APPLIED STOCHASTIC PROCESSES

G.A. Pavliotis
Department of Mathematics
Imperial College London
London SW7 2AZ, UK

January 18, 2009

Contents

Chapter 1

Introduction

1.1 Historical Overview

- The theory of stochastic processes started with Einstein's work on the theory of Brownian motion: *Concerning the motion, as required by the molecular-kinetic theory of heat, of particles suspended in liquids at rest* (1905).
  - explanation of Brown's observation (1827): when suspended in water, small pollen grains are found to be in a very animated and irregular state of motion.
  - Einstein's theory is based on
    - A Markov chain model for the motion of the particle (molecule, pollen grain...).
    - The idea that it makes more sense to talk about the probability of finding the particle at position x at time t, rather than about individual trajectories.
- In his work many of the main aspects of the modern theory of stochastic processes can be found:
  - The assumption of Markovianity (no memory) expressed through the Chapman-Kolmogorov equation.
  - The Fokker–Planck equation (in this case, the diffusion equation).
  - The derivation of the Fokker-Planck equation from the master (Chapman-Kolmogorov) equation through a Kramers-Moyal expansion.
  - The calculation of a transport coefficient (the diffusion equation) using macroscopic (kinetic theory-based) considerations:
    $$D = \frac{k_B T}{6\pi\eta a}.$$
  - k_B is Boltzmann's constant, T is the temperature, η is the viscosity of the fluid and a is the diameter of the particle.
- Einstein's theory is based on the **Fokker-Planck equation**. Langevin (1908) developed a theory based on a **stochastic differential equation**. The equation of motion for a Brownian particle is
  $$m\frac{d^2x}{dt^2} = -6\pi\eta a\frac{dx}{dt} + \xi,$$

- where ξ is a random force.
- There is complete agreement between Einstein's theory and Langevin's theory.
- The theory of Brownian motion was developed independently by Smoluchowski.
- The approaches of Langevin and Einstein represent the two main approaches in the theory of stochastic processes:
  - Study individual trajectories of Brownian particles. Their evolution is governed by a stochastic differential equation:
$$\frac{dX}{dt} = F(X) + \Sigma(X)\xi(t),$$
  - where $\xi(t)$ is a random force.
- Study the probability $\rho(x,t)$ of finding a particle at position x at time t. This probability distribution satisfies the Fokker-Planck equation:
$$\frac{\partial \rho}{\partial t} = \nabla \cdot (F(x)\rho) + \frac{1}{2}\nabla\nabla : (A(x)\rho),$$
- where $A(x) = \Sigma(x)\Sigma(x)^T$.
- The theory of stochastic processes was developed during the 20th century:
  - Physics:
    - Smoluchowksi.
    - Planck (1917).
    - Klein (1922).
    - Ornstein and Uhlenbeck (1930).
    - Kramers (1940).
    - Chandrasekhar (1943).
    - ...
  - Mathematics:
    - Wiener (1922).
    - Kolmogorov (1931).
    - Itô (1940's).
    - Doob (1940's and 1950's).
    - ...

1.2 The One-Dimensional Random Walk

We let time be discrete, i.e. $t = 0,\ 1,\ \ldots$. Consider the following stochastic process S_n:

- S_0 = 0;
- at each time step it moves to ± 1 with equal probability $\frac{1}{2}$.

In other words, at each time step we flip a fair coin. If the outcome is heads, we move one unit to the right. If the outcome is tails, we move one unit to the left.

Alternatively, we can think of the random walk as a sum of independent random variables:

$$S_n = \sum_{j=1}^{n} X_j,$$

where $X_j \in \{-1, 1\}$ with $\mathbb{P}(X_j = \pm 1) = \frac{1}{2}$.

We can simulate the random walk on a computer:

- We need a **(pseudo)random number generator** to generate n independent random variables which are **uniformly distributed** in the interval [0,1].
- If the value of the random variable is $\geqslant \frac{1}{2}$ then the particle moves to the left, otherwise it moves to the right.
- We then take the sum of all these random moves.
- The sequence $\{S_n\}_{n=1}^{N}$ indexed by the discrete time $T = \{1,\ 2,\ \ldots N\}$ is the **path** of the random walk. We use a linear interpolation (i.e. connect the points $\{n, S_n\}$ by straight lines) to generate a **continuous path**.
- Every path of the random walk is different: it depends on the outcome of a sequence of independent random experiments.
- We can compute statistics by generating a large number of paths and computing averages. For example, $\mathbb{E}(S_n) = 0$, $\mathbb{E}(S_n^2) = n$.
- The paths of the random walk (without the linear interpolation) are not continuous: the random walk has a jump of size 1 at each time step.
- This is an example of a **discrete time, discrete space** stochastic processes.
- The random walk is a **time-homogeneous Markov** process.
- If we take a large number of steps, the random walk starts looking like a continuous time process with continuous paths.

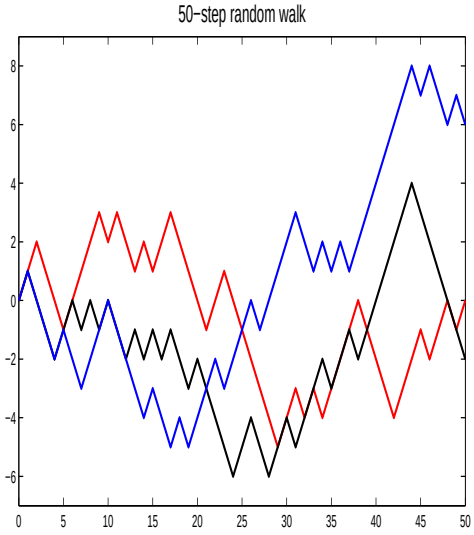


Figure 1.1: Three paths of the random walk of length $N = 50$.

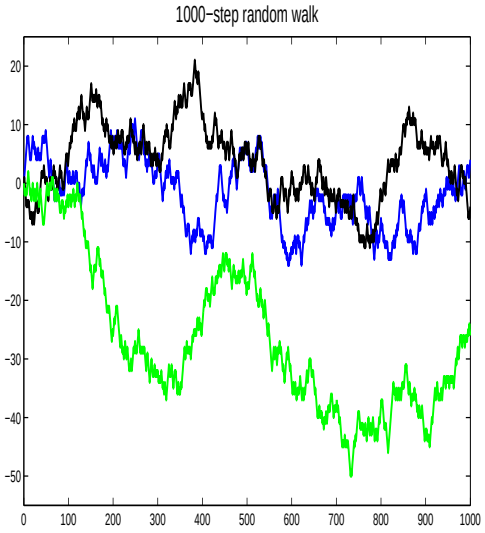


Figure 1.2: Three paths of the random walk of length $N = 50$.

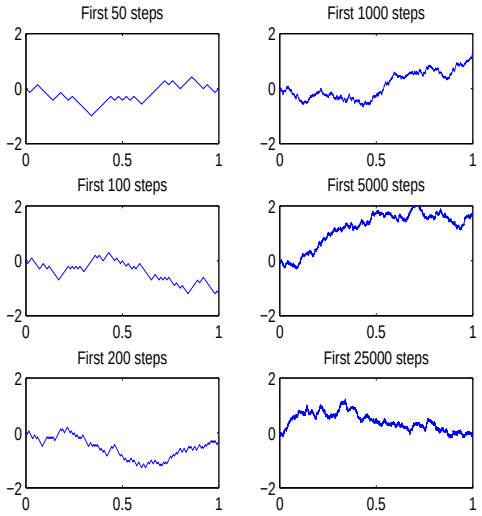


Figure 1.3: Space-time rescaled random walks.

- Consider the sequence of **continuous time** stochastic processes

$$Z_t^n := \frac{1}{\sqrt{n}} S_{nt}.$$

- In the limit as $n \to \infty$, the sequence $\{Z_t^n\}$ converges (in some appropriate sense) to a **Brownian motion** with **diffusion coefficient** $D = \frac{\Delta x^2}{2\Delta t} = \frac{1}{2}$.

- Brownian motion $W(t)$ is a continuous time stochastic processes with continuous paths that starts at 0 ($W(0) = 0$) and has independent, normally. distributed Gaussian increments.

- We can simulate the Brownian motion on a computer using a random number generator that generates normally distributed, independent random variables.

Why introduce randomness in the description of physical systems?

- To describe outcomes of a repeated set of experiments. Think of tossing a coin repeatedly or of throwing a dice.

- To describe a deterministic process for which we have incomplete information. Think of an ODE for which we don't know exactly the initial conditions (weather prediction).

- To describe a complicated deterministic system with many degrees of freedom using a simpler, low dimensional stochastic system. Think of the physical model for Brownian motion (a heavy particle colliding with many small particles).

- To describe a system that is inherently random. Think of quantum mechanics.

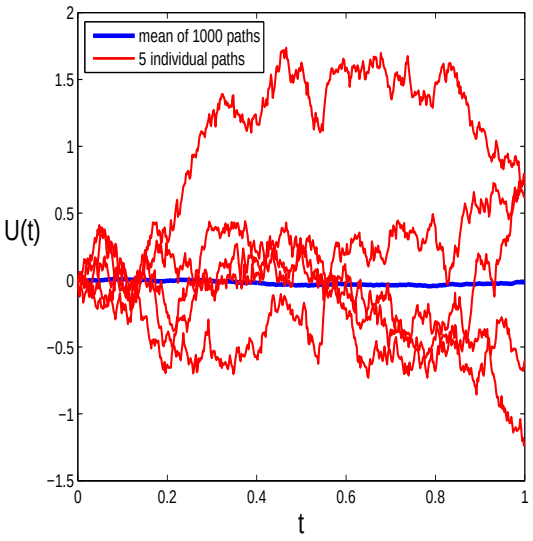


Figure 1.4: Sample Brownian paths.

Chapter 2

Elements of Probability Theory

2.1 Basic Definitions

Definition 2.1.1. *The set of all possible outcomes of an experiment is called the* **sample space** *and is denoted by* Ω.

Example 2.1.2.

- *The possible outcomes of the experiment of tossing a coin are* H *and* T. *The sample space is* $\Omega = \{H, T\}$.
- *The possible outcomes of the experiment of throwing a die are* $1, 2, 3, 4, 5$ *and* 6. *The sample space is* $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- **Events** are subsets of the sample space.
- We would like the unions and intersections of events to also be events.

Definition 2.1.3. *A collection* $\mathcal{F}$ *of* Ω *is called a* **field** *on* Ω *if*

i. $\emptyset \in \mathcal{F}$;

ii. *if* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$;

iii. *If* $A, B \in \mathcal{F}$ *then* $A \cup B \in cF$.

From the definition of a field we immediately deduce that $\mathcal{F}$ is closed under finite unions and finite intersections:

$$A_1, \ldots A_n \in \mathcal{F} \quad \Rightarrow \quad \cup_{i=1}^n A_i \in \mathcal{F}, \quad \cap_{i=1}^n A_i \in \mathcal{F}.$$

When Ω is infinite dimensional then the above definition is not appropriate since we need to consider countable unions of events.

Definition 2.1.4. *A collection* $\mathcal{F}$ *of* Ω *is called a* σ**-field** *or* σ**-algebra** *on* Ω *if*

i. $\emptyset \in \mathcal{F}$;

ii. *if* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$;

iii. If $A_1,\ A_2, \cdots \in \mathcal{F}$ *then* $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A σ-algebra is closed under the operation of taking countable intersections.

Example 2.1.5.

- $\mathcal{F} = \{\emptyset,\ \Omega\}$.
- $\mathcal{F} = \{\emptyset,\ A,\ A^c,\ \Omega\}$ *where* A *is a subset of* Ω.
- *The* **power set** *of* Ω, *denoted by* $\{0,1\}^{\Omega}$ *which contains all subsets of* Ω.

A sub-σ–algebra is a collection of subsets of a σ–algebra which satisfies the axioms of a σ–algebra.

Definition 2.1.6. *A* **probability measure** $\mathbb{P}$ *on the* **measurable space** $(\Omega,\ \mathcal{F})$ *is a function* $\mathbb{P} : \mathcal{F} \mapsto [0,1]$ *satisfying*

i. $\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\Omega) = 1$*;*

ii. For $A_1,\ A_2, \ldots$ *with* $A_i \cap A_j = \emptyset,\ i \neq j$ *then*

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Definition 2.1.7. *The triple* $\big(\Omega,\ \mathcal{F},\ \mathbb{P}\big)$ *comprising a set* Ω*, a* σ*-algebra* $\mathcal{F}$ *of subsets of* Ω *and a probability measure* $\mathbb{P}$ *on* $(\Omega,\ \mathcal{F})$ *is a called a* **probability space**.

- Assume that $\mathbb{E}|X| < \infty$ and let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{F}$. The **conditional expectation** of X with respect to $\mathcal{G}$ is defined to be the function $\mathbb{E}[X|\mathcal{G}] : \Omega \mapsto E$ which is $\mathcal{G}$–measurable and satisfies

$$\int_G \mathbb{E}[X|\mathcal{G}]\, d\mu = \int_G X\, d\mu \quad \forall\, G \in \mathcal{G}.$$

- We can define $\mathbb{E}[f(X)|\mathcal{G}]$ and the conditional probability $\mathbb{P}[X \in F|\mathcal{G}] = \mathbb{E}[I_F(X)|\mathcal{G}]$, where I_F is the indicator function of F, in a similar manner.

2.2 Random Variables

Definition 2.2.1. *Let* $(\Omega, \mathcal{F})$ *and* $(E, \mathcal{G})$ *be two measurable spaces. A function* $X : \Omega \mapsto E$ *such that the event*

$$\{\omega \in \Omega : X(\omega) \in A\} =: \{X \in A\} \tag{2.1}$$

belongs to $\mathcal{F}$ *for arbitrary* $A \in \mathcal{G}$ *is called a measurable function or random variable.*

When E is $\mathbb{R}$ equipped with its Borel σ-algebra, then (2.1) can by replaced with

$$\{X \leqslant x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

- Let X be a random variable (measurable function) from $(\Omega, \mathcal{F}, \mu)$ to $(E, \mathcal{G})$. If E is a metric space then we may define *expectation* with respect to the measure μ by

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mu(\omega).$$

- More generally, let $f : E \mapsto \mathbb{R}$ be $\mathcal{G}$–measurable. Then,

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\, d\mu(\omega).$$

- Let U be a topological space. We will use the notation $\mathcal{B}(U)$ to denote the Borel σ–algebra of U: the smallest σ–algebra containing all open sets of U. Every random variable from a probability space $(\Omega, \mathcal{F}, \mu)$ to a measurable space $(E, \mathcal{B}(E))$ induces a probability measure on E:

$$\mu_X(B) = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(E).$$

  The measure μ_X is called the *distribution* (or sometimes the *law*) of X.

Example 2.2.2. *Let $\mathcal{I}$ denote a subset of the positive integers. A vector $\rho_0 = \{\rho_{0,i},\ i \in \mathcal{I}\}$ is a distribution on $\mathcal{I}$ if it has nonnegative entries and its total mass equals 1: $\sum_{i\in\mathcal{I}} \rho_{0,i} = 1$.*

In the case where $E = \mathbb{R}$, the **distribution function** of the random variable X is the function $F : [0, 1] \mapsto [0, 1]$ given by

$$F_X(x) = \mathbb{P}(X \leqslant x). \tag{2.2}$$

In this case, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X)$ becomes a probability space.

The distribution function $F(x)$ of a random variable has the properties that $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$ and is right continuous.

Definition 2.2.3. *A random variable X with values on $\mathbb{R}$ is called* **discrete** *if it takes values in some countable subset $\{x_0, x_1, x_2, \dots\}$ of $\mathbb{R}$. i.e.: $\mathbb{P}(X = x) \neq x$ only for $x = x_0, x_1, \dots$.*

With a random variable we can associate the **probability mass function** $p_k = \mathbb{P}(X = x_k)$. We will consider nonnegative integer valued discrete random variables. In this case $p_k = \mathbb{P}(X = k),\ k = 0, 1, 2, \dots$.

Example 2.2.4. *The Poisson random variable is the nonnegative integer valued random variable with probability mass function*

$$p_k = \mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-k}, \quad k = 0, 1, 2, \dots,$$

where $\lambda > 0$.

Example 2.2.5. *The binomial random variable is the nonnegative integer valued random variable with probability mass function*

$$p_k = \mathbb{P}(X = k) = \frac{N!}{n!(N-n)!}p^n q^{N-n} \quad k = 0, 1, 2, \dots N,$$

where $p \in (0, 1),\ q = 1 - p$.

Definition 2.2.6. *A random variable X with values on $\mathbb{R}$ is called* **continuous** *if $\mathbb{P}(X = x) = 0 \, \forall x \in \mathbb{R}$.*

Let X be a random variable from $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The (joint) distribution function $F_X \mathbb{R}^d mapsto [0, 1]$ is defined ass

$$F_X(\mathbf{x}) = \mathbb{P}(X \leqslant \mathbf{x}).$$

- We can use the distribution of a random variable to compute expectations and probabilities:

$$\mathbb{E}[f(X)] = \int_S f(x) \, d\mu_X(x)$$

and

$$\mathbb{P}[X \in G] = \int_G d\mu_X(x), \quad G \in \mathcal{B}(E).$$

- When $E = \mathbb{R}^d$ and we can write $d\mu_X(x) = \rho(x) \, dx$, then we refer to $\rho(x)$ as the **probability density function** (pdf), or *density with respect to Lebesque measure* for X.

- When $E = \mathbb{R}^d$ then by $L^p(\Omega; \mathbb{R}^d)$, or sometimes $L^p(\Omega; \mu)$ or even simply $L^p(\mu)$, we mean the Banach space of measurable functions on Ω with norm

$$\|X\|_{L^p} = \Big(\mathbb{E}|X|^p\Big)^{1/p}.$$

- Let X be a nonnegative integer valued random variable with probability mass function p_k. We can compute the expectation of an arbitrary function of X using the formula

$$\mathbb{E}(f(X)) = \sum_{k=0}^{\infty} f(k) p_k.$$

Example 2.2.7.

- *Consider the random variable $X : \Omega \mapsto \mathbb{R}$ with pdf*

$$\gamma_{\sigma,m}(x) := (2\pi\sigma)^{-\frac{1}{2}} \exp\left(-\frac{(x-m)^2}{2\sigma}\right).$$

Such an X is termed a **Gaussian** *or* **normal** *random variable. The mean is*

$$\mathbb{E}X = \int_{\mathbb{R}} x \gamma_{\sigma,m}(x) \, dx = m$$

and the variance is

$$\mathbb{E}(X - m)^2 = \int_{\mathbb{R}} (x - m)^2 \gamma_{\sigma,m}(x) \, dx = \sigma.$$

- *Let $m \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ be symmetric and positive definite. The random variable $X : \Omega \mapsto \mathbb{R}^d$ with pdf*

$$\gamma_{\Sigma,m}(x) := \Big((2\pi)^d det\Sigma\Big)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\langle \Sigma^{-1}(x - m), (x - m)\rangle\right)$$

is termed a **multivariate Gaussian** *or* **normal** *random variable. The mean is*

$$\mathbb{E}(X) = m \tag{2.3}$$

and the covariance matrix is

$$\mathbb{E}\Big((X - m) \otimes (X - m)\Big) = \Sigma. \tag{2.4}$$

- Since the mean and variance specify completely a Gaussian random variable on $\mathbb{R}$, the Gaussian is commonly denoted by $\mathcal{N}(m, \sigma)$. The **standard normal** random variable is $\mathcal{N}(0, 1)$.
- Since the mean and covariance matrix completely specify a Gaussian random variable on $\mathbb{R}^d$, the Gaussian is commonly denoted by $\mathcal{N}(m, \Sigma)$.

Example 2.2.8. *An exponential random variable $T : \Omega \to \mathbb{R}^+$ with rate $\lambda > 0$ satisfies*

$$\mathbb{P}(T > t) = e^{-\lambda t}, \quad \forall t \geqslant 0.$$

We write $T \sim \exp(\lambda)$. The related pdf is

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geqslant 0, \\ 0, & t < 0. \end{cases} \tag{2.5}$$

Notice that

$$\mathbb{E}\, T = \int_{-\infty}^{\infty} t f_T(t) dt = \frac{1}{\lambda} \int_0^{\infty} (\lambda t) e^{-\lambda t} d(\lambda t) = \frac{1}{\lambda}.$$

If the times $\tau_n = t_{n+1} - t_n$ are i.i.d random variables with $\tau_0 \sim \exp(\lambda)$ then, for $t_0 = 0$,

$$t_n = \sum_{k=0}^{n-1} \tau_k$$

and it is possible to show that

$$\mathbb{P}(0 \leqslant t_k \leqslant t < t_{k+1}) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}. \tag{2.6}$$

Chapter 3

Basics of the theory of Stochastic Processes

3.1 Definition of a Stochastic Process

- Let T be an ordered set. A **stochastic process** is a collection of random variables $X = \{X_t; t \in T\}$ where, for each fixed $t \in T$, X_t is a random variable from $(\Omega, \mathcal{F})$ to $(E, \mathcal{G})$.
- The measurable space $\{\Omega, \mathcal{F}\}$ is called the **sample space**. The space $(E, \mathcal{G})$ is called the **state space** .
- In this course we will take the set T to be $[0, +\infty)$.
- The state space E will usually be $\mathbb{R}^d$ equipped with the σ–algebra of Borel sets.
- A stochastic process X may be viewed as a function of both $t \in T$ and $\omega \in \Omega$. We will sometimes write $X(t), X(t, \omega)$ or $X_t(\omega)$ instead of X_t. For a fixed sample point $\omega \in \Omega$, the function $X_t(\omega) : T \mapsto E$ is called a **sample path** (realization, trajectory) of the process X.
- The **finite dimensional distributions** (fdd) of a stochastic process are the distributions of the E^k–valued random variables $(X(t_1), X(t_2), \ldots, X(t_k))$ for arbitrary positive integer k and arbitrary times $t_i \in T$, $i \in \{1, \ldots, k\}$:
$$F(\mathbf{x}) = \mathbb{P}(X(t_i) \leqslant x_i,\ i = 1, \ldots, k)$$
with $\mathbf{x} = (x_1, \ldots, x_k)$.
- We will say that two processes X_t and Y_t are equivalent if they have same finite dimensional distributions.
- From experiments or numerical simulations we can only obtain information about the (fdd) of a process.

Definition 3.1.1. *A* **Gaussian process** *is a stochastic processes for which $E = \mathbb{R}^d$ and all the finite dimensional distributions are Gaussian*

$$\begin{aligned} F(\mathbf{x}) &= \mathbb{P}(X(t_i) \leqslant x_i,\ i = 1, \ldots, k) \\ &= (2\pi)^{-n/2} (det K_k)^{-1/2} \exp\left[-\frac{1}{2}\langle K_k^{-1}(x - \mu_k), x - \mu_k\rangle\right], \end{aligned}$$

for some vector μ_k and a symmetric positive definite matrix K_k.

- A Gaussian process $x(t)$ is characterized by its mean

$$m(t) := \mathbb{E}x(t)$$

and the covariance function

$$C(t,s) = \mathbb{E}\Big(\big(x(t) - m(t)\big) \otimes \big(x(s) - m(s)\big)\Big).$$

- Thus, the first two moments of a Gaussian process are sufficient for a complete characterization of the process.

3.2 Stationary Processes

3.2.1 Strictly Stationary Processes

Definition 3.2.1. *A stochastic process is called* **(strictly) stationary** *if all finite dimensional distributions are invariant under time translation: for any integer k and times $t_i \in T$, the distribution of $(X(t_1), X(t_2), \dots, X(t_k))$ is equal to that of $(X(s+t_1), X(s+t_2), \dots, X(s+t_k))$ for any s such that $s+t_i \in T$ for all $i \in \{1, \dots, k\}$. In other words,*

$$\begin{aligned}&\mathbb{P}(X_{t_1+t} \in A_1, X_{t_2+t} \in A_2 \dots X_{t_k+t} \in A_k)\\ &= \mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2 \dots X_{t_k} \in A_k), \ \ \forall t \in T.\end{aligned}$$

- Let X_t be a strictly stationary stochastic process with finite second moment (i.e. $X_t \in L^2$). The definition of strict stationarity implies that $\mathbb{E}X_t = \mu$, a constant, and $\mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t-s)$. Hence, a strictly stationary process with finite second moment is also stationary in the wide sense. The converse is not true.

Example 3.2.2. *1. A sequence $Y_0, Y_1, \dots$ of independent, identically distributed random variables with mean μ and variance σ^2 is a stationary process with $R(k) = \sigma^2\delta_{k0}$, $\sigma^2 = \mathbb{E}(Y_k)^2$. Let $Z_n = Y_k, k = 0, 1, \dots$. We have*

$$\frac{1}{N}\sum_{j=0}^{N-1} Z_j = \frac{1}{N}\sum_{j=0}^{N-1} Y_j \to \mathbb{E}Y_0 = \mu,$$

by the strong law of large numbers. Hence, we can calculate the mean (as well as higher order moments) using a single sample path of our stochastic process, provided that it is long enough ($N \gg 1$). This is an example of an **ergodic** *stationary processes.*

2. Let Z be a single random variable with mean μ and variance σ^2 and set $Z_j = Z$, $j = 0, 1, 2, \dots$. Then the sequence Z_0, Z_1, $Z_2, \dots$ is a stationary sequence with $R(k) = \sigma^2$. In this case we have that

$$\frac{1}{N}\sum_{j=0}^{N-1} Z_j = \frac{1}{N}\sum_{j=0}^{N-1} Z = Z,$$

which is independent of N and does not converge to the mean of the stochastic processes Z_n. This is an example of a non-ergodic processes.

We will see in the next section that, for stationary processes, ergodicity is related to fast dacay of correlations. In the first of the examples above, there was no correlation between our stochastic processes at different times. On the contrary, in our second example there was very strong correlation between the stochastic process at different times (in fact the stochastic process is given by the same random variable at all times).

3.2.2 Second Order Stationary Processes

Let $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ be a probability space. Let X_t, $t \in T$ (with $T = \mathbb{R}$ or $\mathbb{Z}$) be a real-valued random process on this probability space with finite second moment, $\mathbb{E}|X_t|^2 < +\infty$ (i.e. $X_t \in L^2$).

Definition 3.2.3. *A stochastic process $X_t \in L^2$ is called* **second-order stationary** *or* **wide-sense stationary** *if the first moment $\mathbb{E}X_t$ is a constant and the covariance function $\mathbb{E}(X_t - \mu)(X_s - \mu)$ depends only on the difference $t - s$:*

$$\mathbb{E}X_t = \mu, \quad \mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t - s).$$

The constant μ is the **expectation** of the process X_t. Without loss of generality, we can set $\mu = 0$, since if $\mathbb{E}X_t = \mu$ then the process $Y_t = X_t - \mu$ is mean zero. A mean zero process with be called a **centered** process. The function $C(t)$ is the **covariance** or the **autocorrelation function** of the X_t. Notice that $C(t) = \mathbb{E}(X_t X_0)$, whereas $C(0) = \mathbb{E}(X_t^2)$, which is finite, by assumption. Since we have assumed that X_t is a real valued process, we have that $C(t) = C(-t)$, $t \in \mathbb{R}$.

Remark 3.2.4. *The first two moments of a Gaussian process are sufficient for a complete characterization of the process. A corollary of this is that a second order stationary Gaussian process is also a (strictly) stationary process.*

Continuity properties of the covariance function are equivalent to continuity properties of the paths of X_t in the L^2 sense, i.e.

$$\lim_{h \to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0.$$

Lemma 3.2.5. *Assume that the covariance function $C(t)$ of a second order stationary process is continuous at $t = 0$. Then it is continuous for all $t \in \mathbb{R}$. Furthermore, the continuity of $C(t)$ is equivalent to the continuity of the process X_t in the L^2-sense.*

Proof. Fix $t \in \mathbb{R}$. We calculate:

$$\begin{aligned} |C(t+h) - C(t)|^2 &= |\mathbb{E}(X_{t+h}X_0) - \mathbb{E}(X_t X_0)|^2 = \mathbb{E}|((X_{t+h} - X_t)X_0)|^2 \\ &\leqslant \mathbb{E}(X_0)^2 \mathbb{E}(X_{t+h} - X_t)^2 \\ &= C(0)(\mathbb{E}X_{t+h}^2 + \mathbb{E}X_t^2 - 2\mathbb{E}X_t X_{t+h}) \\ &= 2C(0)(C(0) - C(h)) \to 0, \end{aligned}$$

as $h \to 0$.

Assume now that $C(t)$ is continuous. From the above calculation we have

$$\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \tag{3.1}$$

which converges to 0 as $h \to 0$. Conversely, assume that X_t is L^2-continuous. Then, from the above equation we get $\lim_{h\to 0} C(h) = C(0)$. □

Notice that form (3.1) we immediately conclude that $C(0) > C(h)$, $h \in \mathbb{R}$.

The Fourier transform of the covariance function of a second order stationary process always exists. This enables us to study second order stationary processes using tools from Fourier analysis. To make the link between second order stationary processes and Fourier analysis we will use Bochner's theorem, which applies to all nonnegative functions.

Definition 3.2.6. *A function $f(x) : \mathbb{R} \mapsto \mathbb{R}$ is called nonnegative definite if*

$$\sum_{i,j=1}^{n} f(t_i - t_j) c_i \bar{c}_j \geqslant 0 \tag{3.2}$$

for all $n \in \mathbb{N}$, $t_1, \dots t_n \in \mathbb{R}$, $c_1, \dots c_n \in \mathbb{C}$.

Lemma 3.2.7. *The covariance function of second order stationary process is a nonnegative definite function.*

Proof. We will use the notation $X_t^c := \sum_{i=1}^n X_{t_i} c_i$. We have.

$$\begin{aligned}
\sum_{i,j=1}^{n} C(t_i - t_j) c_i \bar{c}_j &= \sum_{i,j=1}^{n} \mathbb{E} X_{t_i} X_{t_j} c_i \bar{c}_j \\
&= \mathbb{E}\left(\sum_{i=1}^{n} X_{t_i} c_i \sum_{j=1}^{n} X_{t_j} \bar{c}_j \right) = \mathbb{E}\left(X_t^c \bar{X}_t^c \right) \\
&= \mathbb{E}|X_t^c|^2 \geqslant 0.
\end{aligned}$$

□

Theorem 3.2.8. ***(Bochner)*** *There is a one-to-one correspondence between the set of continuous nonnegative definite functions and the set of finite measures on the Borel σ-algebra of $\mathbb{R}$: if ρ is a finite measure, then*

$$C(t) = \int_{\mathbb{R}} e^{ixt} \, \rho(dx) \tag{3.3}$$

in nonnegative definite. Conversely, any nonnegative definite function can be represented in the form (3.3)*.*

Definition 3.2.9. *Let X_t be a second order stationary process with covariance $C(t)$ whose Fourier transform is the measure $\rho(dx)$. The measure $\rho(dx)$ is called the* **spectral measure** *of the process X_t.*

In the following we will assume that the spectral measure is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ with density $f(x)$, $d\rho(x) = f(x)dx$. The Fourier transform $f(x)$ of the covariance function is called the **spectral density** of the process:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} C(t) \, dt.$$

From (3.3) it follows that that the covariance function of a mean zero, second order stationary process is given by the inverse Fourier transform of the spectral density:

$$C(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx.$$

There are various cases where the experimentally measured quantity is the spectral density (or power spectrum) of the stochastic process.

Remark 3.2.10. *The correlation function of a second order stationary process enables us to associate a time scale to X_t, the* **correlation time** τ_{cor}*:*

$$\tau_{cor} = \frac{1}{C(0)} \int_0^{\infty} C(\tau)\, d\tau = \int_0^{\infty} \mathbb{E}(X_\tau X_0)/\mathbb{E}(X_0^2)\, d\tau.$$

The slower the decay of the correlation function, the larger the correlation time is. Of course, we have to assume sufficiently fast decay of correlations so that the correlation time is finite.

Example 3.2.11.

- *Consider the mean zero, second order stationary process with covariance function*

$$R(t) = \frac{D}{\alpha} e^{-\alpha|t|}. \tag{3.4}$$

- *The spectral density of this process is:*

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\frac{D}{\alpha} \int_{-\infty}^{+\infty} e^{-ixt} e^{-\alpha|t|}\, dt \\ &= \frac{1}{2\pi}\frac{D}{\alpha} \left(\int_{-\infty}^{0} e^{-ixt} e^{\alpha t}\, dt + \int_0^{+\infty} e^{-ixt} e^{-\alpha t}\, dt \right) \\ &= \frac{1}{2\pi}\frac{D}{\alpha} \left(\frac{1}{ix+\alpha} + \frac{1}{-ix+\alpha} \right) \\ &= \frac{D}{\pi} \frac{1}{x^2+\alpha^2}. \end{aligned}$$

- *This function is called the* **Cauchy** *or the* **Lorentz** *distribution.*

- *The correlation time is (we have that $R(0) = D/\alpha$)*

$$\tau_{cor} = \int_0^{\infty} e^{-\alpha t}\, dt = \alpha^{-1}.$$

A Gaussian process with an exponential correlation function is of particular importance in the theory and applications of stochastic processes.

Definition 3.2.12. *A Gaussian stochastic process with mean zero and covariance function* (3.4) *is called the* **stationary Ornstein-Uhlenbeck process***.*

The OU process was introduced by Ornstein and Uhlenbeck in 1930 (G.E. Uhlenbeck, L.S. Ornstein, Phys. Rev. 36, 823 (1930)) as a model for the velocity of a Brownian particle. It is of interest to calculate the statistics of the position of the Brownian particle, i.e. of the integral

$$X(t) = \int_0^t Y(s)\, ds, \tag{3.5}$$

where $Y(t)$ denotes the stationary OU process.

Lemma 3.2.13. *Let $Y(t)$ denote the stationary OU process with covariance function* (3.4)*. Then the position process* (3.5) *is a mean zero Gaussian process with covariance function*

$$\mathbb{E}(X(t)X(s)) = 2\min(t,s) + e^{-\min(t,s)} + e^{-max(t,s)} - e^{-|t-s|} - 1.$$

Exercise 3.2.14. *Prove Lemma* (3.2.13)*.*

3.2.3 Ergodic Properties of Stationary Processes

Second order stationary processes are ergodic: time averages equal phase space (ensemble) averages. An example of an ergodic theorem for a stationary processes is the following L^2 (mean-square) ergodic theorem.

Theorem 3.2.15. *Let $\{X_t\}_{t\geqslant 0}$ be a second order stationary process on a probability space Ω, $\mathcal{F}$, $\mathbb{P}$ with mean μ and covariance $R(t)$, and assume that $R(t) \in L^1(0,+\infty)$. Then*

$$\lim_{T\to+\infty} \mathbb{E}\left|\frac{1}{T}\int_0^T X(s)\, ds - \mu\right|^2 = 0. \tag{3.6}$$

For the proof of this result we will first need an elementary lemma.

Lemma 3.2.16. *Let $R(t)$ be an integrable symmetric function. Then*

$$\int_0^T \int_0^T R(t-s)\, dtds = 2\int_0^T (T-s)R(s)\, ds. \tag{3.7}$$

Proof. We make the change of variables $u = t - s$, $v = t + s$. The domain of integration in the t, s variables is $[0,T]\times[0,T]$. In the u, v variables it becomes $[-T,T]\times[0,2(T-|u|)]$. The Jacobian of the transformation is

$$J = \frac{\partial(t,s)}{\partial(u,v)} = \frac{1}{2}.$$

The integral becomes

$$\begin{aligned}\int_0^T \int_0^T R(t-s)\, dtds &= \int_{-T}^T \int_0^{2(T-|u|)} R(u) J\, dvdu \\ &= \int_{-T}^T (T-|u|)R(u)\, du \\ &= 2\int_0^T (T-u)R(u)\, du,\end{aligned}$$

where the symmetry of the function $R(u)$ was used in the last step. □

Proof of Theorem 3.2.15. We use Lemma (3.2.16) to calculate:

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T}\int_0^T X_s\,ds-\mu\right|^2 &= \frac{1}{T^2}\mathbb{E}\left|\int_0^T (X_s-\mu)\,ds\right|^2 \\
&= \frac{1}{T^2}\mathbb{E}\int_0^T\int_0^T (R(t)-\mu)(R(s)-\mu)\,dtds \\
&= \frac{1}{T^2}\int_0^T\int_0^T R(t-s)\,dtds \\
&= \frac{2}{T^2}\int_0^T (T-u)R(u)\,du \\
&\leqslant \frac{2}{T}\int_0^{+\infty}\left(1-\frac{u}{T}\right)R(u) \leqslant \frac{2}{T}\int_0^{+\infty} R(t)\,dt \to 0,
\end{aligned}$$

using the dominated convergence theorem and the assumption $R(\cdot)\in L^1$. □

From the above calculation we conclude that

$$\lim_{T\to+\infty}\int_0^T\left(1-\frac{u}{T}\right)R(u)\,du=\int_0^\infty R(t)\,dt=:D,$$

which is a finite quantity. Assume that $\mu=0$. The above calculation suggests that

$$\lim_{T\to+\infty}\frac{1}{T}\mathbb{E}\left(\int_0^t X(t)\,dt\right)^2=2D.$$

Hence, for $T\gg 1$, we have that

$$\mathbb{E}\left(\int_0^t X(t)\,dt\right)^2\approx 2DT.$$

This implies that, at sufficiently long times, the mean square displacement of the integral of the ergodic stationary process X_t scales linearly in time, with proportionality coefficient $2D$.

