APPLIED STOCHASTIC PROCESSES

G.A. Pavliotis

Department of Mathematics Imperial College London, UK

January 16, 2011

- Lectures: Mondays, 10:00-12:00, Huxley 6M42.
- Office Hours: By appointment.
- Course webpage: `http://www.ma.imperial.ac.uk/~pavl/stoch_proc.htm`
- Text: Lecture notes, available from the course webpage. Also, recommended reading from various textbooks/review articles.
- **The lecture notes are still in progress. Please send me your comments, suggestions and let me know of any typos/errors that you have spotted.**

- This is a basic graduate course on stochastic processes, aimed towards PhD students in applied mathematics and theoretical physics.
- The emphasis of the course will be on the presentation of analytical tools that are useful in the study of stochastic models that appear in various problems in applied mathematics, physics, chemistry and biology.
- The course will consist of three parts: Fundamentals of the theory of stochastic processes, applications (reaction rate theory, surface diffusion....) and non-equilibrium statistical mechanics.

1. PART I: FUNDAMENTALS OF CONTINUOUS TIME STOCHASTIC PROCESSES
   - Elements of probability theory.
   - Stochastic processes: basic definitions, examples.
   - Continuous time Markov processes. Brownian motion
   - Diffusion processes: basic definitions, the generator.
   - Backward Kolmogorov and the Fokker–Planck (forward Kolmogorov) equations.
   - Stochastic differential equations (SDEs); Itô calculus, Itô and Stratonovich stochastic integrals, connection between SDEs and the Fokker–Planck equation.
   - Methods of solution for SDEs and for the Fokker-Planck equation.
   - Ergodic properties and convergence to equilibrium.

2. PART II: APPLICATIONS.
   - Asymptotic problems for the Fokker–Planck equation: overdamped (Smoluchowski) and underdamped (Freidlin-Wentzell) limits.
   - Bistable stochastic systems: escape over a potential barrier, mean first passage time, calculation of escape rates etc.
   - Brownian motion in a periodic potential. Stochastic models of molecular motors.
   - Multiscale problems: averaging and homogenization.

3. PART III: NON-EQUILIBRIUM STATISTICAL MECHANICS.
   - Derivation of stochastic differential equations from deterministic dynamics (heat bath models, projection operator techniques etc.).
   - The fluctuation-dissipation theorem.
   - Linear response theory.
   - Derivation of macroscopic equations (hydrodynamics) and calculation of transport coefficients.

4. ADDITIONAL TOPICS (time permitting): numerical methods, stochastic PDES, Markov chain Monte Carlo (MCMC)

Prerequisites

- Basic knowledge of ODEs and PDEs.
- Elementary probability theory.
- Some familiarity with the theory of stochastic processes.

- Lecture notes will be provided for all the material that we will cover in this course. They will be posted on the course webpage.
- There are many excellent textbooks/review articles on applied stochastic processes, at a level and style similar to that of this course.
- Standard textbooks are
  - Gardiner: Handbook of stochastic methods (1985).
  - Van Kampen: Stochastic processes in physics and chemistry (1981).
  - Horsthemke and Lefever: Noise induced transitions (1984).
  - Risken: The Fokker-Planck equation (1989).
  - Oksendal: Stochastic differential equations (2003).
  - Mazo: Brownian motion: fluctuations, dynamics and fluctuations (2002).
  - Bhatthacharya and Waymire, Stochastic Processes and Applications (1990).

- Other standard textbooks are
  - Nelson: Dynamical theories of Brownian motion (1967). Available from the web (includes a very interesting historical overview of the theory of Brownian motion).
  - Chorin and Hald: Stochastic tools in mathematics and science (2006).
  - Swanzing: Non-equilibrium statistical mechanics (2001).
- The material on multiscale methods for stochastic processes, together with some of the introductory material, will be taken from
  - Pavliotis and Stuart: Multiscale methods: averaging and homogenization (2008).

- Excellent books on the mathematical theory of stochastic processes are
  - Karatzas and Shreeve: Brownian motion and stochastic calculus (1991).
  - Revuz and Yor: Continuous martingales and Brownian motion (1999).
  - Stroock: Probability theory, an analytic view (1993).

- Well known review articles (available from the web) are:
  - Chandrasekhar: Stochastic problems in physics and astronomy (1943).
  - Hanggi and Thomas: Stochastic processes: time evolution, symmetries and linear response (1982).
  - Hanggi, Talkner and Borkovec: Reaction rate theory: fifty years after Kramers (1990).

- 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).
    - . . .

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**.

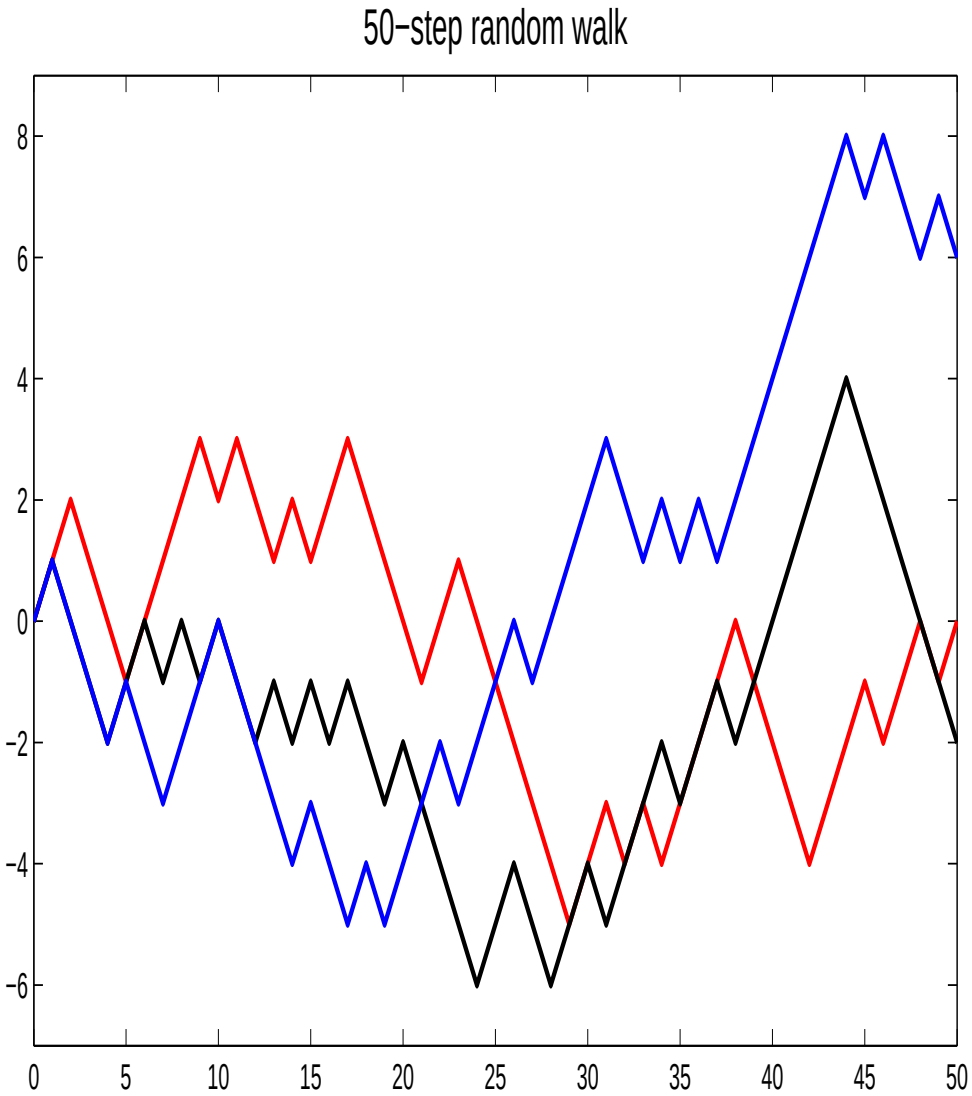

Figure: Three paths of the random walk of length $N = 50$.

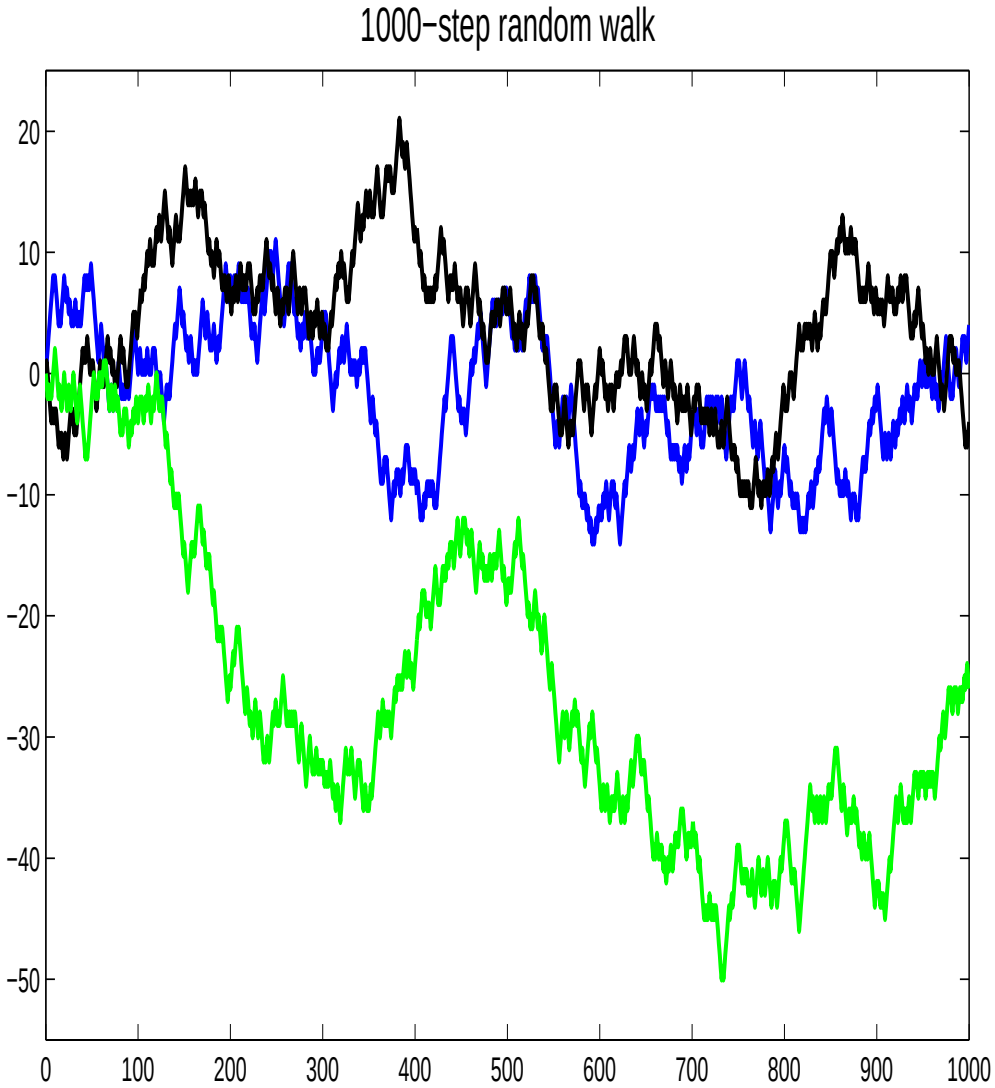

Figure: Three paths of the random walk of length $N = 1000$.

- 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** (the probabilistic law of evolution is independent of time) **Markov** (the future depends only on the present and not on the past) process.
- If we take a large number of steps, the random walk starts looking like a continuous time process with continuous paths.

- 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** $D = \frac{\Delta x^2}{2\Delta t} = \frac{1}{2}$.

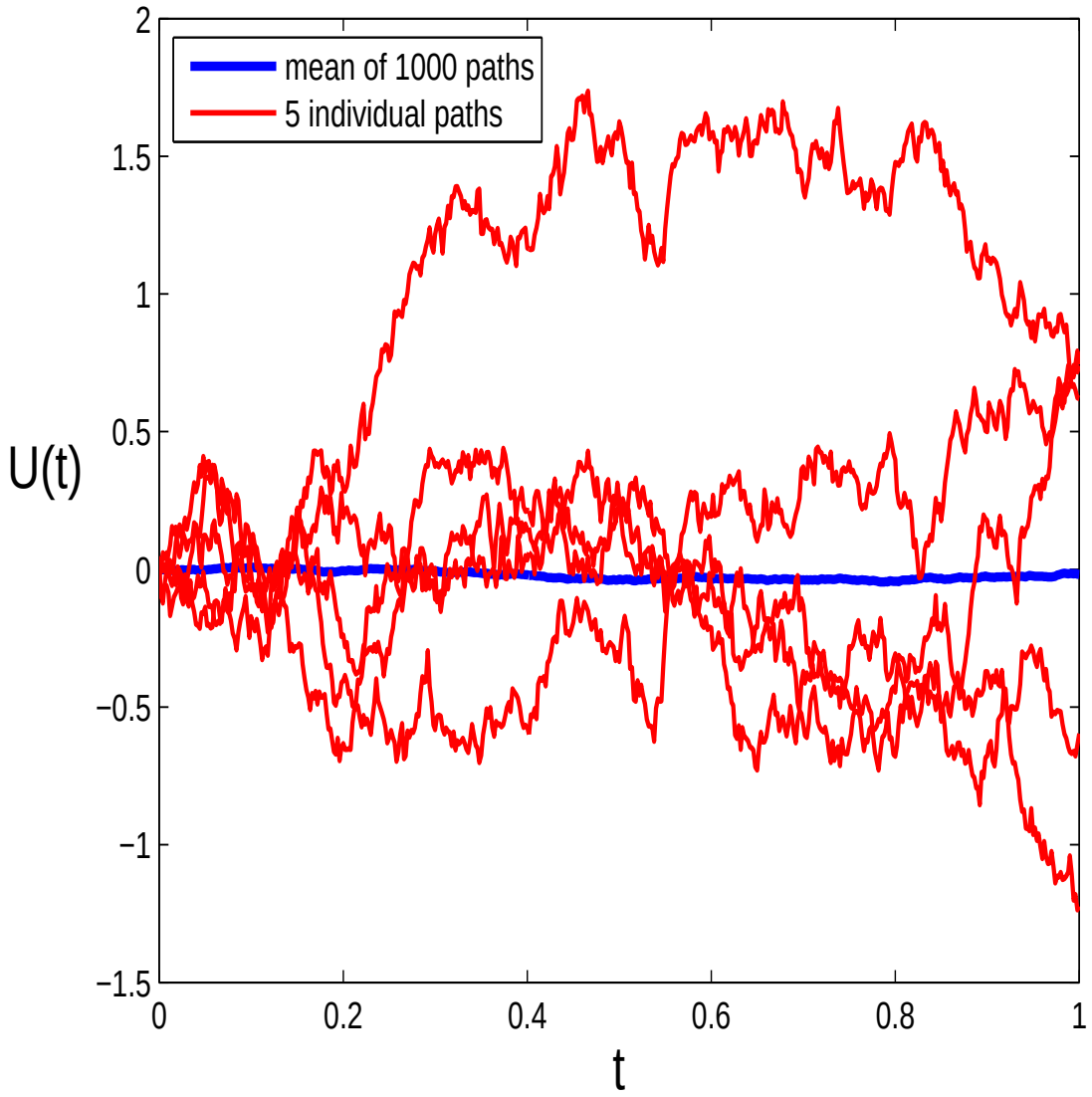

Figure: Sample Brownian paths.

- 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.

- We can write an equation for the evolution of the paths of a Brownian motion X_t with diffusion coefficient D starting at x:

$$dX_t = \sqrt{2D}dW_t, \quad X_0 = x.$$

- This is an example of a **stochastic differential equation**.
- The probability of finding X_t at y at time t, given that it was at x at time $t = 0$, the **transition probability density** $\rho(y, t)$ satisfies the PDE

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial y^2}, \quad \rho(y, 0) = \delta(y - x).$$

- This is an example of the **Fokker-Planck equation**.
- The connection between Brownian motion and the diffusion equation was made by Einstein in 1905.

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 system for which we have incomplete information: we have imprecise knowledge of initial and boundary conditions or of model parameters.
  - ODEs with random initial conditions are equivalent to stochastic processes that can be described using stochastic differential equations.
- To describe systems for which we are not confident about the validity of our mathematical model.

- To describe a dynamical system exhibiting very complicated behavior (chaotic dynamical systems). Determinism versus predictability.
- To describe a high dimensional deterministic system 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.

ELEMENTS OF PROBABILITY THEORY

- A collection of subsets of a set Ω is called a σ–algebra if it contains Ω and is closed under the operations of taking complements and countable unions of its elements.
- A sub-σ–algebra is a collection of subsets of a σ–algebra which satisfies the axioms of a σ–algebra.
- A *measurable space* is a pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ is a σ–algebra of subsets of Ω.
- Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces. A function $X : \Omega \mapsto E$ such that the *event*

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

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

- Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \mapsto [0, 1]$ is called a *probability measure* if $\mu(\emptyset) = 1$, $\mu(\Omega) = 1$ and $\mu(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$ for all sequences of pairwise disjoint sets $\{A_k\}_{k=1}^{\infty} \in \mathcal{F}$.
- The triplet $(\Omega, \mathcal{F}, \mu)$ is called a *probability space*.
- 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

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: $\sum_{i \in \mathcal{I}} \rho_{0,i} = 1$.

