STOCHASTIC PROCESSES AND APPLICATIONS

G.A. Pavliotis
Department of Mathematics
Imperial College London

November 11, 2015

Contents

Chapter 1

Introduction to Stochastic Processes

In this chapter we present some basic results from the theory of stochastic processes and investigate the properties of some of the standard continuous-time stochastic processes. In Section 1.1 we give the definition of a stochastic process. In Section 1.2 we present some properties of stationary stochastic processes. In Section 1.3 we introduce Brownian motion and study some of its properties. Various examples of stochastic processes in continuous time are presented in Section 1.4. The Karhunen-Loeve expansion, one of the most useful tools for representing stochastic processes and random fields, is presented in Section 1.5. Further discussion and bibliographical comments are presented in Section 1.6. Section 1.7 contains exercises.

1.1 Definition of a Stochastic Process

Stochastic processes describe dynamical systems whose time-evolution is of probabilistic nature. The precise definition is given below.[1]

Definition 1.1 (stochastic process)**.** *Let T be an ordered set, $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $(E, \mathcal{G})$ a measurable space. 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 $(\Omega, \mathcal{F}, \mathbb{P})$ to $(E, \mathcal{G})$. Ω is known as the sample space, where E is the state space of the stochastic process X_t.*

The set T can be either discrete, for example the set of positive integers $\mathbb{Z}_+$, or continuous, $T = \mathbb{R}_+$. 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 (realization, trajectory) of the process X.

Definition 1.2 (finite dimensional distributions)**.** *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 $\mathbf{x} = (x_1, \ldots, x_k)$.

[1] The notation and basic definitions from probability theory that we will use can be found in Appendix B.

From experiments or numerical simulations we can only obtain information about the finite dimensional distributions of a process. A natural question arises: are the finite dimensional distributions of a stochastic process sufficient to determine a stochastic process uniquely? This is true for processes with continuous paths [2], which is the class of stochastic processes that we will study in these notes.

Definition 1.3. *We say that two processes X_t and Y_t are equivalent if they have same finite dimensional distributions.*

Gaussian stochastic processes

A very important class of continuous-time processes is that of Gaussian processes which arise in many applications.

Definition 1.4. *A one dimensional continuous time Gaussian process is a stochastic process for which $E = \mathbb{R}$ and all the finite dimensional distributions are Gaussian, i.e. every finite dimensional vector $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ is a $\mathcal{N}(\mu_k, K_k)$ random variable for some vector μ_k and a symmetric nonnegative definite matrix K_k for all $k = 1, 2, \dots$ and for all $t_1, t_2, \dots, t_k$.*

From the above definition we conclude that the finite dimensional distributions of a Gaussian continuous-time stochastic process are Gaussian with probability distribution function

$$\gamma_{\mu_k, K_k}(\mathbf{x}) = (2\pi)^{-n/2} (\det K_k)^{-1/2} \exp\left[-\frac{1}{2} \langle K_k^{-1}(\mathbf{x} - \mu_k), \mathbf{x} - \mu_k \rangle \right],$$

where $\mathbf{x} = (x_1, x_2, \dots x_k)$.

It is straightforward to extend the above definition to arbitrary dimensions. A Gaussian process $x(t)$ is characterized by its mean

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

and the covariance (or autocorrelation) matrix

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

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

It is not difficult to simulate Gaussian stochastic processes on a computer. Given a random number generator that generates $\mathcal{N}(0, 1)$ (pseudo)random numbers, we can sample from a Gaussian stochastic process by calculating the square root of the covariance. A simple algorithm for constructing a skeleton of a continuous time Gaussian process is the following:

- Fix Δt and define $t_j = (j-1)\Delta t,\ j = 1, \dots N$.
- Set $X_j := X(t_j)$ and define the Gaussian random vector $X^N = \left\{ X_j^N \right\}_{j=1}^N$. Then $X^N \sim \mathcal{N}(\mu^N,\, \Gamma^N)$ with $\mu^N = (\mu(t_1), \dots \mu(t_N))$ and $\Gamma_{ij}^N = C(t_i, t_j)$.

[2] In fact, all we need is the stochastic process to be *separable* See the discussion in Section 1.6.

- Then $X^N = \mu^N + \Lambda\,\mathcal{N}(0, I)$ with $\Gamma^N = \Lambda\Lambda^T$.

We can calculate the square root of the covariance matrix C either using the Cholesky factorization, via the spectral decomposition of C, or by using the singular value decomposition (SVD).

Examples of Gaussian stochastic processes

- Random Fourier series: let $\xi_i,\ \zeta_i \sim \mathcal{N}(0,1),\ i = 1, \dots N$ and define

$$X(t) = \sum_{j=1}^{N} \left(\xi_j \cos(2\pi j t) + \zeta_j \sin(2\pi j t)\right).$$

- Brownian motion is a Gaussian process with $m(t) = 0$, $C(t,s) = \min(t,s)$.
- Brownian bridge is a Gaussian process with $m(t) = 0$, $C(t,s) = \min(t,s) - ts$.
- The Ornstein-Uhlenbeck process is a Gaussian process with $m(t) = 0$, $C(t,s) = \lambda\, e^{-\alpha|t-s|}$ with $\alpha,\ \lambda > 0$.

1.2 Stationary Processes

In many stochastic processes that appear in applications their statistics remain invariant under time translations. Such stochastic processes are called *stationary*. It is possible to develop a quite general theory for stochastic processes that enjoy this symmetry property. It is useful to distinguish between stochastic processes for which all finite dimensional distributions are translation invariant (strictly stationary processes) and processes for which this translation invariance holds only for the first two moments (weakly stationary processes).

Strictly Stationary Processes

Definition 1.5. *A stochastic process is called (strictly) stationary if all finite dimensional distributions are invariant under time translation: for any integer k and times $t_i \in T$, the distribution of $(X(t_1), X(t_2), \dots, X(t_k))$ is equal to that of $(X(s+t_1), X(s+t_2), \dots, X(s+t_k))$ for any s such that $s+t_i \in T$ for all $i \in \{1, \dots, k\}$. In other words,*

$$\mathbb{P}(X_{t_1+s} \in A_1, X_{t_2+s} \in A_2 \dots X_{t_k+s} \in A_k) = \mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2 \dots X_{t_k} \in A_k),\ \ \forall s \in T.$$

Example 1.6. Let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables and consider the stochastic process $X_n = Y_n$. Then X_n is a strictly stationary process (see Exercise 1). Assume furthermore that $\mathbb{E}Y_0 = \mu < +\infty$. Then, by the strong law of large numbers, Equation (B.26), we have that

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

almost surely. In fact, the *Birkhoff ergodic theorem* states that, for any function f such that $\mathbb{E}f(Y_0) < +\infty$, we have that

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{j=0}^{N-1} f(X_j) = \mathbb{E}f(Y_0), \tag{1.1}$$

almost surely. The sequence of iid random variables is an example of an ergodic strictly stationary processes.

We will say that a stationary stochastic process that satisfies (1.1) is *ergodic*. For such processes we can calculate expectation values of observable, $\mathbb{E}f(X_t)$ using a single sample path, provided that it is long enough ($N \gg 1$).

Example 1.7. Let Z be a random variable and define the stochastic process $X_n = Z$, $n = 0, 1, 2, \ldots$. Then X_n is a strictly stationary process (see Exercise 2). We can calculate the long time average of this stochastic process:

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

which is independent of N and does not converge to the mean of the stochastic processes $\mathbb{E}X_n = \mathbb{E}Z$ (assuming that it is finite), or any other deterministic number. This is an example of a non-ergodic processes.

Second Order Stationary Processes

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

$$\mathbb{E}(X_{t+s}) = \mathbb{E}X_t, \quad s \in T, \tag{1.2}$$

from which we conclude that $\mathbb{E}X_t$ is constant and

$$\mathbb{E}((X_{t_1+s} - \mu)(X_{t_2+s} - \mu)) = \mathbb{E}((X_{t_1} - \mu)(X_{t_2} - \mu)), \quad s \in T, \tag{1.3}$$

implies that the *covariance function* depends on the difference between the two times, t and s:

$$C(t, s) = C(t - s).$$

This motivates the following definition.

Definition 1.8. *A stochastic process $X_t \in L^2$ is called second-order stationary, wide-sense stationary or weakly stationary if the first moment $\mathbb{E}X_t$ is a constant and the covariance function $\mathbb{E}(X_t - \mu)(X_s - \mu)$ depends only on the difference $t - s$:*

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

The constant μ is the expectation of the process X_t. Without loss of generality, we can set $\mu = 0$, since if $\mathbb{E}X_t = \mu$ then the process $Y_t = X_t - \mu$ is mean zero. A mean zero process is called a centered process. The function $C(t)$ is the *covariance* (sometimes also called autocovariance) 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 $C(t) = C(-t)$, $t \in \mathbb{R}$.

Let now X_t be a strictly stationary stochastic process with finite second moment. 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, in general. It is true, however, for Gaussian processes: since the first two moments of a Gaussian process are sufficient for a complete characterization of the process, a Gaussian stochastic process is strictly stationary if and only if it is weakly stationary.

Example 1.9. Let $Y_0, Y_1, \ldots$ be a sequence of independent, identically distributed random variables and consider the stochastic process $X_n = Y_n$. From Example 1.6 we know that this is a strictly stationary process, irrespective of whether Y_0 is such that $\mathbb{E}Y_0^2 < +\infty$. Assume now that $\mathbb{E}Y_0 = 0$ and $\mathbb{E}Y_0^2 = \sigma^2 < +\infty$. Then X_n is a second order stationary process with mean zero and correlation function $R(k) = \sigma^2 \delta_{k0}$. Notice that in this case we have no correlation between the values of the stochastic process at different times n and k.

Example 1.10. Let Z be a single random variable and consider the stochastic process $X_n = Z$, $n = 0, 1, 2, \ldots$. From Example 1.7 we know that this is a strictly stationary process irrespective of whether $\mathbb{E}|Z|^2 < +\infty$ or not. Assume now that $\mathbb{E}Z = 0$, $EZ^2 = \sigma^2$. Then X_n becomes a second order stationary process with $R(k) = \sigma^2$. Notice that in this case the values of our stochastic process at different times are strongly correlated.

We will see later in this chapter that for second order stationary processes, ergodicity is related to fast decay of correlations. In the first of the examples above, there was no correlation between our stochastic processes at different times and the stochastic process is ergodic. On the contrary, in our second example there is very strong correlation between the stochastic process at different times and this process is not ergodic.

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

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

Lemma 1.11. *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 $t \in \mathbb{R}$ and (without loss of generality) set $\mathbb{E}X_t = 0$. We calculate:

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

as $h \to 0$. Thus, continuity of $C(\cdot)$ at 0 implies continuity for all t.

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

$$\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \tag{1.4}$$

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 (1.4) 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 1.12. *A function $f(x) : \mathbb{R} \mapsto \mathbb{R}$ is called nonnegative definite if*

$$\sum_{i,j=1}^{n} f(t_i - t_j) c_i \bar{c}_j \geqslant 0 \tag{1.5}$$

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

Lemma 1.13. *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 1.14. *[Bochner] Let $C(t)$ be a continuous positive definite function. Then there exists a unique nonnegative measure ρ on $\mathbb{R}$ such that $\rho(\mathbb{R}) = C(0)$ and*

$$C(t) = \int_{\mathbb{R}} e^{i\omega t} \, \rho(d\omega) \quad \forall t \in \mathbb{R}. \tag{1.6}$$

Let X_t be a second order stationary process with autocorrelation function $C(t)$ whose Fourier transform is the measure $\rho(d\omega)$. The measure $\rho(d\omega)$ 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 $S(\omega)$, i.e. $\rho(d\omega) = S(\omega)d\omega$. The Fourier transform $S(\omega)$ of the covariance function is called the *spectral density* of the process:

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\omega} C(t) \, dt. \tag{1.7}$$

From (1.6) it follows that that the autocorrelation function of a mean zero, second order stationary process is given by the inverse Fourier transform of the spectral density:

$$C(t) = \int_{-\infty}^{\infty} e^{it\omega} S(\omega)\, d\omega. \tag{1.8}$$

The autocorrelation function of a second order stationary process enables us to associate a timescale to X_t, the *correlation time* τ_{cor}:

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

The slower the decay of the correlation function, the larger the correlation time is. Notice that when the correlations do not decay sufficiently fast so that $C(t)$ is not integrable, then the correlation time will be infinite.

Example 1.15. Consider a mean zero, second order stationary process with correlation function

$$C(t) = C(0)e^{-\alpha|t|} \tag{1.9}$$

where $\alpha > 0$. We will write $C(0) = \frac{D}{\alpha}$ where $D > 0$. The spectral density of this process is:

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

This function is called the *Cauchy* or the *Lorentz* distribution. The correlation time is (we have that $R(0) = D/\alpha$)

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

A real-valued Gaussian stationary process defined on $\mathbb{R}$ with correlation function given by (1.9) is called the stationary *Ornstein-Uhlenbeck process*. We will study this stochastic process in detail in later chapters. The Ornstein-Uhlenbeck process X_t can be used 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 (we assume that the Brownian particle starts at 0)

$$Z_t = \int_0^t Y_s\, ds, \tag{1.10}$$

The particle position Z_t is a mean zero Gaussian process. Set $\alpha = D = 1$. The covariance function of Z_t is

$$\mathbb{E}(Z_t Z_s) = 2\min(t,s) + e^{-\min(t,s)} + e^{-\max(t,s)} - e^{-|t-s|} - 1. \tag{1.11}$$

Ergodic properties of second-order stationary processes

Second order stationary processes have nice ergodic properties, provided that the correlation between values of the process at different times decays sufficiently fast. In this case, it is possible to show that we can calculate expectations by calculating time averages. An example of such a result is the following.

Proposition 1.16. *Let $\{X_t\}_{t\geqslant 0}$ be a second order stationary process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean μ and covariance $C(t)$, and assume that $C(t) \in L^1(0, +\infty)$. Then*

$$\lim_{T\to+\infty} \mathbb{E}\left|\frac{1}{T}\int_0^T X_s\,ds - \mu\right|^2 = 0. \tag{1.12}$$

For the proof of this result we will first need the following result, which is a property of symmetric functions.

Lemma 1.17. *Let $R(t)$ be an integrable symmetric function. Then*

$$\int_0^T\int_0^T C(t-s)\,dtds = 2\int_0^T (T-s)C(s)\,ds. \tag{1.13}$$

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

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

The integral becomes

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

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

Proof of Theorem 1.16. We use Lemma (1.17) to calculate:

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

using the dominated convergence theorem and the assumption $C(\cdot) \in L^1(0, +\infty)$. □Assume that $\mu = 0$ and define

$$D = \int_0^{+\infty} C(t)\, dt, \tag{1.14}$$

which, from our assumption on $C(t)$, is a finite quantity. [3] The above calculation suggests that, for $t \gg 1$, we have that

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

This implies that, at sufficiently long times, the mean square displacement of the integral of the ergodic second order stationary process X_t scales linearly in time, with proportionality coefficient $2D$. Let now X_t be the velocity of a (Brownian) particle. The particle position Z_t is given by (1.10). From our calculation above we conclude that

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

where

$$D = \int_0^{\infty} C(t)\, dt = \int_0^{\infty} \mathbb{E}(X_t X_0)\, dt \tag{1.15}$$

is the *diffusion coefficient*. Thus, one expects that at sufficiently long times and under appropriate assumptions on the correlation function, the time integral of a stationary process will approximate a Brownian motion with diffusion coefficient D. The diffusion coefficient is an example of a transport coefficient and (1.15) is an example of the Green-Kubo formula: a transport coefficient can be calculated in terms of the time integral of an appropriate autocorrelation function. In the case of the diffusion coefficient we need to calculate the integral of the velocity autocorrelation function. We will explore this topic in more detail in Chapter **??**.

Example 1.18. Consider the stochastic processes with an exponential correlation function from Example 1.15, and assume that this stochastic process describes the velocity of a Brownian particle. Since $C(t) \in L^1(0, +\infty)$ Proposition 1.16 applies. Furthermore, the diffusion coefficient of the Brownian particle is given by

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

Remark 1.19. *Let X_t be a strictly stationary process and let f be such that $\mathbb{E}(f(X_0))^2 < +\infty$. A calculation similar to the one that we did in the proof of Proposition 1.16 enables to conclude that*

$$\lim_{T\to+\infty} \frac{1}{T} \int_0^T f(X_s)\, ds = \mathbb{E}f(X_0), \tag{1.16}$$

in $L^2(\Omega)$. In this case the autocorrelation function of X_t is replaced by

$$C_f(t) = \mathbb{E}\big[\big(f(X_t) - \mathbb{E}f(X_0)\big)\big(f(X_0) - \mathbb{E}f(X_0)\big)\big].$$

[3] Notice however that we do not know whether it is nonzero. This requires a separate argument.

Setting $\overline{f} = f - \mathbb{E}_\pi f$ we have:

$$\lim_{T\to+\infty} T\,\mathrm{Var}_\pi\left(\frac{1}{T}\int_0^T f(X_t)\,dt\right) = 2\int_0^{+\infty} \mathbb{E}_\pi(\overline{f}(X_t)\overline{f}(X_0))\,dt \tag{1.17}$$

We calculate

$$\begin{aligned}
\mathrm{Var}_\pi\left(\frac{1}{T}\int_0^T f(X_t)\,dt\right) &= \mathbb{E}_\pi\left(\frac{1}{T}\int_0^T \overline{f}(X_t)\,dt\right)^2 \\
&= \frac{1}{T^2}\int_0^T\int_0^T \mathbb{E}_\pi\Big(\overline{f}(X_t)\overline{f}(X_s)\Big)\,dtds \\
&=: \frac{1}{T^2}\int_0^T\int_0^T R_{\overline{f}}(t,s)\,dtds \\
&= \frac{2}{T^2}\int_0^T (T-s)R_{\overline{f}}(s)\,ds \\
&= \frac{2}{T}\int_0^T \Big(1-\frac{s}{T}\Big)\,\mathbb{E}_\pi\Big(\overline{f}(X_s)\overline{f}(X_0)\Big)\,ds,
\end{aligned}$$

from which (1.16) follows.

1.3 Brownian Motion

The most important continuous-time stochastic process is Brownian motion. Brownian motion is a process with almost surely continuous paths and independent Gaussian increments. A process X_t has independent increments if for every sequence $t_0 < t_1 < \dots t_n$ the random variables

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

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

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

then the process X_t has stationary independent increments.

Definition 1.20. *A one dimensional standard* Brownian motion $W(t) : \mathbb{R}^+ \to \mathbb{R}$ *is a real valued stochastic process with a.s. continuous paths such that $W(0) = 0$, it has independent increments and for every $t > s \geqslant 0$, the increment $W(t) - W(s)$ has a Gaussian distribution with mean 0 and variance $t - s$, i.e. the density of the random variable $W(t) - W(s)$ is*

$$g(x;t,s) = \Big(2\pi(t-s)\Big)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2(t-s)}\right); \tag{1.18}$$

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

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

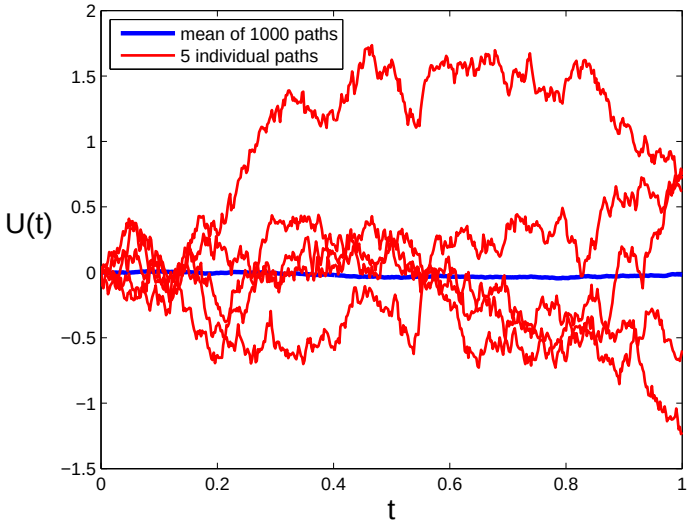


Figure 1.1: Brownian sample paths

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

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

Brownian motion is also referred to as the *Wiener process*. If Figure 1.1 we plot a few sample paths of Brownian motion.

As we have already mentioned, Brownian motion has almost surely continuous paths. More precisely, it has a continuous modification: consider two stochastic processes X_t and Y_t, $t \in T$, that are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The process Y_t is said to be a modification of X_t if $\mathbb{P}(X_t = Y_t) = 1$ for all $t \in T$. The fact that there is a continuous modification of Brownian motion follows from the following result which is due to Kolmogorov.

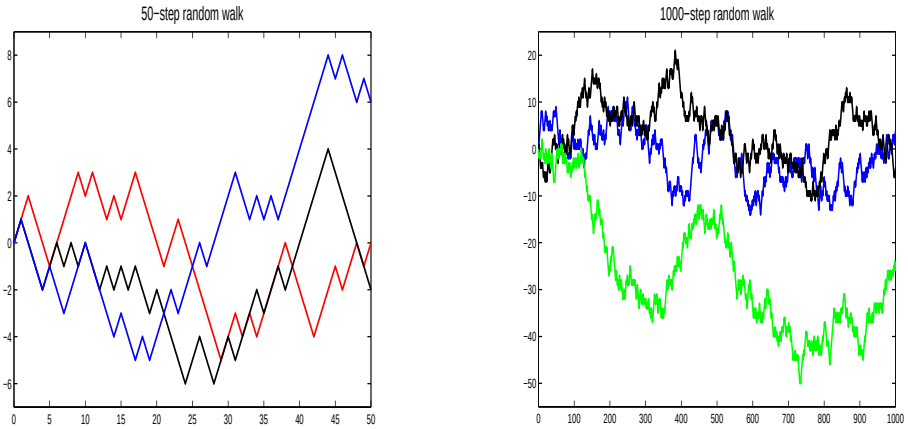
Theorem 1.21. *(Kolmogorov) Let X_t, $t \in [0, \infty)$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that there are positive constants α and β, and for each $T \geqslant 0$ there is a constant $C(T)$ such that*

$$\mathbb{E}|X_t - X_s|^\alpha \leqslant C(T)|t-s|^{1+\beta}, \quad 0 \leqslant s, t \leqslant T. \tag{1.19}$$

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

We can check that (1.19) holds for Brownian motion with $\alpha = 4$ and $\beta = 1$ using (1.18). It is possible to prove rigorously the existence of the Wiener process (Brownian motion):

Theorem 1.22. *(Wiener) There exists an almost surely continuous process W_t with independent increments such and $W_0 = 0$, such that for each $t \geqslant 0$ the random variable W_t is $\mathcal{N}(0, t)$. Furthermore, W_t is almost surely locally Hölder continuous with exponent α for any $\alpha \in (0, \frac{1}{2})$.*



a. $n = 50$ b. $n = 1000$

Figure 1.2: Sample paths of the random walk of length $n = 50$ and $n = 1000$.

Notice that Brownian paths are not differentiable.

We can construct Brownian motion through the limit of an appropriately rescaled random walk: let $X_1, X_2, \ldots$ be iid random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean 0 and variance 1. Define the discrete time stochastic process S_n with $S_0 = 0$, $S_n = \sum_{j=1} X_j$, $n \geqslant 1$. Define now a continuous time stochastic process with continuous paths as the linearly interpolated, appropriately rescaled random walk:

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

where $[\cdot]$ denotes the integer part of a number. Then W_t^n converges weakly, as $n \to +\infty$ to a one dimensional standard Brownian motion. See Figure 1.2.

An alternative definition of the one dimensional standard Brownian motion is that of a Gaussian stochastic process on a probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ with continuous paths for almost all $\omega \in \Omega$, and finite dimensional distributions with zero mean and covariance $\mathbb{E}(W_{t_i} W_{t_j}) = \min(t_i, t_j)$. One can then show that Definition 1.20 follows from the above definition.

For the d-dimensional Brownian motion we have (see (B.7) and (B.8))

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

and

$$\mathbb{E}\Big((W(t) - W(s)) \otimes (W(t) - W(s))\Big) = (t - s)I, \tag{1.20}$$

where I denotes the identity matrix. Moreover,

$$\mathbb{E}\Big(W(t) \otimes W(s)\Big) = \min(t, s)I. \tag{1.21}$$

Although Brownian motion has stationary increments, it is not a stationary process itself Brownian motion

itself. The probability density of the one dimensional Brownian motion is

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

We can easily calculate all moments:

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

In particular, the mean square displacement of Brownian motion grows linearly in time.

Brownian motion is invariant under various transformations in time.

Proposition 1.23. *Let W_t denote a standard Brownian motion in $\mathbb{R}$. Then, W_t has the following properties:*

i. (Rescaling). For each $c > 0$ define $X_t = \frac{1}{\sqrt{c}} W(ct)$. Then $(X_t,\, t \geqslant 0) = (W_t,\, t \geqslant 0)$ in law.

ii. (Shifting). For each $c > 0$ $W_{c+t} - W_c, t \geqslant 0$ is a Brownian motion which is independent of W_u, $u \in [0, c]$.

iii. (Time reversal). Define $X_t = W_{1-t} - W_1$, $t \in [0,1]$. Then $(X_t,\, t \in [0,1]) = (W_t,\, t \in [0,1])$ in law.

iv. (Inversion). Let X_t, $t \geqslant 0$ defined by $X_0 = 0$, $X_t = tW(1/t)$. Then $(X_t,\, t \geqslant 0) = (W_t,\, t \geqslant 0)$ in law.

The equivalence in the above result holds in law and not in a pathwise sense. The proof of this proposition is left as an exercise.

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

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

The mean and variance of X_t are

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

Notice that X_t satisfies the equation

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

This is an example of a *stochastic differential equation*. We will study stochastic differential equations in Chapters 3 and **??**.

1.4 Examples of Stochastic Processes

We present now a few examples of stochastic processes that appear frequently in applications.

The Ornstein-Uhlenbeck process

The stationary Ornstein-Uhlenbeck process that was introduced earlier in this chapter can be defined through the Brownian motion via a time change.

Lemma 1.24. *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 stationary process with mean 0 and correlation function

$$R(t) = e^{-|t|}. \tag{1.22}$$

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

Lemma 1.25. *Let $X(t)$ be a Gaussian stochastic process and let $Y(t) = X(f(t))$ where $f(t)$ is a strictly increasing function. Then $Y(t)$ is also a Gaussian process.*

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

$$\{Y(t_1), Y(t_2), \dots Y(t_N)\} \tag{1.23}$$

is a multivariate Gaussian random variable. Since $f(t)$ is strictly increasing, it is invertible and hence, there exist s_i, $i = 1, \dots N$ such that $s_i = f^{-1}(t_i)$. Thus, the random vector (1.23) can be rewritten as

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

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

Proof of Lemma 1.24. The fact that $V(t)$ is a mean zero process follows immediately from the fact that $W(t)$ is mean zero. To show that the correlation function of $V(t)$ is given by (1.22), we calculate

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

The Gaussianity of the process $V(t)$ follows from Lemma 1.25 (notice that the transformation that gives $V(t)$ in terms of $W(t)$ is invertible and we can write $W(s) = s^{1/2}V(\frac{1}{2}\ln(s))$). □

Brownian Bridge

We can modify Brownian motion so that the resulting processes is fixed at both ends. Let $W(t)$ be a standard one dimensional Brownian motion. We define the Brownian bridge (from 0 to 0) to be the process

$$B_t = W_t - tW_1, \quad t \in [0, 1]. \tag{1.24}$$

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

$$\mathbb{E}B_t = 0, \quad \mathbb{E}(B_tB_s) = \min(s, t) - st, \quad s, t \in [0, 1]. \tag{1.25}$$

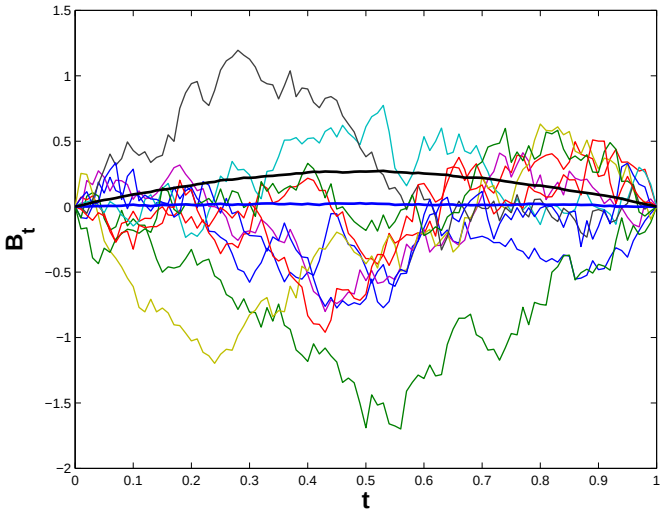


Figure 1.3: Sample paths and first (blue curve) and second (black curve) moment of the Brownian bridge.

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

$$B_t = (1-t)W\left(\frac{t}{1-t}\right), \quad t \in [0,1). \tag{1.26}$$

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

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

We can use the algorithm for simulating Gaussian processes to generate paths of the Brownian bridge process and to calculate moments. In Figure 1.3 we plot a few sample paths and the first and second moments of Brownian bridge.

Fractional Brownian Motion

The fractional Brownian motion is a one-parameter family of Gaussian processes whose increments are correlated.

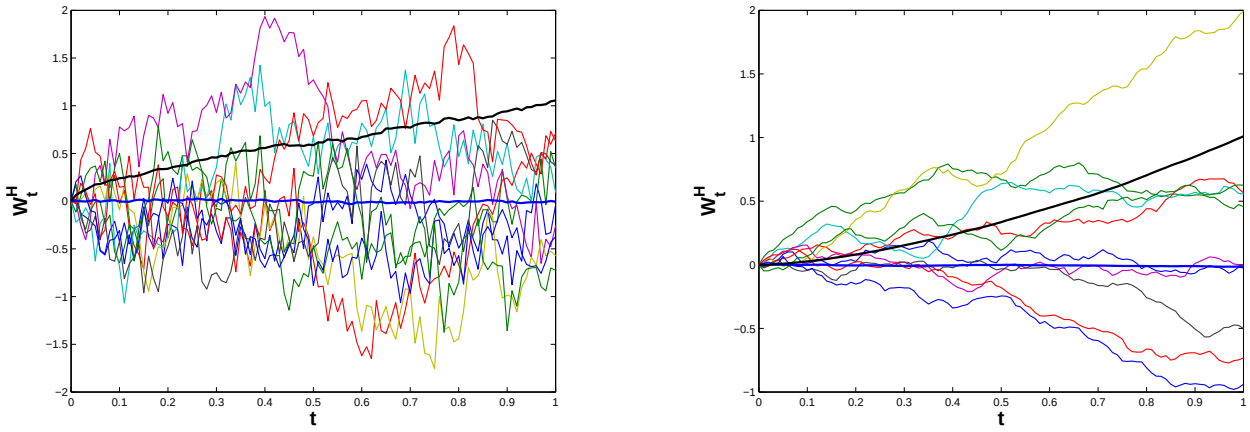
Definition 1.26. *A (normalized) fractional Brownian motion W_t^H, $t \geqslant 0$ with Hurst parameter $H \in (0,1)$ is a centered Gaussian process with continuous sample paths whose covariance is given by*

$$\mathbb{E}(W_t^H W_s^H) = \frac{1}{2}\big(s^{2H} + t^{2H} - |t-s|^{2H}\big). \tag{1.27}$$

The Hurst exponent controls the correlations between the increments of fractional Brownian motion as well as the regularity of the paths: they become smoother as H increases.

Some of the basic properties of fractional Brownian motion are summarized in the following proposition.

Proposition 1.27. *Fractional Brownian motion has the following properties.*



a. $H = 0.3$ b. $H = 0.8$

Figure 1.4: Sample paths of fractional Brownian motion for Hurst exponent $H = 0.3$ and $H = 0.8$ and first (blue curve) and second (black curve) moment.

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

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

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

iv. It has the following self similarity property

$$(W_{\alpha t}^H,\, t \geqslant 0) = (\alpha^H W_t^H,\, t \geqslant 0), \quad \alpha > 0, \tag{1.28}$$

where the equivalence is in law.

The proof of these properties is left as an exercise. In Figure 1.4 we present sample plots and the first two moments of the factional Brownian motion for $H = 0.3$ and $H = 0.8$. As expected, for larger values of the Hurst exponent the sample paths are more regular.

1.5 The Karhunen-Loéve Expansion

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

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

where

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

The convergence is in $L^2(\mathcal{D})$:

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

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

$$\mathbb{E}X_t^2 < +\infty, \quad \mathbb{E}X_t = 0, \quad \lim_{h\to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0. \tag{1.29}$$

For simplicity we will take $T = [0, 1]$. Let $R(t, s) = \mathbb{E}(X_t X_s)$ be the autocorrelation function. Notice that from (1.29) it follows that $R(t, s)$ is continuous in both t and s; see Exercise 20.

Let us assume an expansion of the form

$$X_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) e_n(t), \quad t \in [0, 1] \tag{1.30}$$

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

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

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

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

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

Assuming that an expansion of the form (1.30) exists, we can calculate

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

Consequently, in order to the expansion (1.30) to be valid we need

$$R(t, s) = \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s). \tag{1.31}$$

From equation (1.31) it follows that

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

Consequently, in order for the expansion (1.30) to be valid, $\{\lambda_n,\, e_n(t)\}_{n=1}^{\infty}$ have to be the eigenvalues and eigenfunctions of the integral operator whose kernel is the correlation function of X_t:

$$\int_0^1 R(t,s)e_n(s)\,ds = \lambda_n e_n(t). \tag{1.32}$$

To prove the expansion (1.30) we need to study the eigenvalue problem for the integral operator

$$\mathcal{R}f := \int_0^1 R(t,s)f(s)\,ds. \tag{1.33}$$

We consider it as an operator from $L^2[0,1]$ to $L^2[0,1]$. We can show that this operator is selfadjoint and nonnegative in $L^2(0,1)$:

$$\langle \mathcal{R}f, h\rangle = \langle f, \mathcal{R}h\rangle \quad \text{and} \quad \langle \mathcal{R}f, f\rangle \geqslant 0 \quad \forall\, f,h \in L^2(0,1),$$

where $\langle\cdot,\cdot\rangle$ denotes the $L^2(0,1)$-inner product. It follows that all its eigenvalues are real and nonnegative. Furthermore, it is a compact operator (if $\{\phi_n\}_{n=1}^{\infty}$ is a bounded sequence in $L^2(0,1)$, then $\{\mathcal{R}\phi_n\}_{n=1}^{\infty}$ has a convergent subsequence). The spectral theorem for compact, selfadjoint operators can be used to deduce that $\mathcal{R}$ has a countable sequence of eigenvalues tending to 0. Furthermore, for every $f \in L^2(0,1)$ we can write

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

where $\mathcal{R}f_0 = 0$ and $\{e_n(t)\}$ are the eigenfunctions of the operator $\mathcal{R}$ corresponding to nonzero eigenvalues and where the convergence is in L^2. Finally, Mercer's Theorem states that for $R(t,s)$ continuous on $[0,1]\times[0,1]$, the expansion (1.31) is valid, where the series converges absolutely and uniformly.

