stopping times $\{T_n\}$ with the property that $\sup_n T_n = \infty$ a.s. and such that for any n, $(X_t^{T_n}\mathbf{1}_{\{T_n>0\}}, t \geqslant 0)$ is a martingale.

If $X_0 = 0$, $X_t^{T_n}\mathbf{1}_{\{T_n>0\}} = X_t^{T_n}$. If $T_n > 0$, then $X_t^{T_n}\mathbf{1}_{\{T_n>0\}} = X_t^{T_n}$. Since $X_t^{T_n}\mathbf{1}_{\{T_n=0\}} = X_0\mathbf{1}_{\{T_n=0\}}$, the role of the multiplier, the indicator function $\mathbf{1}_{\{T_n>0\}}$, is to allow us to not impose integrability assumption M_0.

If X_t is a martingale then $\mathbb{E}[X_t] = \mathbb{E}[X_0]$ for all t. If X_t is a local martingale, this may fail to hold. Furthermore given any function $m(t)$ of bounded variation there is a local martingale such that $m(t)$ is its expectation process. A local martingale which is not a martingale is called a strictly local martingale, otherwise it is a true martingale, see [2] for discussions related to this.

Definition 4.4.2 Let (X_t) and (Y_t) be two continuous processes. If for any sequence of partitions with $|\Delta_n| \to 0$,

$$\lim_{n\to\infty}\sum_{j=0}^{\infty}(X_{t\wedge t_{j+1}^n} - X_{t\wedge t_j^n})(Y_{t\wedge t_{j+1}^n} - Y_{t\wedge t_j^n})$$

exists in probability, we define the limit to be $\langle X, Y\rangle_t$.

In particular,

$$\langle X, X\rangle_t \stackrel{P}{=} \lim_{n\to\infty}\sum_{j=0}^{\infty}(X_{t\wedge t_{j+1}^n} - X_{t\wedge t_j^n})^2.$$

For simplicity, we denote $\langle X, X\rangle_t$ by $\langle X\rangle_t$.

Theorem 4.4.3 *For any continuous local martingales (M_t) and (N_t), there exists a unique continuous process $\langle M, N\rangle_t$ of finite variation vanishing at 0 such that $M_tN_t - \langle M, N\rangle_t$ is a continuous local martingale. This process is referred as the quadratic variation of (M_t) and (N_t).*

We would not indulge in the proof, instead we give some examples.

Theorem 4.4.4 *[Burkholder-Davis-Gundy Inequality] For every $p > 0$, there exist universal constants c_p and C_p such that if $(M_t, t \in [0, T])$ is continuous local martingales with $M_0 = 0$,*

$$c_p\mathbb{E}\Big(\langle M, M\rangle_T^{\frac{p}{2}}\Big) \leqslant \mathbb{E}(\sup_{t\leqslant T}|M_t|)^p \leqslant C_p\mathbb{E}\Big(\langle M, M\rangle_T^{\frac{p}{2}}\Big).$$

We may consider also $(M_t, t < \infty)$ in which case the T in the above inequality holds with T replaced by ∞ or by a stopping time.

Exercise 4.4.5 Let (M_t) is a continuous local martingale with $M_0 = 0$. If $\sup_{t<\infty} M_t \in L^1$ show that (M_t) is a martingale.

Definition 4.4.6 A stochastic process $(X_t, t \in I)$ is L^2 bounded if $\sup_{t\in I}\mathbb{E}[X_t]^2 < \infty$.

Exercise 4.4.7 If H, K is an elementary process, the martingale bracket of their martingale Itô integrals are:

$$\langle \int_0^\cdot K_s dB_s, \int_0^\cdot H_s dB_s \rangle_t = \int_0^t H_s K_s ds.$$

Hint. To show this, by Theorem 4.4.3, it suffices to show that

$$\int_0^\cdot K_s dB_s \int_0^\cdot H_s dB_s - \int_0^t H_s K_s ds$$

is a martingale.

Exercise 4.4.8 Let f_t, g_t be a progressively measurable stochastic processes in $L^2(B, T)$. Then

$$\langle \int_0^\cdot f_s \, dB_s, \int_0^\cdot g_s \, dB_s \rangle_t = \int_0^t f_s g_s ds.$$

Hint: First assume that $f = g$, using the property of H^2, then using polarisation.

4.5 Kunita-Watanabe Inequality*

It is possible to define stochastic integration with respect to a local martingale in the same as we defined Itô integration, first with elementary processes, then extending by density. This requires the Kunita-Watanabe inequality

Recall that $\langle M, M \rangle$ correspond to a positive measure and $\langle M, N \rangle$ a signed measure, written as $\mu^+ - \mu^-$ where μ^+, μ^- are positive measures. By $|\langle M, N \rangle|$ we mean the measure corresponds to $\mu^+ + \mu^-$.

Lemma 4.5.1 *Let* $s \leqslant t$*, we define* $\langle M, N \rangle_t^s := \langle M, N \rangle_t - \langle M, N \rangle_s$.

$$\langle M, N \rangle_t^s \leqslant \sqrt{\langle M, M \rangle_t - \langle M, M \rangle_s} \sqrt{\langle N, N \rangle_t - \langle N, N \rangle_s}.$$

Proof For any a, $\langle M - aN \rangle_t \geqslant 0$. This means $\langle M, M \rangle_t + a^2 \langle N, N \rangle_t \geqslant 2a \langle M, N \rangle_t$. Take $a = \sqrt{\frac{\langle M, M \rangle_t}{\langle N, N \rangle_t}}$ to see that

$$\langle M, N \rangle_t \leqslant \sqrt{\langle M, M \rangle_t \langle N, N \rangle_t}.$$

A similar proof shows that for $s < t$:

$$\langle M, N \rangle_t - \langle M, N \rangle_s \leqslant \sqrt{\langle M, M \rangle_t - \langle M, M \rangle_s} \sqrt{\langle N, N \rangle_t - \langle N, N \rangle_s}.$$

□

Let H_s, K_s be measurable functions by which we mean they are Borel measurable functions from $(\mathbf{R}_+ \times \Omega, \mathcal{F}_\infty \otimes \mathcal{B}(\mathbf{R}_+)$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Approximating them by elementary functions leads to the following theorem:

Theorem 4.5.2 *Let (M_t) and (N_t) be two continuous local martingales. Let (H_t) and (K_t) be measurable processes. Then for $t \leqslant \infty$,*

$$\int_0^t |H_s||K_s|\ d|\langle M,N\rangle|_s \leqslant \sqrt{\int_0^t |H_s|^2 d\langle M,M\rangle_s}\sqrt{\int_0^t |K_s|^2 d\langle N,N\rangle_s}, \qquad a.s.$$

The inequality states in particular that the left hand side is finite if the right hand side is. If furthermore $H \in L^1(d\langle M,M\rangle_s)$ and $K \in L^1(d\langle N,N\rangle_s)$,

$$\left|\int_0^t H_s K_s\ d\langle M,N\rangle_s\right| \leqslant \sqrt{\int_0^t |H_s|^2 d\langle M,M\rangle_s}\sqrt{\int_0^t |K_s|^2 d\langle N,N\rangle_s}, \qquad a.s.$$

Proof Let (H_s) and (K_s) be from $\mathcal{E}$, elementary processes. Let $0 = t_1 < \cdots < t_{N+1}$ be a partition such that on each sub-interval, both $H_s(\omega)$ and $K_s(\omega)$ are constant in s. WW write, for $H_0, K_0 \in \mathcal{F}_0, H_i, K_i \in \mathcal{F}_{t_i}$,

$$H_t(\omega) = H_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=1}^N H_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t),$$

and

$$K_t(\omega) = K_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=1}^N K_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t),$$

Then

$$\begin{aligned}
&\left|\int_0^t H_s(\omega)K_s(\omega)d\langle M,N\rangle_s(\omega)\right| \\
&= \left|\sum_i H_i(\omega)K_i(\omega)\langle M,N\rangle_t^s(\omega)\right| \leqslant \sum_i |H_i(\omega)||K_i(\omega)||\langle M,N\rangle_t^s(\omega)| \\
&\leqslant \sqrt{\sum_i |H_i|^2\langle M,M\rangle_t^s(\omega)}\sqrt{\sum_i |K_i|^2\langle N,N\rangle_t^s(\omega)} \\
&= \left(\int_0^\infty (H_s(\omega))^2 d\langle M,M\rangle_s(\omega)\right)^{\frac{1}{2}}\left(\int_0^\infty (K_s(\omega))^2 d\langle N,N\rangle_s(\omega)\right)^{\frac{1}{2}}.
\end{aligned}$$

Take appropriate limit to see the second inequality holds.

Let $\tilde{H}_s = H_s \operatorname{sign}(H_s K_s)\frac{d\langle M,N\rangle_s}{|d\langle M,N\rangle_s|}$, we see that

$$\int_0^t |H_s||K_s|\ d|\langle M,N\rangle|_s = \int_0^t \tilde{H}_s K_s d\langle M,N\rangle_s$$

and apply the second inequality we see the first inequality holds for all bounded measurable functions (H_s) and (K_s). If they are not bounded, we take a sequence of cut off functions for H_s and K_s to see that first inequality always hold. They may however be infinity. □

Apply Hölder Inequality to the above inequality to obtain the following.

Corollary 4.5.3 *[Kunita-Watanabe Inequality] For $t \leqslant \infty$, and $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\begin{aligned}&\mathbb{E}\int_0^t |H_s||K_s||d\langle M, N\rangle|_s \\ &\leqslant \left(\mathbb{E}\left(\int_0^t |H_s|^2 d\langle M, M\rangle_s\right)^{\frac{p}{2}}\right)^{\frac{1}{p}} \left(\mathbb{E}\left(\int_0^t |K_s|^2 d\langle N, N\rangle_s\right)^{\frac{q}{2}}\right)^{\frac{1}{q}}\end{aligned}$$

4.6 Stochastic integration w.r.t. semi-martingales

A continuous semi-martingale is of the form $X_t = M_t + A_t$, where M_t is a continuous local martingale and A_t a continuous finite variation process. The decomposition into the same is unique, up to a almost surely set. It is referred as the Doob-Meyer decomposition of X_t.

Definition 4.6.1 If $X_t = M_t + A_t$ is a continuous semi-martingale and f is a progressively measurable locally bounded stochastic process, we define

$$\int_0^t f_s dX_s = \int_0^t f_s dM_s + \int_0^t f_s dA_s.$$

Proposition 4.6.2 *Let X, Y be continuous semi-martingales. Let f, g, K be locally bounded and progressively measurable. Let $a, b \in \mathbf{R}$.*

1. $\int_0^t (af_s + bg_s)dX_s = a\int_0^t f_s dX_s + b\int_0^t g_s dX_s$.
2. $\int_0^t f_s d(aX_s + bY_s) = a\int_0^t f_s dX_s + b\int_0^t f_s dY_s$.
3. $$\int_0^t f_s d\left(\int_0^s g_r dX_r\right) = \int_0^t f_s g_s dX_s.$$
4. *For any stopping time τ,*

$$\int_0^\tau K_s dX_s = \int_0^\infty \mathbf{1}_{s\leqslant\tau} K_s dX_s = \int_0^\infty K_s dX_s^\tau.$$

5. *If X_s is of bounded total variation on $[0,t]$ so is the integral $\int_0^\cdot K_s dX_s$; and if X_s is a local martingale so is $\int K_s dX_s$.*

Also note that for continuous processes, Riemann sums corresponding to a sequence of partitions whose modulus goes to zero converges to the stochastic integral **in probability**. Note that this convergence does not help with computation. Although there are sub-sequences that converges a.s. we do not know which subsequence of the partition would work and this subsequence would be likely to differ for different integrands and different times.

Proposition 4.6.3 *If (K_t) is left continuous and $\Delta^n : 0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = t$ is a sequence of partition of $[0,t]$ such that their modulus goes to zero, then*

$$\int_0^t K_s dX_s = \lim_{n\to\infty} \sum_{i=1}^{N_n} K_{t_i^n}(X_{t_{i+1}^n} - X_{t_i^n}).$$

The sum converges in probability.

4.7 Itô's formula

Consider an $\mathbf{R}^n$-valued stochastic process $(X_t = (X_t^1, \dots, X_t^n)$. Suppose that it is semi-martingale and H_s a process so that $\int_0^t H_s dX_s$ is defined. We set

$$\int_s^t H_s dX_s = \int_0^t H_s dX_s - \int_0^s H_s dX_s.$$

We denote by $\langle X, X\rangle_t$ the matrix valued process whose entries are $\langle X^i, X^j\rangle_t$.

Let $B_t = (B_t^1, \dots, B_t^m)$ be an m-dimensional Brownian motion. Let $\sigma_k : \mathbf{R}_+ \times \mathbf{R}^n \to \mathbf{R}^n$, $k = 1, \dots, m$, be vector fields on $\mathbf{R}^d$. We consider the stochastic differential equation

$$dx_t = \sum_{k=1}^m \sigma_k(t, x_t) dB_t^k + \sigma_0(t, x_t) dt. \tag{4.1}$$

Proposition 4.7.1 (The product formula) *If X_t and Y_t are real valued semi-martingales,*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y\rangle_t$$

This also provides an understanding, even serve as an definition, for the bracket process,

$$\langle X, Y\rangle_t = X_t Y_t - X_0 Y_0 - \int_0^t X_s dY_s - \int_0^t Y_s dX_s$$

Example 4.7.2 If $B_t = (B_t^2, \dots, B_t^n)$ is an n-dimensional BM, then $\langle B^i, B^j \rangle_t = \delta_{ij} t$, $|B_t|^2 = \sum_i |B_t^i|^2$, and

$$|B_t|^2 = 2 \sum_{i=1}^n \int_0^t B_s^i dB_s^i + nt.$$

Theorem 4.7.3 (Itô's Formula) *Let $X_t = (X_t^1, \dots, X_t^n)$ be a $\mathbf{R}^n$-valued sample continuous semi-martingale and f a C^2 real valued function on $\mathbf{R}^n$ then for $s < t$,*

$$f(X_t) = f(X_s) + \sum_{i=1}^n \int_s^t \frac{\partial f}{\partial x_i}(X_r) dX_r^i + \frac{1}{2} \sum_{i,j=1}^n \int_s^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_r) d\langle X^i, X^j \rangle_r.$$

In short hand,

$$f(X_t) = f(X_s) + \int_s^t (Df)(X_r) dX_r + \frac{1}{2} \int_s^t (D^2 f)(X_r) d\langle X, X \rangle_r.$$

Sketch proof: By the Taylor expansion for C^2 function $f : \mathbf{R} \to \mathbf{R}$:

$$\begin{aligned} f(y) &= f(y_0) + f'(y_0)(y - y_0) + \int_{y_0}^y (y - z) f''(z) dz \\ &= f(y_0) + f'(y_0)(y - y_0) + \frac{1}{2} f''(y_0)(y - y_0)^2 + \int_{y_0}^y (y - z)(f''(z) - f''(y_0)) dz. \end{aligned}$$

The remainder term satisfies the bound

$$\left| \int_{y_0}^y (y - z)(f''(z) - f''(y_0)) dz \right| \leqslant (y - y_0)^2 \sup_{z \in [y_0, y]} |f''(z) - f''(y_0)|.$$

Then Itô's formula follows from

$$\begin{aligned} f(X_t) - f(X_s) = &\sum_{i=0}^{N(n)-1} \Big(f(X_{t_{i+1}^n}) - f(X_{t_i^n}) \Big) \\ = &\sum_{i=0}^{N(n)-1} f'(X_{t_i^n}) \Big(X_{t_{i+1}^n} - X_{t_i^n} \Big) + \frac{1}{2} \sum_{i=0}^{N(n)-1} f''(X_{t_i^n}) \Big(X_{t_{i+1}^n} - X_{t_i^n} \Big)^2 \\ &+ \sum_{i=0}^{N(n)-1} \Big(R(X_{t_{i+1}^n}, X_{t_i^n}) \Big). \end{aligned}$$

It is easy to see that the remainder terms,

$$\sum_{i=0}^{N(n)-1} \sup_{t \in [t_i^n, t_{i+1}^n]} \Big| f''(X_{t_{i+1}^n}) - f''(X_{t_i^n}) \Big| \Big(X_{t_{i+1}^n} - X_{t_i^n} \Big)^2,$$

converges to zero as the partition size is taken to zero.

Remark 4.7.4 If T is a stopping time, apply Itô's formula to $Y_t = X_{T\wedge t}$ to see that

$$f(X_{T\wedge t}) = f(X_0) + \int_0^{T\wedge t} (Df)_{X_r} dX_r + \frac{1}{2}\int_0^{T\wedge t} (D^2 f)_{x_r} d\langle X, X\rangle_r.$$

Example 4.7.5 If $B_t = (B_t^2, \dots, B_t^n)$ is an n-dimensional BM, then $\langle B^i, B^j\rangle_t = \delta_{ij} t$, $|B_t|^2 = \sum_i |B_t^i|^2$, and

$$|B_t|^2 = 2\sum_{i=1}^n \int_0^t B_s^i dB_s^i + nt.$$

Example 4.7.6 Let (M_t) be a continuous semi-martingale. Then $X_t = e^{M_t - \frac{1}{2}\langle M,M\rangle_t}$ satisfies the equation:

$$X_t = e^{M_0} + \int_0^t X_s dM_s.$$

Let $Y_t := M_t - \frac{1}{2}\langle M, M\rangle_t$, then $\langle Y, Y\rangle_t = \langle M, M\rangle_t$, and $X_t = e_y^Y$. Let $f(x) = e^x$ and apply Itô's formula to the function f and the process (Y_t),

$$\begin{aligned} X_t = e^{Y_t} =& e^{Y_0} + \int_0^t e^{Y_s} dY_s - \frac{1}{2}\int_0^t e^{Y_s} d\langle Y, Y\rangle_s \\ =& e^{M_0} + \int_0^t e^{Y_s} dM_s + \frac{1}{2}\int_0^t e^{Y_s} d\langle M, M\rangle_s - \frac{1}{2}\int_0^t e^{Y_s} d\langle Y, Y\rangle_s \\ =& e^{M_0} + \int_0^t X_s dM_s. \end{aligned}$$

Definition 4.7.7 If (M_t) is a continuous local martingale, $e^{M_t - \frac{1}{2}\langle M,M\rangle_t}$ is a continuous local martingale and is called the exponential martingale of M_t.

Theorem 4.7.8 *Let (X_t) be a continuous semi-martingale. Assume that $\frac{\partial}{\partial t}F(t,x)$ and $\frac{\partial^2}{\partial x_i \partial x_j}F(t,x)$, $i, j = 1, \dots, d$, exist and are continuous functions. Then*

$$\begin{aligned} F(t, X_t) =& F(0, X_0) + \int_0^t \frac{\partial F}{\partial s}(s, X_s) ds + \int_0^t DF(s, X_s) dX_s \\ &+ \frac{1}{2}\int_0^t D^2 F(s, X_s) d\langle X_s, X_s\rangle. \end{aligned}$$

4.8 Stochastic Differential Equations

Let us fix a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Let $B_t = (B_t^1, \dots, B_t^m)$ be an standard $\mathcal{F}_t$-Brownian motion on $\mathbf{R}^n$ (with $B_0 = 0$). Let $\sigma_i, \sigma_0 : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$ be locally bounded Borel measurable functions.

Definition 4.8.1 An adapted continuous stochastic process (x_t), with initial condition x_0, is said to be a (strong) solution to

$$dx_t = \sum_{i=1}^{m} \sigma_i(t, x_t) dB_t^i + \sigma_0(t, x_t) dt, \tag{4.2}$$

if for any $t > 0$, the following makes sense and holds almost surely

$$x_t = x_0 + \sum_{k=1}^{m} \int_0^t \sigma_k(s, x_s) dB_s^k + \int_0^t \sigma_0(s, x_s) ds.$$

We allow solution (x_t) to be defined up to a life time $t < \tau(x_0)$. A solution is defined up to time τ, if for all stopping times $T < \tau$, the following makes sense and holds almost surely

$$x_T = x_0 + \sum_{k=1}^{m} \int_0^T \sigma_k(s, x_s) dB_s^k + \int_0^T b(s, x_s) ds.$$

Remark 4.8.2 This concept of a local solution can be incorporated into the above definition by introducing the one point compactification, $\mathbf{R}^d \cup \{\Delta\}$, of $\mathbf{R}^d$, where Δ is an alien state. The compactification is a topological space with the open sets to consist of open sets of $\mathbf{R}^d$ and sets of the form $(\mathbf{R}^d \backslash K) \cup \{\Delta\}$ where K denotes a compact subset of $\mathbf{R}^d$. Given a process $(X_t, t < \tau)$ on $\mathbf{R}^d$ we define a process $(\hat{X}_t, t \geqslant 0)$ on $\mathbf{R}^d \cup \{\Delta\}$:

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

If $(X_t, t < \tau)$ is a continuous process on $\mathbf{R}^d$, then $(\hat{X}_t)$ is a continuous process on $\mathbf{R}^d \cup \Delta$. Define $\hat{W}(\mathbf{R}^d) \equiv C([0, T]; \mathbf{R}^d \cup \Delta)$ whose elements satisfy that: if $Y_s = \Delta$ then $Y_t(\omega) = \Delta$ for all $t \geqslant s$. The last condition means that once a process enters Δ, it does not return.

Definition 4.8.3 The SDE is said to have no explosion, if for any initial condition, there exists a solution defined a.s. for all time t. Otherwise it explodes in finite time with positive probability.

Definition 4.8.4 The SDE is said to be pathwise unique, if for any two solutions (X_t) and (Y_t) with the same initial condition $X_0 = Y_0$ a.s., then $X_t = Y_t$ a.s. for all time t.

4.9 Basic existence and uniqueness theorem

Theorem 4.9.1 *Suppose that $\sigma_i, b : \mathbf{R}^d \to \mathbf{R}^d$ are Lipschitz continuous. Then for each $x_0 \in \mathbf{R}^d$ there exists a unique continuous $\mathcal{F}_t^B$-adapted stochastic process such that*

$$x_t = x_0 + \sum_{k=1}^m \int_0^t \sigma_k(x_s) dB_s^k + \int_0^t b(x_s) ds$$

for all t a.s. Furthermore for each t, x_t is $\mathcal{F}_t^B$-measurable.

Proof Fix $T > 1$. Define, for all $t \in [0, T]$,

$$\begin{aligned} x_t^{(0)} &= x_0, \\ x_t^{(n)} &= x_0 + \sum_{k=1}^m \int_0^t \sigma_k(x_r^{(n-1)}) dB_r^k + \int_0^t b(x_r^{(n-1)}) dr. \end{aligned} \tag{4.3}$$

We note that,

$$\begin{aligned} \mathbb{E} \sup_{t \leqslant u} |x_t^{(1)} - x_0|^2 &\leqslant 2\mathbb{E} \sup_{t \leqslant u} \left| \sum_{k=1}^m \int_0^t \sigma_k(x_0) B_t^k \right|^2 + 2u\mathbb{E}|b(x_0)|^2 \\ &\leqslant 2^m \sum_{k=1}^m \mathbb{E} \left| \sigma_k(x_0) B_u^k \right|^2 + 2u|b(x_0)|^2 \\ &\leqslant 2^m \sum_{k=1}^m \tilde{C}^2 (1 + |x_0|)^2 \mathbb{E}(B_u^k)^2 + 2u\tilde{C}^2(1 + |x_0|^2) \\ &= (2^m mu + 2u)(\tilde{C})^2 (1 + |x_0|)^2 = C_0, \end{aligned}$$

where $\tilde{C}$ is the common linear growth constants for σ_k and b. By induction and analogous estimation, $\mathbb{E} \sup_{t \leqslant u} |x_t^{(n)}|^2$ is finite and the stochastic integrals make sense. By construction each $(x_t^{(n)})$ is sample continuous and is adapted to the filtration of (B_t).

We estimate the differences between iterations:

$$\begin{aligned} &\mathbb{E} \sup_{s \leqslant t} |x_s^{(n+1)} - x_s^{(n)}|^2 \\ &= \mathbb{E} \sup_{s \leqslant t} \left| \sum_{k=1}^m \int_0^s \left(\sigma_k(x_r^{(n)}) - \sigma(x_r^{(n-1)} \right) dB_r^k + \int_0^s \left(b(x_r^{(n)}) - b(x_r^{(n-1)} \right) dr \right|^2. \\ &\leqslant 2\mathbb{E} \sup_{s \leqslant t} \left(\sum_{k=1}^m \left| \int_0^s \left(\sigma_k(x_r^{(n)}) - \sigma_k(x_r^{(n-1)} \right) dB_r^k \right| \right)^2 + 2\mathbb{E} \sup_{s \leqslant t} \left| \int_0^s \left(b(x_r^{(n)}) - b(x_r^{(n-1)}) \right) dr \right|^2 \\ &\leqslant 2^m \sum_{k=1}^m \mathbb{E} \sup_{s \leqslant t} \left| \int_0^s \left(\sigma_k(x_r^{(n)}) - \sigma_k(x_r^{(n-1)} \right) dB_r^k \right|^2 + 2\mathbb{E} \sup_{s \leqslant t} \left| \int_0^s \left(b(x_r^{(n)}) - b(x_r^{(n-1)}) \right) dr \right|^2. \end{aligned}$$

Let K be the common Lipschitz constant for σ_k and b. By Lemma **??**,

$$\begin{aligned}
&\mathbb{E}\sup_{s\leqslant t}|x_s^{(n+1)}-x_s^{(n)}|^2\\
&\leqslant 2^m C\sum_{k=1}^m \mathbb{E}\left(\int_0^t \left|\sigma_k(x_r^{(n)})-\sigma_k(x_r^{(n-1)})\right|^2 dr\right)+2T\mathbb{E}\int_0^t \left|b(x_r^{(n)})-b(x_r^{(n-1)})\right|^2 dr.\\
&\leqslant 2^m C\sum_{k=1}^m K^2\int_0^t \mathbb{E}\left|x_r^{(n)}-x_r^{(n-1)}\right|^2 dr+2TK^2\int_0^t \mathbb{E}\left|x_r^{(n)}-x_r^{(n-1)}\right|^2 dr.
\end{aligned}$$

Let

$$D=2^m CmK^2+2TK^2,$$

Then

$$\begin{aligned}
\mathbb{E}\sup_{s\leqslant t}|x_s^{(n+1)}-x_s^{(n)}|^2 &\leqslant D\int_0^t \mathbb{E}\left|x_r^{(n)}-x_r^{(n-1)}\right|^2 dr\\
&\leqslant D\int_0^t \mathbb{E}\sup_{r\leqslant s_1}\left|x_r^{(n)}-x_r^{(n-1)}\right|^2 ds_1\\
&\leqslant D^2\int_0^t\int_0^{s_1} \mathbb{E}\sup_{r\leqslant s_2}\left|x_r^{(n-1)}-x_r^{(n-2)}\right|^2 ds_2 ds_1\\
&\leqslant D^n\int_0^t\int_0^{s_1}\dots\int_0^{s_{n-1}} \mathbb{E}\sup_{r\leqslant s_n}\left|x_r^{(1)}-x_r^{(0)}\right|^2 ds_n\dots ds_2 ds_1.
\end{aligned}$$

By induction we see that

$$\mathbb{E}\sup_{s\leqslant t}|x_s^{(n+1)}-x_s^{(n)}|^2\leqslant C_1\frac{D^nT^n}{n!}.$$

whee $C_1=T\mathbb{E}\sup_{t\leqslant T}|x_t^{(1)}-x_0|^2\leqslant TC_0$. By Minkowski inequality,

$$\sqrt{\mathbb{E}\left(\sum_{k=1}^\infty \sup_{s\leqslant t}|x_s^{(k+1)}-x_s^{(k)}|\right)^2}\leqslant \sum_{k=1}^\infty\left(\mathbb{E}\sup_{s\leqslant t}|x_s^{(k+1)}-x_s^{(k)}|^2\right)^{\frac12}<\infty.$$

By Fatou's lemma,

$$\begin{aligned}
&\sum_{k=1}^\infty\left(\mathbb{E}\sup_{s\leqslant t}|x_s^{(k+1)}-x_s^{(k)}|^2\right)^{\frac12}\\
&\leqslant \sum_{k=1}^\infty\left(\mathbb{E}\sup_{s\leqslant t}|x_s^{(k+1)}-x_s^{(k)}|^2\right)^{\frac12}\leqslant C_1\sum_{k=1}^\infty\sqrt{\frac{D^kT^k}{k!}}<\infty.
\end{aligned}$$

In particular for almost surely all ω,

$$\sum_{k=1}^\infty \sup_{s\leqslant t}|x_s^{(k+1)}(\omega)-x_s^{(k)}(\omega)|<\infty.$$

For such ω, $\{x_s^{(n)}(\omega)\}$ is a Cauchy sequence in $C([0,t];\mathbf{R}^d)$. Let $x_t(\omega)=\lim_{n\to\infty} x_t^{(n)}(\omega)$. The process is continuous in time by the uniform convergence.

We take $n\to\infty$ in

$$x_t^{(n)} = x_0 + \sum_{k=1}^m \int_0^t \sigma_k(x_r^{(n-1)})dB_r^k + \int_0^t b(x_r^{(n-1)})dr.$$

As $n\to\infty$, $\int_0^t \sigma_k(x_s^{(n)})dB_s^k \to \int_0^t \sigma(x_s)dB_s$ in probability. There will be an almost sure convergent subsequence and this proves that

$$x_t = x_0 + \sum_{k=1}^m \int_0^t \sigma_k(x_s)dB_s^k + \int_0^t b(x_s)ds.$$

Since each $(x_t^{(n)})$ is a adapted to the filtration of (B_t), so is its limit.

(2) Uniqueness. Let (x_t) and (y_t) be two solutions with $x_0 = y_0$ a.s. Let C^* be a constant.

$$\begin{aligned}
&\mathbb{E}\sup_{s\leqslant t}|x_s-y_s|^2\\
&= \mathbb{E}\sup_{s\leqslant t}\left|\sum_{k=1}^m\int_0^s \Big(\sigma_k(x_r)-\sigma(y_r)\Big)dB_r^k + \int_0^s \Big(b(x_r)-b(y_r)\Big)dr\right|^2\\
&\leqslant \mathbb{E}\sup_{s\leqslant t}\left(\sum_{k=1}^m\left|\int_0^s \Big(\sigma_k(x_r)-\sigma_k(y_r)\Big)dB_r^k\right| + \left|\int_0^s \Big(b(x_r)-b(y_r)\Big)dr\right|\right)^2\\
&\leqslant 2^m\sum_{k=1}^m \mathbb{E}\left(\sup_{s\leqslant t}\left|\int_0^s \Big(\sigma_k(x_r)-\sigma_k(y_r)\Big)dB_r^k\right|\right)^2 + 2\mathbb{E}\left(\sum_{k=1}^m\sup_{s\leqslant t}\left|\int_0^s \Big(b(x_r)-b(y_r)\Big)dr\right|\right)^2\\
&\leqslant 2^m C^*\sum_{k=1}^m \mathbb{E}\left(\int_0^t |\sigma_k(x_r)-\sigma_k(y_r)|^2dr\right) + 2T\mathbb{E}\int_0^t |b(x_r)-b(y_r)|^2dr\\
&\leqslant 2^m C^*\sum_{k=1}^m K^2\int_0^t \mathbb{E}|x_r-y_r|^2dr + 2TK^2\int_0^t \mathbb{E}|x_r-y_r|^2dr\\
&\leqslant (2^m m C^* K^2 T + 2TK^2)\int_0^t \mathbb{E}\left(\sup_{r\leqslant s}|x_r-y_r|^2\right)ds,
\end{aligned}$$

By Grownall's inequality,

$$\mathbb{E}\sup_{s\leqslant t}|x_s-y_s|^2 = 0.$$

In particular, $\sup_{s\leqslant t}|x_s-y_s|^2=0$ almost surely. □

Lemma 4.9.2 (Grownall's Inequality/Gronwall's Lemma) *Let $T>0$. Suppose that $f:[0,T]\to\mathbf{R}_+$ is a locally bounded Borel function such that there are two real numbers*

C and K such that for all $0 \leqslant t$,

$$f(t) \leqslant C + K \int_0^t f(s)ds.$$

Then

$$f(t) \leqslant Ce^{Kt}, \quad t \leqslant T$$

In particular if $C = 0$, $f(t) = 0$ for all $t \leqslant T$.

Definition 4.9.3 A solution $(x_t, t < \tau)$ of an SDE is a maximal solution if $(y_t, t < \bar{\tau})$ is any other solution on the same probability space with the same driving noise and with $x_0 = y_0$ a.s., then $\tau \geqslant \bar{\tau}$ a.s.. We say that τ is the explosion time or the life time of (x_t).

By localisation, or cut off the functions σ_k we have the following theorem:

Theorem 4.9.4 *Suppose that for $k = 1, \ldots, m$, $\sigma_k : \mathbf{R}^d \to \mathbf{R}^d$ and $b : \mathbf{R}^d \to \mathbf{R}^d$ are locally Lipschitz continuous , i.e. for each $N \in \mathbf{N}$, there exists a number K_N such that for all x, y with $|x| \leqslant N, |y| \leqslant N$,*

$$|\sigma_k(x) - \sigma_k(y)| \leqslant K_N|x - y|, \qquad |b(x) - b(y)| \leqslant K_N|x - y|.$$

Then there is a maximal solution $(x_t, t < \tau)$. If $(x_t, t < \tau)$ and $(y_t, t < \zeta)$ be two maximal solutions with the same initial value $x \in \mathbf{R}^d$, then $\tau = \zeta$ a.s. and (x_t) and (y_t) are indistinguishable.

Example 4.9.5 Consider $\dot{x}(t) = ax(t)$ on $\mathbf{R}$ where $a \in \mathbf{R}$. Let $x_0 \in \mathbf{R}$. Then $x(t) = x_0 e^{at}$ is a solution with initial value x_0. It is defined for all $t \geqslant 0$.