Assume that X_t is the velocity of a (Brownian) particle. In this case, the integral of X_t

$$Z_t=\int_0^t X_s\,ds,$$

represents the particle position. From our calculation above we conclude that

$$\mathbb{E}Z_t^2=2Dt.$$

where

$$D=\int_0^\infty R(t)\,dt=\int_0^\infty \mathbb{E}(X_tX_0)\,dt \tag{3.8}$$

is the **diffusion coefficient**. Thus, one expects that at sufficiently long times and under appropriate assumptions on the correlation function, the time integral of a stationary process will approximate a Brownian motion with diffusion coefficient D. The diffusion coefficient is an example of a **transport coefficient** and (3.8) is an example of the **Green-Kubo formula**: a transport coefficient can be calculated in terms of the time integral of an appropriate autocorrelation function. In the case of the diffusion coefficient we need to calculate the integral of the velocity autocorrelation function.

Example 3.2.17. *Consider the stochastic processes with an exponential correlation function from Example 3.2.11, and assume that this stochastic process describes the velocity of a Brownian particle. Since* $R(t) \in L^1(0, +\infty)$ *Theorem 3.2.15 applies. Furthermore, the diffusion coefficient of the Brownian particle is given by*

$$\int_0^{+\infty} R(t)\, dt = R(0)\tau_c^{-1} = \frac{D}{\alpha^2}.$$

3.3 Brownian Motion

- The most important continuous time stochastic process is **Brownian motion**. Brownian motion is a mean zero, continuous (i.e. it has continuous sample paths: for a.e $\omega \in \Omega$ the function X_t is a continuous function of time) process with independent Gaussian increments.

- A process X_t has **independent increments** if for every sequence $t_0 < t_1 < \dots t_n$ the random variables

  $$X_{t_1} - X_{t_0},\, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$$

  are independent.

- If, furthermore, for any $t_1,\, t_2,\, s \in T$ and Borel set $B \subset \mathbb{R}$

  $$\mathbb{P}(X_{t_2+s} - X_{t_1+s} \in B) = \mathbb{P}(X_{t_2} - X_{t_1} \in B)$$

  then the process X_t has **stationary independent increments**.

Definition 3.3.1. • *A one dimensional standard* Brownian motion $W(t) : \mathbb{R}^+ \to \mathbb{R}$ *is a real valued stochastic process such that*

i. $W(0) = 0$.

ii. $W(t)$ *has independent increments.*

iii. For every $t > s \geqslant 0$ $W(t) - W(s)$ *has a Gaussian distribution with mean* 0 *and variance* $t - s$. *That is, the density of the random variable* $W(t) - W(s)$ *is*

$$g(x; t, s) = \Big(2\pi(t-s)\Big)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2(t-s)}\right); \tag{3.9}$$

- *A* d*–dimensional standard Brownian motion* $W(t) : \mathbb{R}^+ \to \mathbb{R}^d$ *is a collection of* d *independent one dimensional Brownian motions:*

  $$W(t) = (W_1(t), \dots, W_d(t)),$$

  where $W_i(t), i = 1, \dots, d$ *are independent one dimensional Brownian motions. The density of the Gaussian random vector* $W(t) - W(s)$ *is thus*

  $$g(x; t, s) = \Big(2\pi(t-s)\Big)^{-d/2} \exp\left(-\frac{\|x\|^2}{2(t-s)}\right).$$

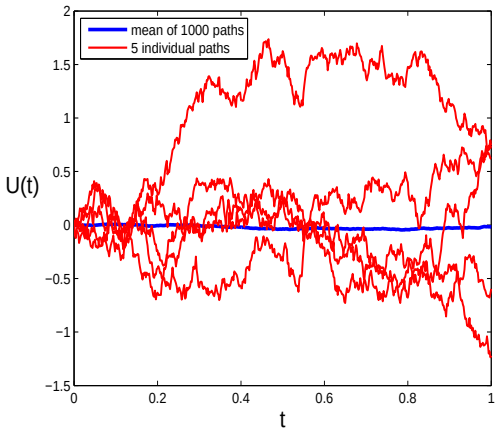


Figure 3.1: Brownian sample paths

Brownian motion is sometimes referred to as the **Wiener process** .

Brownian motion has continuous paths. More precisely:

Definition 3.3.2. *Let X_t and Y_t, $t \in T$, be two stochastic processes defined on the same probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$. The process Y_t is said to be a modification of X_t if $\mathbb{P}(X_t = Y_t) = 1\ \forall t \in T$. Then there is a modification of Brownian motion such that all paths are continuous.*

Lemma 3.3.3. *There is a continuous modification of Brownian motion.*

This follows from a theorem due to Kolmogorov.

Theorem 3.3.4. *(Kolmogorov) Let X_t, $t \in [0, \infty)$ be a stochastic process on a probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$. Suppose that there are positive constants α and β, and for each $T \geqslant 0$ there is a constant $C(T)$ such that*

$$\mathbb{E}|X_t - X_s|^\alpha \leqslant C(T)|t - s|^{1+\beta}, \quad 0 \leqslant s, t \leqslant T. \tag{3.10}$$

There there exists a continuous modification Y_t of the process X_t.

Exercise 3.3.5. *Use Theorem* (3.3.4) *to show that Brownian motion has a continuous modification.*

Remark 3.3.6. *Equivalently, we could have defined the one dimensional standard Brownian motion as a stochastic process on a probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ with continuous paths for almost all $\omega \in \Omega$, and Gaussian finite dimensional distributions with zero mean and covariance $\mathbb{E}(W_{t_i} W_{t_j}) = \min(t_i, t_j)$. One can then show that Definition 3.3.1 and continuity of paths are equivalent to the above definition.*

It is possible to prove rigorously the existence of the Wiener process (Brownian motion):

Theorem 3.3.7. *(Wiener) There exists an almost-surely continuous process W_t with independent increments such and $W_0 = 0$, such that for each $t \geqslant$ the random variable W_t is $\mathcal{N}(0, t)$. Furthermore, W_t is almost surely locally Hölder continuous with exponent α for any $\alpha \in (0, \frac{1}{2})$.*

Notice that Brownian paths are not differentiable.

We can also construct Brownian motion through the limit of the random walk: let $X_1,\, X_2, \dots$ be iid random variables on a probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ with mean 0 and variance 1. Define the discrete time stochastic process S_n with $S_0 = 0$, $S_n = \sum_{j=1} X_j^n$, $n \geqslant n$. Define not a continuous time stochastic process with continuous paths as the linearly interpolated, appropriately rescaled random walk:

$$W_t^n = \frac{1}{\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sqrt{n}} X_{[nt]+1},$$

where $[\cdot]$ denotes the integer part of a number. Then W_t^n converges weakly to a one dimensional standard Brownian motion.

Brownian motion is a Gaussian process. For the d–dimensional Brownian motion, and for I the $d \times d$ dimensional identity, we have (see (2.3) and (2.4))

$$\mathbb{E}W(t) = 0 \quad \forall t \geqslant 0$$

and

$$\mathbb{E}\Big((W(t) - W(s)) \otimes (W(t) - W(s)) \Big) = (t-s)I. \tag{3.11}$$

Moreover,

$$\mathbb{E}\Big(W(t) \otimes W(s) \Big) = \min(t,s) I. \tag{3.12}$$

From the formula for the Gaussian density $g(x, t-s)$, eqn. (3.9), we immediately conclude that $W(t) - W(s)$ and $W(t+u) - W(s+u)$ have the same pdf. Consequently, Brownian motion has stationary increments. Notice, however, that Brownian motion itself **is not** a stationary process. Since $W(t) = W(t) - W(0)$, the pdf of $W(t)$ is

$$g(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.$$

We can easily calculate all moments of the Brownian motion:

$$\begin{aligned} \mathbb{E}(x^n(t)) &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2t}\, dx \\ &= \begin{cases} 1.3 \dots (n-1) t^{n/2}, & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases} \end{aligned}$$

Exercise 3.3.8. *Let $a_1, \dots a_n$ and $s_1, \dots s_n$ be positive real numbers. Calculate the mean and variance of the random variable*

$$X = \sum_{i=1}^{n} a_i W(s_i).$$

Exercise 3.3.9. *Let $W(t)$ be the standard one-dimensional Brownian motion and let $\sigma,\, s_1,\, s_2 > 0$. Calculate*

i. $\mathbb{E}e^{\sigma W(t)}$.

ii. $\mathbb{E}\big(\sin(\sigma W(s_1)) \sin(\sigma W(s_2)) \big)$.

Notice that if W_t is a standard Brownian motion, then so is the rescaled processes

$$X_t = \frac{1}{\sqrt{c}} W_{ct}.$$

To prove this we show that the two processes have the same distribution function:

$$\begin{aligned}
\mathbb{P}\left(X_t \leqslant x\right) &= \mathbb{P}\left(W_{ct} \leqslant \sqrt{c}x\right) \\
&= \int_{-\infty}^{\sqrt{c}x} \frac{1}{\sqrt{2\pi ct}} \exp\left(-\frac{z^2}{2ct}\right) dz \\
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi ct}} \exp\left(-\frac{y^2}{2t}\right) \sqrt{c}\, dy \\
&= \mathbb{P}\left(W_t \leqslant x\right).
\end{aligned}$$

We can also add a drift and change the diffusion coefficient of the Brownian motion: we will define a Brownian motion with drift μ and variance σ^2 as the process

$$X_t = \mu t + \sigma W_t.$$

Notice that

$$\mathbb{E}X_t = \mu t, \quad \mathbb{E}(X_t - \mathbb{E}X_t)^2 = \sigma^2 t.$$

We can define the OU process through the Brownian motion via a time change.

Lemma 3.3.10. *Let $W(t)$ be a standard Brownian motion and consider the process*

$$V(t) = e^{-t} W(e^{2t}).$$

Then $V(t)$ is a Gaussian second order stationary process with mean 0 and covariance

$$R(s,t) = e^{-|t-s|}. \tag{3.13}$$

For the proof of this result we first need to show that time changed Gaussian processes are also Gaussian.

Lemma 3.3.11. *Let $X(t)$ be a Gaussian stochastic process and let $Y(t) = X(f(t))$ where $f(t)$ is a strictly increasing function. Then $Y(t)$ is also a Gaussian process.*

Proof. We need to show that, for all positive integers N and all sequences of times $\{t_1, t_2, \dots t_N\}$ the random vector

$$\{Y(t_1), Y(t_2), \dots Y(t_N)\} \tag{3.14}$$

is has is a multivariate Gaussian random variable. Since $f(t)$ is strictly increasing, it is invertible and hence, there exist s_i, $i = 1, \dots N$ such that $s_i = f^{-1}(t_i)$. Thus, the random vector (3.14) can be rewritten as

$$\{X(s_1), X(s_2), \dots X(s_N)\},$$

which is Gaussian for all N and all choices of times $s_1, s_2, \dots s_N$. Hence $Y(t)$ is also Gaussian. □

Proof of Lemma 3.3.10. The fact that $V(t)$ is mean zero follows immediately from the fact that $W(t)$ is mean zero. To show that the correlation function of $V(t)$ is given by (3.13), we calculate

$$\begin{aligned}\mathbb{E}(V(t)V(s)) &= e^{-t-s}\mathbb{E}(W(e^{2t})W(e^{2s})) = e^{-t-s}\min(e^{2t}, e^{2s}) \\ &= e^{-|t-s|}.\end{aligned}$$

The Gaussianity of the $V(t)$ follows from Lemma 3.3.11 (notice that the transformation that gives $V(t)$ in terms of $W(t)$ is invertible and we can write $W(s) = s^{1/2}V(\frac{1}{2}\ln(s))$). □

Exercise 3.3.12. *Use Lemma 3.3.10 to obtain the probability density function of the stationary Ornstein-Uhlenbeck process.*

Exercise 3.3.13. *Calculate the mean and the correlation function of the integral of a standard Brownian motion*

$$Y_t = \int_0^t W_s\, ds.$$

Exercise 3.3.14. *Show that the process*

$$Y_t = \int_t^{t+1} (W_s - W_t)\, ds, \quad t \in \mathbb{R},$$

is second order stationary.

3.4 Other Examples of Stochastic Processes

3.4.1 Brownian Bridge

Let $W(t)$ be a standard one dimensional Brownian motion. We define the **Brownian bridge** (from 0 to 0) to be the process

$$B_t = W_t - tW_1, \quad t \in [0, 1]. \tag{3.15}$$

Notice that $B_0 = B_1 = 0$. Equivalently, we can define the Brownian bridge to be the continuous Gaussian process $\{B_t : 0 \leqslant t \leqslant 1\}$ such that

$$\mathbb{E}B_t = 0, \quad \mathbb{E}((B_t - \mathbb{E}B_t)(B_s - \mathbb{E}B_s)) = \min(s, t) - st, \quad s, t \in [0, 1]. \tag{3.16}$$

Another, equivalent definition of the Brownian bridge is through an appropriate time change of the Brownian motion:

$$B_t = t(1-t)W\left(\frac{t}{1-t}\right), \quad t \in [0, 1). \tag{3.17}$$

Conversely, we can write the Brownian motion as a time change of the Brownian bridge:

$$W_t = (t+1)B\left(\frac{t}{1+t}\right), \quad t \geqslant 0.$$

Exercise 3.4.1. *Show the equivalence between definitions* (3.15), (3.16) *and* (3.17).

Exercise 3.4.2. *Use* (3.15) *to show that the probability density function of the Brownian bridge is*

$$f(x,t) = (2\pi(1-t))^{1/2} \exp\left(-\frac{|x-1|^2}{2(1-t)}\right).$$

Exercise 3.4.3. *Give the definition of the Brownian Bridge from* 0 *to* a*, where* $a \in \mathbb{R}$*. Calculate the mean and covariance of this Brownian bridge.*

3.4.2 Fractional Brownian Motion

Definition 3.4.4. *A (normalized) fractional Brownian motion* W_t^H, $t \geqslant 0$ *with Hurst parameter* $H \in (0,1)$ *is a centered Gaussian process with continuous sample paths whose covariance is given by*

$$\mathbb{E}(W_t^H W_s^H) = \frac{1}{2}\big(s^{2H} + t^{2H} - |t-s|^{2H}\big). \tag{3.18}$$

Fractional Brownian motion has the following properties.

i. When $H = \frac{1}{2}$, $W_t^{\frac{1}{2}}$ *becomes the standard Brownian motion.*

ii. $W_0^H = 0$, $\mathbb{E}W_t^H = 0$, $\mathbb{E}(W_t^H)^2 = |t|^{2H}$, $t \geqslant 0$.

iii. It has stationary increments, $\mathbb{E}(W_t^H - W_s^H)^2 = |t-s|^{2H}$.

iv. It has the following self similarity property

$$(W_{\alpha t}^H,\, t \geqslant 0) = (\alpha^H W_t^H,\, t \geqslant 0), \quad \alpha > 0,$$

where the equivalence is in law.

3.4.3 The Poisson Process

- Another fundamental continuous time process is the **Poisson process** :

  Definition 3.4.5. *The Poisson process with intensity* λ*, denoted by* $N(t)$*, is an integer-valued, continuous time, stochastic process with independent increments satisfying*

$$\mathbb{P}[(N(t) - N(s)) = k] = \frac{e^{-\lambda(t-s)}\big(\lambda(t-s)\big)^k}{k!}, \quad t > s \geqslant 0,\, k \in \mathbb{N}.$$

- Both Brownian motion and the Poisson process are **homogeneous** (or **time-homogeneous**): the increments between successive times s and t depend only on $t-s$.

Exercise 3.4.6. *Use Theorem* (3.3.4) *to show that there does not exist a continuous modification of the Poisson process.*

3.5 The Karhunen-Loéve Expansion

Let $f \in L^2(\Omega)$ where Ω is a subset of $\mathbb{R}^d$ and let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis in $L^2(\Omega)$. Then, it is well known that f can be written as a series expansion:

$$f = \sum_{n=1}^{\infty} f_n e_n,$$

where

$$f_n = \int_{\Omega} f(x) e_n(x)\, dx.$$

The convergence is in $L^2(\Omega)$:

$$\lim_{N\to\infty} \|f(x) - \sum_{n=1}^{N} f_n e_n(x)\|_{L^2(\Omega)} = 0.$$

It turns out that we can obtain a similar expansion for an L^2 mean zero process which is continuous in the L^2 sense:

$$\mathbb{E}X_t^2 < +\infty, \quad \mathbb{E}X_t = 0, \quad \lim_{h\to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0. \tag{3.19}$$

For simplicity we will take $T = [0, 1]$. Let $R(t, s) = \mathbb{E}(X_t X_s)$ be the correlation function.

Exercise 3.5.1. *Show that the correlation function of a process X_t satisfying* (3.19) *is continuous in both t and s.*

Let us assume that an expansion of the form

$$X_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) e_n(t), \quad t \in [0, 1] \tag{3.20}$$

where $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis in $L^2([0, 1])$. The random variables ξ_n are calculated as

$$\begin{aligned} \int_0^1 X_t e_k(t)\, dt &= \int_0^1 \sum_{n=1}^{\infty} \xi_n e_n(t) e_k(t)\, dt \\ &= \sum_{n=1}^{\infty} \xi_n \delta_{nk} = \xi_k, \end{aligned}$$

where we assumed that we can interchange the summation and integration. We will assume that these random variables are orthogonal:

$$\mathbb{E}(\xi_n \xi_m) = \lambda_n \delta_{nm},$$

where $\{\lambda_n\}_{n=1}^{\infty}$ are positive numbers that will be determined later.

Assuming that an expansion of the form (3.20) exists, we can calculate

$$\begin{aligned} R(t,s) = \mathbb{E}(X_t X_s) &= \mathbb{E}\left(\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} \xi_k e_k(t)\xi_\ell e_\ell(s)\right) \\ &= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} \mathbb{E}\left(\xi_k\xi_\ell\right) e_k(t)e_\ell(s) \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t)e_k(s). \end{aligned}$$

Consequently, in order to the expansion (3.20) to be valid we need

$$R(t,s) = \sum_{k=1}^{\infty} \lambda_k e_k(t)e_k(s). \tag{3.21}$$

From equation (3.21) it follows that

$$\begin{aligned} \int_0^1 R(t,s)e_n(s)\,ds &= \int_0^1 \sum_{k=1}^{\infty} \lambda_k e_k(t)e_k(s)e_n(s)\,ds \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t) \int_0^1 e_k(s)e_n(s)\,ds \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t)\delta_{kn} \\ &= \lambda_n e_n(t). \end{aligned}$$

Hence, in order for the expansion (3.20) to be valid, $\{\lambda_n,\, e_n(t)\}_{n=1}^{\infty}$ have to be the eigenvalues and eigenfunctions of the integral operator whose kernel is the correlation function of X_t:

$$\int_0^1 R(t,s)e_n(s)\,ds = \lambda_n e_n(t). \tag{3.22}$$

Hence, in order to prove the expansion (3.20) we need to study the eigenvalue problem for the integral operator $\mathcal{R} : L^2[0,1] \mapsto L^2[0,1]$. It easy to check that this operator is **self-adjoint** ($(\mathcal{R}f, h) = (f, \mathcal{R}h)$ for all $f,\, h \in L^2(0,1)$) and **nonnegative** ($\mathcal{R}f, f \geqslant 0$ for all $f \in L^2(0,1)$). Hence, all its eigenvalues are real and nonnegative. Furthermore, it is a **compact** operator (if $\{\phi_n\}_{n=1}^{\infty}$ is a bounded sequence in $L^2(0,1)$, then $\{\mathcal{R}\phi_n\}_{n=1}^{\infty}$ has a weakly convergent subsequence). The spectral theorem for compact, self-adjoint operators implies that $\mathcal{R}$ has a countable sequence of eigenvalues tending to 0. Furthermore, for every $f \in L^2(0,1)$ we can write

$$f = f_0 + \sum_{n=1}^{\infty} f_n e_n(t),$$

where $\mathcal{R}f_0 = 0$, $\{e_n(t)\}$ are the eigenfunctions of $\mathcal{R}$ corresponding to nonzero eigenvalues and the convergence is in L^2. Finally, **Spencer's Theorem** states that for $R(t,s)$ continuous on $[0,1] \times [0,1]$, the expansion (3.21) is valid, where the series converges absolutely and uniformly.

Exercise 3.5.2. *Let X_t be a stochastic process satisfying* (3.19) *and $R(t,s)$ its correlation function. Show that the integral operator $\mathcal{R} : L^2[0,1] \mapsto L^2[0,1]$*

$$\mathcal{R}f := \int_0^1 R(t,s)f(s)\,ds \tag{3.23}$$

is self-adjoint and nonnegative. Show that all of its eigenvalues are real and nonnegative. Show that eigenfunctions corresponding to different eigenvalues are orthogonal.

Exercise 3.5.3. *Let H be a Hilbert space. An operator $\mathcal{R} : H \to H$ is said to be* **Hilbert-Schmidt** *if there exists a complete orthonormal sequence $\{\phi_n\}_{n=1}^{\infty}$ in H such that*

$$\sum_{n=1}^{\infty} \|\mathcal{R}e_n\|^2 < \infty.$$

Let $\mathcal{R} : L^2[0,1] \mapsto L^2[0,1]$ be the operator defined in (3.23) *with $R(t,s)$ being continuous both in t and s. Show that it is a Hilbert-Schmidt operator.*

Now we are ready to prove (3.20).

Theorem 3.5.4. *(Karhunen-Loéve). Let $\{X_t,\, t \in [0,1]\}$ be an L^2 process with zero mean and continuous correlation function $R(t,s)$. Let $\{\lambda_n,\, e_n(t)\}_{n=1}^{\infty}$ be the eigenvalues and eigenfunctions of the operator $\mathcal{R}$ defined in* (3.23)*. Then*

$$X_t = \sum_{n=1}^{\infty} \xi_n e_n(t), \quad t \in [0,1], \tag{3.24}$$

where

$$\xi_n = \int_0^1 X_t e_n(t)\,dt, \quad \mathbb{E}\xi_n = 0, \quad \mathbb{E}(\xi_n\xi_m) = \lambda\delta_{nm}. \tag{3.25}$$

The series converges in L^2 to $X(t)$, uniformly in t.

Proof. The fact that $\mathbb{E}\xi_n = 0$ follows from the fact that X_t is mean zero. The orthogonality of the random variables $\{\xi_n\}_{n=1}^{\infty}$ follows from the orthogonality of the eigenfunctions of $\mathcal{R}$:

$$\begin{aligned}
\mathbb{E}(\xi_n\xi_m) &= \mathbb{E}\int_0^1\int_0^1 X_tX_s e_n(t)e_m(s)\,dtds \\
&= \int_0^1\int_0^1 R(t,s)e_n(t)e_m(s)\,dsdt \\
&= \lambda_n\int_0^1 e_n(s)e_m(s)\,ds \\
&= \lambda_n\delta_{nm}.
\end{aligned}$$

Consider now the partial sum $S_N = \sum_{n=1}^{N} \xi_n e_n(t)$.

$$\begin{aligned}
\mathbb{E}|X_t - S_N|^2 &= \mathbb{E}X_t^2 + \mathbb{E}S_N^2 - 2\mathbb{E}(X_t S_N) \\
&= R(t,t) + \mathbb{E}\sum_{k,\ell=1}^{N} \xi_k \xi_\ell e_k(t) e_\ell(t) - 2\mathbb{E}\left(X_t \sum_{n=1}^{N} \xi_n e_n(t) \right) \\
&= R(t,t) + \sum_{k=1}^{N} \lambda_k |e_k(t)|^2 - 2\mathbb{E}\sum_{k=1}^{N} \int_0^1 X_t X_s e_k(s) e_k(t)\, ds \\
&= R(t,t) - \sum_{k=1}^{N} \lambda_k |e_k(t)|^2 \to 0,
\end{aligned}$$

by Spencer's theorem. □

Exercise 3.5.5. *Let X_t a mean zero second order stationary process with continuous covariance $R(t)$. Show that*

$$\sum_{n=1}^{\infty} \lambda_n = T\, R(0).$$

Remark 3.5.6. *Let X_t be a Gaussian second order process with continuous covariance $R(t,s)$. Then the random variables $\{\xi_k\}_{k=1}^{\infty}$ are Gaussian, since they are defined through the time integral of a Gaussian processes. Furthermore, since they are Gaussian and orthogonal, they are also independent. Hence, for Gaussian processes the Karhunen-Loéve expansion becomes:*

$$X_t = \sum_{k=1}^{+\infty} \sqrt{\lambda_k} \xi_k e_k(t), \tag{3.26}$$

where $\{\xi_k\}_{k=1}^{\infty}$ are independent $\mathcal{N}(0,1)$ random variables.

Example 3.5.7. The Karhunen-Loéve Expansion for Brownian Motion. *The correlation function of Brownian motion is $R(t,s) = \min(t,s)$. The eigenvalue problem $\mathcal{R}\psi_n = \lambda_n \psi_n$ becomes*

$$\int_0^1 \min(t,s)\psi_n(s)\, ds = \lambda_n \psi_n(t).$$

Let us assume that $\lambda_n > 0$ (it is easy to check that 0 is not an eigenvalue). Upon setting $t = 0$ we obtain $\psi_n(0) = 0$. The eigenvalue problem can be rewritten in the form

$$\int_0^t s\psi_n(s)\, ds + t\int_t^1 \psi_n(s)\, ds = \lambda_n \psi_n(t).$$

We differentiate this equation once:

$$\int_t^1 \psi_n(s)\, ds = \lambda_n \psi_n'(t).$$

We set $t = 1$ in this equation to obtain the second boundary condition $\psi_n'(1) = 0$. A second differentiation yields;

$$-\psi_n(t) = \lambda_n \psi_n''(t),$$

where primes denote differentiation with respect to t. Thus, in order to calculate the eigenvalues and eigenfunctions of the integral operator whose kernel is the covariance function of Brownian motion, we need to solve the Sturm-Liouville problem

$$-\psi_n(t) = \lambda_n \psi_n''(t), \quad \psi(0) = \psi'(1) = 0.$$

It is easy to check that the eigenvalues and (normalized) eigenfunctions are

$$\psi_n(t) = \sqrt{2}\sin\left(\frac{1}{2}(2n+1)\pi t\right), \quad \lambda_n = \left(\frac{2}{(2n+1)\pi}\right)^2.$$

Thus, the Karhunen-Loéve expansion of Brownian motion on $[0,1]$ is

$$W_t = \sqrt{2}\sum_{n=1}^{\infty} \xi_n \frac{2}{(2n+1)\pi}\sin\left(\frac{1}{2}(2n+1)\pi t\right). \tag{3.27}$$

Exercise 3.5.8. *Calculate the Karhunen-Loeve expansion for a second order stochastic process with correlation function $R(t,s) = ts$.*

Exercise 3.5.9. *Calculate the Karhunen-Loeve expansion of the Brownian bridge on $[0,1]$.*

Exercise 3.5.10. *Let X_t, $t \in [0,T]$ be a second order process with continuous covariance and Karhunen-Loéve expansion*

$$X_t = \sum_{k=1}^{\infty} \xi_k e_k(t).$$

Define the process

$$Y(t) = f(t)X_{\tau(t)}, \quad t \in [0,S],$$

where $f(t)$ is a continuous function and $\tau(t)$ a continuous, nondecreasing function with $\tau(0) = 0$, $\tau(S) = T$. Find the Karhunen-Loéve expansion of $Y(t)$, in terms of the KL expansion of X_t. Use this in order to calculate the KL expansion of the Ornstein-Uhlenbeck process.

Exercise 3.5.11. *Calculate the Karhunen-Loéve expansion of a centered Gaussian stochastic process with covariance function $R(s,t) = \cos(2\pi(t-s))$.*

3.6 The Path Space

- Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, (E,ρ) a metric space and let $T = [0,\infty)$. Let $\{X_t\}$ be a stochastic process from $(\Omega,\mathcal{F},\mu)$ to (E,ρ) with continuous sample paths.
- The above means that for every $\omega \in \Omega$ we have that $X_t \in C_E := C([0,\infty); E)$.
- The space of continuous functions C_E is called the *path space* of the stochastic process.

- We can put a metric on E as follows:

$$\rho_E(X^1, X^2) := \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{0 \leqslant t \leqslant n} \min\big(\rho(X_t^1, X_t^2), 1\big).$$

- We can then define the Borel sets on C_E, using the topology induced by this metric, and $\{X_t\}$ can be thought of as a random variable on $(\Omega, \mathcal{F}, \mu)$ with state space $(C_E, \mathcal{B}(C_E))$.

- The probability measure $\mathbb{P}X_t^{-1}$ on $(C_E, \mathcal{B}(C_E))$ is called the *law* of $\{X_t\}$.

- The law of a stochastic process is a probability measure on its path space.

  Example 3.6.1. *The space of continuous functions C_E is the path space of Brownian motion (the Wiener process). The law of Brownian motion, that is the measure that it induces on $C([0, \infty), \mathbb{R}^d)$, is known as the* **Wiener measure**.

3.7 Discussion and Bibliography

The kind of analysis presented in Section 3.2.3 was initiated by G.I. Taylor in [25]. The spectral theorem for compact, self-adjoint operators which was needed in the proof of the Karhunen-Loéve theorem can be found in [19]. A similar result is valid for **random fields**. See [23] and the reference therein.

Chapter 4

Markov Processes

4.1 Introduction and Examples

In this chapter we will study some of the basic properties of Markov stochastic processes. Roughly speaking, a Markov process is a stochastic process that retains no memory of where it has been in the past: only the current state of a Markov process can influence where it will go next. A bit more precisely: a Markov process is a stochastic process for which, given the present, past and future are statistically independent.

Perhaps the simplest example of a Markov process is that of a random walk in one dimension. We defined the one dimensional random walk as the sum of independent, mean zero and variance 1 random variables ξ_i, $i = 1, \dots$:

$$X_N = \sum_{n=1}^{N} \xi_n, \quad X_0 = 0.$$

Let $i_1, \dots i_2, \dots$ be a sequence of integers. Then, clearly

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_1 = i_1, \dots X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) \tag{4.1}$$

In words, the probability that the random walk will be at i_{n+1} at time $n+1$ depends only on its current value and not on how it got there.

Consider the case where the ξ_n's are independent Bernoulli random variables with $\mathbb{E}\xi_n = \pm 1 = \frac{1}{2}$. Then, $\mathbb{P}(X_{n+1} = i_{n+1} | X_1 = i_1, \dots X_n = i_n) = \frac{1}{2}$ if $i_{n+1} = i_n \pm 1$ and 0 otherwise.

The random walk is an example of a **discrete time Markov chain**:

Definition 4.1.1. *A stochastic process $\{S_n; n \in \mathbb{N}\}$ and state space is $S = \mathbb{Z}$ is called a discrete time Markov chain provided that the* **Markov property** (4.1) *is satisfied.*

Consider now a continuous-time stochastic process X_t with state space $S = \mathbb{Z}$ and denote by $\{X_s,\, s \leqslant t\}$ the collection of values of the stochastic process up to time t. We will say that X_t is a Markov processes provided that

$$\mathbb{P}(X_{t+h} = i_{t+h} | \{X_s,\, s \leqslant t\}) = \mathbb{P}(X_{t+h} = i_{t+h} | X_t = i_t), \tag{4.2}$$

for all $h \geqslant 0$. A continuous-time, discrete state space Markov process is called a **continuous-time Markov chain**.

Example 4.1.2. *The Poisson process is a continuous-time Markov chain with*

$$\mathbb{P}(N_{t+h} = j | N_t = i) = \begin{cases} 0 & \text{if } j < i, \\ \frac{e^{-\lambda s}(\lambda s)^{j-i}}{(j-i)!}, & \text{if } j \geqslant i. \end{cases}$$

Similarly, we can define a Markov process for a continuous-time Markov process whose state space is $\mathbb{R}$. In this case, the above definitions become

$$\mathbb{P}(X_{t+h} \in \Gamma | \{X_s,\, s \leqslant t\}) = \mathbb{P}(X_{t+h} \in \Gamma | X_t = x) \tag{4.3}$$

for all Borel sets Γ.

Example 4.1.3. *The Brownian motion is a Markov process with* **conditional probability density**

$$p(y,t|x,s) := p(W_t = y | W_s = x) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{|x-y|^2}{2(t-s)}\right). \tag{4.4}$$

Example 4.1.4. *The Ornstein-Uhlenbeck process $V_t = e^{-t}W(e^{2t})$ is a Markov process with* **conditional probability density**

$$p(y,t|x,s) := p(V_t = y | V_s = x) = \frac{1}{\sqrt{2\pi(1-e^{-2(t-s)})}} \exp\left(-\frac{|y - xe^{-(t-s)}|^2}{2(1-e^{-2(t-s)})}\right). \tag{4.5}$$

To prove (4.5) *we use the formula for the distribution function of the Brownian motion to calculate, for $t > s$,*

$$\begin{aligned} \mathbb{P}(V_t \leqslant y | V_s = x) &= \mathbb{P}(e^{-t}W(e^{2t}) \leqslant y | e^{-s}W(e^{2s}) = x) \\ &= \mathbb{P}(W(e^{2t}) \leqslant e^t y | W(e^{2s}) = e^s x) \\ &= \int_{-\infty}^{e^t y} \frac{1}{\sqrt{2\pi(e^{2t}-e^{2s})}} e^{-\frac{|ze^t - xe^s|}{2(e^{2t}-e^{2s})}}\, dz \\ &= \int_{-\infty}^{y} \sqrt{2\pi e^{2t}(1-e^{-2(t-s)})}\, e^{-\frac{|\rho e^t - xe^s|}{2(e^{2t}(1-e^{-2(t-s)})}}\, d\rho \\ &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi(1-e^{-2(t-s)})}} e^{-\frac{|\rho - x|}{2(1-e^{-2(t-s)})}}\, d\rho. \end{aligned}$$

Consequently, the transition probability density for the OU process is given by the formula

$$\begin{aligned} p(y,t|x,s) &= \frac{\partial}{\partial y}\mathbb{P}(V_t \leqslant y | V_s = x) \\ &= \frac{1}{\sqrt{2\pi(1-e^{-2(t-s)})}} \exp\left(-\frac{|y - xe^{-(t-s)}|^2}{2(1-e^{-2(t-s)})}\right). \end{aligned}$$

Markov stochastic processes appear in a variety of applications in physics, chemistry, biology and finance. In this and the next chapter we will develop various analytical tools for studying them. In particular, we will see that we can obtain an equation for the **transition probability**

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n), \quad \mathbb{P}(X_{t+h} = i_{t+h} | X_t = i_t), \quad p(X_t + h = y | X_t = x), \tag{4.6}$$

which will enable us to study the evolution of a Markov process. This equation will be called the **Chapman-Kolmogorov** equation.

We will be mostly concerned with **time-homogeneous** Markov processes, i.e. processes for which the conditional probabilities are invariant under time shifts. For time-homogeneous discrete-time Markov chains we have

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) =: p_{ij}.$$

We will refer to the matrix $P = \{p_{ij}\}$ as the **transition matrix**. It is each to check that the transition matrix is a **stochastic matrix**, i.e. it has nonnegative entries and $\sum_j p_{ij} = 1$. Similarly, we can define the n-step transition matrix $P_n = \{p_{ij}(n)\}$ as

$$p_{ij}(n) = \mathbb{P}(X_{m+n} = j | X_m = i).$$

We can study the evolution of a Markov chain through the Chapman-Kolmogorov equation:

$$p_{ij} = (m + n) = \sum_k p_{ik}(m) p_{kj}(n). \tag{4.7}$$

Indeed, let $\mu_i^{(n)} := \mathbb{P}(X_n = i)$. The (possibly infinite dimensional) vector μ^n determines the state of the Markov chain at time n. A simple consequence of the Chapman-Kolmogorov equation is that we can write an evolution equation for the vector $\mu^{(n)}$

$$\mu^{(n)} = \mu^{(0)} P^n, \tag{4.8}$$

where P^n denotes the nth power of the matrix P. Hence in order to calculate the state of the Markov chain at time n all we need is the initial distribution μ^0 and the transition matrix P. Componentwise, the above equation can be written as

$$\mu_j^{(n)} = \sum_i \mu_i^{(0)} \pi_{ij}(n).$$

Consider now a continuous time Markov chain with transition probability

$$p_{ij}(s, t) = \mathbb{P}(X_t = j | X_s = i), \quad s \leqslant t.$$

If the chain is homogeneous, then

$$p_{ij}(s, t) = p_{ij}(0, t - s) \quad \text{for all } i, j, s, t.$$

In particular,

$$p_{ij}(t) = \mathbb{P}(X_t = j | X_0 = i).$$

The Chapman-Kolmogorov equation for a continuous time Markov chain is

$$\frac{dp_{ij}}{dt} = \sum_k p_{ik}(t) g_{kj}, \tag{4.9}$$

where the matrix G is called the **generator** of the Markov chain. Equation (4.9) can also be written in matrix notation:

$$\frac{dP}{dt} = P_t G.$$

The generator of the Markov chain is defined as

$$G = \lim_{h\to 0} \frac{1}{h}(P_h - I).$$

Let now $\mu_t^i = \mathbb{P}(X_t = i)$. The vector μ_t is the distribution of the Markov chain at time t. We can study its evolution using the equation

$$\mu_t = \mu_0 P_t.$$

Thus, as in the case if discrete time Markov chains, the evolution of a continuous time Markov chain is completely determined by the initial distribution and and transition matrix.