- 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

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

Example

- 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

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

Example (Continued)

- 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

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

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

$$\mathbb{E}(X) = m \tag{1}$$

and the covariance matrix is

$$\mathbb{E}\Big((X-m)\otimes(X-m)\Big) = \Sigma. \tag{2}$$

- 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 $\mathcal{N}(m, \Sigma)$.

The Characteristic Function

- Many of the properties of (sums of) random variables can be studied using the Fourier transform of the distribution function.
- The **characteristic function** of the random variable X is defined to be the Fourier transform of the distribution function

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda} \, d\mu_X(\lambda) = \mathbb{E}(e^{itX}). \tag{3}$$

- The characteristic function determines uniquely the distribution function of the random variable, in the sense that there is a one-to-one correspondance between $F(\lambda)$ and $\phi(t)$.
- The characteristic function of a $\mathcal{N}(m, \Sigma)$ is

$$\phi(t) = e^{\langle m,t\rangle - \frac{1}{2}\langle t,\Sigma t\rangle}.$$

Lemma

Let $\{X_1, X_2, \ldots X_n\}$ be independent random variables with characteristic functions $\phi_j(t)$, $j = 1, \ldots n$ and let $Y = \sum_{j=1}^{n} X_j$ with characteristic function $\phi_Y(t)$. Then

$$\phi_Y(t) = \Pi_{j=1}^{n} \phi_j(t).$$

Lemma

Let X be a random variable with characteristic function $\phi(t)$ and assume that it has finite moments. Then

$$E(X^k) = \frac{1}{i^k} \phi^{(k)}(0).$$

Types of Convergence and Limit Theorems

- One of the most important aspects of the theory of random variables is the study of limit theorems for sums of random variables.
- The most well known limit theorems in probability theory are the law of large numbers and the central limit theorem.
- There are various different types of convergence for sequences or random variables.

Definition

Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of random variables. We will say that

(a) Z_n converges to Z with probability one (almost surely) if

$$\mathbb{P}\Big(\lim_{n\to+\infty} Z_n = Z\Big) = 1.$$

(b) Z_n converges to Z in probability if for every $\varepsilon > 0$

$$\lim_{n\to+\infty} \mathbb{P}\big(|Z_n - Z| > \varepsilon\big) = 0.$$

(c) Z_n converges to Z in L^p if

$$\lim_{n\to+\infty} \mathbb{E}\big[\big|Z_n - Z\big|^p\big] = 0.$$

(d) Let $F_n(\lambda), n = 1, \cdots + \infty$, $F(\lambda)$ be the distribution functions of $Z_n\, n = 1, \cdots + \infty$ and Z, respectively. Then Z_n converges to Z in distribution if

$$\lim_{n\to+\infty} F_n(\lambda) = F(\lambda)$$

- Let $\{X_n\}_{n=1}^{\infty}$ be iid random variables with $\mathbb{E}X_n = V$. Then, the **strong law of large numbers** states that average of the sum of the iid converges to V with probability one:

$$\mathbb{P}\left(\lim_{n\to+\infty}\frac{1}{N}\sum_{n=1}^{N}X_n = V\right) = 1.$$

- The strong law of large numbers provides us with information about the behavior of a sum of random variables (or, a large number or repetitions of the same experiment) on average.
- We can also study fluctuations around the average behavior. Indeed, let $\mathbb{E}(X_n - V)^2 = \sigma^2$. Define the centered iid random variables $Y_n = X_n - V$. Then, the sequence of random variables $\frac{1}{\sigma\sqrt{N}}\sum_{n=1}^{N} Y_n$ converges in distribution to a $\mathcal{N}(0,1)$ random variable:

$$\lim_{n\to+\infty}\mathbb{P}\left(\frac{1}{\sigma\sqrt{N}}\sum_{n=1}^{N}Y_n \leqslant a\right) = \int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx.$$

- 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.

ELEMENTS OF THE THEORY OF STOCHASTIC PROCESSES

- 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

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.

Let $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ 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

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 second moment $\mathbb{E}(X_t X_s)$ depends only on the difference $t - s$:

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

- The constant m is called the **expectation** of the process X_t. We will set $m = 0$.
- The function $C(t)$ is called 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 $C(t) = C(-t)$, $t \in \mathbb{R}$.

Continuity properties of the covariance function are equivalent to continuity properties of the paths of X_t.

Lemma

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:

$$\begin{aligned} |C(t+h) - C(t)|^2 &= |\mathbb{E}(X_{t+h}X_0) - \mathbb{E}(X_tX_0)|^2 = \mathbb{E}|((X_{t+h} - X_t)X_0)|^2 \\ &\leqslant \mathbb{E}(X_0)^2\mathbb{E}(X_{t+h} - X_t)^2 \\ &= C(0)(\mathbb{E}X_{t+h}^2 + \mathbb{E}X_t^2 - 2\mathbb{E}X_tX_{t+h}) \\ &= 2C(0)(C(0) - C(h)) \to 0, \end{aligned}$$

as $h \to 0$.
Assume now that $C(t)$ is continuous. From the above calculation we have

$$\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \tag{4}$$

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 (4) 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

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

$$\sum_{i,j=1}^{n} f(t_i - t_j) c_i \bar{c}_j \geqslant 0 \tag{5}$$

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

Lemma

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

*(**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*

$$C(t) = \int_{\mathbb{R}} e^{ixt}\, \rho(dx) \tag{6}$$

in nonnegative definite. Conversely, any nonnegative definite function can be represented in the form (6)*.*

Definition

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 (6) 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:

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

- In most cases, the experimentally measured quantity is the spectral density (or power spectrum) of the stochastic process.

- 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. We have to assume sufficiently fast decay of correlations so that the correlation time is finite.

Example

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

$$R(t) = \frac{D}{\alpha} e^{-\alpha |t|}. \tag{7}$$

- The spectral density of this process is:

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

Example (Continued)

- This function is called the **Cauchy** or the **Lorentz** distribution.
- The Gaussian stochastic process with covariance function (7) is called the **stationary Ornstein-Uhlenbeck process**.
- The correlation time is (we have that $C(0) = D/(\alpha)$)

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

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

$$X(t) = \int_0^t Y(s)\, ds, \tag{8}$$

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

Lemma

Let $Y(t)$ denote the stationary OU process with covariance function (7)*. Then the position process* (8) *is a mean zero Gaussian process with covariance function*

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

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

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

$$\lim_{T\to+\infty} \mathbb{E}\left|\frac{1}{T}\int_0^T X(s)\,ds - \mu\right|^2 = 0. \tag{9}$$

Proof.

We have

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

using the dominated convergence theorem and the assumption $R(\cdot)\in L^1$. In the above we used the fact that R is a symmetric function, together with the change of variables $u=t-s$, $v=t$ and an integration over v. □

Assume that $\mu = 0$. From the above calculation we can conclude that, under the assumption that $R(\cdot)$ decays sufficiently fast at infinity, for $t \gg 1$ we have that

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

where

$$D = \int_0^\infty R(t)\,dt$$

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.
This kind of analysis was initiated in G.I. Taylor, *Diffusion by Continuous Movements* Proc. London Math. Soc..1922; s2-20: 196-212.

Definition

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 $(X(t_1), X(t_2), \ldots, X(t_k))$ is equal to that of $(X(s+t_1), X(s+t_2), \ldots, X(s+t_k))$ for any s such that $s + t_i \in T$ for all $i \in \{1, \ldots, k\}$. In other words,

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

- 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.

Remarks

1. A sequence $Y_0, Y_1, \ldots$ of independent, identically distributed random variables is a stationary process with $R(k) = \sigma^2 \delta_{k0}$, $\sigma^2 = \mathbb{E}(Y_k)^2$.
2. Let Z be a single random variable with known distribution and set $Z_j = Z$, $j = 0, 1, 2, \ldots$. Then the sequence $Z_0, Z_1, Z_2, \ldots$ is a stationary sequence with $R(k) = \sigma^2$.
3. 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.

- 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 $t_0 < t_1 ... t_n$ the random variables

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

  are independent.
- If, furthermore, for any t_1, t_2 and Borel set $B \subset \mathbb{R}$

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

  is independent of s, then the process X_t has **stationary independent increments**.

Definition

- A one dimensional standard *Brownian motion* $W(t) : \mathbb{R}^+ \to \mathbb{R}$ is a real valued stochastic process with the following properties:
  1. $W(0) = 0$;
  2. $W(t)$ is continuous;
  3. $W(t)$ has independent increments.
  4. 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

$$g(x; t, s) = \Big(2\pi(t-s)\Big)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2(t-s)}\right); \qquad (10)$$

Definition (Continued)

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

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

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

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

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

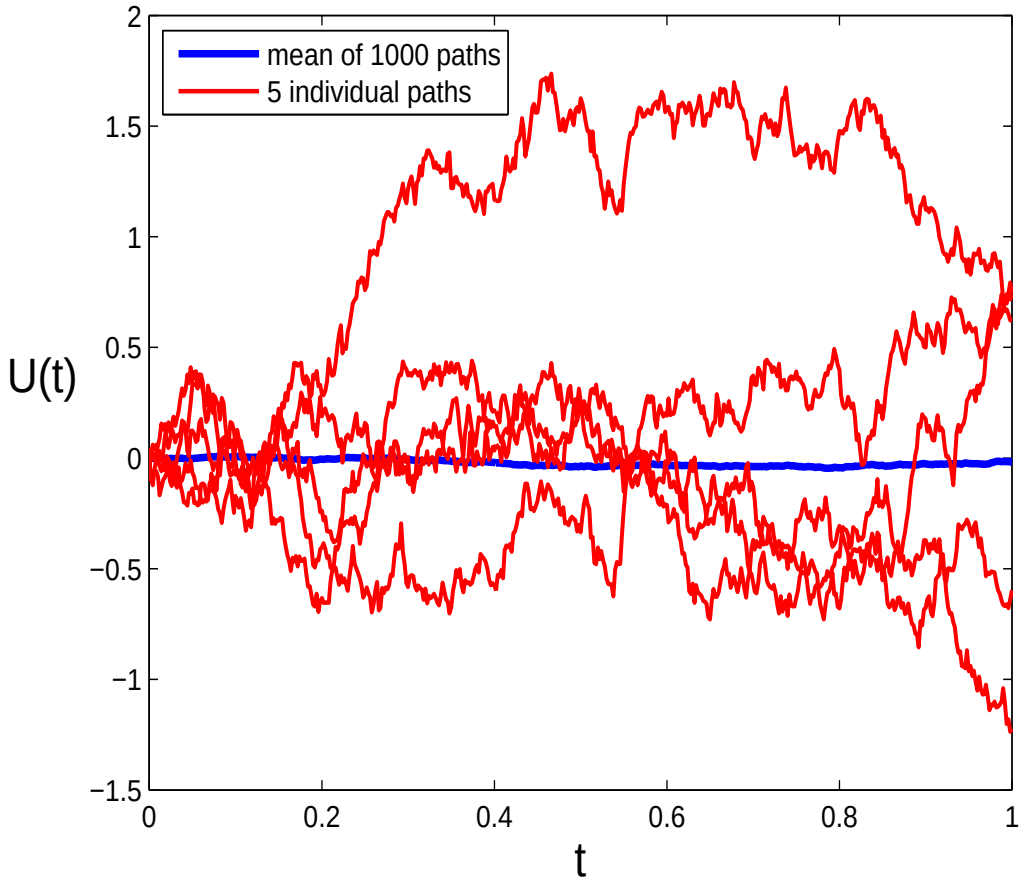

Figure: Brownian sample paths

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

Theorem

(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.

Brownian motion is a Gaussian process. For the d–dimensional Brownian motion, and for I the $d \times d$ dimensional identity, we have (see (1) and (2))

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

and

$$\mathbb{E}\Big((W(t) - W(s)) \otimes (W(t) - W(s))\Big) = (t - s)I. \tag{11}$$

Moreover,

$$\mathbb{E}\Big(W(t) \otimes W(s)\Big) = \min(t, s)I. \tag{12}$$

- From the formula for the Gaussian density $g(x, t-s)$, eqn. (10), 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:

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

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

Lemma

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

$$K(s,t) = e^{-|t-s|}.$$

Definition

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

$$\mathbb{E}(W_t^H W_s^H) = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right). \tag{13}$$

Fractional Brownian motion has the following properties.

1. When $H = \frac{1}{2}$, $W_t^{\frac{1}{2}}$ becomes the standard Brownian motion.
2. $W_0^H = 0$, $\mathbb{E}W_t^H = 0$, $\mathbb{E}(W_t^H)^2 = |t|^{2H}$, $t \geqslant 0$.
3. It has stationary increments, $\mathbb{E}(W_t^H - W_s^H)^2 = |t-s|^{2H}$.
4. 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.

The Poisson Process

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

Definition

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 $t - s$.

- A very useful result is that we can expand every centered ($\mathbb{E}X_t = 0$) stochastic process with continuous covariance (and hence, every L^2-continuous centered stochastic process) into a random Fourier series.
- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_t\}_{t\in T}$ a centered process which is mean square continuous. Then, the **Karhunen-Loéve** theorem states that X_t can be expanded in the form

$$X_t = \sum_{n=1}^{\infty} \xi_n \phi_n(t), \tag{14}$$

- where the ξ_n are an orthogonal sequence of random variables with $\mathbb{E}|\xi_k|^2 = \lambda_k$, where $\{\lambda_k\ \phi_n\}_{k=1}^{\infty}$ are the eigenvalues and eigenfunctions of the integral operator whose kernel is the covariance of the processes X_t. The convergence is in $L^2(\mathbb{P})$ for every $t \in T$.
- If X_t is a Gaussian process then the ξ_k are independent Gaussian random variables.

- Set $T = [0, 1]$.Let $\mathcal{K} : L^2(T) \to L^2(T)$ be defined by

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

- The kernel of this integral operator is a continuous (in both s and t), symmetric, nonnegative function. Hence, the corresponding integral operator has eigenvalues, eigenfunctions

$$\mathcal{K}\phi_n = \lambda_n, \quad \lambda_n \geqslant 0, \quad (\phi_n, \phi_m)_{L^2} = \delta_{nm},$$

- such that the covariance operator can be expanded in a uniformly convergent series

$$B(t,s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t)\phi_n(s).$$

- The random variables ξ_n are defined as

$$\xi_n = \int_0^1 X(t)\phi_n(t)\,dt.$$

- The orthogonality of the eigenfunctions of the covariance operator implies that

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

- Thus the random variables are orthogonal. When $X(t)$ is Gaussian, then ξ_k are Gaussian random variables. Furthermore, since they are also orthogonal, they are independent Gaussian random variables.

Example

The Karhunen-Loéve Expansion for Brownian Motion We set $T = [0, 1]$.The covariance function of Brownian motion is $C(t, s) = \min(t, s)$. The eigenvalue problem $\mathcal{C}\psi_n = \lambda_n \psi_n$ becomes

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

Or,

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

We differentiate this equation twice:

$$\int_t^1 \psi_n(s)\, ds = \lambda_n \psi_n'(t) \quad \text{and} \quad -\psi_n(t) = \lambda_n \psi_n''(t),$$

where primes denote differentiation with respect to t.

Example

From the above equations we immediately see that the right boundary conditions are $\psi(0) = \psi'(1) = 0$. The eigenvalues and 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

$$W_t = \frac{\sqrt{2}}{\pi}\sum_{n=1}^{\infty}\xi_n\frac{2}{(2n+1)\pi}\sin\left(\frac{1}{2}(2n+1)\pi t\right). \tag{15}$$

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

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**.

MARKOV STOCHASTIC PROCESSES

- Let $(\Omega, \mathcal{F})$ be a measurable space and T an ordered set. Let $X = X_t(\omega)$ be a stochastic process from the sample space $(\Omega, \mathcal{F})$ to the state space $(E, \mathcal{G})$. It is a function of two variables, $t \in T$ and $\omega \in \Omega$.
- For a fixed $\omega \in \Omega$ the function $X_t(\omega)$, $t \in T$ is the **sample path** of the process X associated with ω.
- 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 $\mathcal{K}$.
- 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** $\{X_t,\ t \in T\}$.

- 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

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

- We say that 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

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

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma | X_s) \tag{16}$$

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

- The filtration $\mathcal{F}_t^X$ is generated by events of the form $\{\omega | X_{s_1} \in B_1, X_{s_2} \in B_2, \ldots X_{s_n} \in B_n, \}$ 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

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

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*.
- A typical example of a Markov process is the random walk: in order to find the position $x(t+1)$ of the random walker at time $t+1$ it is enough to know its position $x(t)$ at time t: how it got to $x(t)$ is irrelevant.
- 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.

- 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, Γ) and satisfies the **Chapman–Kolmogorov** equation

$$P(\Gamma, t | x, s) = \int_E P(\Gamma, t | y, u) P(dy, u | x, s). \tag{17}$$

- 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:
  1. 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

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1). \tag{18}$$

  2. 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

$$\mathbb{E}(f(X)|\mathcal{G}) = \int_{\mathbb{R}} f(x) P_X(dx|\mathcal{G}). \tag{19}$$

Now we use the Markov property, together with equations (18) and (19) 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.
- 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

$$P(t + s, x, \Gamma) = \int_E P(s, x, dz) P(t, z, \Gamma). \tag{20}$$

- 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

$$\begin{aligned}
&\mathbb{P}(X_0 \in \Gamma_1, X(t_1) \in \Gamma_1, \ldots, X_{t_n} \in \Gamma_n) \\
&= \int_{\Gamma_0} \int_{\Gamma_1} \cdots \int_{\Gamma_{n-1}} P(t_n - t_{n-1}, y_{n-1}, \Gamma_n) P(t_{n-1} - t_{n-2}, y_{n-2}, dy_{n-1}) \\
&\cdots \times P(t_1, y_0, dy_1) \nu(dy_0). \qquad (21)
\end{aligned}$$

Theorem

(Ethier and Kurtz 1986, Sec. 4.1) Let $P(t, x, \Gamma)$ satisfy (20) 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 (21).