Let $\varphi_t(x_0) = x_0 e^{at}$. Then $(t, x) \mapsto \varphi_t(x)$ is continuous and $\varphi_{t+s}(x_0) = \varphi_t(\varphi_s(x))$.

Example 4.9.6 Linear Equation. let $a, b \in \mathbf{R}$. Let $d = m = 1$. Then

$$x(t) = x_0 e^{aB_t - \frac{a^2}{2}t + bt}$$

solves

$$dx_t = a\, x_t\, dB_t + b\, x_t\, dt, \quad x(0) = x_0.$$

The solution exists for all time.

Is this solution unique? The answer is yes. Let y_t be a solution starting from the same point, we could compute and prove that $\mathbb{E}|x_t - y_t|^2 = 0$ for all t, which implies that $x_t = y_t$ a.s. for all t.

Example 4.9.7 Consider a particle of mass 1, subject to a force which is proportional to its own speed, is subject to $\dot{v}_t = -kv_t$. Its random perturbation equation is the Langevin equation:

$$dv_t(\omega) = -kv_t(\omega)dt + dB_t(\omega).$$

For each realisation of the noise (that means for each ω), the solution is an Ornstein-Uhlenbeck process,

$$v_t(\omega) = v_0 e^{-kt} + \int_0^t e^{-k(t-r)} dB_r(\omega).$$

Apply Itô's formula to $e^{-kt}x_t$ we obtain:

$$\begin{aligned} e^{kt}x_t &= x_0 + \int_0^t k e^{ks} x_s ds + \int_0^t e^{ks} dx_s \\ &= x_0 + \int_0^t e^{ks} dB_s. \end{aligned}$$

Multiply both sides by e^{-kt} to conclude.

Example 4.9.8 (1) Small Perturbation. Let $\epsilon > 0$ be a small number,

$$x_t^\epsilon = x_0 + \int_0^t b(x_s^\epsilon) ds + \epsilon B_t.$$

As $\epsilon \to 0$, $x_t^\epsilon \to x_t$. (Exercise)

(2) Let $y_t^\epsilon = y_0 + \epsilon \int_0^t b(y_s^\epsilon) ds + \sqrt{\epsilon} W_t$. Assume that b are bounded, as $\epsilon \to 0$, y_t^ϵ on any finite time interval converges uniformly in time on any finite time interval $[0, t]$, $\mathbb{E} \sup_{0 \leqslant s \leqslant t} (y_s^\epsilon - y_0) \to 0$.

It is worth noticing that Itô's formula and Itø's isometry work with stopping times. Itô isometry may fail for random non-stopping times.

Example 4.9.9 Let $\tau^1 = \inf\{t \geqslant 0 : \sup_{0 \leqslant s \leqslant t} |B_s| \geqslant 1\}$.

$$S = \begin{cases} 0, & \sup_{0 \leqslant s \leqslant t} |B_s| \leqslant 1 \\ \tau^1, & \tau^1 \leqslant 1. \end{cases}$$

Then $B_S = 1$ if $\tau^1 \leqslant 1$ and $B_S = B_0 = 0$ otherwise. Then

$$\mathbb{E}(\int_0^S dB_s) = \mathbb{E}(B_S \mathbf{1}_{\{\tau^1 \leqslant 1\}} = \mathbb{P}(\tau^1 \leqslant 1) \neq 0.$$

4.10 Stratonovich integral

Let $\Delta : 0 = t_0 < t_1 < \cdots < t_N = t$ be a sequence of partition of $[0, t]$ and denote: $|\Delta| = \max_i |t_{i+1} - t_i|$. We have seen that

$$\int_0^t f(s)dB_s = \lim_{|\Delta|\to 0} \sum_i f(t_i)(B_{t_{i+1}} - B_{t_i}),$$

where the limit is taken in probability.

On the other hand, if f and g are continuous and g is a function of finite total variation, then the Riemann-Stieljes integral is defined by

$$\int_0^t f(s)dg_s = \lim \sum f(s_j^*)(g_{s_{j+1}} - g_{s_j})$$

where the limit is independent of the choices of $s_j \in [s_j, s_{j+1}]$. In general, an integral $\int_0^t f_s dg_s$ can be defined as a continuous map from $\mathbf{C}^\alpha \times \mathbf{C}^\beta \to \mathbf{C}^\gamma$ if and only if $\alpha + \beta > 1$. Consequently, $\int_0^t f(s)dB_s$ is not in general an almost sure limit. The choice of the evaluation point for the integrand is also relevant. The sum

$$\sum_i \frac{1}{2}(f(t_{i+1} + f(t_i))(B_{t_{i+1}} - B_{t_i}) = \sum_i f(t_i)(B_{t_{i+1}} - B_{t_i}) + \frac{1}{2}\sum_i (f(t_{i+1} - f(t_i))(B_{t_{i+1}} - B_{t_i}),$$

the first sum converges to the Itô integral, while the second to the bracket $\frac{1}{2}\langle f, B\rangle_t$.

Definition 4.10.1 Let (x_t) be continuous semi-martingale and (y_t) a stochastic process such that the integral below makes sense. The Stratonovich integral is defined as:

$$\int_0^t x_s \circ dx_s := \int_0^t y_s dx_s + \frac{1}{2}\langle x, y\rangle_t.$$

Note that if x_t, y_t be real valued semi-martingales, $f : \mathbf{R} \to \mathbf{R}$ be $\mathbf{C}^2$, then

$$\langle f(y_\cdot), x\rangle_t = \int_0^t f'(y_s)d\langle y, x\rangle_s.$$

Particular interesting is the case where $x_t = B_t$ be a Brownian motion, and y_t a solution to

$$dy_t = Y(y_t)dB_t + Y_0(y_t)dt.$$

Then,

$$\langle f(y_\cdot), B\rangle_t = \int_0^t f'(y_s)Y(y_s)ds.$$

Lemma 4.10.2 *Let $B_t = (B_t^1, \dots, B_t^m)$ be a m-dimensional Brownian motion. There is the following Itô's formula for solution of the equation*

$$dx_t = \sum_{j=1}^m X_j(x_t) \circ dB_t^j + X_0(x_t)dt.$$

Let $f : \mathbf{R}^d \to \mathbf{R}$ be any $\mathbf{C}^3$ function, and $X_j \in \mathbf{C}^2$ and X_0 Lipschitz continuous such that there is a global solution. Then,

$$f(x_t) = f(x_0) + \int_0^t Df(x_s)(X(x_s)dB_s) + \int_0^t \mathcal{L}f(x_s)ds \tag{4.4}$$

where

$$\mathcal{L}f(x) = \frac{1}{2}\sum X_i X_i f + X_0 f$$

and $X(x_s)dB_s$ is shorthand for $\sum_j X_j(x_s)dB_s^j$.

For solutions of the Itô integral

$$dx_t = \sum_{j=1}^m X_j(x_t) \circ dB_t^j + X_0(x_t)dt,$$

(4.4) holds with

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

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

4.11 weak solution, explosion, and uniqueness*

Definition 4.11.1 A d-dimensional stochastic process $(x_t, t < \tau)$, where $\tau \leqslant \infty$, on a probability space $(\Omega, \mathcal{G}, P)$ is a (weak) solution to the SDE (of Markovian type)

$$dx_t = \sigma(t, x_t)dB_t + b(t, x_t)dt. \tag{4.5}$$

If there exists a filtration $(\mathcal{F}_t)$ such that

(1) x_t is adapted to $\mathcal{F}_t$,

(2) a $\mathcal{F}_t$ Brownian motion $B_t = (B_t^1, \dots, B_t^m)$ with $B_0 = 0$;

(3) for all stopping times $T < \tau$, the following makes sense and holds almost surely

$$x_T = x_0 + \sum_{k=1}^m \int_0^T \sigma_k(s, x_s)dB_s^k + \int_0^T b(s, x_s)ds.$$

We may replace (3) by (3')

(3') an adapted continuous stochastic process $x.$ in $C([0,\infty); \mathbf{R}^d \cup \{\Delta\})$, s.t. for all $t \geqslant 0$,

$$x_t = x_0 + \sum_{k=1}^m \int_0^t \sigma_k(s, x_s) dB_s^k + \int_0^t b(s, x_s) ds, a.s.$$

In essence the SDE holds on $\{t < \tau(\omega)\}$. The maximal time τ, up to which a solution is defined is the explosion time, the solution $(x_t, t < \tau)$ is the maximal solution.

Definition 4.11.2 A solution is a global solution is its life time infinite. We say that the SDE does not explode from x_0 if its solution from x_0 is global. We say that the SDE does not explode if all of its solutions are global.

Definition 4.11.3 A solution (x_t, B_t) on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is said to be a strong solution, if x_t is adapted to the filtration of B_t for each t. By a weak solution we mean one which is not strong.

Definition 4.11.4 If, whenever (x_t) and $(\tilde{x}_t)$ are two solutions with $x_0 = \tilde{x}_0$ almost surely, the probability distribution of $\{x_t : t \geqslant 0\}$ is the same as the probability distribution of $\{\tilde{x}_t, t \geqslant 0\}$, we say that uniqueness in law holds.

Uniqueness in law implies the following stronger conclusion: whenever x_0 and $\tilde{x}_0$ have the same distribution, the corresponding solutions have the same law.

Definition 4.11.5 We say pathwise uniqueness of solution holds for an SDE, If whenever (x_t) and $(\tilde{x}_t)$ are two solutions for the SDE on the same probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with the same driving Brownian motion (B_t) and same initial data ($x_0 = \tilde{x}_0$ a.s.), then $x_t = \tilde{x}_t$ for all $t \geqslant 0$ almost surely.

Before giving an example, we state Lévy's martingale characterization Theorem. In dimension 1, it is as follows. An $\mathcal{F}_t$ adapted continuous real valued stochastic process B_t vanishing at 0 is a standard $\mathcal{F}_t$-Brownian motion if and only if (B_t) is an $\mathcal{F}_t$-martingale with quadratic variation t.

Theorem 4.11.6 *Let T be a finite stopping time. Then $(B_{T+s} - B_T, s \geqslant 0)$ is a Brownian motion.*

Definition 4.11.7 An n dimensional stochastic process $(X_t^1, \dots, X_t^n)$ is a $\mathcal{F}_t$ local-martingale if each component is a $\mathcal{F}_t$ local-martingale.

Theorem 4.11.8 *[Lévy's Martingale Characterization Theorem] An $(\mathcal{F}_t)$ adapted sample continuous stochastic process (B_t) in $\mathbf{R}^d$ vanishing at 0 is a $(\mathcal{F}_t)$-Brownian motion if and only if each (B_t) is a $(\mathcal{F}_t)$ local martingale and $\langle B^i, B^j\rangle_t = \delta_{i,j} t$.*

Theorem 4.11.9 (Dambis, Dubins-Schwartz) *Let $\mathcal{F}_t$ be a right continuous filtration. Let (M_t) be a continuous local martingale vanishing at 0 such that $\langle M, M\rangle_\infty = \infty$. Define*

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

Then M_{T_t} is an $\mathcal{F}_{T_t}$ Brownian motion and $M_t = B_{\langle M,M\rangle_t}$ a.s..

The condition on the bracket assures that the time change T_t is almost surely finite for all t. Apply Lévy's Characterization Theorem, Theorem **??**, for Brownian motions.

Example 4.11.10 (Tanaka's SDE) Let

$$\mathrm{sign}(x) = \begin{cases} -1, & \text{if } x \leqslant 0 \\ 1, & \text{if } x > 0. \end{cases}$$

and consider the following Tanaka's SDE, defined on $\mathbf{R}$,

$$dx_t = \mathrm{sign}(x_t) dB_t.$$

If (x_t) is a solution of Tanaka's SDE with initial x_0, then $x_t - x_0 = \int_0^t \mathrm{sgn}(x_s) dB_s$ is a Brownian motion, by Lévy Characterisation Theorem. The distribution of $(x_s, s \leqslant t)$ is the Wiener measure on $C_{x_0}([0,t]; \mathbf{R}^d)$, the space of continuous functions with initial value x_0. So uniqueness in law holds.

If (x_t) solves Tanaka's equation $x_t = \int_0^t x_s dB_s$ (initial value 0), then so does $(-x_s)$. Pathewise uniquness fails.

We construct a solution. In fact, let (W_t) be a Brownian motion on any probability space with $B_0 = 0$ and let $x_0 \in \mathbf{R}^d$, we claim that $x + W_t$ solves Tanaka's equation driven by a Brownian motion B which we specify below. Define

$$B_t = \int_0^t \mathrm{sign}(x + W_s)\, dW_s.$$

This is a local martingale with quadratic variation t and hence a Brownian motion. Furthermore

$$\int_0^t \mathrm{sign}(x + W_s) dB_s = \int_0^t dW_s = W_t.$$

Thus

$$x + W_t = x + \int_0^t \mathrm{sign}(x + W_s) dB_s,$$

as claimed. Taking $x = 0$, it is clear that $B_t = \int_0^t \mathrm{sign}(W_s)\, dW_s$ contains less information than W_t.

Example 4.11.11 ODE $\dot{x}_t = (x_t)^\alpha dt$, $\alpha < 1$, which has two solutions from zero: the trivial solution 0 and $x_t = (1-\alpha)^{\frac{1}{1-\alpha}} t^{\frac{1}{1-\alpha}}$. Both uniqueness fails.

Example 4.11.12 Dimension $d = 1$. Consider $dx_t = \sigma(x_t)dW_t$. Suppose that σ is Hölder continuous of order α, $|\sigma(x)-\sigma(y)| \leqslant c|x-y|^\alpha$ for all x, y. If $\alpha \geqslant 1/2$ then pathwise uniqueness holds for $dx_t = \sigma(x_t)dW_t$. If $\alpha < 1/2$ uniqueness no longer holds. For $\alpha > 1/2$ this goes back to Skorohod (62-65) and Tanaka(64). The $\alpha = 1/2$ case is credited to Yamada-Watanabe, and referred as the square root problem in mathematical finance modelling.

4.11.1 The Yamada-Watanabe Theorem

The following beautiful, and somewhat surprising, theorem of Yamada and Watanabe, states that the existence of a weak solution for any initial distribution together with pathwise uniqueness implies the existence of a unique strong solution.

Proposition 4.11.13 *If pathwise uniqueness holds then any solution is a strong solution and uniqueness in law holds.*

For the precise meaning of 'universally measurable' see P163 of Ikeda-Watanabe's book [8].

Theorem 4.11.14 (The Yamada-Watanabe Theorem) *If for each initial probability distribution there is a weak solution to the SDE and suppose that pathwise uniqueness holds then there exists a unique strong solution. By this we meant that there is a progressively measurable map: $F : \mathbf{R}^d \times W_0^m \to \hat{W}^d$, where the σ-algebras are 'universally complete', such that*

1. *for any probability measure μ on $\mathbf{R}^d$ there exists $\tilde{F}$ that is measurable w.r.t. $\mathcal{B}(\mathbf{R}^d \times W_0^m)^{\mu \times P}$ s.t. $F(x,\omega) = \tilde{F}(x,\omega)$ a.s.. If $\xi_0 \in \mathcal{F}_0$ we set $F(\xi_0, B) = \tilde{F}(\xi_0, B)$.*
2. *For any BM (B_t) on a probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and any $\xi_0 \in \mathcal{F}_0$, $x_t = \tilde{F}_t(\xi_0, B_t)$ is a solution to the SDE with driving noise (B_t) and initial value ξ_0.*
3. *If x_t is a solution to the equation with driving noise (B_t), then $x_t = F_t(x_0, B)$ a.s.*

In another word, for any B_t, and $x_0 \in \mathbf{R}^d$, $F_t(x_0, B)$ is a solution with the driving noise B_t. If x_t is a solution on a filtered probability space with driving noise B_t, then $x_t = F_t(x, B)$ a.s.

We do not prove this theorem, but refer to Ikeda-Watanabe and Revuz-Yor. The following observation is important for the proof of the Yamada-Watanabe Theorem.

Given two solutions on two probability space we could build them on the same probability space: $\hat{W}^d \times \hat{W}^d \times W^m$.

Lemma 4.11.15 *Let f, g be locally bounded predictable processes (measurable with respect to the filtration generated by left continuous processes), and B, W continuous semi-martingales. If $(f, B) = (g, W)$ in distribution then*

$$(f, B, \int_0^t f_s dB_s) \stackrel{law}{=} (g, W, \int_0^t g_s dW_s),$$

i.e. they have the same probability distribution.

See exercise 5.16 Revuz-Yor.

Theorem 4.11.16 *Let (M_t) be a continuous local martingale. The exponential martingale $N_t := e^{M_t - \frac{1}{2}\langle M,M\rangle_t}$ is a martingale if and only if $\mathbb{E}(N_t) = 1$ for all t.*

Proof If (N_t) is a martingale, the statement that its expectation is constant in t follows from the definition. We prove the converse. Since (N_t) is a continuous local martingale, it is a super-martingale. Indeed for a reducing sequence of stopping times T_n and any pair of real numbers $0 \leqslant s \leqslant t$, we apply Fatou's lemma:

$$\mathbb{E}(N_t | \mathcal{F}_s) \leqslant \lim_{n\to\infty} \mathbb{E}(N_t^{T_n} | \mathcal{F}_s) = \lim_{n\to\infty} N_s^{T_n} = N_s.$$

Let T be a stopping time bounded by a positive number K. By the optional stopping theorem,

$$\mathbb{E}(N_T) \geqslant \mathbb{E}(N_K) = 1, \qquad \mathbb{E}(N_T) \leqslant \mathbb{E}(N_0) = 1.$$

Thus $\mathbb{E}(N_T) = 1$ and (N_t) is a martingale. □

This can be generalized to stochastic processes that is not positive valued. Let (M_t) be a continuous local martingale with $\mathbb{E}|M_0| < \infty$. Suppose that the family $\{M_T^-, T$ bounded stopping times $\}$ is uniformly integrable. Then (M_t) is a super martingale. It is a martingale if and only if $\mathbb{E}M_t = \mathbb{E}M_0$, see Prop. 2.2 in **[?]**

4.12 Cocycle property and Markov property

Suppose that the SDE

$$dx_t = \sum_{j=1}^m X_j(t, x_t) \circ dB_t^j + X_0(t, x_t) dt,$$

has a unique strong solution and the SDE is complete (i..e from each initial point the solution exists for all time). We denote by $\varphi_t(x)$ the function given by the Yamada-Watanabe theorem, which is the solution to the SDE with initial x, and it is continuous in time and adapted. Furthermore, if η is independent of $\mathcal{F}_t^s = \sigma(W_r - W_s : r \geqslant s)$, denote by $\varphi_{s,t}(\eta)$ the solution to:

$$x_t = \eta + \sum_k \int_s^t X_k(x_r) dW_r^k + \int_s^t X_r(x_r) dr, \quad t \geqslant s.$$

The solution theory for $\varphi_t(x)$ remains to hold. We note $\varphi_{0,t}(x) \equiv \varphi_0(x)$. Then the solutions of the SDE is a, time in-homogeneous, Markov process if uniqueness of solutions holds. Its transition probability is given by

$$P_{s,t}(x, A) = \mathbb{P}(\varphi_{s,t}(x) \in A).$$

We shall focus on the time homogeneous case.

Theorem 4.12.1 *Assume that pathwise uniqueness holds and the SDE*

$$dx_t = \sum_{j=1}^m X_j(x_t) \circ dB_t^j + X_0(x_t) dt,$$

is complete. Then, the following cocycle property holds for any $x \in \mathbf{R}^n$:

$$\varphi_{s,t}(\varphi_s(x,\omega),\omega) = \varphi_{0,t}(x,\omega), \qquad \forall s \leqslant t,$$

almost surely. Furthermore,

$$\varphi_{s,t}(x,\omega) = \varphi_{0,t-s}(\varphi_s(x,\omega), \theta_s\omega).$$

Consequently, the solution is a (time homogenesoud) Markov process with transition probability

$$P_t(x, A) = \mathbb{P}(\varphi_{s,s+t}(x,\omega)).$$

Proof

$$\varphi_t(x) = \varphi_s(x) + \sum_k \int_s^t X_k(x_r) dW_r^k + \int_s^t X_r(x_r) dr.$$

By parhwise uniqueness, $\varphi_{s,t}(\varphi_s(x,\omega),\omega) = \varphi_{0,t}(x,\omega)$. Let η be independent of $\mathcal{F}_t^s$, consider the equation:

$$x_t = \eta + \sum_k \int_s^t X_k(x_r) dW_r^k + \int_s^t X_r(x_r) dr.$$

Define $\tilde{W}_r = \theta_s W(r) = W_{s+r} - W_s$. Then, by a change of variable,

$$x_t = \eta + \sum_k \int_0^{t-s} X_k(x_{u+s}) dW^k_{u+s} + \int_0^{t-s} X_r(x_{u+s}) du.$$

which means

$$x_{s+(t-s)} = \eta + \sum_k \int_0^{t-s} X_k(x_{u+s}) d\tilde{W}^k_u + \int_0^{t-s} X_r(x_{u+s}) du,$$

consequently by pathwise uniqueness: $\varphi_{s,t}(x,\omega) = \varphi_{0,t-s}(\eta, \theta_s\omega)$. Set $P_t(x, A) = \mathbb{P}(\varphi_t(x) \in A)$, then by the cocycle property,

$$\mathbb{P}(\varphi_{s+t}(x) \in A|\mathcal{F}_s) = \mathbb{P}(\varphi_{0,t}(x_s, \theta_s\omega) \in A|\mathcal{F}_s) = P_t(\varphi_s(x), A),$$

almost surely, proving that the solution is a time homogoneous Markov process with transition probabilities $P_t(x, \cdot)$. □

4.13 The Markov semi-group

We continue to study the SDE $dx_t = \sum_{j=1}^m X_j(x_t) dB^j_t + X_0(x_t)dt$, under the assumption of completeness and pathwise uniqueness. Let $\mathcal{F}^s_t = \sigma\{(W_r - W_s : s \leqslant r \leqslant t\}$ and $\varphi_{s,t}(-)$ the solution flow with initial time s. Recall that $P_t(x, A) = \mathbb{E}(\varphi_t(x) \in A)$ and $\mathbb{P}(\varphi_{s+t}(x) \in A|\mathcal{F}_s) = P_t(\varphi_s(x), A)$. Define the semi-group :

$$T_t f(x) = \int_{\mathbf{R}^d} f(y) P_t(x, dy).$$

Lemma 4.13.1 *Let η be a random variable on $\mathbf{R}^d$, independent of $\mathcal{F}^s_t$. Suppose that pathwise uniqueess and non-explosion holds. Let $x_t := \varphi_{s,t}(\eta)$ denote the solution flow $\varphi_{s,s}(\eta) = \eta$. Then for any function $f \in C^2$, we have*

$$f(x_t) = f(\eta) + \int_s^t df(x_r)(X_i(x_r)\, dW^i_r)) + \int_s^t \mathcal{L}f(x_r)dr,$$

where for $a_{i,j}(x) = \sum_{k=1}^m X^i_k(x) X^j_k(x)$,

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

If $f \in BC^2$ and if X_k, where $k = 0, 1, \dots, m$, are locally Lipschitz continuous and grow at most linearly, then

$$T_t f(x) - f(x) = \int_0^t T_s \mathcal{L} f(x) dr, \tag{4.6}$$

Proof The formula for $f(x_t)$ follows from Itô's formula. If $\sum_{k=1}^m |X_k|$ grows at most linearly and $\sup_{r\leqslant T} \mathbb{E}(|\varphi_{s,r}|^2) < \infty$ (this is the case if all the vector fields X_i are locally Lipschitz continuous and grow at most linearly), then for any $f \in \mathrm{B}C^2$, we have

$$\mathbb{E}[f(x_t)] = \mathbb{E}[f(\eta)] + \int_s^t \mathbb{E}\mathcal{L}f(x_r)dr.$$

Taking $\eta = x$, leading to

$$T_t f(x) = f(x) + \int_0^t T_s \mathcal{L} f(x) dr,$$

as claimed. □

4.14 Strongly continuous semi-group on $\mathbf{C}_0$

We now want to show that $\mathcal{L}$ is the generator of T_t and to determine the domain of $\mathcal{L}$. If $s \mapsto T_s f$ as a continuous map on a subspace E of bounded measurable functions (i.e. T_s is strongly continuous), then fundamental theorem of Calculus allows to conclude that $\lim_{t\to 0} \frac{1}{t}(T_t f - f) \to \mathcal{L}f$ in the supremum norm. Note that the heat semi-group is not strongly continuous on bounded measurable function, it is however continuous on C_K^∞, smooth function with compact support.

4.14.1 Strongly continuous semi-group on a Banach space

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

1. *If $x \in E$ and $t > 0$, then*

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

and

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

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

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

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

Proof (i) We have

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

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

(ii) If $x \in \mathbb{D}(\mathcal{L})$ 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, $T_t x \in \mathbb{D}(L)$ 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 $\frac{1}{t}\int_0^t T_s x ds \in \mathbb{D}(\mathcal{L})$ 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 $x \in \overline{\mathbb{D}(\mathcal{L})}$.

Finally we show that $\mathcal{L}$ is closed. Let $(x_n) \subset \mathbb{D}(\mathcal{L})$, $x_n \to x$, 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, $x \in \mathbb{D}(\mathcal{L})$ and $Lx = y$. Consequently, $\mathcal{L}$ is closed. □

4.14.2 Strongly continuous semigroup arising from SDEs

Definition 4.14.2 A Markov-semigroup T_t is said to have the C_0-property if when it take $\mathbf{C}_0(\mathcal{X}) \to \mathbf{C}_0(\mathcal{X})$. Equivalently, $T_t : \mathbf{C}_0(\mathcal{X}) \to \mathbf{C}_0(\mathcal{X})$ is strongly continuous.

Note that one way to show the C_0 property is to show that $\lim_{x\to\infty}\varphi_t(x) = 0$ where $\varphi_t(x)$ denotes the solution flow to the SDE. Observe also that

$$\mathbb{P}(\frac{1}{|\varphi_t(x)|^2+1} > M) \leqslant \frac{1}{M}\mathbb{E}(\frac{1}{|\varphi_t(x)|^2+1}),$$

which converges to zero, as $x \to \infty$, under the linear growth and locally Lipschitz continuous condition.

To show $T_tC_0 \subset C_0$ it is equivalent to showing that $\lim_{x\to\infty} P_t(x, K) = 0$.

Lemma 4.14.3 *Let T_t be a Markov semigroup and x_t a time homogeneous Markov process such that $T_tf(x) = \mathbb{E}[f(x_{t+s})|\mathcal{F}_s)$. Then T_t restricts to $\mathbf{C}_0(\mathcal{X})$ if and only if for any compact set K,*

$$\lim_{x\to\infty} \mathbb{P}(X_t \in K) = 0.$$

Proof Let $f \in \mathbf{C}_0$ and K be a compact set. Then

$$|T_f(x)| \leqslant |\mathbb{E}[\mathbb{E}(f(x_t)|x_0 = x)\mathbf{1}_{x_t\in K})]| + |\mathbb{E}(f(x_t)|x_0 = x)\mathbf{1}_{x_t\notin K}) \leqslant \|f\|_\infty \mathbb{P}(x_t \in K) + \sup_{x\notin K}|f(x)|.$$

For any $\epsilon > 0$ there exists K_ϵ such that outside of which $|f| \leqslant \epsilon/2$. Thus, taking $\lim_{x\to\infty}\mathbb{P}(X_t \in K) = 0$, □

Example 4.14.4 The heat semi-group is a strongly continuous semi-group on C_0. We first take $f \in C_K^\infty$. Since $T_tf(x) = \mathbb{E}f(x + B_t)$ 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

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

Suppose f has compact support K, T_sf'' 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.

Proposition 4.14.5 *Suppose that the vector fields are locally Lipschitz continuous and grow at most linearly, and suppose that its corresponding Markov-semigroup has the C_0-property. Let $\mathcal{L}$ denotes the Markov generator of T_t. Then any $f \in C_K^2$ is in the domain of $\mathcal{L}$ and*

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

Furthermore, $T_t f$ solves backward Kolmogorov equation.

$$\frac{du}{dt} = \mathcal{L} u_t, \qquad u(0, \cdot) = f.$$

Proof The SDE is conservative and the solutions have finite p-th moments, bounded uniformly in

Suppose that T_t has the $\mathbf{C}_0$-property, Under the conditions on X_i, $\mathcal{L}f \in \mathbf{C}_0$, $t \mapsto T_t(\mathcal{L}f)$ is continuous. Apply the fundamental theorem of calculus to the following identity on $\mathbf{C}_0$:

$$T_t f = f + \int_0^t T_s \mathcal{L} f dr,$$

to see that $\mathcal{L}f$ is indeed the derivative of $T_t f$ at $t = 0$. Consequently $\mathbf{C}_K^2$ is in the domain of the generator. Furthermore, if $\mathcal{L}$ is the generator of a semi-group T_t, and if $f \in \mathbb{D}(\mathcal{L})$, then $T_s f \in \mathbb{D}(\mathcal{L})$ by Theorem 4.14.1 and $u_t := T_t f$ solves the Cauchy problem

$$\frac{du}{dt} = \mathcal{L} u_t, \qquad u(0, \cdot) = f.$$

This completes the proof. □

Example 4.14.6 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 4.14.7 Check that T_t preserves the space $\mathbf{C}_0(\mathbf{R}^n)$.

Definition 4.14.8 The L^2-adjoint of $\mathcal{L}$, denote by $\mathcal{L}^*$, is a linear operator defined as follows. $g \in \mathcal{L}^2$ is in the domain of $\mathcal{L}^*$ if

$$\int_{\mathbf{R}^d} \mathcal{L} f g dx = \int_{\mathbf{R}^d} f \mathcal{L}^* g dx$$

holds for any $f \in \mathrm{D}om(\mathcal{L})$.

We observe that for the generator of the SDE,

$$\mathcal{L}^* f(x) = -\operatorname{div}(bf) + \frac{1}{2}\sum_{i,j} a_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}$$

where

$$\operatorname{div}(bf) = f \operatorname{div} b + \langle b, \nabla f\rangle.$$

Suppose that the probability measure of x_t is absolutely continuous with respect to the Lebesgue measure. We write:

$$P_t(x+0, dx) = p_t(x_0, x)dx.$$

By (4.6),

$$\int_{\mathbf{R}^d} f(y)p_t(x_0,x)dx - f(x_0) = \int_0^t \int_{\mathbf{R}^d} \mathcal{L}f(y)p_t(x_0,x)dx.$$

We therefore expect that p_t solves the Fokker-Planck equation (Kolmogorov's forward equation):

$$\frac{\partial p_t(x_0,x)}{\partial t} = \mathcal{L}^* p_t(x_0,x), \qquad p_0(x_0,\cdot) = \delta_{x_0}. \tag{4.7}$$

4.14.3 Strong complete, Feller Property, and Strong Markov property

Definition 4.14.9 An SDE is strongly complete if for every initial point there exists a version of the solution which we denote by $\varphi_t(x,\omega)$ satisfying the following property: for almost surely all ω,

$$(t,x) \mapsto \varphi_t(x,\omega)$$

is continuous from $\mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$.

If X_j are Lipschitz continuous, strong completeness holds. The following equation has a global solution from any initial point.

$$\begin{aligned} dx_t &= (y_t^2 - x_t^2)\, dW_t^1 + 2x_t y_t\, dW_t^2, \\ dy_t &= -2x_t y_t\, dW_t^1 + (x_t^2 - y_t^2)\, dW_t^2, \end{aligned} \tag{4.8}$$

where (W_t^1, W_t^2) is a Brownian motion on $\mathbf{R}^2$. It is not strong complete

Proposition 4.14.10 *If X_i are Lipschitz continuous, then the SDE is strongly complete.*

Lemma 4.14.11 *If strong completeness holds (indeed we only need to assume that $x \mapsto \varphi_t(x,\omega)$ is continuous in probability for every t), Feller property holds for T_t. Consequently, the solutions are strong Markov processes.*

The Feller property follows from the dominated convergence theorem.

4.15 Martingale Considerations

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

Theorem 4.15.1 *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 $t \geqslant 0$.*

Proof Let us denote the generator by L. Since $\mathbb{D}(L)$ is dense, and T_t, S_t are continuous linear operators, it is sufficient to show that $T_t = S_t$ on $\mathbb{D}(L)$. Note that $S_0 = T_0$. Take x in the domain, then, for each $r \geqslant 0$, $S_r x, T_r x \in \mathbb{D}(\mathcal{L})$. 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 4.14.1 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 $\mathbf{C}_K^\infty$ is sufficient to identify the transition functions. Then if the martingale problem is well posed we are in good business.

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 4.15.2 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 remains unsolved for a general Riemannian manifold.

Example 4.15.3 (BM on $\mathbf{R}_+$, 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 $\mathbf{R}$ 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 $x \geqslant 0$,

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

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

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

Definition 4.15.4 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 separable, s.t.

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

we say that X_t corresponds to T_t.

Consequently, if $\mu = \mathcal{L}(X_0)$ is the initial distribution, $\mathbb{E}[T_t f(X_0)] = \int_{\mathcal{X}} T_t f(x)\mu(dx)$. The solution of an SDE with Lipschitz continuous coefficients corresponds to its semi-group.

Proposition 4.15.5 *[3, pp161] Let $\mathcal{X}$ be a separable metric space and let $E \subset \mathcal{B}_b(\mathcal{X})$ be closed sub-space which is measure determining. Let (T_t) be a semigroup of bounded linear operators on E and (X_t) be a Markov process on $\mathcal{X}$, corresponding to T_t, and with initial distribution μ. Then T_t and μ uniquely determine the finite dimensional distributions of (X_t).*

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

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

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

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

is separating on $\Pi_{i=1}^n \mathcal{X}$, see Theorem 2.2.8. We claim for any $n \geqslant 1$, $f_1, \dots, f_n \in E$, and $0 \leqslant t_1 < \cdots < t_n$,

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

which means that the finite dimensional distribution of X_t is determined. We prove the above claim by induction. 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 $t_1 < t_2 < \cdots < t_n$,

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

this concludes the proof. □

Proposition 4.15.6 *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 $f \in \mathbb{D}(\mathcal{L})$,*

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

is a martingale.