Consider now the case a continuous time Markov process with continuous state space and with continuous paths. As we have seen in Example 4.1.3 the Brownian motion is an example of such a process. It is a standard result in the theory of partial differential equations that the conditional probability density of the Brownian motion (4.4) is the fundamental solution of the diffusion equation:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}, \quad \lim_{t\to s} p(y,t|x,s) = \delta(y-x). \tag{4.10}$$

Exercise 4.1.5. *Show that* (4.4) *is the solution of initial value problem* (4.10) *as well as of the final value problem*

$$-\frac{\partial p}{\partial s} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \quad \lim_{s\to t} p(y,t|x,s) = \delta(y-x).$$

Similarly, the conditional distribution of the OU process satisfies the initial value problem

$$\frac{\partial p}{\partial t} = \frac{\partial (yp)}{\partial y} + \frac{\partial^2 p}{\partial y^2}, \quad \lim_{t\to s} p(y,t|x,s) = \delta(y-x). \tag{4.11}$$

The Brownian motion and the OU process are examples of a **diffusion process**. A diffusion process is a continuous time Markov process with continuous paths. We will see in Chapter 5, that the conditional probability density $p(y,t|x,s)$ of a diffusion process satisfies the **forward Kolmogorov** or **Fokker-Planck equation**

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}\left(a(y,t)p\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(y,t)p), \quad \lim_{t\to s} p(y,t|x,s) = \delta(y-x). \tag{4.12}$$

as well as the **backward Kolmogorov equation**

$$-\frac{\partial p}{\partial s} = -a(x,s)\frac{\partial p}{\partial x} + \frac{1}{2}b(x,s)\frac{\partial^2 p}{\partial x^2}, \quad \lim_{t\to s} p(y,t|x,s) = \delta(y-x). \tag{4.13}$$

for appropriate functions $a(y,t)$, $b(y,t)$. Hence, a diffusion process is determined uniquely from these two functions.

Exercise 4.1.6. *Use* (4.5) *to show that the forward and backward Kolmogorov equations for the OU process are*

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial y}(yp) + \frac{1}{2}\frac{\partial^2 p}{\partial y^2}$$

and

$$-\frac{\partial p}{\partial s} = -x\frac{\partial p}{\partial x} + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}.$$

4.2 Definition of a Markov Process

In Section 4.1 we gave the definition of Markov process whose time is either discrete or continuous, and whose state space is the set of integers. We also gave several examples of Markov chains as well as of processes whose state space is the real line. In this section we give the precise definition of a Markov process with $t \in T$, a general index set and $S = E$, an arbitrary metric space. We will use this formulation in the next section to derive the Chapman-Kolmogorov equation.

In order to state the definition of a continuous-time Markov process that takes values in a metric space we need to introduce various new concepts. For the definition of a Markov process we need to use the conditional expectation of the stochastic process conditioned on all past values. We can encode all past information about a stochastic process into an appropriate collection of σ-algebras. Our setting will be that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an ordered set T. Let $X = X_t(\omega)$ be a stochastic process from the sample space $(\Omega, \mathcal{F})$ to the state space $(E, \mathcal{G})$, where E is a metric space (we will usually take E to be either $\mathbb{R}$ or $\mathbb{R}^d$). Remember that the stochastic process is a function of two variables, $t \in T$ and $\omega \in \Omega$.

We start with the definition of a σ–algebra generated by a collection of sets.

Definition 4.2.1. *Let $\mathcal{K}$ be a collection of subsets of Ω. The smallest σ–algebra on Ω which contains $\mathcal{K}$ is denoted by $\sigma(\mathcal{K})$ and is called the σ–***algebra generated** *by $\mathcal{K}$.*

Definition 4.2.2. *Let $X_t : \Omega \mapsto E$, $t \in T$. The smallest σ–algebra $\sigma(X_t,\, t \in T)$, such that the family of mappings $\{X_t,\, t \in T\}$ is a stochastic process with sample space $(\Omega, \sigma(X_t,\, t \in T))$ and state space $(E, \mathcal{G})$, is called the σ–***algebra generated by*** $\{X_t,\, t \in T\}$.

In other words, the σ–algebra generated by X_t is the smallest σ–algebra such that X_t is a measurable function (random variable) with respect to it: the set

$$\left(\omega \in \Omega :\, X_t(\omega) \leqslant x\right) \in \sigma(X_t, t \in T)$$

for all $x \in \mathbb{R}$ (we have assumed that $E = X$).

Definition 4.2.3. *A* **filtration** *on $(\Omega, \mathcal{F})$ is a nondecreasing family $\{\mathcal{F}_t, t \in T\}$ of sub–σ–algebras of $\mathcal{F}$: $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s \leqslant t$.*

We set $\mathcal{F}_\infty = \sigma(\cup_{t \in T} \mathcal{F}_t)$. The **filtration generated by** X_t, where X_t is a stochastic process, is

$$\mathcal{F}_t^X := \sigma\left(X_s; s \leqslant t\right).$$

Definition 4.2.4. *A stochastic process $\{X_t; t \in T\}$ is* **adapted** *to the filtration $\{\mathcal{F}_t\} := \{\mathcal{F}_t, t \in T\}$ if for all $t \in T$, X_t is an $\mathcal{F}_t$–measurable random variable.*

Definition 4.2.5. *Let $\{X_t\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with values in E and let $\mathcal{F}_t^X$ be the filtration generated by $\{X_t\}$. Then $\{X_t\}$ is a* **Markov process** *if*

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma | X_s) \tag{4.14}$$

for all $t, s \in T$ with $t \geqslant s$, and $\Gamma \in \mathcal{B}(E)$.

Remark 4.2.6. *The filtration $\mathcal{F}_t^X$ is generated by events of the form $\{\omega | X_{s_1} \in B_1,\, X_{s_2} \in B_2,\, \dots X_{s_n} \in B_n,\}$ with $0 \leqslant s_1 < s_2 < \cdots < s_n \leqslant s$ and $B_i \in \mathcal{B}(E)$. The definition of a Markov process is thus equivalent to the hierarchy of equations*

$$\mathbb{P}(X_t \in \Gamma | X_{t_1}, X_{t_2}, \dots X_{t_n}) = \mathbb{P}(X_t \in \Gamma | X_{t_n}) \quad \textit{a.s.}$$

for $n \geqslant 1$, $0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant t$ and $\Gamma \in \mathcal{B}(E)$.

Roughly speaking, the statistics of X_t for $t \geqslant s$ are completely determined once X_s is known; information about X_t for $t < s$ is superfluous. In other words: **a Markov process has no memory**. More precisely: when a Markov process is conditioned on the present state, then there is no memory of the past. *The past and future of a Markov process are statistically independent when the present is known.*

Remark 4.2.7. *A non Markovian process X_t can be described through a Markovian one Y_t by enlarging the state space: the additional variables that we introduce account for the memory in the X_t. This "Markovianization" trick is very useful since there are many more tools for analyzing Markovian process.*

Example 4.2.8. *The velocity of a Brownian particle is modeled by the stationary Ornstein-Uhlenbeck process $Y_t = e^{-t} W(e^{2t})$. The particle position is given by the integral of the OU process (we take $X_0 = 0$)*

$$X_t = \int_0^t Y_s \, ds.$$

The particle position depends on the past of the OU process and, consequently, is not a Markov process. However, the joint position-velocity process $\{X_t, Y_t\}$ is. Its transition probability density $p(x, y, t | x_0, y_0)$ satisfies the forward Kolmogorov equation

$$\frac{\partial p}{\partial t} = -p \frac{\partial p}{\partial x} + \frac{\partial}{\partial y}(yp) + \frac{1}{2} \frac{\partial^2 p}{\partial y^2}.$$

4.3 The Chapman-Kolmogorov Equation

With a Markov process $\{X_t\}$ we can associate a function $P : T \times T \times E \times \mathcal{B}(E) \to \mathbb{R}^+$ defined through the relation

$$\mathbb{P}\left[X_t \in \Gamma | \mathcal{F}_s^X\right] = P(s, t, X_s, \Gamma),$$

for all $t, s \in T$ with $t \geqslant s$ and all $\Gamma \in \mathcal{B}(E)$. Assume that $X_s = x$. Since $\mathbb{P}\left[X_t \in \Gamma | \mathcal{F}_s^X\right] = \mathbb{P}\left[X_t \in \Gamma | X_s\right]$ we can write

$$P(\Gamma, t | x, s) = \mathbb{P}\left[X_t \in \Gamma | X_s = x\right].$$

The **transition function** $P(t, \Gamma | x, s)$ is (for fixed $t,\ x\, s$) a probability measure on E with $P(t, E | x, s) = 1$; it is $\mathcal{B}(E)$–measurable in x (for fixed $t,\ s,\ \Gamma$) and satisfies the **Chapman–Kolmogorov** equation

$$P(\Gamma, t | x, s) = \int_E P(\Gamma, t | y, u) P(dy, u | x, s). \tag{4.15}$$

for all $x \in E$, $\Gamma \in \mathcal{B}(E)$ and $s, u, t \in T$ with $s \leqslant u \leqslant t$. The derivation of the Chapman-Kolmogorov equation is based on the assumption of Markovianity and on properties of the conditional probability. Let

$(\Omega, \mathcal{F}, \mu)$ be a probability space, X a random variable from $(\Omega, \mathcal{F}, \mu)$ to $(E, \mathcal{G})$ and let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$. Then

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1). \tag{4.16}$$

Given $\mathcal{G} \subset \mathcal{F}$ we define the function $P_X(B|\mathcal{G}) = P(X \in B|\mathcal{G})$ for $B \in \mathcal{F}$. Assume that f is such that $\mathbb{E}(f(X)) < \infty$. Then

$$\mathbb{E}(f(X)|\mathcal{G}) = \int_{\mathbb{R}} f(x) P_X(dx|\mathcal{G}). \tag{4.17}$$

Now we use the Markov property, together with equations (4.16) and (4.17) and the fact that $s < u \Rightarrow \mathcal{F}_s^X \subset \mathcal{F}_u^X$ to calculate:

$$\begin{aligned}
P(\Gamma, t|x, s) &:= \mathbb{P}(X_t \in \Gamma | X_s = x) = \mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) \\
&= \mathbb{E}(I_\Gamma(X_t)|\mathcal{F}_s^X) = \mathbb{E}(\mathbb{E}(I_\Gamma(X_t)|\mathcal{F}_s^X)|\mathcal{F}_u^X) \\
&= \mathbb{E}(\mathbb{E}(I_\Gamma(X_t)|\mathcal{F}_u^X)|\mathcal{F}_s^X) = \mathbb{E}(\mathbb{P}(X_t \in \Gamma|X_u)|\mathcal{F}_s^X) \\
&= \mathbb{E}(\mathbb{P}(X_t \in \Gamma|X_u = y)|X_s = x) \\
&= \int_{\mathbb{R}} P(\Gamma, t|X_u = y) P(dz, u|X_s = x) \\
&=: \int_{\mathbb{R}} P(\Gamma, t|y, u) P(dy, u|x, s).
\end{aligned}$$

$I_\Gamma(\cdot)$ denotes the indicator function of the set Γ. We have also set $E = \mathbb{R}$. The CK equation is an integral equation and is the fundamental equation in the theory of Markov processes. Under additional assumptions we will derive from it the Fokker-Planck PDE, which is the fundamental equation in the theory of diffusion processes, and will be the main object of study in this course.

Definition 4.3.1. *A Markov process is* **homogeneous** *if*

$$P(t, \Gamma|X_s = x) := P(s, t, x, \Gamma) = P(0, t - s, x, \Gamma).$$

We set $P(0, t, \cdot, \cdot) = P(t, \cdot, \cdot)$. The Chapman–Kolmogorov (CK) equation becomes

$$P(t + s, x, \Gamma) = \int_E P(s, x, dz) P(t, z, \Gamma). \tag{4.18}$$

Let X_t be a homogeneous Markov process and assume that the **initial distribution** of X_t is given by the probability measure $\nu(\Gamma) = P(X_0 \in \Gamma)$ (for deterministic initial conditions–$X_0 = x$– we have that $\nu(\Gamma) = I_\Gamma(x)$). The transition function $P(x, t, \Gamma)$ and the initial distribution ν determine the finite dimensional distributions of X by

$$\begin{aligned}
&\mathbb{P}(X_0 \in \Gamma_1,\, X(t_1) \in \Gamma_1,\, \ldots, X_{t_n} \in \Gamma_n) \\
&= \int_{\Gamma_0} \int_{\Gamma_1} \cdots \int_{\Gamma_{n-1}} P(t_n - t_{n-1}, y_{n-1}, \Gamma_n) P(t_{n-1} - t_{n-2}, y_{n-2}, dy_{n-1}) \\
&\cdots \times P(t_1, y_0, dy_1) \nu(dy_0).
\end{aligned} \tag{4.19}$$

Theorem 4.3.2. *([3, Sec. 4.1]) Let $P(t, x, \Gamma)$ satisfy* (4.18) *and assume that (E, ρ) is a complete separable metric space. Then there exists a Markov process X in E whose finite-dimensional distributions are uniquely determined by* (4.19).

Let X_t be a homogeneous Markov process with initial distribution $\nu(\Gamma) = P(X_0 \in \Gamma)$ and transition function $P(x, t, \Gamma)$. We can calculate the probability of finding X_t in a set Γ at time t:

$$\mathbb{P}(X_t \in \Gamma) = \int_E P(x, t, \Gamma)\nu(dx).$$

Thus, the initial distribution and the transition function are sufficient to characterize a homogeneous Markov process. Notice that they do not provide us with any information about the actual paths of the Markov process. The transition probability $P(\Gamma, t|x, s)$ is a probability measure. Assume that it has a density for all $t > s$:

$$P(\Gamma, t|x, s) = \int_\Gamma p(y, t|x, s)\, dy.$$

Clearly, for $t = s$ we have $P(\Gamma, s|x, s) = I_\Gamma(x)$. The Chapman-Kolmogorov equation becomes:

$$\int_\Gamma p(y, t|x, s)\, dy = \int_{\mathbb{R}} \int_\Gamma p(y, t|z, u)p(z, u|x, s)\, dzdy,$$

and, since $\Gamma \in \mathcal{B}(\mathbb{R})$ is arbitrary, we obtain the equation

$$p(y, t|x, s) = \int_{\mathbb{R}} p(y, t|z, u)p(z, u|x, s)\, dz. \tag{4.20}$$

The transition probability density is a function of 4 arguments: the initial position and time $x,\ s$ and the final position and time y, t.

In words, the CK equation tells us that, for a Markov process, the transition from x, s to y, t can be done in two steps: first the system moves from x to z at some intermediate time u. Then it moves from z to y at time t. In order to calculate the probability for the transition from (x, s) to (y, t) we need to sum (integrate) the transitions from all possible intermediary states z. The above description suggests that a Markov process can be described through a **semigroup** of operators, i.e. a one-parameter family of linear operators with the properties

$$P_0 = I, \quad P_{t+s} = P_t \circ P_s \ \ \forall\, t, s \geqslant 0.$$

A semigroup of operators is characterized through its **generator**.

Indeed, let $P(t, x, dy)$ be the transition function of a homogeneous Markov process. It satisfies the CK equation (4.18):

$$P(t + s, x, \Gamma) = \int_E P(s, x, dz)P(t, z, \Gamma).$$

Let $X := C_b(E)$ and define the operator

$$(P_t f)(x) := \mathbb{E}(f(X_t)|X_0 = x) = \int_E f(y)P(t, x, dy).$$

This is a linear operator with

$$(P_0 f)(x) = \mathbb{E}(f(X_0)|X_0 = x) = f(x) \ \Rightarrow \ P_0 = I.$$

Furthermore:

$$\begin{aligned}(P_{t+s}f)(x) &= \int f(y)P(t+s,x,dy)\\ &= \int\int f(y)P(s,z,dy)P(t,x,dz)\\ &= \int\left(\int f(y)P(s,z,dy)\right)P(t,x,dz)\\ &= \int (P_sf)(z)P(t,x,dz)\\ &= (P_t\circ P_sf)(x).\end{aligned}$$

Consequently:

$$P_{t+s} = P_t \circ P_s.$$

4.4 The Generator of a Markov Processes

- Let (E,ρ) be a metric space and let $\{X_t\}$ be an E-valued homogeneous Markov process. Define the one parameter family of operators P_t through

$$P_tf(x) = \int f(y)P(t,x,dy) = \mathbb{E}[f(X_t)|X_0 = x]$$

  for all $f(x) \in C_b(E)$ (continuous bounded functions on E).

- Assume for simplicity that $P_t : C_b(E) \to C_b(E)$. Then the one-parameter family of operators P_t forms a **semigroup** of operators on $C_b(E)$.

- We define by $\mathcal{D}(\mathcal{L})$ the set of all $f \in C_b(E)$ such that the strong limit

$$\mathcal{L}f = \lim_{t\to 0}\frac{P_tf - f}{t},$$

  exists.

Definition 4.4.1. *The operator $\mathcal{L} : \mathcal{D}(\mathcal{L}) \to C_b(E)$ is called the* **infinitesimal generator** *of the operator semigroup P_t.*

Definition 4.4.2. *The operator $\mathcal{L} : C_b(E) \to C_b(E)$ defined above is called the* generator *of the Markov process $\{X_t\}$.*

- The study of operator semigroups started in the late 40's independently by Hille and Yosida. Semigroup theory was developed in the 50's and 60's by Feller, Dynkin and others, mostly in connection to the theory of Markov processes.

- Necessary and sufficient conditions for an operator $\mathcal{L}$ to be the generator of a (contraction) semigroup are given by the Hille-Yosida theorem (e.g. Evans Partial Differential Equations, AMS 1998, Ch. 7).

- The semigroup property and the definition of the generator of a semigroup imply that, formally at least, we can write:

$$P_t = \exp(\mathcal{L}t).$$

- Consider the function $u(x,t) := (P_t f)(x)$. We calculate its time derivative:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{d}{dt}(P_t f) = \frac{d}{dt}\left(e^{\mathcal{L}t} f\right) \\ &= \mathcal{L}\left(e^{\mathcal{L}t} f\right) = \mathcal{L}P_t f = \mathcal{L}u. \end{aligned}$$

- Furthermore, $u(x,0) = P_0 f(x) = f(x)$. Consequently, $u(x,t)$ satisfies the initial value problem

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(x,0) = f(x). \tag{4.21}$$

- When the semigroup P_t is the transition semigroup of a Markov process X_t, then equation (4.21) is called the **backward Kolmogorov equation**. It governs the evolution of an **observable**

$$u(x,t) = \mathbb{E}(f(X_t)|X_0 = x).$$

- Thus, given the generator of a Markov process $\mathcal{L}$, we can calculate all the statistics of our process by solving the backward Kolmogorov equation.

- In the case where the Markov process is the solution of a stochastic differential equation, then the generator is a second order elliptic operator and the backward Kolmogorov equation becomes an initial value problem for a parabolic PDE.

- The space $C_b(E)$ is natural in a probabilistic context, but other Banach spaces often arise in applications; in particular when there is a measure μ on E, the spaces $L^p(E;\mu)$ sometimes arise. We will quite often use the space $L^2(E;\mu)$, where μ will is the **invariant measure** of our Markov process.

- The generator is frequently taken as the starting point for the definition of a homogeneous Markov process.

- Conversely, let P_t be a **contraction semigroup** (Let X be a Banach space and $T : X \to X$ a bounded operator. Then T is a contraction provided that $\|Tf\|_X \leqslant \|f\|_X \ \forall\, f \in X$), with $\mathcal{D}(P_t) \subset C_b(E)$, closed. Then, under mild technical hypotheses, there is an E–valued homogeneous Markov process $\{X_t\}$ associated with P_t defined through

$$\mathbb{E}[f(X(t)|\mathcal{F}_s^X)] = P_{t-s} f(X(s))$$

for all $t, s \in T$ with $t \geqslant s$ and $f \in \mathcal{D}(P_t)$.

Example 4.4.3. *The Poisson process is a homogeneous Markov process.*

Example 4.4.4. *The one dimensional Brownian motion is a homogeneous Markov process. The transition function is the Gaussian defined in the example in Lecture 2:*

$$P(t,x,dy) = \gamma_{t,x}(y)dy, \;\; \gamma_{t,x}(y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|x-y|^2}{2t}\right).$$

The semigroup associated to the standard Brownian motion is the heat semigroup $P_t = e^{\frac{t}{2}\frac{d^2}{dx^2}}$. The generator of this Markov process is $\frac{1}{2}\frac{d^2}{dx^2}$.

- Notice that the **transition probability density** $\gamma_{t,x}$ of the one dimensional Brownian motion is the fundamental solution (Green's function) of the heat (diffusion) PDE

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}.$$

4.5 The Adjoint Semigroup

- The semigroup P_t acts on bounded measurable functions.
- We can also define the adjoint semigroup P_t^* which acts on probability measures:

$$P_t^*\mu(\Gamma) = \int_{\mathbb{R}} \mathbb{P}(X_t \in \Gamma | X_0 = x)\, d\mu(x) = \int_{\mathbb{R}} p(t,x,\Gamma)\, d\mu(x).$$

- The image of a probability measure μ under P_t^* is again a probability measure. The operators P_t and P_t^* are adjoint in the L^2-sense:

$$\int_{\mathbb{R}} P_t f(x)\, d\mu(x) = \int_{\mathbb{R}} f(x)\, d(P_t^*\mu)(x). \tag{4.22}$$

- We can, formally at least, write

$$P_t^* = \exp(\mathcal{L}^* t),$$

- where $\mathcal{L}^*$ is the L^2-adjoint of the generator of the process:

$$\int \mathcal{L} f h\, dx = \int f \mathcal{L}^* h\, dx.$$

- Let $\mu_t := P_t^*\mu$. This is the **law** of the Markov process and μ is the initial distribution. An argument similar to the one used in the derivation of the backward Kolmogorov equation (4.21) enables us to obtain an equation for the evolution of μ_t:

$$\frac{\partial \mu_t}{\partial t} = \mathcal{L}^*\mu_t, \quad \mu_0 = \mu.$$

- Assuming that $\mu_t = \rho(y,t)\, dy, \quad \mu = \rho_0(y)\, dy$ this equation becomes:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho, \quad \rho(y,0) = \rho_0(y). \tag{4.23}$$

- This is the **forward Kolmogorov** or **Fokker-Planck** equation. When the initial conditions are deterministic, $X_0 = x$, the initial condition becomes $\rho_0 = \delta(y - x)$.
- Given the initial distribution and the generator of the Markov process X_t, we can calculate the transition probability density by solving the Forward Kolmogorov equation. We can then calculate all statistical quantities of this process through the formula

$$\mathbb{E}(f(X_t)|X_0 = x) = \int f(y)\rho(t, y; x)\, dy.$$

- We will derive rigorously the backward and forward Kolmogorov equations for Markov processes that are defined as solutions of stochastic differential equations later on.
- We can study the evolution of a Markov process in two different ways:
- Either through the evolution of **observables** (Heisenberg/Koopman)

$$\frac{\partial(P_t f)}{\partial t} = \mathcal{L}(P_t f),$$

- or through the evolution of **states** (Schrödinger/Frobenious-Perron)

$$\frac{\partial(P_t^* \mu)}{\partial t} = \mathcal{L}^*(P_t^* \mu).$$

- We can also study Markov processes at the level of trajectories. We will do this after we define the concept of a stochastic differential equation.

4.6 Ergodic Markov processes

- A very important concept in the study of limit theorems for stochastic processes is that of **ergodicity**.
- This concept, in the context of Markov processes, provides us with information on the long–time behavior of a Markov semigroup.

  Definition 4.6.1. *A Markov process is called* **ergodic** *if the equation*

$$P_t g = g, \quad g \in C_b(E) \quad \forall t \geqslant 0$$

  has only constant solutions.

- Roughly speaking, ergodicity corresponds to the case where the semigroup P_t is such that $P_t - I$ has only constants in its null space, or, equivalently, to the case where the generator $\mathcal{L}$ has only constants in its null space. This follows from the definition of the generator of a Markov process.
- Under some additional compactness assumptions, an ergodic Markov process has an *invariant measure* μ with the property that, in the case $T = \mathbb{R}^+$,

$$\lim_{t \to +\infty} \frac{1}{t} \int_0^t g(X_s)\, ds = \mathbb{E}g(x),$$

- where $\mathbb{E}$ denotes the expectation with respect to μ.
- This is a physicist's definition of an ergodic process: **time averages equal phase space averages**.
- Using the adjoint semigroup we can define an invariant measure as the solution of the equation

$$P_t^* \mu = \mu.$$

- If this measure is unique, then the Markov process is ergodic.
- Using this, we can obtain an equation for the invariant measure in terms of the adjoint of the generator L^*, which is the generator of the semigroup P_t^*. Indeed, from the definition of the generator of a semigroup and the definition of an invariant measure, we conclude that a measure μ is invariant if and only if

$$\mathcal{L}^* \mu = 0$$

- in some appropriate generalized sense ($(\mathcal{L}^* \mu, f) = 0$ for every bounded measurable function).
- Assume that $\mu(dx) = \rho(x)\, dx$. Then the **invariant density** satisfies the **stationary Fokker-Planck equation**

$$\mathcal{L}^* \rho = 0.$$

- The invariant measure (distribution) governs the long-time dynamics of the Markov process.

4.7 Stationary Markov Processes

- If X_0 is distributed according to μ, then so is X_t for all $t > 0$. The resulting stochastic process, with X_0 distributed in this way, is **stationary** .
- In this case the transition probability density (the solution of the Fokker-Planck equation) is independent of time: $\rho(x, t) = \rho(x)$.
- Consequently, the statistics of the Markov process is independent of time.

Example 4.7.1. *The one dimensional Brownian motion* **is not** *an ergodic process: The null space of the generator $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$ on $\mathbb{R}$ is not one dimensional!*

Example 4.7.2. *Consider a one-dimensional Brownian motion on $[0, 1]$, with periodic boundary conditions. The generator of this Markov process $\mathcal{L}$ is the differential operator $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$, equipped with periodic boundary conditions on $[0, 1]$. This operator is self-adjoint. The null space of both $\mathcal{L}$ and $\mathcal{L}^*$ comprises constant functions on $[0, 1]$. Both the backward Kolmogorov and the Fokker-Planck equation reduce to the heat equation*

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2}$$

with periodic boundary conditions in $[0, 1]$. Fourier analysis shows that the solution converges to a constant at an exponential rate.

Example 4.7.3.

- *The one dimensional* **Ornstein-Uhlenbeck (OU) process** *is a Markov process with generator*

$$\mathcal{L} = -\alpha x \frac{d}{dx} + D\frac{d^2}{dx^2}.$$

- *The null space of $\mathcal{L}$ comprises constants in x. Hence, it is an ergodic Markov process. In order to calculate the invariant measure we need to solve the stationary Fokker–Planck equation:*

$$\mathcal{L}^*\rho = 0, \quad \rho \geqslant 0, \quad \|\rho\|_{L_1(\mathbb{R})} = 1. \tag{4.24}$$

- *Let us calculate the L^2-adjoint of $\mathcal{L}$. Assuming that f, h decay sufficiently fast at infinity, we have:*

$$\begin{aligned}\int_{\mathbb{R}} \mathcal{L}f h\, dx &= \int_{\mathbb{R}} \left[(-\alpha x \partial_x f)h + (D\partial_x^2 f)h\right]\, dx \\ &= \int_{\mathbb{R}} \left[f\partial_x(\alpha x h) + f(D\partial_x^2 h)\right]\, dx =: \int_{\mathbb{R}} f\mathcal{L}^* h\, dx,\end{aligned}$$

- *where*

$$\mathcal{L}^* h := \frac{d}{dx}\,(axh) + D\frac{d^2 h}{dx^2}.$$

- *We can calculate the invariant distribution by solving equation* (4.24).

- *The invariant measure of this process is the Gaussian measure*

$$\mu(dx) = \sqrt{\frac{\alpha}{2\pi D}} \exp\left(-\frac{\alpha}{2D}x^2\right)\, dx.$$

- *If the initial condition of the OU process is distributed according to the invariant measure, then the OU process is a stationary Gaussian process.*

- Let X_t be the 1d OU process and let $X_0 \sim \mathcal{N}(0, D/\alpha)$. Then X_t is a mean zero, Gaussian second order stationary process on $[0,\infty)$ with correlation function

$$R(t) = \frac{D}{\alpha} e^{-\alpha|t|}$$

and spectral density

$$f(x) = \frac{D}{\pi}\frac{1}{x^2+\alpha^2}.$$

Furthermore, the OU process is **the only** real-valued mean zero Gaussian second-order stationary Markov process defined on $\mathbb{R}$.

Chapter 5

Diffusion Processes

5.1 Definition of a Diffusion Process

- A Markov process consists of three parts: a drift (deterministic), a random process and a jump process.
- A diffusion process is a Markov process that has continuous sample paths (trajectories). Thus, it is a Markov process with no jumps.
- A diffusion process can be defined by specifying its first two moments:

Definition 5.1.1. *A Markov process X_t with transition probability $P(\Gamma, t|x, s)$ is called a* **diffusion process** *if the following conditions are satisfied.*

i. (Continuity). For every x and every $\varepsilon > 0$

$$\int_{|x-y|>\varepsilon} P(dy, t|x, s) = o(t-s) \tag{5.1}$$

uniformly over $s < t$.

ii. (Definition of drift coefficient). There exists a function $a(x, s)$ such that for every x and every $\varepsilon > 0$

$$\int_{|y-x|\leqslant\varepsilon} (y-x)P(dy, t|x, s) = a(x, s)(t-s) + o(t-s). \tag{5.2}$$

uniformly over $s < t$.

iii. (Definition of diffusion coefficient). There exists a function $b(x, s)$ such that for every x and every $\varepsilon > 0$

$$\int_{|y-x|\leqslant\varepsilon} (y-x)^2 P(dy, t|x, s) = b(x, s)(t-s) + o(t-s). \tag{5.3}$$

uniformly over $s < t$.

Remark 5.1.2. *In Definition 5.1.1 we had to truncate the domain of integration since we didn't know whether the first and second moments exist. If we assume that there exists a $\delta > 0$ such that*

$$\lim_{t\to s} \frac{1}{t-s} \int_{\mathbb{R}^d} |y-x|^{2+\delta} P(s, x, t, dy) = 0, \tag{5.4}$$

then we can extend the integration over the whole $\mathbb{R}^d$ and use expectations in the definition of the drift and the diffusion coefficient. Indeed, ,let $k = 0, 1, 2$ and notice that

$$\begin{aligned}\int_{|y-x|>\varepsilon} & |y-x|^k P(s,x,t,dy) \\ &= \int_{|y-x|>\varepsilon} |y-x|^{2+\delta}|y-x|^{k-(2+\delta)} P(s,x,t,dy) \\ &\leqslant \frac{1}{\varepsilon^{2+\delta-k}} \int_{|y-x|>\varepsilon} |y-x|^{2+\delta} P(s,x,t,dy) \\ &\leqslant \frac{1}{\varepsilon^{2+\delta-k}} \int_{\mathbb{R}^d} |y-x|^{2+\delta} P(s,x,t,dy).\end{aligned}$$

Using this estimate together with (5.4) *we conclude that:*

$$\lim_{t\to s} \frac{1}{t-s} \int_{|y-x|>\varepsilon} |y-x|^k P(s,x,t,dy) = 0, \quad k = 0, 1, 2.$$

This implies that assumption (5.4) *is sufficient for the sample paths to be continuous ($k = 0$) and for the replacement of the truncated integrals in* (5.2) *and* (5.3) *by integrals over $\mathbb{R}$ ($k = 1$ and $k = 2$, respectively). The definitions of the drift and diffusion coefficients become:*

$$\lim_{t\to s} \mathbb{E}\left(\frac{X_t - X_s}{t-s} \Big| X_s = x \right) = a(x,s) \tag{5.5}$$

and

$$\lim_{t\to s} \mathbb{E}\left(\frac{|X_t - X_s|^2}{t-s} \Big| X_s = x \right) = \mathbf{b}(x,s) \tag{5.6}$$

5.2 The Backward and Forward Kolmogorov Equations

Under mild regularity assumptions we can show that a diffusion process is completely determined from its first two moments. In particular, we will obtain partial differential equations that govern the evolution of the conditional expectation of an arbitrary function of a diffusion process X_t, $u(x,s) = \mathbb{E}(f(X_t)|X_s = x)$, as well as of the transition probability density $p(y,t|x,s) = \mathbb{P}(X_t = y|X_s = s)$. These are the **backward** and **forward** Kolmogorov equations.

5.2.1 The Backward Kolmogorov Equation

Theorem 5.2.1. *(Kolmogorov) Let $f(x) \in C_b(\mathbb{R})$ and let*

$$u(x,s) := \mathbb{E}(f(X_t)|X_s = x) = \int f(y) P(dy,t|x,s) \in C_b^2(\mathbb{R}).$$

Assume furthermore that the functions $a(x,s)$, $b(x,s)$ are continuous in both x and s. Then $u(x,s) \in C^{2,1}(\mathbb{R}\times\mathbb{R}^+)$ and it solves the **final value problem**

$$-\frac{\partial u}{\partial s} = a(x,s)\frac{\partial u}{\partial x} + \frac{1}{2} b(x,s)\frac{\partial^2 u}{\partial x^2}, \quad \lim_{s\to t} u(s,x) = f(x). \tag{5.7}$$

Proof. First we notice that, the continuity assumption (5.1), together with the fact that the function $f(x)$ is bounded imply that

$$\begin{aligned} u(x,s) &= \int_{\mathbb{R}} f(y)\, P(dy,t|x,s) \\ &= \int_{|y-x|\leqslant\varepsilon} f(y)P(dy,t|x,s) + \int_{|y-x|>\varepsilon} f(y)P(dy,t|x,s) \\ &\leqslant \int_{|y-x|\leqslant\varepsilon} f(y)P(dy,t|x,s) + \|f\|_{L^\infty} \int_{|y-x|>\varepsilon} P(dy,t|x,s) \\ &= \int_{|y-x|\leqslant\varepsilon} f(y)P(dy,t|x,s) + o(t-s). \end{aligned}$$

We add and subtract the final condition $f(x)$ and use the previous calculation to obtain:

$$\begin{aligned} u(x,s) &= \int_{\mathbb{R}} f(y)P(dy,t|x,s) = f(x) + \int_{\mathbb{R}} (f(y)-f(x))P(dy,t|x,s) \\ &= f(x) + \int_{|y-x|\leqslant\varepsilon} (f(y)-f(x))P(dy,t|x,s) + \int_{|y-x|>\varepsilon} (f(y)-f(x))P(dy,t|x,s) \\ &= f(x) + \int_{|y-x|\leqslant\varepsilon} (f(y)-f(x))P(dy,t|x,s) + o(t-s). \end{aligned}$$

Now the final condition follows from the fact that $f(x) \in C_b(\mathbb{R})$ and the arbitrariness of ε.

Now we show that $u(s,x)$ solves the backward Kolmogorov equation. We use the Chapman-Kolmogorov equation (4.15) to obtain

$$\begin{aligned} u(x,\sigma) &= \int_{\mathbb{R}} f(z)P(dz,t|x,\sigma) && (5.8) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(z)P(dz,t|y,\rho)P(dy,\rho|x,\sigma) \\ &= \int_{\mathbb{R}} u(y,\rho)P(dy,\rho|x,\sigma). && (5.9) \end{aligned}$$

The Taylor series expansion of the function $u(s,x)$ gives

$$u(z,\rho) - u(x,\rho) = \frac{\partial u(x,\rho)}{\partial x}(z-x) + \frac{1}{2}\frac{\partial^2 u(x,\rho)}{\partial x^2}(z-x)^2(1+\alpha_\varepsilon), \quad |z-x| \leqslant \varepsilon, \tag{5.10}$$

where

$$\alpha_\varepsilon = \sup_{\rho,|z-x|\leqslant\varepsilon} \left| \frac{\partial^2 u(x,\rho)}{\partial x^2} - \frac{\partial^2 u(z,\rho)}{\partial x^2} \right|.$$

Notice that, since $u(x,s)$ is twice continuously differentiable in x, $\lim_{\varepsilon\to 0}\alpha_\varepsilon = 0$.