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:

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

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

$$p(y, t|x, s) = \int_{\mathbb{R}} p(y, t|z, u) p(z, u|x, s)\, dz. \tag{22}$$

- 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 (20):

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

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:

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

Consequently:

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

- 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

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

  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

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

  exists.

Definition

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

Definition

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

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(x,0) = f(x). \tag{23}$$

- When the semigroup P_t is the transition semigroup of a Markov process X_t, then equation (23) 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

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

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

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

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

Example

The Ornstein-Uhlenbeck process $V_t = e^{-t} W(e^{2t})$ is a homogeneous Markov process. The transition probability density is the Gaussian

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

The semigroup associated to the Ornstein-Uhlenbeck process is $P_t = e^{t(-x\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2})}$. The generator of this Markov process is $\mathcal{L} = -x\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2}$.

- Notice that the **transition probability density** $p(t,x)$ of the 1d OU process is the fundamental solution (Green's function) of the heat (diffusion) PDE

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

- 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:

$$\int_{\mathbb{R}} P_t f(x)\, d\mu(x) = \int_{\mathbb{R}} f(x)\, d(P_t^*\mu)(x). \tag{24}$$

- 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 (23) 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:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad \rho(y,0) = \rho_0(y). \tag{25}$$

- 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.

- 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

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.

- 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

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

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

- 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:

$$\mathcal{L}^*\rho = 0, \quad \rho \geqslant 0, \quad \|\rho\|_{L_1(\mathbb{R})} = 1. \tag{26}$$

Example (Continued)

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

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

- where

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

- We can calculate the invariant distribution by solving equation (26).
- The invariant measure of this process is the Gaussian measure

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

- 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 $\mathbb{R}$.

DIFFUSION PROCESSES

- 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

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

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

$$\int_{|x-y|>\varepsilon} P(dy, t|x, s) = o(t-s) \tag{27}$$

uniformly over $s < t$.

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

$$\int_{|y-x|\leqslant\varepsilon} (y-x)P(dy, t|x, s) = a(x, s)(s-t) + o(s-t). \tag{28}$$

uniformly over $s < t$.

Definition (Continued)

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

$$\int_{|y-x|\leqslant\varepsilon} (y-x)^2 P(dy,t|x,s) = b(x,s)(s-t) + o(s-t). \qquad (29)$$

uniformly over $s < t$.

In Definition 39 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

$$\lim_{t \to s} \frac{1}{t-s} \int_{\mathbb{R}^d} |y-x|^{2+\delta} P(s,x,t,dy) = 0, \tag{30}$$

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 (30) 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 (30) is sufficient for the sample paths to be continuous ($k=0$) and for the replacement of the truncated integrals in (28) and (29) by integrals over $\mathbb{R}^d$ ($k=1$ and $k=2$, respectively). The definitions of the drift and diffusion coefficients become:

$$\lim_{t\to s}\mathbb{E}\left(\frac{X_t-X_s}{t-s}\Big|X_s=x\right)=\mathbf{a}(x,s) \tag{31}$$

and

$$\lim_{t\to s}\mathbb{E}\left(\frac{(X_t-X_s)\otimes(X_t-X_s)}{t-s}\Big|X_s=x\right)=\mathbf{b}(x,s) \tag{32}$$

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 (31) and (32) 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).

Theorem

(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**

$$-\frac{\partial u}{\partial s} = a(x,s)\frac{\partial u}{\partial x} + \frac{1}{2}b(x,s)\frac{\partial^2 u}{\partial x^2}, \quad \lim_{s \to t} u(s,x) = f(x). \tag{33}$$

Proof:
First we notice that, the continuity assumption (27), 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:

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

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 (17) to obtain

$$\begin{aligned} u(x,\sigma) &= \int_{\mathbb{R}} f(z)P(dz,t|x,\sigma) && (34)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(z)P(dz,t|y,\rho)P(dy,\rho|x,\sigma) \\ &= \int_{\mathbb{R}} u(y,\rho)P(dy,\rho|x,\sigma). && (35) \end{aligned}$$

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

$$u(z,\rho)-u(x,\rho) = \frac{\partial u(x,\rho)}{\partial x}(z-x)+\frac{1}{2}\frac{\partial^2 u(x,\rho)}{\partial x^2}(z-x)^2(1+\alpha_\varepsilon), \quad |z-x| \leqslant \varepsilon \tag{36}$$

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 (35) with (36) to calculate

$$\frac{u(x,s)-u(x,s+h)}{h} =$$

$$\frac{1}{h}\left(\int_{\mathbb{R}} P(dy,s+h|x,s)u(y,s+h)-u(x,s+h)\right)$$

$$= \frac{1}{h}\int_{\mathbb{R}} P(dy,s+h|x,s)(u(y,s+h)-u(x,s+h))$$

$$= \frac{1}{h}\int_{|x-y|<\varepsilon} P(dy,s+h|x,s)(u(y,s+h)-u(x,s))+o(1)$$

$$= \frac{\partial u}{\partial x}(x,s+h)\frac{1}{h}\int_{|x-y|<\varepsilon}(y-x)P(dy,s+h|x,s)$$

$$+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(x,s+h)\frac{1}{h}\int_{|x-y|<\varepsilon}(y-x)^2P(dy,s+h|x,s)(1+\alpha_\varepsilon)+o(1)$$

$$= a(x,s)\frac{\partial u}{\partial x}(x,s+h)+\frac{1}{2}b(x,s)\frac{\partial^2 u}{\partial x^2}(x,s+h)(1+\alpha_\varepsilon)+o(1).$$

Equation (33) 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

$$\int_{\mathbb{R}} f(y)\left(\frac{\partial p(y,t|x,s)}{\partial s} + \mathcal{A}_{s,x}p(y,t|x,s)\right) = 0 \tag{37}$$

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

$$-\frac{\partial p(y,t|x,s)}{\partial s} = a(x,s)\frac{\partial p(y,t|x,s)}{\partial x} + \frac{1}{2}b(x,s)\frac{\partial^2 p(y,t|x,s)}{\partial x^2}. \tag{38}$$

Notice that the variation is with respect to the "backward" variables x, s. We can also obtain an equation with respect to the "forward" variables y, t, the Forward Kolmogorov equation .

Now we derive the forward Kolmogorov equation or **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

(Kolmogorov) Assume that conditions (27), (28), (29) *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*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}\left(a(t, y)p\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(t, y)p\right), \quad \lim_{t \to s} p(t, y|x, s) = \delta(x - y). \tag{39}$$

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

$$\lim_{h \to 0} \frac{1}{h} \left(\int f(y) p(y, s+h|x, s)\, ds - f(x) \right) = a(x,s) f_x(x) + \frac{1}{2} b(x,s) f_{xx}(x), \tag{40}$$

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

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

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. □

Assume now that initial distribution of X_t is $\rho_0(x)$ and set $s = 0$ (the initial time) in (39). Define

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

We multiply the forward Kolmogorov equation (39) by $\rho_0(x)$ and integrate with respect to x to obtain the equation

$$\frac{\partial p(y,t)}{\partial t} = -\frac{\partial}{\partial y}\left(a(y,t)p(y,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(y,t)p(t,y)\right), \tag{41}$$

together with the initial condition

$$p(y,0) = \rho_0(y). \tag{42}$$

The solution of equation (41), 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 (39) as the Green's function for the PDE (41).

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:

$$\mathbb{E}(f(X_t, X_s)) = \int \int f(y, x) p(y, t | x, s) p(x, s)\, dxdy. \tag{43}$$

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,

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

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

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

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

$$\lim_{t \to s} \mathbb{E}\left(\frac{X_t - X_s}{t - s} \Big| X_s = x \right) = \mathbf{a}(x, s) \tag{44}$$

and

$$\lim_{t \to s} \mathbb{E}\left(\frac{(X_t - X_s) \otimes (X_t - X_s)}{t - s} \Big| X_s = x \right) = \mathbf{b}(x, s) \tag{45}$$

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

THE FOKKER-PLANCK EQUATION

- Consider a diffusion process on $\mathbb{R}^d$ with time-independent drift and diffusion coefficients. The Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = -\sum_{j=1}^{d} \frac{\partial}{\partial x_j}(a_i(x)p) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(b_{ij}(x)p), \; t > 0, \, x \in \mathbb{R}^d, \tag{46a}$$

$$p(x,0) = f(x), \quad x \in \mathbb{R}^d. \tag{46b}$$

- Write it in **non-divergence form**:

$$\frac{\partial p}{\partial t} = \sum_{j=1}^{d} \tilde{a}_j(x)\frac{\partial p}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d} \tilde{b}_{ij(x)} \frac{\partial^2 p}{\partial x_i \partial x_j} + \tilde{c}(x)u, \; t > 0, \, x \in \mathbb{R}^d, \tag{47a}$$

$$p(x,0) = f(x), \quad x \in \mathbb{R}^d, \tag{47b}$$

- where

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

- The diffusion matrix is always nonnegative. We will assume that it is actually positive, i.e. we will impose the **uniform ellipticity** condition:

$$\sum_{i,j=1}^{d} b_{ij}(x)\xi_i\xi_j \geqslant \alpha\|\xi\|^2, \quad \forall\, \xi \in \mathbb{R}^d, \tag{48}$$

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

$$\|b(x)\| \leqslant M, \ \|\tilde{a}(x)\| \leqslant M(1+\|x\|), \ \|\tilde{c}(x)\| \leqslant M(1+\|x\|^2). \tag{49}$$

- We will call a solution to the Cauchy problem for the Fokker–Planck equation (47) a **classical solution** if:
  1. $u \in C^{2,1}(\mathbb{R}^d, \mathbb{R}^+)$.
  2. $\forall T > 0$ there exists a $c > 0$ such that

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

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

Theorem

Assume that conditions (48) *and* (49) *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, δ *so that*

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

- This estimate enables us to multiply the Fokker-Planck equation by monomials x^n and then to integrate over $\mathbb{R}^d$ and to integrate by parts.
- For a proof of this theorem see Friedman, Partial Differential Equations of Parabolic Type, Prentice-Hall 1964.

The FP equation as a conservation law

- We can define the **probability current** to be the vector whose ith component is

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

- The Fokker–Planck equation can be written 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:

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

Boundary conditions for the Fokker–Planck equation

- So far we have been studying the FP equation on $\mathbb{R}^d$.
- 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.
- We can also study the FP equation in a bounded domain with appropriate boundary conditions.
- We can have **absorbing, reflecting** or **periodic boundary** conditions.

- Consider the FP equation posed in $\Omega \subset \mathbb{R}^d$ where Ω is a bounded domain with smooth boundary.
- Let **J** denote the probability current and let **n** be the unit normal vector to the surface.
  1. We specify reflecting boundary conditions by setting
  $$\mathbf{n} \cdot \mathbf{J}(x,t) = 0, \quad \text{on } \partial\Omega.$$
  2. We specify absorbing boundary conditions by setting
  $$p(x,t) = 0, \quad \text{on } \partial\Omega.$$
  3. When the coefficient of the FP equation are periodic functions, we might also want to consider periodic boundary conditions, i.e. the solution of the FP equation is periodic in x with period equal to that of the coefficients.

- Reflecting BC correspond to the case where a particle which evolves according to the SDE corresponding to the FP equation gets reflected at the boundary.
- Absorbing BC correspond to the case where a particle which evolves according to the SDE corresponding to the FP equation gets absorbed at the boundary.
- There is a complete classification of boundary conditions in one dimension, the **Feller classification**: the BC can be **regular, exit, entrance** and **natural**.

Examples of Diffusion Processes

- Set $a(t,x) \equiv 0$, $b(t,x) \equiv 2D > 0$. The Fokker-Planck equation becomes:

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

- This is the heat equation, which is the Fokker-Planck equation for Brownian motion (Einstein, 1905). Its solution is

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

- Assume that the initial distribution is

$$p_W(x,s|y,s) = W(y,s).$$

- The solution of the Fokker-Planck equation for Brownian motion with this initial distribution is

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

- The Gaussian distribution is the **fundamental solution** (Green's function) of the heat equation (i.e. the Fokker-Planck equation for Brownian motion).

- Set $a(t,x) = -\alpha x$, $b(t,x) \equiv 2D > 0$:

$$\frac{\partial p}{\partial t} = \alpha \frac{\partial (xp)}{\partial x} + D \frac{\partial^2 p}{\partial x^2}.$$

- This is the Fokker-Planck equation for the Ornstein-Uhlenbeck process (Ornstein-Uhlenbeck, 1930). Its solution is

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

- Proof: take the Fourier transform in x, use the method of characteristics and take the inverse Fourier transform.

- Set $y = 0$, $s = 0$. Notice that

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

- Thus, in the limit where the friction coefficient goes to 0, we recover distribution function of BM from the DF of the OU processes.
- Notice also that

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

- Thus, the Ornstein-Uhlenbeck process is an ergodic Markov process. Its invariant measure is Gaussian.
- We can calculate all moments of the OU process.

- Define the nth moment of the OU process:

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

- Let $n = 0$. We integrate the FP equation over $\mathbb{R}$ to obtain:

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

- 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:

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

- Consequently: **probability is conserved**.

- 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:

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

- 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:

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

- The stationary moments of the OU process are:

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

- We have that

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

- If the initial conditions of the OU process are stationary, then:

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

- set $a(x) = \mu x, \;\; b(x) = \frac{1}{2}\sigma x^2$. This is the **geometric Brownian motion**. The generator of this process is

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

- 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:

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

- 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:

$$\lim_{t\to\infty} M_n(t) = \begin{cases} 0, & \frac{\sigma^2}{2} < -\frac{\mu}{n-1}, \\ M_n(0), & \frac{\sigma^2}{2} = -\frac{\mu}{n-1}, \\ +\infty, & \frac{\sigma^2}{2} > -\frac{\mu}{n-1}. \end{cases}$$

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:
$$\mathcal{L} = -\nabla V(x) \cdot \nabla + D\Delta. \tag{50}$$
- 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:

$$\frac{\partial p}{\partial t} = \nabla \cdot (\nabla V p) + D \Delta p. \tag{51}$$

- 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.

Theorem

Assume that $V(x)$ is smooth and that

$$e^{-V(x)/D} \in L^1(\mathbb{R}^d). \tag{52}$$

Then the Markov process with generator (50) *is ergodic. The unique invariant distribution is the* **Gibbs distribution**

$$p(x) = \frac{1}{Z} e^{-V(x)/D}, \tag{53}$$

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

Let $p(x,t)$ be the solution of the Fokker-Planck equation (51)*, assume that* (52) *holds and let $\rho(x)$ be the Gibbs distribution* (187)*. Define $h(x,t)$ through*

$$p(x,t) = h(x,t)\rho(x).$$

Then the function h satisfies the **backward Kolmogorov equation***:*

$$\frac{\partial h}{\partial t} = -\nabla V \cdot \nabla h + D\Delta h, \quad h(x,0) = p(x,0)\rho^{-1}(x). \tag{54}$$

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 (54).
- 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

Assume that $V(x)$ is a smooth potential and assume that condition (52) *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 $f, \in C_0^2(\mathbb{R}^d)$. 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

$$(\mathcal{L}f, f)_\rho = -D\|\nabla f\|_\rho^2,$$

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

$$0 = -D\|\nabla f\|_\rho^2,$$

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

$$(-\mathcal{L}f, f)_\rho = D\|\nabla f\|_\rho^2. \tag{55}$$

- 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

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:

$$\int_{\mathbb{R}^d} f\rho = 0 \quad \Rightarrow \quad \|\nabla f\|_\rho \geqslant \sqrt{\lambda}\|f\|_\rho. \tag{56}$$

Theorem

Assume that $p(x,0) \in L^2(e^{V/D})$. *Then the solution* $p(x,t)$ *of the Fokker-Planck equation* (51) *converges to the Gibbs distribution exponentially fast:*

$$\|p(\cdot,t) - Z^{-1}e^{-V}\|_{\rho^{-1}} \leqslant e^{-\lambda D t}\|p(\cdot,0) - Z^{-1}e^{-V}\|_{\rho^{-1}}. \qquad (57)$$

Proof.

We Use (54), (55) and (56) 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

$$\|h(\cdot,t)-1\|_\rho \leqslant e^{-\lambda D t}\|H(\cdot,0)-1\|_\rho.$$

This, and the definition of h, $p = \rho h$, lead to (57). □

- 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

Let p denote the solution of the Fokker–Planck equation (51) *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.