Proof Let $s < t$. The cádlág property of X_r implies that $r \mapsto \mathcal{L}f(X_r)$ is measurable. Note that f and $\mathcal{L}f$, both belongs to $E \subset \mathcal{B}_b(\mathcal{X})$, are bounded, integrals are finite. It is then trivial to see that

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

In the last step we used the fact that, for a strongly continuous semigroup, $\mathcal{L}T_t f = T_t\mathcal{L}f = \frac{d}{dt}T_t f$, for every $f \in \mathbb{D}(\mathcal{L})$. This completes the proof. □

The converse holds if $E = \mathbf{C}_0(\mathcal{X})$.

Proposition 4.15.7 *Let T_t be a strongly continuous semigroup on $\mathbf{C}_0(\mathcal{X})$ 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 $f, g \in \mathbf{C}_0(\mathcal{X})$ and*

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

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

Proof Note that $g \in \mathbf{C}_0(\mathcal{X})$ is bounded. 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,

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

Take $X_0 = \delta_x$, then $f(x) = \int_{\mathcal{X}} f(y)\delta_x(dy) = \mathbb{E}[f(X_0)]$. Since X_t is a Markov process corresponding to T_t with initial point x and since f is continuous,

$$T_t f(x) = \mathbb{E}[T_t f(X_0)] = \mathbb{E}[\mathbb{E}[f(X_t) \mid X_0]] = \mathbb{E}[f(X_t)].$$

By Fubini's theorem,

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

Since $r \mapsto T_r g$ is continuous on $\mathbf{C}_0$, the right hand side converges to $\frac{1}{t}\int_0^t T_r g dr \to g$ and $\mathcal{L}f = g$. □

4.15.1 Diffusion operator an and Martingale problem

Let $X = (X_1, \ldots, X_m)$ be the $d \times m$-matrix with column vector given by the vector fields $X_1, \ldots, X_m$, set $A = XX^T$. Writing $A = (a_{ij})$, denote $\mathcal{L}f(x) = \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 f}{\partial x_i x_j}(x) + df(X_0)$. It is routine to require the solution, to the SDE ,to have continuous path or has cadlág sample paths. In the former case, we would limit our solutions to the space $\mathbf{C}(\mathbf{R}_+, \mathcal{X})$ of continuous paths, or to $\mathbf{C}(\mathbf{R}_+, \mathcal{X})$ where $T > 0$. For any x, $(a_{ij}(x))$ is a non-negative symmetric matrix. Such an operator is referred as diffusion operator.

Definition 4.15.8 A continuous process x on $\mathbf{R}^d$ or its probability distribution, denoted by $\mathbb{P}_\mu$ where $\mu = \mathcal{L}(X_0)$, is said to solve the local martingale problem for $\mathcal{L}$, if

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

is a local martingale (for the natural filtration of x_t) for every $f \in \mathbf{C}_K^\infty$.

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 $\mathbf{R}^n$.

Definition 4.15.9 The local martingale problem for $\mathcal{L}$ is said to be unique if any two solutions to the martingale problem, with the same initial distribution, have the same probability law. It is said to be well posed if for any initial distribution there exists exactly one solution.

The questions whether the martingale problem is well posed is a fundamental question, which leads essentially to the strong Markov property. Being a martingale is a property of finite dimensional distributions. Indeed M_t^f is a martingale if and only if for any $0 \leqslant t_1 < t_2 < \cdots < t_{n+1}$, and $g_i \in \mathcal{B}_b(\mathcal{X})$, the following holds:

$$\mathbb{E}\Big[\Big(f(x_{t_{n+1}}) - f(x_{t_n}) - \int_{t_n}^{t_{n+1}} \mathcal{L}f(x_r)dr\Big)\Pi_{i=1}^n g_i(x_{t_i})\Big] = 0.$$

For the proof of the next theorem we need the following (local) martingale representation theorem:

Theorem 4.15.10 (Integral representation) *Let M_t be a continuous local martingale with values in $\mathbf{R}^d$, vanishing at time zero, with quadratic variation*

$$\langle M^i, M^j\rangle_t = \sum_{k=1}^m \int_0^t \sigma_k^i(s)\sigma_k^j(s)ds$$

where σ_k^i are progressively measurable stochastic processes. Then there exists a Brownian motion W_t such that

$$M_t^i = \sum_{k=1}^m \int_0^t \sigma_k^i dW_s^k.$$

Proof If $m = d$ and if the matrix $\sigma = (\sigma_1, \dots, \sigma_d)$ is invertible, we simply set $W_t = \int_0^t \sigma^{-1}(s)dM_s$. Otherwise, we let $\Pi : \mathbf{R}^m \to \mathbf{R}^m$ and $\Pi^\perp : \mathbf{R}^m \to \mathbf{R}^m$ denote respectively the orthogonal projection from $\mathbf{R}^m$ to its subspaces $\ker(\sigma)$ and $[\ker(\sigma)]^\perp$, the kernel of $\sigma = (\sigma_1, \dots, \sigma_m)$ and its orthogonal kernel. Observe that the quadratic variation matrix is $\langle M, M\rangle = \int_0^t (\sigma(s)\sigma^T(s))ds$ where σ^T denotes the transpose of σ. Let $\tilde{W}_t$ be a Brownian motion on $\mathbf{R}^n$, independent of M_t, and set

$$W_t = \int_0^t \sigma(s)^{-1}\Pi^\perp dM_s + \int_0^t \Pi_s d\tilde{W}_s.$$

Then W_t is a Brownian motion, as its quadratic variation is:

$$\sigma^{-1}\Pi(\sigma\sigma^T)(\sigma^{-1}\Pi)^T + \Pi^\perp(\Pi^\perp)^T = I_{d\times d}.$$

In addition,

$$\sum_{k=1}^m \int_0^t \sigma^i(s)dW_s^i := \int_0^t \sigma(s)dM_s = \int_0^t \Pi^\perp dM_s + \int_0^t \sigma(s)\Pi_s dM_s = \int_0^t \Pi^\perp dM_s = M_s.$$

We have used that $\int_0^t \sigma(s)\Pi_s^\perp d\tilde{W}_s = 0$. □

Exercise 4.15.11 If M_t is a martingale on $\mathbf{R}^m$ with quadratic variation $\int_0^t A(s)ds$ and $\sigma : \mathbf{R}^m \to \mathbf{R}^d$ is a continuous and adapted process, show that $N_t := \int_0^t \sigma(s)dB_s$ has quadratic variation $\int_0^t (\sigma A\sigma^T)(s)ds$. Hint: Work with individually entries $\langle N^i, N^j\rangle_t$.

Theorem 4.15.12 *Let X_i be progressively measurable. Consider the canonical probability space $\mathbf{C}(\mathbf{R}_+, \mathbf{R}^d)$ endowed with a probability measure $\mathbb{P}$. Then the SDE has a weak solution with distribution $\mathbb{P}$ if and only if $\mathbb{P}$ solve the local martingale problem for $\mathcal{L}$. In particular, if x_t solve the local martingale problem for $\mathcal{L}$, there exists a Brownian motion W_t such that*

$$x_t = x_0 + \int_0^t b(x_s)ds + \int_0^t X_k(x_s)dW_s^k.$$

Proof The 'only if' part is trivial and follows from Itô's formula.

If $\mathbb{P}$ solve the local martingale problem for $\mathcal{L}$, then we take f_n^i be a sequence of smooth functions with compact supports such that $f_n^i(x) = x_i$ for $x \in B_n$ and $f_n^i \to f$. By the assumption:

$$f_n^i(x_t) - f_n^i(x_0) - \int_0^t \mathcal{L} f_n^i(x_s)ds$$

is a local martingale. Using stopping time and by the definition of local martingales we see that

$$M_t^i := x_t^i - x_0^i - \int_0^t b^i(x_s)ds$$

is a local martingale – observe that $\mathcal{L}f = df(b)$. By a similar consideration applied to $g_{ij}(x) = x^i x^j$ we see that

$$M_{i,j} := x_t^i x_t^j - x_0^i x_0^j - \frac{1}{2}\sum_{k=1}^m \int_0^t \sigma_k^i(s)\sigma_k^j(x_s)ds - \int_0^t [x_s^j b^i(x_s) + x_s^j b^i(x_s)]ds,$$

is a local martingale. We have used the fact that $\mathcal{L}(g_{ij}) = \frac{1}{2}\sum_{k=1}^m \sigma_k^i \sigma_k^j + x^j b^i + x^i b^j$. Furthermore,

$$\langle M^i, M^j \rangle_t = \langle x^i, x^j \rangle_t.$$

On one hand, by Itô's formula:

$$\begin{aligned} x_t^i x_t^j &= x_0^i x_0^j + \int_0^t x_s^i dx_s^j + \int_0^t x_s^j dx_s^i + \langle x^i, x^j \rangle_t \\ &= x_0^i x_0^j + \int_0^t x_s^i dM_s^j + \int_0^t x_s^j dM_s^i + \langle x^i, x^j \rangle_t + \int_0^t (x_s^i b^j(x_s) + x_s^j b^i(x_s))ds. \end{aligned}$$

Consequently,

$$M_{ij} = \int_0^t x_s^i dM_s^j + \int_0^t x_s^j dM_s^i + \langle M^i, M^j \rangle_t - \frac{1}{2}\sum_{k=1}^m \int_0^t \sigma_k^i(s)\sigma_k^j(x_s)ds.$$

Since M_{ij} and the first two terms on the right hand side are martingales, by the uniqueness of semi-martingale decomposition, it is necessary that

$$\langle M^i, M^j \rangle_t = \frac{1}{2}\sum_{k=1}^m \int_0^t \sigma_k^i(s)\sigma_k^j(x_s)ds.$$

There exists a Brownian motion W_t such that

$$M_t^i = \int_0^t X_k^i(x_s)dW_s^k.$$

Consequently,

$$x_t^i = x_0^i + \int_0^t b^i(x_s)ds + \int_0^t X_k^i(x_s)dW_s^k,$$

proving that x_t is the solution to the SDE driven by W_t. □

The following theorem illustrates that uniqueness implies Markovian property.

Theorem 4.15.13 *[3, 174] Let $\mathcal{L} : \mathcal{B}_b(\mathcal{X}) \to \mathcal{B}_b(\mathcal{X})$ be a linear operator. Suppose that for each probability measure μ on $\mathcal{X}$, any two solutions X, Y of the martingale problem for $(\mathcal{L}, \mu)$ satisfy that for any $t > 0$,*

$$\mathbb{P}(X_t \in A) = \mathbb{P}(Y_t \in A), \qquad \forall A \in \mathcal{B}(\mathcal{X}).$$

Then, any solution of the martingale problem (MP) for $\mathcal{L}$ with respect to a filtration $\mathcal{G}_t$ is Markov process with respect to $\mathcal{G}_t$. Furthermore uniqueness holds for the martingale problem for $\mathcal{L}$.

Proof Let (X_t) be a solution of the martingale problem (MP), on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for $\mathcal{L}$ with respect to a filtration $\mathcal{G}_t$. Then for any $f, g_k \in B_b(\mathcal{X})$ and for any $r \geqslant 0$, $0 \leqslant t_1 < t_2 < \cdots < t_n < t_{n+1}$,

$$\mathbb{E}\Big[\Big(f(X_{r+t_{n+1}}) - f(X_{r+t_n}) - \int_{t_n}^{t_{n+1}} f(X_{r+s})ds\Big) \;\Big|\; \mathcal{G}_r\Big] = 0. \tag{4.9}$$

We show that for any $t \geqslant 0$, $r \geqslant 0$, and $f \in \mathcal{B}_b(\mathcal{X})$,

$$\mathbb{E}[f(X_{t+r})|\mathcal{G}_r] = \mathbb{E}[f(X_{t+r})|X_r].$$

Equivalently, we show that for any $\Gamma \in \mathcal{G}_r$,

$$\int_\Gamma f(X_{t+r})d\mathbb{P} = \int_\Gamma \mathbb{E}[f(X_{t+r})|X_r]d\mathbb{P}. \tag{4.10}$$

Assume that $\mathbb{P}(\Gamma) > 0$, let us define two probability measures on $(\Omega, \mathcal{F})$ as follows:

$$\mathbb{P}_1(B) = \frac{\mathbb{P}(\Gamma \cap B)}{P(\Gamma)} = \frac{\int_\Gamma \mathbb{E}[\mathbf{1}_B|\mathcal{G}_r]d\mathbb{P}}{P(\Gamma)}, \qquad \mathbb{P}_1(B) = \frac{\int_\Gamma \mathbb{E}[\mathbf{1}_B|X_r]d\mathbb{P}}{P(\Gamma)}.$$

Then,

$$\int_\Gamma f(X_{t+r})d\mathbb{P} = \frac{1}{\mathbb{P}(\Gamma)}\int_\Omega f(X_{t+r})d\mathbb{P}_1 \qquad \int_\Gamma \mathbb{E}[f(X_{t+r})|X_r]d\mathbb{P} = \frac{1}{\mathbb{P}(\Gamma)}\int_\Omega f(X_{t+r})d\mathbb{P}_2.$$

Let $Y_t = X_{t+r}$, the required identity, (4.10), can be written as

$$\int_\Omega f(Y_t)d\mathbb{P}_2 = \int_\Omega f(Y_t)d\mathbb{P}_1,$$

which follows from the uniqueness assumption on the marginal distributions, if we can show that Y_t solves the MP for $\mathcal{L}$ on $(\Omega, \mathcal{F}, \mathbb{P}_1)$ and on $(\Omega, \mathcal{F}, \mathbb{P}_2)$. Note that

$$\begin{aligned}
&\int_\Gamma \Big(f(X_{r+t_{n+1}}) - f(X_{r+t_n}) - \int_{t_n}^{t_{n+1}} f(X_{r+s})ds\Big)d\mathbb{P}_2 \\
&= \mathbb{P}(\Gamma)\int_\Gamma \mathbb{E}\Big(f(X_{r+t_{n+1}}) - f(X_{r+t_n}) - \int_{t_n}^{t_{n+1}} f(X_{r+s})ds \;\Big|\; X_r\Big)d\mathbb{P} \\
&= \mathbb{P}(\Gamma)\int_\Gamma \mathbb{E}\Big[\mathbb{E}\Big(f(X_{r+t_{n+1}}) - f(X_{r+t_n}) - \int_{t_n}^{t_{n+1}} f(X_{r+s})ds \;\Big|\; \mathcal{G}_r\Big] \;\Big|\; X_r\Big)d\mathbb{P} = 0,
\end{aligned}$$

showing that Y_t solves the MP for $\mathcal{L}$ on $(\Omega, \mathcal{F}, \mathbb{P}_2)$, similarly on $(\Omega, \mathcal{F}, \mathbb{P}_1)$. □

The following is proved in Theorem 5.1.20.

Theorem 4.15.14 *Let X^TX and X_0 be bounded and continuous. Then for any initial probability distribution there exists a martingale solution for $\mathcal{L}$, on $\mathbf{C}(\mathbf{R}_+)$. If the local martingale problem for δ_x has a unique solution, then there is a unique solution for any initial distribution. Well-posedness of the local martingale problem for $\mathcal{L}$ implies that the solution is a Markov process.*

See [**?**, pp 295, Corollary 3.4] and [**?**, pp.419, Thm. 21.9, Thm 21.10 om pp 420].

The following theorem is similar to [3, pp 234, Theorem 8.10].

Theorem 4.15.15 *Let X_i be continuous, consider the SDE driven by (X_i). Suppose that weak uniqueness holds and that the solution is global. Suppose that the $C(\mathbf{R}_+, \mathcal{X})$ local martingale problem for $(\mathcal{L}, \mu)$ has at most one solution. Suppose that x_n is a sequence of adapted stochastic processes with sample paths in $C(\mathbf{R}_+; \mathcal{X})$, and is relatively compact, with $\mathcal{L}(x_n(0)) \to \mu$, a probability measure. Let $M \subset \mathrm{BC}$ be a measure separating set. Suppose that for each $f \in \mathbf{C}_K^\infty$,*

$$\limsup_{n\to\infty} \mathbb{E}[\big(f(x_n(t)) - f(x_n(s)) - \int_s^t \mathcal{L}f(x_u)du\big)\Pi_{i=1}^k h_i(x_{t_i})] = 0$$

for all $h_i \in M$, $0 \leqslant t_1 < t_2 < \dots t_k \leqslant s < t$. Then there exists a solution x to the martingale problem for $(\mathcal{L}, \mu)$, and the distributions of x_n converge to that of x (weakly).

Proof Since x_n is relatively compact we only need to identify its accumulation points. Assume that $x_n \to x$. Then by taking $n \to \infty$ we see that

$$\mathbb{E}[\big(f(x(t)) - f(x(s)) - \int_s^t \mathcal{L}f(x_u)du\big)\Pi_{i=1}^k h_i(x_{t_i})] = 0,$$

and $f(x_t) - f(x_s) - \int_s^t \mathcal{L} f(x_u)du$ is a martingale. Consequently x_t solves the martingale problem for $\mathcal{L}$. By weak uniqueness, it is a Martingale with generator $\mathcal{L}$. □

4.16 Ellipticity

The operator $\mathcal{L}$ is elliptic if for any x and any $\xi \in \mathbf{R}^n$,

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

It is strictly elliptic if there exists $c > 0$ for any x and any $\xi \in \mathbf{R}^n$, 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 $\int_{\mathbf{R}^n} \mathcal{L} f g dx = 0$ 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 4.16.1 The Brownian motion on $\mathbf{R}^n$ has no finite invariant probability measure. Its only invariant measure is dx. It has no non-constant harmonic functions.

Example 4.16.2 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 4.16.3 Let $b : \mathbf{R}^n \to \mathbf{R}^n$ and $V : \mathbf{R}^n \to \mathbf{R}$ 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 4.16.4 A diffusion process is a continuous strong Markov process.

4.17 Appendix: 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 $(R_\lambda, \lambda \geqslant 0)$ of the semi-group and show that it is the inverse to $\lambda - \mathcal{L}$.

Definition 4.17.1 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 4.17.2 *If T_t is a strongly continuous contraction semi-group E, then R_λ is a strongly continuous contraction resolvent 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 $\mathbf{R}$ 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

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

Now $\tau_1 + \tau_2$ is distributed as

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

Using the semigroup property,

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

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

Example 4.17.3 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 4.17.4 *Let T_t a strongly continuous contraction semi-group E with generator $\mathcal{L}$, then the following statements hold for any $\lambda > 0$.*

1. *For any $x \in E$, $R_\lambda x \in \mathbb{D}(\mathcal{L})$;*
2. *For any $x \in \mathbb{D}(\mathcal{L})$, $\mathcal{L}R_\lambda x = R_\lambda \mathcal{L}x$.*
3. *Any number $\lambda > 0$ belongs to the resolvent set $\varrho(\mathcal{L})$ 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,

$$\|R_\lambda x\| = \left\|\int_0^\infty e^{-\lambda t} T_t x \; dt\right\| \leqslant \int_0^\infty e^{-\lambda t} \, dt \; \|x\| \leqslant \frac{1}{\lambda}\|x\|, \tag{4.11}$$

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

$$\mathcal{L}(R_\lambda x) = \lambda R_\lambda x - x, \tag{4.12}$$

proving

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

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

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

$$\begin{aligned} R_\lambda(\mathcal{L}x) &\overset{\text{definition}}{=} \lim_{s\to\infty} \int_0^s e^{-\lambda t} T_t(\mathcal{L}x)\, dt \\ &= \lim_{s\to\infty} \int_0^s \mathcal{L}(e^{-\lambda t} T_t x)\, dt = \lim_{s\to\infty} \mathcal{L}\Big(\overbrace{\int_0^s T_t(e^{-\lambda t} x)\, dt}^{R_\lambda^s} \Big). \end{aligned}$$

We used part (i) of Theorem 4.14.1. 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 4.14.1, $\mathcal{L}(R_\lambda^s) \to \mathcal{L}R_\lambda$, concluding

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

the latter follows from (4.12). 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 $\varrho(\mathcal{L})$ 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 4.17.5 *Let $\mathcal{L}$ be the generator of a strongly continuous contraction semigroup on E. Then $\varrho(\mathcal{L}) \supset \{\lambda : Re(\lambda) > 0\}$, for such λ,*

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

Example 4.17.6 Let $E = \{f : \mathbf{R}_+ \to \mathbf{R}_+ :$ 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 $\mathbb{D}(\mathcal{L})$. Now, $T_t f = e^{\lambda t} f$ and $\mathcal{L}f = \lambda f$, so the resolvent set $\varrho(\mathcal{L})$ is the right half of the plane.

4.17.1 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 4.17.7 Consider a linear operator $A : \mathbb{D}(A) \subset E \to E$.

- A is said to be dissipative if

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

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

$$\|(\lambda - A)^{-1}\| \leqslant \frac{1}{\lambda}. \tag{4.13}$$

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

Exercise 4.17.8 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, **[?]**:

Theorem 4.17.9 *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 $\varrho(\mathcal{L}) = \{\lambda \in \mathbf{C} : (\lambda - \mathcal{L}) : E \to E$ is bijection$\}$. If $\lambda \in \varrho(\mathcal{L})$ 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 4.17.10 *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 $x \in \mathbb{D}(\mathcal{L})$,

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

Proof Let $x \in \mathbb{D}(\mathcal{L})$ 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 $\mathbb{D}(\mathcal{L})$ is dense, this holds for all $x \in E$. □ Let us write $R_\lambda = (\lambda - \mathcal{L})^{-1}$, then

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

Definition 4.17.11 $\mathcal{L}_\lambda$ is said to be the Yosida approximation for $\mathcal{L}$.

Lemma 4.17.12 *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,*

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

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 4.17.13 (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.*

1. *$\mathcal{L}$ is closed and densely defined.*
2. *$\mathcal{L}$ is M-dissipative.*

Proof $\Longrightarrow$ The only if part follows from Proposition 4.15.1 and Theorem 4.14.1.

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

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

By Lemma 4.17.12,

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

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

By Theorem 4.15.1, any positive number $\lambda \in \varrho(\mathcal{A})$, $(\lambda - \mathcal{A})$ is a bijection, and

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

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

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

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

Corollary 4.17.14 *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: [**?**, **?**].

Corollary 4.17.15 *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 $\varrho(\mathcal{L}) \supset (0, \infty)$. 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 **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: $T_t = e^{t\mathcal{L}}$. □

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

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 4.17.16 Let E be a vector space of functions with values in K (where $K = \mathbf{R}$ or $\mathbf{C}$). A linear functional ℓ on E is a linear map $\ell : E \to K$. A positive linear functional $\ell : E \to \mathbf{R}$ is a linear functional such that $\ell(f) \geqslant 0$ whenever $f \in E$ is a function with $f \geqslant 0$ 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 4.17.17 1. A sequence x_n in a normed space E is said to convergent (strongly convergent) if $\|x_n - x\| \to 0$ for some $x \in E$.

2. 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 $\mu(f_n) := \int_E f_n d\mu$ 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 4.17.18 (Riesz-Markov) *Let $\mathcal{X}$ be a locally compact metric space. Then the dual of the $\mathbf{C}_0(\mathcal{X})$ is the space of signed Borel measures on $\mathcal{X}$ with finite total variation. In particular, if $\ell : \mathbf{C}_0(\mathcal{X}) \to \mathbf{R}$ is a positive linear functional, then there exists a unique Borel measure μ on $\mathcal{X}$ with finite total variation such that*

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

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

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

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

Note. References for this section are: [**?**, **?**, **?**]

4.17.3 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

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

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 theorem 4.17.20 and to develop a theory for Markov semigroups (T_t) which leave $\mathbf{C}_0(\mathcal{X})$ 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 6.1 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 [17, Thm 2.7, pp91], [10], [18, Section III.7].

Theorem 4.17.19 *If (X_t) is a Markov process with transition semigroup (T_t), which is strongly continuous on $\mathbf{C}_0(\mathcal{X})$, then there exists a càdlàg modification of (X_t), which is a $(\mathcal{F}_t^+)$-Markov process with the same transition semigroup.*

Corollary 4.17.20 *If $\mathcal{X}$ is locally compact and $T_t : \mathbf{C}_0(\mathcal{X}) \to \mathbf{C}_0(\mathcal{X})$, $t \geqslant 0$, 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*

$$T_t f(x) = \int_{\mathcal{X}} f(y) P_t(x, dy) \qquad \forall f \in \mathbf{C}_0(\mathcal{X}). \tag{4.14}$$

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 $f \in \mathbf{C}(\mathcal{X}) \mapsto T_t f(x) \in \mathbf{R}$ 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 4.17.19 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. □

Exercise 4.17.21 Write down a Markov process for which $\mathcal{E} \neq \mathcal{B}_b(\mathcal{X})$.

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

Exercise 4.17.22 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 $\mathcal{X} \to \mathbb{P}(\mathcal{X})$ 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 $T_t\big(\mathbf{C}_0(\mathcal{X})\big) \subset \mathbf{C}_0(\mathcal{X})$ where

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

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 $\mathbf{C}_0(\mathcal{X}) = \{0\}$. Nonetheless, we have the following result, see e.g. [17, Prop 2.4, pp89]:

Lemma 4.17.23 *Let (T_t) be the Markov transition semigroup of a right continuous Markov process with $T_t\big(\mathbf{C}_0(\mathcal{X})\big) \subset \mathbf{C}_0(\mathcal{X})$. Then (T_t) is strongly continuous on $\mathbf{C}_0(\mathcal{X})$.*

(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 $f \in \mathbf{C}_0(\mathcal{X})$.)

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

We then show that $R_\alpha(\mathbf{C}_0(\mathcal{X}))$ is dense in $\mathbf{C}_0(\mathcal{X})$. If not, since $\mathbf{C}_0(\mathcal{X})^*$ 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

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

It follows by the *(first) resolvent identity*

$$R_\beta = R_\alpha - (\beta - \alpha) R_\alpha R_\beta, \qquad \forall \alpha, \beta > 0, \tag{4.15}$$

we have

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

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

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

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

4.17.4 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 $\mathcal{F}_t^+ := \bigcap_{r>t} \mathcal{F}_r$ denote its right-continuous version.

Proposition 4.17.24 *Let (X_t) be a Markov process with right-continuous sample paths. If its transition semigroup (T_t) leaves $\mathrm{BC}(\mathcal{X})$ or $\mathbf{C}_0(\mathcal{X})$-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

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

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

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

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

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

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

Let τ be a stopping time and recall that

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

defines a σ-field. The following two lemmas are standard: With this one can show that

Lemma 4.17.25 *If (X_t) is adapted and right-continuous, then $X_\tau \mathbf{1}_{\tau<\infty} \in \mathcal{F}_\tau$.*

This follows from approximation of the stopping time as follows.

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

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

The next theorem shows that Feller processes are strong Markov:

Theorem 4.17.26 *Let (X_t) be a right-continuous Markov process whose transition function leaves either $\mathbf{C}_0(\mathcal{X})$ 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:*

$$\mathbb{E}\big[\Phi\circ\theta_\tau \mathbf{1}_{\{\tau<\infty\}}\,|\,\mathcal{F}_\tau\big] = \mathbf{1}_{\{\tau<\infty\}}\mathbb{E}_{X_\tau}[\Phi], \tag{4.16}$$

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 $\{t_k : k\in\mathbf{N}\}$ with $0\leqslant t_1<t_2<\cdots<\cdots\leqslant\infty$. Then, using Theorem 3.4.14, we get for each $B\in\mathcal{F}_\tau$,

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

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:

$$\mathbb{E}\big[f(X_{t+\tau})\mathbf{1}_{\{\tau<\infty\}}|\mathcal{F}_\tau\big] = T_t f(X_\tau)\mathbf{1}_{\{\tau<\infty\}}. \tag{4.17}$$

Now assume a general τ, for the approximating sequence of theorem 3.5.3,

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

By the right-continuity of X and the Feller property of T_t, for any $f\in BC$ (or $f\in\mathbf{C}_0(\mathcal{X})$), (4.17) 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. □

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

Example 4.17.27 Let us return to Example **??**, 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:

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

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

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

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

Chapter 5

Weak convergence and solutions of martingale problems

We have previously discussed weak convergence, Prohorov's theorem, tightness. In this set of lecture we cover weak convergence on the space of continuous processes, touching on Cádlág processes.

In our next 2 lectures, we shall cover the following only: the Ascoli-Arzelá Theorem for tightness (Theorem 5.1.9), Kolomgorov's theorem for tightness (Theorem 5.1.15), and an application an application to the existence of martingale solution (Theorem 5.1.20).

5.1 Weak convergence

A family of random variables / stochastic processes is said to converge weakly, otherwise known as convergent in distribution, if their probability distributions converge. To prove that a family of stochastic processes is weakly convergent, we follow two steps:

(1) We demonstrate that the family of their probability measures forms a relatively compact subset of a suitable function space; in our context, this is typically the Wiener space.

(2) We show that all accumulation points of the family are identical, ensuring convergence to a single limit in the distribution sense.

We emphasise the standing assumption that the state space $\mathcal{X}$ of the random variables under consideration is a complete, separable metric space, and $\mathcal{B}(\mathcal{X})$ is its Borel σ-algebra. Recall that a subset of a topological space is relatively compact it its closure is compact. A set is compact if every cover of the set by open set contains a finite sub-cover. In a complete separable metric space, a set is relatively compact if and only if it is sequentially compact, meaning that every sequence from the set has a convergent subsequence in $\mathcal{X}$.

Let $\mathbb{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$.

Definition 5.1.1 Let $\mathbb{P}_n, \mathbb{P} \in \mathbb{P}(\mathcal{X})$. Suppose that for every bounded continuous function $f : \mathcal{X} \to \mathbf{R}$,

$$\int_{\mathcal{X}} f d\mathbb{P}_n \to \int_{\mathcal{X}} f d\mathbb{P}, \tag{5.1}$$

we say that $\mathbb{P}_n$ converges **weakly** to $\mathbb{P}$, denoted by $\mathbb{P}_n \stackrel{(w)}{\to} \mathbb{P}$.

An equivalent criterion is that (5.1) holds for every bounded Lipschitz continuous function. This is substantiated by the fact that for every closed set $F \subset \mathcal{X}$,

$$f_\epsilon(x) = (1 - \frac{1}{\epsilon} d(x, F)),$$

where $d(x, A) = \inf_{y \in A} d(x, y)$ denote the distance from x to A, is a bounded Lipschitz continuous function. Moreover, $\mathbf{1}_F \leqslant f_\epsilon \leqslant \mathbf{1}_{F^\epsilon}$, where $F_\epsilon = \{x : d(x, F) \leqslant \epsilon\}$. Therefore, (5.1) holding for all bounded continuous functions implies one of the equivalent statements in the Portmanteau Theorem: $\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F)$ for all closed sets F.

Weak convergence is preserved by continuous mappings.

Proposition 5.1.2 *Let $f : \mathcal{X} \to \tilde{X}$ be a continuous map between metric spaces. If $\mathbb{P}_n, P \in \mathbb{P}(\mathcal{X})$ with $\mathbb{P}_n \to \mathbb{P}$, then the pushed forward measures satisfy: $f^*\mathbb{P}_n \to f * \mathbb{P}$.*

Exercise 5.1.3 Suppose that x_n, y_n, x are random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a complete separable metric space $(\mathcal{X}, d)$, and assume the following conditions:

(1) $x_n \to x$ weakly.

(2) $d(x_n, y_n) \to 0$.

Prove that $y_n \to y$ weakly.

Hint: The weak convergence of $x_n \to x$ can be charecterised by the condition: $\limsup_{n\to\infty} \mathbb{P}(x_n \in F) \leqslant \mathbb{P}(x \in F)$ for any closed set F in $\mathcal{X}$. Note that if you define $F_\epsilon = \{y : d(y, F) \leqslant \epsilon\}$ as the ϵ-expansion of F, it holds that $F_\epsilon \downarrow F$ and

$$\mathbb{P}(y_n \in F) \leqslant \mathbb{P}(x_n \in F^\epsilon) + \mathbb{P}(d(x_n, y_n) > \epsilon).$$

5.1.1 Tightness

For probability measures, there is the concept of a probability measure being tight and a family of probability measures being tight (uniformly tight).

Definition 5.1.4 A measure is **tight** if, for any $\epsilon > 0$, there exists a compact set K such that $\mathbb{P}(K) > 1 - \epsilon$. Similarly, a family of probability measures is **tight** if, for any $\epsilon > 0$, there exists a compact set K such that $\mathbb{P}(K) > 1 - \epsilon$ for every measure $\mathbb{P}$ in the family.

Any finite family of probability measures on a complete separable metric space is tight.

Theorem 5.1.5 (Prohorov theorem) *On a complete separable metric space, a set of probability measures is tight iff it is relatively compact.*

Corollary 5.1.6 *If $A = \{\mu_\epsilon : \epsilon \in (0,1]\}$ is a tight family of Borel probability measures on $\mathcal{X}$, and each weakly convergent sequence $\{P_{\epsilon_n}\}$, where $\epsilon \downarrow 0$, from A has a convergent subsequence with limit μ, then $\mu_\epsilon \to \mu$ weakly.*

Proof Suppose that μ_ϵ does not converge to μ weakly. Then, there exists a bounded continuous function $f : \mathcal{X} \to \mathbf{R}$ such that

$$\int_{\mathcal{X}} f d\mu_\epsilon - \int_{\mathcal{X}} f d\mu \not\to 0.$$

Consequently, for some $\epsilon > 0$, there exists an decreasing sequence ϵ_n with $\mu_{\epsilon_n} \in A$ such that

$$\Big|\int_{\mathcal{X}} f d\mu_{\epsilon_n} - \int_{\mathcal{X}} f d\mu\Big| > \epsilon.$$

However, applying the tightness assumption, we find a sub-sequence $\mu_{\epsilon_{n_k}}$ such that $\int_{\mathcal{X}} f d\mu_{\epsilon_{n_k}} \to \int_{\mathcal{X}} f d\mu$. This results in a contradiction, as the inequality above would be violated by the convergent subsequence. Hence, μ_ϵ must converge weakly to μ. □

5.1.2 Tightness on the Wiener space

A reference for this section is Chapter 13, Revuz-Yor. We first study the weak convergence of continuous stochastic processes to continuous stochastic processes. A continuous stochastic process on $\mathbf{R}^d$ with time horizon $[0, T]$ has trajectories in the

Wiener space $\mathbf{C}([0,T];\mathbf{R}^d)$, while the later with the supremum norm is a separable Banach space. On $\mathbf{C}(\mathbf{R}_+;\mathbf{R}^d)$, we may define the metric:

$$d(\omega,\omega') = \sum_{n=1}^{\infty} 2^{-n} \frac{\sup_{t\leqslant n} |\omega(t)-\omega'(t)|}{1+\sup_{t\leqslant n} |\omega(t)-\omega'(t)|}.$$

We denote by I the time interval $[0,T]$ or $\mathbf{R}_+$ and W^d the Wiener space on either.