Now we are ready to prove (1.30).

Theorem 1.28. *(Karhunen-Loéve). Let $\{X_t,\, t\in[0,1]\}$ be an L^2 process with zero mean and continuous correlation function $R(t,s)$. Let $\{\lambda_n,\, e_n(t)\}_{n=1}^{\infty}$ be the eigenvalues and eigenfunctions of the operator $\mathcal{R}$ defined in* (1.33)*. Then*

$$X_t = \sum_{n=1}^{\infty} \xi_n e_n(t), \quad t\in[0,1], \tag{1.34}$$

where

$$\xi_n = \int_0^1 X_t e_n(t)\,dt, \quad \mathbb{E}\xi_n = 0, \quad \mathbb{E}(\xi_n\xi_m) = \lambda\delta_{nm}. \tag{1.35}$$

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

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

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

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

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

by Mercer's theorem. □

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

$$X_t = \sum_{k=1}^{+\infty} \sqrt{\lambda_k}\xi_k e_k(t), \tag{1.36}$$

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

Example 1.29. The Karhunen-Loéve Expansion for Brownian Motion. The correlation function of Brownian motion is $R(t,s) = \min(t,s)$. The eigenvalue problem $\mathcal{R}\psi_n = \lambda_n\psi_n$ becomes

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

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

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

We differentiate this equation once:

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

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

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

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

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

We can calculate the eigenvalues and (normalized) eigenfunctions are

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

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

$$W_t = \sqrt{2}\sum_{n=1}^{\infty}\xi_n\frac{2}{(2n-1)\pi}\sin\left(\frac{1}{2}(2n-1)\pi t\right). \tag{1.37}$$

1.6 Discussion and Bibliography

The material presented in this chapter is very standard and can be found in any any textbook on stochastic processes. Consult, for example [48, 47, 49, 32]. The proof of Bochner's theorem 1.14 can be found in [50], where additional material on stationary processes can be found. See also [48].

The Ornstein-Uhlenbeck process was introduced by Ornstein and Uhlenbeck in 1930 as a model for the velocity of a Brownian particle [101]. An early reference on the derivation of formulas of the form (1.15) is [99].

Gaussian processes are studied in [1]. Simulation algorithms for Gaussian processes are presented in [6]. Fractional Brownian motion was introduced in [64].

The spectral theorem for compact, selfadjoint operators that we used in the proof of the Karhunen-Loéve expansion can be found in [84]. The Karhunen-Loéve expansion can be used to generate random fields, i.e. a collection of random variables that are parameterized by a spatial (rather than temporal) parameter x. See [29]. The Karhunen-Loéve expansion is useful in the development of numerical algorithms for partial differential equations with random coefficients. See [92].

We can use the Karhunen-Loéve expansion in order to study the L^2-regularity of stochastic processes. First, let R be a compact, symmetric positive definite operator on $L^2(0,1)$ with eigenvalues and normalized eigenfunctions $\{\lambda_k,\, e_k(x)\}_{k=1}^{+\infty}$ and consider a function $f \in L^2(0,1)$ with $\int_0^1 f(s)\,ds = 0$. We can define the one parameter family of Hilbert spaces H^α through the norm

$$\|f\|_\alpha^2 = \|R^{-\alpha} f\|_{L^2}^2 = \sum_k |f_k|^2 \lambda^{-\alpha}.$$

The inner product can be obtained through polarization. This norm enables us to measure the regularity of the function $f(t)$.[4] Let X_t be a mean zero second order (i.e. with finite second moment) process with continuous autocorrelation function. Define the space $\mathcal{H}^\alpha := L^2((\Omega, P), H^\alpha(0,1))$ with (semi)norm

$$\|X_t\|_\alpha^2 = \mathbb{E}\|X_t\|_{H^\alpha}^2 = \sum_k |\lambda_k|^{1-\alpha}. \tag{1.38}$$

Notice that the regularity of the stochastic process X_t depends on the decay of the eigenvalues of the integral operator $\mathcal{R}\cdot := \int_0^1 R(t,s)\cdot\, ds$.

As an example, consider the L^2-regularity of Brownian motion. From Example 1.29 we know that $\lambda_k \sim k^{-2}$. Consequently, from (1.38) we get that, in order for W_t to be an element of the space $\mathcal{H}^\alpha$, we need that

$$\sum_k |\lambda_k|^{-2(1-\alpha)} < +\infty,$$

from which we obtain that $\alpha < 1/2$. This is consistent with the Hölder continuity of Brownian motion from Theorem 1.22. [5]

1.7 Exercises

1. Let $Y_0, Y_1, \ldots$ be a sequence of independent, identically distributed random variables and consider the stochastic process $X_n = Y_n$.

   (a) Show that X_n is a strictly stationary process.

   (b) Assume that $\mathbb{E}Y_0 = \mu < +\infty$ and $\mathbb{E}Y_0^2 = \sigma^2 < +\infty$. Show that

   $$\lim_{N\to+\infty} \mathbb{E}\left|\frac{1}{N}\sum_{j=0}^{N-1} X_j - \mu\right| = 0.$$

   (c) Let f be such that $\mathbb{E}f^2(Y_0) < +\infty$. Show that

   $$\lim_{N\to+\infty} \mathbb{E}\left|\frac{1}{N}\sum_{j=0}^{N-1} f(X_j) - f(Y_0)\right| = 0.$$

[4] Think of R as being the inverse of the Laplacian with periodic boundary conditions. In this case H^α coincides with the standard fractional Sobolev space.

[5] Notice, however, that Wiener's theorem refers to a.s. Hölder continuity, whereas the calculation presented in this section is about L^2-continuity.

2. Let Z be a random variable and define the stochastic process $X_n = Z,\ n = 0, 1, 2, \dots$. Show that X_n is a strictly stationary process.

3. Let $A_0, A_1, \dots A_m$ and $B_0, B_1, \dots B_m$ be uncorrelated random variables with mean zero and variances $\mathbb{E}A_i^2 = \sigma_i^2$, $\mathbb{E}B_i^2 = \sigma_i^2,\ i = 1, \dots m$. Let $\omega_0, \omega_1, \dots \omega_m \in [0, \pi]$ be distinct frequencies and define, for $n = 0, \pm 1, \pm 2, \dots$, the stochastic process

$$X_n = \sum_{k=0}^{m} \Big(A_k \cos(n\omega_k) + B_k \sin(n\omega_k) \Big).$$

Calculate the mean and the covariance of X_n. Show that it is a weakly stationary process.

4. Let $\{\xi_n\ :\ n = 0, \pm 1, \pm 2, \dots\}$ be uncorrelated random variables with $\mathbb{E}\xi_n = \mu$, $\mathbb{E}(\xi_n - \mu)^2 = \sigma^2,\ n = 0, \pm 1, \pm 2, \dots$. Let $a_1, a_2, \dots$ be arbitrary real numbers and consider the stochastic process

$$X_n = a_1\xi_n + a_2\xi_{n-1} + \dots a_m\xi_{n-m+1}.$$

   (a) Calculate the mean, variance and the covariance function of X_n. Show that it is a weakly stationary process.

   (b) Set $a_k = 1/\sqrt{m}$ for $k = 1, \dots m$. Calculate the covariance function and study the cases $m = 1$ and $m \to +\infty$.

5. Let $W(t)$ be a standard one dimensional Brownian motion. Calculate the following expectations.

   (a) $\mathbb{E}e^{iW(t)}$.

   (b) $\mathbb{E}e^{i(W(t)+W(s))},\ t, s, \in (0, +\infty)$.

   (c) $\mathbb{E}(\sum_{i=1}^{n} c_i W(t_i))^2$, where $c_i \in \mathbb{R},\ i = 1, \dots n$ and $t_i \in (0, +\infty),\ i = 1, \dots n$.

   (d) $\mathbb{E}e^{\left[i\left(\sum_{i=1}^{n} c_i W(t_i)\right)\right]}$, where $c_i \in \mathbb{R},\ i = 1, \dots n$ and $t_i \in (0, +\infty),\ i = 1, \dots n$.

6. Let W_t be a standard one dimensional Brownian motion and define

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

   (a) Show that B_t is a Gaussian process with

$$\mathbb{E}B_t = 0, \quad \mathbb{E}(B_t B_s) = \min(t, s) - ts.$$

   (b) Show that, for $t \in [0, 1)$ an equivalent definition of B_t is through the formula

$$B_t = (1 - t)W\left(\frac{t}{1 - t}\right).$$

   (c) Calculate the distribution function of B_t.

7. Let X_t be a mean-zero second order stationary process with autocorrelation function

$$R(t) = \sum_{j=1}^{N} \frac{\lambda_j^2}{\alpha_j} e^{-\alpha_j |t|},$$

where $\{\alpha_j,\, \lambda_j\}_{j=1}^N$ are positive real numbers.

   (a) Calculate the spectral density and the correlaction time of this process.

   (b) Show that the assumptions of Theorem 1.16 are satisfied and use the argument presented in Section 1.2 (i.e. the Green-Kubo formula) to calculate the diffusion coefficient of the process $Z_t = \int_0^t X_s\, ds$.

   (c) Under what assumptions on the coefficients $\{\alpha_j,\, \lambda_j\}_{j=1}^N$ can you study the above questions in the limit $N \to +\infty$?

8. Show that the position of a Brownian particle whose velocity is described by the stationary Ornstein-Uhlenbeck process, Equation (1.10) is a mean zero Gaussian stochastic process and calculate the covariance function.

9. Let $a_1, \ldots a_n$ and $s_1, \ldots s_n$ be positive real numbers. Calculate the mean and variance of the random variable

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

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

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

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

11. Let W_t be a one dimensional Brownian motion and let $\mu, \sigma > 0$ and define

$$S_t = e^{t\mu + \sigma W_t}.$$

   (a) Calculate the mean and the variance of S_t.

   (b) Calculate the probability density function of S_t.

12. Prove proposition 1.23.

13. Use Lemma 1.24 to calculate the distribution function of the stationary Ornstein-Uhlenbeck process.

14. Calculate the mean and the correlation function of the integral of a standard Brownian motion

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

15. Show that the process

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

is second order stationary.

16. Let $V_t = e^{-t}W(e^{2t})$ be the stationary Ornstein-Uhlenbeck process. Give the definition and study the main properties of the Ornstein-Uhlenbeck bridge.

17. The autocorrelation function of the velocity $Y(t)$ a Brownian particle moving in a harmonic potential $V(x) = \frac{1}{2}\omega_0^2 x^2$ is

$$R(t) = e^{-\gamma|t|}\Big(cos(\delta|t|) - \frac{1}{\delta}\sin(\delta|t|)\Big),$$

where γ is the friction coefficient and $\delta = \sqrt{\omega_0^2 - \gamma^2}$.

(a) Calculate the spectral density of $Y(t)$.

(b) Calculate the mean square displacement $\mathbb{E}(X(t))^2$ of the position of the Brownian particle $X(t) = \int_0^t Y(s)\,ds$. Study the limit $t \to +\infty$.

18. Show the scaling property (1.28) of the fractional Brownian motion.

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

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

Use Theorem (1.21) to show that there does not exist a continuous modification of this process.

20. Show that the correlation function of a process X_t satisfying (1.29) is continuous in both t and s.

21. Let X_t be a stochastic process satisfying (1.29) and $R(t,s)$ its correlation function. Show that the integral operator $\mathcal{R} : L^2[0,1] \mapsto L^2[0,1]$ defined in (1.33),

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

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

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

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

Let $\mathcal{R} : L^2[0,1] \mapsto L^2[0,1]$ be the operator defined in (1.33) with $R(t,s)$ being continuous both in t and s. Show that it is a Hilbert-Schmidt operator.

23. Let X_t a mean zero second order stationary process defined in the interval $[0,T]$ with continuous covariance $R(t)$ and let $\{\lambda_n\}_{n=1}^{+\infty}$ be the eigenvalues of the covariance operator. Show that

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

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

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

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

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

Define the process

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

where $f(t)$ is a continuous function and $\tau(t)$ a continuous, nondecreasing function with $\tau(0) = 0$, $\tau(S) = T$. Find the Karhunen-Loéve expansion of $Y(t)$, in an appropriate weighted L^2 space, in terms of the KL expansion of X_t. Use this in order to calculate the KL expansion of the Ornstein-Uhlenbeck process.

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

28. Use the Karhunen-Loeve expansion to generate paths of

    (a) the Brownian motion on $[0,1]$;

    (b) the Brownian bridge on $[0,1]$;

    (c) the Ornstein-Uhlenbeck on $[0,1]$.

    Study computationally the convergence of the Karhunen-Loéve expansion for these processes. How many terms do you need to keep in the expansion in order to calculate accurate statistics of these processes? How does the computational cost compare with that of the standard algorithm for simulating Gaussian stochastic processes?

29. (See [29].) Consider the Gaussian random field $X(x)$ in $\mathbb{R}$ with covariance function

$$\gamma(x,y) = e^{-a|x-y|} \tag{1.39}$$

where $a > 0$.

    (a) Simulate this field: generate samples and calculate the first four moments.

    (b) Consider $X(x)$ for $x \in [-L,L]$. Calculate analytically the eigenvalues and eigenfunctions of the integral operator $\mathcal{K}$ with kernel $\gamma(x,y)$,

$$\mathcal{K}f(x) = \int_{-L}^{L} \gamma(x,y) f(y)\, dy.$$

    Use this in order to obtain the Karhunen-Loéve expansion for X. Plot the first five eigenfunctions when $a = 1$, $L = -0.5$. Investigate (either analytically or by means of numerical experiments) the accuracy of the KL expansion as a function of the number of modes kept.

(c) Develop a numerical method for calculating the first few eigenvalues and eigenfunctions of $\mathcal{K}$ with $a = 1$, $L = -0.5$. Use the numerically calculated eigenvalues and eigenfunctions to simulate $X(x)$ using the KL expansion. Compare with the analytical results and comment on the accuracy of the calculation of the eigenvalues and eigenfunctions and on the computational cost.

Chapter 2

Diffusion Processes

In this chapter we study some of the basic properties of Markov stochastic processes and, in particular, of diffusion processes. In Section 2.1 we present various examples of Markov processes, in discrete and continuous time. In Section 2.2 we give the precise definition of a Markov process. In Section 2.2 we derive the Chapman-Kolmogorov equation, the fundamental equation in the theory of Markov processes. In Section 2.3 we introduce the concept of the generator of a Markov process. In Section 2.4 we study ergodic Markov processes. In Section 2.5 we introduce diffusion processes and we derive the forward and backward Kolmogorov equations. Discussion and bibliographical remarks are presented in Section 2.6 and exercises can be found in Section 2.7.

2.1 Examples of Markov processes

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

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

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

Let $i_1, i_2, \dots$ be a sequence of integers. Then, for all integers n and m we have that[1]

$$\mathbb{P}(X_{n+m} = i_{n+m} | X_1 = i_1, \dots X_n = i_n) = \mathbb{P}(X_{n+m} = i_{n+m} | X_n = i_n). \tag{2.1}$$

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

[1] In fact, it is sufficient to take $m = 1$ in (2.1). See Exercise 1.

The random walk is an example of a *discrete time Markov chain*: We will say that a stochastic process $\{S_n; n \in \mathbb{N}\}$ with state space $S = \mathbb{Z}$ is a discrete time Markov chain provided that the Markov property (2.1) is satisfied.

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

$$\mathbb{P}(X_{t+h} = i_{t+h} | \{X_s,\, s \leqslant t\}) = \mathbb{P}(X_{t+h} = i_{t+h} | X_t), \tag{2.2}$$

for all $h \geqslant 0$. A continuous-time, discrete state space Markov process is called a continuous-time Markov chain. A standard example of a continuous-time Markov chain is the Poisson process of rate λ with

$$\mathbb{P}(N_{t+h} = j | N_t = i) = \begin{cases} 0 & \text{if } j < i, \\ \frac{e^{-\lambda s}(\lambda s)^{j-i}}{(j-i)!}, & \text{if } j \geqslant i. \end{cases} \tag{2.3}$$

Similarly, we can define a continuous-time Markov process with state space is $\mathbb{R}$, as a stochastic process whose future depends on its present state and not on how it got there:

$$\mathbb{P}(X_{t+h} \in \Gamma | \{X_s,\, s \leqslant t\}) = \mathbb{P}(X_{t+h} \in \Gamma | X_t) \tag{2.4}$$

for all Borel sets Γ. In this book we will consider continuous-time Markov processes for which a *conditional probability density* exists:

$$\mathbb{P}(X_{t+h} \in \Gamma | X_t = x) = \int_\Gamma p(y, t+h|x, t)\, dy. \tag{2.5}$$

Example 2.1. The Brownian motion is a Markov process with conditional probability density given by the following formula

$$\mathbb{P}(W_{t+h} \in \Gamma | W_t = x) = \int_\Gamma \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{|x-y|^2}{2h}\right) dy. \tag{2.6}$$

The Markov property of Brownian motion follows from the fact that it has independent increments.

Example 2.2. The stationary Ornstein-Uhlenbeck process $V_t = e^{-t} W(e^{2t})$ is a Markov process with conditional probability density

$$p(y, t|x, s) = \frac{1}{\sqrt{2\pi(1 - e^{-2(t-s)})}} \exp\left(-\frac{|y - xe^{-(t-s)}|^2}{2(1 - e^{-2(t-s)})}\right). \tag{2.7}$$

To prove (2.7) we use the formula for the distribution function of the Brownian motion to calculate, for $t > s$,

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

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

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

The Markov property enables us to obtain an evolution equation for the *transition probability* for a discrete-time or continuous-time Markov chain

$$\mathbb{P}(X_{n+1} = i_{n+1}|X_n = i_n), \quad \mathbb{P}(X_{t+h} = i_{t+h}|X_t = i_t), \tag{2.8}$$

or for the transition probability density defined in (2.5). This equation is the *Chapman-Kolmogorov equation*. Using this equation we can study the evolution of a Markov process.

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

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

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

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

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

$$p_{ij}(m+n) = \sum_k p_{ik}(m)p_{kj}(n). \tag{2.9}$$

Indeed, let $\mu_i^{(n)} := \mathbb{P}(X_n = i)$. The (possibly infinite dimensional) vector $\mu^{(n)}$ determines the state of the Markov chain at time n. From the Chapman-Kolmogorov equation we can obtain a formula for the evolution of the vector $\mu^{(n)}$

$$\mu^{(n)} = \mu^{(0)}P^n, \tag{2.10}$$

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

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

Consider now a continuous-time Markov chain with transition probability

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

If the chain is homogeneous, then

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

In particular,

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

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

$$\frac{dp_{ij}}{dt} = \sum_k p_{ik}(t) g_{kj}, \tag{2.11}$$

where the matrix G is called the *generator* of the Markov chain that is defined as

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

with P_t denoting the matrix $\{p_{ij}(t)\}$. Equation (2.11) can also be written in matrix form:

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

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

$$\mu_t = \mu_0 P_t.$$

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

Consider now the case a continuous-time Markov process with continuous state space and with continuous paths. As we have seen in Example 2.1 the Brownian motion is such a process. The conditional probability density of the Brownian motion (2.6) is the fundamental solution (Green's function) of the diffusion equation:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}, \quad \lim_{t \to s} p(y, t|x, s) = \delta(y - x). \tag{2.12}$$

Similarly, the conditional distribution of the Ornstein-Uhlenbeck process satisfies the initial value problem

$$\frac{\partial p}{\partial t} = \frac{\partial (yp)}{\partial y} + \frac{1}{2}\frac{\partial^2 p}{\partial y^2}, \quad \lim_{t \to s} p(y, t|x, s) = \delta(y - x). \tag{2.13}$$

The Brownian motion and the OU process are examples of a *diffusion process*: a continuous-time Markov process with continuous paths. A precise definition will be given in Section 2.5, where we will also derive evolution equations for the conditional probability density $p(y, t|x, s)$ of an arbitrary diffusion process, the forward Kolmogorov (Fokker-Planck) (2.54) and backward Kolmogorov (2.47) equations.

2.2 Markov Processes and the Chapman-Kolmogorov equation

In Section 2.1 we gave the definition of Markov process whose time is either discrete or continuous, and whose state space is countable. We also gave several examples of Markov chains as well as of processes whose state space is the real line. In this section we give the precise definition of a Markov process with $t \in \mathbb{R}_+$ and with state space is $\mathbb{R}^d$. We also introduce the Chapman-Kolmogorov equation.

In order to give the definition of a Markov process we need to use the conditional expectation of the stochastic process conditioned on all past values. We can encode all past information about a stochastic process into an appropriate collection of σ-algebras. Let $(\Omega, \mathcal{F}, \mu)$ denote a probability space and consider a stochastic process $X = X_t(\omega)$ with $t \in \mathbb{R}_+$ and state space $(\mathbb{R}^d, \mathcal{B})$ where $\mathcal{B}$ denotes the Borel σ-algebra. We define the σ-algebra generated by $\{X_t,\, t \in \mathbb{R}_+\}$, denoted by $\sigma(X_t,\, t \in \mathbb{R}_+)$, to be the smallest σ-algebra such that the family of mappings $\{X_t,\, t \in \mathbb{R}_+\}$ is a stochastic process with sample space $(\Omega, \sigma(X_t,\, t \in \mathbb{R}_+))$ and state space $(\mathbb{R}^d, \mathcal{B})$.[2] In other words, the σ-algebra generated by X_t is the smallest σ-algebra such that X_t is a measurable function (random variable) with respect to it.

We define now a filtration on $(\Omega, \mathcal{F})$ to be a nondecreasing family $\{\mathcal{F}_t, t \in \mathbb{R}_+\}$ of sub-σ-algebras of $\mathcal{F}$:

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \quad \text{for } s \leqslant t.$$

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

$$\mathcal{F}_t^X := \sigma\left(X_s; s \leqslant t\right). \tag{2.14}$$

Now we are ready to give the definition of a Markov process.

Definition 2.3. *Let X_t be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with values in $\mathbb{R}^d$ and let $\mathcal{F}_t^X$ be the filtration generated by $\{X_t; t \in \mathbb{R}_+\}$. Then $\{X_t; t \in \mathbb{R}_+\}$ is a Markov process provided that*

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma | X_s) \tag{2.15}$$

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

We remark that the filtration $\mathcal{F}_t^X$ is generated by events of the form $\{\omega | X_{t_1} \in \Gamma_1,\, X_{t_2} \in \Gamma_2,\, \dots X_{t_n} \in \Gamma_n,\}$ with $0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant t$ and $\Gamma_i \in \mathcal{B}(\mathbb{R}^d)$. The definition of a Markov process is thus equivalent to the hierarchy of equations

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

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

We also remark that it is sometimes possible to describe a non-Markovian process X_t in terms of a Markovian process Y_t in a higher dimensional state space. The additional variables that we introduce account for the memory in the X_t. This is possible when the non-Markovian process has finite memory that can be represented by a finite number of additional degrees of freedom. We will use this approach in Chapter **??** when we derive stochastic differential equations from deterministic dynamical systems with random initial conditions, see Definition **??**.

As an example, consider a Brownian particle whose velocity is described by the stationary Ornstein-Uhlenbeck process $Y_t = e^{-t} W(e^{2t})$, see (1.10) and (1.11). The particle position is given by the integral of the Ornstein-Uhlenbeck process

$$X_t = X_0 + \int_0^t Y_s \, ds.$$

[2] In later chapters we will also consider Markov processes with state space being the a subset of $\mathbb{R}^d$, for example the unit torus.

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

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

The Chapman-Kolmogorov Equation

With every continuous-time Markov process X_t[3] defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $(\mathbb{R}^d, \mathcal{B})$ we can associate the *transition function*

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

for all $t, s \in \mathbb{R}_+$ with $t \geqslant s$ and all $\Gamma \in \mathcal{B}(\mathbb{R}^d)$. It is a function of 4 arguments, the initial time s and position X_s and the final time t and the set Γ. The transition function $P(t, \Gamma|x, s)$ is, for fixed t, x s, a probability measure on $\mathbb{R}^d$ with $P(t, \mathbb{R}^d|x, s) = 1$; it is $\mathcal{B}(\mathbb{R}^d)$-measurable in x, for fixed t, s, Γ and satisfies the Chapman-Kolmogorov equation

$$P(\Gamma, t|x, s) = \int_{\mathbb{R}^d} P(\Gamma, t|y, u)P(dy, u|x, s). \tag{2.16}$$

for all $x \in \mathbb{R}^d$, $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ and $s, u, t \in \mathbb{R}_+$ with $s \leqslant u \leqslant t$. 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 derivation of the Chapman-Kolmogorov equation is based on the Markovian assumption and on properties of conditional probability. We can formally derive the Chapman-Kolmogorov equation as follows: We use the Markov property, together with Equations (B.9) and (B.10) from Appendix B and the fact that $s < u \Rightarrow \mathcal{F}_s^X \subset \mathcal{F}_u^X$ to calculate:

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

where $I_\Gamma(\cdot)$ denotes the indicator function of the set Γ. In words, the Chapman-Kolmogorov equation tells us that for a Markov process the transition from x at time s to the set Γ at time t can be done in two steps: first the system moves from x to y at some intermediate time u. Then it moves from y to Γ at time t. In

[3] We always take $t \in \mathbb{R}_+$.

order to calculate the probability for the transition from x at time s to Γ at time t we need to sum (integrate) the transitions from all possible intermediate states y.

The transition function and the initial distribution of X_t are sufficient to uniquely determine a Markov process. In fact, a process X_t is a Markov process with respect to its filtration $\mathcal{F}_t^X$ defined in (2.14) with transition function $P(t, \cdot|s, \cdot)$ and initial distribution ν ($X_0 \sim \nu$) if and only if for all $0 = t_0 < t_1 < \cdots < t_n$ and bounded measurable functions f_j, $j = 0, \dots N$ we have

$$\mathbb{E}_\nu \prod_{j=0}^{n} f_j(X_{t_j}) = \int_{\mathbb{R}^d} f_0(x_0)\nu(dx_0) \prod_{j=1}^{n} \int_{\mathbb{R}^d} f_j(x_j) P(dx_j, t_j|x_{j-1}, t_{j-1}), \tag{2.17}$$

, where we have used the notation $\mathbb{E}_\nu$ to emphasize the dependence of the expectation on the initial distribution ν. The proof that a Markov process with transition function P satisfies (2.17) follows from the Chapman-Kolmogorov equation (2.16) and an induction argument. In other words, the finite dimensional distributions of X_t are uniquely determined by the initial distribution and the transition function:

$$\mathbb{P}(X_0 \in dx_0,\, X_{t_1} \in dx_1,\, \dots, X_{t_n} \in dx_n) = \nu(dx_0) \prod_{j=1}^{n} P(dx_j, t_j|x_{j-1}, t_{j-1}). \tag{2.18}$$

In this book we will consider Markov processes for which the transition function has a density with respect to the Lebesgue measure:

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

We will refer to $p(y, t|x, s)$ as the *transition probability density*. It is a function of four arguments, the initial position and time x, s and the final position and time y, t. For $t = s$ we have $P(\Gamma, s|x, s) = I_\Gamma(x)$. The Chapman-Kolmogorov equation becomes:

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

and, since $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ is arbitrary, we obtain the Chapman-Komogorov equation for the transition probability density:

$$p(y, t|x, s) = \int_{\mathbb{R}^d} p(y, t|z, u) p(z, u|x, s)\, dz. \tag{2.19}$$

When the transition probability density exists, and assuming that the initial distribution ν has a density ρ, we can write $\mathbb{P}(X_0 \in dx_0,\, X_{t_1} \in dx_1,\, \dots, X_{t_n} \in dx_n) = p(x_0, t_0, \dots x_n, t_n) \prod_{j=0}^{n} dx_j$ and we have

$$p(x_0, t_0, \dots x_n, t_n) = \rho(x_0) \prod_{j=1}^{n} p(x_j, t_j|x_{j-1}, t_{j-1}). \tag{2.20}$$

The above formulas simplify when the (random) law of evolution of the Markov process X_t does not change in time. In this case the conditional probability in (2.15) depends on the initial and final time t and s only through their difference: we will say that a Markov process is *time-homogeneous* if the transition function $P(\cdot, t|\cdot, s)$ depends only on the difference between the initial and final time $t - s$:

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

for all $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. For time-homogeneous Markov processes with can fix the initial time, $s = 0$. The Chapman-Kolmogorov equation for a time-homogeneous Markov process becomes

$$P(t+s, x, \Gamma) = \int_{\mathbb{R}^d} P(s, x, dz) P(t, z, \Gamma). \tag{2.21}$$

Furthermore, formulas (2.17) and (2.18) become

$$\mathbb{E}_\nu \prod_{j=0}^{n} f_j(X_{t_j}) = \int_{\mathbb{R}^d} f_0(y_0)\mu(dy_0) \prod_{j=1}^{n} \int_{\mathbb{R}^d} f_j(y_j) P(t_j - t_{j-1}, y_{j-1}, dy_j), \tag{2.22}$$

and

$$\mathbb{P}(X_0 \in dx_0,\, X_{t_1} \in dx_1,\, \ldots, X_{t_n} \in dx_n) = \nu(dx_0) \prod_{j=1}^{n} P(t_j - t_{j-1}, y_{j-1}, dy_j), \tag{2.23}$$

respectively. Given the initial distribution ν and the transition function $P(x, t, \Gamma)$ of a Markov process X_t, we can calculate the probability of finding X_t in a set Γ at time t:

$$\mathbb{P}(X_t \in \Gamma) = \int_{\mathbb{R}^d} P(x, t, \Gamma)\nu(dx).$$

Furthermore, for an observable f we can calculate the expectation using the formula

$$\mathbb{E}_\nu f(X_t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x) P(t, x_0, dx)\nu(dx_0). \tag{2.24}$$

The Poisson process, defined in (2.3) is a homogeneous Markov process. Another example of a time-homogenous Markov process is Brownian motion. The transition function is the Gaussian

$$P(t, x, dy) = \gamma_{t,x}(y)dy, \;\; \gamma_{t,x}(y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|x-y|^2}{2t}\right). \tag{2.25}$$

Let now X_t be time-homogeneous Markov process and assume that the transition probability density exists, $P(t, x, \Gamma) = \int_\Gamma p(t, x, y)\, dy$. The Chapman-Kolmogorov equation $p(t, x, y)$ reads

$$p(t+s, x, y) = \int_{\mathbb{R}^d} p(s, x, z) p(t, z, y)\, dz. \tag{2.26}$$

2.3 The Generator of a Markov Processes

Let X_t denote a time-homogeneous Markov process. The Chapman-Kolmogorov equation (2.21) suggests that a time-homogeneous Markov process can be described through a semigroup of operators, i.e. a one-parameter family of linear operators with the properties

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

Indeed, let $P(t,\cdot,\cdot)$ be the transition function of a homogeneous Markov process and let $f \in C_b(\mathbb{R}^d)$, the space of continuous bounded functions on $\mathbb{R}^d$ and define the operator

$$(P_t f)(x) := \mathbb{E}(f(X_t)|X_0 = x) = \int_{\mathbb{R}^d} f(y)P(t,x,dy). \tag{2.27}$$

This is a linear operator with

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

which means that $P_0 = I$. Furthermore:

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

Consequently:

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

The semigroup P_t defined in (2.27) is an example of a Markov semigroup. We can study properties of a time-homogeneous Markov process X_t by studying properties of the Markov semigroup P_t.

Let now X_t be a Markov process in $\mathbb{R}^d$ and let P_t denote the corresponding semigroup defined in (2.27). We consider this semigroup acting on continuous bounded functions and assume that $P_t f$ is also a $C_b(\mathbb{R}^d)$ function. We define by $\mathcal{D}(\mathcal{L})$ the set of all $f \in C_b(E)$ such that the strong limit

$$\mathcal{L}f := \lim_{t\to 0}\frac{P_t f - f}{t} \tag{2.28}$$

exists. The operator $\mathcal{L} : \mathcal{D}(\mathcal{L}) \to C_b(\mathbb{R}^d)$ is called the (infinitesimal) generator of the operator semigroup P_t. We will also refer to $\mathcal{L}$ as the generator of the Markov process X_t.

The semigroup property and the definition of the generator of a Markov semigroup (2.28) imply that, formally, we can write:

$$P_t = e^{t\mathcal{L}}.$$

Consider the function $u(x,t) := (P_t f)(x) = \mathbb{E}(f(X_t)|X_0 = x)$. We calculate its time derivative:

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

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

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \tag{2.29a}$$

$$u(x,0) = f(x). \tag{2.29b}$$

Equation (2.29) is the *backward Kolmogorov equation*. It governs the evolution of the expectation value of an observable $f \in C_b(\mathbb{R}^d)$. At this level this is formal since we do not have a formula for the generator $\mathcal{L}$ of the Markov semigroup. In the case where the Markov process is the solution of a stochastic differential equation, then the generator is a second order elliptic differential operator and the backward Kolmogorov equation becomes an initial value problem for a parabolic PDE. See Section 2.5 and Chapter 4.

As an example consider the Brownian motion in one dimension. The transition function is given by (2.25), the fundamental solution of the heat equation in one dimension. The corresponding Markov semigroup is the heat semigroup $P_t = \exp\left(\frac{t}{2}\frac{d^2}{dx^2}\right)$. The generator of the one dimensional Brownian motion is the one dimensional Laplacian $\frac{1}{2}\frac{d^2}{dx^2}$. The backward Kolmogorov equation is the heat equation

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

The Adjoint Semigroup

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

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

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

$$\int_{\mathbb{R}} P_t f(x)\, d\mu(x) = \int_{\mathbb{R}} f(x)\, d(P_t^*\mu)(x). \tag{2.30}$$

We can, write

$$P_t^* = e^{t\mathcal{L}^*}, \tag{2.31}$$

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 X_t be a Markov process with generator X_t with $X_0 \sim \mu$ and let P_t^* denote the adjoint semigroup defined in (2.31). We define

$$\mu_t := P_t^*\mu. \tag{2.32}$$

This is the *law* of the Markov process. An argument similar to the one used in the derivation of the backward Kolmogorov equation (2.29) 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 both the initial distribution μ and the law of the process μ_t have a density with respect to Lebesgue measure, $\rho_0(\cdot)$ and $\rho(t, \cdot)$, respectively, this equation becomes:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho, \quad \rho(y, 0) = \rho_0(y). \tag{2.33}$$

This is the *forward Kolmogorov* equation. When the initial conditions are deterministic, $X_0 = x$, the initial condition becomes $\rho_0 = \delta(x - y)$. As with the backward Kolmogorov equation (2.29) this equation is still formal, still we do not have a formula for the adjoint $\mathcal{L}^*$ of the generator of the Markov process X_t. In Section 2.5 we will derive the forward and backward Kolmogorov equations and a formula for the generator $\mathcal{L}$ for diffusion processes.