Eigenfunction Expansions

- Consider the generator of a gradient flow with a uniformly convex potential
$$\mathcal{L} = -\nabla V \cdot \nabla + D\Delta. \tag{58}$$
- We know that
  1. $\mathcal{L}$ is a non-positive self-adjoint operator on L^2_ρ.
  2. 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 $-\mathcal{L}$:
$$-\mathcal{L}f_n = \lambda_n f_n, \quad n = 0, 1, \ldots$$

- 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:

$$\phi = \phi_0 + \sum_{n=1}^{\infty} \phi_n f_n, \quad \phi_n = (\phi, f_n)_\rho \tag{59}$$

- with $(f_n, f_m)_\rho = \delta_{nm}$ and $\phi_0 \in \mathcal{N}(\mathcal{L})$.
- This enables us to solve the time dependent Fokker–Planck equation in terms of an eigenfunction expansion.
- Consider the backward Kolmogorov equation (54).
- We assume that the initial conditions $h_0(x) = \phi(x) \in L^2_\rho$ and consequently we can expand it in the form (59).

- We look for a solution of (54) 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:

$$\frac{\partial h}{\partial t} = \sum_{n=0}^{\infty} \dot{h}_n f_n = \mathcal{L}\left(\sum_{n=0}^{\infty} h_n f_n\right) \tag{60}$$

$$= \sum_{n=0}^{\infty} -\lambda_n h_n f_n. \tag{61}$$

- 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

$$\dot{h}_n = -\lambda_n h_n, \quad n = 0, 1, \ldots$$

- 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

$$p(x,t) = Z^{-1} e^{-\beta V(x)} \left(1 + \sum_{n=1}^{\infty} e^{-\lambda_n t} \phi_n f_n \right).$$

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, $\approx$ 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 statinary).
- 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.

Reduction to a Schrödinger Equation

Lemma

The Fokker–Planck operator for a gradient flow can be written in the self-adjoint form

$$\frac{\partial p}{\partial t} = D\nabla \cdot \left(e^{-V/D} \nabla \left(e^{V/D} p \right) \right). \tag{62}$$

Define now $\psi(x,t) = e^{V/2D} p(x,t)$. *Then* ψ *solves the PDE*

$$\frac{\partial \psi}{\partial t} = D\Delta\psi - U(x)\psi, \quad U(x) := \frac{|\nabla V|^2}{4D} - \frac{\Delta V}{2}. \tag{63}$$

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

1. From equation (62) shows that the FP operator can be written in the form
$$\mathcal{L}^* \cdot = D\nabla \cdot \left(e^{-V/D} \nabla \left(e^{V/D} \cdot \right) \right).$$
2. The operator that appears on the right hand side of eqn. (63) has the form of a **Schrödinger operator**:
$$-\mathcal{H} = -D\Delta + U(x).$$
3. 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.
4. In particular, the weak noise asymptotics $D \ll 1$ is equivalent to the semiclassical approximation from quantum mechanics.

Proof.

We calculate

$$\begin{aligned} D\nabla\cdot\left(e^{-V/D}\nabla\left(e^{V/D}f\right)\right) &= D\nabla\cdot\left(e^{-V/D}\left(D^{-1}\nabla Vf+\nabla f\right)e^{V/D}\right)\\ &= \nabla\cdot\left(\nabla Vf+D\nabla f\right)=\mathcal{L}^*f. \end{aligned}$$

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

1. We can rewrite the Schrödinger operator in the form

$$\mathcal{H} = D\mathcal{A}^*\mathcal{A}, \quad \mathcal{A} = \nabla + \frac{\nabla U}{2D}, \quad \mathcal{A}^* = -\nabla + \frac{\nabla U}{2D}.$$

2. These are **creation** and **annihilation** operators. They can also be written in the form

$$\mathcal{A}\cdot = e^{-U/2D}\nabla\left(e^{U/2D}\cdot\right), \quad \mathcal{A}^*\cdot = e^{U/2D}\nabla\left(e^{-U/2D}\cdot\right)$$

3. 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.

The OU Process and Hermite Polynomials

- The generator of the OU process is

$$\mathcal{L} = -y\frac{d}{dy} + D\frac{d^2}{dy^2} \tag{64}$$

- The OU process is an ergodic Markov process whose unique invariant measure is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ with density $\rho(y) \in C^\infty(\mathbb{R})$

$$\rho(y) = \frac{1}{\sqrt{2\pi D}} e^{-\frac{y^2}{2D}}.$$

- Let L^2_ρ denote the closure of C^∞ functions with respect to the norm

$$\|f\|^2_\rho = \int_{\mathbb{R}} f(y)^2 \rho(y)\, dy.$$

- The space L^2_ρ is a Hilbert space with inner product

$$(f,g)_\rho = \int_{\mathbb{R}} f(y) g(y) \rho(y)\, dy.$$

- Consider the eigenvalue problem for $\mathcal{L}$:

$$-\mathcal{L}f_n = \lambda_n f_n.$$

- The operator $\mathcal{L}$ has discrete spectrum in L^2_ρ:

$$\lambda_n = n, \quad n \in \mathcal{N}.$$

- The corresponding eigenfunctions are the normalized Hermite polynomials:

$$f_n(y) = \frac{1}{\sqrt{n!}} H_n\left(\frac{y}{\sqrt{D}}\right), \tag{65}$$

where

$$H_n(y) = (-1)^n e^{\frac{y^2}{2}} \frac{d^n}{dy^n}\left(e^{-\frac{y^2}{2}}\right).$$

- The first few Hermite polynomials are:

$$\begin{aligned} H_0(y) &= 1, \\ H_1(y) &= y, \\ H_2(y) &= y^2 - 1, \\ H_3(y) &= y^3 - 3y, \\ H_4(y) &= y^4 - 3y^2 + 3, \\ H_5(y) &= y^5 - 10y^3 + 15y. \end{aligned}$$

- H_n is a polynomial of degree n.
- Only odd (even) powers appear in H_n when n is odd (even).

Lemma

The eigenfunctions $\{f_n(y)\}_{n=1}^{\infty}$ of the generator of the OU process $\mathcal{L}$ satisfy the following properties.

1. *They form an orthonormal set in L^2_ρ:*

$$(f_n, f_m)_\rho = \delta_{nm}.$$

2. *$\{f_n(y) : n \in \mathcal{N}\}$ is an orthonormal basis in L^2_ρ.*
3. *Define the creation and annihilation operators on $C^1(\mathbb{R})$ by*

$$A_+\phi(y) = \frac{1}{\sqrt{D(n+1)}}\left(-D\frac{d\phi}{dy}(y) + y\phi(y)\right), \quad y \in \mathbb{R}$$

and

$$A_-\phi(y) = \sqrt{\frac{D}{n}}\frac{d\phi}{dy}(y).$$

Then

$$A_+f_n = f_{n+1} \quad \text{and} \quad A_-f_n = f_{n-1}. \tag{66}$$

4. The eigenfunctions f_n satisfy the following recurrence relation

$$yf_n(y) = \sqrt{Dn}f_{n-1}(y) + \sqrt{D(n+1)}f_{n+1}(y), \quad n \geqslant 1. \tag{67}$$

5. The function

$$H(y;\lambda) = e^{\lambda y - \frac{\lambda^2}{2}}$$

is a generating function for the Hermite polynomials. In particular

$$H\left(\frac{y}{\sqrt{D}};\lambda\right) = \sum_{n=0}^{\infty} \frac{\lambda^n}{\sqrt{n!}} f_n(y), \quad \lambda \in \mathbb{C},\, y \in \mathbb{R}.$$

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

$$\mathcal{L} = p \cdot \nabla_q - \nabla_q V \nabla_p + \gamma(-p\nabla_p + D\Delta_p). \tag{68}$$

- 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**

$$\ddot{X}_t = -\nabla V(X_t) - \gamma \dot{X}_t + \sqrt{2\gamma D}\dot{W}_t. \tag{69}$$

- 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

Assume that $V(x)$ is smooth and that

$$e^{-V(x)/D} \in L^1(\mathbb{R}^d).$$

Then the Markov process with generator (140) *is ergodic. The unique invariant distribution is the* **Maxwell-Boltzmann distribution**

$$\rho_\beta(p,q) = \frac{1}{Z} e^{-\beta H(p,q)} \tag{70}$$

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 quite 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.
- See also C. Villani, *Hypocoercivity*, AMS 2008.

Consider a second order elliptic operator of the form

$$\mathcal{L} = X_0 + \sum_{i=1}^{d} X_i^* X_i,$$

where $\{X_j\}_{j=0}^{d}$ are vector fields (first order differential operators).
Define the Lie algebras

$$\begin{aligned}
\mathcal{A}_0 &= \text{Lie}(X_1, X_1, \dots X_d), \\
\mathcal{A}_1 &= \text{Lie}([X_0, X_1], [X_0, X_2], \dots [X_0, X_d]), \\
\dots &= \dots, \\
\mathcal{A}_k &= \text{Lie}([X_0, X]; X \in \mathcal{A}_{k-1}), \quad k \geqslant 1
\end{aligned}$$

Define

$$\mathcal{H} = \text{Lie}(\mathcal{A}_0, \mathcal{A}_1, \dots)$$

Hörmander's Hypothesis. $\mathcal{H}$ is full in the sense that, for every $x \in \mathbb{R}^d$ the vectors $\{H(x) : H \in \mathcal{H}\}$ span $\mathbb{T}_x\mathbb{R}^d$:

$$\text{span}\{H(x); H \in \mathcal{H}\} = \mathbb{T}_x\mathbb{R}^d.$$

Theorem (Hörmander/Kolmogorov.)

Under the above hypothesis, the diffusion process X_t with generator $\mathcal{L}$ has a transition density

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

where $p(\cdot,\cdot,\cdot)$ is a smooth function on $(0,+\infty)\times\mathbb{R}^d\times\mathbb{R}^d$. Moreover, the function p satisfies Kolmogorov's forward and backward equations

$$\frac{\partial}{\partial t} p(\cdot,x,\cdot) = \mathcal{L}_y^* p(\cdot,x,\cdot), \quad x\in\mathbb{R}^d \tag{71}$$

and

$$\frac{\partial}{\partial t} p(\cdot,\cdot,y) = \mathcal{L}_x p(\cdot,\cdot,y), \quad y\in\mathbb{R}^d. \tag{72}$$

For every $x\in\mathbb{R}^d$ the function $p(\cdot,x,\cdot)$ is the fundamental solution (Green's function) of (71), such that, for each $f\in C_0(\mathbb{R}^d)$, $\lim_{t\to 0}\int_{\mathbb{R}^d} p(t,x,y) f(y)\, dy = f(x)$.

Example

Consider the SDE

$$\ddot{x} = \sqrt{2}\dot{W}.$$

Write it as a first order system

$$\dot{x} = y, \quad \dot{y} = \sqrt{2}\dot{W}.$$

The generator of this process is

$$\mathcal{L} = X_0 + X_1^2, \quad X_0 = y\partial_x, \;\; X_1 = \partial_y.$$

Consequently $[X_0, X_1] = -\partial_x$ and

$$\text{Lie}(X_1, [X_0, X_1]) = \text{Lie}(-\partial_x, \partial_y),$$

which spans $\mathbb{T}_x\mathbb{R}^2$.

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

$$\frac{\partial h}{\partial t} = -\mathcal{A}h + \gamma \mathcal{S}h$$

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:

$$\mathcal{L} = \mathcal{A} + \gamma\beta^{-1} \sum_{i=1}^{d} X_i^* X_i. \tag{73}$$

We calculate the commutators between the vector fields in (73):

$$[\mathcal{A}, X_i] = \frac{\partial}{\partial q_i}, \quad [X_i, X_j] = 0, \quad [X_i, X_j^*] = \beta\delta_{ij}.$$

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

$$\mathcal{L} = \beta^{-1} \sum_{i=1}^{d} \left(X_i^* Y_i - Y_i^* X_i + \gamma X_i^* X_i \right).$$

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 (183) 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

$$\frac{\partial h}{\partial t} = \mathcal{L}h. \tag{74}$$

- 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 (74) into the basis of Hermite basis:

$$h(p,q,t) = \sum_{n=0}^{\infty} h_n(q,t) f_n(p), \tag{75}$$

- where $f_n(p) = 1/\sqrt{n!} H_n(p)$.

- Our plan is to substitute (75) into (74) 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).

- There are very few potentials for which we can calculate the eigenvalues and eigenfunctions of the generator of the Markov process $\{q(t), p(t)\}$.
- We can calculate everything explicitly for the quadratic (harmonic) potential

$$V(q) = \frac{1}{2}\omega_0^2 q^2. \tag{76}$$

- The Langevin equation is

$$\ddot{q} = -\omega_0^2 q - \gamma\dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W} \tag{77}$$

or

$$\dot{q} = p, \quad \dot{p} = -\omega_0^2 q - \gamma p + \sqrt{2\gamma\beta^{-1}}\dot{W}. \tag{78}$$

- This is a linear equation that can be solved explicitly (The solution is a Gaussian stochastic process).

- The generator of the process $\{q(t), p(t)\}$

$$\mathcal{L} = p\partial_q - \omega_0^2 q\partial_p + \gamma(-p\partial_p + \beta^{-1}\partial_p^2). \tag{79}$$

The Fokker-Planck operator is

$$\mathcal{L} = p\partial_q - \omega_0^2 q\partial_p + \gamma(-p\partial_p + \beta^{-1}\partial_p^2). \tag{80}$$

The process $\{q(t), p(t)\}$ is an ergodic Markov process with Gaussian invariant measure

$$\rho_\beta(q,p)\,dqdp = \frac{\beta\omega_0}{2\pi} e^{-\frac{\beta}{2}p^2 - \frac{\beta\omega_0^2}{2}q^2}. \tag{81}$$

- 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:

$$a^- = \beta^{-1/2}\partial_p, \quad a^+ = -\beta^{-1/2}\partial_p + \beta^{1/2}p \tag{82}$$

- and

$$b^- = \omega_0^{-1}\beta^{-1/2}\partial_q, \quad b^+ = -\omega_0^{-1}\beta^{-1/2}\partial_q + \omega_0\beta^{1/2}p. \tag{83}$$

- 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

$$\widehat{\mathcal{L}} = -a^+a^- - b^+b^- \tag{84}$$

- is the generator of the OU process in two dimensions.

Exercises

1. Calculate the eigenvalues and eigenfunctions of $\widehat{\mathcal{L}}$. Show that there exists a transformation that transforms $\widehat{\mathcal{L}}$ into the Schrödinger operator of the two-dimensional quantum harmonic oscillator.
2. Show that the operators $a^{\pm}$, $b^{\pm}$ satisfy the commutation relations

$$[a^{+}, a^{-}] = -1, \tag{85a}$$

$$[b^{+}, b^{-}] = -1, \tag{85b}$$

$$[a^{\pm}, b^{\pm}] = 0. \tag{85c}$$

- Using the operators $a^{\pm}$ and $b^{\pm}$ we can write the generator $\mathcal{L}$ in the form

$$\mathcal{L} = -\gamma a^{+}a^{-} - \omega_0(b^{+}a^{-} - a^{+}b^{-}), \tag{86}$$

- We want to find first order differential operators $c^{\pm}$ and $d^{\pm}$ so that the operator (86) becomes

$$\mathcal{L} = -Cc^{+}c^{-} - Dd^{+}d^{-} \tag{87}$$

- for some appropriate constants C and D.
- 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**

$$[c^{+}, c^{-}] = -1, \tag{88a}$$

$$[d^{+}, d^{-}] = -1, \tag{88b}$$

$$[c^{\pm}, d^{\pm}] = 0. \tag{88c}$$

- The operators $c^{\pm}$ and $d^{\pm}$ should be given as linear combinations of the old operators $a^{\pm}$ and $b^{\pm}$. They should be of the form

$$c^{+} = \alpha_{11} a^{+} + \alpha_{12} b^{+}, \tag{89a}$$

$$c^{-} = \alpha_{21} a^{-} + \alpha_{22} b^{-}, \tag{89b}$$

$$d^{+} = \beta_{11} a^{+} + \beta_{12} b^{+}, \tag{89c}$$

$$d^{-} = \beta_{21} a^{-} + \beta_{22} b^{-}. \tag{89d}$$

- Notice that the $c-$ and d^{-} **are not** the adjoints of c^{+} and d^{+}.
- If we substitute now these equations into (87) and equate it with (86) and into the commutation relations (88) 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

$$\lambda_{1,2} = \frac{\gamma \pm \delta}{2}, \quad \delta = \sqrt{\gamma^2 - 4\omega_0^2}. \tag{90}$$

- The eigenvalues satisfy the relations

$$\lambda_1 + \lambda_2 = \gamma, \quad \lambda_1 - \lambda_2 = \delta, \lambda_1\lambda_2 = \omega_0^2. \tag{91}$$

Lemma

Let $\mathcal{L}$ be the generator (86) *and let $c^{\pm}$, $d^{\pm}$ be the operators*

$$c^{+} = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^{+} + \sqrt{\lambda_2}b^{+}\right), \quad c^{-} = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^{-} - \sqrt{\lambda_2}b^{-}\right), \tag{92a}$$

$$d^{+} = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_2}a^{+} + \sqrt{\lambda_1}b^{+}\right), \quad d^{-} = \frac{1}{\sqrt{\delta}}\left(-\sqrt{\lambda_2}a^{-} + \sqrt{\lambda_1}b^{-}\right). \tag{92b}$$

Then $c^{\pm}$, $d^{\pm}$ satisfy the canonical commutation relations (88) *as well as*

$$[\mathcal{L}, c^{\pm}] = -\lambda_1 c^{\pm}, \quad [\mathcal{L}, d^{\pm}] = -\lambda_2 d^{\pm}. \tag{93}$$

Furthermore, the operator $\mathcal{L}$ can be written in the form

$$\mathcal{L} = -\lambda_1 c^{+}c^{-} - \lambda_2 d^{+}d^{-}. \tag{94}$$

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 (93). Now we calculate

$$\begin{aligned}
\mathcal{L} &= -\lambda_1 c^+ c^- - \lambda_2 d^+ d^- \\
&= -\frac{\lambda_2^2 - \lambda_1^2}{\delta} a^+ a^- + 0 b^+ b^- + \frac{\sqrt{\lambda_1 \lambda_2}}{\delta} (\lambda_1 - \lambda_2) a^+ b^- \\
&\quad + \frac{1}{\delta} \sqrt{\lambda_1 \lambda_2} (-\lambda_1 + \lambda_2) b^+ a^- \\
&= -\gamma a^+ a^- - \omega_0 (b^+ a^- - a^+ b^-),
\end{aligned}$$

which is precisely (86). In the above calculation we used (91). □

Theorem

The eigenvalues and eigenfunctions of the generator of the Markov process $\{q, p\}$ (78) *are*

$$\lambda_{nm} = \lambda_1 n + \lambda_2 m = \frac{1}{2}\gamma(n+m) + \frac{1}{2}\delta(n-m), \quad n, m = 0, 1, \ldots \tag{95}$$

and

$$\phi_{nm}(q,p) = \frac{1}{\sqrt{n!m!}}(c^+)^n (d^+)^m 1, \quad n, m = 0, 1, \ldots \tag{96}$$

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 (exercise)

$$[\mathcal{L},(c^+)^n] = -n\lambda_1(c^+)^n \quad \text{and} \quad [\mathcal{L},(d^+)^m] = -m\lambda_1(d^+)^m. \tag{97}$$

We use (97) to calculate

$$\begin{aligned}\mathcal{L}(c^+)^n(d^+)^n\mathbf{1} &= (c^+)^n\mathcal{L}(d^+)^m\mathbf{1} - n\lambda_1(c^+)^n(d^+m)\mathbf{1}\\ &= (c^+)^n(d^+)^m\mathcal{L}\mathbf{1} - m\lambda_2(c^+)^n(d^+m)\mathbf{1} -\\ &\quad n\lambda_1(c^+)^n(d^+m)\mathbf{1}\\ &= -n\lambda_1(c^+)^n(d^+m)\mathbf{1} - m\lambda_2(c^+)^n(d^+m)\mathbf{1}\end{aligned}$$

from which (95) and (96) follow.