We combine now (5.9) with (5.10) to calculate

$$\begin{aligned}
\frac{u(x,s)-u(x,s+h)}{h} &= \frac{1}{h}\left(\int_{\mathbb{R}} P(dy,s+h|x,s)u(y,s+h)-u(x,s+h)\right)\\
&= \frac{1}{h}\int_{\mathbb{R}} P(dy,s+h|x,s)(u(y,s+h)-u(x,s+h))\\
&= \frac{1}{h}\int_{|x-y|<\varepsilon} P(dy,s+h|x,s)(u(y,s+h)-u(x,s))+o(1)\\
&= \frac{\partial u}{\partial x}(x,s+h)\frac{1}{h}\int_{|x-y|<\varepsilon}(y-x)P(dy,s+h|x,s)\\
&\quad +\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(x,s+h)\frac{1}{h}\int_{|x-y|<\varepsilon}(y-x)^2P(dy,s+h|x,s)(1+\alpha_\varepsilon)+o(1)\\
&= a(x,s)\frac{\partial u}{\partial x}(x,s+h)+\frac{1}{2}b(x,s)\frac{\partial^2 u}{\partial x^2}(x,s+h)(1+\alpha_\varepsilon)+o(1).
\end{aligned}$$

Equation (5.7) follows by taking the limits $\varepsilon\to 0,\ h\to 0$. □

Assume now that the transition function has a density $p(y,t|x,s)$. In this case the formula for $u(x,s)$ becomes

$$u(x,s)=\int_{\mathbb{R}} f(y)p(y,t|x,s)\,dy.$$

Substituting this in the backward Kolmogorov equation we obtain

$$\int_{\mathbb{R}} f(y)\left(\frac{\partial p(y,t|x,s)}{\partial s}+\mathcal{A}_{s,x}p(y,t|x,s)\right)=0 \tag{5.11}$$

where

$$\mathcal{A}_{s,x}:=a(x,s)\frac{\partial}{\partial x}+\frac{1}{2}b(x,s)\frac{\partial^2}{\partial x^2}.$$

Since (5.11) is valid for arbitrary functions $f(y)$, we obtain a partial differential equations for the transition probability density:

$$-\frac{\partial p(y,t|x,s)}{\partial s}=a(x,s)\frac{\partial p(y,t|x,s)}{\partial x}+\frac{1}{2}b(x,s)\frac{\partial^2 p(y,t|x,s)}{\partial x^2}. \tag{5.12}$$

Notice that the variation is with respect to the "backward" variables $x,\ s$. We will obtain an equation with respect to the "forward" variables $y,\ t$ in the next section.

5.2.2 The Forward Kolmogorov Equation

In this section we will obtain the forward Kolmogorov equation. In the physics literature is called the **Fokker-Planck equation**. We assume that the transition function has a density with respect to Lebesgue measure.

$$P(\Gamma,t|x,s)=\int_{\Gamma} p(t,y|x,s)\,dy.$$

Theorem 5.2.2. *(Kolmogorov) Assume that conditions* (5.1), (5.2), (5.3) *are satisfied and that* $p(y,t|\cdot,\cdot),\ a(y,t), b(y,t) \in C^{2,1}(\mathbb{R}\times\mathbb{R}^+)$. *Then the transition probability density satisfies the equation*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}\left(a(t,y)p\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(t,y)p\right), \quad \lim_{t\to s} p(t,y|x,s) = \delta(x-y). \tag{5.13}$$

Proof. Fix a function $f(y) \in C_0^2(\mathbb{R})$. An argument similar to the one used in the proof of the backward Kolmogorov equation gives

$$\lim_{h\to 0}\frac{1}{h}\left(\int f(y)p(y,s+h|x,s)\,ds - f(x)\right) = a(x,s)f_x(x) + \frac{1}{2}b(x,s)f_{xx}(x), \tag{5.14}$$

where subscripts denote differentiation with respect to x. On the other hand

$$\begin{aligned}
\int f(y)\frac{\partial}{\partial t}p(y,t|x,s)\,dy &= \frac{\partial}{\partial t}\int f(y)p(y,t|x,s)\,dy \\
&= \lim_{h\to 0}\frac{1}{h}\int \left(p(y,t+h|x,s) - p(y,t|x,s)\right)f(y)\,dy \\
&= \lim_{h\to 0}\frac{1}{h}\left(\int p(y,t+h|x,s)f(y)\,dy - \int p(z,t|s,x)f(z)\,dz\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(\int\int p(y,t+s|z,t)p(z,t|x,s)f(y)\,dydz - \int p(z,t|s,x)f(z)\,dz\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(\int p(z,t|x,s)\left(\int p(y,t+h|z,t)f(y)\,dy - f(z)\right)\right)dz \\
&= \int p(z,t|x,s)\left(a(z,t)f_z(z) + \frac{1}{2}b(z)f_{zz}(z)\right)dz \\
&= \int\Big(-\frac{\partial}{\partial z}(a(z)p(z,t|x,s)) + \frac{1}{2}\frac{\partial^2}{\partial z^2}(b(z)p(z,t|x,s)\Big)f(z)\,dz.
\end{aligned}$$

In the above calculation used the Chapman-Kolmogorov equation. We have also performed two integrations by parts and used the fact that, since the test function f has compact support, the boundary terms vanish.

Since the above equation is valid for every test function $f(y)$, the forward Kolmogorov equation follows. □

Exercise 5.2.3. *Prove equation* (5.14).

Assume now that initial distribution of X_t is $\rho_0(x)$ and set $s = 0$ (the initial time) in (5.13). Define

$$p(y,t) := \int p(y,t|x,0)\rho_0(x)\,dx.$$

We multiply the forward Kolmogorov equation (5.13) by $\rho_0(x)$ and integrate with respect to x to obtain the equation

$$\frac{\partial p(y,t)}{\partial t} = -\frac{\partial}{\partial y}\left(a(y,t)p(y,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(y,t)p(t,y)\right), \tag{5.15}$$

together with the initial condition

$$p(y,0) = \rho_0(y). \tag{5.16}$$

The solution of equation (5.15), provides us with the probability that the diffusion process X_t, which initially was distributed according to the probability density $\rho_0(x)$, is equal to y at time t. Alternatively, we can think of the solution to (5.13) as the Green's function for the PDE (5.15).

Quite often we need to calculate joint probability densities. For, example the probability that $X_{t_1} = x_1$ and $X_{t_2} = x_2$. From the properties of conditional expectation we have that

$$\begin{aligned} p(x_1, t_1, x_2, t_2) &= \mathbb{P}X_{t_1} = x_1, X_{t_2} = x_2) \\ &= \mathbb{P}(X_{t_1} = x_1 | X_{t_2} = x_2)\mathbb{P}(X_{t_2} = x_2) \\ &= p(x_1, t_1 | x_2 t_2) p(x_2, t_2). \end{aligned}$$

Using the joint probability density we can calculate the statistics of a function of the diffusion process X_t at times t and s:

$$\mathbb{E}(f(X_t, X_s)) = \int\int f(y, x) p(y, t|x, s) p(x, s)\, dxdy. \tag{5.17}$$

The **autocorrelation function** at time t and s is given by

$$\mathbb{E}(X_t X_s) = \int\int yx p(y, t|x, s) p(x, s)\, dxdy.$$

In particular,

$$\mathbb{E}(X_t X_0) = \int\int yx p(y, t|x, 0) p(x, 0)\, dxdy.$$

Notice also that from (5.17) it follows that

$$\begin{aligned} \mathbb{E}(f(X_t)) &= \int\int f(y) p(y, t|x, 0) p(x, 0)\, dxdy \\ &= \int f(y) p(y, t)\, dy, \end{aligned}$$

where $p(y, t)$ is the solution of (5.15).

5.3 The Fokker-Planck Equation in Arbitrary Dimensions

The drift and diffusion coefficients of a diffusion process in $\mathbb{R}^d$ are defined as:

$$\lim_{t \to s} \frac{1}{t - s} \int_{|y-x|<\varepsilon} (y - x) P(dy, t|x, s) = \mathbf{a}(x, s)$$

and

$$\lim_{t \to s} \frac{1}{t - s} \int_{|y-x|<\varepsilon} (y - x) \otimes (y - x) P(dy, t|x, s) = \mathbf{b}(x, s).$$

The drift coefficient $\mathbf{a}(x, s)$ is a d-dimensional vector field and the diffusion coefficient $\mathbf{b}(x, s)$ is a $d \times d$ symmetric matrix (second order tensor). The generator of a d dimensional diffusion process is

$$\begin{aligned} \mathcal{L} &= a(s, x) \cdot \nabla + \frac{1}{2} b(s, x) : \nabla\nabla \\ &= \sum_{j=1}^{d} a_j \frac{\partial}{\partial x_j} + \frac{1}{2} \sum_{i,j=1}^{d} b_{ij} \frac{\partial^2}{\partial x_j^2}. \end{aligned}$$

Exercise 5.3.1. *Derive rigorously the forward and backward Kolmogorov equations in arbitrary dimensions.*

Assuming that the first and second moments of the multidimensional diffusion process exist, we can write the formulas for the drift vector and diffusion matrix as

$$\lim_{t\to s}\mathbb{E}\left(\frac{X_t-X_s}{t-s}\Big|X_s=x\right)=\mathbf{a}(x,s) \tag{5.18}$$

and

$$\lim_{t\to s}\mathbb{E}\left(\frac{(X_t-X_s)\otimes(X_t-X_s)}{t-s}\Big|X_s=x\right)=\mathbf{b}(x,s) \tag{5.19}$$

Notice that from the above definition it follows that the diffusion matrix is symmetric and nonnegative definite.

5.4 Connection with Stochastic Differential Equations

Notice also that the continuity condition can be written in the form

$$\mathbb{P}\left(|X_t-X_s|\geqslant\varepsilon|X_s=x\right)=o(t-s).$$

Now it becomes clear that this condition implies that the probability of large changes in X_t over short time intervals is small. Notice, on the other hand, that the above condition implies that the sample paths of a diffusion process **are not differentiable**: if they where, then the right hand side of the above equation would have to be 0 when $t-s\ll 1$. The sample paths of a diffusion process have the regularity of Brownian paths. A Markovian process **cannot be** differentiable: we can define the derivative of a sample paths only with processes for which the past and future are not statistically independent when conditioned on the present.

Let us denote the expectation conditioned on $X_s=x$ by $\mathbb{E}^{s,x}$. Notice that the definitions of the drift and diffusion coefficients (5.5) and (5.6) can be written in the form

$$\mathbb{E}^{s,x}(X_t-X_s)=\mathbf{a}(x,s)(t-s)+o(t-s).$$

and

$$\mathbb{E}^{s,x}\Big((X_t-X_s)\otimes(X_t-X_s)\Big)=\mathbf{b}(x,s)(t-s)+o(t-s).$$

Consequently, the drift coefficient defines the **mean velocity vector** for the stochastic process X_t, whereas the diffusion coefficient (tensor) is a measure of the local magnitude of fluctuations of X_t-X_s about the mean value. hence, we can write locally:

$$X_t-X_s\approx\mathbf{a}(s,X_s)(t-s)+\sigma(s,X_s)\,\xi_t,$$

where $\mathbf{b}=\sigma\sigma^{\mathbf{T}}$ and ξ_t is a mean zero Gaussian process with

$$E^{s,x}(\xi_t\otimes\xi_s)=(t-s)I.$$

Since we have that

$$W_t-W_s\sim\mathcal{N}(0,(t-s)I),$$

we conclude that we can write locally:

$$\Delta X_t \approx \mathbf{a}(s, X_s)\Delta t + \sigma(s, X_s)\Delta W_t.$$

Or, replacing the differences by differentials:

$$dX_t = \mathbf{a}(t, X_t)dt + \sigma(t, X_t)dW_t.$$

Hence, the sample paths of a diffusion process are governed by a **stochastic differential equation** (SDE).

5.5 Discussion and Bibliography

The argument used in the derivation of the forward and backward Kolmogorov equations goes back to Kolmogorov's original work. More material on diffusion processes can be found in [11], [14].

Chapter 6

The Fokker-Planck Equation

In the previous chapter we derived the backward and forward (Fokker-Planck) Kolmogorov equations and we showed that all statistical properties of a diffusion process can be calculated from the solution of the Fokker-Planck equation. [1] In this chapter we will study various properties of this equation such as existence and uniqueness of solutions, long time asymptotics, boundary conditions and spectral properties of the Fokker-Planck operator. We will also study in some detail various examples. We will restrict attention to time-homogeneous diffusion processes, for which the drift and diffusion coefficients do not depend on time.

6.1 Basic Properties of the FP Equation

6.1.1 Existence and Uniqueness of Solutions

Consider a homogeneous diffusion process on $\mathbb{R}^d$ with drift vector and diffusion matrix $\mathbf{a}(x)$ and $\mathbf{b}(x)$. The Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = -\sum_{j=1}^{d} \frac{\partial}{\partial x_j}(a_i(x)p) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(b_{ij}(x)p),\ t > 0,\ x \in \mathbb{R}^d, \tag{6.1a}$$

$$p(x,0) = f(x), \quad x \in \mathbb{R}^d. \tag{6.1b}$$

Since $f(x)$ is the probability density of the initial condition (which is a random variable), we have that

$$f(x) \geqslant 0, \quad \text{and} \quad \int_{\mathbb{R}^d} f(x)\,dx = 1.$$

We can also write the equation in **non-divergence form**:

$$\frac{\partial p}{\partial t} = \sum_{j=1}^{d} \tilde{a}_j(x)\frac{\partial p}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d} \tilde{b}_{ij(x)}\frac{\partial^2 p}{\partial x_i \partial x_j} + \tilde{c}(x)u,\ t > 0,\ x \in \mathbb{R}^d, \tag{6.2a}$$

$$p(x,0) = f(x), \quad x \in \mathbb{R}^d, \tag{6.2b}$$

[1] In this chapter we will call the equation Fokker-Planck, which is more customary in the physics literature. rather forward Kolmogorov, which is more customary in the mathematics literature.

where

$$\tilde{a}_i(x) = -a_i(x) + \sum_{j=1}^{d} \frac{\partial b_{ij}}{\partial x_j}, \quad \tilde{c}_i(x) = \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 b_{ij}}{\partial x_i \partial x_j} - \sum_{i=1}^{d} \frac{\partial a_i}{\partial x_i}.$$

By definition (see equation (5.19)), the diffusion matrix is always symmetric and nonnegative. We will assume that it is actually uniformly positive definite, i.e. we will impose the **uniform ellipticity** condition:

$$\sum_{i,j=1}^{d} b_{ij}(x)\xi_i\xi_j \geqslant \alpha\|\xi\|^2, \quad \forall\, \xi \in \mathbb{R}^d, \tag{6.3}$$

Furthermore, we will assume that the coefficients $\tilde{a},\ b,\ \tilde{c}$ are smooth and that they satisfy the growth conditions

$$\|b(x)\| \leqslant M,\ \|\tilde{a}(x)\| \leqslant M(1 + \|x\|),\ \|\tilde{c}(x)\| \leqslant M(1 + \|x\|^2). \tag{6.4}$$

Definition 6.1.1. *We will call a solution to the Cauchy problem for the Fokker–Planck equation* (6.2) *a* **classical solution** *if:*

i. $u \in C^{2,1}(\mathbb{R}^d, \mathbb{R}^+)$.

ii. $\forall T > 0$ *there exists a* $c > 0$ *such that*

$$\|u(t,x)\|_{L^\infty(0,T)} \leqslant ce^{\alpha\|x\|^2}$$

iii. $\lim_{t\to 0} u(t,x) = f(x)$.

It is a standard result in the theory of parabolic partial differential equations that, under the regularity and uniform ellipticity assumptions, the Fokker-Planck equation has a unique smooth solution. Furthermore, the solution can be estimated in terms of an appropriate heat kernel (i.e. the solution of the heat equation on $\mathbb{R}^d$).

Theorem 6.1.2. *Assume that conditions* (6.3) *and* (6.4) *are satisfied, and assume that* $|f| \leqslant ce^{\alpha\|x\|^2}$. *Then there exists a unique classical solution to the Cauchy problem for the Fokker–Planck equation. Furthermore, there exist positive constants* $K,\ \delta$ *so that*

$$|p|,\ |p_t|,\ \|\nabla p\|,\ \|D^2 p\| \leqslant Kt^{(-n+2)/2} \exp\left(-\frac{1}{2t}\delta\|x\|^2\right). \tag{6.5}$$

Notice that from estimates (6.5) it follows that all moments of a uniformly elliptic diffusion process exist. In particular, we can multiply the Fokker-Planck equation by monomials x^n and then to integrate over $\mathbb{R}^d$ and to integrate by parts. No boundary terms will appear, in view of the estimate (6.5).

Remark 6.1.3. *The solution of the Fokker-Planck equation is nonnegative for all times, provided that the initial distribution is nonnegative. This is follows from the* **maximum principle** *for parabolic PDEs.*

6.1.2 The FP equation as a conservation law

The Fokker-Planck equation is in fact a conservation law: it expresses the law of conservation of probability. To see this we define the **probability current** to be the vector whose ith component is

$$J_i := a_i(x)p - \frac{1}{2}\sum_{j=1}^{d} \frac{\partial}{\partial x_j}\big(b_{ij}(x)p\big). \tag{6.6}$$

We use the probability current to write the Fokker–Planck equation as a **continuity equation**:

$$\frac{\partial p}{\partial t} + \nabla \cdot J = 0.$$

Integrating the FP equation over $\mathbb{R}^d$ and integrating by parts on the right hand side of the equation we obtain

$$\frac{d}{dt}\int_{\mathbb{R}^d} p(x,t)\,dx = 0.$$

Consequently:

$$\|p(\cdot,t)\|_{L^1(\mathbb{R}^d)} = \|p(\cdot,0)\|_{L^1(\mathbb{R}^d)} = 1. \tag{6.7}$$

Hence, the total probability is conserved, as expected. Equation (6.7) simply means that

$$\mathbb{E}(X_t \in \mathbb{R}^d) = 1, \quad t \geqslant 0.$$

6.1.3 Boundary conditions for the Fokker–Planck equation

When studying a diffusion process that can take values on the whole of $\mathbb{R}^d$, then we study the pure initial value (Cauchy) problem for the Fokker-Planck equation, equation (6.1). The boundary condition was that the solution decays sufficiently fast at infinity. For ergodic diffusion processes this is equivalent to requiring that the solution of the backward Kolmogorov equation is an element of $L^2(\mu)$ where μ is the invariant measure of the process. There are many applications where it is important to study stochastic process in bounded domains. In this case it is necessary to specify the value of the stochastic process (or equivalently of the solution to the Fokker-Planck equation) on the boundary.

To understand the type of boundary conditions that we can impose on the Fokker-Planck equation, let us consider the example of a random walk on the domain $\{0, 1, \dots N\}$.[2] When the random walker reaches either the left or the right boundary we can either set

i. $X_0 = 0$ or $X_N = 0$, which means that the particle gets absorbed at the boundary;

ii. $X_0 = X_1$ or $X_N = X_{N-1}$, which means that the particle is reflected at the boundary;

iii. $X_0 = X_N$, which means that the particle is moving on a circle (i.e., we identify the left and right boundaries).

[2] Of course, the random walk is not a diffusion process. However, as we have already seen the Brownian motion can be defined as the limit of an appropriately rescaled random walk. A similar construction exists for more general diffusion processes.

Hence, we can have **absorbing, reflecting** or **periodic** boundary conditions.

Consider the Fokker-Planck equation posed in $\Omega \subset \mathbb{R}^d$ where Ω is a bounded domain with smooth boundary. Let $\mathbf{J}$ denote the probability current and let $\mathbf{n}$ be the unit outward pointing normal vector to the surface. The above boundary conditions become:

i. The transition probability density vanishes on an absorbing boundary:

$$p(x,t) = 0, \quad on\ \partial\Omega.$$

ii. There is no net flow of probability on a reflecting boundary:

$$\mathbf{n} \cdot \mathbf{J}(x,t) = 0, \quad on\ \partial\Omega.$$

iii. The transition probability density is a periodic function in the case of periodic boundary conditions.

Notice that, using the terminology customary to PDEs theory, absorbing boundary conditions correspond to Dirichlet boundary conditions and reflecting boundary conditions correspond to Neumann. Of course, on consider more complicated, mixed boundary conditions.

Consider now a diffusion process in one dimension on the interval $[0, L]$. The boundary conditions are

$$p(0,t) = p(L,t) = 0 \quad \text{absorbing},$$

$$J(0,t)) = J(L,t) = 0 \quad \text{reflecting},$$

$$p(0,t) = p(L,t) \quad \text{periodic},$$

where the probability current is defined in (6.6). An example of mixed boundary conditions would be absorbing boundary conditions at the left end and reflecting boundary conditions at the right end:

$$p(0,t) = J(L,t) = 0.$$

There is a complete classification of boundary conditions in one dimension, the **Feller classification**: the BC can be **regular, exit, entrance** and **natural**.

6.2 Examples of Diffusion Processes

6.2.1 Brownian Motion

Set $a(y,t) \equiv 0$, $b(y,t) \equiv 2D > 0$. This diffusion process is the Brownian motion with diffusion coefficient D. The Fokker-Planck equation is:

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial y^2}, \quad p(y,s|x,s) = \delta(x-y). \tag{6.8}$$

As we have already seen, the solution to this equation is

$$p_W(y,s|x,t) = \frac{1}{\sqrt{4\pi D(s-t)}} \exp\left(-\frac{(y-x)^2}{4D(s-t)}\right).$$

Notice that using the Fokker-Planck equation for the Brownian motion we can immediately show that the mean squared displacement scales linearly in time:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}x^2 &= \frac{d}{dt}\int_{\mathbb{R}} x^2 p(x,t)\,dx \\ &= D\int_{\mathbb{R}} x^2 \frac{\partial^2 p(x,t)}{\partial x^2}\,dx \\ &= D\int_{\mathbb{R}} p(x,t)\,dx = 2D.\end{aligned}$$

Consequently,

$$\mathbb{E}x^2 = 2Dt.$$

Assume now that the initial distribution of the Brownian motion is $\rho_0(x)$. The solution of the Fokker-Planck equation for the Brownian motion with this initial distribution is

$$P_W(y,t) = \int p(y,t|x,0)W(x)\,dy.$$

We can also consider Brownian motion in a bounded domain, with either absorbing, reflecting or periodic boundary conditions. Set $D = \frac{1}{2}$ and consider the Fokker-Planck equation (6.8) on $[0,1]$ with absorbing boundary conditions:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \quad p(0,t) = p(1,t) = 0. \tag{6.9}$$

We look for a solution to this equation in a sine Fourier series:

$$p(x,t) = \sum_{k=1}^{\infty} p_n(t)\sin(n\pi x). \tag{6.10}$$

Notice that the boundary conditions are automatically satisfied. The initial condition it

$$p(x,0) = \delta(x - x_0),$$

where we have assumed that $W_0 = x_0$. The Fourier coefficients of the initial conditions are

$$p_n(0) = 2\int_0^1 \delta(x - x_0)\sin(n\pi x)\,dx = 2\sin(n\pi x_0).$$

We substitute the expansion (6.10) into (6.9) and use the orthogonality properties of the Fourier basis to obtain the equations

$$\dot{p}_n = -\frac{n^2\pi^2}{2}p_n \quad n = 1, 2, \ldots$$

The solution of this equation is

$$p_n(t) = p_n(0)e^{-\frac{n^2\pi^2}{2}t}.$$

Consequently, the transition probability density for the Brownian motion on $[0,1]$ with absorbing boundary conditions is

$$p(x,t|x_0,0) = 2\sum_{n=1}^{\infty} e^{-\frac{n^2\pi^2}{2}t}\sin n\pi x_0 \sin(n\pi x).$$

Notice that

$$\lim_{t\to\infty} p(x,t|x_0,0) = 0.$$

This is not surprising, since all Brownian particles will eventually get absorbed at the boundary.

Exercise 6.2.1. *Consider the Brownian motion with $D = \frac{1}{2}$ on the interval [0,1] with reflecting boundary conditions. The Fokker-Planck equation is*

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \quad \partial_x p(0,t) = \partial_x p(1,t) = 0.$$

i. Find the transition probability density for a Brownian particle starting at x_0.

ii. Show that the process is ergodic and find the invariant distribution $p_s(x)$.

iii. Calculate the stationary autocorrelation function

$$\mathbb{E}(W(t)W(0)) = \int_0^1 \int_0^1 x x_0 p(x,t|x_0,0) p_s(x_0),\, dx dx_0.$$

6.2.2 The Onrstein-Uhlenbeck Process

We set now $a(y,t) = -\alpha y,\ b(y,t) \equiv 2D > 0$ into the Fokker-Planck equation to obtain

$$\frac{\partial p}{\partial t} = \alpha \frac{\partial (yp)}{\partial x} + D\frac{\partial^2 p}{\partial y^2}. \tag{6.11}$$

This is the Fokker-Planck equation for the Ornstein-Uhlenbeck process. The corresponding stochastic differential equation is

$$dX_t = -\alpha X_t + \sqrt{2D} dW_t.$$

So, in addition to Brownian motion there is a linear force pulling the particle towards the origin. This is the reason why the OU process is ergodic.

The transition probability density for the the OU process is

$$p_{OU}(y,t|x,s) = \sqrt{\frac{\alpha}{2\pi D(1-e^{-2\alpha(t-s)})}} \exp\left(-\frac{\alpha(y-e^{-\alpha(t-s)}x)^2}{2D(1-e^{-2\alpha(t-s)})}\right). \tag{6.12}$$

We obtained this formula in Example (4.1.4) (for $\alpha = D = 1$) by using the fact that the OU process can be defined through the a time change of the Brownian motion. We can also derive it by solving equation (6.11).

Exercise 6.2.2. *Solve equation* (6.11) *by taking the Fourier transform, using the method of characteristics for first order PDEs and taking the inverse Fourier transform.*

We set $x = 0,\ s = 0$ in (6.12):

$$p_{OU}(y,t|0,0) = \sqrt{\frac{\alpha}{2\pi D(1-e^{-2\alpha t})}} \exp\left(-\frac{\alpha y^2}{2D(1-e^{-2\alpha t})}\right).$$

We have that

$$\lim_{\alpha\to 0} p_{OU}(y,t) = p_W(y,t).$$

Thus, in the limit where the friction coefficient goes to 0, we recover the transition probability of Brownian motion from the transition probability of the OU processes. Notice also that

$$\lim_{t\to+\infty} p_{OU}(y,t) = \sqrt{\frac{\alpha}{2\pi D}}\exp\left(-\frac{\alpha y^2}{2D}\right).$$

As we have already seen, the Ornstein-Uhlenbeck process is an ergodic Markov process, and its invariant measure is Gaussian.

Using (6.12) we can calculate all moments of the OU process We can calculate all moments of the OU process

$$\mathbb{E}((X_t)^n) \quad = \quad \int\int y^n p(y,t|x,0)p_0(x)\,dxdy,$$

where $p_0(x)$ is the initial distribution. We will calculate the moments by using the Fokker-Planck equation, rather than the explicit formula for the transition probability density.

Define the nth moment of the OU process:

$$M_n = \int_{\mathbb{R}} y^n p(y,t)\,dy, \quad n = 0,1,2,\ldots,$$

where $p(y,t) = \int_{\mathbb{R}} p(y,t|x,0)p_0(x)\,dx$. Let $n = 0$. We integrate the FP equation over $\mathbb{R}$ to obtain:

$$\int \frac{\partial p}{\partial t} = \alpha\int \frac{\partial(yp)}{\partial y} + D\int \frac{\partial^2 p}{\partial y^2} = 0,$$

after an integration by parts and using the fact that $p(x,t)$ decays sufficiently fast at infinity. Consequently:

$$\frac{d}{dt}M_0 = 0 \quad \Rightarrow\ M_0(t) = M_0(0) = 1.$$

In other words:

$$\frac{d}{dt}\|p\|_{L^1(\mathbb{R})} = 0 \quad \Rightarrow\ \int_{\mathbb{R}} p(y,t)\,dy = \int_{\mathbb{R}} p(y,t=0)\,dy = 1.$$

Consequently: **probability is conserved**, as we have already shown. Let $n = 1$. We multiply the FP equation for the OU process by x, integrate over $\mathbb{R}$ and perform and integration by parts to obtain:

$$\frac{d}{dt}M_1 = -\alpha M_1.$$

Consequently, the first moment converges **exponentially fast** to 0:

$$M_1(t) = e^{-\alpha t}M_1(0).$$

Let now $n \geqslant 2$. We multiply the FP equation for the OU process by x^n and integrate by parts (once on the first term on the RHS and twice on the second) to obtain:

$$\frac{d}{dt}\int y^n p = -\alpha n\int y^n p + Dn(n-1)\int y^{n-2}p.$$

Or, equivalently:

$$\frac{d}{dt}M_n = -\alpha n M_n + Dn(n-1)M_{n-2}, \quad n \geqslant 2.$$

This is a first order linear inhomogeneous differential equation. We can solve it using the variation of constants formula:

$$M_n(t) = e^{-\alpha n t}M_n(0) + Dn(n-1)\int_0^t e^{-\alpha n(t-s)}M_{n-2}(s)\,ds. \tag{6.13}$$

We can use this formula, together with the formulas for the first two moments in order to calculate all higher order moments in an iterative way. For example, for $n = 2$ we have

$$\begin{aligned} M_2(t) &= e^{-2\alpha t}M_2(0) + D\int_0^t e^{-2\alpha(t-s)}M_0(s)\,ds \\ &= e^{-2\alpha t}M_2(0) + \frac{D}{2\alpha}e^{-2\alpha t}(e^{2\alpha t}-1) \\ &= \frac{D}{2\alpha} + e^{-2\alpha t}\left(M_2(0) - \frac{D}{2\alpha}\right). \end{aligned}$$

Consequently, the second moment converges exponentially fast to its stationary value $\frac{D}{2\alpha}$. The stationary moments of the OU process are:

$$\begin{aligned} \langle y^n\rangle_{OU} &= \sqrt{\frac{\alpha}{2\pi D}}\int_{\mathbb{R}} y^n e^{-\frac{\alpha y^2}{2D}}\,dx \\ &= \begin{cases} 1.3\ldots(n-1)\left(\frac{D}{\alpha}\right)^{n/2}, & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases} \end{aligned}$$

It is not hard to check that

$$\lim_{t\to\infty} M_n(t) = \langle y^n\rangle_{OU}. \tag{6.14}$$

Furthermore, if the initial conditions of the OU process are stationary, then:

$$M_n(t) = M_n(0) = \langle x^n\rangle_{OU}. \tag{6.15}$$

Exercise 6.2.3. *Use* (6.13) *to show* (6.14) *and* (6.15).

Exercise 6.2.4. *Show that the autocorrelation function of the stationary Ornstein-Uhlenbeck is*

$$\begin{aligned} \mathbb{E}(X_t X_0) &= \int_{\mathbb{R}}\int_{\mathbb{R}} x x_0 p_{OU}(x,t|x_0,0)p_s(x_0)\,dx dx_0 \\ &= \frac{D}{2\alpha}e^{-\alpha|t|}, \end{aligned}$$

where $p_s(x)$ denotes the invariant Gaussian distribution.

6.2.3 The Geometric Brownian Motin

We set $a(x) = \mu x,\ \ b(x) = \frac{1}{2}\sigma^2 x^2$. This is the **geometric Brownian motion**. The corresponding stochastic differential equation is

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t.$$

This equation is one of the basic models of mathematical finance. The coefficient σ is called the **volatility**. The generator of this process is

$$\mathcal{L} = \mu x \frac{\partial}{\partial x} + \frac{\sigma x^2}{2}\frac{\partial^2}{\partial x^2}.$$

Notice that this operator is not uniformly elliptic. The Fokker-Planck equation of the geometric Brownian motion is:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu x) + \frac{\partial^2}{\partial x^2}\left(\frac{\sigma^2 x^2}{2} p\right).$$

We can easily obtain an equation for the nth moment of the geometric Brownian motion:

$$\frac{d}{dt}M_n = \Big(\mu n + \frac{\sigma^2}{2} n(n-1)\Big) M_n, \quad n \geqslant 2.$$

The solution of this equation is

$$M_n(t) = e^{(\mu + (n-1)\frac{\sigma^2}{2})nt} M_n(0), \quad n \geqslant 2$$

and

$$M_1(t) = e^{\mu t} M_1(0).$$

Notice that the nth moment might diverge as $t \to \infty$, depending on the values of μ and σ. Consider for example the second moment and assume that $\mu < 0$. We have

$$M_n(t) = e^{(2\mu + \sigma^2)t} M_2(0),$$

which diverges when $\sigma^2 + 2\mu > 0$.

6.3 The OU Process and Hermite Polynomials

The Ornstein-Uhlenbeck process is one of the few stochastic processes for which we can calculate explicitly the solution of the corresponding SDE, the solution of the Fokker-Planck equation as well as the eigenfunctions of the generator of the process. In this section we will show that the eigenfunctions of the OU process are the Hermite polynomials. We will also study various properties of the generator of the OU process. In the next section we will show that many of the properties of the OU process (ergodicity, self-adjointness of the generator, exponentially fast convergence to equilibrium, real, discrete spectrum) are shared by a large class of diffusion processes, namely those for which the drift term can be written in terms of the gradient of a smooth functions. We will refer to these processes as **gradient flows**.

The generator of the d-dimensional OU process is (we set the drift coefficient equal to 1)

$$\mathcal{L} = -p \cdot \nabla_p + \beta^{-1}\Delta_p \tag{6.16}$$

We have already seen that the OU process is an ergodic Markov process whose unique invariant measure is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ with Gaussian density $\rho \in C^\infty(\mathbb{R}^d)$

$$\rho_\beta(p) = \frac{1}{(2\pi\beta^{-1})^{d/2}} e^{-\beta\frac{|p|^2}{2}}.$$

The natural function space for studying the generator of the OU process is the L^2-space weighted by the invariant measure of the process. This is a separable Hilbert space with norm

$$\|f\|_\rho^2 := \int_{\mathbb{R}^d} f^2 \rho_\beta \, dp.$$

and corresponding inner product

$$(f,h)_\rho = \int_{\mathbb{R}} f h \rho_\beta \, dp.$$

Similarly, we can define weighted L^2-spaced involving derivatives.

Exercise 6.3.1. *Give the definition of the Sobolev space $H^p(\mathbb{R}^d;\rho_\beta)$.*

The reason why this is the right function space is that the generator of the OU process becomes a self-adjoint operator in this space. In fact, $\mathcal{L}$ defined in (6.16) has many nice properties that are summarized in the following proposition.

Proposition 6.3.2. *The operator $\mathcal{L}$ has the following properties:*

- *For every $f,\ h \in C_0^2(\mathbb{R}^d) \cap L^2_\rho(\mathbb{R}^d)$,*

$$(\mathcal{L}f, h)_\rho = (f, \mathcal{L}h)_\rho = -\beta^{-1} \int_{\mathbb{R}^d} \nabla f \cdot \nabla h \rho_\beta \, dp. \tag{6.17}$$

- *$\mathcal{L}$ is a non-positive operator on L^2_ρ.*
- *$\mathcal{L}f = 0$ iff $f \equiv$ const.*
- *For every $f \in C_0^2(\mathbb{R}^d) \cap L^2_\rho(\mathbb{R}^d)$ with $\int f\rho_\beta = 0$,*

$$(-\mathcal{L}f, f)_\rho \geqslant \|f\|_\rho^2 \tag{6.18}$$

Proof. Equation (6.17) follows from an integration by parts:

$$\begin{aligned}
(\mathcal{L}f, h)_\rho &= \int -p\cdot\nabla f h \rho_\beta \, dp + \beta^{-1} \int \Delta f h \rho_\beta \, dp \\
&= \int -p\cdot\nabla f h \rho_\beta \, dp - \beta^{-1} \int \nabla f \cdot \nabla h \rho_\beta \, dp + \int -p\cdot\nabla f h \rho_\beta \, dp \\
&= -\beta^{-1} (\nabla f, \nabla h)_\rho.
\end{aligned}$$

Non-positivity of $\mathcal{L}$ follows from (6.17) upon setting $h = f$:

$$(\mathcal{L}f, f)_\rho = -\beta^{-1} \|\nabla f\|_\rho^2 \leqslant 0.$$

Similarly, multiplying the equation $\mathcal{L}f = 0$ by $f\rho_\beta$, integrating over $\mathbb{R}^d$ and using (6.17) gives

$$\|f\|_\rho = 0,$$

from which we deduce that $f \equiv$ const. The spectral gap follows from (6.17), together with the Poincaré inequality for Gaussian measures:

$$\int_{\mathbb{R}^d} f^2 \rho_\beta \, dp \leqslant \beta^{-1} \int_{\mathbb{R}^d} |\nabla f|^2 \rho_\beta \, dp \tag{6.19}$$

for every $f \in H^1(\mathbb{R}^d; \rho_\beta)$ with $\int f\rho_\beta = 0$. Indeed, upon combining (6.17) with (6.19) we obtain:

$$\begin{aligned} (\mathcal{L}f, f)_\rho &= -\beta^{-1} |\nabla f|_\rho^2 \\ &\leqslant -\|f\|_\rho^2 \end{aligned}$$

□

The spectral gap of the generator of the OU process, which is equivalent to the compactness of its resolvent, implies that $\mathcal{L}$ has discrete spectrum. Furthermore, since it is also a self-adjoint operator, we have that its eigenfunctions form a countable orthonormal basis for the separable Hilbert space L^2_ρ. In fact, we can calculate the eigenvalues and eigenfunctions of the generator of the OU process in one dimension[3]

Theorem 6.3.3. *Consider the eigenvalue problem for the generator of the OU process in one dimension*

$$-\mathcal{L}f_n = \lambda_n f_n. \tag{6.20}$$

Then the eigenvalues of $\mathcal{L}$ are the nonnegative integers:

$$\lambda_n = n, \quad n = 0,\, 1,\, 2,\, \ldots.$$

The corresponding eigenfunctions are the normalized **Hermite polynomials***:*

$$f_n(p) = \frac{1}{\sqrt{n!}} H_n\left(\sqrt{\beta} p\right), \tag{6.21}$$

where

$$H_n(p) = (-1)^n e^{\frac{p^2}{2}} \frac{d^n}{dp^n}\left(e^{-\frac{p^2}{2}}\right). \tag{6.22}$$

Hermite polynomials appear very frequently in applications and they also play a fundamental role in analysis. It is possible to prove that the Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}^d, \rho_\beta)$ without using the fact that they are the eigenfunctions of the a symmetric operator with compact resolvent[4] A typical result along these lines is [24, Lemma 2.3.4] (we use the notation $\rho_1 = \rho$):

[3]The multidimensional problem can be treated similarly by taking tensor products of the eigenfunctions of the one dimensional problem. See Exercise.