2.4 Ergodic Markov processes

In Sect. 1.2, we studied stationary stochastic processes, and we showed that such processes satisfy a form of the law of large numbers, Theorem 1.16. In this section we introduce a class of Markov processes for which the phase-space average, with respect to an appropriate probability measure, the *invariant measure* equals the long time average. Such Markov processes are called *ergodic*. Ergodic Markov processes are characterized by the fact that an invariant measure, see Eq. (2.36) below, exists and is unique.

For the precise definition of an ergodic Markov process we require that all shift invariant sets, i.e. all sets of the form $(X_{t_1} \in \Gamma_1, X_{t_2} \in \Gamma_2, \ldots, X_{t_k} \in \Gamma_k)$ that are invariant under time shifts, are trivial, i.e. they have probability 0 or 1. For our purposes it is more convenient to describe ergodic Markov processes in terms of the properties of their generators and of the corresponding Markov semigroup.

We will consider a Markov process X_t in $\mathbb{R}^d$ with generator $\mathcal{L}$ and Markov semigroup P_t. We will say that X_t is ergodic provided that 0 is a simple eigenvalue of $\mathcal{L}$ or, equivalently, provided that the equation

$$\mathcal{L}g = 0 \tag{2.34}$$

has only constant solutions. Consequently, we can study the ergodic properties of a Markov process X_t by studying the null space of its generator. From (2.34), and using the definition of the generator of a Markov process (2.28), we deduce that a Markov process is ergodic if the equation

$$P_t g = g, \tag{2.35}$$

has only constant solutions for all $t \geqslant 0$. Using the adjoint semigroup, we can define an *invariant measure* as a probability measure that is invariant under the time evolution of X_t, i.e., a fixed point of the semigroup P_t^*:

$$P_t^* \mu = \mu. \tag{2.36}$$

This equation is the L^2-adjoint of the equation $P_t g = g$ in (2.35). If there is a unique probability measure satisfying (2.36), then the Markov process is ergodic (with respect to the measure μ). Using this, we can obtain an equation for the invariant measure in terms of the adjoint $\mathcal{L}^*$ of the generator, which is the generator of the semigroup P_t^*. Assume, for simplicity, that the measure μ has a density ρ with respect to Lebesgue measure. We divide (2.36) by t and pass to the limit as $t \to 0$ to obtain

$$\mathcal{L}^* \rho = 0. \tag{2.37}$$

When X_t is a diffusion process, this equation is the *stationary Fokker–Planck equation*. Equation (2.37), which is the adjoint of (2.34), can be used to calculate the invariant distribution ρ, i.e., the density of the invariant measure μ.

The invariant measure (distribution) governs the long-time dynamics of the Markov process. In particular, when $X_0 \sim \mu_0$ initially, we have that

$$\lim_{t\to+\infty} P_t^* \mu_0 = \mu. \tag{2.38}$$

Furthermore, the long-time average of an observable f converges to the equilibrium expectation with respect to the invariant measure

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T f(X_s)\,ds = \int f(x)\,\mu(dx).$$

This is the definition of an ergodic process that is quite often used in physics: the long-time average equals the phase-space average.

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; see Sect. 1.2.

Example 2.4. *Brownian motion in $\mathbb{R}^d$ is not an ergodic Markov process. On the other hand, if we consider it in a bounded domain with appropriate boundary conditions, then it becomes an ergodic process. 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 spaces of both $\mathcal{L}$ and $\mathcal{L}^*$ comprise constant functions on $[0,1]$. Both the backward Kolmogorov and the Fokker–Planck equation reduce to the heat equation*

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2} \tag{2.39}$$

with periodic boundary conditions in $[0,1]$. We can solve the heat equation (2.39) *using Fourier analysis to deduce that the solution converges to a constant at an exponential rate.*

Example 2.5. *The one-dimensional Ornstein–Uhlenbeck 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:

$$\mathcal{L}^*\rho = 0, \quad \rho \geqslant 0, \quad \int \rho(x)\,dx = 1. \tag{2.40}$$

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

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

where

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

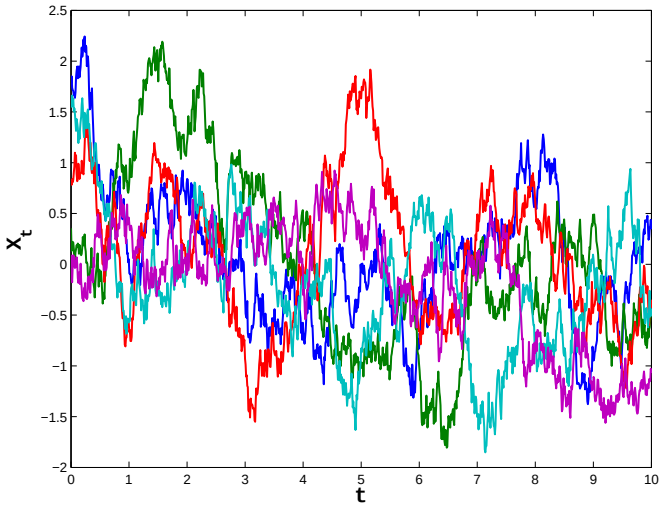


Figure 2.1: Sample paths of the Ornstein–Uhlenbeck process

We can calculate the invariant distribution by solving Eq. (2.40)*. The invariant measure of this process is the Gaussian measure*

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

If the initial condition of the Ornstein–Uhlenbeck process is distributed according to the invariant measure, then the Ornstein–Uhlenbeck process is a stationary Gaussian process. Let X_t denote the one-dimensional Ornstein–Uhlenbeck process with $X_0 \sim \mathcal{N}(0, D/\alpha)$. 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

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

The Ornstein–Uhlenbeck process is the only real-valued mean-zero Gaussian second-order stationary Markov process with continuous paths defined on $\mathbb{R}$. This is the content of Doob's theorem. See Exercise 6. A few paths of the stationary Ornstein–Uhlenbeck process are presented in Fig. 2.1.

2.5 Diffusion processes and the forward and backward Kolmogorov equations

A Markov process consists of three parts: a drift, a random part 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, together with the requirement that there are no jumps. We start with the definition of a diffusion process in one dimension.

Definition 2.6. *A Markov process X_t in $\mathbb{R}$ with transition function $P(\Gamma, t|x, s)$ is called a diffusion process if the following conditions are satisfied.*

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

$$\int_{|x-y|>\varepsilon} P(dy, t|x, s) = o(t-s) \tag{2.41}$$

uniformly over $s < t$.

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

$$\int_{|y-x|\leqslant\varepsilon} (y-x)P(dy, t|x, s) = b(x, s)(t-s) + o(t-s). \tag{2.42}$$

uniformly over $s < t$.

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

$$\int_{|y-x|\leqslant\varepsilon} (y-x)^2 P(dy, t|x, s) = \Sigma(x, s)(t-s) + o(t-s). \tag{2.43}$$

uniformly over $s < t$.

In Definition 2.6 we truncated the domain of integration since do not know whether the first and second moments of X_t are finite. If we assume that there exists a $\delta > 0$ such that

$$\lim_{t\to s} \frac{1}{t-s} \int_{\mathbb{R}^d} |y-x|^{2+\delta} P(dy, t|x, s) = 0, \tag{2.44}$$

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

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

Using this estimate together with (2.44) we conclude that:

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

This implies that Assumption (2.44) is sufficient for the sample paths to be continuous ($k = 0$) and for the replacement of the truncated integrals in (2.42) and (2.43) by integrals over $\mathbb{R}$ ($k = 1$ and $k = 2$, respectively).

Assuming that the first two moment exist, we can write the formulas for the drift and diffusion coefficients in the following form:

$$\lim_{t\to s} \mathbb{E}\left(\frac{X_t - X_s}{t-s} \Big| X_s = x \right) = b(x, s) \tag{2.45}$$

and

$$\lim_{t\to s}\mathbb{E}\left(\frac{|X_t-X_s|^2}{t-s}\Big|X_s=x\right)=\Sigma(x,s). \tag{2.46}$$

The backward Kolmogorov equation

We can now use the definition of the diffusion process in order to obtain an explicit formula for the generator of a diffusion process and to derive a partial differential equation for the conditional expectation $u(x,s)=\mathbb{E}(f(X_t)|X_s=x)$, as well as for the transition probability density $p(y,t|x,s)$. These are the backward and forward Kolmogorov equations. We will derive these equations for one dimensional diffusion processes. The extension to multidimensional diffusion processes is discussed later in this section. In the following we will assume that $u(x,s)$ is a smooth function of x and s.[4]

Theorem 2.7. *(Kolmogorov) Let $f(x)\in C_b(\mathbb{R})$ and let*

$$u(x,s):=\mathbb{E}(f(X_t)|X_s=x)=\int f(y)P(dy,t|x,s),$$

with t fixed. Assume furthermore that the functions $b(x,s)$, $\Sigma(x,s)$ are smooth in both x and s. Then $u(x,s)$ solves the final value problem, for $s\in[0,t]$,

$$-\frac{\partial u}{\partial s}=b(x,s)\frac{\partial u}{\partial x}+\frac{1}{2}\Sigma(x,s)\frac{\partial^2 u}{\partial x^2},\quad u(t,x)=f(x). \tag{2.47}$$

Proof. First we notice that, the continuity assumption (2.41), together with the fact that the function $f(x)$ is bounded imply that

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

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

$$\begin{aligned}
u(x,s) &= \int_{\mathbb{R}} f(y)P(dy,t|x,s)=f(x)+\int_{\mathbb{R}}(f(y)-f(x))P(dy,t|x,s)\\
&= f(x)+\int_{|y-x|\leqslant\varepsilon}(f(y)-f(x))P(dy,t|x,s)+\int_{|y-x|>\varepsilon}(f(y)-f(x))P(dy,t|x,s)\\
&\overset{(2.41)}{=} f(x)+\int_{|y-x|\leqslant\varepsilon}(f(y)-f(x))P(dy,t|x,s)+o(t-s).
\end{aligned}$$

[4] In fact, all we need is that $u\in C^{2,1}(\mathbb{R}\times\mathbb{R}_+)$. This can be proved using our assumptions on the transition function, on f and on the drift and diffusion coefficients.

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 (2.47). We use the Chapman-Kolmogorov equation (2.16) to obtain

$$\begin{aligned} u(x,\sigma) &= \int_{\mathbb{R}} f(z)P(dz,t|x,\sigma) = \int_{\mathbb{R}}\int_{\mathbb{R}} f(z)P(dz,t|y,\rho)P(dy,\rho|x,\sigma) \\ &= \int_{\mathbb{R}} u(y,\rho)P(dy,\rho|x,\sigma). \end{aligned} \tag{2.48}$$

We use Taylor's theorem to obtain

$$u(z,\rho) - u(x,\rho) = \frac{\partial u(x,\rho)}{\partial x}(z-x) + \frac{1}{2}\frac{\partial^2 u(x,\rho)}{\partial x^2}(z-x)^2(1+\alpha_\varepsilon), \quad |z-x| \leqslant \varepsilon, \tag{2.49}$$

where

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

and $\lim_{\varepsilon\to 0}\alpha_\varepsilon = 0$.

We combine now (2.48) with (2.49) to calculate

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

Equation (2.47) follows by taking the limits $\varepsilon \to 0,\ h \to 0$. □

Notice that the backward Kolmogorov equation (2.47) is a final value problem for a partial differential equation of parabolic type. For time-homogeneous diffusion processes, that is where the drift and the diffusion coefficients are independent of time, $b = b(x)$ and $\Sigma = \Sigma(x)$, we can rewrite it as an initial value problem. Let $T = t - s$ and introduce the function $U(x,T) = u(x,t-s)$. The backward Kolmogorov equation now becomes

$$\frac{\partial U}{\partial T} = b(x)\frac{\partial U}{\partial x} + \frac{1}{2}\Sigma(x)\frac{\partial^2 U}{\partial x^2}, \quad U(x,0) = f(x). \tag{2.50}$$

In the time-homogeneous case we can set the initial time $s = 0$. We then have that the conditional expectation $u(x,t) = \mathbb{E}(f(X_t|X_0 = x)$ is the solution to the initial value problem

$$\frac{\partial u}{\partial t} = b(x)\frac{\partial u}{\partial x} + \frac{1}{2}\Sigma(x)\frac{\partial^2 u}{\partial x^2}, \quad u(x,0) = f(x). \tag{2.51}$$

The differential operator that appears on the right hand side of (2.51) is the generator of the diffusion process X_t. In this book we will use the backward Kolmogorov in the form (2.51).[5]

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 (2.47) we obtain

$$\int_{\mathbb{R}} f(y)\left(\frac{\partial p(y,t|x,s)}{\partial s} + \mathcal{L}_{s,x}p(y,t|x,s)\right) = 0 \tag{2.52}$$

where

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

Equation (2.52) is valid for arbitrary continuous bounded functions f. COnsequently, from (2.52) we obtain a partial differential equation for the transition probability density:

$$-\frac{\partial p(y,t|x,s)}{\partial s} = b(x,s)\frac{\partial p(y,t|x,s)}{\partial x} + \frac{1}{2}\Sigma(x,s)\frac{\partial^2 p(y,t|x,s)}{\partial x^2}. \tag{2.53}$$

Notice that the variation is with respect to the "backward" variables $x,\ s$.

The forward Kolmogorov equation

Assume that the transition function has a density with respect to the Lebesgue measure which a smooth function of its arguments

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

We can obtain an equation with respect to the "forward" variables $y,\ t$, the forward Kolmogorov or Fokker-Planck equation.

Theorem 2.8. *(Kolmogorov) Assume that conditions* (2.41), (2.42), (2.43) *are satisfied and that* $p(y,t|\cdot,\cdot),\ b(y,t),\ \Sigma(y,t)$ *are smooth functions of* $y,\ t$. *Then the transition probability density is the solution to the initial value problem*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}\left(b(t,y)p\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\Sigma(t,y)p\right), \quad p(s,y|x,s) = \delta(x-y). \tag{2.54}$$

Proof. The initial condition follows from the definition of the transition probability density $p(y,t|x,s)$. Fix now 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

$$\lim_{h\to 0}\frac{1}{h}\left(\int f(y)p(y,s+h|x,s)\,ds - f(x)\right) = b(x,s)\frac{df}{dx}(x) + \frac{1}{2}\Sigma(x,s)\frac{d^2f}{dx^2}(x), \tag{2.55}$$

[5]The backward Kolmogorov equation can also be derived using Itô's formula. See Chapter 3.

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

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

In the above calculation we 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 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 (2.54). Define

$$p(y,t) := \int p(y,t|x,0)\rho_0(x)\,dx. \tag{2.56}$$

We multiply the forward Kolmogorov equation (2.54) by $\rho_0(x)$ and integrate with respect to x to obtain the equation

$$\frac{\partial p(y,t)}{\partial t} = -\frac{\partial}{\partial y}\left(a(y,t)p(y,t)\right)+\frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(y,t)p(t,y)\right), \tag{2.57}$$

together with the initial condition

$$p(y,0) = \rho_0(y). \tag{2.58}$$

The solution of equation (2.57), 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 (2.54) as the Green's function for the partial differential equation (2.57). Using (2.57) we can calculate the expectation of an arbitrary function of the diffusion process X_t:

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

where $p(y,t)$ is the solution of (2.57).

The solution of the Fokker-Planck equation provides us with the transition probability density. The Markov property enables to calculate joint probability densities using Equation (2.20). For example, let X_t

denote a diffusion process with $X_0 \sim \pi$, let $0 = t_0 < t_1 \cdots < t_n$ and let $f(x_0, \ldots, x_n)$ a measurable function. Denoting by $\mathbb{E}_\pi$ the expectation with respect to π, we have

$$\mathbb{E}_\pi f(X_{t_0}, X_{t_1}, \ldots X_{t_n}) = \int \ldots \int f(x_0, \ldots x_n)\pi(x_0)\, dx_0 \prod_{j=1}^{n} p(x_j, t_j | x_{j-1}, t_{j-1}) dx_j. \tag{2.59}$$

In particular, the autocorrelation function of X_t at time t and 0 is given by the formula

$$C(t) := \mathbb{E}_\pi(X_t X_0) = \int \int yxp(y, t|x, 0)\pi(x)\, dxdy. \tag{2.60}$$

Multidimensional Diffusion Processes

The backward and forward Kolmogorov equations can be derived for multidimensional diffusion processes using the same calculations and arguments that were used in the proofs of Theorems 2.7 and 2.8. Let X_t be a diffusion process in $\mathbb{R}^d$. The drift and diffusion coefficients of a diffusion process in $\mathbb{R}^d$ are defined as:

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

and

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

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

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

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

$$\lim_{t\to s} \mathbb{E}\left(\frac{X_t - X_s}{t-s}\Big| X_s = x\right) = \mathbf{b}(x, s) \tag{2.61}$$

and

$$\lim_{t\to s} \mathbb{E}\left(\frac{(X_t - X_s) \otimes (X_t - X_s)}{t-s}\Big| X_s = x\right) = \mathbf{\Sigma}(x, s). \tag{2.62}$$

The backward and forward Kolmogorov equations for mutidimensional diffusion processes are

$$-\frac{\partial u}{\partial s} = \mathbf{b}(x, s) \cdot \nabla_x u + \frac{1}{2}\mathbf{\Sigma}(x, s) : \nabla_x \nabla_x u, \quad u(t, x) = f(x), \tag{2.63}$$

and

$$\frac{\partial p}{\partial t} = \nabla_y \cdot \left(-\mathbf{b}(t, \mathbf{y})p + \frac{1}{2}\nabla_y \cdot \big(\mathbf{\Sigma}(t, \mathbf{y})p\big)\right), \quad p(\mathbf{y}, s|\mathbf{x}, s) = \delta(x - y). \tag{2.64}$$

As for one-dimensional time-homogeneous diffusion processes, the backward Kolmogorov equation for a time-homogeneous multidimensional diffusion process can be written as an initial value problem:

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(x,0) = f(x), \tag{2.65}$$

for $u(x,t) = \mathbb{E}(f(X_t)|X_0 = x)$. For time-homogeneous processes we can fix the initial time $s = 0$ in the forward Kolmogorov equation. Assuming furthermore that X_0 is a random variable with probability density $\rho_0(x)$ the forward Kolmogorov equation becomes

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \quad p(x,0) = \rho_0(x), \tag{2.66}$$

for the transition probability density $p(y,t)$. In this book we will use the forward and backward Kolmogorov equations in the form (2.65) and (2.66).

2.6 Discussion and Bibliography

Markov chains in both discrete and continuous time are studied in [74, 98]. A standard reference on Markov processes is [19]. The proof that the transition function and the initial distribution of X_t are sufficient to uniquely determine a Markov process, Equation (2.17) can be found in [83, Prop. 1.4, Ch. III]. See also [19, Thm 1.1, Ch. 4].

Operator semigroups are the main analytical tool for the study of diffusion processes, see for example [60]. Necessary and sufficient conditions for an operator $\mathcal{L}$ to be the generator of a (contraction) semigroup are given by the Hille-Yosida theorem [21, Ch. 7].

The space $C_b(E)$ is natural in a probabilistic context for the study of Markov semigroups, but other function 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 is an invariant measure of the Markov process. Markov semigroups can be extended from the space of bounded continuous functions to the space $L^p(E;\mu)$ for any $p \geqslant 1$. The proof of this result, which follows from Jensen's inequality and the Hahn-Banach theorem, can be found in [34, Prop 1.14].

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 $f \in \mathcal{D}(P_t)$.

The argument used in the derivation of the forward and backward Kolmogorov equations goes back to Kolmogorov's original work. See [30] and [38]. A more modern approach to the derivation of the forward equation from the Chapman-Kolmogorov equation can be found in [96, Ch. 1]. The connection between Brownian motion and the corresponding Fokker-Planck equation which is the heat equation was made by Einstein [18]. Many of the early papers on the theory of stochastic processes have been reprinted in [17].

Early papers on Brownian motion, including the original papers by Fokker and by Planck, are available from `http://www.physik.uni-augsburg.de/theo1/hanggi/History/BM-History.html`. Very interesting historical comments can also be found in [72] and [68].

We can also derive backward and forward Kolmogorov equations for continuous-time Markov processes with jumps. For such processes, an additional nonlocal in space term (an integral operator) appears in the Kolmogorov equations that accounts for the jumps.[6] Details can be found in [28].

A diffusion process is characterized by the (almost sure) continuity of its paths and by specifying the first two moments. A natural question that arises is whether other types of stochastic processes can be defined by specifying a fixed number of moments higher than two. It turns out that this is not possible: we either need to retain two or all (i.e. infinitely many) moments. Specifying a finite number of moments, greater than two, leads to inconsistencies. This is the content of Pawula's theorem . [77]. Example 2.4 can also be found in [76, Ch. 6].

The duality between the backward and forward Kolmogorov equations, the duality between studying the evolution of observables and states, is similar to the duality between the Heisenberg (evolution of observables) and Schrödinger (evolution of states) representations of quantum mechanics or the Koopman (evolution of observables) and Frobenius-Perron (evolution of states) operators in the theory of dynamical systems, i.e. on the duality between the study of the evolution of observables and of states. See [82, 52, 100].

2.7 Exercises

1. Let $\{X_n\}$ be a stochastic process with state space $S = \mathbb{Z}$. Show that it is a Markov process if and only if for all n
$$\mathbb{P}(X_{n+1} = i_{n+1}|X_1 = i_1, \dots X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1}|X_n = i_n).$$

2. Show that (2.6) is the solution of initial value problem (2.12) as well as of the final value problem
$$-\frac{\partial p}{\partial s} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \quad \lim_{s\to t} p(y,t|x,s) = \delta(y-x).$$

3. Use (2.7) to show that the forward and backward Kolmogorov equations for the OU process are
$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial y}(yp) + \frac{1}{2}\frac{\partial^2 p}{\partial y^2}$$
and
$$-\frac{\partial p}{\partial s} = -x\frac{\partial p}{\partial x} + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}.$$

4. Let $W(t)$ be a standard one dimensional Brownian motion, let $Y(t) = \sigma W(t)$ with $\sigma > 0$ and consider the process
$$X(t) = \int_0^t Y(s)\, ds.$$
Show that the joint process $\{X(t),\, Y(t)\}$ is Markovian and write down the generator of the process.

[6] The generator of a Markov process with jumps is necessarily nonlocal: a local (differential) operator $\mathcal{L}$ corresponds to a Markov process with continuous paths. See [96, Ch. 1].

5. Let $Y(t) = e^{-t}W(e^{2t})$ be the stationary Ornstein-Uhlenbeck process and consider the process

$$X(t) = \int_0^t Y(s)\, ds.$$

Show that the joint process $\{X(t),\, Y(t)\}$ is Markovian and write down the generator of the process.

6. (a) Let $X,\, Y$ be mean zero Gaussian random variables with $\mathbb{E}X^2 = \sigma_X^2,\, \mathbb{E}Y^2 = \sigma_Y^2$ and correlation coefficient ρ (the correlation coefficient is $\rho = \frac{\mathbb{E}(XY)}{\sigma_X \sigma_Y}$). Show that

$$\mathbb{E}(X|Y) = \frac{\rho\sigma_X}{\sigma_Y} Y.$$

(b) Let X_t be a mean zero stationary Gaussian process with autocorrelation function $R(t)$. Use the previous result to show that

$$\mathbb{E}[X_{t+s}|X_s] = \frac{R(t)}{R(0)} X(s), \quad s, t \geqslant 0.$$

(c) Use the previous result to show that the only stationary Gaussian Markov process with continuous autocorrelation function is the stationary OU process.

7. Show that a Gaussian process X_t is a Markov process if and only if

$$\mathbb{E}(X_{t_n}|X_{t_1} = x_1, \ldots X_{t_{n-1}} = x_{n-1}) = \mathbb{E}(X_{t_n}|X_{t_{n-1}} = x_{n-1}).$$

8. Prove equation (2.55).

9. Derive the initial value problem (2.57), (2.58).

10. Prove Theorems 2.7 and 2.8 for multidimensional diffusion processes.

Chapter 3

Introduction to Stochastic Differential Equations

In this chapter we study diffusion processes at the level of paths. In particular, we study stochastic differential equations driven by Gaussian white noise, defined formally as the derivative of Brownian motion. In Section 3.1 we introduce stochastic differential equations. In Section 3.2 we introduce the Itô and Stratonovich stochastic integrals. In Section 3.3 we present the concept of a solution for a stochastic differential equation. The generator, Itô's formula and the connection with the Fokker-Planck equation are covered in Section 3.4. Examples of stochastic differential equations are presented in Section 3.5. Lamperti's transformation and Girsanov's theorem are discussed briefly in Section 3.6. Linear stochastic differential equations are studied in Section 3.7. Bibliographical remarks and exercises can be found in Sections 3.8 and 3.9, respectively.

3.1 Introduction

We consider stochastic differential equations (SDEs) of the form

$$\frac{dX(t)}{dt} = b(t,\, X(t)) + \boldsymbol{\sigma}(t,\, X(t))\, \xi(t), \quad X(0) = x. \tag{3.1}$$

where $X(t) \in \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$. We use the notation $\xi(t) = \frac{dW}{dt}$ to denote (formally) the derivative of Brownian motion in $\mathbb{R}^m$, i.e. the white noise process which is a (generalized) mean zero Gaussian vector-valued stochastic process with autocorrelation function

$$\mathbb{E}\left(\xi_i(t)\xi_j(s)\right) = \delta_{ij}\delta(t-s), \quad i,j = 1\ldots m. \tag{3.2}$$

The initial condition x can be either deterministic or a random variable which is independent of the Brownian motion $W(t)$, in which case there are two different, independent sources of randomness in (3.1). We will use different notations for the solution of an SDE:

$$X(t), \;\; X_t \;\text{ or }\; X_t^x.$$

The latter notation will be used when we want to emphasize the dependence of the solution of the initial conditions.

We will consider mostly autonomous SDEs, i.e. equations whose coefficients do not depend explicitly on time. When we study stochastic resonance and Brownian motors in Sections **??** and **??**, respectively, it will be necessary to consider SDEs with time dependent coefficients. It is also often useful to consider SDEs in bounded domains, for example in a box of size L, $[0, L]^d$ with periodic boundary conditions; see Section **??** on Brownian morion in periodic potentials.

The amplitude of the noise in (3.1) may be independent of the state of the system, $\sigma(x) \equiv \sigma$, a constant; in this case we will say that the noise in (3.1) is *additive*. When the amplitude of the noise depends on the state of the system we will say that the noise in (3.1) is *multiplicative*. In the modeling of physical systems using stochastic differential equations additive noise is usually due to thermal fluctuations whereas multiplicative noise is due to noise in some control parameter.

Example 3.1. Consider the Landau-Stuart equation

$$\frac{dX}{dt} = X(\alpha - X^2),$$

where α is a parameter. Assume that this parameter fluctuates randomly in time or that we are uncertain about its actual value. Modeling this uncertainty as white noise, $\alpha \mapsto \alpha + \sigma\xi$ we obtain the stochastic Landau equation with multiplicative noise:

$$\frac{dX_t}{dt} = X_t(\alpha - X_t^2) + \sigma X_t\, \xi(t). \tag{3.3}$$

It is important to note that an equation of the form (3.3) is not sufficient to determine uniquely the stochastic process X_t: we also need to determine how we chose to interpret the noise in the equation, e.g. whether the noise in (3.3) is Itô or Stratonovich. This is a separate modeling issue that we will address in Section **??**.

Since the white noise process $\xi(t)$ is defined only in a generalized sense, equation (3.1) is only formal. We will usually write it in the form

$$dX(t) = \mathbf{b}(t,\, X(t))\, dt + \boldsymbol{\sigma}(t,\, X(t))\, dW(t), \tag{3.4}$$

together with the initial condition $X(0) = x$, or, componentwise,

$$dX_i(t) = b_i(t,\, X(t))\, dt + \sum_{j=1}^{m} \boldsymbol{\sigma}_{ij}(t,\, X(t))\, dW_j(t), \quad j = 1, \ldots, d, \tag{3.5}$$

together with the initial conditions. In fact, the correct interpretation of the SDE (3.4) is as a stochastic integral equation

$$X(t) = x + \int_0^t \mathbf{b}(t,\, X(t))\, dt + \int_0^t \boldsymbol{\sigma}(t,\, X(t))\, dW(t). \tag{3.6}$$

Even when writing the SDE as an integral equation, we are still facing several mathematical difficulties. First we need to give an appropriate definition of the stochastic integral

$$I(t) := \int_0^t \boldsymbol{\sigma}(t,\, X(t))\, dW(t), \tag{3.7}$$

or, more generally,

$$I(t) := \int_0^t h(t)\, dW(t), \tag{3.8}$$

for a sufficiently large class of functions. Since Brownian motion is not of bounded variation, the integral (3.7) cannot be defined as a Riemann-Stieljes integral in a unique way. As we will see in the next section, different Riemann-Stieljes approximations lead to different stochastic integrals that, in turn, lead to stochastic differential equations with different properties.

After defining the stochastic integral in (3.6) we need to give a proper definition of a solution to an SDE. In particular, we need to give a definition that takes into account the randomness due to the Brownian motion and the initial conditions. Furthermore, we need to take into account the fact that, since Brownian motion is not regular but only Hölder continuous with exponent $\alpha < 1/2$, solutions to an SDE of the form (3.1) cannot be very regular. As in the case of partial differential equations, there are different concepts of solution for an SDE of the form (3.1).

After having given an appropriate definition for the stochastic integral and developed an existence and uniqueness theory of solutions to SDEs, we would like to be able to calculate (the statistics of) functionals of the solution to the SDE. Let $X(t)$ be the solution of (3.4) and let $f(t, x)$ be a sufficiently regular function of t and x. We want to derive an equation for the function

$$z(t) = f(t, X(t)).$$

In the absence of noise we can easily obtain an equation for $z(t)$ by using the chain rule. However, the stochastic forcing in (3.1) and the lack of regularity of Brownian motion imply that the chain rule has to be modified appropriately. This is, roughly speaking, due to the fact that $\mathbb{E}(dW(t))^2 = dt$ and, consequently, second order differentials need to be kept when calculating the differential of $z(t)$. It turns out that whether a correction to the chain rule from standard calculus is needed depends on how we interpret the stochastic integral (3.7).

Furthermore, we would like to be able to calculate the statistics of solutions to SDEs. In Chapter 2 we saw that we can calculate the expectation value of an observable

$$u(x, t) = \mathbb{E}(f(X_t^x)|X_0^x = x) \tag{3.9}$$

by solving the backward Kolmogorov equation (2.47). In this chapter we will see that the backward Kolmogorov equation is a consequence of the *Itô's formula*, the chain rule of *Itô stochastic calculus*.

Quite often it is important to be able to evaluate the statistics of solutions to SDEs at appropriate random times, the so called *stopping times*. An example of a stopping time is the *first exit time* of the solution of an SDE of the form (3.1), which is defined as the first time the diffusion X_t^x exits an open domain $D \in \mathbb{R}^d$, with $x \in D$:

$$\tau_D = \inf_{t>0} \Big\{ X_t \notin D \Big\}. \tag{3.10}$$

The statistics of the first exit time will be needed in the calculation of the escape time of a diffusion process from a metastable state, see Chapter **??**.

There is also an important modeling issue that we need to address: white noise is a stochastic process with zero correlation time. As such, it can only be thought of as an idealization of the noise that appears

in physical, chemical and biological systems. We are interested, therefore, in understanding the connection between an SDE driven by white noise with equations where a more physically realistic noise is present. The question then is whether an SDE driven by white noise (3.1) can be obtained from an equation driven by noise with a non trivial correlation structure through an appropriate limiting procedure. We will see in Section **??** that it is possible to describe more general classes of noise within the framework of diffusion processes, by adding additional variables. Furthermore, we can obtain the SDE (3.1) in the limit of zero correlation time, and for an appropriate definition of the stochastic integral.

3.2 The Itô and Stratonovich Stochastic Integrals

In this section we define stochastic integrals of the form

$$I(t) = \int_0^t f(s)\, dW(s), \tag{3.11}$$

where $W(t)$ is a standard one dimensional Brownian motion and $t \in [0, T]$. We are interested in the case where the integrand is a stochastic process whose randomness depends on the Brownian motion $W(t)$-think of the stochastic integral in (3.6)-and, in particular, that it is *adapted* to the *filtration* $\mathcal{F}_t$ (see 2.14) generated by the Brownian motion $W(t)$, i.e. that it is an $\mathcal{F}_t$ measurable function for all $t \in [0, T]$. Roughly speaking, this means that the integrand depends only the past history of the Brownian motion with respect to which we are integrating in (3.11). Furthermore, we will assume that the random process $f(\cdot)$ is square integrable:

$$\mathbb{E}\left(\int_0^T f(s)^2\, ds\right) < \infty.$$

Our goal is to define the stochastic integral $I(t)$ as the L^2-limit of a Riemann sum approximation of (3.11). To this end, we introduce a partition of the interval $[0, T]$ by setting $t_k = k\Delta t,\ k = 0, \ldots K - 1$ and $K\Delta t = t$; we also define a parameter $\lambda \in [0, 1]$ and set

$$\tau_k = (1 - \lambda)t_k + \lambda t_{k+1}, \quad k = 0, \ldots K - 1. \tag{3.12}$$

We define now stochastic integral as the $L^2(\Omega)$ limit (Ω denoting the underlying probability space) of the Riemann sum approximation

$$I(t) := \lim_{K\to\infty} \sum_{k=0}^{K-1} f(\tau_k)\left(W(t_{k+1}) - W(t_k)\right). \tag{3.13}$$

Unlike the case of the Riemann-Stieltjes integral when we integrate against a smooth deterministic function, the result in (3.13) depends on the choice of $\lambda \in [0, 1]$ in (3.12). The two most common choices are $\lambda = 0$, in which case we obtain the *Itô stochastic integral*

$$I_I(t) := \lim_{K\to\infty} \sum_{k=0}^{K-1} f(t_k)\left(W(t_{k+1}) - W(t_k)\right). \tag{3.14}$$

The second choice is $\lambda = \frac{1}{2}$, which leads to the *Stratonovich stochastic integral*:

$$I_S(t) := \lim_{K\to\infty} \sum_{k=0}^{K-1} f\left(\frac{1}{2}(t_k + t_{k+1})\right)\left(W(t_{k+1}) - W(t_k)\right). \tag{3.15}$$

We will use the notation

$$I_S(t) = \int_0^t f(s) \circ dW(s),$$

to denote the Stratonovich stochastic integral. In general the Itô and Stratonovich stochastic integrals are different. When the integrand $f(t)$ depends on the Brownian motion $W(t)$ through $X(t)$, the solution of the SDE in (3.1), a formula exists for converting one stochastic integral into another.

When the integrand in (3.14) is a sufficiently smooth function, then the stochastic integral is independent of the parameter λ and, in particular, the Itô and Stratonovich stochastic integrals coincide.

Proposition 3.2. *Assume that there exist C, $\delta > 0$ such that*

$$\mathbb{E}\big(f(t) - f(s)\big)^2 \leqslant C|t-s|^{1+\delta}, \quad 0 \leqslant s,\, t \leqslant T. \tag{3.16}$$

Then the Riemann sum approximation in (3.13) *converges in $L^1(\Omega)$ to the same value for all $\lambda \in [0,1]$.*

The interested reader is invited to provide a proof of this proposition.