Exercise

1. Show that

$$[\mathcal{L}, (c^{\pm})^n] = -n\lambda_1 (c^{\pm})^n, \quad [\mathcal{L}, (d^{\pm})^n] = -n\lambda_1 (d^{\pm})^n, \quad (98)$$

$$[c^-, (c^+)^n] = n(c^+)^{n-1}, \quad [d^-, (d^+)^n] = n(d^+)^{n-1}. \quad (99)$$

Remark In terms of the operators $a^{\pm}$, $b^{\pm}$ the eigenfunctions of $\mathcal{L}$ are

$$\phi_{nm} = \sqrt{n!m!}\delta^{-\frac{n+m}{2}} \lambda_1^{n/2} \lambda_2^{m/2} \sum_{\ell=0}^{n} \sum_{k=0}^{m} \frac{1}{k!(m-k)!\ell!(n-\ell)!} \times$$

$$\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{k-\ell}{2}} (a^+)^{n+m-k-\ell} (b^+)^{\ell+k} 1.$$

The first few eigenfunctions are

$$\phi_{00} = 1.$$

$$\phi_{10} = \frac{\sqrt{\beta}\left(\sqrt{\lambda_1}p + \sqrt{\lambda_2}\omega_0 q\right)}{\sqrt{\delta}}.$$

$$\phi_{01} = \frac{\sqrt{\beta}\left(\sqrt{\lambda_2}p + \sqrt{\lambda_1}\omega_0 q\right)}{\sqrt{\delta}}$$

$$\phi_{11} = \frac{-2\sqrt{\lambda_1}\sqrt{\lambda_2} + \sqrt{\lambda_1}\beta p^2\sqrt{\lambda_2} + \beta p\lambda_1\omega_0 q + \omega_0\beta q\lambda_2 p + \sqrt{\lambda_2}{\omega_0}^2\beta q^2\sqrt{\lambda_1}}{\delta}.$$

$$\phi_{20} = \frac{-\lambda_1 + \beta p^2\lambda_1 + 2\sqrt{\lambda_2}\beta p\sqrt{\lambda_1}\omega_0 q - \lambda_2 + {\omega_0}^2\beta q^2\lambda_2}{\sqrt{2}\delta}.$$

$$\phi_{02} = \frac{-\lambda_2 + \beta p^2\lambda_2 + 2\sqrt{\lambda_2}\beta p\sqrt{\lambda_1}\omega_0 q - \lambda_1 + {\omega_0}^2\beta q^2\lambda_1}{\sqrt{2}\delta}.$$

- The eigenfunctions are not orthonormal.
- The first eigenvalue, corresponding to the constant eigenfunction, is 0:

$$\lambda_{00} = 0.$$

- The operator $\mathcal{L}$ is not self-adjoint and consequently, we do not expect its eigenvalues to be real.
- 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

$$\lambda_{nm} = \frac{\gamma}{2}(n+m) + i\omega(n-m).$$

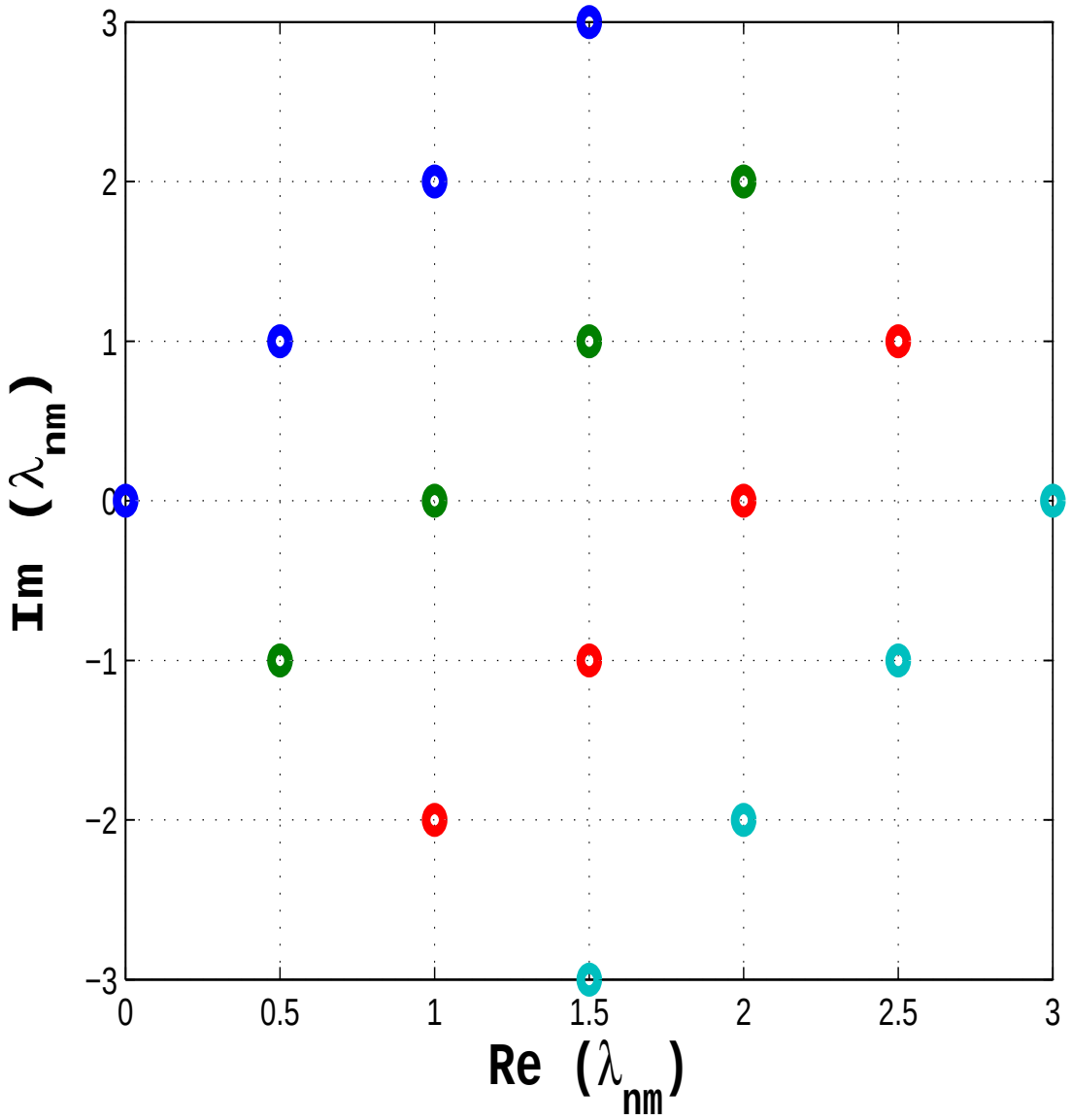

Figure: First few eigenvalues of $\mathcal{L}$ for $\gamma = \omega = 1$.

- In Figure 34 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:

$$\lambda_{nn} = \gamma n.$$

- 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 $\gamma \to +\infty$, the eigenvalues of the generator $\mathcal{L}$ converge to the eigenvalues of the generator of the OU process:

$$\lambda_{nm} = \gamma n + \frac{\omega_0^2}{\gamma}(n-m) + O(\gamma^{-3}).$$

- 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 $\mathcal{L}$.
- 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

$$(c^+)^* = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^- + \sqrt{\lambda_2}b^-\right), \quad (c^-)^* = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_1}a^+ - \sqrt{\lambda_2}b^+\right) \tag{100a}$$

$$(d^+)^* = \frac{1}{\sqrt{\delta}}\left(\sqrt{\lambda_2}a^- + \sqrt{\lambda_1}b^-\right), \quad (d^-)^* = \frac{1}{\sqrt{\delta}}\left(-\sqrt{\lambda_2}a^+ + \sqrt{\lambda_1}b^-\right) \tag{100b}$$

Exercise

1. Show by direct substitution that $\widehat{\mathcal{L}}$ can be written in the form

$$\widehat{\mathcal{L}} = -\lambda_1 (c^-)^* (c^+)^* - \lambda_2 (d^-)^* (d^+)^*.$$

2. Calculate the commutators

$$[(c^+)^*, (c^-)^*], \quad [(d^+)^*, (d^-)^*], \quad [(c^\pm)^*, (d^\pm)^*], \quad [\widehat{\mathcal{L}}, (c^\pm)^*], \quad [\widehat{\mathcal{L}}, (d^\pm)^*]$$

1. $\widehat{\mathcal{L}}$ has the same eigenvalues as $\mathcal{L}$:

$$-\widehat{\mathcal{L}}\psi_{nm} = \lambda_{nm}\psi_{nm},$$

2. where λ_{nm} are given by (95).
3. The eigenfunctions are

$$\psi_{nm} = \frac{1}{\sqrt{n!m!}} ((c^-)^*)^n ((d^-)^*)^m \mathbf{1}. \tag{101}$$

Theorem

The eigenfunctions of $\mathcal{L}$ and $\widehat{\mathcal{L}}$ satisfy the biorthonormality relation

$$\int\int \phi_{nm}\psi_{\ell k}\rho_\beta \, dpdq = \delta_{n\ell}\delta_{mk}. \tag{102}$$

Proof. We will use formulas (98). 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 $n \geqslant \ell$)

$$(c^-)^\ell (c^+)^n \mathbf{1} = n(n-1)\dots(n-\ell+1)(c^+)^{n-\ell}. \tag{103}$$

We have

$$\begin{aligned}
\int\int \phi_{nm}\psi_{\ell k}\rho_\beta \, dpdq &= \frac{1}{\sqrt{n!m!\ell!k!}} \int\int ((c^+))^n ((d^+))^m \mathbf{1} ((c^-)^*)^\ell ((d^-)^*)^k \mathbf{1}\rho_\beta \, dpdq \\
&= \frac{n(n-1)\dots(n-\ell+1)m(m-1)\dots(m-k+1)}{\sqrt{n!m!\ell!k!}} \times \\
&\quad \int\int ((c^+))^{n-\ell} ((d^+))^{m-k} \mathbf{1}\rho_\beta \, dpdq \\
&= \delta_{n\ell}\delta_{mk},
\end{aligned}$$

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

$$\mathcal{L}^*(f\rho_\beta) = \rho\widehat{\mathcal{L}}f$$

- 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

$$\psi^*_{nm} = \rho_\beta\phi_{nm} = \rho_\beta\frac{1}{\sqrt{n!m!}}((c^-)^*)^n((d^-)^*)^m\mathbf{1}. \qquad (104)$$

STOCHASTIC DIFFERENTIAL EQUATIONS (SDEs)

- 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

$$\frac{dz}{dt} = h(z) + \gamma(z)\frac{dW}{dt}, \quad z(0) = z_0. \tag{105}$$

- 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 (105) 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 (105) is as an integral equation for $z(t) \in C(\mathbb{R}^+, \mathcal{Z})$:

$$z(t) = z_0 + \int_0^t h(z(s))ds + \int_0^t \gamma(z(s))dW(s). \tag{106}$$

- In order to make sense of this equation we need to define the stochastic integral against $W(s)$.

The Itô Stochastic Integral

- For the rigorous analysis of stochastic differential equations it is necessary to define stochastic integrals of the form

$$I(t) = \int_0^t f(s)\, dW(s), \tag{107}$$

- 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 (107):

$$I(t) := \lim_{K\to\infty} \sum_{k=1}^{K-1} f(t_{k-1})\left(W(t_k) - W(t_{k-1})\right), \tag{108}$$

- 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 (108).
- 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*

$$\mathbb{E}|I(t)|^2 = \int_0^t \mathbb{E}|f(s)|_F^2 ds, \tag{109}$$

- where $|\cdot|_F$ denotes the Frobenius norm $|A|_F = \sqrt{tr(A^T A)}$.
- 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

- 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

$$\langle I \rangle_t = \int_0^t (f(s) \otimes f(s))\, ds.$$

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 (107), namely

$$I_{strat}(t) := \lim_{K\to\infty} \sum_{k=1}^{K-1} \frac{1}{2}\Big(f(t_{k-1}) + f(t_k)\Big)\left(W(t_k) - W(t_{k-1})\right), \quad (110)$$

- 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 (110).
- 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 (110) 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.

Existence and Uniqueness of solutions for SDEs

Definition

By a solution of (105) we mean a $\mathcal{Z}$-valued stochastic process $\{z(t)\}$ on $t \in [0, T]$ with the properties:

1. $z(t)$ is continuous and $\mathcal{F}_t-$adapted, where the filtration is generated by the Brownian motion $W(t)$;
2. $h(z(t)) \in L^1((0, T)), \gamma(z(t)) \in L^2((0, T))$;
3. equation (105) 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

$$\mathbb{P}(x_1(t) = x_2(t), \quad \forall t \in [0.T]) = 1.$$

- It is well known that existence and uniqueness of solutions for ODEs (i.e. when $\gamma \equiv 0$ in (105)) 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

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 (105) *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

- The Stratonovich analogue of (105) is

$$\frac{dz}{dt} = h(z) + \gamma(z) \circ \frac{dW}{dt}, \quad z(0) = z_0. \tag{111}$$

- By this we mean that $z \in C(\mathbb{R}^+, \mathcal{Z})$ satisfies the integral equation

$$z(t) = z(0) + \int_0^t h(z(s))ds + \int_0^t \gamma(z(s)) \circ dW(s). \tag{112}$$

- By using definitions (108) and (110) it can be shown that z satisfying the Stratonovich SDE (111) also satisfies the Itô SDE

$$\frac{dz}{dt} = h(z) + \frac{1}{2}\nabla \cdot \left(\gamma(z)\gamma(z)^T\right) - \frac{1}{2}\gamma(z)\nabla \cdot \left(\gamma(z)^T\right) + \gamma(z)\frac{dW}{dt}, \tag{113a}$$

$$z(0) = z_0, \tag{113b}$$

- 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 (105), then we get the equation

$$\frac{dz}{ds} = \frac{1}{c}h(z) + \frac{1}{\sqrt{c}}\gamma(z)\frac{dW}{ds}, \quad z(0) = z_0.$$

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

$$\frac{dx}{dt} = h(x) + \frac{1}{\varepsilon} f(x) y, \tag{114a}$$

$$\frac{dy}{dt} = -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2D}{\varepsilon^2}} \frac{dV}{dt}, \tag{114b}$$

- 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 (114a) 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

$$\lim_{\varepsilon\to 0}\frac{y(t)}{\varepsilon} = \sqrt{\frac{2D}{\alpha^2}}\frac{dV}{dt}. \tag{115}$$

- Another way of seeing this is by solving (114b) for y/ε:

$$\frac{y}{\varepsilon} = \sqrt{\frac{2D}{\alpha^2}}\frac{dV}{dt} - \frac{\varepsilon}{\alpha}\frac{dy}{dt}. \tag{116}$$

- If we neglect the $\mathcal{O}(\varepsilon)$ term on the right hand side then we arrive, again, at the heuristic (115).
- Both of these arguments lead us to conjecture the limiting Itô SDE:

$$\frac{dX}{dt} = h(X) + \sqrt{\frac{2D}{\alpha}} f(X)\frac{dV}{dt}. \tag{117}$$

- 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

$$\frac{dX}{dt} = h(X) + \sqrt{\frac{2D}{\alpha}} f(X) \circ \frac{dV}{dt}. \tag{118}$$

- 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

Assume that the initial conditions for $y(t)$ are stationary and that the function f is smooth. Then the solution of eqn (114a) *converges, in the limit as $\varepsilon \to 0$ to the solution of the Stratonovich SDE* (118)*.*

Remarks

1. It is possible to prove pathwise convergence under very mild assumptions.
2. The generator of a Stratonovich SDE has the from

$$\mathcal{L}_{strat} = h(x)\partial_x + \frac{D}{\alpha} f(x)\partial_x \left(f(x)\partial_x\right).$$

3. Consequently, the Fokker-Planck operator of the Stratonovich SDE can be written in divergence form:

$$\mathcal{L}^*_{strat}\cdot = -\partial_x \left(h(x)\cdot\right) + \frac{D}{\alpha}\partial_x \left(f^2(x)\partial_x \cdot\right).$$

4. 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.