Definition 5.1.7 • The finite dimensional distributions of a measure on the Wiener space are the pushed forward measures: $(\pi_{t_1,\dots,t_n})_*\mu$ where $n \in \mathbf{N}_0$, $t_1,\dots,t_n \in I$ and

$$\pi_{t_1,\dots,t_n} : \omega \in W^d \longrightarrow (\omega(t_1),\dots,\omega(t_n)) \in \mathbf{R}^{nd}$$

are the multi-coordinate projections.

- The finite dimensional distributions of a continuous stochastic process $(X_t, t \in [0,T])$ are that of its probability distribution on $\mathbf{C}([0,T];\mathbf{R}^d)$.
- If for a sequence of stochastic processes (X_n), for every collection $(t_1,\dots,t_n)$, $(X^n_{t_1},\dots,X^n_{t_n})$ converges in law, the sequence of stochastic processes is said to converge in finite dimensional distributions.

Furthermore, the collection of cylindrical sets of the form

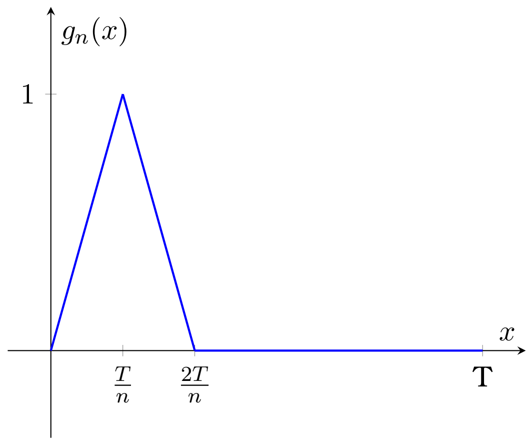
$$\{\omega : \omega(t_i) \in A_i, i = 1,\dots,n\},$$

generates the Borel σ algebra. Here $A_i \in \mathcal{B}(\mathbf{R}^d)$ and t_i is in the time interval. This can be verified with the fact that projections are continuous functions, the cylindrical set is the pre-images of $\Pi_{i=1}^n A_i$ by $\pi_{t_1,\dots,t_n}$ is in the Borel σ-algebra. Conversely, any closed ball is countable intersection of cylindrical sets:

$$\{\tilde\omega : |\omega-\tilde\omega|_\infty \leqslant a\} = \{\tilde\omega : |\omega(q)-\tilde\omega(q)| \leqslant a, \forall q \in \mathbb{Q}\}$$

where $\mathbb{Q}$ is the set of rational numbers and thus the Borel measurable sets are in the σ-algebra generated by the cylindrical sets. Consequently, the measures on the Wiener space are determined by their values on cylindrical sets, and therefore the finite dimensional distributions of a stochastic process uniquely determines its probability distribution.

However if $\mu_n \in \mathbb{P}(\mathbf{C}([0,T];\mathbf{R}^d)$, the convergence of the finite dimensional distributions of μ_n does not necessarily imply the weak convergence of μ_n. Take for example $\mu = \delta_g$, $\mu_n = \delta_{g_n}$ where $g \equiv 0$ and g_n are continuous piecewise linear determined by: $g_n(x) = 0$, $g_n(\frac{T}{n}) = 1$, and $g_n(\frac{2T}{n}) = 0$.



Then, $\mu_n \to \mu$ in finite dimensional distributions on $\mathbf{C}([0,T];\mathbf{R}^d)$. Specifically, for any given $(t_1,\dots,t_n)$, where $t_i > 0$, $\delta_g(\omega(t_i) \in A_i, i = 1,\dots,n)$ takes the values $\{0,1\}$, it is 1 only if $0 \in \cap_i A_i$. One can choose n sufficiently large so that $\frac{2T}{n} \geqslant \min(t_1,\dots,t_n)$. For such n, $\delta_{g_n}(\omega(t_i) \in A_i, i = 1,\dots,n) = 1$ only if $0 \in \cap_i A_i$.

However, μ_n does not converge weakly. For instance, consider the bounded continuous function $\Phi(f) = \min(|f|_\infty, 1)$ on the Wiener space. In this case,

$$\int_{\mathbf{C}([0,T];\mathbf{R}^d)} \Phi(f)\mu(df) = \Phi(g) = 0, \qquad \int_{\mathbf{C}([0,T];\mathbf{R}^d)} \Phi(f)\mu_n(df) = \Phi(g_n) = 1,$$

as $|g_n|_\infty = 1$.

To investigate tightness of measures on the Wiener space, we describe its relatively compact sets. For $\delta > 0$, and for any function $f \in \mathbf{C}([0,T];\mathbf{R}^d)$, define its modulus of continuity as follows:

$$V_\delta(f) = \sup_{s,t\in[0,T]:|t-s|\leqslant\delta} |f(t) - f(s)|. \tag{5.2}$$

A function : $[0,T] \to \mathbf{R}^d$ is uniformly continuous if and only if $\lim_{\delta\to 0} V_\delta(f) = 0$.

Observe that $\delta \mapsto V_\delta(f)$ is an increasing function. Moreover, since f is uniformly continuous, $\lim_{\delta\to 0} V_\delta(f) = 0$. Furthermore, $|V_\delta(f) - V_\delta(g)| \leqslant 2|f-g|_\infty$.

The following characterisation for relatively compact subsets follows from the Arzelá-Ascoli Theorem:

Proposition 5.1.8 (Ascoli-Arzelá Theorem) *A subset D of $\mathbf{C}([0,T];\mathbf{R}^d)$ is relatively compact if and only if:*

(1) $\sup_{f\in D} |f(0)| < \infty$*;*

(2) Uniform continuity, uniformly over D, i.e.:

$$\lim_{\delta\to 0}\sup_{f\in D} V_\delta(f) = 0.$$

Passing to the tightness of measures, we present the following theorem. Recall that $\pi_0 : \mathbf{C}([0,T];\mathbf{R}^d) \to \mathbf{R}^d$ denotes the projection $\pi_0(f) = f(0)$, and $(\pi_0)_*\mu(A) = \mu(\{\omega : \omega(0) \in A\})$ denotes its pushed forward measure on $\mathbf{R}^d$.

Theorem 5.1.9 (Ascoli-Arzelá Theorem for tightness) *A family $A \subset \mathbb{P}(\mathbf{C}([0,T];\mathbf{R}^d)$ is tight if and only if the following holds:*

(1) The set of measures $\{(\pi_0)_*\mu : \mu \in A\}$ on $\mathbf{R}^d$ is tight.

(2) $V_\delta \to 0$ in probability, uniformly over A, meaning that for any $\eta > 0$,

$$\lim_{\delta\to 0}\sup_{\mu\in A} \mu(\{V_\delta(f) \geqslant \eta\}) = 0.$$

Proof Firstly assume that A is relatively compact, therefore it is tight by Prohorov's theorem. For any $\epsilon > 0$, there exists K_ϵ compact with $\mu(K_\epsilon) \geqslant 1-\epsilon$.

Since K_ϵ is compact, according to Proposition 5.1.13, there exists C_ϵ such that $|\omega(0)| \leqslant C_\epsilon$ for any $\mu \in A$, hence $\mu(\omega : |\omega(0)| > C_\epsilon) \leqslant \mu(K_\epsilon^c) < \epsilon$. Furthermore, for any $\eta > 0$, there exists δ such that $V_\delta(f) \leqslant \eta$ for all $f \in K_\epsilon$, hence fro $\mu \in A$,

$$\mu(\omega : V_\delta(\omega) \geqslant \eta) \leqslant \mu(K_\epsilon^c) \leqslant \epsilon,$$

proving (ii).

Conversely assuming (i) and (ii) holds. For any $\epsilon > 0$, we construct a relatively compact subset K_ϵ. Firstly there exists $C_\epsilon > 0$, such that $\mu(|\omega(0)| \geqslant C_\epsilon) < \frac{1}{2}\epsilon$. Set $A_{C_\epsilon} = \{\omega : |\omega(0)| \leqslant C_\epsilon)$. For any $\eta = \frac{1}{m}$, there exists $\delta_{m,\epsilon}$ such that

$$\mu(\omega : V_{\delta_{\epsilon,m}}(\omega) \geqslant \frac{1}{m}) < 2^{-m+1}\epsilon.$$

Set

$$K_{\epsilon,m} = \{\omega : V_{\delta_{\epsilon,m}}(\omega) \leqslant \frac{1}{m}\}$$

and

$$K_\epsilon = A_{C_\epsilon} \cap (\cap_{m=1}^\infty K_{\epsilon,m}).$$

Then

$$\mu(K_\epsilon^c) \leqslant \mathbb{P}(A_{C_\epsilon}) + \sum_{k=1}^m \mu(K_{\epsilon,m}) < \epsilon$$

On K_ϵ, $|\omega(0)| \leqslant C_\epsilon$ and for any $\eta > 0$, choose k_0 with $\frac{1}{k_0} < \eta$, for any $\delta < \delta_{\epsilon,m}$,

$$\lim_{\delta\to 0} V_\delta(\omega) \leqslant V_{\delta_{\epsilon,m}}(\omega) \leqslant \frac{1}{k_0} < \eta,$$

and $\lim_{\delta\to 0} V_\delta(\omega) = 0$, so K_ϵ is compact by Proposition 5.1.13.

If K is a compact set, then $\sup_{\omega\in K} |\omega(0)| \leqslant C$ for some constant C and for any η there exists δ_0 such that

$$\sup_{\omega\in K} w_{\delta_0}(\omega) < \eta.$$

Since w_δ increases with δ, the above holds for any $\tilde{\delta} > \delta$. Let us denote this set by $K_{C,\delta,\eta}$.

If $\{\mathbb{P}_n\}$ is relatively compact, hence tight then, for any $\epsilon > 0$, there exist C, η, η such that $P_n(K_{C,\delta,\eta}) > 1 - \epsilon$. This proves the two conditions hold for any $n \geqslant 0$.

If condition (i) and (ii) are satisfied by $\{\mathbb{P}_n\}_{n\geqslant n_0}$, since $\{\mathbb{P}_n\}_{n\leqslant n_0}$ is tight and satisfy (i) and (ii) uniformly we may assume the conditions are satisfied for all n. □

Note that a family $A \subset \mathbb{P}(\mathbf{C}([0,T];\mathbf{R}^d)$ is relatively compact if and only every sequence from A is relatively compact.

Remark 5.1.10 The two conditions can be rephrased as follows:

(i) For any $\epsilon > 0$ there exists a number $C > 0$, such that $\mu(|\omega(0)| \leqslant C) \geqslant 1 - \epsilon$ for any $\mu \in A$.

(ii) For any $\epsilon > 0$ and $\eta > 0$, there exist $\delta > 0$ such that for any $\mu \in A$,

$$\mu(\omega : V_\delta(\omega) \leqslant \eta) \geqslant 1 - \epsilon.$$

For (ii) it is sufficient to note that w_δ decreases with δ so for any $\tilde{\delta} < \delta$,

$$\mu(\omega : V_{\tilde{\delta}}(\omega) \leqslant \eta) \geqslant \mu(\omega : V_\delta(\omega) \leqslant \eta) \geqslant 1 - \epsilon.$$

For a sequence $A = \{\mathbb{P}_n\}$, (2) can be rephrased as: for any $\epsilon > 0, \eta > 0$ there exists n_0 and δ such that for $n \geqslant n_0$,

$$\mathbb{P}_n(\omega : V_\delta(\omega) \leqslant \eta) \geqslant 1 - \epsilon.$$

Exercise 5.1.11 let $\mathcal{X}$, $\mathcal{Y}$ be complete separable metric spaces. Show that if A is tight on $\mathcal{X}$ and $f : \mathcal{X} \to \mathcal{Y}$ is continuous, then $\{f^*\mu : \mu \in A\}$ is tight.

5.1.3 On Infinite Horizon Wiener Space

Remark 5.1.12 In the above consideration we have taken the time interval in the Wiener space being $[0,T]$. If it is $\mathbf{R}_+$, we only need to consider the modulus of continuity up to each time $[0,N]$ where $N \in \mathbf{N}_0$.

To work with $\mathbf{C}(\mathbf{R}_+;\mathbf{R}^d)$, we take

$$V_\delta^N(f) = \sup_{s,t\in[0,N]:|t-s|\leqslant\delta} |f(t)-f(s)|.$$

Proposition 5.1.13 (Arzelá-Ascoli Theorem -II) *A subset D of $\mathbf{C}(\mathbf{R}_+;\mathbf{R}^d)$ is relatively compact if and only if:*

(1) $\sup_{f\in D}|f(0)| < \infty$*;*

(2) For any N,

$$\lim_{\delta\to 0}\sup_{f\in D} V_\delta^N(f) = 0.$$

Theorem 5.1.14 *A family $A \subset \mathbb{P}(\mathbf{C}([0,T];\mathbf{R}^d))$ is tight if and only if the following holds:*

(1) The set of measures $\{\pi_0^*\mu : \mu \in A\}$ on $\mathbf{R}^d$ is tight.

(2) For every $\epsilon > 0, \eta > 0$ and $N \in \mathbf{N}$, there exists a δ with

$$\sup_{\mu\in A} \mu(\{V_\delta^N(f) \geqslant \eta\}) < \epsilon.$$

The proof in the last theorem is as before except that we would take further intersection in the construction of the relatively compact set: Let $C_{N,\epsilon}$ and $\delta_{N,\epsilon,m}$ be numbers such that for all μ,

$$\mu(|\omega(0)| > C_{N,\epsilon}) \leqslant 2^{-N-1}\epsilon$$

and

$$\mu(\delta_{N,\epsilon,m}(\omega) > \frac{1}{m}) \leqslant 2^{-m-N-1}\epsilon.$$

Set

$$K_\epsilon = \cap_N\{|\omega(0)| \leqslant C_{N,\epsilon} \cap K_{\delta_{N,\epsilon,m}}\}.$$

Then K_ϵ is compact and $\mu(K_\epsilon^c) \leqslant \epsilon$.

5.1.4 Kolmogorov's Theorem –tightness

Since the modulus of continuity $|f(s)-f(t)| \leqslant \sup_{s\neq t} \frac{d(f(x),f(y))}{|s-t|^\alpha}|s-t|^\alpha \leqslant |f|_\alpha |s-t|^\alpha$, if $|f|_\alpha$ is uniformly bounded then the family is equi-continuous. The control the Hölder norm of the stochastic process we use Kolmogorov's theorem.

Theorem 5.1.15 *[Kolmogorov's Theorem for tightness] Let (X^n) be a sequence of $\mathbf{R}^d$-valued continuous processes such that*

(1) The family of initial laws $\mathcal{L}(X_0^n)$ is tight.

(2) There exists numbers $p>1, \beta>0, \alpha>\frac{d}{p}$ such that $\alpha p > d$ and for every $s,t \in [0,N]$, and every n,

$$\|X_s^n - X_t^n\|_p \leqslant \beta |s-t|^\alpha,$$

then the set of laws of $\{X^n\}$ is weakly relatively compact.

Proof We only need to show that for any $\epsilon>0$ and $\eta>0$, there exist δ and an integrer n_0 such that for $n \geqslant n_0$,

$$\mathbb{P}_n(\{V_\delta(\omega) \geqslant \eta\}) < \epsilon.$$

Apply Markov-inequality for condition (2).

$$\mathbb{P}(\sup_{|s-t|<\delta} |X_s^n - X_t^n| \geqslant \eta) \equiv \mathbb{P}_n(\{V_\delta(X^n) \geqslant \eta\}) \leqslant \frac{1}{\eta^p}\mathbb{E} \sup_{|s-t|<\delta} |X_s^n - X_t^n|^p.$$

Recall Theorem 2.8.16 applied to $(X_t^n)_{t\in[0,T]}$: If for some $\alpha>0$, $p>1$, and $C>0$, the following holds:

$$\sup_{s\neq t} \|X_s^n - X_t^n\|_p \leqslant \beta |t-s|^\alpha,$$

then for any $\gamma < \alpha - \frac{d}{p}$,

$$\Big\| \sup_{s\neq t} \frac{|X_s^n - X_t^n|}{|s-t|^\gamma} \Big\|_p \leqslant \beta \tilde{C}.$$

Hence $\lim_{\delta\to 0} \mathbb{E} \sup_{|s-t|<\delta} |X_s^n - X_t^n|^p = 0$, completing the proof. □

5.1.5 Applications

Definition 5.1.16 A sequence of continuous stochastic processes is said to converge in distribution if their probability distributions on the Wiener space converge weakly.

Lemma 5.1.17 *Let $f_n, f : \mathcal{X} \to \mathcal{Y}$ be measurable maps on Banach spaces and such that $f_n(x_n) \to f(x)$ for any sequence x_n in $\mathcal{X}$ converging to x. Then f is continuous and f_n converges to f locally uniformly.*

Proof Suppose that f has a discontinuity at x, then there eixsts $\delta > 0$ and $y_n \to x$ such that

$$|f(y_n) - f(x)| \geqslant \delta.$$

Let $N_0 = 1$, and since $f_n(y_k) \to f(y_k)$

$$N_k = \inf\{N \geqslant N_{k-1} : |f_n(y_k) - f(x)| \geqslant \frac{\delta}{2}, \forall n \geqslant N\} < \infty.$$

Define $x_n = y_k$ if $n \in [N_k, N_{k+1})$ then $x_n \to x$. But $|f_n(x_n) - f(x)| \geqslant \frac{\delta}{2}$ by the construction, contradicting with the assumption and proving that f must be a continuous functions.

Suppose that f_n does nto converge locally uniformlly. Then for any relatively compact set K, there exists $\delta > 0$ and $x_n \in K$ such that

$$|f_{n_k}(x_k) - f(x_k)| > \delta.$$

Now x_k has a convergent subsequence, which we denote by y_k, with limit y. We have

$$|f_{n_k}(y_k) - f(y_k)| > \delta.$$

Since f converges locally uniformlly, there exists N with $|f(y_n) - f(y)| < \frac{\delta}{2}$ for any $k > N$ and

$$|f_{n_k}(y_k) - f(y)| > \delta/2,$$

contradicts the assumption. We have showed therefore for any K compact , any δ there exists N such tht $|f_n(x) - f(x)| < \delta$ for all $x \in K$. □

The following is an extension of the continuous mapping theorem given in Theorem 5.1.18.

Proposition 5.1.18 (Continuous Mapping Theorem) *Let $f_n, f : \mathcal{X} \to \mathcal{Y}$ be measurable maps between metric spaces such that $f_n(x_n) \to f(x)$ for any sequence x_n in $\mathcal{X}$ converging to x. If $\mu_n, \mu \in \mathbb{P}(\mathcal{X})$ with $\mu_n \to \mu$, Then $(f_n)_*\mu_n \to f_*\mu$. In partricular, if ξ_n are random variables converging to ξ in distribution, then $f_n(\xi_n)$ converges to $f(\xi)$ in distribution.*

Proof Denote $\nu_n = (f_n)_*\mu_n$ and $\nu = f_*\mu$. By Portmanteau theorem it is sufficient to show for any $G \subset \mathcal{Y}$ open,

$$\liminf_{n\to\infty} \nu_n(G) \geqslant \nu(G).$$

Fix such an open set $G \subset \mathcal{Y}$. For any $x \in f^{-1}(G)$, there exists a neighbourhood U and a number m such that for all $k \geqslant m$, $f_k(U) \subset G$. Consequently, $x \subset \cap_{k=m}^\infty (f_k^{-1}(G))^o$, where the supersript denotes the interior of a set, in particular it is an open set. Thus,

$$f^{-1}(G) \subset \cup_{m=1}^\infty \cap_{k=m}^\infty (f_k^{-1}(G))^o.$$

Consequently,

$$f_*\mu(G) = \mu(f^{-1}(G)) \leqslant \sup_m \mu(\cap_{k=m}^\infty (f_k^{-1}(G))^o) \leqslant \sup_m \liminf_{n\to\infty} \mu_n(\cap_{k=m}^\infty (f_k^{-1}(G))^o)).$$

We have used $\mu_n \to \mu$. Finally we obtain,

$$f_*\mu(G) \leqslant \sup_m \liminf_{n\to\infty} \mu_n(\cap_{k=m}^\infty (f_k^{-1}(G))^o)) \leqslant \liminf_{n\to\infty} \mu_n(f_n^{-1}(G)) = \liminf_{n\to\infty} (f_n)_*\mu_n(G),$$

which completes the proof. □

We reiterate the following theorem:

Theorem 5.1.19 *Let $X^n, X : \Omega \to W^d$ be measurable functions. Set $X_t^n(\omega) = X^n(\omega)(t)$, $X_t(\omega) = X(\omega)(t)$. Suppose that (X^n) converges in finite dimensional distribution to (X) and if*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}(V_\delta(X^n) \geqslant \epsilon) = 0,$$

then the stochastic processes converge weakly to X on W^d.

Proof Since $X_0^n \to X_0$ in distribution, $\{X_0^n, n \in \mathbf{N}\}$ is tight. Thus both conditions for tightness hold. □

Let $\mathcal{F}_t^0$ denote the σ-algebra on the Wiener space generated by the coordinate process. Let $[a]$ denote the integer part of the number a. Denote by $\mathcal{M}_d^+$ the set of symmetric non-negative definite $d \times d$ matrices.

Theorem 5.1.20 *Let $X_i : \mathbf{R}^d \to \mathbf{R}^d$, $i = 0, 1 \ldots, m$, be bounded continuous, then for any probability measure μ on $\mathbf{R}^d$, there exists a solution to the martingale problem for $\mathcal{L}$ where*

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

$(a_{ij}) = XX^T$, and $X = (X^1, \ldots, X^m)$.

Proof Consider the canonical probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the standard filtration from the coordinate process, on which we have a random variable X_0 with distribution

μ, and an independent Brownian motion B on $\mathbf{R}^d$. We define an approximation for $dx_t = X_0(x_t)dt + \sum_{k=1}^m X_i(x_t)dW_t^k$ as follows. Define for $t \in (0, \epsilon]$:

$$x_t^\epsilon = x_0 + \int_0^t X_0(x_0)ds + \sum_k \int_0^t X_k(x_0)dW_s^k$$

On $(\epsilon, 2\epsilon]$ define,

$$x_t^\epsilon = x_\epsilon^\epsilon + \int_\epsilon^t X_0(x^\epsilon(\epsilon))ds + \sum_k \int_0^t X_k(x^\epsilon(\epsilon))dW_s^k,$$

and iteratedly we defien x_t^ϵ for all t. For any $\epsilon > 0$ and any path $\gamma \in \mathbf{C}(\mathbf{R}_+, \mathbf{R}^d)$, we define: the time dependent vector fields $X_i^\epsilon(t, \gamma)$ as folllows:

$$X_i^\epsilon(t, \gamma) = \sum_{j=1}^\infty X_i(\gamma(j\epsilon))\mathbf{1}_{(j\epsilon,(j+1)\epsilon]}(t).$$

Then x_t^ϵ solvs:

$$x_t^\epsilon = x_0 + \int_0^t X_0^\epsilon(s, x_s^\epsilon) + \int_0^t X_k^\epsilon(s, x_s^\epsilon)dW_s^k.$$

Write

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

Denote by $\mathbb{P}^\epsilon$ the law of x^ϵ on the Wiener space, then

$$f(x_t^\epsilon) - f(x_0) - \int_0^t \mathcal{L}_s^\epsilon f(x_s^\epsilon)ds$$

is a martingale.

Take $\epsilon = \frac{1}{n}$, we change all indices from ϵ to n. We show that $\mathbb{P}^n$ is weakly relatively compact. Firstly, condition (1) is satisfied. Condition (2) follows by BDG inequality. For any n, and $\delta > 0, s \geqslant 0$,

$$\mathbb{E}\sup_{0\leqslant r\leqslant\delta}|x_s^n - x_{s+r}^n|^p \leqslant \delta^{\frac{p}{2}}.$$

For any $p > 1$, we choose $p > d$ to obtain the tightness of $\mathbb{P}^n$.

Then there exists a sequence x^{n_k} that converges weakly to some process x. For simplicity we denote the sub-sequence by x^n as well, and suppose that x^n converges weakly to a stochastic process x. We denote its distribution on W^d by $\mathbb{P}$.

Take a smooth and compactly supported function f, $s < t$, and $g : \mathbf{C}([0, s]; \mathbf{R}^d) \to \mathbf{R}$. We want to show that

$$\mathbb{E}\Big[\Big(f(x_t) - f(x_s) - \int_s^t \mathcal{L}_s(s, x_r)dr\Big)g(x)\Big] = 0.$$

We already know that, for each n,

$$\mathbb{E}\Big(f(x_t^n) - f(x_s^n) - \int_s^t \mathcal{L}_s^n(s, x_r^n)dr\Big)g(x^n) = 0.$$

We invoke the continuous mapping theorem. Write:

$$\Phi(\gamma) = \Big(f(\gamma_t) - f(\gamma_s) - \int_s^t \mathcal{L}_s(s, \gamma_r)dr\Big)g(\gamma),$$

and

$$\Phi_n(\gamma^n) := \Big(f(\gamma_t^n) - f(\gamma_s^n) - \int_s^t \mathcal{L}_s^n(s, \gamma_r^n)dr\Big)g(\gamma^n).$$

Note that for fixed time interval, If $\gamma^n \to \gamma$, $\Phi_n(\gamma_n) \to \Phi(\gamma)$. If $x^n \to x$ in W^d in distribution, then $\Phi_n(x^n) \to \Phi(x)$ in distribution, it follows from the boundedness of Φ_n, that $\mathbb{E}[\Phi(x) = \lim_{n\to\infty} \mathbb{E}[\Phi_n(x^n)] = 0$, so P solves the martingale problem. □

Remark 5.1.21 If we know the martingale problem has at most one solution, then the limiting point would be unique, and this woulld be the weak limit of the processes X^n (the weak limit does exist in this case).

Theorem 5.1.22 *[Billingsley, pp.83] Suppose that* $0 = t_0 < t_1 < \cdots < t_k = 1$ *and*

$$\min_{1<i<k}(t_i - t_{i-1}) \geqslant \delta.$$

Then for any $f : [0, T] \to \mathbf{R}^d$*,*

$$V_f(\delta) \leqslant 3 \max_{1\leqslant i\leqslant k} \sup_{s\in[t_{i-1},t_i]} |f(s) - f(t_{i-1})|$$

and for any probability measure $\mathbb{P}$*,*

$$\mathbb{P}(V_f(\delta) \geqslant 3\epsilon) \leqslant \sum_{i=1}^k \mathbb{P}(\sup_{s\in[t_{i-1},t_i]} |f(s) - f(t_{i-1})| \geqslant \epsilon).$$

Proof Let $M = 3\max_{1\leqslant i\leqslant k}\sup_{s\in[t_{i-1},t_i]}|x(s) - x(t_{i-1})|$. If $|s - t| < \delta$, then they either belong to the same sub-interval $[t_i, t_{i+1}]$, in which case

$$|x(s) - x(t)| \leqslant |x(s) - x(t_i)| + |x(t_i) - x(t)| \leqslant 2M,$$

ot they lie in adjacent intervals, $s \in [t_{i-1}, t_i]$ and $t \in [t_i, t_{i+1}]$. In the latter case,

$$|x(s) - x(t)| \leqslant |x(s) - x(t_{i-1})| + |x(t_i) - -x(t_{i-1})| + |x(t_i) - x(t)| \leqslant 3M,$$

showing $V_f(\delta) \leqslant 3M$. Finally, $V_f(\delta) \geqslant 3\epsilon$ implies that $\max_{1\leqslant i\leqslant k}\sup_{s\in[t_{i-1},t_i]}|x(s) - x(t_{i-1})| \geqslant \epsilon$, and the last inequality holds.

□

Lemma 5.1.23 *[16, pp 517, Lemma (1.7), Chapter 13]. Consider a sequence $\{\mathbb{P}_n\}$ from $\mathbb{P}(\mathcal{X})$. Condition (2) of Theorem 5.1.14 is implied by the following condition:*

For any N, $\epsilon, \eta > 0$ there exists a number $\delta \in (0,1)$ and n_0 such that

$$\frac{1}{\delta}\mathbb{P}_n(\{\omega : |\sup_{t\leqslant s\leqslant t+\delta} |(\omega(s) - \omega(t)| \geqslant \eta\}) \leqslant \epsilon,$$

for any $n \geqslant n_0$, and all $t \in [0, N]$.

Proof Want to show that for every ϵ there exists a $\delta > 0$ such that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}_n(V_\delta(x) \geqslant \epsilon) = 0.$$

Let $\eta > 0$, and $\epsilon > 0$, then there exists $\delta \in (0,1)$, such that

$$\frac{1}{\delta}\mathbb{P}_n(\sup_{s\in[t,t+\delta]} |x(s) - x(t_{i-1})| \geqslant \frac{\epsilon}{3}) \leqslant \frac{\eta}{N}.$$

By the previous theorem, taking $t_i = i\delta$, where $i < [N/\delta]$.

$$\mathbb{P}_n(V_\delta(f) \geqslant \epsilon) \leqslant \sum_{i=1}^{[N/\delta]} \mathbb{P}_n(\sup_{s\in[t_{i-1},t_i]} |x(s) - x(t_{i-1} \wedge N)| \geqslant \frac{\epsilon}{3}) \leqslant [\frac{N}{\delta}]\,\frac{\delta}{N}\eta \leqslant \eta,$$

asserting the statement. □

Given a sequence of mean zero identically distributed real valued random variables ξ_n with variance σ^2. Set

$$S_0 = 0, \qquad S_n = \sum_{k=1}^n \xi_k.$$

Let $X^n(\omega)$ be defined by fixing its values at $\frac{i}{n}$ to be : $X^n(\omega) = \frac{S_i(\omega)}{\sigma\sqrt{n}}$, and linear interpolations between values at $\frac{i}{n}$. Namely $X^n(\omega)$ is a continuous piecewise linear function given by

$$X_0^n = 0, \qquad X^n(t)(\omega) = \frac{S_{[nt]}(\omega)}{\sigma\sqrt{n}} + (nt - [nt])\frac{\xi_{[nt]+1}(\omega)}{\sigma\sqrt{n}}, \qquad t \in [\frac{i}{n}, \frac{(i+1)}{n}].$$

Lemma 5.1.24 *[Billingsley, pp. 88] Suppose that $\{X_n\}$ is stationary, and*

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} \lambda^2\mathbb{P}(\max_{k\leqslant n} |S_k| \geqslant \lambda\sigma\sqrt{n}) = 0.$$

Then $\{X^n\}$ is tight.

Proof We want to show that for any $\epsilon > 0$, there exists δ such that

$$\limsup_{n\to\infty} \mathbb{P}(V_\delta(X^n) \geqslant 3\epsilon) \to 0.$$

By the previous lemmas, Lemma 5.1.22, for a partition of $[0,1]$, $0 = t_0 < t_1 < \cdots < t_N = 1$, of size $\delta > 0$,

$$V_{X^n}(\delta) \leqslant 3 \max_{1\leqslant i\leqslant N} \sup_{s\in[t_{i-1},t_i]} |X^n(s) - X^n(t_{i-1})|.$$

Applying Lemma 5.1.22, and a brutal estimate of the supremum by the sum:

$$\begin{aligned}\mathbb{P}(V_\delta(X^n) \geqslant 3\epsilon) &\leqslant \mathbb{P}(\max_{1\leqslant i\leqslant N} \sup_{s\in[t_{i-1},t_i]} |X^n(s) - X^n(t_{i-1})| \geqslant \epsilon)\\ &\leqslant \sum_{1\leqslant i\leqslant N} \mathbb{P}(\sup_{s\in[t_{i-1},t_i]} |X^n(s) - X^n(t_{i-1})| \geqslant \epsilon).\end{aligned}$$

By the definition, X^n is obtained by interpolation on sub-intervals of size $\frac{1}{n}$. We shall first take $n \to \infty$, then take δ to infinity. Thus we may assume that $\frac{1}{n}$ is small. Denote by m the maximal number of sub-intervals of size $\frac{1}{n}$ that fit into $[t_i, t_{i+1}]$, then $m = [\delta n]$ so $\delta \sim \frac{m}{n}$.