Example 3.3. Consider the Langevin equation (see Chapter **??**) with a space dependent friction coefficient:

$$m\ddot{q}_t = -\nabla V(q_t) - \gamma(q_t)\dot{q}_t + \sqrt{2\gamma(q_t)}\dot{W}_t,$$

where q denotes the particle position with mass m. Writing the Langevin equation as a system of first order SDEs we have

$$\begin{aligned} m dq_t &= p_t\, dt, && \text{(3.17a)}\\ dp_t &= -\nabla V(q_t)\, dt - \gamma(q_t) p_t\, dt + \sqrt{2\gamma(q_t)}\, dW_t. && \text{(3.17b)} \end{aligned}$$

Assuming that the potential and the friction coefficients are smooth functions, the particle position q is a differentiable function of time.[1] Consequently, according to Proposition 3.2, the Itô and Stratonovich stochastic integrals in (3.17) coincide. This is not true in the limit of small mass, $m \to 0$. See the Section **??** and the discussion in Section 3.8.

The Itô stochastic integral $I(t)$ is almost surely continuous in t. As expected, it satisfies the linearity property

$$\int_0^T \Big(\alpha f(t) + \beta g(t)\Big)\, dW(t) = \alpha \int_0^T f(t)\, dW(t) + \beta \int_0^T g(t)\, dW(t), \quad \alpha,\, \beta \in \mathbb{R}$$

for all square integrable functions $f(t)$, $g(t)$. Furthermore, the Itô stochastic integral satisfies the *Itô isometry*

$$\mathbb{E}\left(\int_0^T f(t)\, dW(t)\right)^2 = \int_0^T \mathbb{E}|f(t)|^2\, dt, \tag{3.18}$$

[1] In fact, it is a $C^{1+\alpha}$ function of time, with $\alpha < 1/2$.

from which it follows that, for all square integrable functions $f,\ g$,[2]

$$\mathbb{E}\left(\int_0^T h(t)\,dW(t)\int_0^T g(s)\,dW(s)\right)=\mathbb{E}\int_0^T h(t)g(t)\,dt.$$

The Itô stochastic integral is a *martingale* :

Definition 3.4. *Let $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration defined on the probability space $(\Omega,\mathcal{F},\mu)$ and let $\{\mathcal{M}_t\}_{t\in[0,T]}$ adapted to $\mathcal{F}_t$ with $\mathcal{M}_t\in L^1(0,T)$. We say that $\mathcal{M}_t$ is an $\mathcal{F}_t$ martingale if*

$$\mathbb{E}[\mathcal{M}_t|\mathcal{F}_s]=\mathcal{M}_s\quad\forall\, t\geqslant s.$$

For the Itô stochastic integral we have

$$\mathbb{E}\int_0^t f(s)\,dW(s)=0 \tag{3.19}$$

and

$$\mathbb{E}\left[\int_0^t f(\ell)\,dW(\ell)|\mathcal{F}_s\right]=\int_0^s f(\ell)\,dW(\ell)\quad\forall\, t\geqslant s, \tag{3.20}$$

where $\mathcal{F}_s$ denotes the filtration generated by $W(s)$. The quadratic variation of this martingale is

$$\langle I\rangle_t=\int_0^t (f(s))^2\,ds.$$

The proofs of all these properties and the study of the Riemann sum (3.13) proceed as follows: first, these properties are proved for the simplest possible functions, namely step functions for which we can perform explicit calculations using the properties of Brownian increments. Then, an approximation step is used to show that square integrable functions can be approximated by step functions. The details of these calculations can be found in the references listed in Section 3.8.

The above ideas are readily generalized to the case where $W(t)$ is a standard d- dimensional Brownian motion and $f(t)\in\mathbb{R}^{m\times d}$ for each $t>0$. In the multidimensional case the Itô isometry takes the form

$$\mathbb{E}|I(t)|^2=\int_0^t \mathbb{E}|f(s)|_F^2 ds, \tag{3.21}$$

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

Whether we choose the Itô or Stratonovich interpretation of the stochastic integral in (3.6) is a modeling issue that we will address later in Section **??**. Both interpretations have their advantages: the Itô stochastic integral is a martingale and we can use the well developed theory of martingales to study its properties; in particular, there are many inequalities and limit theorems for martingales that are very useful in the rigorous study of qualitative properties of solutions to SDEs. On the other hand, the Stratonovich stochastic integral leads to the standard Newton-Leibniz chain rule, as opposed to the Itô stochastic integral where a correction to the Leibniz chain rule is needed. Furthermore, as we will see in Section **??**, SDEs driven by noise with non-zero correlation time converge, in the limit as the correlation time tends to 0, to the Stratonivich SDE.

[2]We use Itô isometry with $f=h+g$.

When the stochastic integral in (3.4) or (3.6) is an Itô integral we will refer to the SDE as an *Itô SDE*, whereas when it is a Stratonovich integral we will refer to the SDE as a *Stratonovich stochastic differential equation*. There are other interpretations of the stochastic integral, that arise in applications e.g. the Klimontovich (kinetic) stochastic integral that corresponds to the choice $\lambda = 1$ in (3.12).

An Itô SDE can be converted into a Stratonovich SDE and the other way around. This transformation involves the addition (or subtraction) of a drift term. We calculate this correction to the drift in one dimension. Consider the Itô SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t. \tag{3.22}$$

We want to write it in Stratonovich form:

$$dX_t = \widehat{b}(X_t)\,dt + \widehat{\sigma}(X_t) \circ dW_t. \tag{3.23}$$

Let us calculate the Stratonovich stochastic integral in (3.23). We have, with $\alpha = \frac{1}{2}$, and using the notation $\Delta W_j = W(t_{j+1}) - W(t_j)$ and similarly for ΔX_j as well as a Taylor series expansion,

$$\begin{aligned}
\int_0^t \widehat{\sigma}(X_t) \circ dW_t &\approx \sum_j \widehat{\sigma}(X(j\Delta t + \alpha\Delta t))\,\Delta W_j \\
&\approx \sum_j \widehat{\sigma}(X(j\Delta t))\,\Delta W_j + \alpha \sum_j \frac{d\widehat{\sigma}}{dx}(X(j\Delta t))\Delta X_j\,\Delta W_j \\
&\approx \int_0^t \widehat{\sigma}(X(t))\,dW(t) + \alpha \sum_j \frac{d\widehat{\sigma}}{dx}(X(j\Delta t))\Big(b(X_j)\,\Delta t_j + \sigma(X_j)\Delta W_j\Big)\,\Delta W_j \\
&\approx \int_0^t \widehat{\sigma}(X(t))\,dW(t) + \alpha \int_0^t \frac{d\widehat{\sigma}}{dx}(X(t))\sigma(X(t))dt.
\end{aligned} \tag{3.24}$$

In the above calculation we have used the formulas $\mathbb{E}\big(\Delta t \Delta W_j\big) = 0$ and $\mathbb{E}\big(\Delta W_j\big)^2 = \Delta t$; see Section 3.4. This calculation suggests that the Stratonovich stochastic integral, when evaluated at the solution of the Itô SDE (3.22), is equal to the Itô stochastic integral plus a drift correction. Notice that the above calculation provides us with a correction for arbitrary choices of the parameter $\alpha \in [0, 1]$. The above heuristic argument can be made rigorous: we need to control the difference between $\int_0^t \widehat{\sigma}(X_t) \circ dW_t$ and the righthand side of (3.24) in $L^2(\Omega)$; see Exercise 2.

Substituting (3.24) in the Stratonovich SDE (3.23) we obtain

$$dX_t = \Big(\widehat{b}(X_t) + \frac{1}{2}\frac{d\widehat{\sigma}}{dx}(X_t)\sigma(X_t)\Big)\,dt + \widehat{\sigma}(X_t)\,dW_t.$$

This is the Itô equation (3.22). Comparing the drift and the diffusion coefficients we deduce that

$$\widehat{\sigma} = \sigma \quad \text{and} \quad \widehat{b} = b - \frac{1}{2}\sigma'\sigma. \tag{3.25}$$

Consequently, the Itô SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t$$

is equivalent to the Stratonovich SDE

$$dX_t = \Big(b(X_t) - \frac{1}{2}\sigma'(X_t)\sigma(X_t)\Big)\,dt + \sigma(X_t) \circ dW_t$$

Conversely, the Stratonovich SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t) \circ dW_t$$

is equivalent to the Itô SDE

$$dX_t = \Big(b(X_t) + \frac{1}{2}\sigma'(X_t)\sigma(X_t)\Big)\,dt + \sigma(X_t)\,dW_t.$$

The correction to the drift $\frac{1}{2}\sigma'\sigma$ is called the *Itô-to-Stratonovich correction*. Similar formulas can be obtained in arbitrary dimensions. The multidimensional Itô SDE

$$dX_t = \mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t)\,dW_t, \tag{3.26}$$

where $\mathbf{b} : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\boldsymbol{\sigma} : \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ can be transformed to the Stratonovich SDE

$$dX_t = \big(\mathbf{b}(X_t) - \mathbf{h}(X_t)\big)\,dt + \boldsymbol{\sigma}(X_t) \circ dW_t, \tag{3.27}$$

where the correction drift h is given by the formula

$$h_i(x) = \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{m} \sigma_{jk}(x)\frac{\partial \sigma_{ik}}{\partial x_j}(x), \quad i = 1, \ldots, d. \tag{3.28}$$

Conversely, the multidimensional Stratonovich SDE

$$dX_t = \mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t) \circ dW_t, \tag{3.29}$$

can be transformed into the Itô SDE

$$dX_t = \big(\mathbf{b}(X_t) + \mathbf{h}(X_t)\big)\,dt + \boldsymbol{\sigma}(X_t)\,dW_t, \tag{3.30}$$

with h given by (3.28). The Itô-to-Stratonovich correction can be written in index-free notation:

$$\mathbf{h}(x) = \frac{1}{2}\big[\nabla \cdot \boldsymbol{\Sigma}(x) - (\boldsymbol{\sigma}\nabla \cdot \boldsymbol{\sigma}^T)(x)\big], \quad \boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T. \tag{3.31}$$

To see this, we first note that

$$\big(\nabla \cdot \boldsymbol{\Sigma}\big)_i = \sum_k \frac{\partial \boldsymbol{\sigma}_{ik}}{\partial x_k} = \sum_{k=1}^{d}\sum_{\ell=1}^{m}\left(\frac{\partial \sigma_{i\ell}}{\partial x_k}\sigma_{k\ell} + \sigma_{i\ell}\frac{\partial \sigma_{k\ell}}{\partial x_k}\right). \tag{3.32}$$

On the other hand,

$$\big(\boldsymbol{\sigma}\nabla \cdot \boldsymbol{\sigma}^T\big)_i = \sum_{k=1}^{d}\sum_{\ell=1}^{m} \boldsymbol{\sigma}_{i\ell}\frac{\partial \sigma_{k\ell}}{\partial x_k}. \tag{3.33}$$

The equivalence between (3.31) and (3.28) follows upon subtracting (3.33) from (3.32). Notice also that we can also write

$$\mathbf{h} \cdot \ell = \frac{1}{2}\boldsymbol{\sigma}^T : \nabla\big(\boldsymbol{\sigma}^T\ell\big), \tag{3.34}$$

for all vectors $\ell \in \mathbb{R}^d$.

Notice that in order to be able to transform a Stratonovich SDE into and Itô SDE we need to assume differentiability of the matrix $\boldsymbol{\sigma}$, an assumption that is not necessary for the existence and uniqueness of solutions to an SDE; see Section 3.3.

3.3 Solutions of Stochastic Differential Equations

In this section we present, without proof, a basic existence and uniqueness result for stochastic differential equations of the form

$$dX_t = \mathbf{b}(t,\, X_t)\, dt + \boldsymbol{\sigma}(t,\, X_t)\, dW_t, \quad X(0) = x, \tag{3.35}$$

where $\mathbf{b}(\cdot,\cdot) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\boldsymbol{\sigma}\cdot,\cdot : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times n}$ are measurable vector-valued and matrix-valued functions, respectively, and W_t denotes standard Brownian motion in $\mathbb{R}^n$. We assume that the initial condition is a random variable which is independent of the Brownian motion W_t. We will denote by $\mathcal{F}_t$ the filtration generated by the Brownian motion W_t.

We will use the following concept of a solution to (3.35).

Definition 3.5. *A process X_t with continuous paths defined on the probability space $(\Omega, \mathcal{F}, P)$ is called a* **strong solution** *to the SDE* (3.35) *if:*

(i) *X_t is almost surely continuous and adapted to the filtration $\mathcal{F}_t$.*

(ii) *$\mathbf{b}(\cdot,\, X_\cdot) \in L^1((0,T);\mathbb{R}^d)$ and $\boldsymbol{\sigma}(\cdot,\, X_\cdot) \in L^2((0,T);\mathbb{R}^{d\times n})$ almost surely.*

(iii) *For every $t \geqslant 0$ the stochastic integral equation*

$$X_t = x + \int_0^t \mathbf{b}(s,\, X_s)\, ds + \int_0^t \boldsymbol{\sigma}(s,\, X_s)\, dW_s, \quad X(0) = x \tag{3.36}$$

holds almost surely.

The assumptions that we have to impose on the drift and diffusion coefficients in (3.35) so that a unique strong solutions exists are similar to the Lipschitz continuity and linear growth assumptions that are familiar from the existence and uniqueness theory of (deterministic) ordinary differential equations. In particular, we make the following two assumptions on the coefficients: there exists a positive constant C such that, for all $x \in \mathbb{R}^d$ and $t \in [0,T]$,

$$|\mathbf{b}(t,x)| + |\boldsymbol{\sigma}(t,x)|_F \leqslant C\big(1 + |x|\big) \tag{3.37}$$

and for all $x,\, y \in \mathbb{R}^d$ and $t \in [0,T]$

$$|\mathbf{b}(t,x) - \mathbf{b}(t,y)| + |\boldsymbol{\sigma}(t,x) - \boldsymbol{\sigma}(t,y)|_F \leqslant C|x-y|, \tag{3.38}$$

Notice that for globally Lipschitz vector and matrix fields $\mathbf{b}$ and $\boldsymbol{\sigma}$ (i.e. when (3.38) is satisfied), the linear growth condition (3.37) is equivalent to the requirement that $|\mathbf{b}(t,0)|$ and $|\boldsymbol{\sigma}(t,0)|_F$ are bounded for all $t \geqslant 0$.

Under these assumptions a *global, unique* solution exists for the SDE (3.35). By uniqueness of strong solutions we mean that, if X_t and Y_t are strong solutions to (3.35), then

$$X_t = Y_t \quad \text{for all } t \text{ almost surely.}$$

Theorem 3.6. *Let* $\mathbf{b}(\cdot,\cdot)$ *and* $\boldsymbol{\sigma}(\cdot,\cdot)$ *satisfy Assumptions* (3.37) *and* (3.38). *Assume furthermore that the initial condition* x *is a random variable independent of the Brownian motion* W_t *with*

$$\mathbb{E}|x|^2 < \infty.$$

Then the SDE (3.35) *has a unique strong solution* X_t *with*

$$\mathbb{E}\left[\int_0^t |X_s|^2\, ds\right] < \infty \tag{3.39}$$

for all $t > 0$.

Using Gronwall's inequality we can also obtain a quantitative estimate in (3.39), as a function of the second moment of the solution that increases exponentially in time.

The solution of the SDE (3.35) satisfies the Markov property and had continuous paths: it is a diffusion process. In fact, solutions of SDEs are precisely the diffusion processes that we studied in Chapter 2.

The Stratonovich analogue of (3.35) is

$$dX_t = \mathbf{b}(t,\, X_t)\, dt + \boldsymbol{\sigma}(t,\, X_t) \circ dW_t, \quad X(0) = x, \tag{3.40}$$

As with the Itô SDE, the correct interpretation of (3.35) is in the sense of the integral equation

$$X_t = x + \int_0^t \mathbf{b}(s,\, X_s)\, ds + \int_0^t \boldsymbol{\sigma}(s,\, X_s) \circ dW_s, \quad X(0) = x. \tag{3.41}$$

Using the Itô-to-Stratonovich transformation (3.28), we can write (3.40) as an Itô SDE and then use Theorem 3.6 to prove existence and uniqueness of strong solutions. Notice, however, that for this we need to assume differentiability of the diffusion matrix $\boldsymbol{\sigma}$ and that we need to check that conditions (3.37) and (3.38) are satisfied for the modified drift in (3.30).

Just as with nonautonomous ordinary differential equations, it is possible to rewrite a stochastic differential equation with time dependent coefficients as a time-homogeneous equation by adding one additional variable. Consider, for simplicity, the one dimensional equation

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t. \tag{3.42}$$

We introduce the auxiliary variable τ_t to write (3.42) as the system of SDEs

$$dX_t = b(\tau_t, X_t)\, dt + \sigma(\tau_t, X_t)\, dW_t, \tag{3.43a}$$

$$d\tau_t = dt. \tag{3.43b}$$

Thus we obtain a system of two homogeneous stochastic differential equations for the variables $(X_t,\, \tau_t)$. Notice that noise acts only in the equation for X_t. The generator of the diffusion process $(X_t,\, \tau_t)$ (see Section 3.4) is

$$\mathcal{L} = \frac{\partial}{\partial \tau} + b(\tau, x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(\tau, x)\frac{\partial^2}{\partial x^2}. \tag{3.44}$$

Quite often we are interested in studying the long time properties of the solution of a stochastic differential equation. For this it is useful to rescale time so that we can focus on the long time scales. Using the scaling property of Brownian motion, see Theorem 1.23,

$$W(ct) = \sqrt{c}W(t),$$

we have that, if $s = ct$, then

$$\frac{dW}{ds} = \frac{1}{\sqrt{c}}\frac{dW}{dt},$$

the equivalence being, of course, in law. Hence, if we scale time to $s = ct$, the SDE

$$dX_t = \mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t)\,dW_t$$

becomes

$$dX_t = \frac{1}{c}\mathbf{b}(X_t)\,dt + \frac{1}{\sqrt{c}}\boldsymbol{\sigma}(X_t)\,dW_t,$$

together with the initial condition $X_0 = x$. We will use such a change of the time scale when we study Brownian motors in Section **??**.

3.4 Itô's formula

In Chapter 2 we showed that to a diffusion process X_t we can associate a second order differential operator, the generator of the process. Consider the Itô stochastic differential equation

$$dX_t = \mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t)\,dW_t, \tag{3.45}$$

where for simplicity we have assumed that the coefficients are independent of time. X_t is a diffusion process with drift $\mathbf{b}(x)$ and diffusion matrix

$$\boldsymbol{\Sigma}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)^T. \tag{3.46}$$

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

$$\mathcal{L} = \mathbf{b}(x)\cdot\nabla + \frac{1}{2}\boldsymbol{\Sigma}(x) : D^2, \tag{3.47}$$

where D^2 denotes the Hessian matrix. The generator can also be written as

$$\mathcal{L} = \sum_{j=1}^{d} b_j(x)\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d}\Sigma_{ij}\frac{\partial^2}{\partial x_i \partial x_j}$$

Using now the generator we can write Itô's formula . This formula enables us to calculate the rate of change in time of functions $V : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ evaluated at the solution of an $\mathbb{R}^d$-valued SDE. First, recall that in the absence of noise the rate of change of V can be written as

$$\frac{d}{dt}V(t,x(t)) = \frac{\partial V}{\partial t}(t,x(t)) + \mathcal{A}V(t,x(t)), \tag{3.48}$$

where $x(t)$ is the solution of the ODE $\dot{x} = b(x)$ and $\mathcal{A}$ denotes the (backward) *Liouville operator*

$$\mathcal{A} = b(x) \cdot \nabla. \tag{3.49}$$

Let now X_t be the solution of (3.45) and assume that the white noise $\dot{W}$ is replaced by a smooth function $\zeta(t)$. Then the rate of change of $V(t, X_t)$ is given by the formula

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \mathcal{A}V(t, X_t) + \langle \nabla V(t, X_t), \sigma(X_t)\zeta(t) \rangle \,, \tag{3.50}$$

This formula is no longer valid when the equation (3.45) is driven by white noise and not a smooth function. In particular, the Leibniz formula (3.50) has to be modified by the addition of a drift term that accounts for the lack of smoothness of the noise: the Liouville operator $\mathcal{A}$ given by (3.49) in (3.50) has to be replaced by the generator $\mathcal{L}$ given by (3.47). We have already encountered this additional term in the previous chapter, when we derived the backward Kolmogorov equation. Formally we can write

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \mathcal{L}V(t, X_t) + \left\langle \nabla V(t, X_t), \sigma(X_t)\frac{dW}{dt} \right\rangle ,$$

The precise interpretation of the expression for the rate of change of V is in integrated form.

Lemma 3.7. (Itô's Formula) *Assume that the conditions of Theorem 3.6 hold. Let X_t be the solution of* (3.45) *and let $V \in C^2(\mathbb{R}^d)$. Then the process $V(X_t)$ satisfies*

$$\begin{aligned} V(t, X_t) &= V(X_0) + \int_0^t \frac{\partial V}{\partial s}(s, X_s)\, ds + \int_0^t \mathcal{L}V(s, X_s)\, ds \\ &\quad + \int_0^t \langle \nabla V(s, X_s), \sigma(X_s)\, dW_s \rangle \,. \end{aligned} \tag{3.51}$$

The presence of the additional term in the drift is not very surprising, in view of the fact that the Brownian differential scales like the square root of the differential in time: $\mathbb{E}(dW_t)^2 = dt$. In fact, we can write Itô's formula in the form

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \langle \nabla V(t, X_t), \dot{X}_t \rangle + \frac{1}{2}\langle \dot{X}_t, D^2V(t, X_t)\dot{X}_t \rangle, \tag{3.52}$$

or,

$$dV(t, X_t) = \frac{\partial V}{\partial t}\, dt + \sum_{i=1}^{d} \frac{\partial V}{\partial x_i}\, dX_i + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 V}{\partial x_i \partial x_j}\, dX_i dX_j, \tag{3.53}$$

where we have suppressed the argument (t, X_t) from the right hand side. When writing (3.53) we have used the convention $dW_i(t)\, dW_j(t) = \delta_{ij}\, dt,\ dW_i(t)\, dt = 0,\ i, j = 1, \ldots d$. Thus, we can think of (3.53) as a generalization of Leibniz' rule (3.50) where second order differentials are kept. We can check that, with the above convention, Itô's formula follows from (3.53).

The proof of Itô's formula is essentially the same as the proof of the validity of the backward Kolmogorov equation that we presented in Chapter 2, Theorem 2.7. Conversely, after having proved Itô's formula, then the backward Kolmogorov and Fokker-Planck (forward Kolmogorov) equations follow as corollaries. Let $\phi \in C^2(\mathbb{R}^d)$ and denote by X_t^x the solution of (3.45) with $X_0^x = x$. Consider the function

$$u(x, t) = \mathbb{E}\phi(X_t^x) := \mathbb{E}\big(\phi(X_t^x) | X_0^x = x\big), \tag{3.54}$$

where the expectation is with respect to all Brownian driving paths. We take the expectation in (3.51), use the martingale property of the stochastic integral, Equation (3.20), and differentiate with respect to time to obtain the backward Kolmogorov equation

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \tag{3.55}$$

together with the initial condition $u(x,0) = \phi(x)$. In this formal derivation we need to assume that the expectation $\mathbb{E}$ and the generator $\mathcal{L}$ commute; this can be justified since, as we have already seen, the expectation of a functional of the solution to the stochastic differential equation is given by the semigroup generated by $\mathcal{L}$: $(\mathcal{P}_t f)(x) = (e^{t\mathcal{L}} f)(x) = \mathbb{E} f(X_t^x)$.

The Feyman-Kac formula

Itô's formula can be used in order to obtain a probabilistic description of solutions to more general partial differential equations of parabolic type. Let X_t^x be a diffusion process with drift $b(\cdot)$, diffusion $\Sigma(\cdot) = \sigma\sigma^T(\cdot)$ and generator $\mathcal{L}$ with $X_0^x = x$ and let $f \in C_0^2(\mathbb{R}^d)$ and $V \in C(\mathbb{R}^d)$, bounded from below. Then the function

$$u(x,t) = \mathbb{E}\left(e^{-\int_0^t V(X_s^x)\,ds} f(X_t^x)\right), \tag{3.56}$$

is the solution to the initial value problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mathcal{L}u - Vu, & \text{(3.57a)} \\ u(0,x) &= f(x). & \text{(3.57b)} \end{aligned}$$

To derive this result, we introduce the variable $Y_t = \exp\left(-\int_0^t V(X_s)\,ds\right)$ we rewrite the SDE for X_t as

$$\begin{aligned} dX_t^x &= b(X_t^x)\,dt + \sigma(X_t^x)\,dW_t, \quad X_0^x = x, & \text{(3.58a)} \\ dY_t^x &= -V(X_t^x)\,dt, \quad Y_0^x = 0. & \text{(3.58b)} \end{aligned}$$

The process $\{X_t^x,\, Y_t^x\}$ is a diffusion process with generator

$$\mathcal{L}_{x,y} = \mathcal{L} - V(x)\frac{\partial}{\partial_y}.$$

We can write

$$\mathbb{E}\left(e^{-\int_0^t V(X_s^x)\,ds} f(X_t^x)\right) = \mathbb{E}(\phi(X_t^x, Y_t^x)),$$

where $\phi(x,y) = f(x)e^y$. We apply now Itô's formula to this function (or, equivalently, write the backward Kolmogorov equation for the function $u(x,y) = \mathbb{E}(\phi(X_t, Y_t))$) to obtain (3.57).

The representation formula (3.56) for the solution of the initial value problem (3.57) is called the *Feynman-Kac formula*. It is very useful both for the theoretical analysis of initial value problems for parabolic PDEs of the form (3.57) as well as for their numerical solution using a Monte Carlo approach, based on solving numerically (3.58). See Section **??**.

Derivation of the Fokker-Planck Equation

Starting from Itô's formula we can easily obtain the backward Kolmogorov equation (3.55). Using now the backward Kolmogorov equation and the fact that the Fokker-Planck operator is the L^2-adjoint of the generator, we can obtain the Fokker-Planck (forward Kolmogorov) equation that we will study in detail in Chapter 4. This line of argument provides us with an alternative, and perhaps simpler, derivation of the forward and backward Kolmogorov equations to the one presented in Section 2.5.

The Fokker-Planck operator corresponding to the stochastic differential equation (3.45) reads

$$\mathcal{L}^* \cdot = \nabla \cdot \big(- \mathbf{b}(x) \cdot + \frac{1}{2} \nabla \cdot \big(\mathbf{\Sigma} \cdot \big) \big). \tag{3.59}$$

We can derive this formula from the formula for the generator $\mathcal{L}$ of X_t and two integrations by parts:

$$\int_{\mathbb{R}^d} \mathcal{L} f h \, dx = \int_{\mathbb{R}^d} f \mathcal{L}^* h \, dx,$$

for all f, $h \in C_0^2(\mathbb{R}^d)$.

Proposition 3.8. *Let X_t denote the solution of the Itô SDE* (3.45) *and assume that the initial condition X_0 is a random variable, independent of the Brownian motion driving the SDE, with density $\rho_0(x)$. Assume that the law of the Markov process X_t has a density $\rho(x,t) \in C^{2,1}(\mathbb{R}^d \times (0,+\infty))$[3], then ρ is the solution of the initial value problem for the Fokker-Planck equation*

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho \quad \text{for } (x,t) \in \mathbb{R}^d \times (0,\infty), \tag{3.60a}$$

$$\rho = \rho_0 \quad \text{for } x \in \mathbb{R}^d \times \{0\}. \tag{3.60b}$$

Proof. Let $\mathbb{E}^\mu$ denote the expectation with respect to the product measure induced by the measure μ with density ρ_0 on X_0 and the Wiener measure of the Brownian motion that is driving the SDE. Averaging over the random initial conditions, distributed with density $\rho_0(x)$, we find

$$\begin{aligned} \mathbb{E}^\mu \phi(X_t) &= \int_{\mathbb{R}^d} \phi(x,t) \rho_0(x) \, dx \\ &= \int_{\mathbb{R}^d} (e^{\mathcal{L}t} \phi)(x) \rho_0(x) \, dx \\ &= \int_{\mathbb{R}^d} (e^{\mathcal{L}^* t} \rho_0)(x) \phi(x) \, dx. \end{aligned}$$

But since $\rho(x,t)$ is the density of X_t we also have

$$\mathbb{E}^\mu \phi(X_t) = \int_{\mathbb{R}^d} \rho(z,t) \phi(x) \, dx.$$

Equating these two expressions for the expectation at time t we obtain

$$\int_{\mathbb{R}^d} (e^{\mathcal{L}^* t} \rho_0)(x) \phi(x) \, dx = \int_{\mathbb{R}^d} \rho(x,t) \phi(x) \, dx.$$

[3] This is the case, for example, when the SDE has smooth coefficients and the diffusion matrix $\Sigma = \sigma \sigma^T$ is strictly positive definite.

We use a density argument so that the identity can be extended to all $\phi \in L^2(\mathbb{R}^d)$. Hence, from the above equation we deduce that

$$\rho(x,t) = \left(e^{\mathcal{L}^* t}\rho_0\right)(x).$$

Differentiation of the above equation gives (3.60a). Setting $t = 0$ gives the initial condition (3.60b). □

The chain rule for Stratonovich equations

For a Stratonovich stochastic differential equations the rules of standard calculus apply:

Proposition 3.9. *Let $X_t : \mathbb{R}^+ \mapsto \mathbb{R}^d$ be the solution of the Stratonovich SDE*

$$dX_t = \mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t) \circ dW_t, \tag{3.61}$$

where $\mathbf{b} : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$. Then the generator and Fokker-Planck operator of X_t are, respectively:

$$\mathcal{L}\cdot = \mathbf{b}\cdot\nabla\cdot + \frac{1}{2}\boldsymbol{\sigma}^T : \nabla(\boldsymbol{\sigma}^T\nabla\cdot) \tag{3.62}$$

and

$$\mathcal{L}^*\cdot = \nabla\cdot\big(-\mathbf{b}\cdot + \frac{1}{2}\boldsymbol{\sigma}\nabla\cdot(\boldsymbol{\sigma}^T\cdot)\big). \tag{3.63}$$

Furthermore, the Newton-Leibniz chain rule applies.

Proof. We will use the summation convention and also use the notation ∂_j for the partial derivative with respect to x_j. We use the formula for the Itô-to-Stratonovich correction, Equation (3.31), to write

$$\begin{aligned}
\mathcal{L} &= b_j\partial_j + \frac{1}{2}\sigma_{jk}\partial_j\sigma_{ik}\partial_i f + \frac{1}{2}\sigma_{ik}\sigma_{jk}\partial_i\partial_j f \\
&= b_j\partial_j f_j + \frac{1}{2}\sigma_{jk}\partial_j\big(\sigma_{ik}\partial_i f\big) \\
&= b\cdot\nabla + \frac{1}{2}\boldsymbol{\sigma}^T : \nabla\big(\boldsymbol{\sigma}^T\nabla f\big).
\end{aligned}$$

To obtain the formula for the Fokker-Planck operator, let f, h be two $C_0^2(\mathbb{R}^d)$ functions. We perform two integrations by parts to obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} \boldsymbol{\sigma}_{jk}\partial_j\big(\boldsymbol{\sigma}_{ik}\partial_i f\big)h\,dx &= \int_{\mathbb{R}^d} f\partial_i\Big(\boldsymbol{\sigma}_{ik}\partial_j\big(\boldsymbol{\sigma}_{jk}h\big)\Big)\,dx \\
&= \int_{\mathbb{R}^d} f\partial_i\Big(\boldsymbol{\sigma}_{ik}\big(\nabla\cdot\big(\boldsymbol{\sigma}^T h\big)\big)_k\Big)\,dx \\
&= \int_{\mathbb{R}^d} f\partial_i\Big(\boldsymbol{\sigma}\cdot\nabla\cdot\big(\boldsymbol{\sigma}^T h\big)\Big)_i\,dx \\
&= \int_{\mathbb{R}^d} f\nabla\cdot\Big(\boldsymbol{\sigma}\cdot\nabla\cdot\big(\boldsymbol{\sigma}^T h\big)\Big)\,dx.
\end{aligned}$$

To show that the Stratonovich SDE satisfies the standard chain rule, let $\mathbf{h} : \mathbb{R}^d \mapsto \mathbb{R}^d$ be an invertible map, let $y = \mathbf{h}(x)$ and let $x = \mathbf{h}^{-1}(y) =: \mathbf{g}(y)$. Let $\mathbf{J} = \nabla_x\mathbf{h}$ denote the Jacobian of the transformation. We introduce the notation

$$\widehat{\mathbf{b}}(y) = \mathbf{b}(\mathbf{g}(y)), \quad \widehat{\boldsymbol{\sigma}}(y) = \boldsymbol{\sigma}(\mathbf{g}(y)), \quad \widehat{\mathbf{J}}(y) = \mathbf{J}(\mathbf{g}(y)). \tag{3.64}$$

We need to show that, for $Y_t = \mathbf{h}(X_t)$,

$$dY_t = \widehat{\mathbf{J}}(Y_t)\Big(\widehat{\mathbf{b}}(Y_t)\,dt + \widehat{\boldsymbol{\sigma}}(Y_t) \circ dW_t\Big). \tag{3.65}$$

It is important to note that in this equation the (square root of the) diffusion matrix is $\widehat{\mathbf{J}}(y)\widehat{\boldsymbol{\sigma}}(y)$. In particular, the Stratonovich correction in (3.65) is (see (3.28))

$$b_\ell^S(y) = \frac{1}{2}\big(\widehat{J}_{j\rho}\widehat{\sigma}_{\rho k}\big)(y)\frac{\partial\big(\widehat{J}_{\ell m}\widehat{\sigma}_{mk}\big)}{\partial y_j}(y), \quad \ell = 1, \dots d, \tag{3.66}$$

or, using (3.31),

$$\mathbf{b}^S(y) = \frac{1}{2}\big[\nabla \cdot \big(\widehat{\mathbf{J}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{J}}^T\big) - \widehat{\mathbf{J}}\widehat{\boldsymbol{\sigma}}\nabla \cdot \big(\widehat{\boldsymbol{\sigma}}^T\widehat{\mathbf{J}}^T\big)\big](y).$$

To prove this, we first transform the Stratonovich SDE for X_t into an Itô SDE using (3.31). We then apply Itô's formula to calculate

$$\begin{aligned} dY_\ell &= \big(\mathcal{L}h_\ell\big)(X_t)\,dt + \big(\partial_i h_\ell \sigma_{ij}\big)(X_t)\,dW_j \\ &= J_{\ell i}(X_t)\big(b_i(X_t)\,dt + \sigma_{ij}(X_t)\,dW_j\big) \\ &\quad + \frac{1}{2}\sigma_{jk}(X_t)\partial_j\big(\sigma_{ik}(X_t)J_{\ell i}(X_t)\big)\,dt, \end{aligned} \tag{3.67}$$

for $\ell = 1, \dots d$. Equivalently, using (3.62),

$$\begin{aligned} dY_t &= \nabla_x \mathbf{h}(X_t)\big(\mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t)\,dW_t\big) \\ &\quad + \frac{1}{2}\boldsymbol{\sigma}^T(X_t) : \nabla\big(\boldsymbol{\sigma}^T(X_t)\nabla_x \mathbf{h}(X_t)\big)\,dt. \end{aligned} \tag{3.68}$$

Now we need to rewrite the righthand side of this equation as a function of y. This follows essentially from the inverse function theorem, and in particular the fact that the Jacobian of the inverse transformation (from y to x) is given by the inverse of the Jacobian J of the transformation $y = \mathbf{h}(x)$:

$$\widehat{J}(y) = \big(\nabla_y \mathbf{g}\big)^{-1}.$$

In particular,

$$\nabla_x f(x) = \widehat{J}(y)\,\nabla_y \widehat{f}(y). \tag{3.69}$$

For the first term, and using the notation (3.64), we have

$$\nabla_x \mathbf{h}(X_t)\big(\mathbf{b}(X_t)\,dt + \boldsymbol{\sigma}(X_t)\,dW_t\big) = \widehat{\mathbf{J}}(Y_t)\Big(\widehat{\mathbf{b}}(Y_t)\,dt + \widehat{\boldsymbol{\sigma}}(Y_t)\,dW_t\Big).$$

Now we need to show that the second term on the righthand side of (3.67) is equal to the Stratonovich correction $\mathbf{b}^S(y)$ in (3.65) that is given by (3.66). This follows by applying the chain rule (3.69) to the righthand sider of (3.68) and using (3.34). Equivalently, using (3.67):

$$\begin{aligned} dY_\ell &= \widehat{J}_{\ell i}(Y_t)\big(\widehat{b}_i(Y_t)\,dt + \widehat{\sigma}_{ij}(Y_t)\,dW_j\big) \\ &\quad + \frac{1}{2}\widehat{J}_{j\rho}\widehat{\sigma}_{\rho k}(Y_t)\partial_j\big(\widehat{\sigma}_{ik}(Y_t)\widehat{J}_{\ell i}(X_t)\big)\,dt \\ &= \widehat{J}_{\ell i}(Y_t)\big(\widehat{b}_i(Y_t)\,dt + \widehat{\sigma}_{ij}(Y_t) \circ dW_j\big). \end{aligned}$$

□

3.5 Examples of Stochastic Differential Equations

Brownian Motion

We consider a stochastic differential equation with no drift and constant diffusion coefficient:

$$dX_t = \sqrt{2\sigma}\, dW_t, \quad X_0 = x, \tag{3.70}$$

where the initial condition can be either deterministic or random, independent of the Brownian motion W_t. The solution is:

$$X_t = x + \sqrt{2\sigma} W_t.$$

This is just Brownian motion starting at x with diffusion coefficient σ.