[4]In fact, the Poincaré inequality for Gaussian measures can be proved using the fact that that the Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}^d, \rho_\beta)$.

Proposition 6.3.4. *For each $\lambda \in \mathbb{C}$, set*

$$H(p;\lambda) = e^{\lambda p - \frac{\lambda^2}{2}}, \quad p \in \mathbb{R}.$$

Then

$$H(p;\lambda) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} H_n(p), \quad p \in \mathbb{R}, \tag{6.23}$$

where the convergence is both uniform on compact subsets of $\mathbb{R} \times \mathbb{C}$, and for λ's in compact subsets of $\mathbb{C}$, uniform in $L^2(\mathbb{C};\rho)$. In particular, $\{f_n(p) := \frac{1}{\sqrt{n!}} H_n(\sqrt{\beta}p) : n \in \mathbb{N}\}$ is an orthonormal basis in $L^2(\mathbb{C};\rho_\beta)$.

From (6.22) it is clear that H_n is a polynomial of degree n. Furthermore, only odd (even) powers appear in $H_n(p)$ when n is odd (even). Furthermore, the coefficient multiplying p^n in $H_n(p)$ is always 1. The orthonormality of the modified Hermite polynomials $f_n(p)$ defined in (6.21) implies that

$$\int_{\mathbb{R}} f_n(p) f_m(p) \rho_\beta(p)\, dp = \delta_{nm}.$$

The first few Hermite polynomials and the corresponding rescaled/normalized eigenfunctions of the generator of the OU process are:

$$\begin{aligned}
H_0(p) &= 1, & f_0(p) &= 1,\\
H_1(p) &= p, & f_1(p) &= \sqrt{\beta}p,\\
H_2(p) &= p^2 - 1, & f_2(p) &= \frac{\beta}{\sqrt{2}}p^2 - \frac{1}{\sqrt{2}},\\
H_3(p) &= p^3 - 3p, & f_3(p) &= \frac{\beta^{3/2}}{\sqrt{6}}p^3 - \frac{3\sqrt{\beta}}{\sqrt{6}}p\\
H_4(p) &= p^4 - 3p^2 + 3, & f_4(p) &= \frac{1}{\sqrt{24}}\left(\beta^2 p^4 - 3\beta p^2 + 3\right)\\
H_5(p) &= p^5 - 10p^3 + 15p, & f_5(p) &= \frac{1}{\sqrt{120}}\left(\beta^{5/2}p^5 - 10\beta^{3/2}p^3 + 15\beta^{1/2}p\right).
\end{aligned}$$

The proof of Theorem 6.3.3 follows essentially from the properties of the Hermite polynomials. First, notice that by combining (6.21) and (6.23) we obtain

$$H(\sqrt{\beta}p,\lambda) = \sum_{n=0}^{+\infty} \frac{\lambda^n}{\sqrt{n!}} f_n(p)$$

We differentiate this formula with respect to p to obtain

$$\lambda\sqrt{\beta} H(\sqrt{\beta}p,\lambda) = \sum_{n=1}^{+\infty} \frac{\lambda^n}{\sqrt{n!}} \partial_p f_n(p),$$

since $f_0 = 1$. From this equation we obtain

$$\begin{aligned} H(\sqrt{\beta}p, \lambda) &= \sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{\sqrt{\beta}\sqrt{n!}} \partial_p f_n(p) \\ &= \sum_{n=0}^{+\infty} \frac{\lambda^n}{\sqrt{\beta}\sqrt{(n+1)!}} \partial_p f_{n+1}(p) \end{aligned}$$

from which we deduce that

$$\frac{1}{\sqrt{\beta}} \partial_p f_k = \sqrt{k} f_{k-1}. \tag{6.24}$$

Similarly, if we differentiate (6.23) with respect to λ we obtain

$$(p-\lambda)H(p;\lambda) = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} pH_k(p) - \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} H_{k-1}(p) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} H_{k+1}(p)$$

from which we obtain the recurrence relation

$$pH_k = H_{k+1} + kH_{k-1}.$$

Upon rescaling, we deduce that

$$pf_k = \sqrt{\beta^{-1}(k+1)} f_{k+1} + \sqrt{\beta^{-1}k} f_{k-1}. \tag{6.25}$$

We combine now equations (6.24) and (6.25) to obtain

$$\left(\sqrt{\beta}p - \frac{1}{\sqrt{\beta}} \partial_p \right) f_k = \sqrt{k+1} f_{k+1}. \tag{6.26}$$

Now we observe that

$$\begin{aligned} -\mathcal{L} f_n &= \left(\sqrt{\beta}p - \frac{1}{\sqrt{\beta}} \partial_p \right) \frac{1}{\sqrt{\beta}} \partial_p f_n \\ &= \left(\sqrt{\beta}p - \frac{1}{\sqrt{\beta}} \partial_p \right) \sqrt{n} f_{n-1} = n f_n. \end{aligned}$$

The operators $\left(\sqrt{\beta}p - \frac{1}{\sqrt{\beta}} \partial_p \right)$ and $\frac{1}{\sqrt{\beta}} \partial_p$ play the role of **creation** and **annihilation** operators. In fact, we can generate all eigenfunctions of the OU operator from the **ground state** $f_0 = 0$ through a repeated application of the creation operator.

Proposition 6.3.5. *Set $\beta = 1$ and let $a^- = \partial_p$. Then the L^2_ρ-adjoint of a^+ is*

$$a^+ = -\partial_p + p.$$

Then the generator of the OU process can be written in the form

$$\mathcal{L} = -a^+ a^-.$$

Furthermore, a^+ and a^- satisfy the following commutation relation

$$[a^+, a^-] = -1$$

Define now the creation and annihilation operators on $C^1(\mathbb{R})$ by

$$S^+ = \frac{1}{\sqrt{(n+1)}} a^+$$

and

$$S^- = \frac{1}{\sqrt{n}} a^-.$$

Then

$$S^+ f_n = f_{n+1} \quad \textit{and} \quad S^- f_n = f_{n-1}. \tag{6.27}$$

In particular,

$$f_n = \frac{1}{\sqrt{n!}} (a^+)^n 1 \tag{6.28}$$

and

$$1 = \frac{1}{\sqrt{n!}} (a^-)^n f_n. \tag{6.29}$$

Proof. let $f,\ h \in C^1(\mathbb{R}) \cap L^2_\rho$. We calculate

$$\int \partial_p f h \rho = -\int f \partial_p (h\rho) \tag{6.30}$$

$$= \int f\big(-\partial_p + p\big) h\rho. \tag{6.31}$$

Now,

$$-a^+ a^- = -(-\partial_p + p)\partial_p = \partial_p - p\partial_p = \mathcal{L}.$$

Similarly,

$$a^- a^+ = -\partial_p^2 + p\partial_p + 1.$$

and

$$[a^+, a^-] = -1$$

Forumlas (6.27) follow from (6.24) and (6.26). Finally, formulas (6.28) and (6.29) are a consequence of (6.24) and (6.26), together with a simple induction argument. □

Notice that upon using (6.28) and (6.29) and the fact that a^+ is the adjoint of a^- we can easily check the orthonormality of the eigenfunctions:

$$\int f_n f_m \, \rho = \frac{1}{\sqrt{m!}} \int f_n (a^-)^m 1 \, \rho$$

$$= \frac{1}{\sqrt{m!}} \int (a^-)^m f_n \, \rho$$

$$= \int f_{n-m} \, \rho = \delta_{nm}.$$

From the eigenfunctions and eigenvalues of $\mathcal{L}$ we can easily obtain the eigenvalues and eigenfunctions of $\mathcal{L}^*$, the Fokker-Planck operator.

Lemma 6.3.6. *The eigenvalues and eigenfunctions of the Fokker-Planck operator*

$$\mathcal{L}^*\cdot = \partial_p^2 \cdot + \partial_p(p\cdot)$$

are

$$\lambda_n^* = -n, \quad n = 0, 1, s, \ldots \quad \textit{and} \quad f_n^* = \rho f_n.$$

Proof. We have

$$\begin{aligned} \mathcal{L}^*(\rho f_n) &= f_n \mathcal{L}^* \rho + \rho \mathcal{L} f_n \\ &= -n\rho f_n. \end{aligned}$$

□

An immediate corollary of the above calculation is that we can the nth eigenfunction of the Fokker-Planck operator is given by

$$f_n^* = \rho(p)\frac{1}{n!}(a^+)^n 1.$$

Consider now the Fokker-Planck equation with an arbitrary initial condition.

We can relate the generator of the OU process with the Schrödinger operator

6.4 Gradient Flows

Gradient Flows

- Let $V(x) = \frac{1}{2}\alpha x^2$. The generator of the OU process can be written as:

$$\mathcal{L} = -\partial_x V \partial_x + D\partial_x^2.$$

- Consider diffusion processes with a potential $V(x)$, not necessarily quadratic:

$$\mathcal{L} = -\nabla V(x) \cdot \nabla + D\Delta. \tag{6.32}$$

- This is a **gradient flow** perturbed by noise whose strength is $D = k_B T$ where k_B is Boltzmann's constant and T the absolute temperature.

- The corresponding stochastic differential equation is

$$dX_t = -\nabla V(X_t)\, dt + \sqrt{2D}\, dW_t.$$

- The corresponding FP equation is:

$$\frac{\partial p}{\partial t} = \nabla \cdot (\nabla V p) + D\Delta p. \tag{6.33}$$

- It is not possible to calculate the time dependent solution of this equation for an arbitrary potential. We can, however, always calculate the stationary solution.

Definition 6.4.1. *A potential V will be called* **confining** *if* $\lim_{x\to+\infty} V(x) = +\infty$ *and*

$$e^{-\beta V(x)} \in L^1(\mathbb{R}^d). \tag{6.34}$$

for all $\beta \in \mathbb{R}^+$.

Proposition 6.4.2. *Let $V(x)$ be a smooth confining potential. Then the Markov process with generator* (6.32) *is ergodic. The unique invariant distribution is the* **Gibbs distribution**

$$p(x) = \frac{1}{Z} e^{-V(x)/D} \tag{6.35}$$

where the normalization factor Z is the **partition function**

$$Z = \int_{\mathbb{R}^d} e^{-V(x)/D}\, dx.$$

- The fact that the Gibbs distribution is an invariant distribution follows by direct substitution. Uniqueness follows from a PDEs argument (see discussion below).
- It is more convenient to "normalize" the solution of the Fokker-Planck equation wrt the invariant distribution.

Theorem 6.4.3. *Let $p(x,t)$ be the solution of the Fokker-Planck equation* (6.33)*, assume that* (6.34) *holds and let $\rho(x)$ be the Gibbs distribution* (6.35)*. Define $h(x,t)$ through*

$$p(x,t) = h(x,t)\rho(x).$$

Then the function h satisfies the **backward Kolmogorov equation***:*

$$\frac{\partial h}{\partial t} = -\nabla V \cdot \nabla h + D\Delta h, \quad h(x,0) = p(x,0)\rho^{-1}(x). \tag{6.36}$$

Proof. The initial condition follows from the definition of h. We calculate the gradient and Laplacian of p:

$$\nabla p = \rho \nabla h - \rho h D^{-1} \nabla V$$

and

$$\Delta p = \rho \Delta h - 2\rho D^{-1} \nabla V \cdot \nabla h + h D^{-1} \Delta V \rho + h |\nabla V|^2 D^{-2} \rho.$$

We substitute these formulas into the FP equation to obtain

$$\rho \frac{\partial h}{\partial t} \;=\; \rho\Big(- \nabla V \cdot \nabla h + D\Delta h \Big),$$

from which the claim follows. □

- Consequently, in order to study properties of solutions to the FP equation, it is sufficient to study the backward equation (6.36).
- The generator $\mathcal{L}$ is self-adjoint, in the right function space.

- We define the weighted L^2 space L^2_ρ:

$$L^2_\rho = \{f| \int_{\mathbb{R}^d} |f|^2 \rho(x)\, dx < \infty\},$$

- where $\rho(x)$ is the Gibbs distribution. This is a Hilbert space with inner product

$$(f,h)_\rho = \int_{\mathbb{R}^d} fh\rho(x)\, dx.$$

Theorem 6.4.4. *Assume that $V(x)$ is a smooth potential and assume that condition* (6.34) *holds. Then the operator*

$$\mathcal{L} = -\nabla V(x)\cdot\nabla + D\Delta$$

is self-adjoint in L^2_ρ. Furthermore, it is non-positive, its kernel consists of constants.

Proof. Let $f, \in C^2_0(\mathbb{R}^d)$. We calculate

$$\begin{aligned}(\mathcal{L}f,h)_\rho &= \int_{\mathbb{R}^d}(-\nabla V\cdot\nabla + D\Delta)fh\rho\, dx \\ &= \int_{\mathbb{R}^d}(\nabla V\cdot\nabla f)h\rho\, dx - D\int_{\mathbb{R}^d}\nabla f\nabla h\rho\, dx - D\int_{\mathbb{R}^d}\nabla f h\nabla\rho\, dx \\ &= -D\int_{\mathbb{R}^d}\nabla f\cdot\nabla h\rho\, dx,\end{aligned}$$

from which self-adjointness follows. □

If we set $f = h$ in the above equation we get

$$(\mathcal{L}f,f)_\rho = -D\|\nabla f\|^2_\rho,$$

which shows that $\mathcal{L}$ is non-positive.

Clearly, constants are in the null space of $\mathcal{L}$. Assume that $f \in \mathcal{N}(\mathcal{L})$. Then, from the above equation we get

$$0 = -D\|\nabla f\|^2_\rho,$$

and, consequently, f is a constant.

Remark

- The expression $(-\mathcal{L}f,f)_\rho$ is called the **Dirichlet form** of the operator $\mathcal{L}$. In the case of a gradient flow, it takes the form

$$(-\mathcal{L}f,f)_\rho = D\|\nabla f\|^2_\rho. \tag{6.37}$$

- Using the properties of the generator $\mathcal{L}$ we can show that the solution of the Fokker-Planck equation converges to the Gibbs distribution exponentially fast.

- For this we need the following result.

Theorem 6.4.5. *Assume that the potential V satisfies the convexity condition*

$$D^2V \geqslant \lambda I.$$

Then the corresponding Gibbs measure satisfies the Poincaré inequality with constant λ:

$$\int_{\mathbb{R}^d} f\rho = 0 \quad \Rightarrow \quad \|\nabla f\|_\rho \geqslant \sqrt{\lambda}\|f\|_\rho. \tag{6.38}$$

Theorem 6.4.6. *Assume that $p(x,0) \in L^2(e^{V/D})$. Then the solution $p(x,t)$ of the Fokker-Planck equation* (6.33) *converges to the Gibbs distribution exponentially fast:*

$$\|p(\cdot,t) - Z^{-1}e^{-V}\|_{\rho^{-1}} \leqslant e^{-\lambda Dt}\|p(\cdot,0) - Z^{-1}e^{-V}\|_{\rho^{-1}}. \tag{6.39}$$

Proof. We Use (6.36), (6.37) and (6.38) to calculate

$$\begin{aligned} -\frac{d}{dt}\|(h-1)\|_\rho^2 &= -2\left(\frac{\partial h}{\partial t}, h-1\right)_\rho = -2\left(\mathcal{L}h, h-1\right)_\rho \\ &= \left(-\mathcal{L}(h-1), h-1\right)_\rho = 2D\|\nabla(h-1)\|_\rho \\ &\geqslant 2D\lambda\|h-1\|_\rho^2. \end{aligned}$$

Our assumption on $p(\cdot,0)$ implies that $h(\cdot,0) \in L^2_\rho$. Consequently, the above calculation shows that

$$\|h(\cdot,t) - 1\|_\rho \leqslant e^{-\lambda Dt}\|H(\cdot,0) - 1\|_\rho.$$

This, and the definition of h, $p = \rho h$, lead to (6.39). □

- The assumption
$$\int_{\mathbb{R}^d} |p(x,0)|^2 Z^{-1}e^{V/D} < \infty$$
- is very restrictive (think of the case where $V = x^2$).
- The function space $L^2(\rho^{-1}) = L^2(e^{-V/D})$ in which we prove convergence is not the right space to use. Since $p(\cdot,t) \in L^1$, ideally we would like to prove exponentially fast convergence in L^1.
- We can prove convergence in L^1 using the theory of **logarithmic Sobolev inequalities**. In fact, we can also prove convergence in **relative entropy**:
$$H(p|\rho_V) := \int_{\mathbb{R}^d} p\ln\left(\frac{p}{\rho_V}\right)\,dx.$$
- The relative entropy norm controls the L^1 norm:
$$\|\rho_1 - \rho_2\|_{L^1} \leqslant CH(\rho_1|\rho_2)$$
- Using a logarithmic Sobolev inequality, we can prove exponentially fast convergence to equilibrium, assuming only that the relative entropy of the initial conditions is finite.

Theorem 6.4.7. *Let p denote the solution of the Fokker–Planck equation* (6.33) *where the potential is smooth and uniformly convex. Assume that the the initial conditions satisfy*

$$H(p(\cdot,0)|\rho_V) < \infty.$$

Then p converges to the Gibbs distribution exponentially fast in relative entropy:

$$H(p(\cdot,t)|\rho_V) \leqslant e^{-\lambda D t} H(p(\cdot,0)|\rho_V).$$

- Convergence to equilibrium for **kinetic equations**, both linear and non-linear (e.g., the Boltzmann equation) has been studied extensively.
- It has been recognized that the relative entropy plays a very important role.
- For more information see
- *On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis* by P.A. Markowich and C. Villani, 1999.

6.5 Eigenfunction Expansions

- Consider the generator of a gradient stochastic flow with a uniformly convex potential

$$\mathcal{L} = -\nabla V \cdot \nabla + D\Delta. \tag{6.40}$$

- We know that
  i. $\mathcal{L}$ is a non-positive self-adjoint operator on L^2_ρ.
  ii. It has a spectral gap:
  $$(\mathcal{L}f, f)_\rho \leqslant -D\lambda \|f\|^2_\rho$$
  where λ is the Poincaré constant of the potential V.
- The above imply that we can study the spectral problem for $-\mathcal{L}$:

$$-\mathcal{L}f_n = \lambda_n f_n, \quad n = 0, 1, \dots$$

- The operator $-\mathcal{L}$ has real, discrete spectrum with

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots$$

- Furthermore, the eigenfunctions $\{f_j\}_{j=1}^{infty}$ form an orthonormal basis in L^2_ρ: we can express every element of L^2_ρ in the form of a generalized Fourier series:

$$\phi = \sum_{n=0}^{\infty} \phi_n f_n, \quad \phi_n = (\phi, f_n)_\rho \tag{6.41}$$

- with $(f_n, f_m)_\rho = \delta_{nm}$.
- This enables us to solve the time dependent Fokker–Planck equation in terms of an eigenfunction expansion.
- Consider the backward Kolmogorov equation (6.36).
- We assume that the initial conditions $h_0(x) = \phi(x) \in L^2_\rho$ and consequently we can expand it in the form (6.41).
- We look for a solution of (6.36) in the form

$$h(x,t) = \sum_{n=0}^{\infty} h_n(t) f_n(x).$$

- We substitute this expansion into the backward Kolmogorov equation:

$$\frac{\partial h}{\partial t} = \sum_{n=0}^{\infty} \dot{h}_n f_n = \mathcal{L}\left(\sum_{n=0}^{\infty} h_n f_n\right) \tag{6.42}$$

$$= \sum_{n=0}^{\infty} -\lambda_n h_n f_n. \tag{6.43}$$

- We multiply this equation by f_m, integrate wrt the Gibbs measure and use the orthonormality of the eigenfunctions to obtain the sequence of equations

$$\dot{h}_n = -\lambda_n h_n, \quad n = 0, 1,$$

- The solution is

$$h_0(t) = \phi_0, \quad h_n(t) = e^{-\lambda_n t}\phi_n, \ n = 1, 2, \ldots$$

- Notice that

$$\begin{aligned} 1 &= \int_{\mathbb{R}^d} p(x,0)\, dx = \int_{\mathbb{R}^d} p(x,t)\, dx \\ &= \int_{\mathbb{R}^d} h(x,t) Z^{-1} e^{\beta V}\, dx = (h,1)_\rho = (\phi,1)_\rho \\ &= \phi_0. \end{aligned}$$

- Consequently, the solution of the backward Kolmogorov equation is

$$h(x,t) = 1 + \sum_{n=1}^{\infty} e^{-\lambda_n t}\phi_n f_n.$$

- This expansion, together with the fact that all eigenvalues are positive ($n \geqslant 1$), shows that the solution of the backward Kolmogorov equation converges to 1 exponentially fast.
- The solution of the Fokker–Planck equation is

$$p(x,t) = Z^{-1} e^{-\beta V(x)} \left(1 + \sum_{n=1}^{\infty} e^{-\lambda_n t}\phi_n f_n\right).$$

6.6 Self-adjointness

Self-adjointness

- The Fokker–Planck operator of a general diffusion process is not self-adjoint in general.
- In fact, it is self-adjoint **if and only if** the drift term is the gradient of a potential (Nelson, ≈ 1960). This is also true in infinite dimensions (Stochastic PDEs).
- Markov Processes whose generator is a self-adjoint operator are called **reversible**: for all $t \in [0, T]$ X_t and X_{T-t} have the same transition probability (when X_t is stationary).
- Reversibility is equivalent to the invariant measure being a Gibbs measure.
- See *Thermodynamics of the general duffusion process: time-reversibility and entropy production,* H. Qian, M. Qian, X. Tang, J. Stat. Phys., 107, $(5/6)$, 2002, pp. 1129–1141.

6.7 Reduction to a Schrödinger Equation

Reduction to a Schrödinger Equation

Lemma 6.7.1. *The Fokker–Planck operator for a gradient flow can be written in the self-adjoint form*

$$\frac{\partial p}{\partial t} = D\nabla \cdot \left(e^{-V/D} \nabla \left(e^{V/D} p \right) \right). \tag{6.44}$$

Define now $\psi(x,t) = e^{V/2D} p(x,t)$. Then ψ solves the PDE

$$\frac{\partial \psi}{\partial t} = D\Delta\psi - U(x)\psi, \quad U(x) := \frac{|\nabla V|^2}{4D} - \frac{\Delta V}{2}. \tag{6.45}$$

Let $\mathcal{H} := -D\Delta + U$. Then $\mathcal{L}^*$ and $\mathcal{H}$ have the same eigenvalues. The nth eigenfunction ϕ_n of $\mathcal{L}^*$ and the nth eigenfunction ψ_n of $\mathcal{H}$ are associated through the transformation

$$\psi_n(x) = \phi_n(x) \exp\left(\frac{V(x)}{2D} \right).$$

Remarks 6.7.2. *i. From equation* (6.44) *shows that the FP operator can be written in the form*

$$\mathcal{L}^* \cdot = D\nabla \cdot \left(e^{-V/D} \nabla \left(e^{V/D} \cdot \right) \right).$$

ii. The operator that appears on the right hand side of eqn. (6.45) *has the form of a* **Schrödinger operator***:*

$$-\mathcal{H} = -D\Delta + U(x).$$

iii. The spectral problem for the FP operator can be transformed into the spectral problem for a Schrödinger operator. We can thus use all the available results from quantum mechanics to study the FP equation and the associated SDE.

iv. In particular, the weak noise asymptotics $D \ll 1$ is equivalent to the semiclassical approximation from quantum mechanics.

Proof. We calculate

$$\begin{aligned} D\nabla \cdot \left(e^{-V/D} \nabla \left(e^{V/D} f \right) \right) &= D\nabla \cdot \left(e^{-V/D} \left(D^{-1} \nabla V f + \nabla f \right) e^{V/D} \right) \\ &= \nabla \cdot (\nabla V f + D \nabla f) = \mathcal{L}^* f. \end{aligned}$$

Consider now the eigenvalue problem for the FP operator:

$$-\mathcal{L}^* \phi_n = \lambda_n \phi_n.$$

Set $\phi_n = \psi_n \exp\left(-\frac{1}{2D} V\right)$. We calculate $-\mathcal{L}^* \phi_n$:

$$\begin{aligned} -\mathcal{L}^* \phi_n &= -D\nabla \cdot \left(e^{-V/D} \nabla \left(e^{V/D} \psi_n e^{-V/2D} \right) \right) \\ &= -D\nabla \cdot \left(e^{-V/D} \left(\nabla \psi_n + \frac{\nabla V}{2D} \psi_n \right) e^{V/2D} \right) \\ &= \left(-D\Delta \psi_n + \left(-\frac{|\nabla V|^2}{4D} + \frac{\Delta V}{2D} \right) \psi_n \right) e^{-V/2D} = e^{-V/2D} \mathcal{H} \psi_n. \end{aligned}$$

From this we conclude that $e^{-V/2D} \mathcal{H} \psi_n = \lambda_n \psi_n e^{-V/2D}$ from which the equivalence between the two eigenvalue problems follows. □

Remarks 6.7.3. *i. We can rewrite the Schrödinger operator in the form*

$$\mathcal{H} = D\mathcal{A}^* \mathcal{A}, \quad \mathcal{A} = \nabla + \frac{\nabla U}{2D}, \quad \mathcal{A}^* = -\nabla + \frac{\nabla U}{2D}.$$

ii. These are **creation** *and* **annihilation** *operators. They can also be written in the form*

$$\mathcal{A} \cdot = e^{-U/2D} \nabla \left(e^{U/2D} \cdot \right), \quad \mathcal{A}^* \cdot = e^{U/2D} \nabla \left(e^{-U/2D} \cdot \right)$$

iii. The forward the backward Kolmogorov operators have the same eigenvalues. Their eigenfunctions are related through

$$\phi_n^B = \phi_n^F \exp\left(-V/D\right),$$

where ϕ_n^B and ϕ_n^F denote the eigenfunctions of the backward and forward operators, respectively.

6.8 The Klein-Kramers-Chandrasekhar Equation

- Consider a diffusion process in two dimensions for the variables q (position) and momentum p. The generator of this Markov process is

$$\mathcal{L} = p \cdot \nabla_q - \nabla_q V \nabla_p + \gamma(-p\nabla_p + D\Delta_p). \tag{6.46}$$

- The $L^2(dpdq)$-adjoint is

$$\mathcal{L}^*\rho = -p \cdot \nabla_q \rho - \nabla_q V \cdot \nabla_p \rho + \gamma \left(\nabla_p (p\rho) + D\Delta_p \rho \right).$$

- The corresponding FP equation is:

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p.$$

- The corresponding stochastic differential equations is the **Langevin equation**

$$\ddot{X}_t = -\nabla V(X_t) - \gamma \dot{X}_t + \sqrt{2\gamma D}\dot{W}_t. \tag{6.47}$$

- This is Newton's equation perturbed by dissipation and noise.
- The Klein-Kramers-Chandrasekhar equation was first derived by Kramers in 1923 and was studied by Kramers in his famous paper "Brownian motion in a field of force and the diffusion model of chemical reactions", Physica 7(1940), pp. 284-304.
- Notice that $\mathcal{L}^*$ is not a uniformly elliptic operator: there are second order derivatives only with respect to p and not q. This is an example of a degenerate elliptic operator. It is, however, **hypoelliptic**: we can still prove existence and uniqueness of solutions for the FP equation, and obtain estimates on the solution.
- It is not possible to obtain the solution of the FP equation for an arbitrary potential.
- We can calculate the (unique normalized) solution of the stationary Fokker-Planck equation.

Theorem 6.8.1. *Let $V(x)$ be a smooth confining potential. Then the Markov process with generator* (8.6) *is ergodic. The unique invariant distribution is the* **Maxwell-Boltzmann distribution**

$$\rho(p,q) = \frac{1}{Z} e^{-\beta H(p,q)} \tag{6.48}$$

where

$$H(p,q) = \frac{1}{2}\|p\|^2 + V(q)$$

is the **Hamiltonian**, $\beta = (k_B T)^{-1}$ *is the* **inverse temperature** *and the normalization factor Z is the* **partition function**

$$Z = \int_{\mathbb{R}^{2d}} e^{-\beta H(p,q)}\, dpdq.$$

- It is possible to obtain rates of convergence in either a weighted L^2-norm or the relative entropy norm.

$$H(p(\cdot,t)|\rho) \leqslant Ce^{-\alpha t}.$$

- The proof of this result is very complicated, since the generator $\mathcal{L}$ is degenerate and non-selfadjoint.
- See F. Herau and F. Nier, *Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential*, Arch. Ration. Mech. Anal., 171(2),(2004), 151–218.

Let $\rho(q,p,t)$ be the solution of the Kramers equation and let $\rho_\beta(q,p)$ be the Maxwell-Boltzmann distribution. We can write

$$\rho(q,p,t) = h(q,p,t)\rho_\beta(q,p),$$

where $h(q,p,t)$ solves the equation

$$\frac{\partial h}{\partial t} = -\mathcal{A}h + \gamma S h \tag{6.49}$$

where

$$\mathcal{A} = p\cdot\nabla_q - \nabla_q V\cdot\nabla_p, \quad \mathcal{S} = -p\cdot\nabla_p + \beta^{-1}\Delta_p.$$

The operator $\mathcal{A}$ is antisymmetric in $L^2_\rho := L^2(\mathbb{R}^{2d};\rho_\beta(q,p))$, whereas $\mathcal{S}$ is symmetric.

Let $X_i := -\frac{\partial}{\partial p_i}$. The L^2_ρ-adjoint of X_i is

$$X_i^* = -\beta p_i + \frac{\partial}{\partial p_i}.$$

We have that

$$\mathcal{S} = \beta^{-1}\sum_{i=1}^d X_i^* X_i.$$

Consequently, the generator of the Markov process $\{q(t),\, p(t)\}$ can be written in Hörmander's "sum of squares" form:

$$\mathcal{L} = \mathcal{A} + \gamma\beta^{-1}\sum_{i=1}^d X_i^* X_i. \tag{6.50}$$

We calculate the commutators between the vector fields in (6.50):

$$[\mathcal{A}, X_i] = \frac{\partial}{\partial q_i}, \quad [X_i, X_j] = 0, \quad [X_i, X_j^*] = \beta\delta_{ij}.$$

Consequently,

$$\text{Lie}(X_1,\dots X_d, [\mathcal{A},X_1],\dots[\mathcal{A},X_d]) = \text{Lie}(\nabla_p,\nabla_q)$$

which spans $\mathbb{T}_{p,q}\mathbb{R}^{2d}$ for all $p,\, q\in\mathbb{R}^d$. This shows that the generator $\mathcal{L}$ is a hypoelliptic operator.

Let now $Y_i = -\frac{\partial}{\partial p_i}$ with L^2_ρ-adjoint $Y_i^* = \frac{\partial}{\partial q_i} - \beta\frac{\partial V}{\partial q_i}$. We have that

$$X_i^* Y_i - Y_i^* X_i = \beta\left(p_i\frac{\partial}{\partial q_i} - \frac{\partial V}{\partial q_i}\frac{\partial}{\partial p_i}\right).$$

Consequently, the generator can be written in the form

$$\mathcal{L} = \beta^{-1}\sum_{i=1}^d \left(X_i^* Y_i - Y_i^* X_i + \gamma X_i^* X_i\right). \tag{6.51}$$

Notice also that

$$\mathcal{L}_V := -\nabla_q V\nabla_q + \beta^{-1}\Delta_q = \beta^{-1}\sum_{i=1}^d Y_i^* Y_i.$$

- The phase-space Fokker-Planck equation can be written in the form

$$\frac{\partial \rho}{\partial t} + p \cdot \nabla_q \rho - \nabla_q V \cdot \nabla_p \rho = Q(\rho, f_B)$$

- where the *collision operator* has the form

$$Q(\rho, f_B) = D\nabla \cdot \left(f_B \nabla \left(f_B^{-1} \rho \right) \right).$$

- The Fokker-Planck equation has a similar structure to the Boltzmann equation (the basic equation in the kinetic theory of gases), with the difference that the collision operator for the FP equation is linear.

- Convergence of solutions of the Boltzmann equation to the Maxwell-Boltzmann distribution has also been proved. See

- L. Desvillettes and C. Villani: On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. Invent. Math. 159, 2 (2005), 245-316.

- We can study the backward and forward Kolmogorov equations for (9.7) by expanding the solution with respect to the Hermite basis.

- We consider the problem in 1d. We set $D = 1$. The generator of the process is:

$$\begin{aligned} \mathcal{L} &= p\partial_q - V'(q)\partial_p + \gamma\left(-p\partial_p + \partial_p^2\right). \\ &=: \mathcal{L}_1 + \gamma \mathcal{L}_0, \end{aligned}$$

- where

$$\mathcal{L}_0 := -p\partial_p + \partial_p^2 \quad \text{and} \quad \mathcal{L}_1 := p\partial_q - V'(q)\partial_p.$$

- The backward Kolmogorov equation is

$$\frac{\partial h}{\partial t} = \mathcal{L} h. \tag{6.52}$$

- The solution should be an element of the weighted L^2-space

$$L^2_\rho = \left\{ f | \int_{\mathbb{R}^2} |f|^2 Z^{-1} e^{-\beta H(p,q)}\, dpdq < \infty \right\}.$$

- We notice that the invariant measure of our Markov process is a product measure:

$$e^{-\beta H(p,q)} = e^{-\beta \frac{1}{2}|p|^2} e^{-\beta V(q)}.$$

- The space $L^2(e^{-\beta\frac{1}{2}|p|^2}\, dp)$ is spanned by the Hermite polynomials. Consequently, we can expand the solution of (6.52) into the basis of Hermite basis:

$$h(p,q,t) = \sum_{n=0}^{\infty} h_n(q,t) f_n(p), \tag{6.53}$$

- where $f_n(p) = 1/\sqrt{n!}H_n(p)$.

- Our plan is to substitute (6.53) into (6.52) and obtain a sequence of equations for the coefficients $h_n(q,t)$.

- We have:

$$\mathcal{L}_0 h = \mathcal{L}_0 \sum_{n=0}^{\infty} h_n f_n = -\sum_{n=0}^{\infty} n h_n f_n$$

- Furthermore

$$\mathcal{L}_1 h = -\partial_q V \partial_p h + p\partial_q h.$$

- We calculate each term on the right hand side of the above equation separately. For this we will need the formulas

$$\partial_p f_n = \sqrt{n} f_{n-1} \quad \text{and } p f_n = \sqrt{n} f_{n-1} + \sqrt{n+1} f_{n+1}.$$

$$\begin{aligned} p\partial_q h &= p\partial_q \sum_{n=0}^{\infty} h_n f_n = p\partial_p h_0 + \sum_{n=1}^{\infty} \partial_q h_n p f_n \\ &= \partial_q h_0 f_1 + \sum_{n=1}^{\infty} \partial_q h_n \left(\sqrt{n} f_{n-1} + \sqrt{n+1} f_{n+1}\right) \\ &= \sum_{n=0}^{\infty} (\sqrt{n+1}\partial_q h_{n+1} + \sqrt{n}\partial_q h_{n-1}) f_n \end{aligned}$$

- with $h_{-1} \equiv 0$.