5. 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 61 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

$$\rho(y) = \sqrt{\frac{\alpha}{2\pi D}}e^{-\frac{\alpha y^2}{2D}}. \tag{119}$$

The backward Kolmogorov equation is

$$\frac{\partial u^\varepsilon}{\partial t} = \left(\frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2\right)u^\varepsilon. \tag{120}$$

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 (120) 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:

$$u_1(x,y,t) = \frac{1}{\alpha} f(x)\partial_x uy + \psi_1(x,t).$$

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

$$dX = \quad \sqrt{2D} \quad dV$$

- A Stratonovich SDE

$$dX(t) = f(X(t))\,dt + \sigma(X(t)) \circ dW(t) \tag{121}$$

- 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

$$dX(t) = f(X(t))\,dt + \sigma(X(t))dW(t) \tag{122}$$

- 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.

- Given the function $\gamma(z)$ in the SDE (105) we define

$$\Gamma(z) = \gamma(z)\gamma(z)^T. \tag{123}$$

- The **generator** $\mathcal{L}$ is then defined as

$$\mathcal{L}v = h \cdot \nabla v + \frac{1}{2}\Gamma : \nabla\nabla v. \tag{124}$$

- This operator, equipped with a suitable domain of definition, is the generator of the Markov process given by (105).
- The formal L^2-adjoint operator $\mathcal{L}^*$

$$\mathcal{L}^*v = -\nabla \cdot (hv) + \frac{1}{2}\nabla \cdot \nabla \cdot (\Gamma v).$$

- 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

(Itô's Formula) *Assume that the conditions of Theorem 60 hold. Let* $x(t)$ *solve* (105) *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

$$v(z,t) = \mathbb{E}\big(\phi(z(t))|z(0) = z\big), \tag{125}$$

- 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 (121) and let $V(x) \in C^2(\mathbb{R})$. Then

$$dV(X(t)) = \frac{dV}{dx}(X(t))\left(f(X(t))\,dt + \sigma(X(t)) \circ dW(t)\right).$$

  - Consider the Stratonovich SDE (121) 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:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (f\rho) + \frac{1}{2}\nabla \cdot (\sigma \nabla \cdot (\sigma \rho))). \tag{126}$$

1. The SDE for Brownian motion is:

$$dX = \sqrt{2\sigma}dW, \quad X(0) = x.$$

- The Solution is:

$$X(t) = x + \sqrt{2\sigma}W(t).$$

2. 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:

$$\begin{aligned} dX(t)^n &= \mathcal{L}X(t)^n\,dt + \sqrt{2\lambda}\partial X(t)^n\,dW \\ &= -\alpha n X(t)^n\,dt + \lambda n(n-1)X(t)^{n-2}\,dt \\ &\quad + n\sqrt{2\lambda}X(t)^{n-1}\,dW. \end{aligned}$$

- 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:

$$M_n(t) = x^n + \int_0^t \left(-\alpha n M_n(s) + \lambda n(n-1)M_{n-2}(s)\right)\,ds.$$

3. Consider the **geometric Brownian motion**

$$dX(t) = \mu X(t)\,dt + \sigma X(t)\,dW(t), \tag{127}$$

- 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

$$X(t) = X(0)\exp\left((\mu - \frac{\sigma^2}{2})t + \sigma W(t)\right). \tag{128}$$

- 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 (128) 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.

- Consider the **Landau equation**:

$$\frac{dX_t}{dt} = X_t(c - X_t^2), \quad X_0 = x. \tag{129}$$

- 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:

$$\frac{dX_t}{dt} = X_t(c - X_t^2) + \sqrt{2\sigma}\frac{dW_t}{dt}, \quad X_0 = x. \tag{130}$$

- This equation defines an ergodic Markov process on $\mathbb{R}$: There exists a unique invariant distribution:

$$\rho(x) = Z^{-1}e^{-V(x)/\sigma}, \quad Z = \int_{\mathbb{R}} e^{-V(x)/\sigma}\, dx, \quad V(x) = \frac{1}{2}cx^2 - \frac{1}{4}x^4.$$

- $\rho(x)$ 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.

$$\frac{dX_t}{dt} = X_t(c - X_t^2) + \sqrt{2\sigma} X_t \frac{dW_t}{dt}, \quad X_0 = x. \tag{131}$$

- 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 (131). 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 (131) into an SDE with additive noise:

$$dY_t = \left[(c - \sigma) - e^{2Y_t} \right] dt + \sigma\,dW_t. \tag{132}$$

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.

Theorem

Assume that ϕ is chosen sufficiently smooth so that the **backward Kolmogorov equation**

$$\begin{aligned} \frac{\partial v}{\partial t} &= \mathcal{L}v \quad \text{for } (z,t) \in \mathcal{Z} \times (0,\infty), \\ v &= \phi \quad \text{for } (z,t) \in \mathcal{Z} \times \{0\}, \end{aligned} \tag{133}$$

has a unique classical solution $v(x,t) \in C^{2,1}(\mathcal{Z} \times (0,\infty),)$. Then v is given by (125) *where $z(t)$ solves* (106)*.*

- Now we can derive rigorously the Fokker-Planck equation.

Theorem

Consider equation (106) *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**

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho \quad \text{for } (z,t) \in \mathcal{Z} \times (0,\infty), \tag{134a}$$

$$\rho = \rho_0 \quad \text{for } z \in \mathcal{Z} \times \{0\}. \tag{134b}$$

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 (134a).
- Setting $t = 0$ gives the initial condition (134b).

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:

1. Small noise asymptotics at finite time intervals.
2. Small noise asymptotics/large times (rare events): the theory of large deviations, escape from a potential well, exit time problems.
3. Small and large friction asymptotics for the Fokker-Planck equation: The Freidlin–Wentzell (underdamped) and Smoluchowski (overdamped) limits.
4. Large time asymptotics for the Langevin equation in a periodic potential: homogenization and averaging.
5. Stochastic systems with two characteristic time scales: multiscale problems and methods.

- We will study various asymptotic limits for the Langevin equation (we have set $m = 1$)

$$\ddot{q} = -\nabla V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W}. \tag{135}$$

- 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 (183) is dominated by diffusion: the Langevin equation (183) 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)

$$\kappa = \nu \exp\left(-\beta E_b\right), \tag{136}$$

- 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 (183):

$$q^\gamma(t) = \lambda_\gamma q(t/\mu_\gamma).$$

- This rescaled process satisfies the equation

$$\ddot{q}^\gamma = -\frac{\lambda_\gamma}{\mu_\gamma^2}\partial_q V(q^\gamma/\lambda_\gamma) - \frac{\gamma}{\mu_\gamma}\dot{q}^\gamma + \sqrt{2\gamma\lambda_\gamma^2\mu_\gamma^{-3}\beta^{-1}}\dot{W}, \tag{137}$$

- 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 (137) becomes

$$\gamma^{-2}\ddot{q}^\gamma = -\partial_q V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\beta^{-1}}\dot{W}. \tag{138}$$

- 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 (138) can be approximated by the solution to

$$\dot{q} = -\partial_q V + \sqrt{2\beta^{-1}}\dot{W}.$$

- $\lambda_\gamma = 1, \quad \mu_\gamma = \gamma, \quad \gamma \ll 1$:

$$\ddot{q}^\gamma = -\gamma^{-2}\nabla V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\gamma^{-2}\beta^{-1}}\dot{W}. \tag{139}$$

- 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 (139) converges to a stochastic process on a graph.

- We consider the rescaled Langevin equation (138):

$$\varepsilon^2 \ddot{q}^\gamma(t) = -\nabla V(q^\gamma(t)) - \dot{q}^\gamma(t) + \sqrt{2\beta^{-1}}\dot{W}(t), \tag{140}$$

- 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 (140), converges to $q(t)$, the solution of the Smoluchowski equation

$$\dot{q} = -\nabla V + \sqrt{2\beta^{-1}}\dot{W}. \tag{141}$$

- We write (140) as a system of SDEs:

$$\dot{q} = \frac{1}{\varepsilon}p, \tag{142}$$

$$\dot{p} = -\frac{1}{\varepsilon}\nabla V(q) - \frac{1}{\varepsilon^2}p + \sqrt{\frac{2}{\beta\varepsilon^2}}\dot{W}. \tag{143}$$

- This systems of SDEs defined a Markov process in phase space. Its generator is

$$\begin{aligned}\mathcal{L}^\varepsilon &= \frac{1}{\varepsilon^2}\left(-p\cdot\nabla_p + \beta^{-1}\Delta_p\right) + \frac{1}{\varepsilon}\left(p\cdot\nabla_q - \nabla_q V\cdot\nabla_p\right)\\ &=: \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1.\end{aligned}$$

- This is a singularly perturbed differential operator.
- We will derive the Smoluchowski equation (141) 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 (142) 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 (141) 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 (140) converges (in some appropriate sense to be explained below) to the solution of the Fokker-Planck equation corresponding to the Smoluchowski equation (141).
- 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 (140) 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 (142) and (143) is

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \mathcal{L}^* \rho \\ &= \frac{1}{\varepsilon}\left(-p\partial_q \rho + \partial_q V(q)\partial_p \rho\right) + \frac{1}{\varepsilon^2}\left(\partial_p(p\rho) + \beta^{-1}\partial_p^2 \rho\right) \\ &=: \left(\frac{1}{\varepsilon^2}\mathcal{L}_0^* + \frac{1}{\varepsilon}\mathcal{L}_1^*\right)\rho. \end{aligned} \tag{144}$$

- 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

$$\rho(p,q,t) = f(p,q,t)\rho_0(p,q). \tag{145}$$

Theorem

The function $f(p,q,t)$ *defined in* (145) *satisfies the equation*

$$\begin{aligned} \frac{\partial f}{\partial t} &= \left[\frac{1}{\varepsilon^2}\left(-p\partial_q + \beta^{-1}\partial_p^2\right) - \frac{1}{\varepsilon}\left(p\partial_q - \partial_q V(q)\partial_p\right)\right] f \\ &=: \left(\frac{1}{\varepsilon^2}\mathcal{L}_0 - \frac{1}{\varepsilon}\mathcal{L}_1\right) f. \end{aligned} \tag{146}$$

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 (144) 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 (146) in the form of a power series in ε:

$$f(p,q,t) = \sum_{n=0}^{\infty} \varepsilon^n f_n(p,q,t). \tag{147}$$

- We substitute this expansion into eqn. (146) to obtain the following system of equations.

$$\begin{aligned}
\mathcal{L}_0 f_0 &= 0, && (148)\\
\mathcal{L}_0 f_1 &= \mathcal{L}_1 f_0, && (149)\\
\mathcal{L}_0 f_2 &= \mathcal{L}_1 f_1 + \frac{\partial f_0}{\partial t} && (150)\\
\mathcal{L}_0 f_{n+1} &= \mathcal{L}_1 f_n + \frac{\partial f_n}{\partial t}, \quad n = 2,3\ldots && (151)
\end{aligned}$$

- The null space of $\mathcal{L}_0$ consists of constants in p. Consequently, from equation (148) we conclude that

$$f_0 = f(q,t).$$

- Now we can calculate the right hand side of equation (149):

$$\mathcal{L}_1 f_0 = p\partial_q f\,.$$

- Equation (149) 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 (150). 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 (150) 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:

$$\frac{\partial f}{\partial t} = -\partial_q V \partial_q f + \beta^{-1}\partial_p^2 f. \tag{152}$$

- Now we solve the equation for f_2. We use (152) to write (150) 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

$$\mathcal{L}_1 f_2 = \frac{1}{2}p^3\partial_q^3 f - p^2\partial_q^2\psi_1 + p\partial_q\psi_2 - \partial_q V\partial_q^2 fp - \partial_q V\partial_q\psi_1.$$

  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 (146):

$$\rho(p,q,t) = Z^{-1} e^{-\beta H(p,q)} \Big(f + \varepsilon(-p\partial_q f) + \mathcal{O}(\varepsilon^2) \Big),$$

- where f is the solution of (152).
- 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 (147) in terms of Hermite functions (the eigenfunctions of the generator of the OU process)

$$f_n(p,q,t) = \sum_{k=0}^{n} f_{nk}(q,t)\phi_k(p), \tag{153}$$

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:

$$\begin{aligned}\rho(x,y,t) &= \rho_0(p,q)\Big(f+\varepsilon(-\sqrt{\beta^{-1}}\partial_q f\phi_1)\\ &\quad +\varepsilon^2\big(\frac{\beta^{-1}}{\sqrt{2}}\partial_q^2 f\phi_2+f_{20}\big)\\ &\quad +\varepsilon^3\Big(-\sqrt{\frac{\beta^{-3}}{3!}}\partial_q^3 f\phi_3+\big(-\sqrt{\beta^{-1}}\widehat{\mathcal{L}}\partial_q^2 f\\ &\quad -\sqrt{\beta^{-1}}\partial_q f_{20}\big)\phi_1\Big)\Big)\\ &\quad +\mathcal{O}(\varepsilon^4),\end{aligned}$$

The Freilin–Wentzell Limit

Consider now the rescaling $\lambda_{\gamma,\varepsilon} = 1$, $\mu_{\gamma,\varepsilon} = \gamma$. The Langevin equation becomes

$$\ddot{q}^\gamma = -\gamma^{-2}\nabla V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\gamma^{-2}\beta^{-1}}\dot{W}. \tag{154}$$

We write equation (154) 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

$$\dot{H} = \left(\beta^{-1} - p^2\right) + \sqrt{2\beta^{-1}p^2}\dot{W} \tag{155a}$$

$$\dot{p}^\gamma = -\gamma^{-1}V'(q^\gamma) - p^\gamma + \sqrt{2\beta^{-1}}\dot{W}. \tag{155b}$$

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:

$$\dot{H} = \left(\beta^{-1} - \langle p^2\rangle\right) + \sqrt{2\beta^{-1}\langle p^2\rangle}\dot{W}. \tag{156}$$

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 $I_i, i = 1, \ldots d$ 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 $\{q^\gamma, p^\gamma\}$:

$$\frac{\partial u^\gamma}{\partial t} = \left(\frac{1}{\gamma}\mathcal{L}_0 + \mathcal{L}_1\right) u^\gamma. \tag{157}$$

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 (157) and equate equal powers in ε to obtain the following sequence of equations:

$$\mathcal{L}_0 u_0 = 0, \tag{158a}$$

$$\mathcal{L}_0 u_1 = -\mathcal{L}_1 u_1 + \frac{\partial u_0}{\partial t}, \tag{158b}$$

$$\mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 + \frac{\partial u_1}{\partial t}. \tag{158c}$$

$$\dots\dots\dots$$

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:

$$\mathcal{N}(\mathcal{L}_0) = \{\text{functions of} H\}. \tag{159}$$

Let us now analyze equations (158). We start with (158a); eqn. (159) implies that u_0 depends on $q,\ p$ through the Hamiltonian function H:

$$u_0 = u(H(p,q),t) \tag{160}$$

Now we proceed with (158b). For this we need to find the solvability condition for equations of the form

$$\mathcal{L}_0 u = f \tag{161}$$

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]

[1] We assume that both u_1 and F decay to 0 as $|p| \to \infty$ to justify the integration by parts that follows.

$$\begin{aligned}\int_{\mathbb{R}^2} \mathcal{L}_0 u F(H(p,q))\,dpdq &= \int_{\mathbb{R}^2} u \mathcal{L}_0^* F(H(p,q))\,dpdq \\ &= \int_{\mathbb{R}^2} u(-\mathcal{L}_0 F(H(p,q)))\,dpdq \\ &= 0, \quad \forall F \in C_b^\infty(\mathbb{R}).\end{aligned}$$

This implies that the **solvability condition** for equation (161) is that

$$\int_{\mathbb{R}^2} f(p,q) F(H(p,q))\,dpdq = 0, \quad \forall F \in C_b^\infty(\mathbb{R}). \tag{162}$$

We use the solvability condition in (158b) to obtain that

$$\int_{\mathbb{R}^2} \left(\mathcal{L}_1 u_1 - \frac{\partial u_0}{\partial t} \right) F(H(p,q))\,dpdq = 0, \tag{163}$$

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

$$\mathcal{L}_1 = \left[(\beta^{-1} - p^2)\partial_H + \beta^{-1}p^2\partial_H^2\right], \tag{164}$$

where

$$p^2 = p^2(H,q) = 2(H - V(q)).$$

We want to change variables in the integral (163) 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 (164), to rewrite eqn. (163) as

$$\int\int \left(\frac{\partial u}{\partial t} + \left[(\beta^{-1} - p^2)\partial_H + \beta^{-1}p^2\partial_H^2\right]u\right)F(H)p^{-1}(H,q)\,dHdq = 0.$$