Ornstein-Uhlenbeck Process

Adding a restoring force to (3.70) we obtain the SDE for the Ornstein-Uhlenbeck process that we have already encountered:

$$dX_t = -\alpha X_t\, dt + \sqrt{2\sigma}\, dW_t, \quad X(0) = x, \tag{3.71}$$

where, as in the previous example, the initial condition can be either deterministic or random, independent of the Brownian motion W_t. We solve this equation using the variation of constants formula:

$$X_t = e^{-\alpha t} x + \sqrt{2\sigma} \int_0^t e^{-\alpha(t-s)} dW_s. \tag{3.72}$$

We can use Itô's formula to obtain equations for the moments of the OU process.[4] The generator is:

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

We apply Itô's formula to the function $f(x) = x^n$ to obtain:

$$\begin{aligned} dX_t^n &= \mathcal{L} X_t^n\, dt + \sqrt{2\sigma} \frac{d}{dx} X_t^n\, dW \\ &= -\alpha n X_t^n\, dt + \sigma n(n-1) X_t^{n-2}\, dt + n\sqrt{2\sigma} X_t^{n-1}\, dW. \end{aligned}$$

Consequently:

$$X_t^n = x^n + \int_0^t \left(-\alpha n X_t^n + \sigma n(n-1) X_t^{n-2} \right)\, dt + n\sqrt{2\sigma} \int_0^t X_t^{n-1}\, dW_t.$$

By taking the expectation in the above equation and using the fact that the stochastic integral is a martingale, in particular (3.19), we obtain an equation for the moments $M_n(t) = \mathbb{E} X_t^n$ of the OU process for $n \geqslant 2$:

$$M_n(t) = M_n(0) + \int_0^t \left(-\alpha n M_n(s) + \sigma n(n-1) M_{n-2}(s) \right)\, ds,$$

[4] In Section 4.2 we will redo this calculation using the Fokker-Planck equation.

where we have considered random initial conditions distributed a according to a distribution $\rho_0(x)$ with finite moments.

A variant of the the SDE (3.71) is the *mean reverting Ornstein-Uhlenbeck process*

$$dX_t = \big(\mu - \alpha X_t\big)\, dt + \sqrt{2\sigma}\, dW_t, \quad X(0) = x. \tag{3.73}$$

We can use the variation of constants formula to solve this SDE; see exercise 7.

Geometric Brownian motion

(See also Section 4.2) Consider the following scalar linear SDE with multiplicative noise

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t, \quad X_0 = x, \tag{3.74}$$

where we use the Itô interpretation of the stochastic differential. The solution to this equation is known as the *geometric Brownian motion*. We can think of it as a very simple model of population dynamics in a fluctuating environment: We can obtain (3.74) from the exponential growth (or decay) model

$$\frac{dX_t}{dt} = \mu(t) X_t. \tag{3.75}$$

We assume that there are fluctuations (or uncertainty) in the growth rate $\mu(t)$. Modelling the uncertainty as Gaussian white noise we can write

$$\mu(t) = \mu + \sigma\xi(t), \tag{3.76}$$

with $\xi(t)$ denoting the white noise process $\frac{dW_t}{dt}$. Substituting now (3.76) into (3.75) we obtain the SDE for the geometric Brownian motion. The multiplicative noise in (3.74) is due to our lack of complete knowledge of the growth parameter $\mu(t)$. This is a general feature in the modelling of physical, biological systems using SDEs: multiplicative noise is quite often associated with fluctuations or uncertainty in the parameters of the model.

The generator of the geometric Brownian motion is

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

The solution of (3.74) is

$$X(t) = x \exp\left(\big(\mu - \frac{\sigma^2}{2}\big)t + \sigma W(t) \right). \tag{3.77}$$

To derive this formula, we apply Itô's formula to the function $f(x) = \log(x)$:

$$\begin{aligned} d\log(X_t) &= \mathcal{L}\Big(\log(X_t)\Big)\, dt + \sigma x \frac{d}{dx}\log(X_t)\, dW_t \\ &= \left(\mu X_t \frac{1}{X_t} + \frac{\sigma^2 X_t^2}{2}\left(-\frac{1}{X_t^2}\right)\right) dt + \sigma\, dW_t \\ &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dW_t. \end{aligned}$$

Consequently:

$$\log\left(\frac{X_t}{X_0}\right) = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W(t),$$

from which (3.77) follows.

Notice that if we interpret the stochastic differential in (3.74) in the Stratonovich sense, then the solution is no longer given by (3.77). To see this, first note that from (3.28) it follows that the Stratonovich SDE

$$dX_t = \mu X_t\, dt + \sigma X_t \circ dW_t, \quad X_0 = x, \tag{3.78}$$

is equivalent to the Itô SDE

$$dX_t = \left(\mu + \frac{1}{2}\sigma^2\right) X_t\, dt + \sigma X_t\, dW_t, \quad X_0 = x. \tag{3.79}$$

Consequently, from (3.77) and replacing μ with $\mu + \frac{1}{2}\sigma^2$ we conclude that the solution of (3.78) is

$$X(t) = x\exp(\mu t + \sigma W(t)). \tag{3.80}$$

Comparing (3.77) with (3.80) we immediately see that the Itô and Stratonovich interpretations of the stochastic integral lead to SDEs with different properties. For example, from (3.77) we observe that the noise in (3.74) can change the qualitative behaviour of the solution: for $\mu > 0$ the solution to the deterministic equation ($\sigma = 0$) increases exponentially as $t \to +\infty$, whereas for the solution of the stochastic equation, it can be shown that it converges to 0 with probability one, provided that $\mu - \frac{\sigma^2}{2}) < 0$.

We now present a list of stochastic differential equations that appear in applications.

- The Cox-Ingersoll-Ross equation:

$$dX_t = \alpha(b - X_t)\, dt + \sigma\sqrt{X_t}\, dW_t, \tag{3.81}$$

  where α, b, σ are positive constants.

- The stochastic Verhulst equation (population dynamics):

$$dX_t = (\lambda X_t - X_t^2)\, dt + \sigma X_t\, dW_t. \tag{3.82}$$

- Coupled Lotka-Volterra stochastic equations:

$$dX_i(t) = X_i(t)\left(a_i + \sum_{j=1}^{d} b_{ij} X_j(t)\right) dt + \sigma_i X_i(t)\, dW_i(t). \tag{3.83}$$

- Protein kinetics:

$$dX_t = (\alpha - X_t + \lambda X_t(1 - X_t))\, dt + \sigma X_t(1 - X_t) \circ dW_t. \tag{3.84}$$

- Dynamics of a tracer particle (turbulent diffusion):

$$dX_t = u(X_t, t)\, dt + \sigma\, dW_t, \quad \nabla \cdot u(x, t) = 0. \tag{3.85}$$

- Josephson junction (pendulum with friction and noise):

$$\ddot{\phi}_t = -\sin(\phi_t) - \gamma\dot{\phi}_t + \sqrt{2\gamma\beta^{-1}}\,\dot{W}_t. \tag{3.86}$$

- Noisy Duffing oscillator (stochastic resonance)

$$\ddot{X}_t = -\beta X_t - \alpha X_t^3 - \gamma\dot{X}_t + A\cos(\omega t) + \sigma\,\dot{W}_t. \tag{3.87}$$

- The stochastic heat equation that we can write formally as

$$\partial_t u = \partial_x^2 u + \partial_t W(x,t), \tag{3.88}$$

where $W(x,t)$ denotes an infinite dimensional Brownian motion. We can consider (3.88) on $[0,1]$ with Dirichlet boundary conditions. We can represent $W(x,t)$ as a Fourier series:

$$W(x,t) = \sum_{k=1}^{+\infty} e_k(x)W_k(t), \tag{3.89}$$

where $W_k(t)$, $k = 1, \cdots + \infty$ are one dimensional independent Brownian motions and $\{e_k(x)\}_{k=1}^{+\infty}$ is the standard orthonormal basis in $L^2(0,1)$ with Dirichlet boundary conditions, i.e. $e_k(x) = \sin(2\pi kx)$, $k = 1, \cdots + \infty$.

3.6 Lamperti's Transformation and Girsanov's Theorem

Stochastic differential equations with a nontrivial drift and multiplicative noise are hard to analyze, in particular in dimensions higher than one. In this section we present two techniques that enable us to map equations with multiplicative noise to equations with additive noise, and to map equations with a nonconstant drift to equations with no drift and multiplicative noise. The first technique, Lamperti's transformation works mostly in one dimension, whereas the second technique, Girsanov's theorem is applicable in arbitrary, even infinite, dimensions.

Lamperti's transformation

For stochastic differential equations in one dimension it is possible to map multiplicative noise to additive noise through a generalisation of the method that we used in order to obtain the solution of the equation for geometric Brownian motion. Consider a one dimensional Itô SDE with multiplicative noise

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t. \tag{3.90}$$

We ask whether there exists a transformation $z = h(x)$ that maps (3.90) into an SDE with additive noise. We apply Itô's formula to obtain

$$dZ_t = \mathcal{L}h(X_t)\,dt + h'(X_t)\sigma(X_t)\,dW_t,$$

where $\mathcal{L}$ denotes the generator of X_t. In order to obtain an SDE with unit diffusion coefficient we need to impose the condition

$$h'(x)\sigma(x) = 1,$$

from which we deduce that

$$h(x) = \int_{x_0}^{x} \frac{1}{\sigma(x)}\, dx, \tag{3.91}$$

where x_0 is arbitrary. We have that

$$\mathcal{L}h(x) = \frac{b(x)}{\sigma(x)} - \frac{1}{2}\sigma'(x).$$

Consequently, the transformed SDE has the form

$$dY_t = b_Y(Y_t)\, dt + dW_t \tag{3.92}$$

with

$$b_Y(y) = \frac{b(h^{-1}(y))}{\sigma(h^{-1}(y))} - \frac{1}{2}\sigma'(h^{-1}(y)).$$

This is called the *Lamperti transformation*. As an application of this transformation, consider the Cox-Ingersoll-Ross equation (3.81)

$$dX_t = (\mu - \alpha X_t)\, dt + \sigma\sqrt{X_t}\, dW_t, \quad X_0 = x > 0.$$

From (3.91) we deduce that

$$h(x) = \frac{2}{\sigma}\sqrt{x}.$$

The generator of this process is

$$\mathcal{L} = (\mu - \alpha x)\frac{d}{dx} + \frac{\sigma^2}{2}x\frac{d^2}{dx^2}.$$

We have that

$$\mathcal{L}h(x) = \left(\frac{\mu}{\sigma} - \frac{\sigma}{4}\right)x^{-1/2} - \frac{\alpha}{\sigma}x^{1/2}.$$

The CIR equation becomes, for $Y_t = \frac{2}{\sigma}\sqrt{X_t}$,

$$\begin{aligned} dY_t &= \left(\frac{\mu}{\sigma} - \frac{\sigma}{4}\right)\frac{1}{\sqrt{X_t}}\, dt - \frac{\alpha}{\sigma}\sqrt{X_t}\, dt + dW_t \\ &= \left[\left(\frac{2\mu}{\sigma^2} - \frac{1}{2}\right)\frac{1}{Y_t} - \frac{\alpha}{2}Y_t\right] dt + dW_t. \end{aligned}$$

When $\mu = \frac{\sigma^2}{4}$ the equation above becomes the Ornstein-Uhlenbeck process for Y_t.

Apart from being a useful tool for obtaining solutions to one dimensional SDEs with multiplicative noise, it is also often used in statistical inference for SDEs. See Section **??** and the discussion in Section **??**. Such a transformation does not exist for arbitrary SDEs in higher dimensions. In particular, it is not possible, in general, to transform a multidimensional Itô SDE with multiplicative noise to an SDE with additive noise; see Exercise 8.

Girsanov's theorem

Conversely, it is also sometimes possible to remove the drift term from an SDE and obtain an equation where only (multiplicative) noise is present. Consider first the following one dimensional SDE:

$$dX_t = b(X_t)\,dt + dW_t. \tag{3.93}$$

We introduce the following functional of X_t

$$M_t = \exp\left(-\frac{1}{2}\int_0^t b^2(X_s)\,ds + \int_0^t b(X_s)\,dW_s\right). \tag{3.94}$$

We can write $M_t = e^{-Y_t}$ where Y_t is the solution to the SDE

$$M_t = e^{-Y_t}, \quad dY_t = \frac{1}{2}b^2(X_t)\,dt + b(X_t)\,dW_t,\ Y_0 = 0. \tag{3.95}$$

We can now apply Itô's formula to obtain the equation

$$dM_t = -M_t b(X_t)\,dW_t. \tag{3.96}$$

Notice that this is a stochastic differential equation without drift.

In fact, under appropriate conditions on the drift $b(\cdot)$, it is possible to show that the law of the process X_t, denoted by $\mathbb{P}$, which is a probability measure over the space of continuous functions, is absolutely continuous with respect to the Wiener measure $\mathbb{P}_W$, the law of the Brownian motion W_t. The *Radon-Nikodym* derivative between these two measures is the inverse of the stochastic process M_t given in (3.95):

$$\frac{d\mathbb{P}}{d\mathbb{P}_W}(X_t) = \exp\left(\frac{1}{2}\int_0^t b^2(X_s)\,ds + \int_0^t b(X_s)\,dW_s\right). \tag{3.97}$$

This is a form of *Girsanov's theorem* . Using now equation (3.93) we can rewrite (3.97) as

$$\frac{d\mathbb{P}}{d\mathbb{P}_W}(X_t) = \exp\left(\int_0^t b(X_s)\,dX_s - \frac{1}{2}\int_0^t |b(X_s)|^2\,ds\right). \tag{3.98}$$

The Girsanov transformation (3.96) or (3.98) enables us to "compare" the process X_t with the Brownian motion W_t. This is a very useful result when the drift function $b(\cdot)$ in (3.93) is known up to parameters that we want to estimate from observations. In this context, the Radon-Nikodym derivative in (3.98) becomes the *likelihood function* . We will study the problem of maximum likelihood parameter estimation for SDEs in Section **??**.

3.7 Linear Stochastic Differential Equations

In this section we study linear SDEs in arbitrary finite dimensions. Let A, $Q \in R^{d\times d}$ be positive definite and positive semidefinite matrices, respectively and let $W(t)$ be a standard Brownian motion in $\mathbb{R}^d$. We will consider the SDE[5]

$$dX(t) = -AX(t)\,dt + \sigma\,dW(t), \tag{3.99}$$

[5]We can also consider the case where $\sigma \in R^{d\times m}$ with $n \neq m$, i.e. when the SDE is driven by an m dimensional Brownian motion.

or, componentwise,

$$dX_i(t) = -\sum_{j=1}^{d} A_{ij} X_j(t) + \sum_{j=1}^{d} \sigma_{ij}\, dW_j(t), \quad i = 1, \dots d,$$

The initial conditions $X(0) = x$ can be taken to be either deterministic or random. The generator of the Markov process $X(t)$ is

$$\mathcal{L} = -Ax \cdot \nabla + \frac{1}{2}\Sigma : D^2,$$

where $\Sigma = \sigma\sigma^T$. The corresponding Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = \nabla \cdot (Axp) + \frac{1}{2}\nabla \cdot \big(\Sigma \nabla p\big). \tag{3.100}$$

The solution of (3.99) is

$$X(t) = e^{-At}x + \int_0^t e^{-A(t-s)}\sigma\, dW(s). \tag{3.101}$$

We use the fact that the stochastic integral is a martingale to calculate the expectation of the process $X(t)$:

$$\mu(t) := \mathbb{E}X(t) = e^{-At}\mathbb{E}x.$$

To simplify the formulas below we will set $\mu(t) = 0$. This is with no loss of generality, since we can define the new process $Y(t) = X(t) - \mu(t)$ that is mean zero.

We think of $X(t) \in \mathbb{R}^{d\times 1}$ as a column vector. Consequently, $X^T(t) \in \mathbb{R}^{1\times n}$ is a row vector. The autocorrelation matrix is

$$R(t,s) = \mathbb{E}(X(t)X^T(s)) = \mathbb{E}(X(t) \otimes X_s).$$

Componentwise:

$$R_{ij}(t,s) = \mathbb{E}(X_i(t)X_j(s)).$$

We will denote the covariance matrix of the initial conditions x by R_0, $\mathbb{E}xx^T =: R_0$.

Proposition 3.10. *The autocorrelation matrix of the process $X(t)$ is given by*

$$R(t,s) = e^{-At}\left(R_0 + \int_0^{\min(t,s)} e^{A\rho}\Sigma e^{A^T\rho}\, d\rho\right) e^{-A^T s}. \tag{3.102}$$

Furthermore, the variance at time t, $\Sigma(t) := R(t,t)$ satisfies the differential equation

$$\frac{d\Sigma(t)}{dt} = -A\Sigma(t) - \Sigma(t)A^T + \Sigma. \tag{3.103}$$

The steady state variance Σ_∞ is the solution of the equation

$$A\Sigma_\infty + \Sigma_\infty A^T = \Sigma. \tag{3.104}$$

The solution to this equation is

$$\Sigma_\infty = \int_0^{+\infty} e^{A\rho}\Sigma e^{A^T\rho}\, d\rho \tag{3.105}$$

The invariant distribution of (3.99) *is Gaussian with mean* 0 *and variance* Σ_∞.

Proof. We will use the notation $S_t = e^{-At}$ and $B_t = QW(t)$. The solution (3.101) can be written in the form

$$X_t = S_t x + \int_0^t S_{t-s}\, dB_s.$$

Consequently, and using the properties of the stochastic integral:

$$\begin{aligned} R(t,s) &= \mathbb{E}\Big((S_t x)(S_s x)^T\Big) + \mathbb{E}\int_0^t \int_0^s \Big(S_{t-\ell}\, dB_\ell\Big)\Big(S_{s-\rho}\, dB_\rho\Big)^T \\ &=: S_t \mathbb{E}(xx^T) S_s + \int_0^t \int_0^s S_{t-\ell} Q \mathbb{E}\Big(dW_\ell dW_\rho^T\Big) Q^T S_{s-\rho}^T \\ &= S_t R_0 S_s + \int_0^t \int_0^s S_{t-\ell} Q \delta(\ell-\rho) Q^T S_{s-\rho}^T\, d\ell d\rho \\ &= S_t R_0 S_s + \int_0^{\min(t,s)} S_{t-\rho} \Sigma S_{s-\rho}^T\, d\rho \\ &= e^{-At}\left(R_0 + \int_0^{\min(t,s)} e^{A\rho}\Sigma e^{A^T\rho}\, d\rho\right) e^{-A^T s}. \end{aligned}$$

From (3.102) it follows that, with $\Sigma_t := R(t,t)$,

$$\Sigma_t = e^{-At}\left(R_0 + \int_0^t e^{A\rho}\Sigma e^{A^T\rho}\, d\rho\right) e^{-A^T t}. \tag{3.106}$$

Upon differentiating this equation we obtain the equation for the variance:

$$\frac{d\Sigma_t}{dt} = -A\Sigma_t - \Sigma_t A^T + \Sigma,$$

with $\Sigma_0 = R_0$. We now set the left hand side of the above equation to 0 to obtain the equation for the steady state variance Σ_∞:

$$A\Sigma_\infty + \Sigma_\infty A^T = \Sigma.$$

□

Equation (3.104) is an example of a *Lyapunov equation*. When Σ is strictly positive definite then it is possible to show that (3.105) is a well defined unique solution of (3.104); see Exercise 4. The situation becomes more complicated when Σ is positive semidefinite. See the discussion in Section 3.8.

We can use Proposition 3.10 us to solve the Fokker-Planck equation (3.100) with initial conditions

$$p(\mathbf{x}, t|\mathbf{x_0}, 0) = \delta(\mathbf{x} - \mathbf{x_0}). \tag{3.107}$$

The solution of (3.99) with $X(0) = x_0$ deterministic is a Gaussian process with mean $\mu(t) = e^{-At}$ and variance Σ_t given by (3.106). Consequently, the solution of the Fokker-Planck equation with initial conditions (3.107):

$$p(x,t|x_0,0) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma(t))}} \exp\left(-\frac{1}{2}\left(x - e^{-At}x_0\right)^T \Sigma^{-1}(t)\left(x - e^{-At}x_0\right)\right). \tag{3.108}$$

This result can also be obtained by using the Fourier transform. See Exercise 6.

3.8 Discussion and Bibliography

Stochastic differential equations and stochastic calculus are treated in many textbooks. See, for example [5, 25, 26, 42, 66, 75, 20] as well as [46, 83, 90, 91]. The reader is strongly encouraged to study carefully the proofs of the construction of the Itô integral, Itô's formula and of the basic existence and uniqueness theorem for stochastic differential equations from the above references.

In this book, we consider stochastic equations in finite dimensions. There is also a very well developed theory of stochastic partial differential equations; see [78]. Proposition 3.2 is taken from [75, Exer. 3.10]. Theorem 3.6 is taken from [75] where the proof be found; see also [49, Ch. 21] and [79, Thm. 3.1.1]. The assumption that the drift and diffusion coefficient is globally Lipschitz can be weakened when a priori bounds on the solution can be found. This is the case, for example, when a *Lyapunov function* can be constructed. See, e.g. [66]. Several examples of stochastic equations that can be solved analytically can be found in [27].

The derivation of Itô's formula is similar to the derivation of the backward Kolmogorov equation that was presented in Chapter 2, in particular Theorem 2.7. The proof can be found in all of the books on stochastic differential equations mentioned above. The Feynman-Kac formula, which his a generalization of Itô's formula, was first derived in the context of quantum mechanics, providing a path integral solution to the Schrödinger equation. Path integrals and functional integration are studied in detail in [94, 31].

The concept of the solution used in 3.6 is that of a *strong solution*. It is also possible to define weak solutions of SDEs, in which case the Brownian motion that is driving the SDE (3.35) is not specified a priori. Roughly speaking, the concept of a weak solution for a stochastic equation is related to the solution of the corresponding Fokker-Planck equation, i.e. the law of the process X_t. One could argue that in some applications such as physics and chemistry the concept of a weak solution is more natural than that of a strong solution, since in these applications one is usually interested in probability distribution functions rather than the actual paths of a diffusion process.

It is not very difficult to construct examples of stochastic differential equations that have weak but not strong solutions. A standard example is that of the *Tanaka equation*

$$dX_t = \operatorname{sgn}(X_t)\, dW_t, \tag{3.109}$$

where sgn denotes the sign function. One can show that this equation has no strong solution (which is not a big surprise, since the assumptions of Theorem 3.6 are not satisfied), but that it does have a unique weak solution. See [75] for the details. The fact that (3.109) is solvable in the weak sense follows from the fact that the Fokker-Planck equation for (3.109) is simply the heat equation, i.e. the Fokker-Planck equation for Brownian motion. Hence, any Brownian motion is a weak solution of the Tanaka equation.

Itô's formula also holds for a particular class of random times, the Markov or *stopping times*. Roughly speaking, a stopping time is a random time τ for which we can decide whether it has already occurred or not based on the information that is available to us, i.e. the solution of our stochastic differential equation up to the present time.[6] Let X_t be a diffusion process with generator $\mathcal{L}$, starting at x. We will denote the

[6]More precisely: Let $\{\mathcal{F}_t\}$ be a filtration. A function $\tau\,\Omega \mapsto [0, +\infty]$ is a called a (strict) stopping time with respect to $\{\mathcal{F}_t\}$ provided that

$$\big\{\omega\,;\, \tau(\omega) \leqslant t\big\} \in \mathcal{F}_t \quad \text{for all } t \geqslant 0.$$

expectation with respect to the law of this process by $\mathbb{E}^x$. Let furthermore $f \in C_0^2(\mathbb{R}^d)$ and let τ be a stopping time with $\mathbb{E}\tau < +\infty$. *Dynkin's formula* reads

$$\mathbb{E}^x f(X_\tau) = f(x) + \mathbb{E}^x \Big[\int_0^\tau \mathcal{L} f(X_s)\, ds \Big]. \tag{3.110}$$

The derivation of this formula and generalizations can be found in [75, Ch. 7, 8]. Dynkin's formula and the Feynman-Kac formula are very useful for deriving partial differential equations for interesting functionals of the diffusion process X_t. Details, in particular for diffusion processes in one dimension, can be found in [47].

It is not possible, in general, to transform a stochastic differential equation with multiplicative noise to one with additive noise in dimensions higher than one. In other words, the Lamperti transformation exists, unless additional assumptions on the diffusion matrix are imposed, only in one dimension. Extensions of this transformation to higher dimensions are discussed in [2]. See also Exercise 8.

Girsanov's theorem is one of the fundamental results in stochastic analysis and is presented in all the standard textbooks, e.g. [46, 83, 43]. A very detailed discussion of Girsanov's theorem and of its connection with the construction of the likelihood function for diffusion processes can be found in [56, Ch. 7]. A form of Girsanov's theorem that is very useful in statistical inference for diffusion processes is the following [51, Sec. 1.1.4], [41, Sec. 1.12]: consider the two equations

$$dX_t = b_1(X_t)\, dt + \sigma(X_t)\, dW_t, \quad X_0 = x^1, \quad t \in [0,T], \tag{3.111a}$$

$$dX_t = b_2(X_t)\, dt + \sigma(X_t)\, dW_t, \quad X_0 = x^2, \quad t \in [0,T], \tag{3.111b}$$

where $\sigma(x) > 0$. We assume that we have existence and uniqueness of strong solutions for both SDEs. Assume that x^1 and x^2 are random variables with densities $f_1(x)$ and $f_2(x)$ with respect to the Lebesgue measure which have the same support, or nonrandom and equal to the same constant. Let P_1 and P_2 denote the laws of these two SDEs. Then these two measures are equivalent[7] and their Radon-Nikodym derivative is

$$\frac{dP_2}{dP_1}(X) = \frac{f_2(X_0)}{f_1(X_0)} \exp\left(\int_0^T \frac{b_2(X_t) - b_1(X_t)}{\sigma^2(X_t)}\, dX_t - \frac{1}{2} \int_0^T \frac{b_2^2(X_t) - b_1^2(X_t)}{\sigma^2(X_t)}\, dt \right). \tag{3.112}$$

Linear stochastic differential equations are studied in most textbooks on stochastic differential equations and stochastic processes, see, e.g. [5, Ch. 8], [86, Ch. 6].

3.9 Exercises

1. Calculate all moments of the geometric Brownian motion (3.74) for the Itô and Stratonovich interpretations of the noise in the equation.

2. Prove rigorously (3.24) by keeping careful track of the error terms.

3. Use Itô's formula to obtain the Feynman-Kac formula (3.56). (*Hint:* Define $Y_t = f(X_t)$ and $Z_t = \exp\left(-\int_0^t q(X_s)\, ds \right)$. Calculate then $d(Y_t\, Z_t)$ and use this to calculate $\mathbb{E}(Y_t\, Z_t)$).

[7] Two probability measures $P_1,\ P_2$ are equivalent on a σ-field $\mathcal{G}$ if $P_1(A) = 0 \implies P_2(A) = 0$ for all $A \in \mathcal{G}$. In this case $\frac{dP_2}{dP_1}$ and $\frac{dP_1}{dP_2}$ exist.

4. Consider Equation (3.104) and assume that all the eigenvalues of the matrix A have positive real parts. Assume furthermore that Σ is symmetric and strictly positive definite. Show that there exists a unique positive solution to the Lyapunov equation (3.104).

5. Obtain (3.103) using Itô's formula.

6. Solve the Fokker-Planck equation (3.100) with initial conditions (3.107) (hint: take the Fourier transform and use the fact that the Fourier transform of a Gaussian function is Gaussian). Assume that the matrices A and Σ commute. Calculate the stationary autocorrelation matrix using the formula

$$\mathbb{E}(X_0^T X_t) = \int \int x_0^T x p(x,t|x_0,0) p_s(x_0)\, dx dx_0$$

and Gaussian integration.

7. Consider the mean reverting Ornstein-Uhlenbeck process

$$dX_t = \big(\mu - \alpha X_t\big)\, dt + \sqrt{2\lambda}\, dW_t, \quad X(0) = x. \tag{3.113}$$

Obtain a solution to this equation. Write down the generator and the Fokker-Planck equation. Obtain the transition probability density, the stationary distribution and formulas for the moments of this process.

8. Consider the two-dimensional Itô stochastic equation

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \tag{3.114}$$

where W_t is a standard two-dimensional Brownian motion and $\sigma(x) \in \mathbb{R}^{2\times 2}$ a uniformly positive definite matrix. Investigate whether a generalisation of the Lamperti transformation (3.91) to two dimensions exist, i.e. whether there exists a transformation that maps (3.114) to an SDE with additive noise. In particular, find conditions on $\sigma(x)$ so that such a transformation exists. What is the analogue of this transformation at the level of the backward and forward Kolmogorov equations? Is such a transformation still possible when we consider a diffusion process in a bounded domain with reflecting, absorbing or periodic boundary conditions?

9. (See [38, Ch. 6])

   (a) Consider the Itô equation

$$dX_t = f(X_t)\, dt + \sigma g(X_t)\, dW_t. \tag{3.115}$$

Define

$$Z(x) = \frac{f(x)}{g(x)} - \frac{1}{2}\frac{dg}{dx}(x) \quad \text{and} \quad \theta(x) = -\frac{1}{\frac{dZ}{dx}(x)}\frac{d}{dx}\left(g(x)\frac{dZ}{dx}(x)\right).$$

Assume that

$$\theta(x) = \text{const} \equiv \theta. \tag{3.116}$$

Define the diffusion process Y_t

$$Y_t = \exp(\theta B(X_t)), \quad B(x) = \int^x \frac{1}{g(z)}\, dz.$$

Show that when (3.116) is satisfied, Y_t is the solution of the linear SDE

$$dY_t = (\alpha + \beta Y_t)\,dt + (\gamma + \sigma Y_t)\,dW_t. \tag{3.117}$$

(b) Apply this transformation to obtain the solution and the transition probability density of the Stratonovich equation

$$dX_t = -\frac{1}{2\sqrt{2}}\tanh(2\sqrt{2}X_t)\,dt + \frac{\sigma}{4}\operatorname{sech}(2\sqrt{2}X_t) \circ dW_t. \tag{3.118}$$

(c) Do the same for the Verhulst SDE

$$dX_t = (\lambda X_t - X_t^2)\,dt + \sigma X_t\,dW_t. \tag{3.119}$$

Chapter 4

The Fokker-Planck Equation

In Chapter 2 we derived the backward and forward (Fokker-Planck) Kolmogorov equations.[1] The Fokker-Planck equation enables us to calculate the transition probability density, which we can use in order to calculate the expectation value of observables of a diffusion process. In this chapter we study various properties of this equation such as existence and uniqueness of solutions, long time asymptotics, boundary conditions and spectral properties of the Fokker-Planck operator. We also study in some detail various examples of diffusion processes and of the associated Fokker-Palnck equation. We will restrict attention to time-homogeneous diffusion processes, for which the drift and diffusion coefficients do not depend on time.

In Section 4.1 we study various basic properties of the Fokker-Planck equation, including existence and uniqueness of solutions and boundary conditions. In Section 4.2 we present some examples of diffusion processes and use the corresponding Fokker-Planck equation in order to calculate statistical quantities such as moments. In Section 4.3 we study diffusion processes in one dimension. In Section 4.4 we study the Ornstein-Uhlenbeck process and we study the spectral properties of the corresponding Fokker-Planck operator. In Section 4.5 we study stochastic processes whose drift is given by the gradient of a scalar function, the so-called Smoluchowski equation . In Section 4.6 we study properties of the Fokker-Planck equation corresponding to reversible diffusions. In Section 4.7 we solve the Fokker-Planck equation for a reversible diffusion using eigenfunction expansions. In Section 4.8 we introduce very briefly Markov Chain Monte Carlo techniques. In Section 4.9 we study the connection between the Fokker-Planck operator, the generator of a diffusion process and the Schrödinger operator. Discussion and bibliographical remarks are included in Section 4.10. Exercises can be found in Section 4.11.