If $s \in [t_i, t_{i+1}]$, where $t_i = i\delta \sim \frac{im}{n}$, suppose that $s \in [(mi+k)/n, (mi+k+1)/n]$. Thus $X^n(t_i) \sim \frac{S_{mi}(\omega)}{\sigma\sqrt{n}}$ and $X_s \sim= \frac{S_{mi+k}(\omega)}{\sigma\sqrt{n}}$, and we have

$$\begin{aligned}X^n(s) - X^n(t_i) &= \frac{S_{mi+k}(\omega)}{\sigma\sqrt{n}} + (ns - im + k)\frac{\xi_{mi+k+1}(\omega)}{\sigma\sqrt{n}} - \frac{S_{mi}(\omega)}{\sigma\sqrt{n}} - (nt_i - im)\frac{\xi_{mi+k}(\omega)}{\sigma\sqrt{n}}.\\ &= \frac{S_{mi+k}(\omega)}{\sigma\sqrt{n}} - \frac{S_{mi}(\omega)}{\sigma\sqrt{n}} + R_i,\end{aligned}$$

where the error term $R_i = (ns - im + k)\frac{\xi_{mi+k+1}(\omega)}{\sigma\sqrt{n}} - (nt_i - im)\frac{\xi_{mi+k}(\omega)}{\sigma\sqrt{n}}$ has mean zero variance less or qual to $\frac{2}{\sigma^2 n^3}$. We have used the fast that ξ_i is a stationary sequence:

$$\frac{S_{mi+k}(\omega)}{\sigma\sqrt{n}} - \frac{S_{mi}(\omega)}{\sigma\sqrt{n}} \overset{\text{(law)}}{=} \frac{S_k(\omega)}{\sigma\sqrt{n}},$$

the latter does not depend on i. Put all together, we have

$$\mathbb{P}(V_\delta(X^n) \geqslant 3\epsilon) \leqslant \sum_{1\leqslant i\leqslant k} \mathbb{P}(\sup_{1\leqslant k\leqslant m+1} \frac{|S_k(\omega)|}{\sigma\sqrt{n}} \geqslant \frac{\epsilon}{2}) + \sum_{1\leqslant i\leqslant k} \mathbb{P}(|R_i| \geqslant \frac{\epsilon}{2}).$$

The second term is controlled by $\sum_{1\leqslant i\leqslant k} \frac{4}{\epsilon^2}\frac{2}{\sigma^2 n^3}$, where keeping $k \sim \frac{1}{\delta}$ fixed, converges to zero as $n \to \infty$. We now put this into the content of our assumption:

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \lambda^2 \mathbb{P}(\max_{k\leqslant n} |S_k| \geqslant \lambda\sigma\sqrt{n}) = 0.$$

It remains to work

$$\sum_{1\leqslant i\leqslant N}\mathbb{P}(\sup_{1\leqslant k\leqslant m+1}|S_k|\geqslant\frac{\epsilon}{2}\sigma\sqrt{n})=\frac{1}{\delta}\mathbb{P}(\sup_{1\leqslant k\leqslant m+1}|S_k|\geqslant\frac{\epsilon}{2}\sigma\sqrt{n}).$$

Set $\lambda=\frac{\epsilon}{2}\sqrt{\frac{n}{m+1}}$, recall that $m=[\frac{\delta}{n}]$, so $\frac{1}{\delta}\leqslant\frac{n}{m}\leqslant\frac{8}{\epsilon^2}\lambda^2$. Then

$$\frac{1}{\delta}\mathbb{P}(\sup_{1\leqslant k\leqslant m+1}|S_k|\geqslant\frac{\epsilon}{2}\sigma\sqrt{n})\leqslant\frac{8}{\epsilon^2}\lambda^2\mathbb{P}(\sup_{1\leqslant k\leqslant m+1}|S_k|\geqslant\lambda\cdot\sigma\sqrt{m+1}),$$

which, by the assumption, converges to zero as $\lambda\to\infty$, completing the proof. □

Before proceeding further we recall the following inequality:

Lemma 5.1.25 (Etemai's inequality) *Let $\{\xi_i\}$ be independent random variables, and $S_k=\sum_{i=1}^k\xi_i$. Then for any number $a>0$,*

$$\mathbb{P}(\max_{1\leqslant k\leqslant m}|S_k|\geqslant 3a)\leqslant 3\max_{1\leqslant k\leqslant m}\mathbb{P}(|S_k|\geqslant a).$$

Lemma 5.1.26 *The conditions of Lemma 5.1.24 is satisfied by $\{X^n\}$.*

Proof For any $\lambda>0$, we want to show that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\lambda^2\mathbb{P}(\max_{k\leqslant n}|S_k|\geqslant\lambda\sigma\sqrt{n})=0.$$

By Etemadi, it is sufficient to show that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\lambda^2\max_{k\leqslant n}\mathbb{P}|S_k|\geqslant\lambda\sigma\sqrt{n})=0.$$

Since $\frac{S_k}{\sigma\sqrt{k}}\to N(0,1)$, denoting η such a standard Gaussian random variable. For Gaussian random variable,

$$\mathbb{P}(|\eta|\geqslant\lambda)\leqslant\frac{\mathbb{E}[\eta^4]}{\lambda^4}=3\lambda^{-4}.$$

We may choose k_0 such that for $k\geqslant k_0$, the error between $\mathbb{P}|S_k|\geqslant\lambda\sigma\sqrt{n})$ and $\mathbb{P}(|\eta|\geqslant\lambda)$ is small. For sufficiently large n, we may assume that $k_0\leqslant n$. For $k\geqslant k_0$,

$$\mathbb{P}|S_k|\geqslant\lambda\sigma\sqrt{n})\leqslant\mathbb{P}|S_k|\geqslant\lambda\sigma\sqrt{k})\leqslant\lambda^{-4}+\mathbb{P}(|\eta|\geqslant\lambda)\leqslant\frac{4}{\lambda^4}.$$

For $k\leqslant k_0\leqslant n$,

$$\mathbb{P}|S_k|\geqslant\lambda\sigma\sqrt{n})\leqslant\frac{\mathbb{E}|S_k|^2}{\lambda^2\sigma^2 n}=\frac{k_0}{\lambda^2 n}.$$

Overall,

$$\mathbb{P}(|S_k|\geqslant\lambda\sigma\sqrt{n})\leqslant\max(\frac{k_0}{\lambda^2 n},\frac{4}{\lambda^4})\to 0$$

as $n\to\infty$, proving the desired inequality. □

Proposition 5.1.27 (Donsker's theorem /Invariance Principle) *Given a sequence of mean zero independent identically distributed real valued random variables ξ_n with variance σ^2. Set*

$$S_0 = 0, \qquad S_n = \sum_{k=1}^n \xi_k.$$

Then X^n, where it is defined by piecewise interpolation by

$$X_t^n(\omega) = \frac{S_{[nt]}(\omega)}{\sigma\sqrt{n}} + (nt - [nt])\frac{\xi_{[nt]+1}(\omega)}{\sigma\sqrt{n}},$$

converges in distribution to the standard Brownian motion.

Proof Since $\{X^n, n \in \mathbf{N}\}$ is relatively compact ,by Lemma 5.1.26, it remains to identify the limiting distributions. Observe that $\frac{[nt]}{n} \to t$, it follows by Markov-Chebeshev inequality that

$$R_n(t) := (nt - [nt])\frac{\xi_{[nt]+1}}{\sigma\sqrt{n}} \to 0.$$

By the central limit theorem, X_t^n converges to the standard normal distribution for every t. To obtain convergence in distribution of the stochastic process, we work with their increments:

$$(X_s^n, X_t^n - X_s^n) = (\frac{1}{\sigma\sqrt{n}}S_{[ns]}, \frac{1}{\sigma\sqrt{n}}(S_{[nt]} - S_{[ns]})) + (R_s^n, R_t^n - R_s^n),$$

Now $(\frac{1}{\sigma\sqrt{n}}S_{[ns]}, \frac{1}{\sigma\sqrt{n}}(S_{[nt]} - S_{[ns]})) \to (N_1, N_2)$, where (N_1, N_2) are independent standard normal distribution with variance s and $t - s$ respectively. By the continuous mapping theorem

$$(\frac{1}{\sigma\sqrt{n}}S_{[ns]}, \frac{1}{\sigma\sqrt{n}}(S_{[nt]}) \to (N_1, N_1 + N_2).$$

Since $(R_s^n, R_t^n - R_s^n) \to (0, 0)$ in probability,

$$(X_s^n, X_t^n) \to (N_1, N_1 + N_2).$$

Similarly one observe that as $n \to \infty$,

$$(X_{t_1}^n, X_{t_2}^n - X_{t_1}^n, \dots, X_{t_k}^n - X_{t_k-1}^n) \to (B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_k-1})$$

in law where B is a one dimensional Brownian motion. Consequently, as $n \to \infty$, (X_t^n) converges in distribution to a Brownian motion. □

5.1.6 Cádlág processes

We briefly discuss the analogue for the space $\mathbb{D}$, of Cádlág processes. Such processes has at mostly countable number of discontinuity, and there can be at most a finite number of points at which the jump exceeds a certain given size. Consequently, $\sup_t |f(t)| < \infty$ for $f \in D$.

Define for any interval I:

$$V(f, I) = \sup_{s,t,\in I} |f(s) - f(t)|.$$

Since Cádlǵ functions are only right continuous, we are interested in intervals of the form $[s, t)$. The idea is to take t to be one of the point of discontinuity thus introducing special partitions of $[0, 1]$.

Let us consider δ-sparse sets: this is a finite collection of points $\{t_i\}$ with

$$\min(t_t - t_{i-1}) > \delta.$$

For $0 < \delta < 1$, define

$$V'_\delta(f) = \inf_{\delta-\text{sparse}\{t_i\}} \max_i \sup_{s,t,\in[t_i,t_{i+1})} |f(s) - f(t)|.$$

The infimum is taken over all δ-spare sets $\{t_i\}$. Then,

$$\lim_{\delta\to 0} V'_\delta(f) = 0$$

is necessary and sufficient for a function f to belong to D. The idea is that we can choose a δ-sparse set containing the points of discontinuous. One can define a metric on D as follows.

Definition 5.1.28 Let Λ be a set consisting of strictly increasing and continuous mapping from $[0, 1]$ onto $[0, 1]$ with $\lambda(0) = 0$ and $\lim_{t\uparrow\infty} \lambda(t) = \infty$. A function in Λ is referred as a time change.

Its possible to define a metric with

$$d(f, g) = \inf_{\lambda\in\Lambda} (\sup_t |\lambda(t) - t| + |f \circ \lambda - g|_\infty\}.$$

This defines a (incomplete) metric on D. Then D is a separable Polish space. (By Polish we mean that there exists a complete metric inducing the same topology).

Definition 5.1.29 The Skorohod topology, on the cádlág space $\mathbb{D}(\mathbf{R}_+; \mathbf{R}^d)$, is characterised by the following convergence. A sequence $f_n \to f$ if and only if there exists a sequence $\lambda_n \in \Lambda$ with the following holds:

- $\sup_t |\lambda_n(t) - t| \to 0$
- $\sup_{t \leqslant N} |f_n \circ \lambda_n - f| \to 0$ for all N.

Proposition 5.1.30 *If f is a continuous function, then a sequence $f_n \in \mathbb{D}$ converge in the Skorohod topology if and only if they converge locally uniformly.*

Proof To see this note that

$$|f_n(t) - f(t)| \leqslant |f_n \circ \lambda_n \circ \lambda_n^{-1}(t) - f \circ \lambda_n^{-1}(t)| + |f \circ \lambda_n^{-1}(t) - f(t)|.$$

Note that if the distance between λ and the identity map is less than 1, then

$$\sup_{t \in [0,N]} |f_n(t) \circ \lambda_n \circ \lambda^{-1}(t) - f \circ \lambda_n^{-1}(t)| \leqslant \sup_{t \in [0,N+1]} |f_n \circ \lambda_n(t) - f(t)| \to 0,$$

by the convergence in $\mathbb{D}$. Also by the uniform continuity of f,

$$\lim_{n \to \infty} \sum_{t \in [0,N]} |f \circ \lambda_n^{-1}(t) - f(t)| \to 0,$$

concluding the proof for the assertion. □

Theorem 5.1.31 *A necessary and sufficient condition for a set in $\mathbb{P}(\mathcal{X})$ to be relatively compact is that:*

$$\sup_{f \in A} \|f\| < \infty$$

and

$$\lim_{\delta \to 0} \sup_{f \in A} V'_\delta(f) = 0.$$

To identify the limit, we only need to show the convergence of the finite dimensional distributions (with times in T_P: the collection of points for which the projection π_t os continuous except for a set of P-measure zero). It is sufficient to identify the limit on a dense subset of $[0, T]$.

An interesting concept is C-tight.

Definition 5.1.32 A sequence of stochastic processes X^n is C-tight if it is tight and that all limit point of the sequence are probability distributions of a continuous process, in other words these probability distribution has full measure on the Wiener space.

This allows to discuss the convergence of discontinuous processes and obtaining another version of Donsker's type theorem for

$$\frac{1}{\sigma\sqrt{n}} S_{[nt]}.$$

Chapter 6

Ergodic Theorems

6.0.1 The adjoint operator

Let $\mathcal{L}$ denote the generator of a Markov process. If π is an invariant probability measure, then $\int T_t f d\pi = \int f d\pi$. If f is in the domain of its generator and if we can exchange the order of integration and differentiation, then $\int \mathcal{L} f d\pi = 0$. This procedure usually holds for $f \in \mathbf{C}_K^\infty$. If the invariant measure is of the form $d\pi = \varrho\, dx$ where $\varrho : \mathcal{X} \to \mathbf{R}$ is a density function, and if $\mathcal{L}^*$ denote the L^2-adjoint of $\mathcal{L}$, then formally, $\int_{\mathcal{X}} f \mathcal{L}^* \varrho\, dx = 0$. Hence we often look for a solution of $\mathcal{L}^* \varrho = 0$ and then proceed to show ϱdx is an invariant measure.

Example 6.0.1 If

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

Then for any C^2 function f with compact support and any function g, we may apply integration by parts formula:

$$\int \mathcal{L} f g\, dx = -\int f \frac{\partial g}{\partial x_k} dx - \int f(g\, \mathrm{div}\, b) dx + \frac{1}{2} \sum_{i,j=1}^{d} \int f \frac{\partial^2 (a_{i,j} g)}{\partial x_i \partial x_j}.$$

Hence

$$\mathcal{L}^* g = -b_k \sum_{k=1}^{d} \frac{\partial g}{\partial x_k} + \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j} \frac{\partial^2 g}{\partial x_i \partial x_j} + \sum_j \sum_i \frac{\partial a_{i,j}}{\partial x_i} \frac{\partial g}{\partial x_j} + \frac{1}{2} \frac{\partial^2 a_{i,j}}{\partial x_i \partial x_j} - g\, \mathrm{div}\, b.$$

For compact manifolds, if $\mathcal{L}$ is smooth and strictly elliptic, then it has a unique invariant probability measure. This follows from the fact that the semi-group is strong Feller.

Example 6.0.2 Let $\mathcal{L}f(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x)dx$, then

$$\int \mathcal{L}f\varrho dx = \int f\mathcal{L}^*\varrho dx = \int f(-(b\varrho)'(x) + \frac{1}{2}(a\varrho)'')dx.$$

Formally, ϱ satisfies,

$$\frac{1}{2}a\varrho'' + (a' - b)\varrho' + (\frac{1}{2}a'' - b')\varrho = 0.$$

6.1 L^p-Semigroups and Invariant Measure

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

$$\mathbb{E}[f(X_{t+s}) \mid \mathcal{F}_s] = T_t f(X_s) \tag{6.1}$$

for any $f \in E$. See Theorem 4.1 in [3, pp182]. Furthermore uniqueness holds for the martingale problem for $\mathcal{L}$ with the initial distribution μ.

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

Lemma 6.1.1

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

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

  In other words,

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

- *(Young Inequality) Let $f, K : \mathbf{R}^n \to \mathbf{R}$ 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 6.1.2 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, first let f be smooth with compact support. For any $\epsilon > 0$ choose $\delta > 0$ so $|f(x) - f(y)| < \epsilon/2$ for $|x - y| < \delta$ and let $K_t(x) = P_t(0, x)$.

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

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

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

Let X be a Markov process on $\mathcal{X}$ with transition semi-group T_t on $\mathcal{B}_b(\mathcal{X})$. Recall that a probability 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 6.1.3 *Let $\pi \in \mathbb{P}(\mathcal{X})$ 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 $L^p(\mathcal{X}, \pi)$.*

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

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

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 $f \in \mathrm{B}C(\mathcal{X}) \cap L_p$, $|T_t f - f|_{L^p} \to 0$ 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. □

6.2 Characterisation of Invariant Measures

The following theorem unifies several notions of invariant measures, see [3, pp239].

Theorem 6.2.1 *Suppose that $\mathcal{L}$ generates a strongly continuous contraction semi-group T_t on E and E is measure determining ($\mathcal{B}_b$ is dense in E), and the martingale problem for $\mathcal{L}$ is well posed. Let (right continuous) process X_t be the solution for the martingale problem for $\mathcal{L}$ with the initial condition π. Then the following is equivalent for a probability measure π.*

1. *The distribution of X_t is μ for all time $t \geqslant 0$.*
2. *$\theta_t X$ and X have the same finite dimensional distributions.*
3. *$\int_{\mathcal{X}} T_t f \, d\pi = \int_{\mathcal{X}} f \, d\pi$, for every $f \in E, t \geqslant 0$.*
4. *$\int_{\mathcal{X}} \mathcal{L} f d\pi = 0$ for any $f \in \mathrm{Dom}(\mathcal{L})$.*

Proof

- (ii) obviously implies (i).
- (i) $\Longrightarrow$ (ii) If $\mathcal{L}(X_t) = \pi$ for some $t > 0$, then $\theta_t X$ is a solution of the martingale problem with the initial value π 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 (6.1),

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

- (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 $f \in \mathbb{D}(\mathcal{L})$,

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

  the right hand side equals

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

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

This completes the proof. □

6.2.1 Lyapunov Function Technique

In the content of the theorem below, a Lyapunov function is a function with $\mathcal{L}V \leqslant K - cV$.

Note that if V is twice continuous, then by Itô's formula, $V(x_t) - \int_0^t \mathcal{L}V(x_s)ds$ is a local martingale. If $\mathcal{L}V \leqslant K - cV$, $V(x_t) - \int_0^t (K - cV(x_s))ds$ is expected to be a super-martingale.

Definition 6.2.2 Let x_t denote a Markov process with generator $\mathcal{L}$. A measurable functions is said to satisfy that $\mathcal{L}V \leqslant K - cV$, if

$$V(x_t) - \int_0^t (K - cV)(x_s)ds \tag{6.2}$$

is a super-martingale, for every starting point x_0. In this content, we say V is a Lyapunov function.

Depending on the problem we also introduce a weighted supremum norm:

$$\|\varphi\|_V = \sup_x \frac{|\varphi(x)|}{1 + V(x)}.$$

A version of the following theorem can be found [7, Thm. 3.6]

Theorem 6.2.3 *Let T_t be a Markov semi-group on $\mathcal{B}_b(\mathcal{X})$ with generator $\mathcal{L}$. Suppose that the following hold:*

- *There exists a measurable function $V : \mathcal{X} \to \mathbf{R}_+$ such that $\mathcal{L}V \leqslant K - cV$ for some positive constants c and K.*
- *For every $R > 0$, there exists a constant $\alpha > 0$, and a positive number t_0, such that the transition kernels satisfy:*

  $$\|P_{t_0}(x, \cdot) - P_{t_0}(y, \cdot)\|_{TV} \leqslant 2(1 - \alpha)$$

  for all $x, y \in \{V(x) + V(y) \leqslant R\}$.

Then, the Markov process has a unique invariant probability measure π and there exists $C > 0$ and $\varrho \in (0, 1)$ such that for every measurable function φ with $\|\varphi\| < \infty$,

$$\|T_t\varphi - \bar{\varphi}\|_V \leqslant Ce^{-\varrho t}\|\varphi - \bar{\varphi}\|_V$$

where $\bar{\varphi} = \int \varphi d\pi$.

Proof There exists a corresponding theorem for discrete time Markov processes in [7, Thm. 3.6] in Convergence of Markov processes. Applying that theorems to $\mathbb{P} = T_{t_0}$, we see that there exists an unique invariant probability μ such that $(T_{t_0 n})_* \mu = \mu$ for every n, It follows immediately that T_t has at most one invariant measure. Furthermore,

$$\|T_{nt_0}\varphi - \bar{\varphi}\|_V \leqslant Ce^{-\varrho t}\|\varphi - \bar{\varphi}\|_V.$$

Now for any t, written as $t = t_0 m + \alpha$ where $\alpha \in [0, t_0)$. Then assuming $\bar{\varphi} = 0$ for simplicity of notation,

$$\|T_t\varphi\|_V = \|T_{t_0 m}T_{t-t_0 m}\varphi\|_V \leqslant Ce^{-\varrho t}\|T_{t-t_0 m}\varphi\|_V \leqslant Ce^{-\varrho t}\|\varphi\|_V,$$

proving the estimate for the time continuous case. □

Example 6.2.4 Consider $\mathcal{L}f(x) = -bxf'(x) + \frac{1}{2}f''(x)$,

$$dx_t = -bx_t dt + dW_t.$$

It is immediate that the Gaussian measure $\pi = N(0, \frac{1}{2b})$ is an invariant probability measure. The transition probability $P_t(x, \cdot)$ is Gaussian with center e^{-tx} and variance $\int_0^t e^{-2(t-s)} ds$. In fact, the solutions are

$$F_t(x) = e^{-bt}x_0 + \int_0^t e^{-b(t-s)dW_s}.$$

Its probability law is Gaussian with center $e^{-bt}x_0$ and variance $\sigma(t)^2 = \frac{1}{2b}(1 - e^{-2bt})$, which converges to $\frac{1}{2b}$. For any bounded measurable function f:

$$P_t f(x) = \frac{1}{\sqrt{2\pi}\sigma(t)} \int e^{-\frac{|y - e^{-2bt}x_0|^2}{2\sigma(t)}} f(y) dy.$$

This is a strong Feller semi-group, and $P_t f$ converges to

$$\int f d\pi = \sqrt{\frac{b}{\pi}} \int e^{-b|y|^2} dy.$$

The transition densities $P_t(x, \cdot)$ and $P_t(y, \cdot)$ have an overlap, their total variation distance is smaller than 2, so the conditions of the theorem hold with the Lyapunov function $V(x) = |x|^2$. If we take the initial condition

$$\int_{-\infty}^t e^s dW_s,$$

which is independent of $(W_t, t \geqslant 0)$, we see that the solution in invariant.

It is known that the set of invariant probability measures for a Feller Markov process is a closed convex hull and the extreme measures, rather the invariant probability measures induced on the path space with the extreme measure as its marginal distributions, are ergodic invariant measures in the sense that the shift invariant sets have measure 0 or 1.

Chapter 7

Discrete time Markov processes

7.0.1 Lyapunov Function test

One simple way of checking that the tightness condition of the Krylov-Bogoliubov theorem holds is to find a so-called Lyapunov function for the system. A Lyapunov function is allowed to take thev value $+\infty$. We clarify what does it mean to integrate a function that might take the value $+\infty$. Let $\mathcal{X}_0 = \{x : V(x) < \infty\}$. If μ is a measure on $\mathcal{X}$ with $\mu(\mathcal{X}_0) = 1$, we define $\int_{\mathcal{X}} V d\mu = \int_{\mathcal{X}_0} V d\mu$, otherwise we set $\int_{\mathcal{X}} V d\mu = \infty$. In particular the assumption that $TV(x) \leqslant \gamma V(x) + C$ implies that $P(x, \mathcal{X}_0) = 1$ for every x with $V(x) < \infty$.

Lemma 7.0.1 *Let P be a transition function on $\mathcal{X}$ and let $V : \mathcal{X} \to \mathbf{R}_+ \cup \{\infty\}$ be a Borel measurable function. Suppose there exist a positive constant $\gamma \in (0, 1)$ and a constant $C > 0$ such that*

$$TV(x) \leqslant \gamma V(x) + C ,$$

for every x such that $V(x) \neq \infty$. Then

$$T^n V(x) \leqslant \gamma^n V(x) + \frac{C}{1-\gamma}. \tag{7.1}$$

Proof This is a simple consequence of the Chapman-Kolmogorov equations:

$$\begin{aligned} T^n V(x) &= \int_{\mathcal{X}} V(y)\, P^n(x, dy) = \int_{\mathcal{X}} TV(y) P^{n-1}(x, dy) = \int_{\mathcal{X}} \int_{\mathcal{X}} V(y)\, P(z, dy)\, P^{n-1}(x, dz) \\ &\leqslant C + \gamma \int_{\mathcal{X}} V(z)\, P^{n-1}(x, dz) \leqslant \ldots \\ &\leqslant C + C\gamma + \ldots + C\gamma^n + \gamma^n V(x) \leqslant \gamma^n V(x) + \frac{C}{1-\gamma} , \end{aligned}$$

completing the proof. □

Typically, $V(x) = |x|^p$ or $V(x) = \log|x|$, etc... These allow us to control $\mathbb{E}|x_n|^p$ etc. Note the following:

- If V is bounded $\mathbb{E}V(x_n) < \infty$ provides no information on tightness of the law of $\{x_n\}$.
  To avoid this assume $V^{-1}([0,a]) := \{y : V(y) \leqslant a\}$ is compact.
- We can allow $V = +\infty$ where x_n does not visit. But V should not be $+\infty$ everywhere, i.e. $V^{-1}(\mathbf{R}_+) \neq \phi$.

Definition 7.0.2 Let $\mathcal{X}$ be a complete separable metric space and let P be a transition probability on $\mathcal{X}$. A Borel measurable function $V : \mathcal{X} \to \mathbf{R}_+ \cup \{\infty\}$ is called a **Lyapunov function** for P if it satisfies the following conditions:

1. $V^{-1}(\mathbf{R}_+) \neq \phi$.
2. For every $a \in \mathbf{R}_+$, the set $\{y : V(y) \leqslant a\}$ is compact.
3. There exist a positive constant $\gamma < 1$ and a constant C such that
$$TV(x) = \int_{\mathcal{X}} V(y)\, P(x, dy) \leqslant \gamma V(x) + C\ ,$$
for every x such that $V(x) \neq \infty$.

With this definition at hand, it is now easy to prove the following results.

Theorem 7.0.3 (Lyapunov function test) *If a transition probability P is Feller and admits a Lyapunov function, then it has an invariant probability measure.*

Proof Let $x_0 \in \mathcal{X}$ be any point such that $V(x_0) \neq \infty$, we show that the sequence of measures $\{P^n(x_0, \cdot)\}$ is tight. For every $a > 0$, let $K_a = \{y \,|\, V(y) \leqslant a\}$, a compact set. By the lemma above,
$$T^n V(x_0) = \int_{\mathcal{X}} V P^n(x, dy) \leqslant \gamma^n V(x) + \frac{C}{1-\gamma}.$$

Tchebycheff's inequality shows that

$$\begin{aligned} P^n(x_0, (K_a)^c) &= \int_{\{V(y)>a\}} P^n(x_0, dy) \leqslant \int_{\{V(y)>a\}} \frac{V(y)}{a} P^n(x_0, dy) \leqslant \frac{1}{a} T^n V(x_0) \\ &\leqslant \frac{1}{a}(V(x_0) + \frac{C}{1-\gamma}). \end{aligned}$$

We have used Lemma 7.0.1 and the fact that $\gamma < 1$. The results follows from convergence of the right hand side, as $a \to \infty$, with rate uniform in n. (More precisely, for every $\varepsilon > 0$ we can now choose $a \geqslant \frac{1}{\epsilon}\Big(V(x_0) + \frac{C}{1-\gamma}\Big)$, then $P^n(x, K_a) \geqslant 1 - \varepsilon$ for every $n \geqslant 0$.) We can now use Krylov-Bogoliubov theorem to conclude. □

The proof the previous theorem suggests that a Lyapunov function V for T allows us to deduce information on its invariant measures. E.g. if $V(x) = |x|^2$ we expect to deduce that π has second moment and the second moment bound $C/(1-\gamma)$, where C and γ are the constants appearing in (7.1). This is indeed the case, as shown by the following proposition:

Proposition 7.0.4 *Let P be a transition probability on $\mathcal{X}$ and let $V\colon \mathcal{X} \to \mathbf{R}_+$ be a measurable function such that there exist constants $\gamma \in (0,1)$ and $C \geqslant 0$ with*

$$\int_{\mathcal{X}} V(y)\, P(x,dy) \leqslant \gamma V(x) + C\ .$$

Then, every invariant measure π for P satisfies

$$\int_{\mathcal{X}} V(x)\, \pi(dx) \leqslant \frac{C}{1-\gamma}\ .$$

Proof Let $M \geqslant 0$ be an arbitrary constant. As a shorthand, we will use the notation $a \wedge b$ to denote the minimum between two numbers a and b. Let $V_M = V \wedge M$. For every $n \geqslant 0$, one then has the following chain of inequalities:

$$\begin{aligned}\int_{\mathcal{X}} V_M(x)\, \pi(dx) &= \int_{\mathcal{X}} V_M(x)\, \big(T^n\pi\big)(dx) = \int_{\mathcal{X}} T^n V_M(x)\, \pi(dx) \\ &\leqslant \int_{\mathcal{X}} (\gamma^n V_M(x) + \frac{C}{1-\gamma})\, \pi(dx)\end{aligned}$$

We have used Jensen's inequality. Since the function on the right hand side is bounded by M, we can apply the Lebesgue dominated convergence theorem. It yields the bound

$$\int_{\mathcal{X}} \big(V(x) \wedge M\big)\, \pi(dx) \leqslant \frac{C}{1-\gamma}\ ,$$

which holds uniformly in M, and the result follows. □

We complete this section with a couple of inequalities which can be handy for applying Lyapunov function methods.

Lemma 7.0.5 *For any $p \geqslant 1$ and any $\delta > 0$ there exists a constant $K > 1$ such that*

$$|1 + x|^p \leqslant K|x|^p + 1 + \delta.$$

Note that if $x \leqslant 0$, $|x+1|^p \leqslant 1 + |x|^p$.

Proof This is clear if $x < 0$. For p an integer, this can also be obtained by apply Young's inequality to terms $|x|^{p'}|y|^{p-p'}$ in the expansion of $|x+y|^p$.

Now we assume $x \geqslant 0$. Let $f(x) = |1+x|^p$. Let $g(x) = K|x|^p + 1 + \delta$. Note that $g(0) > f(0)$. If $f(x) \geqslant g(x)$ for some x, then by the intermediate value theorem there exists a point where they have equal value. Let x_0 be the first point they are equal. Then $x_0 > 0$. Choose $K = \left(|\frac{1}{|x_0|^p} + 1|^p\right)$. Then $f(x) = |x|^p\left(|\frac{1}{|x|^p} + 1|^p\right) \leqslant K|x|^p$ for any $x \geqslant x_0$. □

Young's inequality: for any $\alpha, \beta > 0$ with $\frac{1}{\alpha} + \frac{1}{\beta} = 1$,

$$ab \leqslant \frac{(\epsilon a)^\alpha}{\alpha} + \frac{b^\beta}{\beta \epsilon^\beta}.$$

7.0.2 Application to a random dynamical system

In this section, let (x_n) be a Markov process defined by a recursion relation of the type

$$x_{n+1} = F(x_n, \xi_n)\;, \tag{7.2}$$

for $\{\xi_n\}$ a sequence of independent and identically distributed random variables taking values in a measurable space $\mathcal{Y}$, and all independent of x_0, and $F\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ a Borel measurable function. Then for any $V \in \mathcal{B}_b(X)$,

$$TV(x) = \mathbb{E}[V(F(x, \xi_n))].$$

An effective criteria for the transition probabilities to be Feller is as follows:

Theorem 7.0.6 *Let (x_n) be a Markov process defined by a recursion relation of the type*

$$x_{n+1} = F(x_n, \xi_n)\;,$$

for $\{\xi_n\}$ a sequence of i.i.d. random variables taking values in a measurable space $\mathcal{Y}$ and $F\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$. If the function $F(\cdot, \xi_n)\colon \mathcal{X} \to \mathcal{X}$ is continuous for almost every realisation of ξ (If A is the set of y such that $x \mapsto F(x,y)$ is continuous, then the property that $\mathbb{P}(\xi_n \in A) = 1$ does not depend on n.), then the corresponding transition semigroup is Feller.

Proof Denote by $\hat{\mathbb{P}}$ the law of ξ_n on $\mathcal{Y}$ and by $\varphi\colon \mathcal{X} \to \mathcal{X}$ an arbitrary continuous bounded function. It follows from the definition of the transition semigroup T that

$$\big(T\varphi\big)(x) = \mathbb{E}\big(\varphi(x_{n+1}) \,|\, x_n = x\big) = \mathbb{E}\varphi(F(x, \xi_n)) = \int_{\mathcal{Y}} \varphi\big(F(x,y)\big)\, \hat{\mathbb{P}}(dy)\;.$$

Let now $\{z_n\}$ be a sequence of elements in $\mathcal{X}$ converging to z. Lebesgue's dominated convergence theorem shows that

$$\begin{aligned}\lim_{n\to\infty}\big(T\varphi\big)(z_n) &= \lim_{n\to\infty}\int_{\mathcal{Y}}\varphi\big(F(z_n,y)\big)\,\hat{\mathbb{P}}(dy) = \int_{\mathcal{Y}}\lim_{n\to\infty}\varphi\big(F(z_n,y)\big)\,\hat{\mathbb{P}}(dy)\\ &= \int_{\mathcal{Y}}\varphi\big(F(z,y)\big)\,\hat{\mathbb{P}}(dy) = \big(T\varphi\big)(z)\;,\end{aligned}$$

which implies that $T\varphi$ is continuous and therefore that T is Feller. □

If F is continuous in the first variable for each y, then the Markov process is Feller.

Theorem 7.0.7 *Suppose that the function $F(\cdot,\xi_n)\colon \mathcal{X}\to\mathcal{X}$ is continuous for almost every realisation of ξ_n. If, furthermore, there exists a Borel measurable function $V:\mathcal{X}\to$* **R** *with compact sub-level sets and constants $\gamma\in(0,1)$ and $C\geqslant 0$ such that*

$$\int_{\mathcal{Y}} V\big(F(x,y)\big)\,\hat{\mathbb{P}}(dy)\leqslant \gamma V(x)+C\;,\quad \forall x\in\mathcal{X}\;,$$

where $\hat{\mathbb{P}}$ is the distribution of ξ_n, then the process x has at least one invariant probability measure.