We introduce the notation

$$\langle\cdot\rangle := \int \cdot\, dq.$$

The integration over q can be performed "explicitly":

$$\int \left[\frac{\partial u}{\partial t}\langle p^{-1}\rangle + \left((\beta^{-1}\langle p^{-1}\rangle - \langle p\rangle)\partial_H + \beta^{-1}\langle p\rangle\partial_H^2\right)u\right]F(H)\,dH = 0.$$

This equation should be valid for every smooth function $F(H)$, and this requirement leads to the differential equation

$$\langle p^{-1}\rangle \frac{\partial u}{\partial t} = \left(\beta^{-1}\langle p^{-1}\rangle - \langle p\rangle\right)\partial_H u + \langle p\rangle \beta^{-1}\partial_H^2 u,$$

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:

$$\dot{H} = b(H) + \sigma(H)\dot{W} \tag{165}$$

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 (165) 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

The limiting Fokker–Planck equation for the energy distribution function $\rho(E, t)$ is

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial E} \left(\left(I(E) + \beta^{-1} \frac{\partial}{\partial E} \right) \left(\frac{\omega(E)\rho}{2\pi} \right) \right). \tag{166}$$

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

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= -\frac{\partial}{\partial E}\left(\left(\beta^{-1}\frac{I(E)\omega(E)}{2\pi}\right)\rho\right) + \beta^{-1}\frac{\partial^2}{\partial E^2}\left(\frac{I\omega}{2\pi}\right) \\
&= -\beta^{-1}\frac{\partial \rho}{\partial E} + \frac{\partial}{\partial E}\left(\frac{I\omega}{2\pi}\rho\right) + \beta^{-1}\frac{\partial}{\partial E}\left(\frac{dI}{dE}\frac{\omega\rho}{2\pi}\right) \\
&\quad + \beta^{-1}\frac{\partial}{\partial E}\left(I\frac{\partial}{\partial E}\left(\frac{\omega\rho}{2\pi}\right)\right) \\
&= \frac{\partial}{\partial E}\left(\frac{I\omega}{2\pi}\rho\right) + \beta^{-1}\frac{\partial}{\partial E}\left(I\frac{\partial}{\partial E}\left(\frac{\omega\rho}{2\pi}\right)\right) \\
&= \frac{\partial}{\partial E}\left(\left(I(E) + \beta^{-1}\frac{\partial}{\partial E}\right)\left(\frac{\omega(E)\rho}{2\pi}\right)\right),
\end{aligned}$$

Remarks

1. 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.
2. If we rescale back to the original time-scale we obtain the equation

$$\frac{\partial \rho}{\partial t} = \gamma \frac{\partial}{\partial E}\left(\left(I(E) + \beta^{-1}\frac{\partial}{\partial E}\right)\left(\frac{\omega(E)\rho}{2\pi}\right)\right). \tag{167}$$

We will use this equation later on to calculate the rate of escape from a potential barrier in the energy-diffusion-limited regime.

Exit Time Problems and Escape from a Potential Well

Escape From a Potential Well

- There are many systems in physics, chemistry and biology that exist in at least two stable states:
  - Switching and storage devices in computers.
  - Conformational changes of biological macromolecules (they can exist in many different states).
- The problems that we would like to study 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?
  - How does the system relax to an unstable state?

- The study of bistability and metastability is a very active research area. Topics of current research include:
  - The development of numerical methods for the calculation of various quantities such as reaction rates, transition pathways etc.
  - The study of bistable systems in infinite dimensions.

- We will consider the dynamics of a particle moving in a bistable potential, under the influence of thermal noise in one dimension:

$$\dot{x} = -V'(x) + \sqrt{2k_B T}\,\dot{\beta}. \tag{168}$$

- We will consider potentials that have to local minima, one local maximum (saddle point) and they increase at least quadratically fast 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

$$V(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2 + \frac{1}{4}. \tag{169}$$

- This potential has three local minima, a local maximum at $x = 0$ and two local minima at $x = \pm 1$.

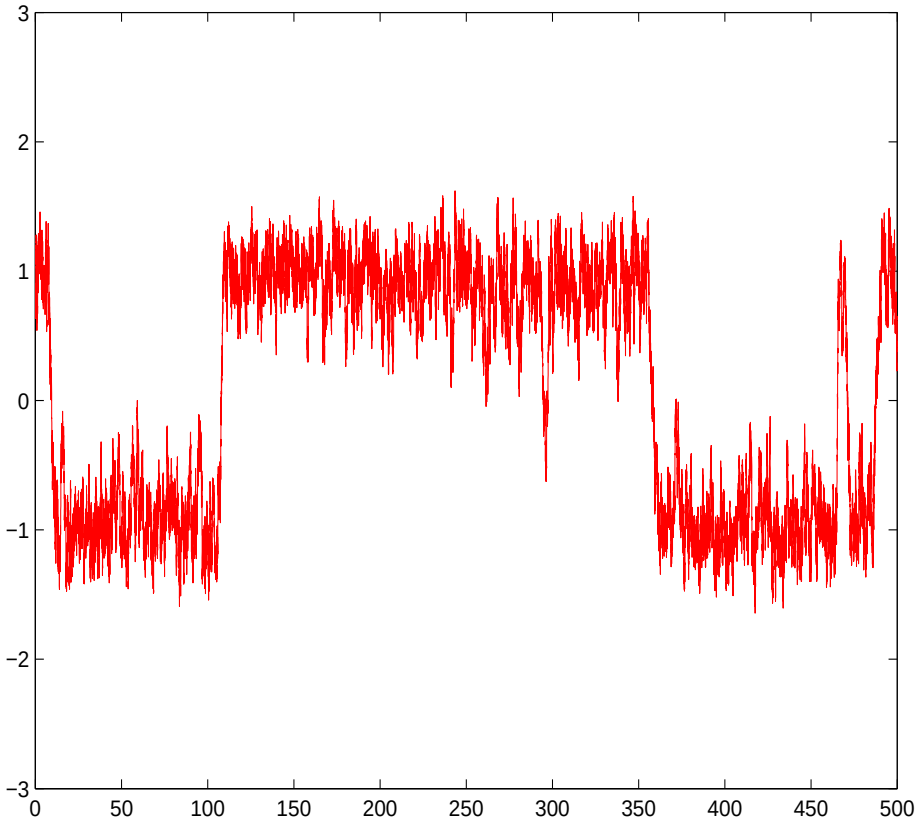

Figure: Sample Path of solution to equation (168).

- 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.
- We will consider the problem of the escape of the particle from the left local minimum a.
- The potential barrier is then defined as

$$E_b = V(b) - V(a).$$

- Our assumption that the thermal fluctuations are weak can be written as

$$\frac{k_B T}{E_b} \ll 1. \tag{170}$$

- Under condition (170), the particle is most likely to be found close to a or c. There it will perform small oscillations around either of the local minima.
- We can study the weak noise asymptotics for finite times using standard perturbation theory.
- We can describe locally the dynamics of the particle by appropriate Ornstein-Uhlenbeck processes.
- 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 $\beta := (k_B T)^{-1}$:
$$\tau = \nu^{-1} \exp(\beta E_b).$$
- the rate with which particles escape from a local minimum of the potential is called the **rate of escape** or the **reaction rate** $\kappa := \tau^{-1}$:
$$\kappa = \nu \exp(-\beta \Delta E). \tag{171}$$
- It is 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**:
- In the absence of thermal fluctuations 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.
- We will study this problem and calculate the escape rate using by calculating the **mean first passage time**.

- The Arrhenius-type factor in the formula for the reaction rate, eqn. (171) is intuitively clear 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 later on.

- Let X_t be a continuous time diffusion process on $\mathbb{R}^d$ whose evolution is governed by the SDE

$$dX_t^x = b(X_t^x)\,dt + \sigma(X_t^x)\,dW_t, \quad X_0^x = x. \tag{172}$$

- 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

$$\tau_D^x = \inf\{t \geqslant 0 \,:\, X_t^x \notin D\}.$$

- 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

The MFPT is the solution of the boundary value problem

$$-\mathcal{L}\tau = 1, \quad x \in D, \tag{173a}$$

$$\tau = 0, \quad x \in \partial D, \tag{173b}$$

where $\mathcal{L}$ is the generator of the SDE 173.

- 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 67 is based on Itô's formula.

Derivation of the BVP for the MFPT

- 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.

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad \rho(X,x,0) = \delta(X-x), \quad \rho|_{\partial D} = 0. \tag{174}$$

- 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 $f(x,t)$:

$$\begin{aligned}
\tau(x) &= \int_0^{+\infty} f(s,x)s\,ds = \int_0^{+\infty} -\frac{dS}{ds}s\,ds \\
&= \int_0^{+\infty} S(s,x)\,ds = \int_0^{+\infty}\int_D \rho(X,x,s)\,dXds \\
&= \int_0^{+\infty}\int_D e^{\mathcal{L}^* s}\delta(X-x)\,dXds \\
&= \int_0^{+\infty}\int_D \delta(X-x)\left(e^{\mathcal{L}s}1\right)dXds = \int_0^{+\infty}\left(e^{\mathcal{L}s}1\right)ds.
\end{aligned}$$

- We apply $\mathcal{L}$ to the above equation to deduce:

$$\begin{aligned}
\mathcal{L}\tau &= \int_0^{+\infty}\left(\mathcal{L}e^{\mathcal{L}t}1\right)dt = \int_0^t \frac{d}{dt}\left(e^{\mathcal{L}t}1\right)dt \\
&= -1.
\end{aligned}$$

- Consider the boundary value problem for the MFPT of the one dimensional diffusion process (168) from the interval (a, b):

$$-\beta^{-1} e^{\beta V} \partial_x \left(e^{-\beta V} \partial_x \tau \right) = 1 \tag{175}$$

- We choose reflecting BC at $x = a$ and absorbing B.C. at $x = b$. We can solve (175) with these boundary conditions by quadratures:

$$\tau(x) = \beta^{-1} \int_x^b dy e^{\beta V(y)} \int_a^y dz e^{-\beta V(z)}. \tag{176}$$

- 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 ∞:

$$\begin{aligned} \int_{-\infty}^z \exp(-\beta V(z))\, dz &\approx \int_{-\infty}^{+\infty} \exp(-\beta V(a)) \\ &\quad \exp\left(-\frac{\beta\omega_0^2}{2}(z-a)^2\right) dz \\ &= \exp\left(-\beta V(a)\right) \sqrt{\frac{2\pi}{\beta\omega_0^2}}, \end{aligned}$$

- where we have used the Taylor series expansion around the minimum:

$$V(z) = V(a) + \frac{1}{2}\omega_0^2 (z-a)^2 + \ldots$$

- 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 half of the particles escape. Consequently, the escape rate (or reaction rate), is given by $\frac{1}{2\tau}$:

$$\kappa = \frac{\omega_0 \omega_b}{2\pi} \exp\left(-\beta E_b\right). \tag{177}$$

- Consider now the problem of escape from a potential well for the Langevin equation

$$\ddot{q} = -\partial_q V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W}. \tag{178}$$

- The reaction rate depends on the fiction coefficient and the temperature. In the overdamped limit ($\gamma \gg 1$) we retrieve (177), appropriately rescaled with γ:

$$\kappa = \frac{\omega_0 \omega_b}{2\pi\gamma} \exp\left(-\beta E_b\right). \tag{179}$$

- We can also obtain a formula for the reaction rate for $\gamma = \mathcal{O}(1)$:

$$\kappa = \frac{\sqrt{\frac{\gamma^2}{4} - \omega_b^2} - \frac{\gamma}{2}}{\omega_b} \frac{\omega_0}{2\pi} \exp\left(-\beta E_b\right). \tag{180}$$

- Naturally, in the limit as $\gamma \to +\infty$ (180) reduces to (179)

- 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, (165) or (166). The result is

$$\kappa = \gamma\beta I(E_b)\frac{\omega_0}{2\pi}e^{-\beta E_b}, \tag{181}$$

- 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 (179) and (181) in the overdamped and underdamped limits, respectively.

- The escape rate can be calculated in closed form only in 1d or in multidimesional problems that can be reduced to a 1d problem (e.g., problems with radial symmetry).
- For more information on the calculation of escape rates and applications of reaction rate theory see
  - P. Hänggi, P. Talkner and M. Borkovec: *Reaction rate theory: fifty years after Kramers*. Rev. Mod. Phys. 62(2) 251-341 (1990).
  - R. Zwanzig: *Nonequilibrium Statistical Mechanics*, Oxford 2001, Ch. 4.
  - C.W. Gardiner: *Handbook of Stochastic Methods*, Springer 2002, Ch. 9.
  - Freidlin and Wentzell: *Random Perturbations of Dynamical Systems*, Springer 1998, Ch. 6.

The Kac-Zwanzig model and the Generalized Langevin Equation

Contents

- Introduction: Einstein's theory of Brownian motion.
- Particle coupled to a heat bath: the Kac-Zwanzig model.
- Coupled particle/field models.
- The Mori-Zwanzig formalism.

- Heavy (Brownian) particle immersed in a fluid.
- Collisions of the heavy particle with the light fluid particles.
- Statistical description of the motion of the Brownian particle.
- Fluid is assumed to be in equilibrium.
- Brownian particle performs an effective random walk:

$$\langle X(t)^2 \rangle = 2Dt.$$

- The constant D is the diffusion coefficient (which should be calculated from first principles, i.e. through Green-Kubo formulas).

- Phenomenological theories based either on an equation for the evolution of the probability distribution function $\rho(q,p,t)$ (the **Fokker-Planck equation**)

$$\partial_t \rho = -p\partial_q \rho + \partial_q V \partial_p \rho + \gamma\left(\partial_p(p\rho) + \beta^{-1}\partial_p^2 \rho\right) \tag{182}$$

- or the a stochastic equation for the evolution of trajectories, the **Langevin equation**

$$\ddot{q} = -V'(q) - \gamma\dot{q} + \sqrt{2\gamma\beta^{-1}}\xi, \tag{183}$$

- where $\xi(t)$ is a white noise process, i.e. a (generalized) mean zero Gaussian Markov process with correlation function

$$\langle \xi(t)\xi(s)\rangle = \delta(t-s).$$

- Notice that the noise and dissipation in the Langevin equation are not independent. They are related through the **fluctuation-dissipation theorem**.
- This is not a coincidence since, as we will see later, noise and dissipation have the same source, namely the interaction between the Brownian particle and its environment (heat bath).
- There are two phenomenological constants, the **inverse temperature** β and the **friction coefficient** γ.
- Our goal is to derive the Fokker-Planck and Langevin equations from first principles and to calculate the phenomenological coefficients β and γ.
- We will consider some simple "particle + environment" systems for which we can obtain rigorously a stochastic equation that describes the dynamics of the Brownian particle.

- We can describe the dynamics of the Brownian particle/fluid system:

$$H(Q_N, P_N; q, p) = H_{BP}(Q_N, P_N) + H_{HB}(q, p) + H_I(Q_N, q), \quad (184)$$

- where $\{q, p\} := \left\{\{q_j\}_{j=1}^N, \{p_j\}_{j=1}^N\right\}$ are the positions and momenta of the fluid particles, N is the number of fluid (**heat bath**) particles (we will need to take the thermodynamic limit $N \to +\infty$).
- The initial conditions of the Brownian particle are taken to be fixed, whereas the fluid is assumed to be initially in equilibrium (Gibbs distribution).

- Goal: eliminate the fluid variables $\{q, p\} := \left\{\{q_j\}_{j=1}^N, \{p_j\}_{j=1}^N\right\}$ to obtain a closed equation for the Brownian particle.
- We will see that this equation is a stochastic integrodifferential equation, the **Generalized Langevin Equation (GLE)** (in the limit as $N \to +\infty$)

$$\ddot{Q} = -V'(Q) - \int_0^t R(t-s)\dot{Q}(s)\,ds + F(t), \tag{185}$$

- where $R(t)$ is the **memory kernel** and $F(t)$ is the **noise**.
- We will also see that, in some appropriate limit, we can derive the Markovian Langevin equation (183).

- Need to model the interaction between the heat bath particles and the coupling between the Brownian particle and the heat bath.
- The simplest model is that of a harmonic heat bath and of linear coupling:

$$H(Q_N, P_N, q, p) = \frac{P_N^2}{2} + V(Q_N) + \sum_{n=1}^{N} \frac{p_n^2}{2m_n} + \frac{1}{2} k_n (q_n - \lambda Q_N)^2 \quad (186)$$

- The initial conditions of the Brownian particle $\{Q_N(0), P_N(0)\} := \{Q_0, P_0\}$ are taken to be deterministic.
- The initial conditions of the heat bath particles are distributed according to the Gibbs distribution, conditional on the knowledge of $\{Q_0, P_0\}$:

$$\mu_\beta(dpdq) = Z^{-1} e^{-\beta H(q,p)}\, dqdp, \quad (187)$$

where β is the inverse temperature. This is a way of introducing the concept of the temperature in the system (through the average kinetic energy of the bath particles).