4.1 Basic properties of the Fokker-Planck equation

We consider a time-homogeneous diffusion process X_t on $\mathbb{R}^d$ with drift vector $\mathbf{b}(x)$ and diffusion matrix $\mathbf{\Sigma}(x)$. We assume that the initial condition X_0 is a random variable with probability density function $\rho_0(x)$. The transition probability density $p(x,t)$, if it exists and is a $C^{2,1}(\mathbb{R}^d \times \mathbb{R}^+)$ function, is the solution of the

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

initial value problem for the Fokker-Planck (backward Kolmogorov) equation that we derived in Chapter 2

$$\begin{aligned} \frac{\partial p}{\partial t} &= \nabla \cdot \Big(-\mathbf{b}(x)p + \frac{1}{2}\nabla \cdot \big(\mathbf{\Sigma}(x)p\big)\Big) && \text{(4.1a)} \\ &= -\sum_{j=1}^{d} \frac{\partial}{\partial x_j}(b_i(x)p) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(\Sigma_{ij}(x)p), \\ p(x,0) &= \rho_0(x). && \text{(4.1b)} \end{aligned}$$

The Fokker-Planck equation (4.1a) can be written in equivalent forms that are often useful. First, we can rewrite it in the form

$$\frac{\partial p}{\partial t} = \nabla \cdot \big(\widehat{\mathbf{b}}(x)p\big) + \frac{1}{2}\nabla \cdot \big(\mathbf{\Sigma}(x)\nabla p\big). \tag{4.2}$$

with

$$\widehat{\mathbf{b}}(x) = -\mathbf{b}(x) + \frac{1}{2}\nabla \cdot \mathbf{\Sigma}(x). \tag{4.3}$$

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

$$\frac{\partial p}{\partial t} = \frac{1}{2}\mathbf{\Sigma}(x) : D^2 p + \tilde{\mathbf{b}}(x) \cdot \nabla p + c(x)p, \tag{4.4}$$

where

$$\tilde{b}(x) = -\mathbf{b}(x) + \frac{1}{2}\nabla \cdot \mathbf{\Sigma}(x), \quad c(x) = -\nabla \cdot \mathbf{b}(x) + \frac{1}{2}\nabla \cdot (\nabla \cdot \mathbf{\Sigma})(x). \tag{4.5}$$

By definition (see equation (2.62)), the diffusion matrix is always symmetric and nonnegative. We will assume that it is uniformly positive definite: there exists a constant $\alpha > 0$ such that

$$\langle \xi, \mathbf{\Sigma}(x)\xi \rangle \geqslant \alpha \|\xi\|^2, \quad \forall\, \xi \in \mathbb{R}^d, \tag{4.6}$$

uniformly in $x \in \mathbb{R}^d$. We will refer to this as the uniform ellipticity assumption . This assumption is sufficient to guarantee the existence of the transition probability density; see the discussion in Section **??**.

Furthermore, we will assume that the coefficients in (4.4) are smooth and that they satisfy the growth conditions

$$\|\mathbf{\Sigma}(x)\| \leqslant M, \ \|\widehat{\widehat{b}}(x)\| \leqslant M(1 + \|x\|), \ \|c(x)\| \leqslant M(1 + \|x\|^2). \tag{4.7}$$

Definition 4.1. *We will call a solution to the initial value problem for the Fokker-Planck equation* (4.4) *a classical solution if:*

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

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

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

iii. $\lim_{t \to 0} p(t,x) = \rho_0(x)$.

We can prove that, under the regularity and uniform ellipticity assumptions, the Fokker-Planck equation has a unique smooth solution. Furthermore, we can obtain pointwise bounds on the solution.

Theorem 4.2. *Assume that conditions* (4.6) *and* (4.7) *are satisfied, and assume that* $|\rho_0(x)| \leqslant ce^{\alpha\|x\|^2}$. *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*

$$|p|,\ |p_t|,\ \|\nabla p\|,\ \|D^2 p\| \leqslant Kt^{(-d+2)/2} \exp\left(-\frac{1}{2t}\delta\|x\|^2\right). \tag{4.8}$$

From estimates (4.8) it follows that all moments of a diffusion process whose diffusion matrix satisfies the uniform ellipticity assumption (4.6) exist. In particular, we can multiply the Fokker-Planck equation by monomials x^n, integrate over $\mathbb{R}^d$ and then integrate by parts. It also follows from the maximum principle for parabolic PDEs that the solution of the Fokker-Planck equation is nonnegative for all times, when the initial condition $\rho_0(x)$ is nonnegative. Since $\rho_0(x)$ is the probability density function of the random variable X_0 it is nonnegative and normalized, $|\rho_0|_{L^1(\mathbb{R}^d)} = 1$. The solution to the Fokker-Planck equation preserves this properties–as we would expect, since it is the transition probability density.

The Fokker-Planck equation can be written as the familiar continuity equation from continuum mechanics. We define the *probability flux (current)* to be the vector

$$\mathbf{J} := \mathbf{b}(\mathbf{x})\mathbf{p} - \frac{\mathbf{1}}{\mathbf{2}}\nabla \cdot \big(\mathbf{\Sigma}(\mathbf{x})\mathbf{p}\big). \tag{4.9}$$

The Fokker-Planck equation can be written in the form

$$\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J} = 0. \tag{4.10}$$

Integrating the Fokker-Planck equation over $\mathbb{R}^d$ and using the divergence theorem on the right hand side of the equation, together with (4.8), we obtain

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

Consequently:

$$\int_{\mathbb{R}^d} p(x,t)\,dx = \int_{\mathbb{R}^d} \rho_0(x)\,dx = 1. \tag{4.11}$$

Hence, the total probability is conserved, as expected.

The stationary Fokker-Planck equation, whose solutions give us the invariant distributions of the diffusion process X_t, can be written in the form

$$\nabla \cdot \mathbf{J}(p_s) = 0. \tag{4.12}$$

Consequently, the equilibrium probability flux is a divergence-free vector field.

Boundary conditions for the Fokker-Planck equation

There are many applications where it is necessary to study diffusion processes in bounded domains. In such cases we need to specify the behavior of the diffusion process on the boundary of the domain. Equivalently, we need to specify the behavior of the transition probability density on the boundary. Diffusion processes in bounded domains lead to initial boundary value problems for the corresponding Fokker-Planck equation.

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

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

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

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

These three different boundary behaviors for the random walk correspond to absorbing, reflecting or periodic boundary conditions.

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

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

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

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

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

iii. Consider the case where $\Omega = [0, L]^d$ and assume that the transition probability function is periodic in all directions with period L. We can then consider the Fokker-Planck equation in $[0, L]^d$ with periodic boundary conditions.

Using the terminology of the theory of partial differential equations, absorbing boundary conditions correspond to Dirichlet boundary conditions and reflecting boundary conditions correspond to Robin boundary conditions. We can, of course, consider mixed boundary conditions where part of the domain is absorbing and part of it is reflecting.

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

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

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

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

where the probability current is defined in (4.9). An example of mixed boundary conditions would be absorbing boundary conditions at the left end and reflecting boundary conditions at the right end:

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

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

4.2 Examples of Diffusion Processes and their Fokker-Planck Equation

There are several examples of diffusion processes in one dimension for which the corresponding Fokker-Planck equation can be solved analytically. We present some examples in this section. In the next section we study diffusion processes in one dimension using eigenfunction expansions.

Brownian motion

First we consider Brownian motion in $\mathbb{R}$. We set $b(x,t) \equiv 0$, $\Sigma(x,t) \equiv 2D > 0$. The Fokker-Planck equation for Brownian motion is the heat equation. We calculate the transition probability density for a Brownian particle that is at x_0 at time s. The Fokker-Planck equation for the transition probability density $p(x,t|x_0,s)$ is:

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}, \quad p(x,s|x_0,s) = \delta(x-x_0). \tag{4.13}$$

The solution to this equation is the Green's function (fundamental solution) of the heat equation:

$$p(x,t|y,s) = \frac{1}{\sqrt{4\pi D(t-s)}} \exp\left(-\frac{(x-x_0)^2}{4D(t-s)}\right). \tag{4.14}$$

Quite often we can obtain information on the properties of a diffusion process, for example, we can calculate moments, without having to solve the Fokker-Planck equation but by only using the structure of the equation. For example, using the Fokker-Planck equation (4.13) we can show that the mean squared displacement of Brownian motion grows linearly in time. Assume that the Brownian particle was at x_0 initially. We calculate, by performing two integrations by parts (which can be justified in view of (4.14))

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}W_t^2 &= \frac{d}{dt}\int_{\mathbb{R}} x^2 p(x,t|x_0,0)\,dx \\
&= D\int_{\mathbb{R}} x^2 \frac{\partial^2 p(x,t|x_0,0)}{\partial x^2}\,dx \\
&= D\int_{\mathbb{R}} p(x,t|x_0,0)\,dx = 2D,
\end{aligned}$$

From this calculation we conclude that the one dimensional Brownian motion W_t with diffusion coefficient D satisfies

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

Assume now that the initial condition W_0 of the Brownian particle is a random variable with distribution $\rho_0(x)$. To calculate the transition probability density of the Brownian particle we need to solve the Fokker-Planck equation with initial condition $\rho_0(x)$. In other words, we need to take the average of the probability density function

$$p(x,t|x_0) := p(x,t|x_0,0)$$

over all initial realizations of the Brownian particle. The solution of the Fokker-Planck equation with $p(x,0) = \rho_0(x)$ is

$$p(x,t) = \int p(x,t|x_0)\rho_0(x_0)\,dx_0. \tag{4.15}$$

Brownian motion with absorbing boundary conditions

We can also consider Brownian motion in a bounded domain, with either absorbing, reflecting or periodic boundary conditions. Consider the Fokker-Planck equation for Brownian motion with diffusion coefficient D (4.13) on $[0,1]$ with absorbing boundary conditions:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \quad p(0,t|x_0) = p(1,t|x_0) = 0, \quad p(x,0|x_0) = \delta(x - x_0), \tag{4.16}$$

In view of the Dirichlet boundary conditions we look for a solution to this equation in a sine Fourier series:

$$p(x,t) = \sum_{k=1}^{\infty} p_n(t)\sin(n\pi x). \tag{4.17}$$

With our choice (4.16) the boundary conditions are automatically satisfied. The Fourier coefficients of the initial condition $\delta(x - x_0)$ are

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

We substitute the expansion (4.17) into (4.16) and use the orthogonality properties of the Fourier basis to obtain the equations

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

The solution of this equation is

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

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

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

Notice that

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

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

Brownian Motion with Reflecting Boundary Condition

Consider now Brownian motion with diffusion coefficient D on the interval $[0,1]$ with reflecting boundary conditions. To calculate the transition probability density we solve the Fokker-Planck equation which in this case is the heat equation on $[0,1]$ with Neumann boundary conditions:

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}, \quad \partial_x p(0,t|x_0) = \partial_x p(1,t|x_0) = 0, \quad p(x,0) = \delta(x - x_0).$$

The boundary conditions are satisfied by functions of the form $cos(n\pi x)$. We look for a solution in the form of a cosine Fourier series

$$p(x,t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n(t)\cos(n\pi x).$$

From the initial conditions we obtain

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

We substitute the expansion into the PDE and use the orthonormality of the Fourier basis to obtain the equations for the Fourier coefficients:

$$\dot{a}_n = -n^2\pi^2 D a_n$$

from which we deduce that

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

Consequently

$$p(x, t|x_0) = 1 + 2\sum_{n=1}^{\infty} \cos(n\pi x_0)\cos(n\pi x)e^{-n^2\pi^2 Dt}.$$

Brownian motion on [0,1] with reflecting boundary conditions is an ergodic Markov process. To see this, let us consider the stationary Fokker-Planck equation

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

The unique normalized solution to this boundary value problem is $p_s(x) = 1$. Indeed, we multiply the equation by p_s, integrate by parts and use the boundary conditions to obtain

$$\int_0^1 \left|\frac{dp_s}{dx}\right|^2 dx = 0,$$

from which it follows that $p_s(x) = 1$. Alternatively, by taking the limit of $p(x, t|x_0)$ as $t \to \infty$ we obtain the invariant distribution:

$$\lim_{t\to\infty} p(x, t|x_0) = 1.$$

Now we can calculate the stationary autocorrelation function:

$$\begin{aligned}
\mathbb{E}(W_t W_0) &= \int_0^1\int_0^1 x x_0 p(x, t|x_0, 0)p_s(x_0)\, dx dx_0 \\
&= \int_0^1\int_0^1 x x_0 \left(1 + 2\sum_{n=1}^{\infty}\cos(n\pi x_0)\cos(n\pi x)e^{-n^2\pi^2 Dt}\right) dx dx_0 \\
&= \frac{1}{4} + \frac{8}{\pi^4}\sum_{n=0}^{+\infty}\frac{1}{(2n+1)^4}e^{-(2n+1)^2\pi^2 Dt}.
\end{aligned}$$

The Ornstein–Uhlenbeck Process

We set now $b(x, t) = -\alpha x$ with $\alpha > 0$ and $\Sigma(x, t) = 2D > 0$ for the drift and diffusion coefficients, respectively. The Fokker-Planck equation for the transition probability density $p(x, t|x_0)$ is

$$\frac{\partial p}{\partial t} = \alpha\frac{\partial(xp)}{\partial x} + D\frac{\partial^2 p}{\partial x^2}, \tag{4.18a}$$

$$p(x, 0|x_0) = \delta(x - x_0). \tag{4.18b}$$

This equation is posed on the real line, and the boundary conditions are that $p(x,t|x_0)$ decays sufficiently fast at infinity; see Definition 4.1. The corresponding stochastic differential equation is

$$dX_t = -\alpha X_t\,dt + \sqrt{2D}\,dW_t, \quad X_0 = x_0. \tag{4.19}$$

In addition to Brownian motion there is a linear force pulling the particle towards the origin. We know that Brownian motion is not a stationary process, since the variance grows linearly in time. By adding a linear damping term, it is reasonable to expect that the resulting process can become stationary. When (4.19) is used to model the velocity or position of a particle, the noisy term on the right hand side of the equation is related to thermal fluctuations. The diffusion coefficient D measures the strength of thermal fluctuations and it is associated with the temperature:

$$D = k_B T =: \beta^{-1}, \tag{4.20}$$

where T denotes the absolute temperature and k_B Boltzmann's constant. We will quite often use the notation β^{-1} and refer to β as the inverse temperature. We will also refer to α in (4.19) as the friction coefficient.

The solution of the stochastic differential equation (4.19) is given by (3.72). It follows then that

$$X_t \sim \mathcal{N}\left(x_0 e^{-\alpha t},\, \frac{D}{\alpha}(1-e^{-2\alpha t})\right). \tag{4.21}$$

See Equation (3.72) and Section 3.7. From this we can immediately obtain the transition probability density, the solution of the Fokker-Planck equation

$$p(x,t|x_0) = \sqrt{\frac{\alpha}{2\pi D(1-e^{-2\alpha t})}}\exp\left(-\frac{\alpha(x-x_0e^{-\alpha t})^2}{2D(1-e^{-2\alpha t})}\right). \tag{4.22}$$

We also studied the stationary Ornstein-Uhlenbeck process in Example (2.2) (for $\alpha = D = 1$) by using the fact that it can be defined as a time change of the Brownian motion. We can also derive it by solving the Fokker-Planck equation (4.18), by taking the Fourier transform of (4.18a), solving the resulting first order PDE using the method of characteristics and then taking the inverse Fourier transform. See Exercise 6 in Chapter 3.

In the limit as the friction coefficient α goes to 0, the transition probability (4.22) converges to the transition probability of Brownian motion. Furthermore, by taking the long time limit in (4.22) we obtain

$$\lim_{t\to+\infty} p(x,t|x_0) = \sqrt{\frac{\alpha}{2\pi D}}\exp\left(-\frac{\alpha x^2}{2D}\right),$$

irrespective of the initial position x_0. This is to be expected, since as we have already seen the Ornstein-Uhlenbeck process is an ergodic Markov process with a Gaussian invariant distribution

$$p_s(x) = \sqrt{\frac{\alpha}{2\pi D}}\exp\left(-\frac{\alpha x^2}{2D}\right). \tag{4.23}$$

Using now (4.22) and (4.23) we obtain the stationary joint probability density

$$\begin{aligned} p_2(x,t|x_0) &= p(x,t|x_0)p_s(x_0) \\ &= \frac{\alpha}{2\pi D\sqrt{1-e^{-2\alpha t}}}\exp\left(-\frac{\alpha(x^2+x_0^2-2xx_0e^{-\alpha t})}{2D(1-e^{-2\alpha t})}\right), \end{aligned}$$

or, starting at an arbitrary initial time s,

$$p_2(x,t|x_0,s) = \frac{\alpha}{2\pi D\sqrt{1-e^{-2\alpha|t-s|}}} \exp\left(-\frac{\alpha(x^2+x_0^2-2xx_0e^{-\alpha|t-s|})}{2D(1-e^{-2\alpha|t-s|})}\right). \tag{4.24}$$

Now we can calculate the stationary autocorrelation function of the Ornstein-Uhlenbeck process

$$\begin{aligned}\mathbb{E}(X_tX_s) &= \int\int xx_0p_2(x,t|x_0,s)\,dxdx_0 \\ &= \frac{D}{\alpha}e^{-\alpha|t-s|}.\end{aligned}$$

The derivation of this formula requires the calculation of Gaussian integrals, similar to the calculations presented in Section B.5. See Exercise 3 and Section 3.7.

Assume now that the initial conditions of the Ornstein-Uhlenbeck process X_t is a random variable distributed according to a distribution $\rho_0(x)$. As in the case of a Brownian particle, the probability density function is given by the convolution integral

$$p(x,t) = \int p(x,t|x_0)\rho_0(x_0)\,dx_0, \tag{4.25}$$

When X_0 is distributed according to the invariant distribution $p_s(x)$, given by (4.23), then the Ornstein-Uhlenbeck process becomes stationary. The solution to the Fokker-Planck equation is now $p_s(x)$ at all times and the joint probability density is given by (4.24).

Knowledge of the transition probability density enables us to calculate all moments of the Ornstein-Uhlenbeck process:

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

In fact, we can calculate the moments by using the Fokker-Planck equation, rather than the explicit formula for the transition probability density. We assume that all moments of the initial distribution exist. We start with $n = 0$. We integrate (4.18a) over $\mathbb{R}$ to obtain:

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

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

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

which simply means that

$$\int_{\mathbb{R}} p(x,t)\,dx = \int_{\mathbb{R}} \rho_0(x)\,dx = 1.$$

Let now $n = 1$. We multiply (4.18a) by x, integrate over $\mathbb{R}$ and perform an integration by parts to obtain:

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

Consequently:

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

Now we consider the case $n \geqslant 2$. We multiply (4.18a) by x^n and integrate by parts, once on the first term and twice on the second on the right hand side of the equation) to obtain:

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

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

$$M_n(t) = e^{-\alpha n t}M_n(0) + Dn(n-1)\int_0^t e^{-\alpha n(t-s)}M_{n-2}(s)\,ds. \tag{4.26}$$

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

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

As expected, all moments of the Ornstein-Uhlenbeck process converge to their stationary values:

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

In fact, it is clear from (4.26) that the moments conerge to their stationary values exponentially fast and that the convergence is faster the closer we start from the stationary distribution–see the formula for $M_2(t)$. Later in this chapter we will study the problem of convergence to equilibrium for more general classes of diffusion processes; see Section 4.6

Geometric Brownian Motion

We set $b(x) = \mu x, \quad \Sigma(x) = \frac{1}{2}\sigma^2 x^2$. This is the geometric Brownian motion that we encountered in Chapters 1 and 3. This diffusion process appears in mathematical finance and in population dynamics. The generator of this process is

$$\mathcal{L} = \mu x\frac{\partial}{\partial x} + \frac{\sigma x^2}{2}\frac{\partial^2}{\partial x^2}. \tag{4.27}$$

Notice that this operator is not uniformly elliptic, since the diffusion coefficient vanishes at $x = 0$. The Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu x) + \frac{\partial^2}{\partial x^2}\left(\frac{\sigma^2 x^2}{2}p\right). \tag{4.28a}$$

$$p(x,0|x_0) = \delta(x - x_0). \tag{4.28b}$$

Since the diffusion coefficient is not uniformly elliptic, it is not covered by Theorem 4.2. The corresponding stochastic differential equation is given by Equation (3.74). As for the Ornstein-Uhlenbeck process, we can use the Fokker-Planck equation in order to obtain equations for the moments of geometric Brownian motion:

$$\frac{d}{dt}M_1 = \mu n M_n, \quad \frac{d}{dt}M_n = \Big(\mu n + \frac{\sigma^2}{2}n(n-1)\Big)M_n, \quad n \geqslant 2.$$

We can solve these equations to obtain

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

and

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

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

$$M_n(t) = e^{(2\mu+\sigma^2)t}M_2(0), \tag{4.29}$$

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

4.3 Diffusion Processes in One Dimension

In this section we study the Fokker-Planck equation for diffusion processes in one dimension in a bounded interval and with reflecting boundary conditions. Let X_t denote a diffusion process in the interval $[\ell,\, r]$ with drift and diffusion coefficients $b(x)$ and $\sigma(x)$, respectively. We will assume that $\sigma(x)$ is positive in $[\ell,\, r]$. The transition probability density $p(x,t|x_0)$ is the solution of the following initial boundary value problem for the Fokker-Planck equation:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left(b(x)p - \frac{1}{2}\frac{\partial}{\partial x}(\sigma(x)p)\right) =: -\frac{\partial J}{\partial x}, \quad x \in (\ell,\, r), \tag{4.30a}$$

$$p(x,0|x_0) = \delta(x-x_0), \tag{4.30b}$$

$$J(\ell,t) = J(r,t) = 0. \tag{4.30c}$$

The reflecting boundary conditions and the assumption on the positivity of the diffusion coefficient ensure that X_t is ergodic. We can calculate the unique invariant probability distribution. The stationary Fokker-Planck equation reads

$$\frac{dJ_s}{dx} = 0, \quad x \in (\ell,\, r), \tag{4.31a}$$

$$J_s(\ell) = J_s(r) = 0, \tag{4.31b}$$

where

$$J_s(x) := J(p_s(x)) = b(x)p_s(x) - \frac{1}{2}\sigma(x)\frac{dp_s}{dx}(x) \tag{4.32}$$

denotes the stationary probability flux, $p_s(x)$ being the stationary probability distribution. We use the reflecting boundary conditions (4.31) to write the stationary Fokker-Planck equation in the form

$$J_s(x) = 0, \quad x \in (\ell,\, r). \tag{4.33}$$

Thus, the stationary probability flux vanishes. This is called the *detailed balance* condition and it will be discussed in Section 4.6. Equation (4.33) gives

$$b(x)p_s(x) - \frac{1}{2}\frac{d}{dx}(\sigma(x)p_s(x)) = 0, \quad x \in (\ell,\, r), \tag{4.34}$$

together with the normalization condition

$$\int_\ell^r p_s(x)\, dx = 1.$$

The detailed balance condition (4.33) results in the stationary Fokker-Planck equation becoming a first order differential equation. The solution to (4.34) can be obtained up to one constant, which is determined from the normalization condition. The solution is

$$p_s(x) = \frac{1}{Z}\frac{1}{\sigma(x)}\exp\left(2\int_\ell^x \frac{b(y)}{\sigma(y)}\, dy\right), \quad Z = \int_\ell^r \left(\frac{1}{\sigma(x)}\exp\left(2\int_\ell^x \frac{b(y)}{\sigma(y)}\, dy\right)\right) dx. \tag{4.35}$$

Now we solve the time dependent Fokker-Planck equation (4.30). We first transform the Fokker-Planck (forward Kolmogorov) equation to the backward Kolmogorov equation, since the boundary conditions for the generator $\mathcal{L}$ of the diffusion process X_t are simpler than those for the Fokker-Planck operator $\mathcal{L}^*$. Let $p \in D(\mathcal{L}^*) := \{p \in \mathcal{C}^2(\ell, r)\,;\, J(p(\ell)) = J(p(r)) = 0\}$, the domain of definition of the Fokker-Planck operator with reflecting boundary conditions. We write $p(x) = f(x)p_s(x)$ and use the stationary Fokker-Planck equation and the detailed balance condition (4.31b) to calculate

$$\begin{aligned} \mathcal{L}^* p &= \frac{d}{dx}\left(-b(x)f(x)p_s(x) + \frac{1}{2}\frac{d}{dx}(\sigma(x)f(x)p_s(x))\right) \\ &= p_s \mathcal{L} f. \end{aligned} \tag{4.36}$$

Furthermore,

$$J(p) = J(fp_s) = -\frac{1}{2}\sigma(x)p_s(x)\frac{df}{dx}(x).$$

In particular, in view of the reflecting boundary conditions and the fact that both the diffusion coefficient and the invariant distribution are positive,

$$\frac{df}{dx}(\ell) = \frac{df}{dx}(r) = 0. \tag{4.37}$$

Consequently, the generator $\mathcal{L}$ of the diffusion process X_t is equipped with Neumann boundary conditions, $D(\mathcal{L}) = (f \in C^2(\ell, r), f'(\ell) = f'(r) = 0)$.

Setting now $p(x, t|x_0) = f(x, t|x_0)p_s(x)$, we obtain the following initial boundary value problem

$$\frac{\partial f}{\partial t} = b(x)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma(x)\frac{\partial^2 f}{\partial x^2} =: \mathcal{L}f, \quad x \in (\ell,\, r), \tag{4.38a}$$

$$f(x, 0|x_0) = p_s^{-1}(x)\delta(x - x_0), \tag{4.38b}$$

$$f'(\ell, t|x_0) = f'(r, t|x_0) = 0. \tag{4.38c}$$

We solve this equation using separation of variables (we suppress the dependence on the initial condition x_0),

$$f(x, t) = \psi(x)c(t).$$

Substituting this into (4.38a) we obtain

$$\frac{\dot{c}}{c} = \frac{\mathcal{L}\psi}{\psi} = -\lambda,$$

where λ is a constant and

$$c(t) = c(0)e^{-\lambda t}, \quad -\mathcal{L}\psi = \psi.$$

Using the superposition principle, we deduce that the solution to the backward Kolmogorov equation (4.38) is

$$f(x,t) = \sum_{n=0}^{+\infty} c_n \psi_n(x) e^{-\lambda_n t}, \tag{4.39}$$

where $\{\lambda_n,\, \psi_n\}_{n=0}^{+\infty}$ are the eigenvalues and eigenfunctions of the generator of X_t equipped with Neumann boundary conditions:

$$-\mathcal{L}\psi_n = \lambda_n \psi_n, \quad \psi_n'(\ell) = \psi_n'(r) = 0. \tag{4.40}$$

The generator $\mathcal{L}$ (with Neumann boundary conditions) is a selfadjoint operator in the space $L^2((\ell, r); p_s(x))$, the space of square integrable functions in the interval $(\ell,\, r)$, weighted by the stationary distribution of the process. This is a Hilbert space with inner product

$$\langle f, h\rangle = \int_\ell^r f(x)h(x)p_s(x)\,dx$$

and corresponding norm $\|f\| = \sqrt{\langle f, f\rangle}$. The selfadjointness of $\mathcal{L}$ follows from (4.36):[3]

$$\int_\ell^r \mathcal{L}f h p_s\,dx = \int_\ell^r f\mathcal{L}^*(hp_s)\,dx = \int_\ell^r f\mathcal{L}h p_s\,dx$$

for all $f,\, h \in D(\mathcal{L})$. Furthermore, $-\mathcal{L}$ is a positive operator: performing an integration by parts and using the stationary Fokker-Planck equation we obtain

$$\int_\ell^r (-\mathcal{L}f) f p_s\,dx = \frac{1}{2}\int_\ell^r |f'|^2 \sigma p_s\,dx.$$

The generator $\mathcal{L}$ has discrete spectrum in the Hilbert space $L^2((\ell, r); p_s(x))$. In addition, the eigenvalues of $-\mathcal{L}$ are real, nonnegative, with $\lambda_0 = 0$ corresponding to the invariant distribution, and can be ordered, $0 = \lambda_0 < \lambda_1 < \lambda_2 < \ldots$. The eigenfunctions of $-\mathcal{L}$ form an orthonormal basis on $L^2((\ell, r); p_s(x))$: a function $f \in L^2((\ell, r); p_s(x))$ can be expanded in a generalized Fourier series $f = \sum f_n \psi_n$ with $f_n = \langle f, \psi_n\rangle$.

The solution to (4.38) is given by (4.39)

$$f(x, t|x_0) = \sum_{n=0}^{+\infty} c_n e^{-\lambda_n t}\psi_n(x),$$

where the constants $\{c_n\}_{n=0}^{+\infty}$ are determined from the initial conditions:

$$\begin{aligned} c_n &= \int_\ell^r f(x,0|x_0)\psi_n(x)p_s(x)\,dx = \int_\ell^r \delta(x-x_0)\psi_n(x)\,dx \\ &= \psi_n(x_0). \end{aligned}$$

[3] In fact, we only prove that $\mathcal{L}$ is symmetric. An additional argument is needed in order to prove that it is self-adjoint. See the comments in Section 4.10.

Putting everything together we obtain a solution to the time dependent Fokker-Planck equation (4.30):

$$p(x,t|x_0) = p_s(x)\sum_{n=0}^{+\infty} e^{-\lambda_n t}\psi_n(x)\psi_m(x_0). \tag{4.41}$$

The main challenge in this approach to solving the Fokker-Planck equation is the calculation of the eigenvalues and eigenfunctions of the generator of X_t in $L^2((\ell,r);p_s(x))$. This can be done either analytically or, in most cases, numerically.

If the initial condition of the diffusion process X_0 is distributed according to a probability distribution $\rho_0(x)$, then the solution of the stationary Fokker-Planck equation is

$$p(x,t) = p_s(x)\sum_{n=0}^{+\infty} c_n e^{-\lambda_n t}\psi_n(x), \quad c_n = \int_\ell^r \psi_n(x)\rho_0(x)\,dx. \tag{4.42}$$

Notice that from the above formula and the fact that all eigenvalues apart from the first are positive we conclude that X_t, starting from an arbitrary initial distribution, converges to its invariant distribution exponentially fast in $L^2((\ell,r);p_s(x))$. We will consider multidimensional diffusion processes for which a similar result can be obtained later in this chapter.

4.4 The Ornstein-Uhlenbeck Process and Hermite Polynomials

The Ornstein-Uhlenbeck process that we already encountered in Section 4.2 is one of the few stochastic processes for which we can calculate explicitly the solution of the corresponding stochastic differential equation, the solution of the Fokker-Planck equation as well as the eigenvalues and eigenfunctions of the generator of the process. In this section we show that the eigenfunctions of the Ornstein-Uhlenbeck process are the Hermite polynomials and study various properties of the generator of the Ornstein-Uhlenbeck process. We will see that it has many of the properties of the generator of a diffusion process in one dimension with reflective boundary conditions that we studied in the previous section. In the next section we will show that many of the properties of the Ornstein-Uhlenbeck process (ergodicity, selfadjointness of the generator, exponentially fast convergence to equilibrium, real, discrete spectrum) are shared by a large class of diffusion processes, namely reversible diffusions.

We consider a diffusion process in $\mathbb{R}^d$ with drift $\mathbf{b}(\mathbf{x}) = -\alpha\mathbf{x}$, $\alpha > 0$ and $\mathbf{\Sigma}(x) = \beta^{-1}I$, where I denotes the $d \times d$ identity matrix. The generator of the d-dimensional Ornstein-Uhlenbeck process is

$$\mathcal{L} = -\alpha p\cdot\nabla_p + \beta^{-1}\Delta_p, \tag{4.43}$$

where, as explained in Section 4.4, β denotes the inverse temperature and α denotes the friction coefficient.

We have already seen that the Ornstein-Uhlenbeck process is an ergodic Markov process whose unique invariant density is the Gaussian

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

We can perform the same transformation as in the previous section: we have that

$$\mathcal{L}^*(h\rho_\beta(p)) = \rho_\beta(p)\mathcal{L}h. \tag{4.44}$$

The initial value problem for the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \quad p(x,0) = p_0(x)$$

becomes

$$\frac{\partial h}{\partial t} = \mathcal{L} h, \quad h(x,0) = \rho_\beta^{-1}(x) p_0(x).$$

Therefore, in order to study the Fokker-Planck equation for the Ornstein-Uhlenbeck process it is sufficient to study the properties of the generator $\mathcal{L}$. As in the previous section, the natural function space for studying the generator of the Ornstein-Uhlenbeck process is the L^2-space weighted by the invariant measure of the process. This is a (separable) Hilbert space with norm

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

and corresponding inner product

$$\langle f, h \rangle_\rho = \int_{\mathbb{R}^d} f h \rho_\beta \, dp.$$

We can also define weighted L^2-spaces involving derivatives, i.e. weighted Sobolev spaces. See Exercise 6.

The generator of the Ornstein-Uhlenbeck process becomes a selfadjoint operator in this space. In fact, $\mathcal{L}$ defined in (4.43) has many nice properties that are summarized in the following proposition.

Proposition 4.3. *The operator $\mathcal{L}$ has the following properties:*

i. For every f, $h \in C^2(\mathbb{R}^d) \cap L^2(\rho_\beta)$,

$$\langle -\mathcal{L} f, h \rangle_\rho = -\beta^{-1} \int_{\mathbb{R}^d} \nabla f \cdot \nabla h \rho_\beta \, dp. \tag{4.45}$$

ii. $\mathcal{L}$ is a nonpositive operator on $L^2(\rho_\beta)$.

iii. The null space of $\mathcal{L}$ consists of constants.