Proof Indeed,

$$P(x,A)=\mathbb{E}(x_1\in A|x_0=x)=\mathbb{E}(F(x_0,\xi_0)\in A|x_0=x)=\int \mathbf{1}_A(F(x,y))\hat{P}(dy).$$

Then P is Feller follows from Theorem 7.0.6. Then the left hand side of the given inequality is TV and V is a Lyapunov function. The existence of an invariant probability measure now follows from the Lyapunov function test. □

Chapter 8

Ergodic Theorem

8.1 Ergodic Theorems

In this small section we introduce/recall some core notions of dynamical systems, these will connect to stationary Markov process viewed on the canonical path space $\mathbf{C}(\mathbf{R}_+, \mathcal{X})$. A Markov chain will be viewed to be on $\mathcal{X}^{\mathbf{N}}$ or two-sided path space $\mathcal{X}^{\mathbf{Z}}$.

Definition 8.1.1 A **dynamical system** consists of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measure preserving measurable map $\theta\colon \Omega \to \Omega$, *i.e.* a map such that $\mathbb{P}(\theta^{-1}(A)) = \mathbb{P}(A)$ for every $A \in \mathcal{F}$ (i.e. $\theta_* \mathbb{P} = \mathbb{P}$).

Definition 8.1.2 Given a measurable transformation θ on $(\Omega, \mathcal{F}, \mathbb{P})$, a set with $\theta^{-1}(A) = A$ is called an invariant set for θ (or θ-invariant). Then the invariant σ-algebra $\mathcal{I} \subset \mathcal{F}$ is defined as

$$\mathcal{I} = \{A \in \mathcal{F} \ : \ \theta^{-1}(A) = A\}.$$

It is clear that $\mathcal{I}$ is again a σ-algebra. In order to emphasise the invariance with respect to θ, we may refer an invariant set as a θ-invariant set.

Definition 8.1.3 A measurable function $f : \Omega \to \mathbf{R}$ is said to be θ-invariant (or simply invariant) if $f \circ \theta = f$.

Exercise 8.1.4 Let $f : \Omega \to \mathbf{R}$ be an $\mathcal{F}$- measurable function. Then f is invariant if and only if f is measurable with respect to the invariant σ-algebra $\mathcal{I}$.

Definition 8.1.5 Given a dynamical system $(\Omega, \mathcal{F}, \mathbb{P})$ and θ. We say θ is ergodic if any θ-invariant set has either measure 0 or measure 1. Note that this is a property of the map θ as well as of the measure $\mathbb{P}$. We also say $\mathbb{P}$ is ergodic (w.r.t. θ).

Proposition 8.1.6 *The following statements are equivalent.*

1. *$\mathbb{P}$ is ergodic (θ is ergodic);*
2. *Every invariant integrable function f is almost surely a constant.*
3. *Every invariant bounded function is almost surely a constant.*

Proof From (2) to (3) is trivial. It remains to show (3) ⇒ (1), and (1) ⇒ (2).

(3) ⇒ (1). Assume that (3) holds. Let $f = \mathbf{1}_A$ where A is an invariant set. Then $\mathbf{1}_A$ in invariant and $\mathbf{1}_A = 1$ or 0 a.e., hence $\mathbf{1}_A = \mathbb{P}(A) \in \{0, 1\}$ and $\mathbb{P}$ is ergodic.

(1) ⇒ (2). Suppose that $\mathbb{P}$ is ergodic, i.e. $\mathbb{P}(A) = 1$ or 0 for any $A \in \mathcal{I}$. Let function f be integrable and invariant, then f is measurable with respect to $\mathcal{I}$.[1] We prove that $f = \mathbb{E}f$ a.e. . Note that the following sets

$$A_+ = \{\omega \in \Omega \,|\, f(\omega) > \mathbb{E}f\}, \qquad A_- = \{\omega \in \Omega \,|\, f(\omega) < \mathbb{E}f\}, \qquad A_0 = \{\omega \in \Omega \,|\, f(\omega) = \mathbb{E}f\},$$

are invariant sets and form a partition of Ω. Therefore, by ergodicity, exactly one of them has measure 1 and the other two must have measure 0. Suppose $\mathbb{P}(A_+) = 1$, then

$$0 = \int_\Omega (f - \mathbb{E}f)\, d\mathbb{P} = \int_{A_+} (f - \mathbb{E}f)\, d\mathbb{P}.$$

Then $f - \mathbb{E}f = 0$ a.s. on A_+, which is a contradiction. Similarly if $\mathbb{P}(A_-) = 1$, we also have $f = \mathbb{E}f$ a.e., hence we must have $\mathbb{P}(A_0) = 1$. □

Theorem 8.1.7 (Birkhoff's Ergodic Theorem) *Let $(\Omega, \mathcal{F}, \mathbb{P}, \theta, \mathcal{I})$ be as above and let $f\colon \Omega \to \mathbf{R}$ be such that $\mathbb{E}|f| < \infty$. Then,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\theta^n \omega) = \mathbb{E}\big(f \,|\, \mathcal{I}\big)$$

almost surely.

Let θ be the shift operator on $\mathcal{X}^{\mathbf{Z}}$, i.e. $\theta(x_\cdot)(n) = x(n+1)$, so that

$$\big(\theta_n x\big)(m) = x(n+m)\ ,$$

and we write $\theta = \theta_1$ and $\theta^{-1} = \theta_{-1}$. As in previous section, we denote by $\mathcal{I}$ the set of all measurable subsets of $\mathcal{X}^{\mathbf{Z}}$ that are invariant under θ,

$$\mathcal{I} = \{C \in \mathcal{B}(\mathcal{X}^z) : \theta^{-1} C = C\}.$$

[1]See Exercise 3 of Problem Sheet 8.

Also let $P = (P(x,\cdot), x \in \mathcal{X})$ be a family of transition probabilities and a probability measure $\pi \in \mathbb{P}(\mathcal{X})$ satisfying $\pi = \int_{\mathcal{X}} P(x,\cdot)\pi(dx)$.

By the definition of stationarity, one has:

Lemma 8.1.8 *The triple $(\mathcal{X}^{\mathbf{Z}}, \mathcal{B}(\mathcal{X}^{Z}), \mathbb{P}_\pi, \theta)$ defines a dynamical system, and θ is continuous (This is called a continuous dynamical system).*

Proof We have already seen that θ is $\mathbb{P}_\pi$-invariant . It is clear that θ is continuous (with respect to the product topology). The product topology is the coarsest topology such that each projection map $\pi_i : \Pi\mathcal{X} \to \mathcal{X}$ is continuous. We only need to test with open sets of the form $\pi_i^{-1}(U)$. It is clear that $\theta^{-1}(\pi_i^{-1}(U))$ is an open set. □

Remember that the measure $\mathbb{P}_\pi$ is **ergodic** if every $A \in \mathcal{I}$ has $\mathbb{P}_\pi(A) \in \{0,1\}$.

Definition 8.1.9 We say that an invariant measure π of a Markov process with associated transition semigroup T is **ergodic** if the corresponding measure $\mathbb{P}_\pi$ is ergodic for θ.

Theorem 8.1.10 *Let $P = P(x,\cdot)$ be a transition probability with an invariant probability measure π. Let $(x_n)_{n\in Z}$ be a time homogeneous Markov process with t.p. P and initial position $x_0 = x$. Then for π-almost every $x \in \mathcal{X}$, the following statements hold:*
1. For any integrable function $f : \mathcal{X}^{\mathbf{Z}} \to \mathbf{R}$,

$$\frac{1}{n}\sum_{k=1}^{n} f(\theta^k x_\cdot(\omega)) \quad \textit{converges for } \mathbb{P}\textit{-a.e. } \omega.$$

2. If furthermore π is ergodic,

$$\frac{1}{n}\sum_{k=1}^{n} f(\theta^k x_\cdot(\omega)) \quad \xrightarrow{n\to\infty} \quad \int_{\mathcal{X}} f \, d\mathbb{P}_\pi \qquad \mathbb{P}\textit{-a.e. } \omega.$$

Proof There are many proofs for this, here we illustrate the use of stopping times. First let $x_0 \sim \pi$ (then the Markov chain with initial condition x_0 is stationary). By Theorem **??**, we have

$$\frac{1}{n}\sum_{k=1}^{n} f(\theta^k x_\cdot) \quad \longrightarrow \quad \bar{f}(x_\cdot), \quad \mathbb{P}\text{-a.e. } \omega.$$

Then by the dominated convergence theorem

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} f(\theta^k x_\cdot) \mid \sigma(x_0)\right] \quad \xrightarrow{n\to\infty} \quad \mathbb{E}\big[\bar{f}(x_\cdot)|\sigma(x_0)\big], \quad \mathbb{P}\text{-a.e. } \omega.$$

From this we deduce that for π-almost every x,

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} f(\theta^k x_\cdot) \mid x_0 = x\right] \xrightarrow{n\to\infty} \mathbb{E}\big[\bar{f}(x_\cdot)|x_0 = x\big], \quad \mathbb{P}\text{-a.e. } \omega.$$

This can be seen by testing the conditional expectation in the previous line with functions of the form $\varphi(x_0)$ and turn the expectation into integration with respect to x_0.

□

8.2 Structure Theorem

Let T be the transition operator for a Markov chain, $I_P = \{\pi \in \mathcal{P}(\mathcal{X}) \,:\, T\pi = \pi\}$ denote the set of invariant probability measures. It is a convex set: If π_1 and π_2 are in I_P, then any of their convex combination is in I_P also.

If T is Feller, then it is a continuous map from $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$ in the topology of weak convergence. Therefore, if π_n is a sequence of invariant probability measures converging weakly to a limit π, one has

$$T\pi = T\lim_{n\to\infty}\pi_n = \lim_{n\to\infty} T\pi_n = \lim_{n\to\infty}\pi_n = \pi \,,$$

so that π is again an invariant probability measure for P. This shows that if T is Feller, then the set I_P is closed (in the topology of weak convergence).

Definition 8.2.1 A probability measure $\pi \in I_P$ is an extremal, of I_P, if π cannot be decomposed as $\pi = t\pi_1 + (1-t)\pi_2$ with $t \in (0,1)$ and $\pi_i \in I_P$ are distinct.

Theorem 8.2.2 *Given a time homogeneous transition probability P, with corresponding transition operator T. With I_P denoting the set of probability measures invariant w.r.t. P, set*

$$\mathcal{E} = \{\pi \in \mathcal{P}(\mathcal{X}) \,:\, T\pi = \pi,\ \pi \text{ is ergodic }\} \subset I_P.$$

Then the following statements hold.

(a) The set I_P is convex and $\mathcal{E}$ is precisely the set of its extremal points.

(b) Any two ergodic invariant probability measures are either identical or mutually singular.

(c) Furthermore, every invariant probability measure $\pi \in I_P$ is a convex combination of ergodic invariant probability measures, i.e. for every invariant measure $\mu \in \mathcal{I}$, there exists a probability measure $\mathcal{Q}_\mu$ on $\mathcal{E}$ such that

$$\mu(A) = \int_{\mathcal{E}} \nu(A)\, \mathcal{Q}_\mu(d\nu) \,.$$

Remark 8.2.3 As a consequence, if a Markov process admits more than one invariant measure, it does admit at least two ergodic (and therefore mutually singular) ones. This leads to the intuition that, in order to guarantee the uniqueness of its invariant measure, it suffices to show that a Markov process explores its state space 'sufficiently thoroughly'.

8.3 Quantitative Ergodic Theorem

Let (y_t) be a Markov process on a metric space $\mathcal{X}$ with transition semi-group $T_t : E \to E$ where $E \subset \mathcal{B}_b(\mathcal{X})$ is a Banach space. Note that, given geometric convergence to the equilibrium of the Markov process,

$$\|T_t f - \bar{f}\|_E \leqslant M e^{-\lambda t} \|f\|_E,$$

one can deduce that the convergence of the expectation of the time average to the spatial average is of the order $1/T$: Denote $\mathbb{E}_y$ taking expectation with respect to the Markov process with initial value y.

$$\begin{aligned}\left\|\mathbb{E}_y\left(\frac{1}{T}\int_0^T f(y_s)ds\right) - \bar{f}\right\|_E =& \left\|\frac{1}{T}\int_0^T (T_s f(y) - \bar{f})ds\right\|_E \\ \leqslant& \frac{M}{T}\|f\|_E \int_0^\infty M e^{-\lambda s} ds \lesssim \frac{1}{T}\frac{1}{\lambda}\|f\|_E.\end{aligned}$$

However we would like to work out the difference between the time average, not the expectation of the time average as given above, from $\bar{f}$.

Lemma 8.3.1 *[Law of large numbers] Let T_t be a Markov semi-group on $\mathcal{X}$. Suppose that the following exponential ergodicity holds:*

$$\|T_t f - \bar{f}\|_\infty \leqslant M e^{-\lambda t} \|f\|_\infty.$$

Let (y_t) be a Markov process corresponding to T_t. Then, for any bounded function f,

$$\left|\frac{1}{T}\int_0^T f(y(s))ds - \bar{f}\right|_{L^2(\Omega)} \leqslant C\frac{1}{\sqrt{T}},$$

where $C = \sqrt{\frac{M}{\lambda}}\|f\|_\infty$.

Proof Without loss of generality, let us assume that $\bar{f} = 0$. Then

$$\mathbb{E}\Big(\frac{1}{T}\int_0^T f(y(s))ds\Big)^2 = \mathbb{E}\Big(\frac{1}{T^2}\int_0^T\int_0^T f(y(s))f(y(r))drds\Big)$$

$$= \frac{1}{T^2}\int_0^T\int_0^s\int_{\mathcal{X}} T_r(fT_{s-r}f)(y)\mu(dy)drds + \frac{1}{T^2}\int_0^T\int_s^r\int_{\mathcal{X}} T_s(fT_{r-s}f(y)\mu(dy))drds$$
$$\leqslant \frac{M}{T^2}\int_0^T\int_0^T e^{-|s-r|\lambda}\|f\|_\infty^2\, ds\, dr \leqslant \frac{M}{T\lambda}(\|f\|_\infty)^2,$$

from which we see the $\frac{1}{\sqrt{T}}$ rate of convergence. □

What we have in mind in the last theorem is an elliptic operator on a compact manifold.

8.4 Functional Central Limit theorem

Given a function f centred with respect to the invariant measure for a Markov process y with generator $\mathcal{L}$, we consider $\sqrt{\epsilon}\int_0^{t/\epsilon} f(y_r)dr$. For simplicity we shall assume that y_t solves the SDE:

$$dy_t = \sum_i Y_i(y_t)dW_t^i + Y_0(y_t)dt. \tag{8.1}$$

If g is a C^2 function solving the Poisson equation and then

$$M_t^g := g(y_t) - g(y_0) - \int_0^t \mathcal{L}g(y_s)ds = \sum_k \int_0^t Dg(Y_k)(y_s)\, dW_s^k$$

is a martingale. If $\mathcal{L}g = f$, we expect that $\sqrt{\epsilon}\int_0^{\frac{t}{\epsilon}} f(y_s)ds = \sqrt{\epsilon}g(y_{\frac{t}{\epsilon}}) - \sqrt{\epsilon}g(y_0) - \sqrt{\epsilon}M^g_{\frac{t}{\epsilon}}$ converges to a Wiener process.

Lemma 8.4.1 *Let T_t be a strongly continuous semi-group on a Banach space E with generator $\mathcal{L}$. Suppose that g is a solution of $\mathcal{L}g = f$ and suppose that $\lim_{t\to\infty} T_t g$ exists which we denote by $\bar{g}$. Then*

$$g = \bar{g} - \int_0^\infty T_t f dt.$$

Proof By the semi-group theory,

$$T_t g - g = \int_0^t \frac{\partial}{\partial s} T_s g ds = \int_0^t T_s \mathcal{L} g ds$$
$$= \int_0^t T_s f ds.$$

Then $\int_0^t T_s f ds$ has a limit which we denote by $\int_0^\infty T_s f ds$. We take $t \to \infty$ to conclude. □

Note that under the ergodic assumption, $T_t f$ converges. Suppose that, on the other hand, $\int_0^t T_s f ds$ converges as $t \to \infty$ in $L^2(dx)$, i.e. the following limit exists:

$$\lim_{t\to\infty} \int_0^t \int_0^t \langle T_s f(x), T_r f(x)\rangle_{L^2} ds dr.$$

Lemma 8.4.2 *Let T_t be a strongly continuous semi-group on a Banach space $E \subset \mathcal{B}_b(\mathcal{X})$ with generator $\mathcal{L}$. Let $f \in E$ be such that $\int_0^t T_s f ds$ and $T_t f$ converges as $t \to \infty$. Then $\int_0^\infty T_s f ds \in \mathbb{D}(\mathcal{L})$ and it solves the Poisson equation $\mathcal{L}(\int_0^\infty T_s f ds) = -f$.*

Proof Note that $\int_0^t T_s f ds \in \mathrm{D}om(\mathcal{L})$, and $\int_0^t T_s f ds \to \int_0^\infty T_s f ds$ by the assumption. Since $\mathcal{L} \int_0^t T_s f ds = (T_t f - f)$, the right hand side converges by assumption. Consequently, $\int_0^\infty T_s f ds$ belongs to the domain of the closed operator $\mathcal{L}$ and

$$\mathcal{L}(\int_0^\infty T_s f ds) = \lim_{t\to\infty} T_t f - f.$$

The convergence is in the supremum norm. Then $\lim_{t\to\infty} \int_0^t T_t f(x) ds$ and $\lim_{t\to\infty} T_t f(x)$ exist for every x. Fixing x, suppose that $a = \lim_{t\to\infty} T_t f(x) \neq 0$. Without loss of generality, we assume$a > 0$. Let T be sufficiently large, then $\int_T^t T_s f(x) ds > \frac{a}{2}(t - T)$ which has no finite limit, as $t \to \infty$. We conclude that $\lim_{t\to\infty} T_t f = 0$. □

Recall that if y_t is a right continuous Markov process and an invariant probability measure π, then its Markov semi-group extends to a positive preserving strongly continuous contraction semi-group on $L^p(\pi)$ where $p \geqslant 1$.

Definition 8.4.3 Let μ be a Borel measure on $\mathcal{X}$ and denote by $L^2(\mathcal{X})$ the space of L^2 functions from $\mathcal{X} \to \mathbf{R}$. A Markov semi-group on $\mathcal{X}$ is said to be reversible with respect to μ if for all $f, g \in L^2(\mu)$,

$$\int_{\mathcal{X}} g T_t f d\mu = \int_{\mathcal{X}} f T_t g d\mu.$$

The measure μ is said to be a reversible measure for T_t. In this case T_t is said to be symmetric on $L^2(\mu)$.

Exercise 8.4.4 If μ is a reversible probability measure for T_t, show that it is an invariant measure. (Recall that $T_t 1 = 1$.)

Theorem 8.4.5 (Spectral Theorem) *Let T be a self-adjoint operator on some separable Hilbert space H. Then, there exists a measure space (E, μ), a unitary operator $K : H \to L^2(E, \mu)$, and a function $\Lambda : E \to \mathbb{R}$ such that*

$$\mathrm{D}om(T) = \{f \in H : \Lambda K f \in L^2(E, \mu)\},$$
$$(KTf)(\lambda) = \Lambda(\lambda)(Kf)(\lambda),$$

where $\lambda \in E$.

Corollary 8.4.6 *Let $T_t : H \to H$ be a strongly continuous symmetric contraction operator with generator $\mathcal{L}$, and $g \in \mathrm{Dom}(\mathcal{L})$ then $\lim_{t\to\infty} T_t g$ exists.*

Proof Since T_t is symmetric, $\mathcal{L}$ is self-adjoint. Since T_t is a contraction semi-group, $\langle \mathcal{L}ff\rangle \leqslant 0$. By the spectral theorem, there exists a measure space (E, μ), whose σ-algebra plays no -significant role here and so omitted, a function $\varphi : E \to \mathbf{R}$, and a unitary operator $K : H \to L^2(E, \mu)$ (an isometry, preserving the norm, and $K^* = K^{-1}$) such that

$$\mathcal{L}f = K^*(\varphi K f)$$

Since $\mathcal{L}$ is negative, $\langle \varphi(Kf), Kf\rangle_{L^2} \leqslant 0$. Since K is unitary, $\langle \varphi(f), f\rangle_{L^2} \leqslant 0$ for all $f \in L^2$, consequently, $\varphi \leqslant 0$. Now

$$T_t g = e^{tK^*\varphi K} g$$

by functional calculus. Since $\varphi \leqslant 0$, $e^{tK^*\varphi K} \leqslant 1$ and $\lim_{t\to\infty} e^{tK^*\varphi K} = \mathbf{1}_{\varphi=0}$. By the dominated convergence theorem, as $t \to \infty$

$$\|T_t g - \mathbf{1}_{\varphi=0}\, g\|_{L^2} \to 0,$$

proving the claim. □

Note that $1_{\varphi=0}\, g$ is in the kernel of $\mathcal{L}$ in the sense that $\mathcal{L}(1_{\varphi=0}\, g) = K^*\varphi 1_{\varphi=0} K g) = 0$. So for any $f \in E$,

$$\frac{1}{t}\int_0^t T_s f ds \to H(f)$$

where $H(f)$ denotes the projection of f to the null space of $\mathcal{L}$.

Let y_t be a stationary Markov process corresponding to a symmetric semi-group T_t.

Lemma 8.4.7 *Let T_t is a reversible strongly continuous contraction Markov semigroup on $L^2(\pi)$, where π is a probability measure on a smooth manifold, with generator $\mathcal{L}$. Let $f : M \to \mathbf{R}$ be a function with $\bar{f} = 0$, $f \in L^2(\pi)$. Suppose that $\mathcal{L}g = f$ has a bounded C^2 solution g. Suppose that*

$$\Sigma^2 := \int_0^\infty \langle T_s f, f\rangle_{L^2(\pi)} ds < \infty,$$

Then,

$$\mathbb{E}(\sqrt{\epsilon}\int_0^{t/\epsilon}\int_0^{t/\epsilon} \mathbb{E}[f(y_s)f(y_r])dsdr)^2$$

increases to $2t\int_0^\infty \langle f, T_s f\rangle_{L^2(\pi)} ds$.

Proof For $s < r$,

$$\mathbb{E}[f(y_s)f(y_r)] = \int_{\mathcal{X}} T_s(fT_{r-s}f)d\pi = \int_{\mathcal{X}} fT_{r-s}fd\pi = \langle f, T_{r-s}f\rangle_{L^2(\pi)}.$$

We first compute

$$\begin{aligned}
&\mathbb{E}(\sqrt{\epsilon}\int_0^{t/\epsilon}\int_0^{t/\epsilon}\mathbb{E}[f(y_s)f(y_r])dsdr)^2\\
&=\epsilon\int_0^{t/\epsilon}\int_0^{r}\mathbb{E}[f(y_s)f(y(r)]\;ds\;dr+\epsilon\int_0^{t/\epsilon}\int_r^{\frac{t}{\epsilon}}\mathbb{E}[f(y_s)f(y(r)]\;ds\;dr.\\
&=\epsilon\int_0^{t/\epsilon}\int_0^{r/\epsilon}\int_{\mathcal{X}}fT_{r-s}fd\pi\;ds\;dr+\epsilon\int_0^{t/\epsilon}\int_r^{\frac{t}{\epsilon}}\int_{\mathcal{X}}fT_{s-r}fd\pi\;ds\;dr\\
&=2\epsilon\int_{\mathcal{X}}f\int_0^{t/\epsilon}\int_0^{u}T_vf\;dvdud\pi=2\int_0^t\int_0^{u'/\epsilon}\langle f,T_vf\rangle dvdu'.
\end{aligned}$$

Suppose that $\int_0^\infty\langle f,T_sf\rangle_{L^2(\pi)}$ exists, since $\int_0^t\langle f,T_sf\rangle_{L^2(\pi)}\geqslant 0$,

$$\mathbb{E}(\sqrt{\epsilon}\int_0^{t/\epsilon}\int_0^{t/\epsilon}\mathbb{E}[f(y_s)f(y_r])dsdr)^2\to 2t\int_0^\infty\langle f,T_sf\rangle_{L^2(\pi)}ds,$$

the left hand side is monotone increasing in ϵ. □

The following can be found in [**?**], see [**?**, Theorem VIII.2.17].

Lemma 8.4.8 *Let $X^\varepsilon=\left(X_1^\varepsilon,X_2^\varepsilon,...\right)$, $\varepsilon>0$ be a family of continuous local martingales starting at 0. Let $B_1,B_2,...$ be independent standard Brownian motions, $\alpha_{ij}\in\mathbf{R}$, $i,j\in\mathbf{N}$ such that $\sum_j\alpha_{ij}^2<\infty$ for all $i\in\mathbf{N}$, $V_i:=\sum_{j=1}^\infty\alpha_{ij}B_j$, $i\in\mathbf{N}$, and $V=(V_1,V_2,...)$. If the quadratic variation $\langle X_k^\varepsilon,X_l^\varepsilon\rangle_t$ converges in law to $\langle V_k,V_l\rangle_t=t\sum_{j=1}^\infty\alpha_{kj}\alpha_{lj}$ for all $k,l\in\mathbf{N}$, $t\geqslant 0$, then X^ε converges to V weakly as $\varepsilon\to 0$, i.e. for each $n\in\mathbf{N}$, $(X_1^\varepsilon,...,X_n^\varepsilon)$ converges in law to $(V_1,...,V_n)$ with respect to the uniform topology on compact intervals.*

The following theorem has numerous extensions.

Theorem 8.4.9 (Functional Central Limit Theorem) *Suppose that T_t is a reversible strongly continuous contraction Markov semigroup on $L^2(\pi)$, where π is a probability measure on a smooth manifold, with generator $\mathcal{L}$. Let $f:M\to\mathbf{R}$ be a function with $\bar{f}=0$, $f\in L^2(\pi)$. Suppose that $\mathcal{L}g=f$ has a bounded C^2 solution g. If*

$$\Sigma^2:=\int_0^\infty\langle T_sf,f\rangle_{L^2(\pi)}ds<\infty,$$

and y_t is the stationary Markov process with transition semi-group T_t, solving the SDE (8.1), then

$$\frac{1}{\sqrt{\epsilon}}\int_0^t f(y_{\frac{s}{\epsilon}})ds=\sqrt{\epsilon}\int_0^{t/\epsilon}f(y_r)dr\to 2\Sigma W_t.$$

Proof Since y_t solves the martingale problem for $\mathcal{L}$,

$$M_t^g := g(y_t) - g(y_0) - \int_0^t \mathcal{L}g(y_s)ds$$

is a local martingale. In fact,

$$\mathbb{E}(M_t^g)^2 \leqslant 4|T_t g|_{L_2}^2 + 4|g|^2 + \int_0^t \int_0^t \mathbb{E}[\mathcal{L}g(y_s)\mathcal{L}g(y_r)]dsdr < \infty.$$

We have used the contraction property: $|T_t g|_{L_2}^2 \leqslant |g|_2^2$ and by the previous computation, the last term is finite. Thus (M_t^g) is in fact an L^2- martingale and $(M_t)^2 - \langle M \rangle_t$ is a martingale. To show that $\sqrt{\epsilon}\langle M \rangle_{t/\epsilon}$ converges, it is sufficient to show their quadratic variations converge. It is easy to see that these converge in expectation.

$$\mathbb{E}(\sqrt{\epsilon}\langle M \rangle_{t/\epsilon} - 2t\Sigma) = \mathbb{E}(g(y_t) - g(y_0) - \sqrt{\epsilon}\int_0^{\frac{t}{\epsilon}} f(y_s)ds)^2 - 2t\Sigma \to 0.$$

Note that $g \in L^2$, and y_t is stationary, so $\sqrt{\epsilon}g(y_{\frac{t}{\epsilon}}) - \sqrt{\epsilon}g(y_0) \to 0$ in L^2. This implies that the martingale process converge weakly to a Wiener process. □

Note that under the assumption of the theorem,

$$\int_0^t \int_0^t \langle T_s f, T_r f \rangle_{L^2(\pi)} dsdr \lesssim t \int_0^\infty |T_s f|_{L^2(\pi)} ds = t\Sigma^2 < \infty.$$

Recall that under the assumptions of Lemma 8.4.2, $g = -\int_0^\infty T_s f ds$, which is sufficiently smooth if $\mathcal{L}$ is nice, e.g. elliptic and smooth, and f has nice properties.

8.4.1 Appendix: Locally Uniform Law of Large Numbers

The motivation for this section comes from with feedback models. The fast variable maybe affected by the slow variable.

$$dY_t^\epsilon = \frac{1}{\sqrt{\epsilon}} \sum_i V_i(x_t^\epsilon, Y_t^\epsilon) d\tilde{W}_t^i + \frac{1}{\epsilon} V_0(x_t^\epsilon, Y_t^\epsilon) dt.$$

We first sort out what is the averaging of long time for a function. Again let us first consider:

$$dY_t^\epsilon = \sum_i V_i(x_{t\epsilon}^\epsilon, Y_t^\epsilon) d\tilde{W}_t^i + \frac{1}{\epsilon} V_0(x_{t\epsilon}^\epsilon, Y_t^\epsilon) dt.$$

We freeze the slow variable in time, as they move slowly, and consider the stochastic equation:

$$dY_t^x = \sum_i V_i(x, Y_t) d\tilde{W}_t^i + \frac{1}{\epsilon} V_0(x, Y_t) dt.$$

We further postulate that for each x, Y_t^x has an invariant probability measure μ^x. From earlier discussions it is reasonable to assume that

$$\left| \frac{1}{t} \int_0^t f(x, y_s) ds - \int f(x, y) \mu_x(dy) \right| \to 0.$$

Example 8.4.10 Consider on S^1 the following sde, where the parameters x taking values in **R**,

$$dy_t = \sin(y_t + x) dB_t + \cos(y_t + x) dt.$$

Its generator is $\mathcal{L}_x = \cos(x+y)\frac{\partial}{\partial y} + \frac{1}{2}\sin^2(x+y)\frac{\partial^2}{\partial y^2}$

Let us begin with a Markov process with an invariant probability measure π and generator $\mathcal{L}$. Suppose that we want to solve the Poisson equation $\mathcal{L}g = f$. Since $\int \mathcal{L}g d\pi = 0$, it is necessary that

$$\int f d\pi = 0.$$

This is called the center condition.

Suppose that $\mathcal{L}$ is a Markov generator for a continuous Markov process on a d-dimensional compact manifold. Assume that $\mathcal{L}$ has a discrete spectrum with eigenvalues

$$0 = \lambda_0 < \lambda_1 < \dots,$$

and the corresponding eigen-functions e_n forms an orthogonal (normal) basis of $L^2(\pi)$

For simplicity we assume that

$$\mathcal{L}f = \frac{1}{2}\sum_k L_{Y_k} L_{Y_k} f + L_{Y_0} f.$$

where Y_i are smooth vector fields. Suppose that $\{Y_1(x), \ldots, Y_m(x)\}$ has dimension d at every point – in this case $\mathcal{L}$ is said to be elliptic. It is also a diffusion operator.

Note that $\mathcal{L}$ applies to any constant returns zero. So $e_0 = 1$. Then any L^2 function f can be written as:

$$f = \sum_{n=0}^{\infty} \langle f, e_n \rangle_{L^2} e_n.$$

The condition $\int f d\pi = 0$ implies that $\langle f, e_0 \rangle_{L^2} = 0$. Then we can solve the Poisson equation $\mathcal{L}g = f$ explicitly. Set $g = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle f, e_n \rangle_{L^2} e_n$. It is clear that we can bound g with bounds on f.

Definition 8.4.11 Let $\mathcal{L}_x$ be a family of Markov operators with a unique invariant probability measure μ_x. We say that $\mathcal{L}_x$ satisfy a locally uniform law of large numbers if the following holds.

(a) $x \mapsto \mu_x$ is locally Lipschitz continuous in the total variation norm.

(b) There exists a positive constant $C(x)$, locally bounded in x, such that for every smooth function $f : G \to \mathbf{R}$ of compact support, there exists a constant $c(f)$ such that

$$\left| \frac{1}{T} \int_t^{t+T} f(z_r^x)\, dr - \int_G f(y) \mu_x(dy) \right|_{L_2(\Omega)} \leqslant C(x) c(f) \frac{1}{\sqrt{T}}, \tag{8.2}$$

where z_r denotes an $\mathcal{L}_x$-diffusion.

Theorem 8.4.12 (Locally Uniform Law of Large Numbers) *Let G be a compact manifold. Suppose that Y_i are bounded, C^∞ with bounded derivatives. Suppose that each*

$$\mathcal{L}_x = \frac{1}{2}\sum_{i=1}^m Y_i^2(x, \cdot) + Y_0(x, \cdot)$$

satisfies Hörmander's condition, and has a unique invariant probability measure μ_x. Then μ_x has a locally uniform law of large numbers.