- Furthermore

$$\begin{aligned} \partial_q V \partial_p h &= \sum_{n=0}^{\infty} \partial_q V h_n \partial_p f_n = \sum_{n=0}^{\infty} \partial_q V h_n \sqrt{n} f_{n-1} \\ &= \sum_{n=0}^{\infty} \partial_q V h_{n+1} \sqrt{n+1} f_n. \end{aligned}$$

- Consequently:

$$\begin{aligned} \mathcal{L} h &= \mathcal{L}_1 + \gamma \mathcal{L}_1 h \\ &= \sum_{n=0}^{\infty} \big(-\gamma n h_n + \sqrt{n+1}\partial_q h_{n+1} \\ &\quad + \sqrt{n}\partial_q h_{n-1} + \sqrt{n+1}\partial_q V h_{n+1} \big) f_n \end{aligned}$$

- Using the orthonormality of the eigenfunctions of $\mathcal{L}_0$ we obtain the following set of equations which determine $\{h_n(q,t)\}_{n=0}^{\infty}$.

$$\begin{aligned}\dot{h}_n &= -\gamma n h_n + \sqrt{n+1}\partial_q h_{n+1} \\ &\quad +\sqrt{n}\partial_q h_{n-1} + \sqrt{n+1}\partial_q V h_{n+1}, \quad n = 0, 1, \ldots\end{aligned}$$

- This is set of equations is usually called the **Brinkman hierarchy** (1956).
- We can use this approach to develop a numerical method for solving the Klein-Kramers equation.
- For this we need to expand each coefficient h_n in an appropriate basis with respect to q.
- Obvious choices are other the Hermite basis (polynomial potentials) or the standard Fourier basis (periodic potentials).
- We will do this for the case of periodic potentials.
- The resulting method is usually called the **continued fraction expansion**. See Risken (1989).
- The Hermite expansion of the distribution function wrt to the velocity is used in the study of various kinetic equations (including the Boltzmann equation). It was initiated by Grad in the late 40's.
- It quite often used in the approximate calculation of transport coefficients (e.g. diffusion coefficient).
- This expansion can be justified rigorously for the Fokker-Planck equation. See
- J. Meyer and J. Schröter, *Comments on the Grad Procedure for the Fokker-Planck Equation*, J. Stat. Phys. 32(1) pp.53-69 (1983).
- This expansion can also be used in order to solve the Poisson equation $-\mathcal{L}\phi = f(p,q)$. See G.A. Pavliotis and T. Vogiannou *Diffusive Transport in Periodic Potentials: Underdamped dynamics*, Fluct. Noise Lett., 8(2) L155-L173 (2008).

6.9 Brownian Motion in a Harmonic Potential

There are very few potentials for which we can solve the Langevin equation or to calculate the eigenvalues and eigenfunctions of the generator of the Markov process $\{q(t),\, p(t)\}$. One case where we can calculate everything explicitly is that of a Brownian particle in a quadratic (harmonic) potential

$$V(q) = \frac{1}{2}\omega_0^2 q^2. \tag{6.54}$$

The Langevin equation is

$$\ddot{q} = -\omega_0^2 q - \gamma\dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W} \tag{6.55}$$

or

$$\ddot{q} = p, \quad \dot{p} = -\omega_0^2 q - \gamma p + \sqrt{2\gamma\beta^{-1}}\dot{W}. \tag{6.56}$$

This is a linear equation that can be solved explicitly. Rather than doing this, we will calculate the eigenvalues and eigenfunctions of the generator, which takes the form

$$\mathcal{L} = p\partial_q - \omega_0^2 q\partial_p + \gamma(-p\partial_p + \beta^{-1}\partial_p^2). \tag{6.57}$$

The Fokker-Planck operator is

$$\mathcal{L} = p\partial_q - \omega_0^2 q\partial_p + \gamma(-p\partial_p + \beta^{-1}\partial_p^2). \tag{6.58}$$

The process $\{q(t),\, p(t)\}$ is an ergodic Markov process with Gaussian invariant measure

$$\rho_\beta(q,p)\, dqdp = \frac{\beta\omega_0}{2\pi} e^{-\frac{\beta}{2}p^2 - \frac{\beta\omega_0^2}{q^2}}. \tag{6.59}$$

For the calculation of the eigenvalues and eigenfunctions of the operator $\mathcal{L}$ it is convenient to introduce creation and annihilation operator in both the position and momentum variables. We set

$$a^- = \beta^{-1/2}\partial_p, \quad a^+ = -\beta^{-1/2}\partial_p + \beta^{1/2}p \tag{6.60}$$

and

$$b^- = \omega_0^{-1}\beta^{-1/2}\partial_q, \quad b^+ = -\omega_0^{-1}\beta^{-1/2}\partial_q + \omega_0\beta^{1/2}p. \tag{6.61}$$

We have that

$$a^+a^- = -\beta^{-1}\partial_p^2 + p\partial_p$$

and

$$b^+b^- = -\beta^{-1}\partial_q^2 + q\partial_q$$

Consequently, the operator

$$\widehat{\mathcal{L}} = -a^+a^- - b^+b^- \tag{6.62}$$

is the generator of the OU process in two dimensions.

Exercise 6.9.1. *Calculate the eigenvalues and eigenfunctions of $\widehat{\mathcal{L}}$. Show that there exists a transformation that transforms $\widehat{\mathcal{L}}$ into the Schr"odinger operator of the two-dimensional quantum harmonic oscillator.*

Exercise 6.9.2. *Show that the operators $a^\pm$, $b^\pm$ satisfy the commutation relations*

$$[a^+, a^-] = -1, \tag{6.63a}$$

$$[b^+, b^-] = -1, \tag{6.63b}$$

$$[a^\pm, b^\pm] = 0. \tag{6.63c}$$

Using now the operators $a^\pm$ and $b^\pm$ we can write the generator $\mathcal{L}$ in the form

$$\mathcal{L} = -\gamma a^+a^- - \omega_0(b^+a^- - a^+b^-), \tag{6.64}$$

which is a particular case of (6.51). In order to calculate the eigenvalues and eigenfunctions of (6.64) we need to make an appropriate change of variables in order to bring the operator $\mathcal{L}$ into the "decoupled" form (6.62). Clearly, this is a linear transformation and can be written in the form

$$\mathbf{Y} = A\mathbf{X}$$

where $\mathbf{X} = (q, p)$ for some 2×2 matrix A. It is somewhat easier to make this change of variables at the level of the creation and annihilation operators. In particular, our goal is to find first order differential operators $c^{\pm}$ and $d^{\pm}$ so that the operator (6.64) becomes

$$\mathcal{L} = -Cc^+c^- - Dd^+d^- \tag{6.65}$$

for some appropriate constants C and D. Since our goal is, essentially, to map $\mathcal{L}$ to the two-dimensional OU process, we require that that the operators $c^{\pm}$ and $d^{\pm}$ satisfy the **canonical commutation relations**

$$[c^+, c^-] = -1, \tag{6.66a}$$

$$[d^+, d^-] = -1, \tag{6.66b}$$

$$[c^{\pm}, d^{\pm}] = 0. \tag{6.66c}$$

The operators $c^{\pm}$ and $d^{\pm}$ should be given as linear combinations of the old operators $a^{\pm}$ and $b^{\pm}$. From the structure of the generator $\mathcal{L}$ (6.64), the decoupled form (6.65) and the commutation relations (6.66) and (6.63) we conclude that $c^{\pm}$ and $d^{\pm}$ should be of the form

$$c^+ = \alpha_{11}a^+ + \alpha_{12}b^+, \tag{6.67a}$$

$$c^- = \alpha_{21}a^- + \alpha_{22}b^-, \tag{6.67b}$$

$$d^+ = \beta_{11}a^+ + \beta_{12}b^+, \tag{6.67c}$$

$$d^- = \beta_{21}a^- + \beta_{22}b^-. \tag{6.67d}$$

Notice that the $c-$ and d^- **are not** the adjoints of c^+ and d^+. If we substitute now these equations into (6.65) and equate it with (6.64) and into the commutation relations (6.66) we obtain a system of equations for the coefficients $\{\alpha_{ij}\}$, $\{\beta_{ij}\}$. In order to write down the formulas for these coefficients it is convenient to introduce the eigenvalues of the deterministic problem

$$\ddot{q} = -\gamma\dot{q} - \omega_0^2 q.$$

The solution of this equation is

$$q(t) = C_1 e^{-\lambda_1 t} + C_2 e^{-\lambda_2 t}$$

with

$$\lambda_{1,2} = \frac{\gamma \pm \delta}{2}, \quad \delta = \sqrt{\gamma^2 - 4\omega_0^2}. \tag{6.68}$$

The eigenvalues satisfy the relations

$$\lambda_1 + \lambda_2 = \gamma, \quad \lambda_1 - \lambda_2 = \delta, \lambda_1\lambda_2 = \omega_0^2. \tag{6.69}$$

Proposition 6.9.3. *Let $\mathcal{L}$ be the generator* (6.64) *and let $c^{\pm}$, d^{pm} be the operators*

$$c^+ = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^+ + \sqrt{\lambda_2}b^+\right), \tag{6.70a}$$

$$c^- = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^- - \sqrt{\lambda_2}b^-\right), \tag{6.70b}$$

$$d^+ = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_2}a^+ + \sqrt{\lambda_1}b^+\right), \tag{6.70c}$$

$$d^- = \frac{1}{\sqrt{\delta}}\left(-\sqrt{\lambda_2}a^- + \sqrt{\lambda_1}b^-\right). \tag{6.70d}$$

Then $c^\pm$, $d^\pm$ satisfy the canonical commutation relations (6.66) *as well as*

$$[\mathcal{L}, c^\pm] = -\lambda_1 c^\pm, \quad [\mathcal{L}, d^\pm] = -\lambda_2 d^\pm. \tag{6.71}$$

Furthermore, the operator $\mathcal{L}$ can be written in the form

$$\mathcal{L} = -\lambda_1 c^+ c^- - \lambda_2 d^+ d^-. \tag{6.72}$$

Proof. first we check the commutation relations:

$$\begin{aligned}
[c^+, c^-] &= \frac{1}{\delta}\left(\lambda_1[a^+, a^-] - \lambda_2[b^+, b^-]\right) \\
&= \frac{1}{\delta}(-\lambda_1 + \lambda_2) = -1.
\end{aligned}$$

Similarly,

$$\begin{aligned}
[d^+, d^-] &= \frac{1}{\delta}\left(-\lambda_2[a^+, a^-] + \lambda_1[b^+, b^-]\right) \\
&= \frac{1}{\delta}(\lambda_2 - \lambda_1) = -1.
\end{aligned}$$

Clearly, we have that

$$[c^+, d^+] = [c^-, d^-] = 0.$$

Furthermore,

$$\begin{aligned}
[c^+, d^-] &= \frac{1}{\delta}\left(-\sqrt{\lambda_1\lambda_2}[a^+, a^-] + \sqrt{\lambda_1\lambda_2}[b^+, b^-]\right) \\
&= \frac{1}{\delta}(-\sqrt{\lambda_1\lambda_2} + -\sqrt{\lambda_1\lambda_2}) = 0.
\end{aligned}$$

Finally:

$$\begin{aligned}
[\mathcal{L}, c^+] &= -\lambda_1 c^+ c^- c^+ + \lambda_1 c^+ c^+ c^- \\
&= -\lambda_1 c^+(1 + c^+ c^-) + \lambda_1 c^+ c^+ c^- \\
&= -\lambda_1 c^+(1 + c^+ c^-) + \lambda_1 c^+ c^+ c^- \\
&= -\lambda_1 c^+,
\end{aligned}$$

and similarly for the other equations in (6.71). Now we calculate

$$\begin{aligned}
\mathcal{L} &= -\lambda_1 c^+ c^- - \lambda_2 d^+ d^- \\
&= -\frac{\lambda_2^2 - \lambda_1^2}{\delta} a^+ a^- + 0 b^+ b^- + \frac{\sqrt{\lambda_1\lambda_2}}{\delta}(\lambda_1 - \lambda_2) a^+ b^- + \frac{1}{\delta}\sqrt{\lambda_1\lambda_2}(-\lambda_1 + \lambda_2) b^+ a^- \\
&= -\gamma a^+ a^- - \omega_0(b^+ a^- - a^+ b^-),
\end{aligned}$$

which is precisely (6.64). In the above calculation we used (6.69). □

Using now (6.72) we can readily obtain the eigenvalues and eigenfunctions of $\mathcal{L}$. From our experience with the two-dimensional OU processes (or, the Schrödinger operator for the two-dimensional quantum harmonic oscillator), we expect that the eigenfunctions should be tensor products of Hermite polynomials. Indeed, we have the following, which is the main result of this section.

Theorem 6.9.4. *The eigenvalues and eigenfunctions of the generator of the Markov process* $\{q,\, p\}$ (6.56) *are*

$$\lambda_{nm} = \lambda_1 n + \lambda_2 m = \frac{1}{2}\gamma(n+m) + \frac{1}{2}\delta(n-m), \quad n, m = 0,\, 1,\, \ldots \tag{6.73}$$

and

$$\phi_{nm}(q,p) = \frac{1}{\sqrt{n!m!}}(c^+)^n (d^+)^m 1, \quad n, m = 0,\, 1,\, \ldots \tag{6.74}$$

Proof. We have

$$\begin{aligned}
[\mathcal{L}, (c^+)^2] &= \mathcal{L}(c^+)^2 - (c^+)^2\mathcal{L} \\
&= (c^+\mathcal{L} - \lambda_1 c^+)c^+ - c^+(\mathcal{L}c^+ + \lambda_1 c^+) \\
&= -2\lambda_1 (c^+)^2
\end{aligned}$$

and similarly $[\mathcal{L}, (d^+)^2] = -2\lambda_1 (c^+)^2$. A simple induction argument now shows that (see Exercise 6.9.5)

$$[\mathcal{L}, (c^+)^n] = -n\lambda_1 (c^+)^n \quad \text{and} \quad [\mathcal{L}, (d^+)^m] = -m\lambda_1 (d^+)^m. \tag{6.75}$$

We use (6.75) to calculate

$$\begin{aligned}
&\mathcal{L}(c^+)^n (d^+)^n \mathbf{1} \\
&\qquad = (c^+)^n \mathcal{L}(d^+)^m \mathbf{1} - n\lambda_1 (c^+)^n (d^+ m)\mathbf{1} \\
&\qquad = (c^+)^n (d^+)^m \mathcal{L}\mathbf{1} - m\lambda_2 (c^+)^n (d^+ m)\mathbf{1} - n\lambda_1 (c^+)^n (d^+ m)\mathbf{1} \\
&\qquad = -n\lambda_1 (c^+)^n (d^+ m)\mathbf{1} - m\lambda_2 (c^+)^n (d^+ m)\mathbf{1}
\end{aligned}$$

from which (6.73) and (6.74) follow. □

Exercise 6.9.5. *Show that*

$$[\mathcal{L}, (c^\pm)^n] = -n\lambda_1 (c^\pm)^n, \quad [\mathcal{L}, (d^\pm)^n] = -n\lambda_1 (d^\pm)^n, \quad [c^-, (c^+)^n] = n(c^+)^{n-1}, \quad [d^-, (d^+)^n] = n(d^+)^{n-1}. \tag{6.76}$$

Remark 6.9.6. *In terms of the operators* $a^\pm$, $b^\pm$ *the eigenfunctions of* $\mathcal{L}$ *are*

$$\phi_{nm} = \sqrt{n!m!}\,\delta^{-\frac{n+m}{2}} \lambda_1^{n/2} \lambda_2^{m/2} \sum_{\ell=0}^{n} \sum_{k=0}^{m} \frac{1}{k!(m-k)!\ell!(n-\ell)!} \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{k-\ell}{2}} (a^+)^{n+m-k-\ell} (b^+)^{\ell+k} 1.$$

The first few eigenfunctions are

$$\phi_{00} = 1.$$

$$\phi_{10} = \frac{\sqrt{\beta}\left(\sqrt{\lambda_1}p + \sqrt{\lambda_2}\omega_0 q\right)}{\sqrt{\delta}}.$$

$$\phi_{01} = \frac{\sqrt{\beta}\left(\sqrt{\lambda_2}p + \sqrt{\lambda_1}\omega_0 q\right)}{\sqrt{\delta}}$$

$$\phi_{11} = \frac{-2\sqrt{\lambda_1}\sqrt{\lambda_2} + \sqrt{\lambda_1}\beta\, p^2\sqrt{\lambda_2} + \beta\, p\lambda_1\omega_0 q + \omega_0\beta\, q\lambda_2 p + \sqrt{\lambda_2}{\omega_0}^2\beta\, q^2\sqrt{\lambda_1}}{\delta}.$$

$$\phi_{20} = \frac{-\lambda_1 + \beta\, p^2\lambda_1 + 2\sqrt{\lambda_2}\beta\, p\sqrt{\lambda_1}\omega_0 q - \lambda_2 + {\omega_0}^2\beta\, q^2\lambda_2}{\sqrt{2}\delta}.$$

$$\phi_{02} = \frac{-\lambda_2 + \beta\, p^2\lambda_2 + 2\sqrt{\lambda_2}\beta\, p\sqrt{\lambda_1}\omega_0 q - \lambda_1 + {\omega_0}^2\beta\, q^2\lambda_1}{\sqrt{2}\delta}.$$

Notice that the eigenfunctions are not orthonormal.

As we already know, the first eigenvalue, corresponding to the constant eigenfunction, is 0:

$$\lambda_{00} = 0.$$

Notice that the operator $\mathcal{L}$ is not self-adjoint and consequently, we do not expect its eigenvalues to be real. Indeed, whether the eigenvalues are real or not depends on the sign of the discriminant $\Delta = \gamma^2 - 4\omega_0^2$. In the **underdamped** regime, $\gamma < 2\omega_0$ the eigenvalues are complex:

$$\lambda_{nm} = \frac{1}{2}\gamma(n+m) + \frac{1}{2}i\sqrt{-\gamma^2 + 4\omega_0^2}(n-m), \quad \gamma < 2\omega_0.$$

This it to be expected, since the underdamped regime the dynamics is dominated by the deterministic Hamiltonian dynamics that give rise to the antisymmetric Liouville operator. We set $\omega = \sqrt{(4\omega_0^2 - \gamma^2)}$, i.e. $\delta = 2i\omega$. The eigenvalues can be written as

$$\lambda_{nm} = \frac{\gamma}{2}(n+m) + i\omega(n-m).$$

In Figure 6.9 we present the first few eigenvalues of $\mathcal{L}$ in the underdamped regime. The eigenvalues are contained in a cone on the right half of the complex plane. The cone is determined by

$$\lambda_{n0} = \frac{\gamma}{2}n + i\omega n \quad \text{and} \quad \lambda_{0m} = \frac{\gamma}{2}m - i\omega m.$$

The eigenvalues along the diagonal are real:

$$\lambda_{nn} = \gamma n.$$

On the other hand, in the **overdamped** regime, $\gamma \geqslant 2\omega_0$ all eigenvalues are real:

$$\lambda_{nm} = \frac{1}{2}\gamma(n+m) + \frac{1}{2}\sqrt{\gamma^2 - 4\omega_0^2}(n-m), \quad \gamma \geqslant 2\omega_0.$$

In fact, in the overdamped limit $\gamma \to +\infty$ (which we will study in Chapter 8), the eigenvalues of the generator $\mathcal{L}$ converge to the eigenvalues of the generator of the OU process:

$$\lambda_{nm} = \gamma n + \frac{\omega_0^2}{\gamma}(n-m) + O(\gamma^{-3}).$$

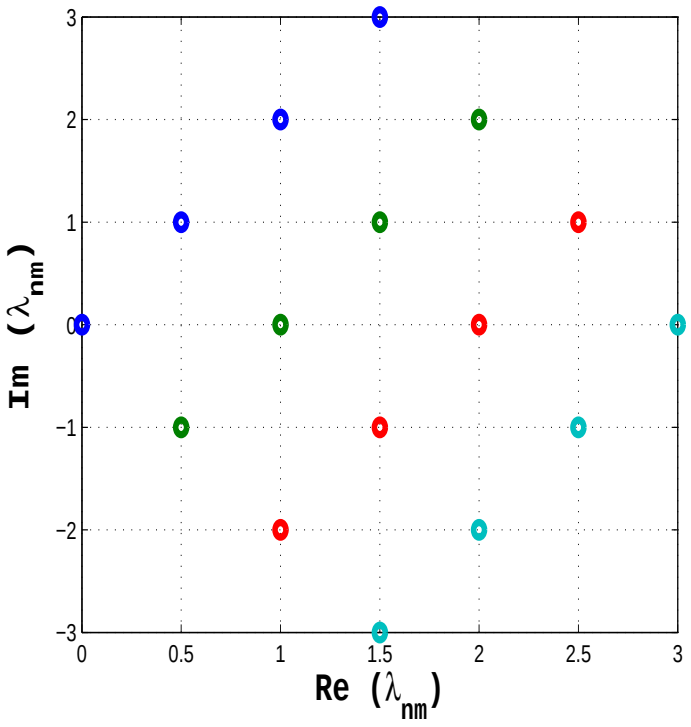


Figure 6.1: First few eigenvalues of $\mathcal{L}$ for $\gamma = \omega = 1$.

This is consistent with the fact that in this limit the solution of the Langevin equation converges to the solution of the OU SDE. See Chapter 8 for details.

The eigenfunctions of $\mathcal{L}$ do not form an orthonormal basis in $L^2_\beta := L^2(\mathbb{R}^2, Z^{-1}e^{-\beta H})$ since $\mathcal{L}$ is not a selfadjoint operator. Using the eigenfunctions/eigenvalues of $\mathcal{L}$ we can easily calculate the eigenfunctions/eigenvalues of the L^2_β adjoint of $\mathcal{L}$. From the calculations presented in Section 6.8 we know that the adjoint operator is

$$\begin{aligned}\widehat{\mathcal{L}} &:= -\mathcal{A} + \gamma S \\ &= -\omega_0(b^+a^- - b^-a^+) + \gamma a^+a^- \\ &= -\lambda_1(c^-)^*(c^+)^* - \lambda_2(d^-)*(d^+)*,\end{aligned}$$

where

$$(c^+)^* = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^- + \sqrt{\lambda_2}b^-\right), \tag{6.77a}$$

$$(c^-)^* = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^+ - \sqrt{\lambda_2}b^+\right), \tag{6.77b}$$

$$(d^+)^* = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_2}a^- + \sqrt{\lambda_1}b^-\right), \tag{6.77c}$$

$$(d^-)^* = \frac{1}{\sqrt{\delta}}\left(-\sqrt{\lambda_2}a^+ + \sqrt{\lambda_1}b^+\right). \tag{6.77d}$$

Exercise 6.9.7. *i. Show by direct substitution that $\widehat{\mathcal{L}}$ can be written in the form*

$$\widehat{\mathcal{L}} = -\lambda_1(c^-)^*(c^+)^* - \lambda_2(d^-)*(d^+)*.$$

ii. Calculate the commutators

$$[(c^+)^*, (c^-)^*],\ \ [(d^+)^*, (d^-)^*],\ \ [(c^\pm)^*, (d^\pm)^*],\ \ [\widehat{\mathcal{L}}, (c^\pm)^*],\ \ [\widehat{\mathcal{L}}, (d^\pm)^*].$$

$\widehat{\mathcal{L}}$ has the same eigenvalues as $\mathcal{L}$:

$$-\widehat{\mathcal{L}}\psi_{nm} = \lambda_{nm}\psi_{nm},$$

where λ_{nm} are given by (6.73). The eigenfunctions are

$$\psi_{nm} = \frac{1}{\sqrt{n!m!}}((c^-)^*)^n((d^-)^*)^m\mathbf{1}. \tag{6.78}$$

Proposition 6.9.8. *The eigenfunctions of $\mathcal{L}$ and $\widehat{\mathcal{L}}$ satisfy the biorthonormality relation*

$$\int\int \phi_{nm}\psi_{\ell k}\rho_\beta\, dpdq = \delta_{n\ell}\delta_{mk}. \tag{6.79}$$

Proof. We will use formulas (6.76). Notice that using the third and fourth of these equations together with the fact that $c^-\mathbf{1} = d^-\mathbf{1} = 0$ we can conclude that (for $n \geqslant \ell$)

$$(c^-)^\ell(c^+)^n\mathbf{1} = n(n-1)\dots(n-\ell+1)(c^+)^{n-\ell}. \tag{6.80}$$

We have

$$\begin{aligned}\int\int \phi_{nm}\psi_{\ell k}\rho_\beta\, dpdq &= \frac{1}{\sqrt{n!m!\ell!k!}}\int\int((c^+))^n((d^+))^m\mathbf{1}((c^-)^*)^\ell((d^-)^*)^k\mathbf{1}\rho_\beta\, dpdq \\ &= \frac{n(n-1)\dots(n-\ell+1)m(m-1)\dots(m-k+1)}{\sqrt{n!m!\ell!k!}}\int\int((c^+))^{n-\ell}((d^+))^{m-k}\mathbf{1}\rho_\beta\, dpdq \\ &= \delta_{n\ell}\delta_{mk},\end{aligned}$$

since all eigenfunctions average to 0 with respect to ρ_β. □

From the eigenfunctions of $\widehat{\mathcal{L}}$ we can obtain the eigenfunctions of the Fokker-Planck operator. Using the formula (see equation (6.49))

$$\mathcal{L}^*(f\rho_\beta) = \rho\widehat{\mathcal{L}}f$$

we immediately conclude that the the Fokker-Planck operator has the same eigenvalues as those of $\mathcal{L}$ and $\widehat{\mathcal{L}}$. The eigenfunctions are

$$\psi*_{nm} = \rho_\beta\phi_{nm} = \rho_\beta\frac{1}{\sqrt{n!m!}}((c^-)^*)^n((d^-)^*)^m\mathbf{1}. \tag{6.81}$$

6.10 Discussion and Bibliography

The proof of existence and uniqueness of classical solutions for the Fokker-Planck equation of a uniformly elliptic diffusion process with smooth drift and diffusion coefficients, Theorem 6.1.2, can be found in [7]. A standard textbook on PDEs, with a lot of material on parabolic PDEs is [4]. The Fokker-Planck equation is studied extensively in Risken's monograph [20]. See also [10] and [14]. The connection between the Fokker-Planck equation and stochastic differential equations is presented in Chapter 7. See also [1, 8, 9].

Diffusion processes in one dimension are studied in [16]. The Feller classification for one dimensional diffusion processes can be also found in [15, 5].

Chapter 7

Stochastic Differential Equations

7.1 Introduction

- In this part of the course we will study stochastic differential equation (SDEs): ODEs driven by Gaussian white noise.
- Let $W(t)$ denote a standard m–dimensional Brownian motion, $h : \mathcal{Z} \to \mathbb{R}^d$ a smooth vector-valued function and $\gamma : \mathcal{Z} \to \mathbb{R}^{d\times m}$ a smooth matrix valued function (in this course we will take $\mathcal{Z} = \mathbb{T}^d$, $\mathbb{R}^d$ or $\mathbb{R}^l \oplus \mathbb{T}^{d-l}$.
- Consider the SDE
$$\frac{dz}{dt} = h(z) + \gamma(z)\frac{dW}{dt}, \quad z(0) = z_0. \tag{7.1}$$
- We think of the term $\frac{dW}{dt}$ as representing Gaussian white noise: a mean-zero Gaussian process with correlation $\delta(t-s)I$.
- The function h in (7.1) is sometimes referred to as the *drift* and γ as the *diffusion coefficient*.
- Such a process exists only as a distribution. The precise interpretation of (7.1) is as an integral equation for $z(t) \in C(\mathbb{R}^+, \mathcal{Z})$:
$$z(t) = z_0 + \int_0^t h(z(s))ds + \int_0^t \gamma(z(s))dW(s). \tag{7.2}$$
- In order to make sense of this equation we need to define the stochastic integral against $W(s)$.

7.2 The Itô Stochastic Integral

The Itô Stochastic Integral

- For the rigorous analysis of stochastic differential equations it is necessary to define stochastic integrals of the form
$$I(t) = \int_0^t f(s)\, dW(s), \tag{7.3}$$

- where $W(t)$ is a standard one dimensional Brownian motion. This is not straightforward because $W(t)$ does not have bounded variation.
- In order to define the stochastic integral we assume that $f(t)$ is a random process, adapted to the filtration $\mathcal{F}_t$ generated by the process $W(t)$, and such that

$$\mathbb{E}\left(\int_0^T f(s)^2\,ds\right) < \infty.$$

- The Itô stochastic integral $I(t)$ is defined as the L^2–limit of the Riemann sum approximation of (7.3):

$$I(t) := \lim_{K\to\infty} \sum_{k=1}^{K-1} f(t_{k-1})\left(W(t_k) - W(t_{k-1})\right), \tag{7.4}$$

- where $t_k = k\Delta t$ and $K\Delta t = t$.
- Notice that the function $f(t)$ is evaluated at the left end of each interval $[t_{n-1}, t_n]$ in (7.4).
- The resulting Itô stochastic integral $I(t)$ is a.s. continuous in t.
- These ideas are readily generalized to the case where $W(s)$ is a standard d dimensional Brownian motion and $f(s) \in \mathbb{R}^{m\times d}$ for each s.
- The resulting integral satisfies the *Itô isometry*

$$\mathbb{E}|I(t)|^2 = \int_0^t \mathbb{E}|f(s)|_F^2 ds, \tag{7.5}$$

- where $|\cdot|_F$ denotes the Frobenius norm $|A|_F = \sqrt{tr(A^TA)}$.
- The Itô stochastic integral is a *martingale*:

$$\mathbb{E}I(t) = 0$$

- and

$$\mathbb{E}[I(t)|\mathcal{F}_s] = I(s) \quad \forall\, t \geqslant s,$$

where $\mathcal{F}_s$ denotes the filtration generated by $W(s)$.

Example 7.2.1.

- *Consider the Itô stochastic integral*

$$I(t) = \int_0^t f(s)\,dW(s),$$

- *where f, W are scalar–valued. This is a martingale with quadratic variation*

$$\langle I\rangle_t = \int_0^t (f(s))^2\,ds.$$

- *More generally, for f, W in arbitrary finite dimensions, the integral $I(t)$ is a martingale with quadratic variation*

$$\langle I\rangle_t = \int_0^t (f(s)\otimes f(s))\,ds.$$

7.2.1 The Stratonovich Stochastic Integral

The Stratonovich Stochastic Integral

- In addition to the Itô stochastic integral, we can also define the Stratonovich stochastic integral. It is defined as the L^2–limit of a different Riemann sum approximation of (7.3), namely

$$I_{strat}(t) := \lim_{K\to\infty} \sum_{k=1}^{K-1} \frac{1}{2}\Big(f(t_{k-1}) + f(t_k)\Big)\left(W(t_k) - W(t_{k-1})\right), \tag{7.6}$$

- where $t_k = k\Delta t$ and $K\Delta t = t$. Notice that the function $f(t)$ is evaluated at both endpoints of each interval $[t_{n-1}, t_n]$ in (7.6).

- The multidimensional Stratonovich integral is defined in a similar way. The resulting integral is written as

$$I_{strat}(t) = \int_0^t f(s) \circ dW(s).$$

- The limit in (7.6) gives rise to an integral which differs from the Itô integral.

- The situation is more complex than that arising in the standard theory of Riemann integration for functions of bounded variation: in that case the points in $[t_{k-1}, t_k]$ where the integrand is evaluated do not effect the definition of the integral, via a limiting process.

- In the case of integration against Brownian motion, which does not have bounded variation, the limits differ.

- When f and W are correlated through an SDE, then a formula exists to convert between them.

7.3 Existence and Uniqueness of solutions for SDEs

Existence and Uniqueness of solutions for SDEs

Definition 7.3.1. *By a solution of* (7.1) *we mean a $\mathcal{Z}$-valued stochastic process $\{z(t)\}$ on $t \in [0, T]$ with the properties:*

i. *$z(t)$ is continuous and $\mathcal{F}_t-$adapted, where the filtration is generated by the Brownian motion $W(t)$;*

ii. *$h(z(t)) \in L^1((0,T)), \gamma(z(t)) \in L^2((0,T))$;*

iii. *equation* (7.1) *holds for every $t \in [0, T]$ with probability* 1.

The solution is called unique if any two solutions $x_i(t), i = 1, 2$ satisfy

$$\mathbb{P}(x_1(t) = x_2(t), \ \ \forall t \in [0.T]) = 1.$$

- It is well known that existence and uniqueness of solutions for ODEs (i.e. when $\gamma \equiv 0$ in (7.1)) holds for globally Lipschitz vector fields $h(x)$.

- A very similar theorem holds when $\gamma \neq 0$.
- As for ODEs the conditions can be weakened, when *a priori* bounds on the solution can be found.

Theorem 7.3.2. *Assume that both $h(\cdot)$ and $\gamma(\cdot)$ are globally Lipschitz on $\mathcal{Z}$ and that z_0 is a random variable independent of the Brownian motion $W(t)$ with*

$$\mathbb{E}|z_0|^2 < \infty.$$

Then the SDE (7.1) *has a unique solution $z(t) \in C(\mathbb{R}^+; \mathcal{Z})$ with*

$$\mathbb{E}\left[\int_0^T |z(t)|^2 \, dt\right] < \infty \quad \forall\, T < \infty.$$

Furthermore, the solution of the SDE is a Markov process.

Remarks 7.3.3.

- *The Stratonovich analogue of* (7.1) *is*

$$\frac{dz}{dt} = h(z) + \gamma(z) \circ \frac{dW}{dt}, \quad z(0) = z_0. \tag{7.7}$$

- *By this we mean that $z \in C(\mathbb{R}^+, \mathcal{Z})$ satisfies the integral equation*

$$z(t) = z(0) + \int_0^t h(z(s))ds + \int_0^t \gamma(z(s)) \circ dW(s). \tag{7.8}$$

- *By using definitions* (7.4) *and* (7.6) *it can be shown that z satisfying the Stratonovich SDE* (7.7) *also satisfies the Itô SDE*

$$\frac{dz}{dt} = h(z) + \frac{1}{2}\nabla \cdot \left(\gamma(z)\gamma(z)^T\right) - \frac{1}{2}\gamma(z)\nabla \cdot \left(\gamma(z)^T\right) + \gamma(z)\frac{dW}{dt}, \tag{7.9a}$$

$$z(0) = z_0, \tag{7.9b}$$

- *provided that $\gamma(z)$ is differentiable.*
- White noise is, in most applications, an idealization of a stationary random process with short correlation time. In this context the Stratonovich interpretation of an SDE is particularly important because it often arises as the limit obtained by using smooth approximations to white noise.
- On the other hand the martingale machinery which comes with the Itô integral makes it more important as a mathematical object.
- It is very useful that we can convert from the Itô to the Stratonovich interpretation of the stochastic integral.
- There are other interpretations of the stochastic integral, e.g. the Klimontovich stochastic integral.
- The Definition of Brownian motion implies the scaling property

$$W(ct) = \sqrt{c}W(t),$$

- where the above should be interpreted as holding in law. From this it follows that, if $s = ct$, then

$$\frac{dW}{ds} = \frac{1}{\sqrt{c}}\frac{dW}{dt},$$

  again in law.

- Hence, if we scale time to $s = ct$ in (7.1), then we get the equation

$$\frac{dz}{ds} = \frac{1}{c}h(z) + \frac{1}{\sqrt{c}}\gamma(z)\frac{dW}{ds}, \quad z(0) = z_0.$$

7.4 Derivation of the Stratonovich SDE

The Stratonovich Stochastic Integral: A first application of multiscale methods

- When white noise is approximated by a smooth process this often leads to Stratonovich interpretations of stochastic integrals, at least in one dimension.

- We use multiscale analysis (singular perturbation theory for Markov processes) to illustrate this phenomenon in a one-dimensional example.

- Consider the equations

$$\frac{dx}{dt} = h(x) + \frac{1}{\varepsilon}f(x)y, \tag{7.10a}$$

$$\frac{dy}{dt} = -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2D}{\varepsilon^2}}\frac{dV}{dt}, \tag{7.10b}$$

- with V being a standard one-dimensional Brownian motion.

- We say that the process $x(t)$ is driven by **colored noise**: the noise that appears in (7.10a) has non-zero correlation time. The correlation function of the colored noise $\eta(t) := y(t)/\varepsilon$ is (we take $y(0) = 0$)

$$R(t) = \mathbb{E}\left(\eta(t)\eta(s)\right) = \frac{1}{\varepsilon^2}\frac{D}{\alpha}e^{-\frac{\alpha}{\varepsilon^2}|t-s|}.$$

- The power spectrum of the colored noise $\eta(t)$ is:

$$\begin{aligned} f^{\varepsilon}(x) &= \frac{1}{\varepsilon^2}\frac{D\varepsilon^{-2}}{\pi}\frac{1}{x^2 + (\alpha\varepsilon^{-2})^2} \\ &= \frac{D}{\pi}\frac{1}{\varepsilon^4 x^2 + \alpha^2} \to \frac{D}{\pi\alpha^2} \end{aligned}$$

- and, consequently,

$$\lim_{\varepsilon\to 0}\mathbb{E}\left(\frac{y(t)}{\varepsilon}\frac{y(s)}{\varepsilon}\right) = \frac{2D}{\alpha^2}\delta(t-s),$$

- which implies the heuristic

$$\lim_{\varepsilon\to 0}\frac{y(t)}{\varepsilon} = \sqrt{\frac{2D}{\alpha^2}}\frac{dV}{dt}. \tag{7.11}$$

- Another way of seeing this is by solving (7.10b) for y/ε:

$$\frac{y}{\varepsilon} = \sqrt{\frac{2D}{\alpha^2}}\frac{dV}{dt} - \frac{\varepsilon}{\alpha}\frac{dy}{dt}. \tag{7.12}$$

- If we neglect the $\mathcal{O}(\varepsilon)$ term on the right hand side then we arrive, again, at the heuristic (7.11).