Proof.

i. Equation (4.45) follows from an integration by parts:

$$\begin{aligned} \langle \mathcal{L} f, h \rangle_\rho &= \int -p \cdot \nabla f h \rho_\beta \, dp + \beta^{-1} \int \Delta f h \rho_\beta \, dp \\ &= \int -p \cdot \nabla f h \rho_\beta \, dp - \beta^{-1} \int \nabla f \cdot \nabla h \rho_\beta \, dp + \int -p \cdot \nabla f h \rho_\beta \, dp \\ &= -\beta^{-1} \langle \nabla f, \nabla h \rangle_\rho. \end{aligned}$$

ii. Non-positivity of $\mathcal{L}$ follows from (4.45) upon setting $h = f$:

$$\langle \mathcal{L} f, f \rangle_\rho = -\beta^{-1} \|\nabla f\|_\rho^2 \leqslant 0. \tag{4.46}$$

iii. Let $f \in \mathcal{N}(\mathcal{L})$ and use (4.46) to deduce that

$$\int_{\mathbb{R}^d} |\nabla f|^2 \rho_\beta \, dx = 0,$$

from which we deduce that $f \equiv$ const. $\square$

The generator of the Ornstein-Uhlenbeck process has a spectral gap: For every $f \in C^2(\mathbb{R}^d) \cap L^2(\rho_\beta)$ we have

$$\langle -\mathcal{L}f, f\rangle_\rho \geqslant \text{Var}(f), \tag{4.47}$$

where $\text{Var}(f) = \int_{\mathbb{R}^d} f^2\rho_\beta - \big(\int_{\mathbb{R}^d} f\rho_\beta\big)^2$. This statement is equivalent to the statement that the Gaussian measure $\rho_\beta(x)\,dx$ satisfies *Poincarè's inequality* :

$$\int_{\mathbb{R}^d} f^2\rho_\beta\,dp \leqslant \beta^{-1}\int_{\mathbb{R}^d} |\nabla f|^2\rho_\beta\,dp \tag{4.48}$$

for all smooth functions with $\int f\rho_\beta = 0$. Poincaré's inequality for Gaussian measures can be proved using the fact that (tensor products of) Hermite polynomials form an orthonormal basis in $L^2(\mathbb{R}^d;\rho_\beta)$. We can also use the fact that the generator of the Ornstein-Uhlenbeck process is unitarily equivalent to the Schrödinger operator for the quantum harmonic oscillator, whose eigenfunctions are the Hermite functions:[4] consider the generator $\mathcal{L}$ in one dimension and set, for simplicity $\beta = 1$, $\alpha = 2$. Then

$$\rho_\beta^{1/2}\big(-\mathcal{L}(h\rho_\beta^{-1/2})\big) = -\frac{d^2h}{dx^2} + x^2h - h := \mathcal{H}h. \tag{4.49}$$

Poincaré's inequality (4.48) follows from the fact that the operator $\widehat{\mathcal{H}} = -\frac{d^2}{dx^2} + x^2$ has a spectral gap,

$$\int_{\mathbb{R}} \widehat{\mathcal{H}}hh\,dx \geqslant \int_{\mathbb{R}} |h|^2\,dx, \tag{4.50}$$

which, in turn, follows from the estimate

$$\|h\|_{L^2}^2 \leqslant 2\|xh\|_{L^2}\left\|\frac{dh}{dx}\right\|_{L^2}, \tag{4.51}$$

for all smooth functions h. From (4.49) it follows that (4.48) is equivalent to

$$\int_{\mathbb{R}} \mathcal{H}hh\,dx \geqslant \int_{\mathbb{R}} h^2\,dx, \quad \int_{\mathbb{R}} he^{-x^2/2}\,dx = 0.$$

We can check that $\sqrt{\rho} = e^{-x^2/2}$ is the first eigenfunction of $\mathcal{H}$, corresponding to the zero eigenvalue, $\lambda_0 = 0$. The centering condition for f is equivalent to the condition that h is orthogonal to the ground state (i.e. the first eigenfunction) of $\mathcal{H}$.

Since now $\widehat{\mathcal{H}} = -\frac{d^2}{dx^2} + x^2$ is a selfadjoint operator in $L^2(\mathbb{R})$ that satisfies the spectral gap estimate (4.50), it has discrete spectrum and its eigenfunctions form an orthonormal basis in $L^2(\mathbb{R})$.[5] Furthermore its eigenvalues are positive (from (4.50) it follows that it is a positive operator) and we can check that its first nonzero eigenvalue is $\lambda_1 = 2$. Let $\{\lambda_n, \phi_n\}$ denote the eigenvalues and eigenfunctions of $\mathcal{H}$ and let h be a smooth L^2 function that is orthogonal to the ground state. We have

$$\int_{\mathbb{R}} \mathcal{H}hh\,dx = \sum_{n=1}^{\infty}\lambda_n h_n^2 \geqslant 2\sum_{n=1}^{\infty} h_n^2, \tag{4.52}$$

[4]The transformation of the generator of a diffusion process to a Schrödinger operator is discussed in detail in Section 4.9.

[5]These eigenfunctions are the Hermite functions.

from which (4.47) or, equivalently, (4.48) follows. From Proposition 4.3 and the spectral gap estimate (4.47) it follows that the generator of the Ornstein-Uhlenbeck process is a selfadjoint operator in $L^2(\rho_\beta)$ with discrete spectrum, nonnegative eigenvalues and its eigenfunctions form an orthonormal basis in $L^2(\rho_\beta)$.

The connection between the generator of the Ornstein-Uhlenbeck process and the Schrödinger operator for the quantum harmonic oscillator can be used in order to calculate the eigenvalues and eigenfunctions of $\mathcal{L}$. We present the results in one dimension. The multidimensional problem can be treated similarly by taking tensor products of the eigenfunctions of the one dimensional problem.

Theorem 4.4. *Consider the eigenvalue problem for the the generator of the one dimensional Ornstein-Uhlenbeck process*

$$\mathcal{L} = -\alpha p \frac{d}{dp} + \beta^{-1} \frac{d^2}{dp^2}, \tag{4.53}$$

in the space $L^2(\rho_\beta)$:

$$-\mathcal{L} f_n = \lambda_n f_n. \tag{4.54}$$

Then the eigenvalues of $\mathcal{L}$ are the nonnegative integers multiplied by the friction coefficient:

$$\lambda_n = \alpha\, n, \quad n = 0,\, 1,\, 2,\, \ldots. \tag{4.55}$$

The corresponding eigenfunctions are the normalized Hermite polynomials :

$$f_n(p) = \frac{1}{\sqrt{n!}} H_n\left(\sqrt{\alpha\beta}\, p\right), \tag{4.56}$$

where

$$H_n(p) = (-1)^n e^{\frac{p^2}{2}} \frac{d^n}{dp^n}\left(e^{-\frac{p^2}{2}}\right). \tag{4.57}$$

Notice that the eigenvalues of $\mathcal{L}$ are independent of the strength of the noise β^{-1}.[6] This is a general property of the spectrum of the generator of linear stochastic differential equations. See Sections 3.7 and **??**.

From (4.57) we can see that H_n is a polynomial of degree n. Furthermore, only odd (even) powers appear in $H_n(p)$ when n is odd (even). In addition, the coefficient multiplying p^n in $H_n(p)$ is always 1. The orthonormality of the modified Hermite polynomials $f_n(p)$ defined in (4.56) implies that

$$\int_{\mathbb{R}} f_n(p) f_m(p) \rho_\beta(p)\, dp = \delta_{nm}.$$

The first few Hermite polynomials and the corresponding rescaled/normalized eigenfunctions of the gener-

[6] Of course, the function space $L^2(\rho_\beta)$ in which we study the eigenvalue problem for $\mathcal{L}$ does depend on β through ρ_β.

ator of the Ornstein-Uhlenbeck process are:

$$\begin{aligned}
H_0(p) &= 1, & f_0(p) &= 1,\\
H_1(p) &= p, & f_1(p) &= \sqrt{\beta}p,\\
H_2(p) &= p^2 - 1, & f_2(p) &= \frac{\alpha\beta}{\sqrt{2}}p^2 - \frac{1}{\sqrt{2}},\\
H_3(p) &= p^3 - 3p, & f_3(p) &= \frac{\alpha\beta^{3/2}}{\sqrt{6}}p^3 - \frac{3\sqrt{\alpha\beta}}{\sqrt{6}}p\\
H_4(p) &= p^4 - 3p^2 + 3, & f_4(p) &= \frac{1}{\sqrt{24}}\left((\alpha\beta)^2p^4 - 3\alpha\beta p^2 + 3\right)\\
H_5(p) &= p^5 - 10p^3 + 15p, & f_5(p) &= \frac{1}{\sqrt{120}}\left((\alpha\beta)^{5/2}p^5 - 10(\alpha\beta)^{3/2}p^3 + 15(\alpha\beta)^{1/2}p\right).
\end{aligned}$$

Proof of Theorem 4.4. We already know that $\mathcal{L}$ has discrete nonnegative spectrum and that its eigenfunctions span $L^2(\rho_\beta)$. We can calculate the eigenvalues and eigenfunctions by introducing appropriate creation and annihilation operators. We define the annihilation operator

$$a^- = \frac{1}{\sqrt{\beta}}\frac{d}{dp} \tag{4.58}$$

and the creation operator

$$a^+ = \sqrt{\beta}\alpha p - \frac{1}{\sqrt{\beta}}\frac{d}{dp}. \tag{4.59}$$

These two operators are $L^2(\rho_\beta)$-adjoint:

$$\langle a^- f, h\rangle_\rho = \langle f, a^+ h\rangle_\rho,$$

for all C^1 functions $f,\ h$ in $L^2(\rho_\beta)$. Using these operators we can write the generator $\mathcal{L}$ in the form

$$\mathcal{L} = -a^+a^-.$$

The eigenvalue problem (4.54) becomes

$$a^+a^-f_n = \lambda_n f_n.$$

Furthermore, we easily check that a^+ and a^- satisfy the following commutation relation

$$[a^+, a^-] = a^+a^- - a^-a^+ = -\alpha.$$

Now we calculate:

$$\begin{aligned}
[\mathcal{L}, a^+] &= \mathcal{L}^+a + -a^+\mathcal{L} = -a^+a^-a^+ + a^+a^+a^-\\
&= -a^+a^-a^+ + a^+(a^-a^+ - \alpha)\\
&= -\alpha.
\end{aligned}$$

We proceed by induction to show that

$$[\mathcal{L}, (a^+)^n] = -\alpha n (a^+)^n. \tag{4.60}$$

Define now

$$\phi_n = (a^+)^n \mathbf{1}, \tag{4.61}$$

where $\phi_0 = \mathbf{1}$ is the "ground state" corresponding to the eigenvalue $\lambda_0 = 0$, $\mathcal{L}_0 \phi_0 = 0$. We use (4.60) and the fact that $\mathcal{L}\mathbf{1} = 0$ to check that ϕ_n is the n-th unnormalized eigenfunction of the generator $\mathcal{L}$:

$$\begin{aligned} -\mathcal{L}\phi_n &= -\mathcal{L}(a^+)^n \mathbf{1} = -(a^+)^n \mathcal{L}\mathbf{1} - [\mathcal{L}, (a^+)^n]\mathbf{1} \\ &= \alpha n (a^+)^n \mathbf{1} = \alpha n \phi_n. \end{aligned}$$

We can check by induction that the eigenfunctions defined in (4.61) are the unormalized Hermite polynomials. We present the calculation of the first few eigenfunctions:

$$\phi_0 = \mathbf{1}, \quad \phi_1 = a^+ \phi_0 = \sqrt{\beta}\alpha p, \quad \phi_2 = a^+ \phi_1 = \beta \alpha^2 p^2 - \alpha.$$

Since a^+ and a^- are $L^2(\rho_\beta)$-adjoint, we have that

$$\langle \phi_n, \phi_m \rangle_\rho = 0, \quad n \neq m.$$

Upon normalizing the $\{\phi_n\}_{n=0}^{+\infty}$ we obtain (4.56). The normalization constant

$$\|\phi_n\|_\rho = \sqrt{\langle (a^+)^n \mathbf{1}, (a^+)^n \mathbf{1} \rangle_\rho}$$

can be calculated by induction. □From the eigenfunctions and eigenvalues of $\mathcal{L}$ and using the transformation (4.44) we conclude that the Fokker-Planck operator of the Ornstein-Uhlenbeck process has the same eigenvalues as the generator $\mathcal{L}$ and the eigenfunctions are obtained by multiplying those of the generator by the invariant distribution:

$$-\mathcal{L}^*(\rho_\beta f_n) = \alpha n \rho_\beta f_n, \quad n = 0, 1, \ldots \tag{4.62}$$

Using the eigenvalues and eigenfunction of the Fokker-Planck operator (or, equivalently, of the generator) we can solve the time dependent problem and obtain a formula for the probability density function (compare with (4.41)):

$$\rho(p,t) = \rho_\beta(p) \sum_{n=0}^{+\infty} c_n e^{-\lambda_n t} f_n(p), \quad c_n = \int_{\mathbb{R}} f_n(p) \rho_0(p)\, dx, \tag{4.63}$$

where $\{f_n\}_{n=0}^{+\infty}$ denote the eigenvalues of the generator. From this formula we deduce that the when starting from an arbitrary initial distribution $\rho_0(p) \in L^2(\mathbb{R}; \rho_\beta^{-1})$ the law of the process converges exponentially fast to the invariant distribution.

4.5 The Smoluchowski Equation

The Ornstein-Uhlenbeck process is an example of an ordinary differential equation with a gradient structure that is perturbed by noise: letting $V(x) = \frac{1}{2}\alpha|x|^2$ we can write the SDE for the Ornstein-Uhlenbeck process in the form

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t.$$

The generator can be written as:

$$\mathcal{L} = -\nabla V(x)\cdot\nabla + \beta^{-1}\Delta. \tag{4.64}$$

The Gaussian invariant distribution of the Ornstein-Uhlenbeck process can be written in the form

$$\rho_\beta(x) = \frac{1}{Z}e^{-\beta V(x)}, \quad Z = \int_{\mathbb{R}^d} e^{-\beta V(x)}\,dx.$$

In the previous section we were able to obtain detailed information on the spectrum of the generator of the Ornstein-Uhlenbeck process, which in turn enabled us to solve the time-dependent Fokker-Planck equation and to obtain (4.63) from which exponentially fast convergence to equilibrium follows. We can obtain similar results, using the same approach for more general classes of diffusion processes whose generator is of the form (4.64), for a quite general class of scalar functions $V(x)$. We will refer to $V(x)$ as the potential.

In this section we consider stochastic differential equations of the form

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t, \quad X_0 = x, \tag{4.65}$$

for more general potentials $V(x)$, not necessarily quadratic. The generator of the diffusion process X_t is

$$\mathcal{L} = -\nabla V(x)\cdot\nabla + \beta^{-1}\Delta, \tag{4.66}$$

Assume that the initial condition for X_t is a random variable with probability density function $\rho_0(x)$. The probability density function of X_t, $\rho(x,t)$ is the solution of the initial value problem for the corresponding Fokker-Planck equation:

$$\frac{\partial\rho}{\partial t} = \nabla\cdot(\nabla V\rho) + \beta^{-1}\Delta\rho, \tag{4.67a}$$

$$\rho(x,0) = \rho_0(x). \tag{4.67b}$$

The Fokker-Planck equation (4.67a) is often called the *Smoluchowski equation*. In the sequel we will refer to this equation as either the Smoluchowski or the Fokker-Planck equation.

It is not possible to calculate the time dependent solution of the Smoluchowski equation for arbitrary potentials. We can, however, always calculate the stationary solution, if it exists.

Definition 4.5. *A potential V will be called* confining *if $\lim_{|x|\to+\infty} V(x) = +\infty$ and*

$$e^{-\beta V(x)} \in L^1(\mathbb{R}^d). \tag{4.68}$$

for all $\beta \in \mathbb{R}^+$.

In other words, a potential $V(\cdot)$ is confining if it grows sufficiently fast at infinity so that (4.68) holds. The simplest example of a confining potential is the quadratic potential. A particle moving in a confining potential according to the dynamics (4.65) cannot escape to infinity, it is confined to move in a bounded area in $\mathbb{R}^d$. It is reasonable, then, to expect that the dynamics (4.65) in a confining potential has nice ergodic properties.

Proposition 4.6. *Let $V(x)$ be a smooth confining potential. Then the Markov process with generator* (4.66) *is ergodic. The unique invariant distribution is the Gibbs distribution*

$$\rho_\beta(x) = \frac{1}{Z} e^{-\beta V(x)} \tag{4.69}$$

where the normalization factor Z is the partition function

$$Z = \int_{\mathbb{R}^d} e^{-\beta V(x)}\, dx. \tag{4.70}$$

The fact that the Gibbs distribution is an invariant distribution follows by direct substitution. In fact, the stationary probability flux vanishes (compare with (4.12)):

$$J(\rho_\beta) = -\beta^{-1}\nabla\rho_\beta - \nabla V \rho_\beta = 0.$$

Uniqueness follows from the fact that the Fokker-Planck operator has a spectral gap; see the discussion later in the section.

Just as with the one dimensional diffusion processes with reflecting boundary conditions and the Ornstein-Uhlenbeck process, we can obtain the solution of the Smoluchowski equation (4.67) in terms of the solution of the backward Kolmogorov equation by a simple transformation. We define $h(x,t)$ through

$$\rho(x,t) = h(x,t)\rho_\beta(x).$$

Then we can check that the function h satisfies the backward Kolmogorov equation:

$$\frac{\partial h}{\partial t} = -\nabla V \cdot \nabla h + \beta^{-1}\Delta h, \quad h(x,0) = \rho_0(x)\rho_\beta^{-1}(x). \tag{4.71}$$

To derive (4.71), we calculate the gradient and Laplacian of the solution to the Fokker-Planck equation:

$$\nabla p = \rho\nabla h - \rho h\beta\nabla V \quad \text{and} \quad \Delta p = \rho\Delta h - 2\rho\beta\nabla V\cdot\nabla h + h\beta\Delta V\rho + h|\nabla V|^2\beta^2\rho.$$

We substitute these formulas into the Fokker-Planck equation to obtain (4.71).

Consequently, in order to study properties of solutions to the Fokker-Planck equation it is sufficient to study the backward equation (4.71). The generator $\mathcal{L}$ is self-adjoint in the right function space, which is the space of square integrable functions, weighted by the invariant density of the process X_t:

$$L^2(\rho_\beta) := \left\{ f | \int_{\mathbb{R}^d} |f|^2 \rho_\beta\, dx < \infty \right\}, \tag{4.72}$$

where ρ_β denotes the Gibbs distribution. This is a Hilbert space with inner product

$$\langle f, h\rangle_\rho := \int_{\mathbb{R}^d} fh\rho_\beta\, dx \tag{4.73}$$

and corresponding norm $\|f\|_\rho = \sqrt{\langle f, f\rangle_\rho}$.

The generator of the Smoluchowski dynamics (4.65) has the same properties as those of the generator of the Ornstein-Uhlenbeck process.

Proposition 4.7. *Assume that $V(x)$ is a smooth potential and assume that condition* (4.68) *holds. Then the operator*

$$\mathcal{L} = -\nabla V(x)\cdot\nabla + \beta^{-1}\Delta$$

is self-adjoint in $\mathcal{H}$. Furthermore, it is nonpositive and its kernel consists of constants.

Proof. Let $f, \in C_0^2(\mathbb{R}^d)\cap L^2(\rho_\beta)$. We calculate

$$\begin{aligned}\langle \mathcal{L}f, h\rangle_\rho &= \int_{\mathbb{R}^d}(-\nabla V\cdot\nabla + \beta^{-1}\Delta)fh\rho_\beta\,dx \\ &= \int_{\mathbb{R}^d}(\nabla V\cdot\nabla f)h\rho_\beta\,dx - \beta^{-1}\int_{\mathbb{R}^d}\nabla f\nabla h\rho_\beta\,dx - \beta^{-1}\int_{\mathbb{R}^d}\nabla f h\nabla\rho_\beta\,dx \\ &= -\beta^{-1}\int_{\mathbb{R}^d}\nabla f\cdot\nabla h\rho_\beta\,dx, \end{aligned} \tag{4.74}$$

from which selfadjointness follows.[7]

If we now set $f = h$ in the above equation we get

$$\langle \mathcal{L}f, f\rangle_\rho = -\beta^{-1}\|\nabla f\|^2_{\rho_\beta},$$

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

$$\int |\nabla f|^2\rho_\beta\,dx = 0,$$

from which we deduce that f is constant. □

The expression

$$D_{\mathcal{L}}(f) := \langle -\mathcal{L}f, f\rangle_\rho \tag{4.75}$$

is called the *Dirichlet form* of the generator $\mathcal{L}$. In the case of a gradient flow, it takes the form

$$D_{\mathcal{L}}(f) = \beta^{-1}\int_{\mathbb{R}^d}|\nabla f|^2\rho_\beta(x)\,dx. \tag{4.76}$$

Several properties of the diffusion process X_t can be studied by looking at the corresponding Dirichlet form.

Using now Proposition 4.7 we can study the problem of convergence to equilibrium for X_t. In particular, we can show that the solution of the Fokker-Planck equation (4.67) for an arbitrary initial distribution $\rho_0(x)$ converges to the Gibbs distribution exponentially fast. To prove this we need a functional inequality that is a property of the potential V. In particular, we need to use the fact that, under appropriate assumptions on V, the Gibbs measure $\mu(dx) = Z^{-1}e^{-\beta V(x)}\,dx$ satisfies a *Poincarè inequality*:

[7] In fact, a complete proof of this result would require a more careful study of the domain of definition of the generator.

Theorem 4.8. *Let $V \in C^2(\mathbb{R}^d)$ and define $\mu(dx) = \frac{1}{Z}e^{-V}\,dx$. If*

$$\lim_{|x|\to+\infty}\left(\frac{|\nabla V(x)|^2}{2} - \Delta V(x)\right) \to +\infty, \tag{4.77}$$

then $\mu(dx)$ satisfies a Poincaré inequality with constant $\lambda > 0$: for every $f \in C^1(\mathbb{R}^d) \cap L^2(\rho_\beta)$ with $\int f\,\mu(dx) = 0$, there exists a constant $\lambda > 0$ such that

$$\lambda \|f\|^2_{L^2(\mu)} \leqslant \|\nabla f\|^2_{L^2(\mu)}. \tag{4.78}$$

For simplicity we will sometimes say that the potential V, rather than the corresponding Gibbs measure, satisfies Poincaré's inequality. Clearly, if $\mu(dx) = \frac{1}{Z}e^{-V}\,dx$ satisfies (4.78), so does $\mu_\beta(dx) = \frac{1}{Z}e^{-\beta V}\,dx$ for all positive β. Examples of potentials (Gibbs measures) that satisfy Poincaré's inequality are quadratic potentials in $\mathbb{R}^d$ of the form $V(x) = \frac{1}{2}x^T Dx$ with $D \in \mathbb{R}^d$ being a strictly positive symmetric matrix and the bistable potential $V(x) = -\frac{x^2}{2} + \frac{x^4}{4}$ in $\mathbb{R}$. A condition that ensures that the probability measure $\mu(dx) = \frac{1}{Z}e^{-V}\,dx$ satisfies Poincaré's inequality with constant λ is the convexity condition

$$D^2 V \geqslant \lambda I. \tag{4.79}$$

This is the Bakry-Emery criterion .

Notice also that, using the definition of the Dirichlet form (4.76) associated with the generator $\mathcal{L}$, we can rewrite Poincaré's inequality in the form

$$\lambda\beta^{-1}\mathrm{Var}(f) \leqslant D_{\mathcal{L}}(f, f), \tag{4.80}$$

for functions f in the domain of definition of the Dirichlet form. The assumption that the potential V satisfies Poincaré's inequality is equivalent of assuming that the generator $\mathcal{L}$ has a *spectral gap* in $L^2(\rho_\beta)$.

The proof of Theorem 4.8 or of the equivalent spectral gap estimate (4.80) is beyond the scope of this book. We remark that, just as in the case of Gaussian measures, we can link (4.80) to the study of an appropriate Schrödinger operator. Indeed, we have that (see Section 4.9) that

$$-\rho_\beta^{-1/2}\mathcal{L}\rho_\beta^{1/2} = -\beta^{-1}\Delta + \big(\frac{\beta}{4}|\nabla V|^2 - \frac{1}{2}\Delta V\big) := -\beta^{-1}\Delta + W(x) =: \mathcal{H}. \tag{4.81}$$

If we can prove that the operator $\mathcal{H}$ has a spectral gap in $L^2(\mathbb{R}^d)$, we can then use the expansion of $L^2(\mathbb{R}^d)$-functions in eigenfunctions of $\mathcal{H}$ to prove (4.80), see estimate (4.52) for the quadratic potential. This amounts to proving the estimate

$$\beta^{-1}\int_{\mathbb{R}^d} |\nabla h|^2\,dx + \int_{\mathbb{R}^d} W(x)h^2\,dx \geqslant \lambda \int_{\mathbb{R}^d} |h|^2\,dx.$$

To prove this, it is certainly sufficient to prove an inequality analogous to (4.52):

$$\|h\|^2_{L^2} \geqslant 2\lambda \|Wh\|_{L^2}\|\nabla h\|_{L^2} \tag{4.82}$$

for all C^1 functions with compact support. It is clear that the behavior of $W(\cdot)$ at infinity, Assumption 4.77, plays a crucial role in obtaining such an estimate.

Poincaré's inequality yields exponentially fast convergence to equilibrium, in the right function space.

Theorem 4.9. *Let $\rho(x,t)$ denote the solution of the Fokker-Planck equation* (4.67) *with $\rho_0(x) \in L^2(\mathbb{R}^d;\rho_\beta^{-1})$ and assume that the potential V satisfies a Poincaré inequality with constant λ. Then $\rho(x,t)$ converges to the Gibbs distribution ρ_β defined in* (4.69) *exponentially fast:*

$$\|\rho(\cdot,t)-\rho_\beta\|_{L^2(\rho_\beta^{-1})} \leqslant e^{-\lambda\beta^{-1}t}\|\rho_0(\cdot)-Z^{-1}e^{-\beta V}\|_{L^2(\rho_\beta^{-1})}. \tag{4.83}$$

Proof. We can rewrite (4.71) for the mean zero function $h-1$:

$$\frac{\partial(h-1)}{\partial t} = \mathcal{L}(h-1).$$

We multiply this equation by $(h-1)\,\rho_\beta$, integrate and use (4.76) to obtain

$$\frac{1}{2}\frac{d}{dt}\|h-1\|_\rho^2 = -D_{\mathcal{L}}(h-1,h-1).$$

We apply now Poincaré's inequality in the form (4.80) to deduce

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|h-1\|_\rho^2 &= -D_{\mathcal{L}}(h-1,h-1)\\ &\leqslant -\beta^{-1}\lambda\|h-1\|_\rho^2.\end{aligned}$$

Our assumption on ρ_0 implies that $h \in L^2(\rho_\beta)$. Consequently, the above calculation shows that

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

This estimate, together with the definition of h through $\rho=\rho_0 h$, leads to (4.83). □

The proof of the above theorem is based on the fact that the weighted L^2-norm of $h-1$ is a *Lyapunov function* for the backward Kolmogorov equation (4.71). In fact, we can construct a whole family of Lyapunov functions for the diffusion process X_t.

Proposition 4.10. *Let $\phi(\cdot) \in C^2(\mathbb{R}^d)$ be a convex function on $\mathbb{R}$ and define*

$$H(h) = \int \phi(h)\rho_\beta\,dx \tag{4.84}$$

with $\rho_\beta = \frac{1}{Z}e^{-\beta V}$. Furthermore, assume that V is a confining potential and let $h(t,\cdot)$ be the solution of (4.71)*. Then*

$$\frac{d}{d}(H(h(t,\cdot))) \leqslant 0. \tag{4.85}$$

Proof. We use (4.74) to calculate

$$\begin{aligned}\frac{d}{dt}H(h(t,\cdot)) &= \frac{d}{dt}\int\phi(h)\,\rho_\beta\,dx = \int\phi'(h)\frac{\partial h}{\partial t}\,\rho_\beta\,dx\\ &= \int\phi'(h)\mathcal{L}h\,\rho_\beta\,dx = -\int\nabla\phi'(h)\nabla h\,\rho_\beta\,dx\\ &= -\int\phi''(h)|\nabla h|^2\,\rho_\beta\,dx\\ &\leqslant 0,\end{aligned}$$

since $\phi(\cdot)$ is convex. □

In Theorem 4.9 we used $\phi(h) = (h-1)^2$. In view of the previous proposition, another choice is

$$\phi(h) = h \ln h - h + 1. \tag{4.86}$$

From a thermodynamic perspective, this a more natural choice: in this case the Lyapunov functional (4.84) becomes the (rather, a multiple of) *free energy functional*

$$F(\rho) = \int V \rho \, dx + \beta^{-1} \int \rho \ln \rho \, dx + \beta^{-1} \ln Z. \tag{4.87}$$

This follows from the following calculation (using the notation $\int f$ instead of $\int_{\mathbb{R}^d} f \, dx$):

$$\begin{aligned} H(\rho) &= \int \phi(h) \rho_\beta \, dx = \int \big(h \ln h - h + 1\big) \rho_\beta \, dx \\ &= \int \rho \ln \left(\frac{\rho}{\rho_\beta} \right) \, dx \\ &= \int \rho \ln \rho \, dx - \int \rho \ln \left(Z^{-1} e^{-\beta V} \right) \, dx \\ &= \beta \int V \rho \, dx + \int \rho \ln \rho \, dx + \ln Z. \end{aligned}$$

We will also refer to the functional

$$H(\rho) = \int \rho \ln \left(\frac{\rho}{\rho_\infty} \right) \, dx$$

as the *relative entropy* between the probability densities ρ and ρ_β. It is possible to prove exponentially fast convergence to equilibrium for the Smoluchowski equation under appropriate assumptions on the potential V. This result requires that that the measure $\frac{1}{Z} e^{-V}$ satisfies a *logarithmic Sobolev inequality* . See the discussion in Section 4.10.

4.6 Reversible Diffusions

The stationary ($X_0 \sim \rho_\beta(x) \, dx$) diffusion process X_t with generator (4.66) that we studied in the previous section is an example of a (time-) reversible Markov process.

Definition 4.11. *A stationary stochastic process X_t is time reversible if its law is invariant under time reversal: for every $T \in (0, +\infty)$ X_t and the time-reversed process X_{T-t} have the same distribution.*

This definition means that the processes X_t and X_{T-t} have the same finite dimensional distributions. Equivalently, for each $N \in \mathbb{N}^+$, a collection of times $0 = t_0 < t_1 \cdots < t_N = T$, and bounded measurable functions with compact support f_j, $j = 0, \ldots N$ we have that

$$\mathbb{E}_\mu \prod_{j=0}^{N} f_j(X_{t_j}) = \mathbb{E}_\mu \prod_{j=0}^{N} f_j(X_{T-t_j}), \tag{4.88}$$

where $\mu(dx)$ denotes the invariant measure of X_t and $\mathbb{E}_\mu$ denotes expectation with respect to μ.

In fact, the Markov property implies that it is sufficient to check (4.88). In particular, reversible diffusion processes can be characterized in terms of the properties of their generator. Indeed, time-reversal, which is a symmetry property of the process X_t, is equivalent to the selfadjointness (which is also a symmetry property) of the generator in the Hilbert space $L^2(\mathbb{R}^d;\mu)$.

Theorem 4.12. *A stationary Markov process X_t in $\mathbb{R}^d$ with generator $\mathcal{L}$ and invariant measure μ is reversible if and only if its generator is selfadjoint in $L^2(\mathbb{R}^d;\mu)$.*

Proof. [8] It is sufficient to show that (4.88) holds if and only if the generator is selfadjoint in $L^2(\mathbb{R}^d;\mu)$. Assume first (4.88). We take $N=1$ and $t_0=0,\ t_1=T$ to deduce that

$$\mathbb{E}_\mu\Big(f_0(X_0)f_1(X_T)\Big)=\mathbb{E}_\mu\Big(f_0(X_T)f_1(X_0)\Big),\quad \forall f_0,\ f_1\in L^2(\mathbb{R}^d;\mu).$$

This is equivalent to

$$\int \left(e^{\mathcal{L}t}f_0(x)\right) f_1(x)\,\mu(dx)=\int f_0(x)\left(e^{\mathcal{L}t}f_1(x)\right)\,\mu(dx),$$

i.e.

$$\langle e^{\mathcal{L}t}f_1,f_2\rangle_{L^2_\mu}=\langle f_1,e^{\mathcal{L}t}f_2\rangle_{L^2_\mu},\quad \forall f_1,\ f_2\in L^2(\mathbb{R}^d;\rho_s). \tag{4.89}$$

Consequently, the semigroup $e^{\mathcal{L}t}$ generated by $\mathcal{L}$ is selfadjoint. Differentiating (4.89) at $t=0$ gives that $\mathcal{L}$ is selfadjoint.