- In order to choose the initial conditions according to $\mu_\beta(dpdq)$ we can take

$$q_n(0) = \lambda Q_0 + \sqrt{\beta^{-1} k_n^{-1}}\xi_n, \quad p_n(0) = \sqrt{m_n \beta^{-1}}\eta_n, \tag{188}$$

  where the $\xi_n\, \eta_n$ are mutually independent sequences of i.i.d. $\mathcal{N}(0,1)$ random variables.
- Notice that we actually consider the Gibbs measure of an effective (renormalized) Hamiltonian.
- Other choices for the initial conditions are possible. For example, we can take $q_n(0) = \sqrt{\beta^{-1} k_n^{-1}}\xi_n$. Our choice of I.C. ensures that the forcing term in the GLE that we will derive is mean zero (see below).

- Hamilton's equations of motion are:

$$\ddot{Q}_N + V'(Q_N) = \sum_{n=1}^{N} k_n(\lambda q_n - \lambda^2 Q_N), \qquad (189a)$$

$$\ddot{q}_n + \omega_n^2(q_n - \lambda Q_N) = 0, \quad n = 1, \ldots N, \qquad (189b)$$

  where $\omega_n^2 = k_n/m_n$.
- The equations for the heat bath particles are second order linear inhomogeneous equations with constant coefficients. Our plan is to solve them and then to substitute the result in the equations of motion for the Brownian particle.

- We can solve the equations of motion for the heat bath variables using the variation of constants formula

$$\begin{aligned} q_n(t) &= q_n(0)\cos(\omega_n t) + \frac{p_n(0)}{m_n\omega_n}\sin(\omega_n t) \\ &\quad + \omega_n\lambda\int_0^t \sin(\omega_n(t-s))Q_N(s)\,ds. \end{aligned}$$

- An integration by parts yields

$$\begin{aligned} q_n(t) &= q_n(0)\cos(\omega_n t) + \frac{p_n(0)}{m_n\omega_n}\sin(\omega_n t) + \lambda Q_N(t) \\ &\quad - \lambda Q_N(0)\cos(\omega_n t) - \lambda\int_0^t \cos(\omega_n(t-s))\dot{Q}_N(s)\,ds. \end{aligned}$$

- We substitute this in equation (189) and use the initial conditions (188) to obtain the **Generalized Langevin Equation**

$$\ddot{Q}_N = -V'(Q_N) - \lambda^2 \int_0^t R_N(t-s)\dot{Q}_N(s)\,ds + \lambda F_N(t), \qquad (190)$$

- where the **memory kernel** is

$$R_N(t) = \sum_{n=1}^{N} k_n \cos(\omega_n t) \qquad (191)$$

- and the **noise process** is

$$\begin{aligned} F_N(t) &= \sum_{n=1}^{N} k_n \left(q_n(0) - \lambda Q_0\right)\cos(\omega_n t) + \frac{k_n p_n(0)}{m_n \omega_n}\sin(\omega_n t) \\ &= \sqrt{\beta^{-1}} \sum_{n=1}^{N} \sqrt{k_n}\left(\xi_n \cos(\omega_n t) + \eta_n \sin(\omega_n t)\right). \qquad (192) \end{aligned}$$

1. The noisy and random term are related through the **fluctuation-dissipation theorem**:

$$\begin{aligned}\langle F_N(t)F_N(s)\rangle &= \beta^{-1}\sum_{n=1}^{N} k_n\big(\cos(\omega_n t)\cos(\omega_n s)\\ &\quad + \sin(\omega_n t)\sin(\omega_n s)\big)\\ &= \beta^{-1}R_N(t-s). \end{aligned} \tag{193}$$

2. The noise $F(t)$ is a mean zero Gaussian process.
3. The choice of the initial conditions (188) for q, p is crucial for the form of the GLE and, in particular, for the fluctuation-dissipation theorem (193) to be valid.
4. The parameter λ measures the strength of the coupling between the Brownian particle and the heat bath.
5. By choosing the frequencies ω_n and spring constants $k_n(\omega)$ of the heat bath particles appropriately we can pass to the limit as $N \to +\infty$ and obtain the GLE with different memory kernels $R(t)$ and noise processes $F(t)$.

- Let $a \in (0,1)$, $2b = 1 - a$ and set $\omega_n = N^a \zeta_n$ where $\{\zeta_n\}_{n=1}^{\infty}$ are i.i.d. with $\zeta_1 \sim \mathcal{U}(0,1)$. Furthermore, we choose the spring constants according to

$$k_n = \frac{f^2(\omega_n)}{N^{2b}},$$

- where the function $f(\omega_n)$ decays sufficiently fast at infinity.
- We can rewrite the dissipation and noise terms in the form

$$R_N(t) = \sum_{n=1}^{N} f^2(\omega_n) \cos(\omega_n t)\, \Delta\omega$$

and

$$F_N(t) = \sum_{n=1}^{N} f(\omega_n) \left(\xi_n \cos(\omega_n t) + \eta_n \sin(\omega_n t)\right) \sqrt{\Delta\omega},$$

- where $\Delta\omega = N^a/N$.

- Using now properties of Fourier series with random coefficients/frequencies and of weak convergence of probability measures we can pass to the limit:

$$R_N(t) \to R(t) \quad \text{in } L^1[0,T],$$

- for a.a. $\{\zeta_n\}_{n=1}^{\infty}$ and

$$F_N(t) \to F(t) \quad \text{weakly in } C([0,T],\mathbb{R}).$$

- The time $T > 0$ if finite but arbitrary.
- The limiting kernel and noise satisfy the fluctuation-dissipation theorem (193):

$$\langle F(t)F(s)\rangle = \beta^{-1}R(t-s). \tag{194}$$

- $Q_N(t)$, the solution of (190) converges weakly to the solution of the limiting GLE

$$\ddot{Q} = -V'(Q) - \lambda^2 \int_0^t R(t-s)\dot{Q}(s)\,ds + \lambda F(t). \tag{195}$$

- The properties of the limiting dissipation and noise are determined by the function $f(\omega)$.
- Ex: Consider the Lorentzian function

$$f^2(\omega) = \frac{2\alpha/\pi}{\alpha^2 + \omega^2} \tag{196}$$

- with $\alpha > 0$. Then

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

- The noise process $F(t)$ is a mean zero stationary Gaussian process with continuous paths and, from (194), exponential correlation function:

$$\langle F(t)F(s)\rangle = \beta^{-1}e^{-\alpha|t-s|}.$$

- Hence, $F(t)$ is the stationary Ornstein-Uhlenbeck process:

$$\frac{dF}{dt} = -\alpha F + \sqrt{2\beta^{-1}\alpha}\frac{dW}{dt}, \tag{197}$$

with $F(0) \sim \mathcal{N}(0, \beta^{-1})$.

- The GLE (195) becomes

$$\ddot{Q} = -V'(Q) - \lambda^2 \int_0^t e^{-\alpha|t-s|}\dot{Q}(s)\,ds + \lambda^2 F(t), \tag{198}$$

- where $F(t)$ is the OU process (197).

- $Q(t)$, the solution of the GLE (195), is not a Markov process, i.e. the future is not statistically independent of the past, when conditioned on the present. The stochastic process $Q(t)$ has memory.
- We can turn (195) into a Markovian SDE by enlarging the dimension of state space, i.e. introducing auxiliary variables.
- We might have to introduce infinitely many variables!
- For the case of the exponential memory kernel, when the noise is given by an OU process, it is sufficient to introduce one auxiliary variable.

- We can rewrite (198) as a system of SDEs:

$$\begin{aligned}
\frac{dQ}{dt} &= P, \\
\frac{dP}{dt} &= -V'(Q) + \lambda Z, \\
\frac{dZ}{dt} &= -\alpha Z - \lambda P + \sqrt{2\alpha\beta^{-1}}\frac{dW}{dt},
\end{aligned}$$

- where $Z(0) \sim \mathcal{N}(0, \beta^{-1})$.
- The process $\{Q(t), P(t), Z(t)\} \in \mathbb{R}^3$ is Markovian.
- It is a degenerate Markov process: noise acts directly only on one of the 3 degrees of freedom.
- We can eliminate the auxiliary process Z by taking an appropriate distinguished limit.

- Set $\lambda = \sqrt{\gamma}\varepsilon^{-1}$, $\alpha = \varepsilon^{-2}$. Equations (200) become

$$\begin{aligned} \frac{dQ}{dt} &= P, \\ \frac{dP}{dt} &= -V'(Q) + \frac{\sqrt{\gamma}}{\varepsilon} Z, \\ \frac{dZ}{dt} &= -\frac{1}{\varepsilon^2} Z - \frac{\sqrt{\gamma}}{\varepsilon} P + \sqrt{\frac{2\beta^{-1}}{\varepsilon^2}} \frac{dW}{dt}. \end{aligned}$$

- We can use tools from singular perturbation theory for Markov processes to show that, in the limit as $\varepsilon \to 0$, we have that

$$\frac{1}{\varepsilon} Z \to \sqrt{2\gamma\beta^{-1}} \frac{dW}{dt} - \gamma P.$$

- Thus, in this limit we obtain the Markovian **Langevin Equation** ($R(t) = \gamma\delta(t)$)

$$\ddot{Q} = -V'(Q) - \gamma\dot{Q} + \sqrt{2\gamma\beta^{-1}} \frac{dW}{dt}. \tag{201}$$

- Whenever the GLE (195) has "finite memory", we can represent it as a Markovian SDE by adding a **finite number** of additional variables.
- These additional variables are solutions of a linear system of SDEs.
- This follows from results in approximation theory.
- Consider now the case where the memory kernel is a bounded analytic function. Its Laplace transform

$$\widehat{R}(s) = \int_0^{+\infty} e^{-st} R(t)\, dt$$

- can be represented as a continued fraction:

$$\widehat{R}(s) = \cfrac{\Delta_1^2}{s + \gamma_1 + \cfrac{\Delta_2^2}{\ldots}}, \quad \gamma_i \geqslant 0, \tag{202}$$

- Since $R(t)$ is bounded, we have that

$$\lim_{s \to \infty} \widehat{R}(s) = 0.$$

- Consider an approximation $R_N(t)$ such that the continued fraction representation terminates after N steps.
- $R_N(t)$ is bounded which implies that

$$\lim_{s\to\infty} \widehat{R}_N(s) = 0.$$

- The Laplace transform of $R_N(t)$ is a rational function:

$$\widehat{R}_N(s) = \frac{\sum_{j=1}^{N} a_j s^{N-j}}{s^N + \sum_{j=1}^{N} b_j s^{N-j}}, \quad a_j, b_j \in \mathbb{R}. \tag{203}$$

- This is the Laplace transform of the autocorrelation function of an appropriate linear system of SDEs. Indeed, set

$$\frac{dx_j}{dt} = -b_j x_j + x_{j+1} + a_j \frac{dW_j}{dt}, \quad j = 1, \ldots, N, \tag{204}$$

- with $x_{N+1}(t) = 0$. The process $x_1(t)$ is a stationary Gaussian process with autocorrelation function $R_N(t)$.

- For $N = 1$ and $b_1 = \alpha$, $a_1 = \sqrt{2\beta^{-1}\alpha}$ we derive the GLE (198) with $F(t)$ being the OU process (197).
- Consider now the case $N = 2$ with $b_i = \alpha_i$, $i = 1, 2$ and $a_1 = 0$, $a_2 = \sqrt{2\beta^{-1}\alpha_2}$. The GLE becomes

$$\begin{aligned}
\ddot{Q} &= -V'(Q) - \lambda^2 \int_0^t R(t-s)\dot{Q}(s)\,ds + \lambda F_1(t), \\
\dot{F}_1 &= -\alpha_1 F_1 + F_2, \\
\dot{F}_2 &= -\alpha_2 F_2 + \sqrt{2\beta^{-1}\alpha_2}\dot{W}_2,
\end{aligned}$$

with

$$\beta^{-1}R(t-s) = \langle F_1(t)F_1(s)\rangle.$$

- We can write (206) as a Markovian system for the variables $\{Q, P, Z_1, Z_2\}$:

$$\begin{aligned} \dot{Q} &= P, \\ \dot{P} &= -V'(Q) + \lambda Z_1(t), \\ \dot{Z}_1 &= -\alpha_1 Z_1 + Z_2, \\ \dot{Z}_2 &= -\alpha_2 Z_2 - \lambda P + \sqrt{2\beta^{-1}\alpha_2}\dot{W}_2. \end{aligned}$$

- Notice that this diffusion process is "more degenerate" than (198): noise acts on fewer degrees of freedom.
- It is still, however, hypoelliptic (Hormander's condition is satisfied): there is sufficient interaction between the degrees of freedom $\{Q, P, Z_1, Z_2\}$ so that noise (and hence regularity) is transferred from the degrees of freedom that are directly forced by noise to the ones that are not.
- The corresponding Markov semigroup has nice regularizing properties. There exists a smooth density.

- Stochastic processes that can be written as a Markovian process by adding a **finite number** of additional variables are called **quasimarkovian**.
- Under appropriate assumptions on the potential $V(Q)$ the solution of the GLE equation is an ergodic process.
- It is possible to study the ergodic properties of a quasimarkovian processes by analyzing the spectral properties of the generator of the corresponding Markov process.
- This leads to the analysis of the spectral properties of hypoelliptic operators.

- When studying the Kac-Zwanzing model we considered a one dimensional Hamiltonian system coupled to a finite dimensional Hamiltonian system with random initial conditions (the harmonic heat bath) and then passed to the theromdynamic limit $N \to \infty$.
- We can consider a small Hamiltonian system coupled to its environment which we model as an infinite dimensional Hamiltonian system with random initial conditions. We have a **coupled particle-field model**.
- The distinguished particle (Brownian particle) is described through the Hamiltonian

$$H_{DP} = \frac{1}{2}p^2 + V(q). \tag{207}$$

- We will model the environment through a classical linear field theory (i.e. the wave equation) with infinite energy:

$$\partial_t^2 \phi(t,x) = \partial_x^2 \phi(t,x). \tag{208}$$

- The Hamiltonian of this system is

$$\mathcal{H}_{HB}(\phi,\pi) = \int \left(|\partial_x \phi|^2 + |\pi(x)|^2\right). \tag{209}$$

- $\pi(x)$ denotes the conjugate momentum field.
- The initial conditions are distributed according to the Gibbs measure (which in this case is a Gaussian measure) at inverse temperature β, which we formally write as

$$"\mu_\beta = Z^{-1} e^{-\beta \mathcal{H}(\phi,\pi)}\, d\phi d\pi". \tag{210}$$

- Care has to be taken when defining probability measures in infinite dimensions.

- Under this assumption on the initial conditions, typical configurations of the heat bath have infinite energy.
- In this way, the environment can pump enough energy into the system so that non-trivial fluctuations emerge.
- We will assume linear coupling between the particle and the field:

$$H_I(q,\phi) = q \int \partial_q \phi(x)\rho(x)\,dx. \tag{211}$$

- where The function $\rho(x)$ models the coupling between the particle and the field.
- This coupling is influenced by the dipole coupling approximation from classical electrodynamics.

- The Hamiltonian of the particle-field model is

$$H(q,p,\phi,\pi) = H_{DP}(p,q) + \mathcal{H}(\phi,\pi) + H_I(q,\phi). \tag{212}$$

- The corresponding Hamiltonian equations of motion are a coupled system of equations of the coupled particle field model.
- Now we can proceed as in the case of the finite dimensional heat bath. We can integrate the equations motion for the heat bath variables and plug the solution into the equations for the Brownian particle to obtain the GLE. The final result is

$$\ddot{q} = -V'(q) - \int_0^t R(t-s)\dot{q}(s) + F(t), \tag{213}$$

- with appropriate definitions for the memory kernel and the noise, which are related through the fluctuation-dissipation theorem.

- Consider now the $N+1$-dimensional Hamiltonian (particle + heat bath) with random initial conditions. The $N+1-$ **probability distribution function** f_{N+1} satisfies the **Liouville equation**

$$\frac{\partial f_{N+1}}{\partial t} + \{f_{N+1}, H\} = 0, \tag{214}$$

- where H is the full Hamiltonian and $\{\cdot,\cdot\}$ is the Poisson bracket

$$\{A, B\} = \sum_{j=0}^{N} \left(\frac{\partial A}{\partial q_j}\frac{\partial B}{\partial p_j} - \frac{\partial B}{\partial q_j}\frac{\partial A}{\partial p_j} \right).$$

- We introduce the Liouville operator

$$L_{N+1}\cdot = -i\{\cdot, H\}.$$

- The Liouville equation can be written as

$$i\frac{\partial f_{N+1}}{\partial t} = L_{N+1} f_{N+1}. \tag{215}$$

- We want to obtain a closed equation for the distribution function of the Brownian particle. We introduce a projection operator which projects onto the distribution function f of the Brownian particle:

$$Pf_{N+1} = f, \quad Pf_{N+1} = h.$$

- The Liouville equation becomes

$$i\frac{\partial f}{\partial t} = PL(f+h), \tag{216a}$$

$$i\frac{\partial h}{\partial t} = (I-P)L(f+h). \tag{216b}$$

- We integrate the second equation and substitute into the first equation. We obtain

$$i\frac{\partial f}{\partial t} = PLf - i\int_0^t PLe^{-i(I-P)Ls}(I-P)Lf(t-s)\,ds + PLe^{-i(I-P)Lt}h(0). \tag{217}$$

- In the Markovian limit (large mass ratio) we obtain the Fokker-Planck equation (182).

- L. Rey-Bellet *Open Classical Systems*.
- D. Givon, R. Kupferman and A.M. Stuart *Extracting Marcoscopic Dynamics: Model Problems and Algorithms*, Nonlinearity **17** (2004) R55-R127.
- R.M. Mazo Brownian Motion, Oxford University Press, 2002.