Furthermore, there exists a positive constant $C(x)$, depending continuously in x, such that for every smooth function $f : G \to \mathbf{R}$,

$$\left| \frac{1}{T} \int_t^{t+T} f(x, z_r^x)\, dr - \int_G f(y) \mu_x(dy) \right|_{L_2(\Omega)} \leqslant C(x) C(f) \frac{1}{\sqrt{T}}. \tag{8.3}$$

Proof We only prove the elliptic case. It is sufficient to work with a fixed $x \in N$. We may assume that $\int_G f(x,y)\mu_x(dy) = 0$. For any smooth function f with $\int_G f(x,y)\,\mu_x(dy) = 0$, $\mathcal{L}_x g(x,\cdot) = f(x,\cdot)$ has a smooth solution. If f is compactly supported in the first variable, so is g. We may then apply Itô's formula to the smooth function $g(x,\cdot)$, allowing us to estimate $\frac{1}{T}\int_0^T f(y_r^x)dr$ whose $L^2(\Omega)$ norm is controlled by the norm of g in C^1 and the norms $|Y_j(x,\cdot)|_\infty$. The $\mathcal{L}_x$ diffusion satisfies the equation:

$$\frac{1}{T}\int_0^T f(x,z_r^x)dr = \frac{1}{T}(g(x,z_T^x) - g(x,y_0)) - \frac{1}{T}\left(\sum_{k=1}^{m_2}\int_0^T dg(x,\cdot)(Y_k(x,z_r^x))dW_r^k\right).$$

Since $|Y_j(x,\cdot)|_\infty$ is bounded, it is sufficient to estimate the stochastic integral term by Burkholder-Davis-Gundy inequality:

$$\mathbb{E}\left(\sum_{k=1}^{m_2}\int_0^T dg(x,\cdot)(Y_k(x,z_r^x))dW_r^k\right)^2 \leqslant \sum_{k=1}^{m_1}|Y_k|_\infty^2\int_0^T \mathbb{E}|dg(x,z_r^x)|^2\,ds.$$

It remains to control the supremum norm of $dg(x,\cdot)$. This follows from elliptic regularity theory. □

8.5 A basic averaging theorem– not covered in class

8.5.1 Invariant Measure

We present here a formal derivation of the formula for the invariant measure. Consider an SDE of the form:

$$dX_t^i = \sum_{i,j}\sigma^{ij}(X_t)dB^j + \sigma_0^i(X_t)dt \tag{8.4}$$

Then, for any function $g(X_t)$ we define $P_t g(X_0) := \mathbb{E}g(X_t)$. The ıinvariant measure μ of $\mathcal{L}$ is defined to be the measure that satisfies, for all functions g:

$$\int P_t g(X)\mu(dX) = \int g(X)\mu(dX)$$

Let us define the differential operator $\mathcal{L} = \frac{1}{2}\sum_{ij}\sigma^{ij}(X)\frac{\partial^2}{\partial X^i\partial X^j} + \sum_i \sigma_0^i(X)\frac{\partial}{\partial X^i}$ such that from Itô's formula applied to $g(X_t)$, we obtain:

$$g(X_t) = g(X_0) + \int_0^t \mathcal{L}g(X_s)ds + \int_0^t \sum_{i,j}\frac{\partial g}{\partial X^i}(X_s)\sigma^{ij}(X_s)dB_s^j \tag{8.5}$$

Taking expectation values in eq. (8.5) (and assuming this commutes through the integral and $\mathcal{L}$), the stochastic integral term vanishes by the martingale property and we obtain:

$$P_t g(X_0) - g(X_0) = \int_0^t \mathcal{L} P_s g(X_0) ds \tag{8.6}$$

Differentiating with respect to time then gives $\dot{P}_t g(X_0) = P_t g(X_0)$ and then after formally integrating we have $P_t g = e^{t\mathcal{L}} g$. Now, if $\mathcal{L}$ is an elliptic operator and if X_t takes values only in a compact set, then there is a unique invariant probability measure. If we suppose this measure to be of the form $\mu(dX) = p(X)dX$ for some density function p then we find:

$$\int \left[e^{t\mathcal{L}} g(X)\right] p(X) dX = \int g(X) p(X) dX \tag{8.7}$$

Upon integrating by parts we find:

$$\int \left[e^{t\mathcal{L}} g(X)\right] p(X) dX = \int g(X) \Big[e^{t\mathcal{L}^*} p(X)\Big] dX \tag{8.8}$$

Where $\mathcal{L}^*$ is the adjoint operator to $\mathcal{L}$ and is given by:

$$\mathcal{L}^* = \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial X^i \partial X^j} \big(\sigma^{ij}(X) \cdot\big) - \sum_i \frac{\partial}{\partial X^i} \big(\sigma_0^i(X) \cdot\big) \tag{8.9}$$

We conclude that $\mathcal{L}^* p = 0$ almost everywhere.

8.5.2 An interactive averaging principle –not covered in class

Let us consider an equation on $\mathbb{T}^d \times \mathbf{R}^d$: for $1 \leqslant i \leqslant d, \theta_t^\epsilon = (\theta_t^{1,\epsilon}, \dots, \theta_t^{m,\epsilon})$, $I_t^\epsilon = (I_t^{1,\epsilon}, \dots, I_t^{m,\epsilon})$,

$$\begin{aligned} d\theta_t^{i,\epsilon} &= \frac{1}{\sqrt{\varepsilon}} \sum_{k=1}^m \omega_k^i (I_t^\varepsilon) dW_t^k + K_\theta^i(\theta_t^\varepsilon, I_t) dt \\ dI_t^{\epsilon,i} &= K_I^i(\theta_t^\epsilon, I_t^\varepsilon) dt \end{aligned}$$

Theorem 8.5.1 *Let $\{\omega_k^i\}$ be smooth functions, and $(\omega_i^i(\theta))$ is a positive matrix at any point $\theta \in \mathbb{T}$. Let K_I^i smooth functions. For any $\beta \geqslant 1, > 0$ and for some function $c(t)$,*

$$\mathbb{E}\Big[\sup_{s \leqslant t} |I_i(s) - f_i(s)|^\beta\Big]^{1/\beta} \leqslant c(t) \epsilon^{1/4} \tag{8.10}$$

Proof This proof is a re-writing of that given in [11]. Throughout the proof we shall use the following inequalities which follow from the properties of norms on $\mathbf{R}^n$ (see

[**?**]). For all $\beta \geqslant 1$:

$$\left(\sum_{i=1}^{N} |x_i|\right)^{\beta} \leqslant N^{\beta-1} \sum_{i=1}^{N} |x_i|^{\beta} \qquad \left(\sum_{i=1}^{N} |x_i|\right)^{1/\beta} \leqslant \sum_{i=1}^{N} |x_i|^{1/\beta} \tag{8.11}$$

We wish to bound the following quantity:

$$H(y_t) - f(t) = \epsilon \int_0^t [g(y_s) - Q(f(bs))] \tag{8.12}$$

Now let us divide the interval of time integration into sub-intervals of length $\Delta t := (tb^{1-q})$:

$$\begin{aligned} &0 = t_0 \leqslant t_1 \leqslant ... \leqslant t_N \leqslant t_{N+1} = t \\ &t_n = n\Delta t \quad (\text{for } n = 0, ..., N); \qquad N := \left[b^{q-1}\right] \end{aligned}$$

Note the following useful bounds:

$$\Delta t \leqslant tb^{1-q} \qquad N \leqslant b^{1-q} \qquad t_{N+1} - t_N \leqslant \Delta t \tag{8.13}$$

After the division into sub-intervals we have:

$$\begin{aligned} H_i(y_t) - f_i(b(t)) = b \sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} & [g_i(y_r) - g_i(F_{t_n,r}(y_{t_n}))]dr & (8.14) \\ + b \sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} & \left[g_i(F_{t_n,r}(y_{t_n})) - Q_i(\tilde{H}(I_{t_n}))\right]dr & (8.15) \\ + b \sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} & \left[Q_i(\tilde{H}(I_{t_n})) - Q_i(\tilde{H}(I_r))\right]dr & (8.16) \\ + b \sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} & \left[Q_i(\tilde{H}(I_r)) - Q_i(f(br))\right]dr & (8.17) \end{aligned}$$

Thus there are four terms to bound; we shall refer to eqs. (8.14) to (8.17) as A_1^i, A_2^i, A_3^i, A_4^i respectively. Consider A_4^i:

$$\left|A_4^i\right| = b\left|\int_0^{t_{N+1}} \left[Q_i(\tilde{H}(I_r)) - Q_i(f(br))\right]dr\right| \leqslant b \sup_D |\nabla Q_i| \int_0^{t_{N+1}} \left|\tilde{H}(I_r) - f(br)\right|dr$$

Then:

$$|H_i(y_t) - f_i(b(t))| \leqslant \left|A_1^i\right| + \left|A_2^i\right| + \left|A_3^i\right| + b \sup_D |\nabla Q_i| \int_0^t |H(y_r) - f(br)|dr$$

If we sum this inequality over i we find:

$$\left|H(y_t) - f(b(t \wedge T^b))\right| \leqslant \sum_{i=1}^{m} \left(\left|A_1^i\right| + \left|A_2^i\right| + \left|A_3^i\right|\right)$$

$$+ b\left(\sum_{i=1}^{m} \sup_D |\nabla Q_i|\right) \int_0^t |H(y_r) - f(br)| dr$$

The Gronwall inequality may be applied to this to give:

$$\left|H(y_t) - f(b(t \wedge T^b))\right| \leqslant e^{bc_1 t} \sum_{i=1}^{m} \left(\left|A_1^i\right| + \left|A_2^i\right| + \left|A_3^i\right|\right)$$

where $c_1 := (\sum_{i=1}^m \sup_D |\nabla Q_i|)$. We may then deduce that:

$$\mathbb{E}\sup_{s\leqslant t} \left|H(y_s) - f(b(s \wedge T^b))\right|^{\beta 1/\beta} \leqslant c_2 e^{bc_1 t} \sup_i \sum_{l=1}^{3} \mathbb{E}\sup_{s\leqslant t} \left|A_l^i\right|^{\beta 1/\beta}$$

where $c_2 := m(3m)^{1-1/\beta}$. Now let us consider A_2:

$$\begin{aligned}
\left|A_2^i\right| &= b \sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} \left[\tilde{g}_i(F_{t_n,r}(\theta_{t_n}), I_{t_n}) - Q_i(\tilde{H}(I_{t_n}))\right] dr \\
&= b\Delta t \sum_{n=0}^{N} \left| \frac{1}{\Delta t} \int_0^{\Delta t} \tilde{g}_i(F_{t_n,r}(\theta_{t_n}), I_{t_n}) dr - \int_{\varphi\left(\mathcal{M}_{\tilde{H}(I_{t_n})}\right)} \tilde{g}_i(\theta, I_{t_n}) d\mu_{\tilde{H}(I_{t_n})}(\theta) \right| \qquad (8.18) \\
&\leqslant b\Delta t (N+1) \frac{c_3}{\sqrt{\Delta t}} \leqslant c_3^i (bt)^{1/2} \left(b^{q/2} + b^{1-q/2}\right) \qquad (8.19)
\end{aligned}$$

In going from eq. (8.18) to eq. (8.19) we have used the law of large numbers as stated in [11, pp. 814]/ From the above it also follows that:

$$\mathbb{E}\sup_{s\leqslant t} \left|A_2^i\right|^{\beta 1/\beta} \leqslant c_3^i (bt)^{1/2} \left(b^{q/2} + b^{1-q/2}\right) \qquad (8.20)$$

Now we turn our attention to A_3:

$$\begin{aligned}
\left|A_3^i\right| &= b \left| \sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} \left[Q_i(\tilde{H}(I_{t_n})) - Q_i(\tilde{H}(I_r))\right] dr \right| \\
&\leqslant b\Delta t \sup_D \left|\nabla(Q_i \circ \tilde{H})\right| \sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} |I_{t_n} - I_{t_n+r}|
\end{aligned}$$

From **??** we can see that:

$$|I_{t_n} - I_{t_n+r}| = b \left| \int_{t_n}^{t_n+r} K_I(\theta_s, I_s) ds \right| \leqslant b\Delta t \sup_D |K_I|$$

which implies:

$$\begin{aligned}
\left|A_3^i\right| &\leqslant c_4^i (N+1)(b\Delta t)^2 \leqslant c_4^i (bt)^2 (b^{1-q} + b^{2-2q}) \\
\implies \qquad \mathbb{E}\sup_{s\leqslant t} \left|A_2^i\right|^{\beta 1/\beta} &\leqslant c_4^i (bt)^2 (b^{1-q} + b^{2-2q})
\end{aligned}$$

where $c_4^i := \sup_{\mathbb{T}^m \times D} |K_I| \sup_D \left|\nabla(Q_i \circ \tilde{H})\right|$. It remains for us to consider A_1:

$$\begin{aligned}
\left|A_1^i\right| &= b\left|\sum_{n=0}^{N} \int_{t_n}^{t_{n+1}} [\tilde{g}_i(\theta_r, I_r) - \tilde{g}_i(F_{t_n,r}(\theta_{t_n}), I_{t_n})]dr\right| \\
&\leqslant b\Delta t \sup_{\mathbb{T}^m \times D} |\nabla \tilde{g}_i| \sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} \left(|\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})|^2 + |I_{t_n+r} - I_{t_n}|^2\right)^{1/2} \\
&\leqslant c_5^i b\Delta t \sum_{n=0}^{N} \left(\sup_{r\in[0,\Delta t]} |\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})| + \sup_{r\in[0,\Delta t]} |I_{t_n+r} - I_{t_n}|\right) \\
&\leqslant c_5^i b\Delta t \sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} |\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})| + c_6(bt)^2(b^{1-q} + b^{2-2q})
\end{aligned}$$

where $c_5^i := \sup_{\mathbb{T}^m \times D} |\nabla \tilde{g}_i|$ and $c_6^i = c_5^i \sup_{\mathbb{T}^m \times D} |K_I|$. This then implies:

$$\begin{aligned}
\mathbb{E}\sup_{s\leqslant t}\left|A_1^i\right|^{\beta\,1/\beta} \leqslant{}& \tilde{c}_5^i (N+1)^{1-1/\beta}(bt)b^{1-q}\mathbb{E}\sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} |\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})|^{\beta^{\,1/\beta}} \\
&+ \tilde{c}_6^i (bt)^2(b^{1-q} + b^{2-2q})
\end{aligned}$$

where $\tilde{c}_5^i := 2^{1-1/\beta} c_5^i$ and $\tilde{c}_6^i = 2^{1-1/\beta} c_6^i$. Now:

$$\begin{aligned}
|\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})| &\leqslant \left|\sum_{k=1}^{m} \int_{t_n}^{t_n+r} [\omega_k(I_s) - \omega_k(I_{t_n})]dB_s^k\right| + b\left|\int_{t_n}^{t_n+r} K_\theta(\theta_s, I_s)ds\right| \\
&\leqslant \sum_{k=1}^{m} \left|\int_{t_n}^{t_n+r} [\omega_k(I_s) - \omega_k(I_{t_n})]dB_s^k\right| + b\Delta t \sup_{\mathbb{T}^m \times D} |K_\theta|
\end{aligned}$$

From which we deduce:

$$\begin{aligned}
\mathbb{E}\sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} |\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})|^{\beta^{\,1/\beta}} \leqslant{}& c_8 \sum_{k=1}^{m} \mathbb{E}\sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} \left|\int_{t_n}^{t_n+r} [\omega_k(I_s) - \omega_k(I_{t_n})]dB_s^k\right|^{\beta\,1/\beta} \\
&+ c_7 c_8 (N+1)^{1/\beta}(bt)b^{1-q}
\end{aligned}$$

where $c_7 := \sup_{\mathbb{T}^m \times D} |K_I|^\beta$ and $c_8 := (m+1)^{1-1/\beta}$. We now apply the Burkholder-Davis-Gundy inequality to find:

$$\begin{aligned}
&\sum_{k=1}^{m} \mathbb{E}\sum_{n=0}^{N} \sup_{r\in[0,\Delta t]} \left|\int_{t_n}^{t_n+r} [\omega_k(I_s) - \omega_k(I_{t_n})]dB_s^k\right|^{\beta\,1/\beta} \\
&\qquad\leqslant c_9 \sum_{k=1}^{m} \mathbb{E}\sum_{n=0}^{N} \left(\int_{t_n}^{t_{n+1}} |\omega_k(I_s) - \omega_k(I_{t_n})|^2 ds\right)^{\beta/2\,1/\beta}
\end{aligned}$$

$$\leqslant c_9 \sum_{k=1}^{m} \mathbb{E} \sum_{n=0}^{N} \left(\Delta t \sup_D |\nabla \omega_k|^2 \sup_{s\in[t_n,t_{n+1}]} |I_s - I_{t_n}|^2 \right)^{\beta/2^{1/\beta}}$$

$$\leqslant c_{11} \sum_{k=1}^{m} \mathbb{E} \sum_{n=0}^{N} \left(b^2(\Delta t)^3\right)^{\beta/2^{1/\beta}} \leqslant \tilde{c}_{11}(bt)^{3/2} b^{1-3q/2}(N+1)^{1/\beta}$$

And so:

$$\mathbb{E} \sup_{r\in[0,\Delta t]} |\theta_{t_n+r} - F_{t_n,t_n+r}(\theta_{t_n})|^{\beta^{1/\beta}} \leqslant \left[c_8\tilde{c}_{11}(bt)^{3/2}b^{1-3q/2} + c_{10}(bt)b^{1-q}\right](N+1)^{1/\beta} \tag{8.21}$$

c_9 is the constant from the BDG inequality and $c_{10} = c_7c_8$. $c_{11} := c_9 \sup_D |\nabla\omega_k| \sup_{\mathbb{T}^m\times D} |K_I|$ and $\tilde{c}_{11} = mc_{11}$. Overall, then, we see:

$$\begin{aligned}\mathbb{E}\sup_{s\leqslant t} \left|A_1^i\right|^{\beta^{1/\beta}} &\leqslant \tilde{c}_5^i\tilde{c}_{11}(N+1)(bt)^{5/2}b^{2-5q/2} + \tilde{c}_5^i c_{10}(N+1)(bt)^2b^{2-2q} + \tilde{c}_6^i(bt)^2(b^{1-q}+b^{2-2q}) \\ &\leqslant c_{12}^i(bt)^{5/2}\left(b^{1-3q/2}+b^{2-5q/2}\right) + c_{13}^i(bt)^2\left(b^{1-q}+b^{2-2q}\right) + \tilde{c}_6^i(bt)^2(b^{1-q}+b^{2-2q})\end{aligned}$$

where $c_{12} := c_8\tilde{c}_5^i\tilde{c}_{11}$ and $c_{13} := \tilde{c}_5^i c_{10}$. We can now put everything together to obtain:

$$\begin{aligned}&\mathbb{E}\sup_{s\leqslant t} \left|H(y_s) - f(b(s\wedge T^b))\right|^{\beta^{1/\beta}} \\ &\quad\leqslant c_2e^{bc_1t}\Big\{\hat{c}_{12}(bt)^{5/2}\left(b^{1-3q/2}+b^{2-5q/2}\right) + \hat{c}_{13}(bt)^{1/2}\left(b^{1-q}+b^{2-2q}\right) + \hat{\tilde{c}}_6(bt)^2(b^{1-q}+b^{2-2q}) \\ &\qquad+\hat{c}_3(bt)^{1/2}b^{q/2} + \hat{c}_4(bt)^2(b^{1-q}+b^{2-2q})\Big\}\end{aligned}$$

where the hats on the constants denote that we have taken the supremum over i. If we re-scale the time $t \to t/b$ we find:

$$\begin{aligned}&\mathbb{E}\sup_{s\leqslant t} \left|H(y_{\frac{s}{b}\wedge T^b}) - f(s\wedge bT^b)\right|^{\beta^{1/\beta}} \\ &\quad\leqslant c_2e^{c_1t}\Big\{\hat{c}_{12}t^{5/2}\left(b^{1-3q/2}+b^{2-5q/2}\right) + \hat{c}_{13}t^{1/2}\left(b^{1-q}+b^{2-2q}\right) + \hat{\tilde{c}}_6t^2(b^{1-q}+b^{2-2q}) \\ &\qquad+\hat{c}_3t^{1/2}b^{q/2} + \hat{c}_4t^2(b^{1-q}+b^{2-2q})\Big\}\end{aligned}$$

The powers of b in the above are$1-\frac{3q}{2}$, $2-\frac{5q}{2}$, $1-q$, $2-2q$, and $\frac{q}{2}$; therefore the order of convergence is maximized if we choose $q = 1/2$. Then:

$$\begin{aligned}&\mathbb{E}\sup_{s\leqslant t} \left|H(y_{\frac{s}{b}\wedge T^b}) - f(s\wedge bT^b)\right|^{\beta^{1/\beta}} \\ &\quad\leqslant \left[b^{1/4}\left(\alpha_1t^{1/2}+\alpha_2t^{5/2}\right) + b^{1/2}\left(\alpha_3t^{1/2}+\alpha_4t^2\right) + b^{3/4}\alpha_5t^{5/2} + b\left(\alpha_6t^{1/2}+\alpha_7t^2\right)\right]e^{c_1t}\end{aligned}$$

for some constants α_i. For $b < 1$ we can write the above as:

$$\mathbb{E}\sup_{s\leqslant t} \left|H(y_{\frac{s}{b}\wedge T^b}) - f(s\wedge bT^b)\right|^{\beta^{1/\beta}} \leqslant b^{1/4}\left(\tilde{\alpha}_1t^{1/2}+\tilde{\alpha}_2t^2+\tilde{\alpha}_3t^{5/2}\right)e^{c_1t} \tag{8.22}$$

for some constants $\tilde{\alpha}_i$. □

8.6 Examples: not covered in class

8.6.1 Perturbation to Stochastic Integrable systems

Let $H : \mathbf{R}^{2d} \to \mathbf{R}$, consider the Hamiltonian equation

$$\dot{q} = \frac{\partial H}{\partial p}, \qquad \dot{p} = -\frac{\partial H}{\partial q},$$

Write $x_t = (q_t, p_t)$ and $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The equation becomes

$$\dot{x}_t = J\nabla H(x_t).$$

The energy H is a constant along the trajectory of the solution : $H(x_t) = H(x_0)$.

Example 8.6.1 Take $H = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2$. The general solution if of the form $q(t) = A\sin(\omega t) + B\cos(\omega t)$. Consider

$$(q(t), p(t)) = A(\sin(\omega t), \cos(\omega t)).$$

The level sets are ellipsoids.

Example 8.6.2 Let $H : \mathbf{R}^2 \to \mathbf{R}$, and $(\nabla H)^\perp$ is the skew gradient of H. Consider

$$dx_t^\epsilon = (\nabla H)^\perp(x_t^\epsilon) \circ dB_t + \epsilon V(x_t)dt.$$

In (H, θ), the action angle coordinates, It reduces to a system of equations where $H \in \mathbf{R}^n$ is the slow variable and $\theta \in S^n$ is the fast variables.

$$\begin{aligned} \frac{d}{dt}H_t &= \epsilon f(H_t^\epsilon, \theta_t^\epsilon), \\ d\theta_t &= X(H_t^\epsilon, \theta_t^\epsilon) \circ dW_t + \epsilon X_0(H_t^\epsilon, \theta_t^\epsilon)dt. \end{aligned}$$

Change time $t \mapsto t/\epsilon$:

$$\begin{aligned} \frac{d}{dt}\tilde{H}_t^\epsilon &= f(\tilde{H}_t^\epsilon, \tilde{\theta}_t^\epsilon), \\ d\tilde{\theta}_t &= \frac{1}{\sqrt{\epsilon}}X(\tilde{H}_t^\epsilon, \tilde{\theta}_t^\epsilon) \circ d\tilde{W}_t + X_0(\tilde{H}_t^\epsilon, \tilde{\theta}_t^\epsilon)dt. \end{aligned}$$

The Poisson bracket of two functions f and g is denoted by $\{f, g\}$, it is given by the formula:

$$\{f, g\} = \langle \nabla f, J\nabla g) = \sum_{i=1}^{d} \left(\frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial f}{\partial q_i} \right).$$

A Hamiltonian system in dimension $2d$ is said to be Liouville integrable if it has independent conserved quantities H_i, which are in involution, i.e. $\{H_i, H_j\} = 0$.

Louville's theorem : Any integrable system on $\mathbf{R}^{2d}$ is solvable by quadratures, i.e. the solution can be expressed explicitly by integrals. Furthermore, if the level set $\{H_i = c_i\}$ is compact and connected, it is diffeomorphic to the d-dimensional torus torus $\mathbb{T}^d$.

Suppose we have $\{H_i\}_{i=1}^n$ in evolution, an interesting model is:

$$dx_t^\epsilon = \sum_{i=1}^n H_i(x_t^\epsilon)dW_t^i + \epsilon V(x_t^\epsilon)dt.$$

In Darboux coordinates we have:

$$\begin{aligned} d\theta_t^{i,\epsilon} &= \frac{1}{\sqrt{\varepsilon}} \sum_{k=1}^m \omega_k^i(I_t^\varepsilon)dW_t^k + K_\theta^i(\theta_t^\varepsilon, I_t)dt \\ dI_t^{\epsilon,i} &= K_I^i(\theta_t^\epsilon, I_t^\varepsilon)dt \end{aligned}$$

8.6.2 Scaling of Riemannian metrics

$SU(2)$ which can be identified with the sphere S^3. The Lie algebra of $SU(2)$ is given by the Pauli matrices

$$X_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

By declaring $\{\frac{1}{\sqrt{\epsilon}}X_1, X_2, X_3\}$ an orthonormal frame we define Berger's metrics g^ϵ. Thus (S^3, g^ϵ) converges to S^2. Consider

$$\mathcal{L}^\epsilon := \frac{1}{\epsilon}(X_1)^2 + Y_0,$$

equivalent the SDE on the Lie group:

$$dg_t = \frac{1}{\sqrt{\epsilon}} X_1^*(g_t) \circ dB_t + Y_0^*(g_t)dt,$$

There is the Hopf fibration $\pi : SU(2) \to S^2(\frac{1}{2})$. Using this structure, we obtain u_t^ϵ such that u_t^ϵ and g_t^ϵ are both on $SU(2)$ and have the same projection on S^3/S^1:

$$\begin{aligned} \dot{u}_t^\epsilon &= (\mathrm{Ad}(h_{t/\epsilon})Y_0)^*(u_t^\epsilon) \\ dh_t &= X_1^*(h_t) \circ dB_t. \end{aligned}$$

Let us denote by $\Delta^{\epsilon}_{S^3}$ and Δ_{S^1} the Laplacians on (S^3, m_ϵ) and on S^1 respectively, and also denote by Δ^h the horizontal Laplacian identified with the Laplacian on $S^2(\frac{1}{2}) = S^3/S^1$.

$$dg_t = \frac{1}{\sqrt{\epsilon}} X_1(g_t) \circ db^1_t + X_2(g_t) \circ db^2_t + X_3(g_t) \circ db^3_t. \tag{8.23}$$

These operators commute and $\Delta^{\epsilon}_{S^3} = \frac{1}{\epsilon}\Delta_{S^1} + \Delta^h$. If $\{X_1, X_2, X_3\}$ are the Pauli matrices, identified with left invariant vector fields, then $\Delta_{S^1} = (X_1)^2$, $\Delta^h = (X_2)^2 + (X_3)^2$. As ϵ approaches 0, any eigenvalues of the Laplacian $\Delta^{\epsilon}_{S^3}$ coming from a non-zero eigenvalue of $\frac{1}{\epsilon}\Delta_{S^1}$ is pushed to the back of the spectrum and an eigenfunction of $\Delta^{\epsilon}_{S^3}$, not constant in the fibre, flies away. In other words the spectrums of S^3 converge to that of S^2. Cheeger, M. Gromov, Tanno-79, L. Bérard-Bergery and J. -P. Bourghignon, H. Urakawa]Urakawa86.

8.6.3 Geodesics

The geodesic equation on the orthonormal frame bundle solves

$$\dot{u}_t = H_{e_i}(u_t)$$

Then the projection of u_t is a geodesic with speed $u_0 e$. The solution to

$$du_t = \sum_{i=1}^{n} H_{e_i}(u_t) \circ dW^i_t$$

projects to a Brownian motion. We consider

$$du_t = \sum_{i=1}^{n} H_{e_i}(u_t) \circ dW^i_t + \epsilon V(u_t) dt.$$

Chapter 9

Appendix

9.1 Metric Spaces

A metric space is a space equipped with a distance function d. A metric is a function $d : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ such that the following holds for any points $x, y, z \in \mathcal{X}$: (i) $d(x, y) = d(y, x)$, (ii) $d(x, y) = 0$ if and only if $x = y$, (iii) $d(x, y) \leqslant d(x, z) + d(yz,)$.

A sequence of points $\{a_n\}$ in a metric space is called a Cauchy sequence if

$$\lim_{m \to \infty} d(a_n, a_{n+m}) \to 0.$$

A metric space $\mathcal{X}$ is **complete** if every Cauchy sequence has a limit in $\mathcal{X}$. A metric space is **separable** if there exists a countable dense subset $\{a_n\}$, meaning that every open ball $B_{x_0}(r) := \{x \in \mathcal{X} : d(x, x_0) < r\}$ contains at least one point from $\{a_n\}$.

A set in the metric space is **open** if whenever it contains a point x, there exists $r > 0$ such that it contains $B_x(r)$. A set U is said to be **closed** if its complement $\mathcal{X} \backslash U$, in $\mathcal{X}$, is an open set. Open balls $B_{x_0}(r)$ are open and closed balls $\bar{B}_{x_0}(r) : \{x \in \mathcal{X} : d(x, x_0) \leqslant r\}$ are closed sets.

Examples of metric space include the Euclidean spaces with $d(x, y) = |x - y|$ is complete separable metric spaces. Any finite dimensional complete Riemannian manifolds, with the Riemannian metric, are complete separable metric spaces. For clarity, in the definition of manifolds, we assume the properties of Hausdorff and second countability. A subset of the metric space with the same metric is a metric space.

Any set $\mathcal{X}$ is a metric space when endowed with the discrete metric: $d(x, y) = 1$ for any $x \neq y$ and $d(x, x) = 0$ for any point x, y.

Definition 9.1.1 • A metric space is said to be **sequentially compact** if every sequence of points in the space has a convergent subsequence.

- It is said to be **totally bounded** (or pre-compact) if, for every number $\epsilon > 0$, the space can be covered by a finite family of open balls of radius ϵ.

Every totally bounded set is separable, in particular every compact metric space is separable.

On a metric space, the following notions of compactness agree:

Proposition 9.1.2 *Let K be a subset of a metric space $\mathcal{X}$. The following statements are equivalent:*

- *Every open covering of K has a finite sub-covering.*
- *K is complete and totally bounded.*
- *Any infinite sequence of distinct points in K has a limit point in K.*

In other words, sequential compactness is equivalent to the space being totally bounded and complete.

If A is a set, define the distance function to A by: $d(x, A) = \inf_{y \in A} d(x, y)$.

Lemma 9.1.3 *Let $A \subset \mathcal{X}$. Then for any $x, z \in A$,*

$$|d(x, A) - (z, A)| \leqslant d(x, z).$$

Proof For any $x, y, z \in \mathcal{X}$, the triangle inequality gives $d(x, y) \leqslant d(x, z) + d(z, y)$. Taking the infimum over $y \in A$, we obtain: $d(x, A) \leqslant d(x, z) + d(z, A)$. This gives $d(x, A) - d(z, A) \leqslant d(x, z)$. The required inequality follows from the symmetry of d. □ In fact $d(x, A) = 0$ if and only if x belongs to the closure of A.

Definition 9.1.4 A space is said to be a Hausdorff space if the following hold:

- (1) [T_1, Fréchet] or ay $x \neq y$ there exist disjoint open sets U and V such that $x \in U$ and $y \in V$.
- (2) [T_4, normal] For any disjoint closed sets C_1 and C_2 there are disjoint open sets U and V such that $C_1 \subset U$ and $C_2 \subset V$.

For a general topological space, T_4 does not necessarily implies T_4/ Tkae the example $\mathcal{X} = \{0, 1\}$ with the topology consists of two open sets $\phi, \mathcal{X}$. The only closed sets are ϕ and $\mathcal{X}$, and one can take the open sets $\phi \subset \phi$ and $\mathcal{X} \subset \mathcal{X}$. But there are no open sets that distinguish the two points.

Proposition 9.1.5 *A metric space is a Hausdorff space.*

Proof Any singleton sets is closed, so T_4 implies T_1. Let A, B be disjoint closed sets. Since distance function to a set is continuous, $\{x : d(x,A) < d(x,B)\}$ and $\{x : d(x,A) > d(x,B)\}$ are open sets, and

$$A \subset \{x : d(x,A) < d(x,B)\}, \qquad B \subset \{x : d(x,A) > d(x,B)\}.$$

□

Metric space has a other nice property: If A and B are closed,

$$f(x) = \frac{d(x,A)}{d(x,A) + d(x,B)}$$

is a continuous function with the property that $f|_A = 0$ and $f|_B = 0$.

Theorem 9.1.6 (Urysohn's lemma) *For any closed and disjoint subsets A and B of $\mathcal{X}$ there exists a continuous function $f : \mathcal{X} \to [0,1]$ such that $f = 0$ on A and $f = 1$ on B.*

The following is taken from Theorem 4.34 [4, pp132].

Theorem 9.1.7 (Tietze Extension Theorem) *Suppose that $\mathcal{X}$ is a locally compact Hausdorff space and K is a compact subset of $\mathcal{X}$. If f is a real-valued continuous function on K, then there exists a continuous function F on $\mathcal{X}$ such that $F = f$ on K and F can be taken to vanish outside of a compact set.*

As usual, $\mathcal{B}(\mathcal{X})$ denotes the Borel σ-algebra on $\mathcal{X}$. It is the smallest collection of subsets of $\mathcal{X}$ that contains all open sets and closed under countable unions and countable intersections.

Definition 9.1.8 Let $\mathcal{X}$ be a metric space. A measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is called a Borel measure.

Theorem 9.1.9 *[14, Thm. 1.7, pp.4] The Borel σ-algebra on a metric space $\mathcal{X}$ is the smallest σ-algebra with respect to which all bounded, continuous, real valued functions on $\mathcal{X}$ are measurable.*

9.1.1 Algebra of functions

An algebra A (over the real number field) is a real vector space with a binary operation $A \times A \to A$, which we denote by $(x, y) \mapsto xy$, satisfying the following bilinear iproperties:

$$(x + y)z = xz + yz$$

$$z(x + y) = zx + zy$$
$$(az)(by) = (ab)(xy)$$

for all $x, y, z \in A$ and $a, b \in \mathbf{R}$.

If $\mathcal{X}$ is a metric space, the space $\mathbf{C}_b(\mathcal{X})$, of bounded continuous real-valued functions on $\mathcal{X}$ is an algebra with the usual pointwise multiplication as its binary operation. The spaces $\mathbf{C}_K(\mathcal{X})$ ((the space of continuous functions with compact support)) and $\mathbf{C}_0(\mathcal{X})$ (the space of continuous functions vanishing at infinity) are subspace of $\mathbf{C}_b(\mathcal{X})$, and are also algebras.