- Both of these arguments lead us to conjecture the limiting Itô SDE:

$$\frac{dX}{dt} = h(X) + \sqrt{\frac{2D}{\alpha}} f(X)\frac{dV}{dt}. \tag{7.13}$$

- In fact, as applied, **the heuristic gives the incorrect limit.**

- whenever white noise is approximated by a smooth process, the limiting equation should be interpreted in the Stratonovich sense, giving

$$\frac{dX}{dt} = h(X) + \sqrt{\frac{2D}{\alpha}} f(X) \circ \frac{dV}{dt}. \tag{7.14}$$

- This is usually called the Wong-Zakai theorem. A similar result is true in arbitrary finite and even infinite dimensions.

- We will show this using singular perturbation theory.

Theorem 7.4.1. *Assume that the initial conditions for $y(t)$ are stationary and that the function f is smooth. Then the solution of eqn* (7.10a) *converges, in the limit as $\varepsilon \to 0$ to the solution of the Stratonovich SDE* (7.14).

Remarks

i. It is possible to prove pathwise convergence under very mild assumptions.

ii. The generator of a Stratonovich SDE has the from

$$\mathcal{L}_{strat} = h(x)\partial_x + \frac{D}{\alpha} f(x)\partial_x \left(f(x)\partial_x\right).$$

iii. Consequently, the Fokker-Planck operator of the Stratonovich SDE can be written in divergence form:

$$\mathcal{L}^*_{strat}\cdot = -\partial_x \left(h(x)\cdot\right) + \frac{D}{\alpha}\partial_x \left(f^2(x)\partial_x \cdot\right).$$

iv. In most applications in physics the white noise is an approximation of a more complicated noise processes with non-zero correlation time. Hence, the physically correct interpretation of the stochastic integral is the Stratonovich one.

v. In higher dimensions an additional drift term might appear due to the noncommutativity of the row vectors of the diffusion matrix. This is related to the **Lévy area correction** in the theory of rough paths.

Proof of Proposition 7.4.1 The generator of the process $(x(t),\, y(t))$ is

$$\begin{aligned}\mathcal{L} &= \frac{1}{\varepsilon^2}\left(-\alpha y\partial_y + D\partial_y^2\right) + \frac{1}{\varepsilon}f(x)y\partial_x + h(x)\partial_x \\ &=: \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2.\end{aligned}$$

The "fast" process is an stationary Markov process with invariant density

$$\rho(y) = \sqrt{\frac{\alpha}{2\pi D}}e^{-\frac{\alpha y^2}{2D}}. \tag{7.15}$$

The backward Kolmogorov equation is

$$\frac{\partial u^\varepsilon}{\partial t} = \left(\frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2\right)u^\varepsilon. \tag{7.16}$$

We look for a solution to this equation in the form of a power series expansion in ε:

$$u^\varepsilon(x,y,t) = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots$$

We substitute this into (7.16) and equate terms of the same power in ε to obtain the following hierarchy of equations:

$$\begin{aligned}-\mathcal{L}_0 u_0 &= 0, \\ -\mathcal{L}_0 u_1 &= \mathcal{L}_1 u_0, \\ -\mathcal{L}_0 u_2 &= \mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t}.\end{aligned}$$

The ergodicity of the fast process implies that the null space of the generator $\mathcal{L}_0$ consists only of constant in y. Hence:

$$u_0 = u(x,t).$$

The second equation in the hierarchy becomes

$$-\mathcal{L}_0 u_1 = f(x)y\partial_x u.$$

This equation is solvable since the right hand side is orthogonal to the null space of the adjoint of $\mathcal{L}_0$ (this is the **Fredholm alterantive**). We solve it using separation of variables:

$$u_1(x,y,t) = \frac{1}{\alpha}f(x)\partial_x u y + \psi_1(x,t).$$

In order for the third equation to have a solution we need to require that the right hand side is orthogonal to the null space of $\mathcal{L}_0^*$:

$$\int_{\mathbb{R}}\left(\mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t}\right)\rho(y)\,dy = 0.$$

We calculate:

$$\int_{\mathbb{R}}\frac{\partial u_0}{\partial t}\rho(y)\,dy = \frac{\partial u}{\partial t}.$$

Furthermore:

$$\int_{\mathbb{R}} \mathcal{L}_2 u_0 \rho(y)\, dy = h(x)\partial_x u.$$

Finally

$$\begin{aligned}
\int_{\mathbb{R}} \mathcal{L}_1 u_1 \rho(y)\, dy &= \int_{\mathbb{R}} f(x) y \partial_x \left(\frac{1}{\alpha} f(x)\partial_x u y + \psi_1(x,t) \right) \rho(y)\, dy \\
&= \frac{1}{\alpha} f(x)\partial_x \left(f(x)\partial_x u \right) \langle y^2 \rangle + f(x)\partial_x \psi_1(x,t) \langle y \rangle \\
&= \frac{D}{\alpha^2} f(x)\partial_x \left(f(x)\partial_x u \right) \\
&= \frac{D}{\alpha^2} f(x)\partial_x f(x)\partial_x u + \frac{D}{\alpha^2} f(x)^2 \partial_x^2 u.
\end{aligned}$$

Putting everything together we obtain the limiting backward Kolmogorov equation

$$\frac{\partial u}{\partial t} = \left(h(x) + \frac{D}{\alpha^2} f(x)\partial_x f(x) \right) \partial_x u + \frac{D}{\alpha^2} f(x)^2 \partial_x^2 u,$$

from which we read off the limiting Stratonovich SDE

$$\frac{dX}{dt} = h(X) + \sqrt{\frac{2D}{\alpha}} f(X) \circ \frac{dV}{dt}.$$

7.5 Itô versus Stratonovich

- A Stratonovich SDE
$$dX(t) = f(X(t))\, dt + \sigma(X(t)) \circ dW(t) \tag{7.17}$$
- can be written as an Itô SDE
$$dX(t) = \left(f(X(t)) + \frac{1}{2} \left(\sigma \frac{d\sigma}{dx} \right) (X(t)) \right) dt + \sigma(X(t))\, dW(t).$$
- Conversely, and Itô SDE
$$dX(t) = f(X(t))\, dt + \sigma(X(t)) dW(t) \tag{7.18}$$
- can be written as a Statonovich SDE
$$dX(t) = \left(f(X(t)) - \frac{1}{2} \left(\sigma \frac{d\sigma}{dx} \right) (X(t)) \right) dt + \sigma(X(t)) \circ dW(t).$$
- The Itô and Stratonovich interpretation of an SDE can lead to equations with very different properties!
- **Multiplicative Noise.**
  - When the diffusion coefficient depends on the solution of the SDE $X(t)$, we will say that we have an equation with **multiplicative noise**.

- Multiplicative noise can lead to **noise induced phase transitions**. See
  - W. Horsthemke and R. Lefever, *Noise-induced transitions*, Springer-Verlag, Berlin 1984.
- This is a topic of current interest for SDEs in infinite dimensions (SPDEs).

Colored Noise

- When the noise which drives an SDE has non-zero correlation time we will say that we have **colored noise**.
- The properties of the SDE (stability, ergodicity etc.) are quite robust under "coloring of the noise". See
  - G. Blankenship and G.C. Papanicolaou, *Stability and control of stochastic systems with wide-band noise disturbances. I*, SIAM J. Appl. Math., **34**(3), 1978, pp. 437–476.
- Colored noise appears in many applications in physics and chemistry. For a review see
  - P. Hanggi and P. Jung *Colored noise in dynamical systems.* Adv. Chem. Phys. **89** 239 (1995).
- In the case where there is an additional small time scale in the problem, in addition to the correlation time of the colored noise, it is not clear what the right interpretation of the stochastic integral (in the limit as both small time scales go to 0). This is usually called the **Itô versus Stratonovich problem**.
- Consider, for example, the SDE
$$\tau \ddot{X} = -\dot{X} + v(X)\eta^{\varepsilon}(t),$$
- where $\eta^{\varepsilon}(t)$ is colored noise with correlation time ε^2.
- In the limit where both small time scales go to 0 we can get either Itô or Stratonovich or neither. See
  - G.A. Pavliotis and A.M. Stuart, *Analysis of white noise limits for stochastic systems with two fast relaxation times*, Multiscale Model. Simul., **4**(1), 2005, pp. 1-35.

7.6 The Generator

- Given the function $\gamma(z)$ in the SDE (7.1) we define
$$\Gamma(z) = \gamma(z)\gamma(z)^T. \tag{7.19}$$
- The **generator** $\mathcal{L}$ is then defined as
$$\mathcal{L}v = h \cdot \nabla v + \frac{1}{2}\Gamma : \nabla\nabla v. \tag{7.20}$$
- This operator, equipped with a suitable domain of definition, is the generator of the Markov process given by (7.1).
- The formal L^2-adjoint operator $\mathcal{L}^*$
$$\mathcal{L}^* v = -\nabla \cdot (hv) + \frac{1}{2}\nabla \cdot \nabla \cdot (\Gamma v).$$

7.7 The Itô Formula

- The **Itô formula** enables us to calculate the rate of change in time of functions $V : \mathcal{Z} \to \mathbb{R}^n$ evaluated at the solution of a $\mathcal{Z}$-valued SDE.

- Formally, we can write:

$$\frac{d}{dt}\Big(V(z(t))\Big) = \mathcal{L}V(z(t)) + \left\langle \nabla V(z(t)), \gamma(z(t))\frac{dW}{dt} \right\rangle .$$

- Note that if W were a smooth time-dependent function this formula would not be correct: there is an additional term in $\mathcal{L}V$, proportional to Γ, which arises from the lack of smoothness of Brownian motion.

- The precise interpretation of the expression for the rate of change of V is in integrated form:

Lemma 7.7.1. (Itô's Formula) *Assume that the conditions of Theorem 7.3.2 hold. Let $x(t)$ solve* (7.1) *and let $V \in C^2(\mathcal{Z}, \mathbb{R}^n)$. Then the process $V(z(t))$ satisfies*

$$V(z(t)) = V(z(0)) + \int_0^t \mathcal{L}V(z(s))ds + \int_0^t \langle \nabla V(z(s)), \gamma(z(s))\, dW(s) \rangle .$$

- Let $\phi : \mathcal{Z} \mapsto \mathbb{R}$ and consider the function

$$v(z,t) = \mathbb{E}\big(\phi(z(t))|z(0) = z\big), \tag{7.21}$$

- where the expectation is with respect to all Brownian driving paths. By averaging in the Itô formula, which removes the stochastic integral, and using the Markov property, it is possible to obtain the Backward Kolmogorov equation.

- For a Stratonovich SDE the rules of standard calculus apply:

  – Consider the Stratonovich SDE (7.17) and let $V(x) \in C^2(\mathbb{R})$. Then

$$dV(X(t)) = \frac{dV}{dx}(X(t))\,(f(X(t))\,dt + \sigma(X(t)) \circ dW(t)) .$$

  – Consider the Stratonovich SDE (7.17) on $\mathbb{R}^d$ (i.e. $f \in \mathbb{R}^d$, $\sigma : \mathbb{R}^n \mapsto \mathbb{R}^d$, $W(t)$ is standard Brownian motion on $\mathbb{R}^n$). The corresponding Fokker-Planck equation is:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (f\rho) + \frac{1}{2}\nabla \cdot (\sigma \nabla \cdot (\sigma\rho))). \tag{7.22}$$

7.8 Examples of SDEs

i. The SDE for Brownian motion is:

$$dX = \sqrt{2\sigma}dW, \quad X(0) = x.$$

- The Solution is:

$$X(t) = x + W(t).$$

ii. The SDE for the Ornstein-Uhlenbeck process is

$$dX = -\alpha X\, dt + \sqrt{2\lambda}\, dW, \quad X(0) = x.$$

- We can solve this equation using the variation of constants formula:

$$X(t) = e^{-\alpha t}x + \sqrt{2\lambda}\int_0^t e^{-\alpha(t-s)}dW(s).$$

- We can use Itô's formula to obtain equations for the moments of the OU process. The generator is:

$$\mathcal{L} = -\alpha x\partial_x + \lambda\partial_x^2.$$

- We apply Itô's formula to the function $f(x) = x^n$ to obtain:

$$\begin{aligned} dX(t)^n &= \mathcal{L}X(t)^n\, dt + \sqrt{2\lambda}\partial X(t)^n\, dW \\ &= -\alpha n X(t)^n\, dt + \lambda n(n-1)X(t)^{n-2}\, dt + n\sqrt{2\lambda}X(t)^{n-1}\, dW. \end{aligned}$$

- Consequently:

$$\begin{aligned} X(t)^n &= x^n + \int_0^t \left(-\alpha n X(t)^n + \lambda n(n-1)X(t)^{n-2}\right)\, dt \\ &\quad + n\sqrt{2\lambda}\int_0^t X(t)^{n-1}\, dW. \end{aligned}$$

- By taking the expectation in the above equation we obtain the equation for the moments of the OU process that we derived earlier using the Fokker-Planck equation:

$$M_n(t) = x^n + \int_0^t \left(-\alpha n M_n(s) + \lambda n(n-1)M_{n-2}(s)\right)\, ds.$$

- Consider the **geometric Brownian motion**

$$dX(t) = \mu X(t)\, dt + \sigma X(t)\, dW(t), \tag{7.23}$$

– where we use the Itô interpretation of the stochastic differential. The generator of this process is

$$\mathcal{L} = \mu x \partial_x + \frac{\sigma^2 x^2}{2} \partial_x^2.$$

– The solution to this equation is

$$X(t) = X(0) \exp\left((\mu - \frac{\sigma^2}{2})t + \sigma W(t) \right). \tag{7.24}$$

- To derive this formula, we apply Itô's formula to the function $f(x) = \log(x)$:

$$\begin{aligned} d\log(X(t)) &= \mathcal{L}\Big(\log(X(t)) \Big)\, dt + \sigma x \partial_x \log(X(t))\, dW(t) \\ &= \left(\mu x \frac{1}{x} + \frac{\sigma^2 x^2}{2} \left(-\frac{1}{x^2} \right) \right) dt + \sigma\, dW(t) \\ &= \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma\, dW(t). \end{aligned}$$

- Consequently:

$$\log\left(\frac{X(t)}{X(0)} \right) = \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t)$$

- from which (7.24) follows.

- Notice that the Stratonovich interpretation of this equation leads to the solution

$$X(t) = X(0) \exp(\mu t + \sigma W(t))$$

- Exercise: calculate all moments of the geometric Brownian motion for the Itô and Stratonovich interpretations of the stochastic integral.

7.8.1 The Stochastic Landau Equation

- Consider the **Landau equation**:

$$\frac{dX_t}{dt} = X_t(c - X_t^2), \quad X_0 = x. \tag{7.25}$$

- This is a gradient flow for the potential $V(x) = \frac{1}{2}cx^2 - \frac{1}{4}x^4$.

- When $c < 0$ all solutions are attracted to the single steady state $X_* = 0$.

- When $c > 0$ the steady state $X_* = 0$ becomes unstable and $X_t \to \sqrt{c}$ if $x > 0$ and $X_t \to -\sqrt{c}$ if $x < 0$.

- Consider additive random perturbations to the Landau equation:

$$\frac{dX_t}{dt} = X_t(c - X_t^2) + \sqrt{2\sigma}\frac{dW_t}{dt}, \quad X_0 = x. \tag{7.26}$$

- This equation defines an ergodic Markov process on $\mathbb{R}$: There exists a unique invariant distribution:

$$\rho(x) = Z^{-1} e^{-V(x)/\sigma}, \quad Z = \int_{\mathbb{R}} e^{-V(x)/\sigma}\, dx, \quad V(x) = \frac{1}{2} c x^2 - \frac{1}{4} x^4.$$

- $\rho(x)$ is a probability density for all values of $c \in \mathbb{R}$.

- The presence of additive noise in some sense "trivializes" the dynamics.

- The dependence of various averaged quantities on c resembles the physical situation of a second order phase transition.

- Consider now multiplicative perturbations of the Landau equation.

$$\frac{dX_t}{dt} = X_t(c - X_t^2) + \sqrt{2\sigma} X_t \frac{dW_t}{dt}, \quad X_0 = x. \tag{7.27}$$

- Where the stochastic differential is interpreted in the Itô sense.

- The generator of this process is

$$\mathcal{L} = x(c - x^2)\partial_x + \sigma x^2 \partial_x^2.$$

- Notice that $X_t = 0$ is always a solution of (7.27). Thus, if we start with $x > 0$ ($x < 0$) the solution will remain positive (negative).

- We will assume that $x > 0$.

- Consider the function $Y_t = \log(X_t)$. We apply Itô's formula to this function:

$$\begin{aligned} dY_t &= \mathcal{L}\log(X_t)\, dt + \sigma X_t \partial_x \log(X_t)\, dW_t \\ &= \left(X_t(c - X_t^2)\frac{1}{X_t} - \sigma X_t^2 \frac{1}{X_t^2} \right) dt + \sigma X_t \frac{1}{X_t}\, dW_t \\ &= (c - \sigma)\, dt - X_t^2\, dt + \sigma\, dW_t. \end{aligned}$$

- Thus, we have been able to transform (7.27) into an SDE with additive noise:

$$dY_t = \left[(c - \sigma) - e^{2Y_t} \right] dt + \sigma\, dW_t. \tag{7.28}$$

  This is a gradient flow with potential

$$V(y) = -\left[(c - \sigma) y - \frac{1}{2} e^{2y} \right].$$

- The invariant measure, if it exists, is of the form

$$\rho(y)\, dy = Z^{-1} e^{-V(y)/\sigma}\, dy.$$

- Going back to the variable x we obtain:

$$\rho(x)\,dx = Z^{-1}x^{(c/\sigma-2)}e^{-\frac{x^2}{2\sigma}}\,dx.$$

- We need to make sure that this distribution is integrable:

$$Z = \int_0^{+\infty} x^\gamma e^{-\frac{x^2}{2\sigma}} < \infty, \quad \gamma = \frac{c}{\sigma} - 2.$$

- For this it is necessary that

$$\gamma > -1 \;\Rightarrow\; c > \sigma.$$

- Not all multiplicative random perturbations lead to ergodic behavior!
- The dependence of the invariant distribution on c is similar to the physical situation of first order phase transitions.
- **Exercise** Analyze this problem for the Stratonovich interpretation of the stochastic integral.
- **Exercise** Study additive and multiplicative random perturbations of the ODE

$$\frac{dx}{dt} = x(c + 2x^2 - x^4).$$

- For more information see M.C. Mackey, A. Longtin, A. Lasota **Noise-Induced Global Asymptotic Stability**, J. Stat. Phys. 60 (5/6) pp. 735-751.

7.9 The Backward Kolmogorov Equation

Theorem 7.9.1. *Assume that ϕ is chosen sufficiently smooth so that the* **backward Kolmogorov equation**

$$\begin{aligned} \frac{\partial v}{\partial t} &= \mathcal{L}v \quad \textit{for}\,(z,t) \in \mathcal{Z} \times (0,\infty), \\ v &= \phi \quad \textit{for}\,(z,t) \in \mathcal{Z} \times \{0\}\,, \end{aligned} \tag{7.29}$$

has a unique classical solution $v(x,t) \in C^{2,1}(\mathcal{Z} \times (0,\infty),)$. Then v is given by (7.21) *where $z(t)$ solves* (7.2).

7.9.1 Derivation of the Fokker-Planck Equation

- Now we can derive rigorously the Fokker-Planck equation.

Theorem 7.9.2. *Consider equation* (7.2) *with $z(0)$ a random variable with density $\rho_0(z)$. Assume that the law of $z(t)$ has a density $\rho(z,t) \in C^{2,1}(\mathcal{Z} \times (0,\infty))$. Then ρ satisfies the* **Fokker-Planck equation**

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho \quad \textit{for}\,(z,t) \in \mathcal{Z} \times (0,\infty), \tag{7.30a}$$

$$\rho = \rho_0 \quad \textit{for}\, z \in \mathcal{Z} \times \{0\}. \tag{7.30b}$$

Proof.

- Let $\mathbb{E}^{\mu}$ denote averaging with respect to the product measure induced by the measure μ with density ρ_0 on $z(0)$ and the independent driving Wiener measure on the SDE itself.
- Averaging over random $z(0)$ distributed with density $\rho_0(z)$, we find

$$\begin{aligned}\mathbb{E}^{\mu}(\phi(z(t))) &= \int_{\mathcal{Z}} v(z,t)\rho_0(z)\,dz \\ &= \int_{\mathcal{Z}} (e^{\mathcal{L}t}\phi)(z)\rho_0(z)\,dz \\ &= \int_{\mathcal{Z}} (e^{\mathcal{L}^{*}t}\rho_0)(z)\phi(z)\,dz.\end{aligned}$$

- But since $\rho(z,t)$ is the density of $z(t)$ we also have

$$\mathbb{E}^{\mu}(\phi(z(t))) = \int_{\mathcal{Z}} \rho(z,t)\phi(z)dz.$$

- Equating these two expressions for the expectation at time t we obtain

$$\int_{\mathcal{Z}} (e^{\mathcal{L}^{*}t}\rho_0)(z)\phi(z)\,dz = \int_{\mathcal{Z}} \rho(z,t)\phi(z)\,dz.$$

- We use a density argument so that the identity can be extended to all $\phi \in L^2(\mathcal{Z})$. Hence, from the above equation we deduce that

$$\rho(z,t) = \left(e^{\mathcal{L}^{*}t}\rho_0\right)(z).$$

- Differentiation of the above equation gives (7.30a).
- Setting $t = 0$ gives the initial condition (7.30b).

□

Chapter 8

The Smoluchowski and Freidlin-Wentzell Limits

- There are very few SDEs/Fokker-Planck equations that can be solved explicitly.
- In most cases we need to study the problem under investigation either approximately or numerically.
- In this part of the course we will develop approximate methods for studying various stochastic systems of practical interest.
- There are many problems of physical interest that can be analyzed using techniques from perturbation theory and asymptotic analysis:

i. Small noise asymptotics at finite time intervals.

ii. Small noise asymptotics/large times (rare events): the theory of large deviations, escape from a potential well, exit time problems.

iii. Small and large friction asymptotics for the Fokker-Planck equation: The Freidlin–Wentzell (underdamped) and Smoluchowski (overdamped) limits.

iv. Large time asymptotics for the Langevin equation in a periodic potential: homogenization and averaging.

v. Stochastic systems with two characteristic time scales: multiscale problems and methods.

8.1 Asymptotics for the Langevin equation

- We will study various asymptotic limits for the Langevin equation (we have set $m = 1$)

$$\ddot{q} = -\nabla V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W}. \tag{8.1}$$

- There are two parameters in the problem, the friction coefficient γ and the inverse temperature β.

- We want to study the qualitative behavior of solutions to this equation (and to the corresponding Fokker-Planck equation).

- There are various asymptotic limits at which we can eliminate some of the variables of the equation and obtain a simpler equation for fewer variables.

- In the large temperature limit, $\beta \ll 1$, the dynamics of (9.7) is dominated by diffusion: the Langevin equation (9.7) can be approximated by free Brownian motion:

$$\dot{q} = \sqrt{2\gamma\beta^{-1}}\dot{W}.$$

- The small temperature asymptotics, $\beta \gg 1$ is much more interesting and more subtle. It leads to exponential, Arrhenius type asymptotics for the reaction rate (in the case of a particle escaping from a potential well due to thermal noise) or the diffusion coefficient (in the case of a particle moving in a periodic potential in the presence of thermal noise)

$$\kappa = \nu \exp\left(-\beta E_b\right), \tag{8.2}$$

- where κ can be either the reaction rate or the diffusion coefficient. The small temperature asymptotics will be studied later for the case of a bistable potential (reaction rate) and for the case of a periodic potential (diffusion coefficient).

- Assuming that the temperature is fixed, the only parameter that is left is the friction coefficient γ. The large and small friction asymptotics can be expressed in terms of a slow/fast system of SDEs.

- In many applications (especially in biology) the friction coefficient is large: $\gamma \gg 1$. In this case the momentum is the fast variable which we can eliminate to obtain an equation for the position. This is the **overdamped** or **Smoluchowski** limit.

- In various problems in physics the friction coefficient is small: $\gamma \ll 1$. In this case the position is the fast variable whereas the energy is the slow variable. We can eliminate the position and obtain an equation for the energy. This is the **underdampled** or **Freidlin-Wentzell** limit.

- In both cases we have to look at sufficiently long time scales.

- We rescale the solution to (9.7):

$$q^\gamma(t) = \lambda_\gamma(t/\mu_\gamma).$$

- This rescaled process satisfies the equation

$$\ddot{q}^\gamma = -\frac{\lambda_\gamma}{\mu_\gamma^2}\partial_q V(q^\gamma/\lambda_\gamma) - \frac{\gamma}{\mu_\gamma}\dot{q}^\gamma + \sqrt{2\gamma\lambda_\gamma^2\mu_\gamma^{-3}\beta^{-1}}\dot{W}, \tag{8.3}$$

- Different choices for these two parameters lead to the overdamped and underdamped limits:

- $\lambda_\gamma = 1, \quad \mu_\gamma = \gamma^{-1}, \quad \gamma \gg 1.$

- In this case equation (8.3) becomes

$$\gamma^{-2}\ddot{q}^\gamma = -\partial_q V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\beta^{-1}}\dot{W}. \tag{8.4}$$

- Under this scaling, the interesting limit is the overdamped limit, $\gamma \gg 1$.
- We will see later that in the limit as $\gamma \to +\infty$ the solution to (8.4) can be approximated by the solution to

$$\dot{q} = -\partial_q V + \sqrt{2\beta^{-1}}\dot{W}.$$

- $\lambda_\gamma = 1, \quad \mu_\gamma = \gamma, \quad \gamma \ll 1$:

$$\ddot{q}^\gamma = -\gamma^{-2}\nabla V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\gamma^{-2}\beta^{-1}}\dot{W}. \tag{8.5}$$

- Under this scaling the interesting limit is the underdamped limit, $\gamma \ll 1$.
- We will see later that in the limit as $\gamma \to 0$ the energy of the solution to (8.5) converges to a stochastic process on a graph.

8.2 The Kramers to Smoluchowski Limit

- We consider the rescaled Langevin equation (8.4):

$$\varepsilon^2\ddot{q}^\gamma(t) = -\nabla V(q^\gamma(t)) - \dot{q}^\gamma(t) + \sqrt{2\beta^{-1}}\dot{W}(t), \tag{8.6}$$

- where we have set $\varepsilon^{-1} = \gamma$, since we are interested in the limit $\gamma \to \infty$, i.e. $\varepsilon \to 0$.
- We will show that, in the limit as $\varepsilon \to 0$, $q^\gamma(t)$, the solution of the Langevin equation (8.6), converges to $q(t)$, the solution of the Smoluchowski equation

$$\dot{q} = -\nabla V + \sqrt{2\beta^{-1}}\dot{W}. \tag{8.7}$$

- We write (8.6) as a system of SDEs:

$$\dot{q} = \frac{1}{\varepsilon}p, \tag{8.8}$$

$$\dot{p} = -\frac{1}{\varepsilon}\nabla V(q) - \frac{1}{\varepsilon^2}p + \sqrt{\frac{2}{\beta\varepsilon^2}}\dot{W}. \tag{8.9}$$

- This systems of SDEs defined a Markov process in phase space. Its generator is

$$\begin{aligned}\mathcal{L}^\varepsilon &= \frac{1}{\varepsilon^2}\big(-p\cdot\nabla_p + \beta^{-1}\Delta\big) + \frac{1}{\varepsilon}\big(p\cdot\nabla_q - \nabla_q V\cdot\nabla_p\big)\\ &=: \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1.\end{aligned}$$

- This is a singularly perturbed differential operator.

- We will derive the Smoluchowski equation (8.7) using a pathwise technique, as well as by analyzing the corresponding Kolmogorov equations.

- We apply Itô's formula to p:

$$\begin{aligned} dp(t) &= \mathcal{L}^\varepsilon p(t)\,dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\partial_p p(t)\,dW \\ &= -\frac{1}{\varepsilon^2}p(t)\,dt - \frac{1}{\varepsilon}\nabla_q V(q(t))\,dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\,dW. \end{aligned}$$

- Consequently:

$$\frac{1}{\varepsilon}\int_0^t p(s)\,ds = -\int_0^t \nabla_q V(q(s))\,ds + \sqrt{2\beta^{-1}}\,W(t) + \mathcal{O}(\varepsilon).$$

- From equation (8.8) we have that

$$q(t) = q(0) + \frac{1}{\varepsilon}\int_0^t p(s)\,ds.$$

- Combining the above two equations we deduce

$$q(t) = q(0) - \int_0^t \nabla_q V(q(s))\,ds + \sqrt{2\beta^{-1}}\,W(t) + \mathcal{O}(\varepsilon)$$

- from which (8.7) follows.

- Notice that in this derivation we assumed that

$$\mathbb{E}|p(t)|^2 \leqslant C.$$

- This estimate is true, under appropriate assumptions on the potential $V(q)$ and on the initial conditions.

- In fact, we can prove a pathwise approximation result:

$$\left(\mathbb{E}\sup_{t\in[0,T]}|q^\gamma(t) - q(t)|^p\right)^{1/p} \leqslant C\varepsilon^{2-\kappa},$$

- where $\kappa > 0$, arbitrary small (it accounts for logarithmic corrections).

- For the rigorous proof see

  - E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University press 1967.

- A similar approximation theorem is also valid in infinite dimensions (i.e. for SPDEs):

– S. Cerrai and M. Freidlin, *On the Smoluchowski-Kramers approximation for a system with an infinite number of degrees of freedom*, Probab. Theory Related Fields, 135 (3), 2006, pp. 363–394.

- The pathwise derivation of the Smoluchowski equation implies that the solution of the Fokker-Planck equation corresponding to the Langevin equation (8.6) converges (in some appropriate sense to be explained below) to the solution of the Fokker-Planck equation corresponding to the Smoluchowski equation (8.7).

- It is important in various applications to calculate corrections to the limiting Fokker-Planck equation.

- We can accomplish this by analyzing the Fokker-Planck equation for (8.6) using singular perturbation theory.

- We will consider the problem in one dimension. This mainly to simplify the notation. The multi–dimensional problem can be treated in a very similar way.

- The Fokker–Planck equation associated to equations (8.8) and (8.9) is

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \mathcal{L}^* \rho \\ &= \frac{1}{\varepsilon}\left(-p\partial_q \rho + \partial_q V(q)\partial_p \rho\right) + \frac{1}{\varepsilon^2}\left(\partial_p(p\rho) + \beta^{-1}\partial_p^2 \rho\right) \\ &=: \left(\frac{1}{\varepsilon^2}\mathcal{L}_0^* + \frac{1}{\varepsilon}\mathcal{L}_1^*\right)\rho. \end{aligned} \tag{8.10}$$

- The invariant distribution of the Markov process $\{q,\, p\}$, if it exists, is

$$\rho_0(p,q) = \frac{1}{Z} e^{-\beta H(p,q)}, \quad Z = \int_{\mathbb{R}^2} e^{-\beta H(p,q)}\, dpdq,$$

- where $H(p,q) = \frac{1}{2}p^2 + V(q)$. We define the function f(p,q,t) through

$$\rho(p,q,t) = f(p,q,t)\rho_0(p,q). \tag{8.11}$$

Theorem 8.2.1. *The function $f(p,q,t)$ defined in* (8.11) *satisfies the equation*

$$\begin{aligned} \frac{\partial f}{\partial t} &= \left[\frac{1}{\varepsilon^2}\left(-p\partial_q + \beta^{-1}\partial_p^2\right) - \frac{1}{\varepsilon}\left(p\partial_q - \partial_q V(q)\partial_p\right)\right] f \\ &=: \left(\frac{1}{\varepsilon^2}\mathcal{L}_0 - \frac{1}{\varepsilon}\mathcal{L}_1\right) f. \end{aligned} \tag{8.12}$$

remark

- This is "almost" the backward Kolmogorov equation with the difference that we have $-\mathcal{L}_1$ instead of $\mathcal{L}_1$. This is related to the fact that $\mathcal{L}_0$ is a symmetric operator in $L^2(\mathbb{R}^2; Z^{-1}e^{-\beta H(p,q)})$, whereas $\mathcal{L}_1$ is antisymmetric.

Proof.

- We note that $\mathcal{L}_0^*\rho_0 = 0$ and $\mathcal{L}_1^*\rho_0 = 0$. We use this to calculate:

$$\begin{aligned}\mathcal{L}_0^*\rho &= \mathcal{L}_0(f\rho_0) = \partial_p(f\rho_0) + \beta^{-1}\partial_p^2(f\rho_0)\\ &= \rho_0 p\partial_p f + \rho_0\beta^{-1}\partial_p^2 f + f\mathcal{L}_0^*\rho_0 + 2\beta^{-1}\partial_p f\partial_p\rho_0\\ &= \left(-p\partial_p f + \beta^{-1}\partial_p^2 f\right)\rho_0 = \rho_0\mathcal{L}_0 f.\end{aligned}$$

- Similarly,

$$\begin{aligned}\mathcal{L}_1^*\rho &= \mathcal{L}_1^*(f\rho_0) = \left(-p\partial_q + \partial_q V\partial_p\right)(f\rho_0)\\ &= \rho_0\left(-p\partial_q f + \partial_q V\partial_p f\right) = -\rho_0\mathcal{L}_1 f.\end{aligned}$$

- Consequently, the Fokker–Planck equation (8.10) becomes

$$\rho_0\frac{\partial f}{\partial t} = \rho_0\left(\frac{1}{\varepsilon^2}\mathcal{L}_0 f - \frac{1}{\varepsilon}\mathcal{L}_1 f\right),$$

- from which the claim follows.

□

- We look for a solution to (8.12) in the form of a power series in ε:

$$f(p,q,t) = \sum_{n=0}^{\infty}\varepsilon^n f_n(p,q,t). \tag{8.13}$$

- We substitute this expansion into eqn. (8.12) to obtain the following system of equations.

$$\begin{aligned}\mathcal{L}_0 f_0 &= 0, &(8.14)\\ -\mathcal{L}_0 f_1 &= -\mathcal{L}_1 f_0, &(8.15)\\ -\mathcal{L}_0 f_2 &= -\mathcal{L}_1 f_1 - \frac{\partial f_0}{\partial t} &(8.16)\\ -\mathcal{L}_0 f_{n+1} &= -\mathcal{L}_1 f_n, \quad n = 2,3\ldots &(8.17)\end{aligned}$$

- The null space of $\mathcal{L}_0$ consists of constants in p. Consequently, from equation (8.14) we conclude that

$$f_0 = f(q,t).$$

- Now we can calculate the right hand side of equation (8.15):

$$\mathcal{L}_1 f_0 = p\partial_q f.$$

- Equation (8.15) becomes:

$$\mathcal{L}_0 f_1 = p\partial_q f.$$

- The right hand side of this equation is orthogonal to $\mathcal{N}(\mathcal{L}_0^*)$ and consequently there exists a unique solution. We obtain this solution using separation of variables:

$$f_1 = -p\partial_q f + \psi_1(q,t).$$

- Now we can calculate the RHS of equation (8.16). We need to calculate $\mathcal{L}_1 f_1$:

$$\begin{aligned} -\mathcal{L}_1 f_1 &= \Big(p\partial_q - \partial_q V \partial_p\Big)\Big(p\partial_q f - \psi_1(q,t)\Big) \\ &= p^2\partial_q^2 f - p\partial_q\psi_1 - \partial_q V \partial_q f. \end{aligned}$$

- The solvability condition for (8.16) is

$$\int_{\mathbb{R}} \Big(- \mathcal{L}_1 f_1 - \frac{\partial f_0}{\partial t}\Big)\rho_{OU}(p)\, dp = 0,$$

- from which we obtain the backward Kolmogorov equation corresponding to the Smoluchowski SDE:

$$\frac{\partial f}{\partial t} = -\partial_q V \partial_q f + \beta^{-1}\partial_p^2 f. \tag{8.18}$$

- Now we solve the equation for f_2. We use (8.18) to write (8.16) in the form

$$\mathcal{L}_0 f_2 = \Big(\beta^{-1} - p^2\Big)\partial_q^2 f + p\partial_q\psi_1.$$

- The solution of this equation is

$$f_2(p,q,t) = \frac{1}{2}\partial_q^2 f(p,q,t)p^2 - \partial_q\psi_1(q,t)p + \psi_2(q,t).$$

- The right hand side of the equation for f_3 is

$$\mathcal{L}_1 f_2 = \frac{1}{2}p^3\partial_q^3 f - p^2\partial_q^2\psi_1 + p\partial_q\psi_2 - \partial_q V\partial_q^2 f p - partial_q V\partial_q\psi_1.$$

  The solvability condition

$$\int_{\mathbb{R}} \mathcal{L}_1 f_2 \rho_{OU}(p)\, dp = 0.$$

- This leads to the equation

$$-\partial_q V\partial_q\psi_1 - \beta^{-1}\partial_q^2\psi_1 = 0$$

- The only solution to this equation which is an element of $L^2(e^{-\beta V(q)})$ is

$$\psi_1 \equiv 0.$$

- Putting everything together we obtain the first two terms in the ε-expansion of the Fokker–Planck equation (8.12):

$$\rho(p,q,t) = Z^{-1}e^{-\beta H(p,q)}\Big(f + \varepsilon(-p\partial_q f) + \mathcal{O}(\varepsilon^2)\Big),$$

- where f is the solution of (8.18).
- Notice that we can rewrite the leading order term to the expansion in the form

$$\rho(p,q,t) = (2\pi\beta^{-1})^{-\frac{1}{2}} e^{-\beta p^2/2} \rho_V(q,t) + \mathcal{O}(\varepsilon),$$

- where $\rho_V = Z^{-1} e^{-\beta V(q)} f$ is the solution of the Smoluchowski Fokker-Planck equation

$$\frac{\partial \rho_V}{\partial t} = \partial_q(\partial_q V \rho_V) + \beta^{-1}\partial_q^2 \rho_V.$$

- It is possible to expand the n-th term in the expansion (8.13) in terms of Hermite functions (the eigenfunctions of the generator of the OU process)

$$f_n(p,q,t) = \sum_{k=0}^{n} f_{nk}(q,t)\phi_k(p), \tag{8.19}$$

  where $\phi_k(p)$ is the k–th eigenfunction of $\mathcal{L}_0$:

$$-\mathcal{L}_0\phi_k = \lambda_k \phi_k.$$

- We can obtain the following system of equations ($\widehat{\mathcal{L}} = \beta^{-1}\partial_q - \partial_q V$):

$$\begin{aligned}
\widehat{\mathcal{L}} f_{n1} &= 0, \\
\sqrt{\frac{k+1}{\beta^{-1}}}\widehat{\mathcal{L}} f_{n,k+1} + \sqrt{k\beta^{-1}}\partial_q f_{n,k-1} &= -k f_{n+1,k}, \quad k = 1, 2 \ldots, n-1, \\
\sqrt{n\beta^{-1}}\partial_q f_{n,n-1} &= -n f_{n+1,n}, \\
\sqrt{(n+1)\beta^{-1}}\partial_q f_{n,n} &= -(n+1) f_{n+1,n+1}.
\end{aligned}$$

- Using this method we can obtain the first three terms in the expansion:

$$\begin{aligned}
\rho(x,y,t) &= \rho_0(p,q)\left(f + \varepsilon(-\sqrt{\beta^{-1}}\partial_q f \phi_1) + \varepsilon^2\left(\frac{\beta^{-1}}{\sqrt{2}}\partial_q^2 f\phi_2 + f_{20}\right)\right. \\
&\quad \left. + \varepsilon^3\left(-\sqrt{\frac{\beta^{-3}}{3!}}\partial_q^3 f\phi_3 + \left(-\sqrt{\beta^{-1}}\widehat{\mathcal{L}}\partial_q^2 f - \sqrt{\beta^{-1}}\partial_q f_{20}\right)\phi_1\right)\right) \\
&\quad + \mathcal{O}(\varepsilon^4),
\end{aligned}$$

8.3 The Freidlin–Wentzell Limit

Consider now the rescaling $\lambda_{\gamma,\varepsilon} = 1$, $\mu_{\gamma,\varepsilon} = \gamma$. The Langevin equation becomes

$$\ddot{q}^\gamma = -\gamma^{-2}\nabla V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\gamma^{-2}\beta^{-1}}\dot{W}. \tag{8.20}$$

We write equation (8.20) as system of two equations

$$\dot{q}^\gamma = \gamma^{-1} p^\gamma, \quad \dot{p}^\gamma = -\gamma^{-1}V'(q^\gamma) - p^\gamma + \sqrt{2\beta^{-1}}\dot{W}.$$

This is the equation for an $\mathcal{O}(1/\gamma)$ Hamiltonian system perturbed by $\mathcal{O}(1)$ noise. We expect that, to leading order, the energy is conserved, since it is conserved for the Hamiltonian system. We apply Itô's formula to the Hamiltonian of the system to obtain

$$\dot{H} = \left(\beta^{-1} - p^2\right) + \sqrt{2\beta^{-1}p^2}\dot{W}$$

with $p^2 = p^2(H, q) = 2(H - V(q))$.

Thus, in order to study the $\gamma \to 0$ limit we need to analyze the following fast/slow system of SDEs

$$\dot{H} = \left(\beta^{-1} - p^2\right) + \sqrt{2\beta^{-1}p^2}\dot{W} \tag{8.21a}$$

$$\dot{p^\gamma} = -\gamma^{-1}V'(q^\gamma) - p^\gamma + \sqrt{2\beta^{-1}}\dot{W}. \tag{8.21b}$$

The Hamiltonian is the slow variable, whereas the momentum (or position) is the fast variable. Assuming that we can average over the Hamiltonian dynamics, we obtain the limiting SDE for the Hamiltonian:

$$\dot{H} = \left(\beta^{-1} - \langle p^2\rangle\right) + \sqrt{2\beta^{-1}\langle p^2\rangle}\dot{W}. \tag{8.22}$$

The limiting SDE lives on the graph associated with the Hamiltonian system. The domain of definition of the limiting Markov process is defined through appropriate boundary conditions (the **gluing conditions**) at the interior vertices of the graph.

- We identify all points belonging to the same connected component of the a level curve $\{x : H(x) = H\}$, $x = (q, p)$.
- Each point on the edges of the graph correspond to a trajectory.
- Interior vertices correspond to separatrices.
- Let $I_i, i = 1, \dots d$ be the edges of the graph. Then (i, H) defines a global coordinate system on the graph.
- For more information see
  - Freidlin and Wentzell, Random Perturbations of Dynamical Systems, Springer 1998.
  - Freidlin and Wentzell, Random Perturbations of Hamiltonian Systems, AMS 1994.
  - Sowers, A Boundary layer theory for diffusively perturbed transport around a heteroclinic cycle, CPAM 58 (2005), no. 1, 30–84.

We will study the small γ asymptotics by analyzing the corresponding backward Kolmogorov equation using singular perturbation theory. The generator of the process $\{q^\gamma, p^\gamma\}$ is

$$\begin{aligned}\mathcal{L}^\gamma &= \gamma^{-1}\left(p\partial_q - \partial_q V\partial_p\right) - p\partial_p + \beta^{-1}\partial_p^2 \\ &= \gamma^{-1}\mathcal{L}_0 + \mathcal{L}_1.\end{aligned}$$

Let $u^\gamma = \mathbb{E}(f(p^\gamma(p,q;t), q^\gamma(p,q;t)))$. It satisfies the backward Kolmogorov equation associated to the process $\{q^\gamma,\, p^\gamma\}$:

$$\frac{\partial u^\gamma}{\partial t} = \left(\frac{1}{\gamma}\mathcal{L}_0 + \mathcal{L}_1\right) u^\gamma. \tag{8.23}$$

We look for a solution in the form of a power series expansion in ε:

$$u^\gamma = u_0 + \gamma u_1 + \gamma^2 u_2 + \dots$$

We substitute this ansatz into (8.23) and equate equal powers in ε to obtain the following sequence of equations:

$$\mathcal{L}_0 u_0 = 0, \tag{8.24a}$$

$$\mathcal{L}_0 u_1 = -\mathcal{L}_1 u_1 + \frac{\partial u_0}{\partial t}, \tag{8.24b}$$

$$\mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 + \frac{\partial u_1}{\partial t}. \tag{8.24c}$$

$$\dots\dots\dots$$

Notice that the operator $\mathcal{L}_0$ is the backward Liouville operator of the Hamiltonian system with Hamiltonian

$$H = \frac{1}{2}p^2 + V(q).$$

We assume that there are no integrals of motion other than the Hamiltonian. This means that the null space of $\mathcal{L}_0$ consists of functions of the Hamiltonian:

$$\mathcal{N}(\mathcal{L}_0) = \big\{\text{functions of} H\big\}. \tag{8.25}$$

Let us now analyze equations (8.24). We start with (8.24a); eqn. (8.25) implies that u_0 depends on $q,\ p$ through the Hamiltonian function H:

$$u_0 = u(H(p,q),t) \tag{8.26}$$

Now we proceed with (8.24b). For this we need to find the solvability condition for equations of the form

$$\mathcal{L}_0 u = f \tag{8.27}$$

My multiply it by an arbitrary smooth function of $H(p,q)$, integrate over $\mathbb{R}^2$ and use the skew-symmetry of the Liouville operator $\mathcal{L}_0$ to deduce:[1]

$$\begin{aligned}
\int_{\mathbb{R}^2} \mathcal{L}_0 u F(H(p,q))\, dpdq &= \int_{\mathbb{R}^2} u\mathcal{L}_0^* F(H(p,q))\, dpdq \\
&= \int_{\mathbb{R}^2} u(-\mathcal{L}_0 F(H(p,q)))\, dpdq \\
&= 0, \quad \forall F \in C_b^\infty(\mathbb{R}).
\end{aligned}$$

[1] We assume that both u_1 and F decay to 0 as $|p| \to \infty$ to justify the integration by parts that follows.

This implies that the **solvability condition** for equation (8.27) is that

$$\int_{\mathbb{R}^2} f(p,q)F(H(p,q))\,dpdq = 0, \quad \forall F \in C_b^\infty(\mathbb{R}). \tag{8.28}$$

We use the solvability condition in (8.24b) to obtain that

$$\int_{\mathbb{R}^2} \left(\mathcal{L}_1 u_1 - \frac{\partial u_0}{\partial t}\right) F(H(p,q))\,dpdq = 0, \tag{8.29}$$

To proceed, we need to understand how $\mathcal{L}_1$ acts to functions of $H(p,q)$. Let $\phi = \phi(H(p,q))$. We have that

$$\frac{\partial \phi}{\partial p} = \frac{\partial H}{\partial p}\frac{\partial \phi}{\partial H} = p\frac{\partial \phi}{\partial H}$$

and

$$\frac{\partial^2 \phi}{\partial p^2} = \frac{\partial}{\partial p}\left(\frac{\partial \phi}{\partial H}\right) = \frac{\partial \phi}{\partial H} + p^2\frac{\partial^2 \phi}{\partial H^2}.$$

The above calculations imply that, when $\mathcal{L}_1$ acts on functions $\phi = \phi(H(p,q))$, it becomes

$$\mathcal{L}_1 = \Big[(\beta^{-1} - p^2)\partial_H + \beta^{-1}p^2\partial_H^2\Big], \tag{8.30}$$

where

$$p^2 = p^2(H,q) = 2(H - V(q)).$$

We want to change variables in the integral (8.29) and go from $(p,\, q)$ to $p,\, H$. The Jacobian of the transformation is:

$$\frac{\partial(p,q)}{\partial(H,q)} = \begin{vmatrix} \frac{\partial p}{\partial H} & \frac{\partial p}{\partial q} \\ \frac{\partial q}{\partial H} & \frac{\partial q}{\partial q} \end{vmatrix} = \frac{\partial p}{\partial H} = \frac{1}{p(H,q)}.$$

We use this, together with (8.30), to rewrite eqn. (8.29) as

$$\int\int \Big(\frac{\partial u}{\partial t} + \Big[(\beta^{-1} - p^2)\partial_H + \beta^{-1}p^2\partial_H^2\Big]u\Big)F(H)p^{-1}(H,q)\,dHdq = 0.$$

We introduce the notation

$$\langle\cdot\rangle := \int \cdot\, dq.$$

The integration over q can be performed "explicitly":

$$\int \Big[\frac{\partial u}{\partial t}\langle p^{-1}\rangle + \big((\beta^{-1}\langle p^{-1}\rangle - \langle p\rangle)\partial_H + \beta^{-1}\langle p\rangle\partial_H^2\big)u\Big]F(H)\,dH = 0.$$

This equation should be valid for every smooth function $F(H)$, and this requirement leads to the differential equation

$$\langle p^{-1}\rangle\frac{\partial u}{\partial t} = \big(\beta^{-1}\langle p^{-1}\rangle - \langle p\rangle\big)\partial_H u + \langle p\rangle\beta^{-1}\partial_H^2 u,$$

or,

$$\frac{\partial u}{\partial t} = \left(\beta^{-1} - \langle p^{-1}\rangle^{-1}\langle p\rangle\right)\partial_H u + \gamma\langle p^{-1}\rangle^{-1}\langle p\rangle\beta^{-1}\partial_H^2 u.$$

Thus, we have obtained the limiting backward Kolmogorov equation for the energy, which is the "slow variable". From this equation we can read off the limiting SDE for the Hamiltonian:

$$\dot{H} = b(H) + \sigma(H)\dot{W} \tag{8.31}$$

where

$$b(H) = \beta^{-1} - \langle p^{-1}\rangle^{-1}\langle p\rangle, \quad \sigma(H) = \beta^{-1}\langle p^{-1}\rangle^{-1}\langle p\rangle.$$

Notice that the noise that appears in the limiting equation (8.31) is multiplicative, contrary to the additive noise in the Langevin equation.

As it well known from classical mechanics, the **action** and **frequency** are defined as

$$I(E) = \int p(q, E)\, dq$$

and

$$\omega(E) = 2\pi \left(\frac{dI}{dE}\right)^{-1},$$

respectively. Using the action and the frequency we can write the limiting Fokker–Planck equation for the distribution function of the energy in a very compact form.

Theorem 8.3.1. *The limiting Fokker–Planck equation for the energy distribution function $\rho(E, t)$ is*

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial E}\left(\left(I(E) + \beta^{-1}\frac{\partial}{\partial E}\right)\left(\frac{\omega(E)\rho}{2\pi}\right)\right). \tag{8.32}$$

Proof. We notice that

$$\frac{dI}{dE} = \int \frac{\partial p}{\partial E}\, dq = \int p^{-1}\, dq$$

and consequently

$$\langle p^{-1}\rangle^{-1} = \frac{\omega(E)}{2\pi}.$$

Hence, the limiting Fokker–Planck equation can be written as

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= -\frac{\partial}{\partial E}\left(\left(\beta^{-1}\frac{I(E)\omega(E)}{2\pi}\right)\rho\right) + \beta^{-1}\frac{\partial^2}{\partial E^2}\left(\frac{I\omega}{2\pi}\right) \\
&= -\beta^{-1}\frac{\partial \rho}{\partial E} + \frac{\partial}{\partial E}\left(\frac{I\omega}{2\pi}\rho\right) + \beta^{-1}\frac{\partial}{\partial E}\left(\frac{dI}{dE}\frac{\omega\rho}{2\pi}\right) + \beta^{-1}\frac{\partial}{\partial E}\left(I\frac{\partial}{\partial E}\left(\frac{\omega\rho}{2\pi}\right)\right) \\
&= \frac{\partial}{\partial E}\left(\frac{I\omega}{2\pi}\rho\right) + \beta^{-1}\frac{\partial}{\partial E}\left(I\frac{\partial}{\partial E}\left(\frac{\omega\rho}{2\pi}\right)\right) \\
&= \frac{\partial}{\partial E}\left(\left(I(E) + \beta^{-1}\frac{\partial}{\partial E}\right)\left(\frac{\omega(E)\rho}{2\pi}\right)\right),
\end{aligned}$$

which is precisely equation (8.32). □

Remarks 8.3.2. *i. We emphasize that the above formal procedure does not provide us with the boundary conditions for the limiting Fokker–Planck equation. We will discuss about this issue in the next section.*

ii. If we rescale back to the original time-scale we obtain the equation

$$\frac{\partial \rho}{\partial t} = \gamma \frac{\partial}{\partial E}\left(\left(I(E) + \beta^{-1}\frac{\partial}{\partial E}\right)\left(\frac{\omega(E)\rho}{2\pi}\right)\right). \tag{8.33}$$

We will use this equation later on to calculate the rate of escape from a potential barrier in the energy-diffusion-limited regime.

Chapter 9

Exit Time Problems and Reaction Rate Theory

9.1 Introduction

There are many systems in physics, chemistry and biology that exist in at least two stable states. Among the many applications we mention the switching and storage devices in computers. Another example is biological macromolecules that can exist in many different states. The problems that we would like to solve are:

- How stable are the various states relative to each other.
- How long does it take for a system to switch spontaneously from one state to another?
- How is the transfer made, i.e. through what path in the relevant state space? There is a lot of important current work on this problem by E, Vanden Eijnden etc.
- How does the system relax to an unstable state?

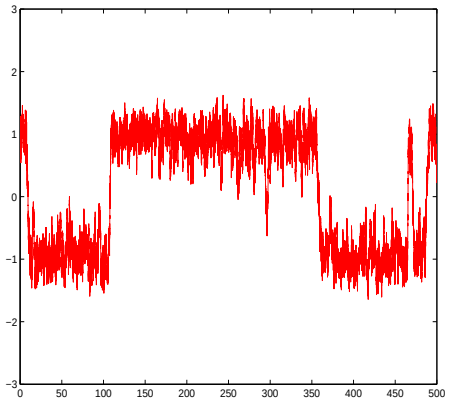
We can separate between the 1d problem, the finite dimensional problem and the infinite dimensional problem (SPDEs). We we will solve completely the one dimensional problem and discuss in some detail about the finite dimensional problem. The infinite dimensional situation is an extremely hard problem and we will only make some remarks. The study of bistability and metastability is a very active research area, in particular the development of numerical methods for the calculation of various quantities such as reaction rates, transition pathways etc.

We will mostly consider the dynamics of a particle moving in a bistable potential, under the influence of thermal noise in one dimension:

$$\dot{x} = -V'(x) + \sqrt{2k_B T}\dot{\beta}. \tag{9.1}$$

An example of the class of potentials that we will consider is shown in Figure. It has to local minima, one local maximum and it increases at least quadratically at infinity. This ensures that the state space is ”compact”, i.e. that the particle cannot escape at infinity. The standard potential that satisfies these assumptions is

$$V(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2 + \frac{1}{4}. \tag{9.2}$$



It is easily checked that this potential has three local minima, a local maximum at $x = 0$ and two local minima at $x = \pm 1$. The values of the potential at these three points are:

$$V(\pm 1) = 0, \quad V(0) = \frac{1}{4}.$$

We will say that the height of the potential barrier is $\frac{1}{4}$. The physically (and mathematically!) interesting case is when the thermal fluctuations are weak when compared to the potential barrier that the particle has to climb over.

More generally, we assume that the potential has two local minima at the points a and c and a local maximum at b. Let us consider the problem of the escape of the particle from the left local minimum a. The potential barrier is then defined as

$$\Delta E = V(b) - V(a).$$

Our assumption that the thermal fluctuations are weak can be written as

$$\frac{k_B T}{\Delta E} \ll 1.$$

In this limit, it is intuitively clear that the particle is most likely to be found at either a or c. There it will perform small oscillations around either of the local minima. This is a result that we can obtain by studying the small temperature limit by using perturbation theory. The result is that we can describe locally the dynamics of the particle by appropriate Ornstein–Uhlenbeck processes. Of course, this result is valid only for finite times: at sufficiently long times the particle can escape from the one local minimum, a say, and surmount the potential barrier to end up at c. It will then spend a long time in the neighborhood of c until it escapes again the potential barrier and end at a. This is an example of a **rare event**. The relevant time scale, the **exit time** or the **mean first passage time** scales exponentially in $\beta := (k_B T)^{-1}$:

$$\tau = \nu^{-1} \exp(\beta \Delta E).$$

It is more customary to calculate the **reaction rate** $\kappa := \tau^{-1}$ which gives the rate with which particles escape from a local minimum of the potential:

$$\kappa = \nu \exp(-\beta \Delta E). \tag{9.3}$$

It is very important to notice that the escape from a local minimum, i.e. a state of local stability, can happen only at positive temperatures: it is a noise assisted event. Indeed, consider the case $T = 0$. The equation of motion becomes

$$\dot{x} = -V'(x), \quad x(0) = x_0.$$

In this case the potential becomes a **Lyapunov function**:

$$\frac{dx}{dt} = V'(x)\frac{dx}{dt} = -(V'(x))^2 < 0.$$

Hence, depending on the initial condition the particle will converge either to a or c. The particle cannot escape from either state of local stability.

On the other hand, at high temperatures the particle does not "see" the potential barrier: it essentially jumps freely from one local minimum to another.

To get a better understanding of the dependence of the dynamics on the depth of the potential barrier relative to temperature, we solve the equation of motion (9.1) numerically. In Figure we present the time series of the particle position. We observe that at small temperatures the particle spends most of its time around $x = \pm 1$ with rapid transitions from -1 to 1 and back.

9.2 Kramers' Theory

9.3 The Mean First Passage Time Approach

The Arrhenius-type factor in the formula for the reaction rate, eqn. (9.3) is intuitively and it has been observed experimentally in the late nineteenth century by Arrhenius and others. What is extremely important both from a theoretical and an applied point of view is the calculation of the prefactor ν, the **rate coefficient**. A systematic approach for the calculation of the rate coefficient, as well as the justification of the Arrhenius kinetics, is that of the mean first passage time method (MFPT). Since this method is of independent interest and is useful in various other contexts, we will present it in a quite general setting and apply it to the problem of the escape from a potential barrier in later sections. We will first treat the one dimensional problem and then extend the theory to arbitrary finite dimensions.

We will restrict ourselves to the case of homogeneous Markov processes. It is not very easy to extend the method to non-Markovian processes.

9.3.1 The Boundary Value Problem for the MFPT

- Let X_t be a continuous time diffusion process on $\mathbb{R}^d$ whose evolution is governed by the SDE

$$dX_t^x = b(X_t^x)\,dt + \sigma(X_t^x)\,dW_t, \quad X_0^x = x. \tag{9.4}$$

- Let D be a bounded subset of $\mathbb{R}^d$ with smooth boundary. Given $x \in D$, we want to know how long it takes for the process X_t to leave the domain D for the first time

$$\tau_D^x = \inf\left\{t \geqslant 0 \,:\, X_t^x \notin D\right\}.$$

- Clearly, this is a random variable. The average of this random variable is called the **mean first passage time** MFPT or the **first exit time**:

$$\tau(x) := \mathbb{E}\tau_D^x.$$

- We can calculate the MFPT by solving an appropriate boundary value problem.

Theorem 9.3.1. *The MFPT is the solution of the boundary value problem*

$$-\mathcal{L}\tau = 1, \quad x \in D, \tag{9.5a}$$

$$\tau = 0, \quad x \in \partial D, \tag{9.5b}$$

where $\mathcal{L}$ is the generator of the SDE 9.5.

- The homogeneous Dirichlet boundary conditions correspond to an **absorbing boundary**: the particles are removed when they reach the boundary. Other choices of boundary conditions are also possible.
- The rigorous proof of Theorem 9.3.1 is based on Itô's formula.

Derivation

- Let $\rho(X, x, t)$ be the probability distribution of the particles that **have not left** the domain D at time t. It solves the FP equation with absorbing boundary conditions.

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho, \quad \rho(X, x, 0) = \delta(X - x), \quad \rho|_{\partial D} = 0. \tag{9.6}$$

- We can write the solution to this equation in the form

$$\rho(X, x, t) = e^{\mathcal{L}^* t}\delta(X - x),$$

- where the absorbing boundary conditions are included in the definition of the semigroup $e^{\mathcal{L}^* t}$.
- The homogeneous Dirichlet (absorbing) boundary conditions imply that

$$\lim_{t\to+\infty} \rho(X, x, t) = 0.$$

- That is: all particles will eventually leave the domain.
- The (normalized) number of particles that are still inside D at time t is

$$S(x, t) = \int_D \rho(X, x, t)\, dx.$$

- Notice that this is a decreasing function of time. We can write

$$\frac{\partial S}{\partial t} = -f(x, t),$$

- where $f(x,t)$ is the **first passage times distribution**.
- The MFPT is the first moment of the distribution $f(x,t)$:

$$\begin{aligned}\tau(x) &= \int_0^{+\infty} f(s,x)s\,ds = \int_0^{+\infty} -\frac{dS}{ds}s\,ds \\ &= \int_0^{+\infty} S(s,x)\,ds = \int_0^{+\infty}\int_D \rho(X,x,s)\,dX ds \\ &= \int_0^{+\infty}\int_D e^{\mathcal{L}^* s}\delta(X-x)\,dX ds \\ &= \int_0^{+\infty}\int_D \delta(X-x)\Big(e^{\mathcal{L}s}1\Big)\,dX ds = \int_0^{+\infty}\Big(e^{\mathcal{L}s}1\Big)\,ds.\end{aligned}$$

- We apply $\mathcal{L}$ to the above equation to deduce:

$$\begin{aligned}\mathcal{L}\tau &= \int_0^{+\infty}\Big(\mathcal{L}e^{\mathcal{L}t}1\Big)\,dt = \int_0^t \frac{d}{dt}\Big(\mathcal{L}e^{\mathcal{L}t}1\Big)\,dt \\ &= -1.\end{aligned}$$

9.4 Escape from a potential barrier

In this section we use the theory developed in the previous section to study the long time/small temperature asymptotics of solutions to the Langevin equation for a particle moving in a one–dimensional potential of the form (9.2):

$$\ddot{x} = -V'(x) - \gamma\dot{x} + \sqrt{2\gamma k_B T}\dot{W}. \tag{9.7}$$

In particular, we justify the Arrhenius formula for the reaction rate

$$\kappa = \nu(\gamma)\exp(-\beta\Delta E)$$

and we calculate the escape rate $\nu = \nu(\gamma)$. In particular, we analyze the dependence of the escape rate on the friction coefficient. We will see that the we need to distinguish between the cases of large and small friction coefficients.

9.4.1 Calculation of the Reaction Rate in the Smoluchowski regime

We consider the Langevin equation (9.7) in the limit of large friction. As we saw in Chapter **??**, in the overdamped limit $\gamma \gg 1$, the solution to (9.7) can be approximated by the solution to the Smoluchowski equation (9.1)

$$\gamma\dot{x} = -V'(x) + \sqrt{2\gamma k_B T}\dot{W}.$$

We want to calculate the rate of escape from the potential barrier in this case. We assume that the particle is initially at x_0 which is near a, the left potential minimum.

- Consider the boundary value problem for the MFPT of the one dimensional diffusion process (9.1) from the interval (a,b):

$$-\beta^{-1}e^{\beta V}\partial_x\Big(e^{-\beta V}\tau\Big) = 1 \tag{9.8}$$

- We choose reflecting BC at $x = a$ and absorbing B.C. at $x = b$. We can solve (9.8) with these boundary conditions by quadratures:

$$\tau(x) = \beta^{-1} \int_x^b dy e^{\beta V(y)} \int_0^y dz e^{-\beta V(z)}. \tag{9.9}$$

- Now we can solve the problem of the escape from a potential well: the reflecting boundary is at $x = a$, the left local minimum of the potential, and the absorbing boundary is at $x = b$, the local maximum.

- We can replace the B.C. at $x = a$ by a repelling B.C. at $x = -\infty$:

$$\tau(x) = \beta^{-1} \int_x^b dy e^{\beta V(y)} \int_{-\infty}^y dz e^{-\beta V(z)}.$$

- When $E_b\beta \gg 1$ the integral wrt z is dominated by the value of the potential near a. Furthermore, we can replace the upper limit of integration by ∞:

$$\begin{aligned} \int_{-\infty}^z \exp(-\beta V(z))\, dz &\approx \int_{-\infty}^{+\infty} \exp(-\beta V(a)) \exp\left(-\frac{\beta\omega_0^2}{2}(z-a)^2\right) dz \\ &= \exp\left(-\beta V(a)\right) \sqrt{\frac{2\pi}{\beta\omega_0^2}}, \end{aligned}$$

- where we have used the Taylor series expansion around the minimum:

$$V(z) = V(a) + \frac{1}{2}\omega_0^2(z-a)^2 + \dots$$

- Similarly, the integral wrt y is dominated by the value of the potential around the saddle point. We use the Taylor series expansion

$$V(y) = V(b) - \frac{1}{2}\omega_b^2(y-b)^2 + \dots$$

- Assuming that x is close to a, the minimum of the potential, we can replace the lower limit of integration by $-\infty$. We finally obtain

$$\begin{aligned} \int_x^b \exp(\beta V(y))\, dy &\approx \int_{-\infty}^b \exp(\beta V(b)) \exp\left(-\frac{\beta\omega_b^2}{2}(y-b)^2\right) dy \\ &= \frac{1}{2} \exp\left(\beta V(b)\right) \sqrt{\frac{2\pi}{\beta\omega_b^2}}. \end{aligned}$$

- Putting everything together we obtain a formula for the MFPT:

$$\tau(x) = \frac{\pi}{\omega_0\omega_b} \exp\left(\beta E_b\right).$$

- The rate of arrival at b is $1/\tau$. Only have of the particles escape. Consequently, the escape rate (or reaction rate), is given by $\frac{1}{2\tau}$:

$$\kappa = \frac{\omega_0\omega_b}{2\pi} \exp\left(-\beta E_b\right).$$

9.4.2 The Intermediate Regime: $\gamma = O(1)$

- Consider now the problem of escape from a potential well for the Langevin equation

$$\ddot{q} = -\partial_q V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W}. \tag{9.10}$$

- The reaction rate depends on the fiction coefficient and the temperature. In the overdamped limit ($\gamma \gg 1$) we retrieve (**??**), appropriately rescaled with γ:

$$\kappa = \frac{\omega_0 \omega_b}{2\pi\gamma} \exp\left(-\beta E_b\right). \tag{9.11}$$

- We can also obtain a formula for the reaction rate for $\gamma = \mathcal{O}(1)$:

$$\kappa = \frac{\sqrt{\frac{\gamma^2}{4} - \omega_b^2} - \frac{\gamma}{2}}{\omega_b} \frac{\omega_0}{2\pi} \exp\left(-\beta E_b\right). \tag{9.12}$$

- Naturally, in the limit as $\gamma \to +\infty$ (9.12) reduces to (9.11)

9.4.3 Calculation of the Reaction Rate in the energy-diffusion-limited regime

- In order to calculate the reaction rate in the underdamped or **energy-diffusion-limited** regime $\gamma \ll 1$ we need to study the diffusion process for the energy, (8.31) or (8.32). The result is

$$\kappa = \gamma\beta I(E_b) \frac{\omega_0}{2\pi} e^{-\beta E_b}, \tag{9.13}$$

- where $I(E_b)$ denotes the action evaluated at b.

- A formula for the escape rate which is valid for all values of friction coefficient was obtained by Melnikov and Meshkov in 1986, J. Chem. Phys 85(2) 1018-1027. This formula requires the calculation of integrals and it reduced to (9.11) and (9.13) in the overdamped and underdamped limits, respectively.

9.5 Extension to Higher Dimensions

9.6 Discussion and Bibliography

The calculation of reaction rates and the stochastic modeling of chemical reactions has been a very active area of research since the 30's. One of the first methods that were developed was that of **transition state theory cite WIGNER**. Kramers developed his theory in his celebrated paper **CITE KRAMERS**. In this chapter we have based our approach to the calculation of the mean first passage time. Our analysis is based mostly on [10, Ch. 5, Ch. 9], [26, Ch. 4] and the excellent review article [13]. We highly recommend this review article for further information on reaction rate theory. See also [12] and the review article of Melnikov (1991).

There are many applications of interest where it is important to calculate reaction rates for non-Markovian Langevin equations of the form

$$\ddot{x} = -V'(x) - \int_0^t b\gamma(t-s)\dot{x}(s)\,ds + \xi(t) \tag{9.14a}$$

$$\langle \xi(t)\xi(0)\rangle = k_B T M^{-1}\gamma(t) \tag{9.14b}$$

We will derive generalized non–Markovian equations of the form (9.14a), together with the fluctuation–dissipation theorem (9.14b), in Chapter **??**. The calculation of reaction rates for the generalized Langevin equation is presented in [12].

The long time/small temperature asymptotics can be studied rigorously by means of the theory of Freidlin-Wentzell [6]. See also [2]. A related issue is that of the small temperature asymptotics for the eigenvalues (in particular, the first eigenvalue) of the generator of the Markov process $x(t)$ which is the solution of

$$\gamma\dot{x} = -\nabla V(x) + \sqrt{2\gamma k_B T}\dot{W}.$$

The theory of Freidlin and Wentzell has also been extended to infinite dimensional problems. This is a very important problem in many applications such as micromagnetics...We refer to **CITE...** for more details.

A systematic study of the problem of the escape from a potential well was developed by Matkowsky, Schuss and collaborators [21, 17, 18]. This approach is based on a systematic use of singular perturbation theory. In particular, the calculation of the transition rate which is uniformly valid in the friction coefficient is presented in [18]. This formula is obtained through a careful analysis of the PDE

$$p\partial_q\tau - \partial_q V\partial_p\tau + \gamma(-p\partial_p + k_B T\partial_p^2)\tau = -1,$$

for the mean first passage time τ. The PDE is equipped, of course, with the appropriate boundary conditions. Singular perturbation theory is used to study the small temperature asymptotics of solutions to the boundary value problem. The formula derived in this paper reduces to the formulas which are valid at large and small values of the friction coefficient at the appropriate asymptotic limits.

The study of rare transition events between long lived metastable states is a key feature in many systems in physics, chemistry and biology. Rare transition events play an important role, for example, in the analysis of the transition between different conformation states of biological macromolecules such as DNA [22]. The study of rare events is one of the most active research areas in the applied stochastic processes. Recent developments in this area involve the transition path theory of W. E and Vanden Eijnden. Various simple applications of this theory are presented in Metzner, Schutte et al 2006. As in the mean first passage time approach, transition path theory is also based on the solution of an appropriate boundary value problem for the so-called commitor function.

Bibliography

[1] L. Arnold. *Stochastic differential equations: theory and applications*. Wiley-Interscience [John Wiley & Sons], New York, 1974. Translated from the German.

[2] N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems*. Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2006. A sample-paths approach.

[3] S.N. Ethier and T.G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[4] L.C. Evans. *Partial Differential Equations*. AMS, Providence, Rhode Island, 1998.

[5] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[6] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of dunamical systems*. Springer-Verlag, New York, 1984.

[7] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[8] A. Friedman. *Stochastic differential equations and applications. Vol. 1*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. Probability and Mathematical Statistics, Vol. 28.

[9] A. Friedman. *Stochastic differential equations and applications. Vol. 2*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. Probability and Mathematical Statistics, Vol. 28.

[10] C. W. Gardiner. *Handbook of stochastic methods*. Springer-Verlag, Berlin, second edition, 1985. For physics, chemistry and the natural sciences.

[11] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes*. Dover Publications Inc., Mineola, NY, 1996.

[12] P. Hanggi. Escape from a metastable state. *J. Stat. Phys.*, 42(1/2):105–140, 1986.

[13] P. Hanggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Modern Phys.*, 62(2):251–341, 1990.

[14] W. Horsthemke and R. Lefever. *Noise-induced transitions*, volume 15 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1984. Theory and applications in physics, chemistry, and biology.

[15] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.

[16] P. Mandl. *Analytical treatment of one-dimensional Markov processes*. Die Grundlehren der mathematischen Wissenschaften, Band 151. Academia Publishing House of the Czechoslovak Academy of Sciences, Prague, 1968.

[17] B. J. Matkowsky, Z. Schuss, and E. Ben-Jacob. A singular perturbation approach to Kramers' diffusion problem. *SIAM J. Appl. Math.*, 42(4):835–849, 1982.

[18] B. J. Matkowsky, Z. Schuss, and C. Tier. Uniform expansion of the transition rate in Kramers' problem. *J. Statist. Phys.*, 35(3-4):443–456, 1984.

[19] Frigyes Riesz and Béla Sz.-Nagy. *Functional analysis*. Dover Publications Inc., New York, 1990. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.

[20] H. Risken. *The Fokker-Planck equation*, volume 18 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1989.

[21] Z. Schuss. Singular perturbation methods in stochastic differential equations of mathematical physics. *SIAM Review*, 22(2):119–155, 1980.

[22] Ch. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In *Handbook of Numerical Analysis (Computational Chemistry), Vol X*, 2003.

[23] C. Schwab and R.A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[24] D.W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993.

[25] G. I. Taylor. Diffusion by continuous movements. *London Math. Soc.*, 20:196, 1921.

[26] R. Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, New York, 2001.