For a manifold $\mathcal{X}$, the sub-spaces of $\mathbf{C}^k$ or $\mathbf{C}^\infty$ functions within the above mentioned spaces, such as $\mathbf{C}_0^k(\mathcal{X})$ and $\mathbf{C}_K^\infty(\mathcal{X})$, are also algebras.

If $\mathcal{X} = \mathbf{R}^n$, the space of polynomials forms an algebra, while the space of linear functions does not.

9.2 Measures

By a measure we mean a σ-finite measure: $\mathcal{X} = \cup_{i=1}^n U_i$ where U_i are measurable sets of finite measure. On the metric space in the context of this note, we assume, in addition, that each ball has finite measure.

Proposition 9.2.1 *Let μ be a probability measure on separable metric space, there exists a unique closed set C such that $\mu(C) = 1$ and if D is a closed set with $\mu(D) = 1$ it is necessary that $C \subset D$. Moreover C is the collection of all points $x \in \mathcal{X}$ with the property that any open set containing x has positive measure.*

Definition 9.2.2 The closed set C in the above proposition is called the support or the spectrum of the measure μ.

Definition 9.2.3 A measure on a metric space is **regular** if the measure of any measurable set is determined by the values of the measures on open sets or on closed sets:

$$\mu(A) = \sup\{\mu(C) : C \subset A, \text{C is closed }\}$$
$$\mu(A) = \inf\{\mu(C) : C \supset A, \text{U is open}\}.$$

Definition 9.2.4 A measure μ is tight if for any $\epsilon > 0$ there exists a compact set $K \subset \mathcal{X}$ such that $\mu(\mathcal{X} \backslash K) < \epsilon$.

Lemma 9.2.5 *If $\mathcal{X}$ is a complete separable metric space, and μ a probability measure. Then for every $\varepsilon > 0$ there exists a compact set $K \subset \mathcal{X}$ such that $\mu(K) \geqslant 1 - \varepsilon$.*

Proof Let $\{r_i\}$ be a countable dense subset of $\mathcal{X}$ and denote by $\mathcal{B}(x, r)$ the ball of radius r centred at x. Note that since $\{r_k\}$ is a dense set, one has $\bigcup_{k>0} \mathcal{B}(r_k, 1/n) = \mathcal{X}$ for every n. Fix $\varepsilon > 0$ and, for every integer $n > 0$, denote by N_n the smallest integer such that

$$\mu\Big(\bigcup_{k \leqslant N_n} \mathcal{B}(r_k, \frac{1}{n})\Big) \geqslant 1 - \frac{\varepsilon}{2^n}\;.$$

Since $\bigcup_{k>0} \mathcal{B}(r_k, 1/n) = \mathcal{X}$, the number N_n is finite for every n. Define now the set K as

$$K = \bigcap_{n \geqslant 0} \bigcup_{k \leqslant N_n} \mathcal{B}(r_k, \frac{1}{n})\;.$$

It is clear that $\mu(K) > 1 - \varepsilon$. Furthermore, K is totally bounded, *i.e.* for every $\delta > 0$ it can be covered by a finite number of balls of radius δ (since it can be covered by N_n balls of radius $1/n$). It is a classical result from topology that in complete separable metric spaces, totally bounded sets have compact closure. □

Definition 9.2.6 (Functions of positive type) A function $f : V \to \mathbf{C}$ where V is a vector space is of **positive type** if for any n vectors $\lambda_1, \ldots, \lambda_n$ in V, the matrix A, with $A_{i,j} = f(\lambda_i - \lambda_j)$, is a positive semi-definite matrix and f is continuous on each finite dimensional subspace of V. Thus

1. $f(\lambda_i - \lambda_j) = \overline{f(\lambda_j - \lambda_i)}$,

2. $\sum_{i,j=1}^{n} f(\lambda_i - \lambda_j)\xi_i\overline{\xi_j} \geqslant 0$ for any $\xi_1, \ldots, \xi_n$ in $\mathbf{C}$.

Lemma 9.2.7 *If f is of positive type, then*

(1) $f(0) \geqslant 0$,

(2) $f(-x) = \overline{f(x)}$, *any* $x \in \mathbf{C}$,

(3) $|f(x)| \leqslant f(0)$ *for all* x, *so* f *is bounded.*

Proof Take $N = 1$ to get (1). Take $N = 2$, $\lambda_1 = 0$ and $\lambda_2 = -x$ to get (2) and

$$\begin{pmatrix} f(0) & f(x) \\ f(-x) & f(0) \end{pmatrix}$$

is positive semi-definite. So $f(0)^2 - f(x)\overline{f(x)} \geqslant 0$, giving part 3. □

9.3 Measure determining sets

Theorem 9.3.1 *Suppose that $(\Omega, \mathcal{F})$ is a measurable space, and* **C** *is a π-system generating $\mathcal{F}$. Let μ and ν be two measures which agree on* **C**.

1. *If $\mu(\Omega) = \nu(\Omega) < \infty$, then $\mu = \nu$*
2. *More generally, if there exists an increasing sequence of subsets $\Omega_k \in$* **C**, *such that $\Omega = \cup_{k\geqslant 1}\Omega_k$ and $\mu(\Omega_k) = \nu(\Omega_k) < \infty$ for all $k \geqslant 1$, then $\mu = \nu$.*

Proof (1) First assume that $\mu(\Omega) = \nu(\Omega) < \infty$. Let

$$\mathcal{G} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}.$$

Then $\Omega \in \mathcal{G}$ by assumption. Moreover, if A_n is a non-decreasing sequence of measurable sets in $\mathcal{G}$,

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu(\bigcup_{n=1}^{\infty} A_n)$$

Thus, $\mathcal{G}$ is closed under taking lower limit. Let $A \subset B$, $A, B \in \mathcal{G}$, then by additive property,

$$\mu(B\backslash A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B\backslash A).$$

This means $B\backslash A \in \mathcal{G}$, and therefore $\mathcal{G}$ is a λ-system containing a π-system generating $\mathcal{F}$. By the $\pi - \lambda$-Theorem, $\mathcal{G} \supset \mathcal{F}$ and $\mu = \nu$ on $\mathcal{F}$, which proves the first point.

(2) For the second point, let $\mathcal{F}_k = \{A \cap \Omega_k : A \in \mathcal{F}\}$ denote the trace σ-algebras on Ω_k, and denote by μ_k and ν_k the restrictions to Ω_k of the measures μ and ν:

$$\forall A \in \mathcal{F},\ \mu_k(A) = \mu(A \cap \Omega_k),\ \nu_k(A) = \nu(A \cap \Omega_k).$$

Applying the first point to μ_k and ν_k, we deduce that $\mu_k = \nu_k$. Therefore, by lower contiunity of measures, we obtain, for all $A \in \mathcal{F}$,

$$\mu(A) = \lim_{k\to\infty} \mu(A \cap \Omega_k) = \lim_{k\to\infty} \nu(A \cap \Omega_k) = \nu(A),$$

completing the proof. □

Example 9.3.2 Let $\Omega = \{1, 2, 3, 4\}$. Let $\mathcal{G} = \{\{1, 2\}, \{1, 3\}, \{3, 4\}\}$, which generates the discrete σ-algebra $\mathcal{F}$ (the power set), but not a π-system. The measure μ and ν agree on $\mathcal{G}$, not on $\mathcal{F}$.

$$\mu(1) = 1/6, \quad \mu(2) = 2/6, \quad \mu(3) = 1/6, \quad \mu(4) = 2/6,$$

$$\nu(1) = 2/6, \quad \nu(2) = 1/6, \quad \nu(3) = 0, \quad \nu(4) = 3/6.$$

9.4 Push forward measure

Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be two measurable spaces. Let μ be a measure on $\mathcal{X}$, and let $f : X \to Y$ be a measurable function. We denote by $f_*\mu$ the **push forward measure** under f, on $(Y, \mathcal{B})$, defined as

$$(f_*\mu)(B) = \mu(f^{-1}(B)),$$

for all $B \in \mathcal{B}$, where $f^{-1}(B) := \{x : f(x) \in B\}$ is the pre-image of B under f. In other words, the measure of B is assigned to be the measure of its pre-image.

Proposition 9.4.1 *Let $(\mathcal{Y}, \mathcal{B})$ be a measurable space, and let $\varphi : \mathcal{Y} \to \mathbf{R}$ a measurable function. Then, we have*

$$\int_{\mathcal{X}} \varphi \circ f \, d\mu = \int_{\mathcal{Y}} \varphi \, d(f_*\mu).$$

This is understood in the sense that φ is integrable with respect to $f_*\mu$ if and only if $\varphi \circ f$ is integrable with respect to μ.

Proof This result holds for indicator functions of measurable sets by the definition of the push forward measure. Applying the monotone convergence theorem on both sides shows that the set of functions with the desired property forms a monotone class. Finally apply the monotone class theorem to conclude the assertion. □

Definition 9.4.2 If $f : \mathcal{X} \to \mathcal{X}$ is a measurable map, we say that μ is **invariant** under f if $f_*(\mu) = \mu$.

We define the direct measure at $x \in \mathcal{X}$ as

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A, \end{cases}$$

for measurable sets $A \subset \mathcal{X}$.

Example 9.4.3 For any transformation $T : \mathbf{R}^n \to \mathbf{R}^n$, we have $T_*(\delta_x) = \delta_{Tx}$. The δ-measure is not invariant under rotations (unless $x = 0$), nor under translations.

Remark 9.4.4 Given a measure ν on $\mathcal{Y}$, there is no universally sensible way to construct a measure on $\mathcal{X}$ from ν and a measurable map $f : \mathcal{X} \to \mathcal{Y}$ in general. However, if $f : \mathbf{R} \to \mathbf{R}$ is injective, we can define the pullback measure, denoted by $f^*\nu$, as the measure $(f^{-1})_*\nu$. In other words, $f^*\nu(A) = (f^{-1})_*\nu(A) = \nu(f(A))$ for measurable sets $A \subset \mathcal{X}$.

On the other hand, consider the case where $\mathcal{Y} = [0,1]$, dx is the Lebesgue measure, and $f : \{1,2\} \to [0,1]$ is defined by $f(1) = 0$ and $f(2) = 0$. In this situation, there is no measure on $\{1,2\}$ whose push forward measure is dx. Therefore, there is no suitable measure μ on $\mathcal{X}$ such that $f_*(\mu) = \nu$.

9.4.1 Distributions of random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability measure. Let $(S, \mathcal{G})$ be a measurable space. A measurable function $X : \Omega \to S$ is called a **random variable**, and S is called its **state space**. The distribution of X is the push forward measure $X_*(\mathbb{P})$ on its state space $\mathcal{X}$, where $X^*(\mathbb{P})(A) = \mathbb{P}(X^{-1}(A))$. The distribution of a random variable encodes its statistical information.

If $\varphi : S \to \mathbf{R}$ and $X : \Omega \to S$ are measurable maps, then

$$\mathbb{E}[\varphi(X)] := \int_\Omega \varphi \circ X \, dP = \int_S \varphi \, d(X_*\mathbb{P}).$$

Proposition 9.4.5 *Let Y be a real-valued, non-negative random variable on $(\Omega, \mathcal{F})$. Then, $Y = 0$ almost surely if $\int_A Y = 0$ for every measurable set $A \in \mathcal{F}$. Consequently, two real valued, integrable random variables Y and Y' are equal almost surely if $\int_A Y = \int_A Y'$ for every measurable set A.*

Proof Suppose that $\{Y \neq 0\} > 0$. Since $\{Y > 0\} = \{Y \neq 0\}$ and has positive measure, there must exist some $a > 0$ such that the set $\{Y > a\}$ has positive measure (otherwise, $\mathbb{P}(Y > 0) = \lim_{n\to\infty} \mathbb{P}(Y > \frac{1}{n}) = 0$), But, then

$$\int_{\{Y>0\}} Y d\mathbb{P} \geqslant a\mathbb{P}(Y > a) > 0,$$

which contradicts the assumption that $\int_A Y = 0$ for any measurable set A. Hence $\mu(\{Y \neq 0\}) = 0$,completing the proof.

For the second statement, let $A = \{Y > Y'\}$. Applying the first result to $(Y - Y')\mathbf{1}_A$, we conclude that $Y = Y'$ on A.By symmetry, this concludes the proof. $\square$

9.5 Conditional Expectations

9.5.1 Absolute continuity of measures

Let P and Q be two measures on a measure space $(\Omega, \mathcal{F})$.

Definition 9.5.1 1. We say that Q is **absolutely continuous** with respect to P if $Q(A) = 0$ for all $A \in \mathcal{F}$ such that $P(A) = 0$. This is denoted as $Q \ll P$.

2. The measures P and$Q$ are said to be **equivalent**, denoted by $Q \sim P$, if they are absolutely continuous with respect to the other.

3. The two measures P and Q are said to be **singular** if $P(A) = 0$ whenever $Q(A) \neq 0$, and $Q(A) = 0$ whenever $P(A) \neq 0$.

Example 9.5.2 Let $\Omega = [0, 1)$ and $A_i^n = [\frac{i}{2^n}, \frac{i+1}{2^n})$ for each $n \in \mathbf{N}$, and $i = 0, 1, \ldots, 2^n - 1$. Let $\mathcal{F}_n$ denote the σ-algebra generated by the sets $A_0^n, A_1^n, \ldots, A_{2^n-1}^n$. Let dx denote the Lebesgue measure, restricted to $\mathcal{F}_n$, and let μ be a measure on $(\Omega, \mathcal{F}_n)$ such that $\mu \ll dx$. Then,

$$\frac{d\mu}{dx}(x) = \sum_i \frac{\mu(A_i^n)}{\mathrm{Leb}(A_i^n)} \mathbf{1}_{A_i^n}(x), \qquad x \in [0, 1),$$

where $\mathbf{1}_{A_i^n}$ in the indicator function of A_i^n.

Example 9.5.3 Let $\Omega = [0, 1]$ and let P denote the Lebesgue measure. Define the measure Q_1 by

$$\frac{dQ_1}{dP} = 2\mathbf{1}_{[0, \frac{1}{2}]}.$$

Then $Q_1 \ll P$, but P is not absolutely continuous with respect to Q_1. Now define Q_2 by $\frac{dQ_2}{dP} = 2\mathbf{1}_{[\frac{1}{2}, 1]}$. The two measures Q_1 and Q_2 are singular.

Theorem 9.5.4 (Radon-Nikodym Theorem) *If $Q \ll P$, there exists a nonnegative measurable function $\Omega \to \mathbf{R}$, which we denote by $\frac{dQ}{dP}$, such that for each measurable set A we have*

$$Q(A) = \int_A \frac{dQ}{dP}(\omega) dP(\omega).$$

The function $\frac{dQ}{dP} : \Omega \to \mathbf{R}$ is called the Radon-Nikodym derivative of Q with respect to P. We also refer to $\frac{dQ}{dP}$ as the density of Q with respect to P. This function is unique.

Note that if Q is a finite measure, then $\frac{dQ}{dP} \in L^1(\Omega, \mathcal{F}, P)$. If P is a probability measure and

$$\int_\Omega \frac{dQ}{dP}(\omega)\, dP(\omega) = 1,$$

then Q is also a probability measure.

Furthermore, if $\frac{dQ}{dP} > 0$, then

$$P(A) = \int_A dP = \int_A \frac{1}{\frac{dQ}{dP}} \frac{dQ}{dP} dP = \int_A \frac{1}{\frac{dQ}{dP}} dQ.$$

It follows that if $Q(A) = 0$ then $P(A) = 0$. Hence, the two measures are equivalent, and

$$\frac{dP}{dQ} \cdot \frac{dQ}{dP} = 1.$$

9.5.2 Conditional expectations

Definition 9.5.5 Let $X \in L^1(\Omega, \mathcal{F}, P)$ be a r.v.. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. A conditional expectation of X given $\mathcal{G}$ is any $\mathcal{G}$-measurable integrable random variable Y such that

$$\int_A X dP = \int_A Y dP, \qquad \forall A \in \mathcal{G} \tag{9.1}$$

Theorem 9.5.6 *Let $X \in L^1(\Omega, \mathcal{F}, P)$.*

(1) If $Y_1, Y_2 \in L^1(\Omega, \mathcal{G}, P)$ are conditional expectations of X then $Y_1 = Y_2$ a.s.

(2) If $a, b \in \mathbf{R}$, $X_1, X_2 \in L^1(\Omega, \mathcal{F}, P)$ then $\mathbb{E}(aX_1 + bX_2|\mathcal{G}) = a\mathbb{E}(X_1|\mathcal{G}) + b\mathbb{E}(X_2|\mathcal{G})$.

(3) The conditional expectation of X given $\mathcal{G}$ exists.

(4) If $X \geqslant 0$, $\mathbb{E}(X|\mathcal{G}) \geqslant 0$.

We denote by $\mathbb{E}(X|\mathcal{G})$ or $\mathbb{E}\{X|\mathcal{G}\}$ any version of the conditional expectation of X given $\mathcal{G}$.

Proof (1) We first prove uniqueness. Let Y_1, Y_2 be variables such that for any $A \in \mathcal{G}$,

$$\int_A (Y_1 - Y_2) dP = 0.$$

This implies that $Y_1 = Y_2$ a.s.

(2) The linearity follows from uniqueness.

(3) and (4). Assume that $X \geqslant 0$. Define $Q(A) = \int_A X(\omega) dP(\omega)$ for $A \in \mathcal{G}$. Then Q is a measure. The measure P restricts to a measure on $\mathcal{G}$. If $P(A) = 0$ then $Q(A) = 0$. By the Radon-Nikodym theorem, there exists a non-negative random variable $\frac{dQ}{dP}$, that belongs to $L^1(\Omega, \mathcal{G}, P)$, such that

$$Q(A) = \int_A X(\omega) dP(\omega) = \int_A \frac{dQ}{dP} dP.$$

Thus $\frac{dQ}{dP}$ satisfies (9.1) and is the conditional expectation of X given $\mathcal{G}$.

This proves (4).

Let $X \in L^1$. Then $X = X^+ - X^-$ where X^+, X^- are positive functions in L^1. By part (2) they have conditional expectations. We define

$$\mathbb{E}\{X|\mathcal{G}\} = \mathbb{E}\{X^+|\mathcal{G}\} - \mathbb{E}\{X^-|\mathcal{G}\}.$$

(The conditional expectation can also be obtained directly by Radon-Nikodym theorem for signed measures). This proves (3).

□

Proposition 9.5.7 *For all bounded $\mathcal{G}$-measurable functions g,*

$$\int_\Omega g(\omega)X(\omega)dP(\omega) = \int_\Omega g(\omega)\mathbb{E}\{X|\mathcal{G}\}(\omega)\, dP(\omega). \tag{9.2}$$

9.5.3 Properties of Conditional Expectations

Proposition 9.5.8 *Let $X, Y \in L^1(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$.*

1. *Positivity Preserving. If $X \leqslant Y$, then $\mathbb{E}(X|\mathcal{G}) \leqslant \mathbb{E}(Y|\mathcal{G})$.*
2. *Linearity. For all $a, b \in \mathbf{R}$,*

   $$\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G}).$$

3. $|\mathbb{E}(X|\mathcal{G})| \leqslant \mathbb{E}(|X|\,|\mathcal{G})$.
4. *If X is $\mathcal{G}$-measurable, $\mathbb{E}(X|G) = X$.*
5. *If $\sigma(X)$ is independent of $\mathcal{G}$, $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$ a.s.*
6. *Taking out what is known: If X is $\mathcal{G}$ measurable, $XY \in L^1$ then*

   $$\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G}).$$

7. $\mathbb{E}(\mathbb{E}(X|\mathcal{G})\,) = \mathbb{E}X$.
8. *Tower property: If $\mathcal{G}_1$ is a sub σ-algebra of $\mathcal{G}_2$ then*

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

9. *Conditional Jensen's Inequality. Let $\varphi : \mathbf{R}^d \to \mathbf{R}$ be a convex function. Then*

   $$\varphi(\mathbb{E}(X|\mathcal{G})) \leqslant \mathbb{E}(\varphi(X)|\mathcal{G}).$$

   For $p \geqslant 1$, $\|\mathbb{E}(X|\mathcal{G})\|_{L_p} \leqslant \|X\|_{L_p}$.

10. *Conditional dominated convergence Theorem. If* $|X_n| \leqslant g \in L^1$ *then*

$$\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G}).$$

11. L^1 *convergence. If* $X_n \to X$ *in* L^1 *then* $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ *in* L^1.

12. *Monotone Convergence Theorem. If* $X_n \geqslant 0$ *and* X_n *increases with* n *then* $\mathbb{E}(X_n|\mathcal{G})$ *increases to* $\mathbb{E}(\lim_{n\to\infty} X_n|\mathcal{G})$.

13. *Fatou's Lemma. If* $X_n \geqslant 0$,

$$\mathbb{E}(\liminf_{n\to\infty} X_n|\mathcal{G}) \leqslant \liminf_{n\to\infty} \mathbb{E}(X_n|\mathcal{G}).$$

14. *Suppose that* $\sigma(X) \vee \mathcal{G}$ *is independent of* $\mathcal{A}$, *then* $\mathbb{E}(X|\mathcal{A} \vee \mathcal{G}) = \mathbb{E}(X|\mathcal{G})$.

Proposition 9.5.9 *Let* $h : E \times E \to \mathbf{R}$ *be an integrable function on a metric space* E. *Let* X, Y *be random variables with state space* E *such that* $h(X,Y) \in L^1$. *Let* $H(y) = \mathbb{E}(h(X,y))$. *Then*

$$\mathbb{E}(h(X,Y)|\sigma(Y)) = H(Y).$$

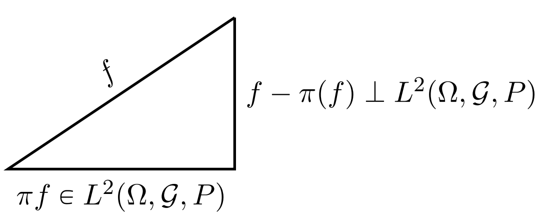
Proposition 9.5.10 *Let* $X : \Omega \to \mathcal{X}$ *and* $Y : \Omega \to \mathcal{Y}$ *be random variables with* X *measurable with respect to* $\mathcal{G} \subset \mathcal{F}$ *and* Y *is independent of* $\mathcal{G}$. *If* $\varphi : \mathcal{X} \times \mathcal{Y} \to \mathbf{R}$ *is a measurable function such that* $\varphi(X,Y)$ *is integrable, then*

$$\mathbb{E}(\varphi(X,Y)|\mathcal{G})(\omega) = \mathbb{E}(\varphi(X(\omega), Y), \quad a.s.$$

9.5.4 Disintegration and Orthogonal Projection

Let $\mathcal{G}$ be a sub-σ-algebra of a σ-algebra $\mathcal{F}$. Since $L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space and $L^2(\Omega, \mathcal{G}, P)$ is a closed subspace of L^2, let π denote the orthogonal projection defined by the projection theorem (§II.2 Functional Analysis [15]),

$$\pi : L^2(\Omega, \mathcal{F}, P) \to L^2(\Omega, \mathcal{G}, P).$$



We will see below that the conditional expectation of an L^2 function is precisely its L^2 orthogonal projection to $L^2(\Omega, \mathcal{G}, P)$. We give below second proof for the existence of conditional expectations.

Proof (1) Let $X \in L^2(\Omega, \mathcal{F}, P)$. Then for any $h \in L^2(\Omega, \mathcal{G}, P)$,

$$\langle X - \pi X, h \rangle_{L^2(\Omega, \mathcal{F}, P)} = 0.$$

This is,

$$\int_\Omega X h dP = \int_\Omega \pi(X) h dP$$

Let $A \in \mathcal{G}$ and take $h = \mathbf{1}_A$ to see that

$$\pi X = \mathbb{E}\{X | \mathcal{G}\}.$$

(2) Let $X \in L^1$ with $X \geqslant 0$. Let $0 \leqslant X_1 \leqslant X_2 \leqslant \dots$ be a sequence of bounded positive functions (increasing with n) converging to X pointwise. Then $X_n \in L^2$, $\{\pi X_n\}$ exists, and are positive. Furthermore for any $A \in \mathcal{G}$,

$$\int_A X_n dP = \int_A \pi X_n dP$$

Since,

$$0 \leqslant \mathbf{1}_A X_1 \leqslant \mathbf{1}_A X_2 \leqslant \dots,$$

$\lim_{n\to\infty} \pi X_n$ exists. By the monotone convergence theorem,

$$\int_A X dP = \lim_{n\to\infty} \int_A X_n dP = \lim_{n\to\infty} \int_A \pi X_n dP = \int_A \lim_{n\to\infty} \pi X_n dP.$$

(3) Finally for $X \in L^1$ not necessarily positive, let $X = X^+ - X^-$ and define $\mathbb{E}\{X|\mathcal{G}\} = \mathbb{E}\{X^+|\mathcal{G}\} - \mathbb{E}\{X^-|\mathcal{G}\}$.

□

Remark 9.5.11 Let $X \in L^2(\Omega, \mathcal{F}, P)$. Then πX is the unique element of $L^2(\Omega, \mathcal{G}, P)$ such that

$$\mathbb{E}|X - \pi X|^2 = \min_{Y \in L^2(\Omega, \mathcal{G}, P)} \mathbb{E}|X - Y|^2.$$

9.5.5 Note on filtering

At this point we note a simple problem from Filtering Theory. Let Y_t be the observation process of a signal process. What is the best estimation for X_t given $\{Y_s, s \leqslant t\}$? We have seen that in the L^2 case, the conditional expectation is an L^2 minimizer. We therefore define the L^2 estimator to be:

$$\hat{X}_t := \mathbb{E}\{X_t|\sigma\{Y_s : 0 \leqslant s \leqslant t\}\}.$$

The concern in filtering is to find the conditional distribution, and the conditional density when it exists, of $X(t)$ given $Y(t)$.

In linear filtering, we assume that

$$X_t(\omega) = X_0(\omega) + W_t(\omega) + \int_0^t F(s)X_s(\omega)ds + \int_0^t f(s)ds$$
$$Y_t(\omega) = \int_0^t H(s)X_s ds + \int_0^t h(s)ds + B_t(\omega).$$

Here $\{(W_t),(B_t)\}$ are independent Brownian motions and both independent of X_0. We assume that $F, f, H, h : \mathbf{R}_+ \to \mathbf{R}$ are bounded measurable functions. This leads to Karman Filter, linear filtering and Zakai equation.

9.6 Uniform Integrability

Let $(\Omega, \mathcal{F}, \mu)$ be a (σ-finite) measure space, and I an index set.

Definition 9.6.1 A family of real-valued measurable functions $(f_\alpha, \alpha \in I)$ is uniformly integrable (u.i.) if

$$\lim_{C\to\infty} \sup_{\alpha\in I} \int_{\{|f_\alpha|\geqslant C\}} |f_\alpha| d\mu = 0.$$

Lemma 9.6.2 (Uniform Integrability of Conditional Expectations) *Let $X : \Omega \to \mathbf{R}$ be in L^1. Then the family of functions*

$$\{\mathbb{E}\{X|\mathcal{G}\} : \mathcal{G} \text{ is a sub } \sigma\text{-algebra of } \mathcal{F}\}$$

is uniformly integrable.

Proof exercise. □

Theorem 9.6.3 *(Vitali Theorem) Let $f_n \in L^p(\mu)$, $p \in [1, \infty]$. Then the following is equivalent.*

1. *$f_n \stackrel{L^p}{\to} f$, i.e. $\lim_{n\to\infty} \|f_n - f\|_p = 0$.*
2. *$\{|f_n|^p\}$ is uniformly integrable and $f_n \to f$ in measure.*
3. *$\int |f_n|^p d\mu \to \int |f|^p d\mu$ and $f_n \to f$ in measure.*

9.7 Uniformly absolute continuity

Let $(S, \mathcal{A}, \mu)$ be a measure space. Let $f, f_\alpha : S \to \mathbf{R}$ be Borel measurable functions.

Proposition 9.7.1 *If $f \in L^1(\mu)$ where μ is a σ-finite measure, for every $\epsilon > 0$ there is $\delta > 0$ such that for all A with $\mu(A) < \delta$,*

$$\int_A |f| d\mu < \epsilon.$$

Proof We define a measure $\nu(A) = \int_A f d\mu$. It is a signed measure with both the positive and negative part absolutely continuous w.r.t. μ. By considering ν^+, ν^- separately, we may and will assume that $f \geqslant 0$ and ν is a positive measure. If the conclusion does not hold, there exists a positive number ϵ such that for each n there is a set A_n with $\mu(A_n) < \frac{1}{2^n}$ and

$$\nu(A_n) = \int_{A_n} |f| d\mu \geqslant \epsilon.$$

Let $A = \cap_{n=1}^\infty \cup_{k=n}^\infty A_k$. Then,

$$\mu(A) = \mu(\cap_{n=1}^\infty \cup_{k=n}^\infty A_k) = \lim_{n\to\infty} \mu(\cup_{k=n}^\infty A_k) = 0.$$

In particular $\int_A f d\mu = 0$. But,

$$\nu(A) = \lim_{n\to\infty} \nu(\cup_{k=n}^\infty A_k) \geqslant \nu(A_n) \geqslant \epsilon.$$

This gives a contradiction. □

Definition 9.7.2 A family of integrable real valued random functions $\{f_\alpha\}$ is uniformly absolutely continuous if for every $\epsilon > 0$ there is a number $\delta > 0$ such that if a measurable set A has $\mu(A) < \delta$ then for all $\alpha \in I$

$$\int_A |f_\alpha| d\mu < \epsilon.$$

Proposition 9.7.3 *Let μ be a finite measure. Let $(f_\alpha, \alpha \in I)$ be a family of integrable real valued functions. The following statements are equivalent:*

(1) $(f_\alpha, \alpha \in I)$ *is uniformly integrable (u.i.)*

(2) $(f_\alpha, \alpha \in I)$ *is L^1 bounded and uniformly absolutely continuous.*

(3) (de la Vallee-Poussin criterion) There exists an increasing convex function $\Phi : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\lim_{x\to\infty} \frac{\Phi(x)}{x} = \infty$ and $\sup_\alpha \mathbb{E}(\Phi(|f_\alpha|)) < \infty$.

Proposition 9.7.4 *Let $(S, \mathcal{A}, \mu)$ be a measure space. Suppose that $f_n : S \to \mathbf{R}$ belongs to L^1.*

1. *If $f_n \to f$ in L^1 then $\{f_n\}$ is L^1 bounded.*
2. *If $f_n \to f$ in L^1, then $\{f_n\}$ is uniformly absolutely continuous. See exercise 11, section 3.2 in [4].*
3. *Suppose that μ is a finite measure. If $f_n \to f$ in measure and $\{f_n\}$ is uniformly absolutely continuous then $f_n \to f$ in L^1.*

Proof By Riesz-Fisher theorem, the L^1 space is a complete Banach space. (1) is obvious.

(2)Suppose that $f_n \to f$ in L^1. For any $\epsilon > 0$ there is $N(\epsilon)$ such that

$$\sup_{n \geqslant N} \int |f_n - f| d\mu < \epsilon/2.$$

Let $\alpha > 0$ be such that if $\mu(A) < \alpha$ then

$$\int_A |f| d\mu < \epsilon/2, \qquad \sup_{k \leqslant N-1} \int_A |f_k| d\mu < \epsilon.$$

For $n \geqslant N$,

$$\int_A |f_n| d\mu \leqslant \int |f_n - f| d\mu + \int_A |f| d\mu < \epsilon.$$

(3) We may assume that $\mu = P$ is a probability measure.

Suppose that $\{f_n\}$ is uniformly absolutely continuous and $f_n \to f$ in measure, i.e. for any $\epsilon > 0$,

$$\lim_{n \to \infty} P(|f_n - f| > \frac{\epsilon}{3}) = 0.$$

Let $\epsilon > 0$. Choose $\delta(\epsilon) > 0$, such that if E is a measurable set with $\mu(E) < \delta$,

$$\sup_n \int_E |f_n| dP < \epsilon/3, \quad \int_E |f| dP < \epsilon/3.$$

There exists $N(\epsilon, \delta)$ such that for $P(|f_n - f| > \epsilon/3) < \delta$ whenever $n \geqslant N(\delta, \epsilon)$. For such n,

$$\int |f_n - f| dP \leqslant \int_{|f_n - f| \leqslant \frac{\epsilon}{3}} |f_n - f| dP + \int_{|f_n - f| > \frac{\epsilon}{3}} |f_n| dP + \int_{|f_n - f| > \frac{\epsilon}{3}} |f| dP < \epsilon.$$

It follows that $f_n \to f$ in L^1.

□

Bibliography

[1] Jagdish Chandra and B. A. Fleishman. On a generalization of the Gronwall-Bellman lemma in partially ordered Banach spaces. *J. Math. Anal. Appl.*, 31:668–681, 1970.

[2] K. D. Elworthy, Xue-Mei Li, and M. Yor. The importance of strictly local martingales; applications to radial Ornstein-Uhlenbeck processes. *Probab. Theory Related Fields*, 115(3):325–355, 1999.

[3] 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.

[4] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[5] W. Gerstner, W.M. Kistler, R. Naud, , and L. Paninski. *Neuronal Dynamics*. Cambridge University Press, 2014.

[6] T. H. Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math. (2)*, 20(4):292–296, 1919.

[7] Martin Hairer. 2021. preprint.

[8] 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.

[9] G. Stephen Jones. Fundamental inequalities for discrete and discontinuous functional equations. *J. Soc. Indust. Appl. Math.*, 12:43–57, 1964.

[10] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer, [Cham], french edition, 2016.

[11] Xue-Mei Li. An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity*, 21(4):803–822, 2008.

[12] Alber M., A. Buganza Tepole, W. R. Cannon, Suvranu De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, William W. Karniadakis L., P. Perdikaris, and E. Petzold, L.and Kuhl. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digital Medicine*, 2(1), 2019.

[13] Edward Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, NJ, 1967.

[14] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press, Inc., New York-London, 1967.

[15] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I. Functional analysis*. Academic Press, New York-London, 1972.

[16] 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, second edition, 1994.

[17] 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.

[18] 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.