Conversely, assume that $\mathcal{L}$ is selfadjoint in $L^2(\mathbb{R}^d;\mu)$. We will use an induction argument. Our assumption of selfadjointness implies that (4.88) is true for $N=1$

$$\mathbb{E}_\mu\prod_{j=0}^{1}f_j(X_{t_j})=\mathbb{E}_\mu\prod_{j=0}^{1}f_j(X_{T-t_j}), \tag{4.90}$$

Assume that it is true for $N=k$. Using Equation (2.22) we have that

$$\begin{aligned}\mathbb{E}_\mu\prod_{j=0}^{k}f_j(X_{t_j}) &= \int\dots\int f_0(x_0)\mu(dx_0)\prod_{j=1}^{k}f_j(x_j)p(t_j-t_{j-1},x_{j-1},dx_j)\\ &= \mathbb{E}_\mu\prod_{n=0}^{k}f_j(X_{t_{j-1}})\\ &= \int\dots\int f_k(x_k)\mu(dx_k)\prod_{j=1}^{k}f_{j-1}(x_{j-1})p(t_j-t_{j-1},x_j,dx_{j-1}),\end{aligned} \tag{4.91}$$

[8] The calculations presented here are rather formal. In particular, we do not distinguish between a symmetric and a selfadjoint operator. For a fully rigorous proof of this result we need to be more careful with issues such as the domain of definition of the generator and its adjoint. See the discussion in Section 4.10.

where $p(t,x,\Gamma)$ denotes the transition function of the Markov process X_t. Now we show that (4.88) it is true for $N = k+1$. We calculate, using (4.90) and (4.91)

$$
\begin{aligned}
\mathbb{E}_\mu \prod_{j=1}^{k+1} f_j(X_{t_j}) &= E_\mu \Pi_{j=1}^{k} f_j(X_{t_j}) f_{k+1}(X_{t_{k+1}}) \\
&\overset{(2.22)}{=} \int \dots \int \mu(dx_0) f_0(x_0) \prod_{j=1}^{k} f_j(x_j) p(t_j - t_{j-1}, x_{j-1}, dx_j) \times \\
&\quad f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \\
&\overset{(4.91)}{=} \int \dots \int \mu(dx_k) f_0(x_k) \prod_{j=1}^{k} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \times \\
&\quad f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \\
&\overset{(4.91)}{=} \int \dots \int \mu(dx_k) f_0(x_k) f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \times \\
&\quad \prod_{j=1}^{k} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \\
&\overset{(4.90)}{=} \int \dots \int \mu(dx_{k+1}) f_0(x_{k+1}) \prod_{j=1}^{k+1} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \times \\
&\quad f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \\
&= \mathbb{E}_\mu \prod_{j=0}^{k+1} f_j(X_{T-t_j}).
\end{aligned}
$$

□

In the previous section we showed that the generator of the Smoluchowski dynamics

$$\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1} \Delta$$

is a selfadjoint operator in $L^2(\rho_\beta)$, which implies that the stationary solution of (4.66) is a reversible diffusion process. More generally, consider the Itô stochastic differential equation in $\mathbb{R}^d$ (see Section 3.2)

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \tag{4.92}$$

The generator of this Markov process is

$$\mathcal{L}\cdot = b(x) \cdot \nabla + \frac{1}{2} \mathrm{Tr}\big(\mathbf{\Sigma}(x) D^2\big), \tag{4.93}$$

where $\mathbf{\Sigma}(x) = \sigma(x)\sigma^T(x)$, which we assume to be strictly positive definite, see (4.6). Note that we can also write

$$\mathrm{Tr}\big(\mathbf{\Sigma}(x) D^2\big) = \mathbf{\Sigma}(x) : \nabla \cdot \nabla.$$

The Fokker-Planck operator is

$$\mathcal{L}^*\cdot = \nabla \cdot \Big(- b(x) \cdot + \frac{1}{2} \nabla \cdot \big(\mathbf{\Sigma}(x) \cdot \big)\Big). \tag{4.94}$$

We assume that the diffusion process has a unique invariant distribution which is the solution of the stationary Fokker-Planck equation

$$\mathcal{L}^*\rho_s = 0. \tag{4.95}$$

The stationary Fokker-Planck equation can be written as (see Equation (4.12))

$$\nabla \cdot J(\rho_s) = 0 \quad \text{with} \quad J(\rho_s) = b\rho_s - \frac{1}{2}\nabla \cdot (\mathbf{\Sigma}\rho_s).$$

Notice that we can write the invariant distribution ρ_s in the form

$$\rho_s = e^{-\Phi}, \tag{4.96}$$

where Φ can be thought of as a generalized potential.[9]

Let now X_t be an ergodic diffusion process with generator (4.93). We consider now the stationary solution of (4.92), i.e. we set $X_0 \sim \rho_s$. Our goal is to find under what conditions on the drift and diffusion coefficients this process is reversible. According to Theorem 4.12 it is sufficient to check under what conditions on the drift and diffusion coefficients the generator $\mathcal{L}$ is symmetric in $\mathcal{H} := L^2(\mathbb{R}^d; \rho_s(x)\,dx)$. Let $f,\, h \in \mathcal{H} \cap C^2(\mathbb{R}^d)$. We perform integrations by parts to calculate:

$$\int b \cdot \nabla f h \rho_s \, dx = -\int f b \cdot \nabla h \rho_s \, dx - \int f h \nabla \cdot (b\rho_s)\, dx$$

and

$$\begin{aligned}
\int \big(\mathbf{\Sigma}\nabla \cdot \nabla f\big) h \rho_s \, dx &= -\int \mathbf{\Sigma}\nabla f \cdot \nabla h \rho_s \, dx - \int \big(\nabla f h\big) \cdot \nabla \cdot (\mathbf{\Sigma}\rho_s)\, dx \\
&= -\int \mathbf{\Sigma}\nabla f \cdot \nabla h \rho_s \, dx + \int \big(f \nabla h\big) \cdot \nabla\big(\mathbf{\Sigma}\rho_s\big)\, dx \\
&\quad + \int f h \nabla \cdot \big(\nabla \cdot (\mathbf{\Sigma}\rho_s)\big)\, dx.
\end{aligned}$$

We combine the above calculations and use the stationary Fokker-Planck equation (4.95) and the definition of the stationary probability flux $J_s := J(p_s)$ to deduce that

$$\begin{aligned}
\langle -\mathcal{L}f, g\rangle_\rho &:= \int \big(-\mathcal{L}f\big) h \rho_s \, dx \\
&= \frac{1}{2}\int \mathbf{\Sigma}\nabla f \cdot \nabla h \rho_s \, dx + \int f \nabla h \cdot J_s \, dx \\
&= \frac{1}{2}\langle \mathbf{\Sigma}\nabla f, \nabla h\rangle_\rho + \langle f, \rho_s^{-1}\nabla h \cdot J_s\rangle_\rho.
\end{aligned}$$

[9] Note that we can incorporate the normalization constant in the definition of Φ. Alternatively, we can write $\rho = \frac{1}{Z}e^{-\Phi}$, $Z = \int e^{-\Phi}\, dx$. See, for example, the formula for the stationary distribution of a one dimensional diffusion process with reflecting boundary conditions, Equation (4.35). We can write it in the form

$$p_s(x) = \frac{1}{Z}e^{-\Phi} \quad \text{with} \quad \Phi = \log(\sigma(x)) - \left(2\int_\ell^x \frac{b(y)}{\sigma(y)}\, dy\right).$$

The generator $\mathcal{L}$ is symmetric if and only if the last term on the righthand side of the above equation vanishes, i.e. if and only if the stationary probability flux vanishes:

$$J(\rho_s) = 0. \tag{4.97}$$

This is the *detailed balance* condition. From the detailed balance condition (4.97) we obtain a relation between the drift vector b, the diffusion matrix $\mathbf{\Sigma}$ and the generalized potential Φ:

$$b = \frac{1}{2}\rho_s^{-1}\nabla \cdot \big(\mathbf{\Sigma}\rho_s\big) = \frac{1}{2}\nabla \cdot \mathbf{\Sigma} - \frac{1}{2}\mathbf{\Sigma}\nabla\Phi. \tag{4.98}$$

We summarize the above calculations

Proposition 4.13. *Let X_t denote the stationary process* (4.93) *with invariant distribution ρ_s, the solution of* (4.95)*. Then X_t is reversible if and only if the detailed balance condition* (4.97) *holds or, equivalently, there exists a scalar function Φ such that* (4.98) *holds.*

Remark 4.14. The term $\frac{1}{2}\nabla \cdot \mathbf{\Sigma}(x)$ in (4.98) is the Itô-to-Stratonovich correction, see Section 3.2. Indeed, if we interpret the noise in (4.92) in the Stratonovich sense,

$$dX_t = b(X_t)\,dt + \sigma(X_t) \circ dW_t, \tag{4.99}$$

then the condition for reversibility (4.98) becomes

$$b(x) = -\frac{1}{2}\mathbf{\Sigma}(x)\nabla\Phi(x). \tag{4.100}$$

For the stationary process X_t, the solution of the Stratonovich stochastic differential equation (4.99) with $X_0 \sim \rho_s$, the following conditions are quivalent:

i. The process X_t is reversible.

ii. The generator $\mathcal{L}$ is selfadjoint in $L^2(\mathbb{R}^d; \rho_s)$.

iii. There exists a scalar function Φ such that (4.100) holds.

Consider now an arbitrary ergodic diffusion process X_t, the solution of (4.92) with invariant distribution ρ_s. We can decompose this process into a reversible and a nonreversible part in the sense that the generator can be decomposed into a symmetric and antisymmetric part in the space $L^2(\mathbb{R}^d; \rho_s)$. To check this, we add the subtract the term $\rho_s^{-1}J_s \cdot \nabla$ from the generator $\mathcal{L}$ and use the formula for the stationary probability flux:

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}\mathbf{\Sigma}\nabla \cdot \nabla + \big(b - \rho_s^{-1}J_s\big) \cdot \nabla + \rho_s^{-1}J_s \cdot \nabla \\
&= \frac{1}{2}\mathbf{\Sigma}\nabla \cdot \nabla + \rho_s^{-1}\nabla \cdot \big(\mathbf{\Sigma}\rho_s\big) \cdot \nabla + \rho_s^{-1}J_s \cdot \nabla \\
&= \frac{1}{2}\rho_s^{-1}\nabla \cdot \big(\mathbf{\Sigma}\rho_s\nabla\big) + \rho_s^{-1}J_s \cdot \nabla \\
&=: \mathcal{S} + \mathcal{A}.
\end{aligned}$$

Clearly, the operator $\mathcal{S} = \frac{1}{2}\rho_s^{-1}\nabla\cdot\big(\mathbf{\Sigma}\rho_s\nabla\big)$ is symmetric in $L^2(\mathbb{R}^d;\rho_s)$. To prove that $\mathcal{A}$ is antisymmetric we use the stationary Fokker-Planck equation written in the form $\nabla\cdot J_s = 0$:

$$\begin{aligned}\langle \mathcal{A}f, h\rangle_\rho &= \int J_s\cdot\nabla f\, h\, dx \\ &= -\int J_s\cdot\nabla f J_s\cdot\nabla h\, dx - \int fh\nabla\cdot J_s\, dx \\ &= -\langle f, \mathcal{A}h\rangle_\rho.\end{aligned}$$

The generator of an arbitrary diffusion process in $\mathbb{R}^d$ can we written in the useful form

$$\mathcal{L} = \rho_s^{-1}J_s\cdot\nabla + \frac{1}{2}\rho_s^{-1}\nabla\cdot\big(\mathbf{\Sigma}\rho_s\nabla\big), \tag{4.101}$$

where the drift (advection) term on the right hand side is antisymmetric, whereas the second order, divergence-form part is symmetric in $L^2(\mathbb{R}^d;\rho_s)$.

4.7 Eigenfunction Expansions for Reversible Diffusions

Let X_t denote the generator of a reversible diffusion. We write the generator $\mathcal{L}$ in the form, see Equation (4.101)

$$\mathcal{L} = \frac{1}{2}\rho_s^{-1}\nabla\cdot\big(\mathbf{\Sigma}\rho_s\nabla\big). \tag{4.102}$$

The corresponding Dirichlet form is

$$D_{\mathcal{L}}(f) := \langle -\mathcal{L}f, f\rangle_\rho = \frac{1}{2}\langle \mathbf{\Sigma}\nabla f, \nabla h\rangle_\rho. \tag{4.103}$$

We assume that the diffusion matrix $\mathbf{\Sigma}$ is uniformly positive definite with constant α, Equation (4.6). This implies that

$$D_{\mathcal{L}}(f) \geqslant \frac{\alpha}{2}\int_{\mathbb{R}^d}|\nabla f|^2 e^{-\Phi}\, dx,$$

where we have introduced the generalized potential Φ, $\rho_s = e^{-\Phi}$. In order to prove that the generator (4.102) has a spectral gap we need to show that the probability measure $\rho_s\, dx$ or, equivalently, the potential Φ, satisfied a Poincaré inequality. For this it is sufficient to show that the generalized potential satisfies assumption (4.77) in Theorem 4.8.

Assume now that the generator $\mathcal{L}$ has a spectral gap:

$$\lambda\mathrm{Var}(f) \leqslant D_{\mathcal{L}}(f,f). \tag{4.104}$$

Then $-\mathcal{L}$ is a nonnegative, selfadjoint operator in $L^2(\mathbb{R}^d;\rho_s)$ with discrete spectrum. The eigenvalue problem for the generator is

$$-\mathcal{L}\phi_n = \lambda_n\phi_n, \quad n = 0, 1, \ldots \tag{4.105}$$

Notice that $\phi_0 = 1$ and $\lambda_0 = 0$. The eigenvalues of the generator are real and nonnegative:

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \ldots$$

Furthermore, the eigenfunctions $\{\phi_j\}_{j=0}^{\infty}$ span $L^2(\mathbb{R}^d;\rho_s)$: we can express every element of $L^2(\mathbb{R}^d;\rho_s)$ in the form of a generalized Fourier series:

$$f=\sum_{n=0}^{\infty}\phi_n f_n,\quad f_n=(\phi_n,\phi_n)_\rho \tag{4.106}$$

with $(\phi_n,\phi_m)_\rho=\delta_{nm}$. This enables us to solve the time dependent Fokker-Planck equation in terms of an eigenfunction expansion, exactly as in the case of a one dimensional diffusion process with reflecting boundary conditions that we studied in Section 4.3. The calculation is exactly the same as for the one dimensional problem: consider first the initial value problem for the transition probability density $p(x,t|x_0)$:

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \tag{4.107a}$$

$$p(x,0|x_0) = \delta(x-x_0). \tag{4.107b}$$

The function $h(x,t|x_0)=p(x,t|x_0)\rho_s^{-1}(x)$ is the solution of the initial value problem

$$\frac{\partial h}{\partial t} = \mathcal{L}h, \tag{4.108a}$$

$$h(x,0|x_0) = \rho_s^{-1}(x)\delta(x-x_0). \tag{4.108b}$$

We solve this equation using separation of variables and the superposition principle. Transforming back, we finally obtain

$$p(x,t|x_0)=\rho_s(x)\left(1+\sum_{\ell=1}^{\infty}e^{-\lambda_\ell t}\phi_\ell(x)\phi_\ell(x_0)\right). \tag{4.109}$$

When the initial condition is a random variable with probability density function $\rho_0(x)$ then the formula for the probability distribution function $p(x,t)$, the solution of the Fokker-Planck equation with $p(x,0)=\rho_0(x)$ is

$$p(x,t)=\rho_s(x)\left(1+\sum_{\ell=1}^{+\infty}e^{-\lambda_\ell t}\phi_\ell(x)\rho_\ell\right),\quad \rho_\ell=\int_{\mathbb{R}^d}\rho_0(x)\phi_\ell(x)\,dx. \tag{4.110}$$

We use now (2.60) to define the stationary autocorrelation matrix

$$C(t):=\mathbb{E}(X_t\otimes X_0)=\int\int x_0\otimes x p(x,t|x_0)\rho_s(x_0)\,dxdx_0. \tag{4.111}$$

Substituting (4.109) into (4.111) we obtain

$$C(t)=\sum_{\ell=0}^{\infty}e^{-\lambda_\ell|t|}\boldsymbol{\alpha}_\ell\otimes\boldsymbol{\alpha}_\ell,\quad \boldsymbol{\alpha}_\ell=\int_{\mathbb{R}^d}x\phi_k(x)\rho_s(x)\,dx, \tag{4.112}$$

with $\lambda_0=1,\ \phi_0=1$. Using now (1.7) we can obtain a formula for the spectral density, which in the multdimensional case is a $d\times d$ matrix. We present here for the formula in one dimension

$$S(\omega)=\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\alpha_k^2\,\lambda_k}{\lambda_k^2+\omega^2}. \tag{4.113}$$

The interested reader is invited to supply the details of these calculations (see Exercise 11). It is important to note that for a reversible diffusion the spectral density is given as a sum of Cauchy-Lorentz functions , which is the spectral density of the Ornstein-Uhlenbeck process.

4.8 Markov Chain Monte Carlo

Suppose that we are given a probability distribution $\pi(x)$ in $\mathbb{R}^d$ that is known up to the normalization constant.[10] Our goal is to sample from this distribution and to calculate expectation values of the form

$$\mathbb{E}_\pi f = \int_{\mathbb{R}^d} f(x)\pi(x)\,dx, \tag{4.114}$$

for particular choices of functions $f(x)$. A natural approach to solving this problem is to construct an ergodic diffusion process whose invariant distribution is $\pi(x)$. We then run the dynamics for a sufficiently long time, until it reaches the stationary regime. The equilibrium expecation (4.114) can be calculated by taking the long time average and using the ergodic theorem:

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T f(X_s)\,ds = \mathbb{E}_\pi f. \tag{4.115}$$

This is an example of the *Markov Chain Monte Carlo (MCMC)* methodology.

There are many different diffusion processes that we can use: we have to choose the drift and diffusion coefficients so that the stationary Fokker-Planck equation is satisfied:

$$\nabla\cdot\big(-b\pi + \frac{1}{2}\nabla\cdot\big(\mathbf{\Sigma}\pi\big)\big) = 0. \tag{4.116}$$

We have to solve the "inverse problem" for this partial differential equation: given its solution $\pi(x)$, we want to find the coefficients $b(x)$ and $\mathbf{\Sigma}(x)$ that this equation is satisfied. Clearly, there are (infinitely) many solutions to this problem. We can restrict the class of drift and diffusion coefficients (and of the corresponding diffusion process) that we consider by imposing the detailed balance condition $J(\pi) = 0$. The stationary Fokker-Planck equation is

$$-b\pi + \frac{1}{2}\nabla\cdot(\mathbf{\Sigma}\pi) = 0. \tag{4.117}$$

Thus, we consider reversible diffusion processes in order to sample from $\pi(x)$. Even when we impose the detailed balance condition, there is still a lot of freedom in choosing the drift and diffusion coefficients. A natural choice is to consider a constant diffusion matrix, $\mathbf{\Sigma} = 2I$. The drift is

$$b = \pi^{-1}\nabla\pi = \nabla\log\pi.$$

This leads to the Smoluchowski dynamics that we studied in Section 4.5.[11]

$$dX_t = \nabla\ln\pi(X_t)\,dt + \sqrt{2}\,dW_t. \tag{4.118}$$

Notice that in order to be able to construct this diffusion process we do not need to know the normalization constant since only the gradient of the logarithm of $\pi(x)$ appears in (4.118). Provided that the "potential"

[10] The calculation of the normalization constant requires the calculation of an integral (the partition function) in a high dimensional space which might be computationally ver expensive.

[11] In the statistics literature this is usually called the Langevin dynamics. We will use this term for the seconder order stochastic differential equation that is obtained after adding dissipation and noise to a Hamiltonian system, see Chapter **??**. Using the terminology that we will introduce there, the dynamics (4.118) corresponds to the overdamped Langevin dynamics.

$V(x) = -\log \pi(x)$ satisfies Poincaré's inequality, Theorem 4.8, we have exponentially fast convergence to the target distribution in $L^2(\mathbb{R}^d; \pi)$, Theorem 4.9. The rate of convergence to the target distribution $\pi(x)$ depends only on (the tails of) the distribution itself, since the Poincaré constant depends only on the potential.

When the target distribution is multimodal (or equivalently, $V(x) = -\log \pi(x)$ has a lot of local minima) convergence to equilibrium for the dynamics (4.118) might be slow (see Chapter **??**). In such a case, it might be useful to modify the dynamics either through the drift or the diffusion in order to facilitate the escape of the dynamics for the local minima of $V(x)$. Ideally, we would like to choose the drift and diffusion coefficients in such a way that the corresponding dynamics converges to the target distribution as quickly as possible.[12] For the reversible dynamics, for which the generator is a self-adjoint operator in $L^2(\mathbb{R}^d; \pi)$, the optimal choice of the diffusion process is the one that maximizes the first nonzero eigenvalue of the generator, since this determines the rate of convergence to equilibrium. The first nonzero eigenvalue can be expressed in terms of the Rayleigh quotient:

$$\lambda_1 = \min_{\phi \in D(\mathcal{L})\ \phi \neq 0} \frac{D_{\mathcal{L}}(\phi)}{\|\phi\|_\rho}, \tag{4.119}$$

where $D_{\mathcal{L}}(\phi) = \langle -\mathcal{L}\phi, \phi \rangle_\rho$. The optimal choice of the drift and diffusion coefficients is the one that maximizes λ_1, subject to the detailed balance condition:

$$\lambda_1 = \max_{b,\, \boldsymbol{\Sigma}\ J(\rho_s)=0} \min_{\phi \in D(\mathcal{L})\ \phi \neq 0} \frac{D_{\mathcal{L}}(\phi)}{\|\phi\|_\rho}. \tag{4.120}$$

We can also introduce perturbations in the drift of the Smoluchowski dynamics that lead to nonreversible diffusions. We consider the following dynamics

$$dX_t^\gamma = (\nabla \pi(X_t^\gamma) + \boldsymbol{\gamma}(X_t^\gamma))\, dt + \sqrt{2}\, dW_t. \tag{4.121}$$

where $\boldsymbol{\gamma}(x)$ a smooth vector field that has to be chosen so that the invariant distribution of (4.121) is still $\pi(x)$. The stationary Fokker-Planck equation becomes

$$\nabla \cdot (\boldsymbol{\gamma}(x)\pi(x)) = 0. \tag{4.122}$$

Consequently, all divergence-free perturbations vector fields, with respect to the target distribution, can be used in order to construct nonreversible ergodic dynamics whose invariant distribution is $\pi(x)$. There exist many such vector fields, for example

$$\boldsymbol{\gamma}(x) = J \nabla \log \pi(x), \quad J = -J^T.$$

We can then ask whether it is possible to accelerate convergence to the target distribution by choosing the nonreversible perturbation, i.e. the matrix J, appropriately. It is reasonable to expect that a nonreversible perturbation that facilitates the escape of the dynamics from the local minima of $V(x) = -\nabla \log \pi(x)$ would speed up convergence of the modified dynamics X_t^γ to equilibrium.

[12] When implementing the MCMC algorithm we need to discretize the stochastic differential equation and take the discretization and Monte Carlo errors into account. See Section **??**.

4.9 Reduction to a Schrödinger Operator

In Sections 4.4 and 4.5 we used the fact that the generator of the Ornstein-Uhlenbeck process and of the generator of the Smoluchowski dynamics can be transformed to a Schrödinger operator, formulas (4.49) and (4.81). We used this transformation in order to study Poincaré inequalities for the Gibbs measure $\mu(dx) = \frac{1}{Z}e^{-V(x)}\,dx$. In this section we study the connection between the generator $\mathcal{L}$, the Fokker-Planck operator and an appropriately defined Schrödinger-like operator in more detail.

We start by writing the generator of an ergodic diffusion process in the form (4.101)

$$\begin{aligned} \mathcal{L}\cdot &= \frac{1}{2}\rho_s^{-1}\nabla\cdot\big(\mathbf{\Sigma}\rho_s\nabla\cdot\big) + \rho_s^{-1}J_s\cdot\nabla\cdot && \text{(4.123a)}\\ &=: \mathcal{S}+\mathcal{A}, && \text{(4.123b)}\end{aligned}$$

where ρ_s denotes the invariant density. The operators $\mathcal{S}$ and $\mathcal{A}$ are symmetric and antisymmetric, respectively, in the function space $L^2(\mathbb{R}^d;\rho_s)$. Using the definition of the $L^2(\mathbb{R}^d)$-adjoint, we can check that the Fokker-Planck operator can be written in the form

$$\begin{aligned} \mathcal{L}^*\cdot &= \frac{1}{2}\nabla\cdot\big(\rho_s\mathbf{\Sigma}\nabla(\rho_s^{-1}\cdot)\big) - J_s\cdot\nabla\big(\rho_s^{-1}\cdot\big), && \text{(4.124a)}\\ &=: \mathcal{S}^*+\mathcal{A}^*. && \text{(4.124b)}\end{aligned}$$

The operators $\mathcal{S}^*$ and $\mathcal{A}^*$ are symmetric and antisymmetric, respectively, in the function space $L^2(\mathbb{R}^d;\rho_s^{-1})$. We introduce the following operator (compare with Equation (4.81))

$$\mathcal{H}\cdot := \rho_s^{1/2}\mathcal{L}\big(\rho_s^{-1/2}\cdot\big), \tag{4.125}$$

acting on twice differentiable functions that belong in $L^2(\mathbb{R}^d)$.

Lemma 4.15. *The operator $\mathcal{H}$ defined in* (4.125) *has the form*

$$\mathcal{H}\cdot = \frac{1}{2}\nabla\cdot\big(\mathbf{\Sigma}\nabla\cdot\big) + W(x)\cdot + \mathcal{A}\cdot, \tag{4.126}$$

where $\mathcal{A}$ denotes the antisymmetric part of $\mathcal{L}$ and the scalar function W is given by the formula

$$W(x) = \sqrt{\rho_s}\mathcal{L}\sqrt{\rho_s^{-1}}. \tag{4.127}$$

Proof. Let $f\in C^2(\mathbb{R}^d)$. We calculate

$$\begin{aligned} \mathcal{H}f &= \rho_s^{1/2}\mathcal{L}\big(\rho_s^{-1/2}f\big) = \rho_s^{1/2}\mathcal{S}\big(\rho_s^{-1/2}f\big) + \rho_s^{1/2}\mathcal{A}\big(\rho_s^{-1/2}f\big)\\ &= \rho_s^{1/2}\mathcal{S}\big(\rho_s^{-1/2}f\big) + \mathcal{A}f + f\sqrt{\rho_s}\mathcal{A}\sqrt{\rho_s^{-1}}, \end{aligned} \tag{4.128}$$

since $\mathcal{A}$ is a first order differential operator. Let now ψ be another C^2 function. We have the identity

$$\begin{aligned} \nabla\cdot\big(\mathbf{\Sigma}\rho_s\nabla(f\psi)\big) &= \nabla\cdot\big(\mathbf{\Sigma}\nabla f\big)\rho_s\psi + \nabla\cdot\big(\mathbf{\Sigma}\rho_s\nabla\psi\big)f\\ &\quad + \big[\mathbf{\Sigma}\psi\nabla\rho_s + 2\mathbf{\Sigma}\rho_s\nabla\psi\big]\cdot f. \end{aligned}$$

In particular, for $\psi = \sqrt{\rho_s^{-1}}$ the second term on the righthand side of the above equation vanishes:

$$\nabla \cdot \Big(\mathbf{\Sigma} \rho_s \nabla \left(f \sqrt{\rho_s^{-1}} \right) \Big) = \nabla \cdot \Big(\mathbf{\Sigma} \nabla f \Big) \sqrt{\rho_s} + \nabla \cdot \Big(\mathbf{\Sigma} \rho_s \nabla \sqrt{\rho_s^{-1}} \Big) f.$$

This equation implies that

$$\rho_s^{1/2} \mathcal{S}\big(\rho_s^{-1/2} f\big) = \frac{1}{2} \nabla \cdot \Big(\mathbf{\Sigma} \nabla f \Big) + \rho_s^{1/2} \mathcal{S}\big(\rho_s^{-1/2}\big).$$

We combine this with (4.128) to obtain (4.126). □

The operator $\mathcal{H}$ given by (4.126) is reminiscent of a Schrödinger operator in a magnetic field. We can write the "effective potential" W in a more explicit form. We use the notation $\rho_s = e^{-\Phi}$, see Equation (4.96). We have

$$\begin{aligned} \rho_s^{1/2} \mathcal{S} \rho_s^{-1/2} &= \frac{1}{2} e^{\Phi/2} \nabla \cdot \Big(\mathbf{\Sigma} e^{-\Phi} \nabla e^{\Phi/2} \Big) \\ &= \frac{1}{4} e^{\Phi/2} \nabla \cdot \Big(\mathbf{\Sigma} e^{-\Phi/2} \nabla \Phi \Big) \\ &= \frac{1}{4} \nabla \cdot (\mathbf{\Sigma} \nabla \Phi) - \frac{1}{8} \mathbf{\Sigma} \nabla \Phi \cdot \nabla \Phi. \end{aligned}$$

Furthermore,

$$\begin{aligned} \rho_s^{1/2} \mathcal{A} \rho_s^{-1/2} &= \rho_s^{-1/2} J_s \cdot \nabla \rho_s^{-1/2} \\ &= \frac{1}{2} \rho_s^{-1} J_s \cdot \nabla \Phi. \end{aligned}$$

We combine these calculations to obtain

$$W(x) = \frac{1}{4} \nabla \cdot (\mathbf{\Sigma} \nabla \Phi) - \frac{1}{8} \mathbf{\Sigma} \nabla \Phi \cdot \nabla \Phi + \frac{1}{2} \rho_s^{-1} J_s \cdot \nabla \Phi. \tag{4.129}$$

For reversible diffusions the stationary probability current vanishes and the operator $\mathcal{H}$ becomes

$$\mathcal{H} \cdot = \frac{1}{2} \nabla \cdot \big(\mathbf{\Sigma} \nabla \cdot \big) + \left(\frac{1}{4} \nabla \cdot (\mathbf{\Sigma} \nabla \Phi) - \frac{1}{8} \mathbf{\Sigma} \nabla \Phi \cdot \nabla \Phi \right) \cdot . \tag{4.130}$$

On the other hand, for nonreversible perturbations of the Smoluchwoski (overdamped Langevin) dynamics, Equation (4.121) whose generator is given by

$$\mathcal{L} = (-\nabla V + \gamma) \cdot \nabla + \Delta, \quad \nabla \cdot \big(\gamma e^{-V} \big) = 0, \tag{4.131}$$

operator $\mathcal{H}$ takes the form

$$\mathcal{H} = \Delta + \left(\frac{1}{2} \Delta V - \frac{1}{4} \big| \nabla V \big|^2 + \frac{1}{2} \gamma \cdot \nabla V \right) + \gamma \cdot \nabla. \tag{4.132}$$

It is important to keep in mind that the three operators $\mathcal{L}$, $\mathcal{L}^*$ and $\mathcal{H}$ are defined in different function spaces, in (dense subsets of) $L^2(\mathbb{R}^d; \rho_s)$, $L^2(\mathbb{R}^d; \rho_s^{-1})$ and $L^2(\mathbb{R}^d)$, respectively. These operators are related trough

a (unitary) transformation.[13] We have already shown this for the map from $\mathcal{L}$ to $\mathcal{H}$, Equation (4.125): define the multiplication operator

$$U_{\mathcal{L},\mathcal{H}} = \sqrt{\rho_s}:\ L^2(\mathbb{R}^d;\rho_s) \mapsto L^2(\mathbb{R}^d).$$

This is a unitary operators from $L^2(\rho_\beta)$ to $L^2(\mathbb{R}^d)$:

$$\langle U_{\mathcal{L},\mathcal{H}} f, U_{\mathcal{L},\mathcal{H}} h\rangle_{L^2} = \langle f,h\rangle_\rho \quad \forall f,\, h \in L^2(\mathbb{R}^d;\rho_s).$$

Clearly, $U_{\mathcal{L},\mathcal{H}} = \sqrt{\rho_s}$. We can then rewrite (4.125) in the form

$$\mathcal{H} = U_{\mathcal{L},\mathcal{H}}\mathcal{L}U^{-1}_{\mathcal{L},\mathcal{H}}.$$

The generator and Fokker-Planck operators are also unitarily equivalent, up to a sign change. We denote the generator defined in (4.123) by $\mathcal{L}_{J_s}$, to emphasize the dependence on the stationary flux J_s. We use (4.123) and (4.124) to calculate, for every $f \in L^2(\mathbb{R}^d;\rho_s^{-1})$

$$\begin{aligned}\rho_s\mathcal{L}_{J_s}\big(\rho_s^{-1}f\big) &= \rho_s\big(\mathcal{S}+\mathcal{A}\big)\big(\rho_s^{-1}f\big)\\ &= \big(\mathcal{S}^*-\mathcal{A}^*\big)f =: \mathcal{L}^*_{-J_s}f.\end{aligned}$$

Introducing then the unitary multiplication operator

$$U_{\mathcal{L},\mathcal{L}^*} = \rho_s:\ L^2(\mathbb{R}^d;\rho_s) \mapsto L^2(\mathbb{R}^d;\rho_s^{-1}),$$

we can write

$$U_{\mathcal{L},\mathcal{L}^*}\mathcal{L}_{J_s}U^{-1}_{\mathcal{L},\mathcal{L}^*} = \mathcal{L}^*_{-J_s}.$$

It is important to note that, under this transformation, the symmetric part of the generator (corresponding to the reversible part of the dynamics) is mapped to the symmetric part of the Fokker-Planck operator whereas the antisymmetric part (corresponding to the nonreversible part of the dynamics) is mapped to minus the antisymmetric part of the Fokker-Planck operator. As an example, consider the generator of the Smoluchowski dynamics, perturbed by the divergence-free vector field γ we have (see Equation (4.131))

$$\mathcal{L}_{-\gamma} = (-\nabla V - \gamma)\cdot\nabla + \Delta, \quad \nabla\cdot\big(\gamma e^{-V}\big) = 0,$$

and similarly for the corresponding Fokker-Planck operator:

$$\begin{aligned}\mathcal{L}^*_\gamma &= \nabla\cdot\Big(\nabla V - \gamma + \nabla\Big) &\text{(4.133a)}\\ &= \Delta + \nabla V\cdot\nabla - \gamma\cdot\nabla - \nabla\cdot\gamma\,, &\text{(4.133b)}\end{aligned}$$

and

$$\mathcal{L}^*_{-\gamma} = \Delta + \nabla V\cdot\nabla + \gamma\cdot\nabla + \nabla\cdot\gamma$$

We summarize these calculations in the following proposition.

[13] Two operators $A_1,\ A_2$ defined in two Hilbert spaces $H_1,\ H_2$ with inner products $\langle\cdot,\cdot\rangle_{H_1},\ \langle\cdot,\cdot\rangle_{H_2}$, respectively, are called unitarily equivalent if there exists a unitary transformation $U:\ H_1 \mapsto H_2$ (i.e. $\langle Uf, Uh\rangle_{H_2} = \langle f,h\rangle_{H_1},\ \forall f,\ h \in H_1$) such that

$$UA_1U^{-1} = A_2.$$

When the operators $A_1,\ A_2$ are unbounded we need to be more careful with their domain of definition.

Proposition 4.16. *The operators $\mathcal{L}_{J_s}$, $\mathcal{L}^*_{-J_s}$ and $\mathcal{H}$ defined on $L^2(\mathbb{R}^d;\rho_s)$, $L^2(\mathbb{R}^d;\rho_s^{-1})$ and $L^2(\mathbb{R}^d)$, respectively, are unitarily equivalent:*

$$\rho_s \mathcal{L}_{J_s} \rho_s^{-1} = \mathcal{L}^*_{-J_s}, \tag{4.134a}$$

$$\sqrt{\rho_s}\mathcal{L}_{J_s}\sqrt{\rho_s^{-1}} = \mathcal{H}, \tag{4.134b}$$

$$\sqrt{\rho_s^{-1}}\mathcal{L}^*_{-J_s}\sqrt{\rho_s} = \mathcal{H}. \tag{4.134c}$$

This proposition is presented graphically in Figure 4.9.

For reversible diffusions, i.e. when the stationary probability flux vanishes, no sign reversal is needed and the generator, the Fokker-Planck operator and the corresponding Schrödinger-like operator $\mathcal{H}$ are unitarily equivalent. Consider now this case, $J_s = 0$ and assume that the generalized potential Φ is such that the generator $\mathcal{L}$ has a spectral gap, Assumption (4.104) holds. Then the generator has discrete nonnegative spectrum and its eigenfucntions span $L^2(\mathbb{R}^d;\rho_s)$, see Section 4.7. Then the three operators $\mathcal{L}$, $\mathcal{L}^*$ and $\mathcal{H}$ have the same eigenvalues and its eigenfunctions are related through a simple transformation. Indeed, let $(T_i,\, H_i)$, $i = 1, 2$ be two unitarily equivalent self-adjoint operators with discrete spectrum in the Hilbert spaces $H_1,\, H_2$. Consider the eigenvalue problem for T_1:

$$T_1\psi_k^1 = \lambda_k^1\psi_k^1, \quad k = 1, \dots$$

Let $U^{-1}T_2U = T_1$. Substituting this formula in the above equation and multiplying the resulting equation by U we deduce that

$$\psi_k^2 = U\psi_k^1 \quad \text{and} \quad \lambda_k^2 = \lambda_k^1.$$

For eigenfunctions of the generator, the Fokker-Planck operator and the operator $\mathcal{H}$, denoted by $\psi_k^{\mathcal{L}}$, $\psi_k^{\mathcal{L}^*}$ and $\psi_k^{\mathcal{H}}$, respectively, are related through the formulas

$$\psi_k^{\mathcal{L}^*} = \rho_s^{-1}\psi_k^{\mathcal{L}} \qquad \psi_k^{\mathcal{H}} = \sqrt{\rho_s^{-1}}\psi_k^{\mathcal{L}}. \tag{4.135}$$

Mapping the eigenvalue problem for the Fokker-Planck operator (or the generator) to the eigenvalue problem for a Schrödinger operator is very useful, since the spectral problem for such operators is very well studied. Similarly, we can map the Fokker-Planck equation to a Schrödinger equation in imaginary time. Let us consider the Smoluchowski Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \beta^{-1}\nabla\cdot\left(e^{-\beta V}\nabla\left(e^{\beta V}p\right)\right). \tag{4.136}$$

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

$$\frac{\partial \psi}{\partial t} = \beta^{-1}\Delta\psi - U(x)\psi, \quad U(x) := \frac{\beta|\nabla V|^2}{4} - \frac{\Delta V}{2}. \tag{4.137}$$

The operator $\mathcal{H}$ can be written as the product of two first order operators:

$$\mathcal{H} = \beta^{-1}\mathcal{A}^*\mathcal{A}, \quad \mathcal{A} = \nabla + \frac{\beta\nabla U}{2}, \quad \mathcal{A}^* = -\nabla + \frac{\beta\nabla U}{2},$$

or

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

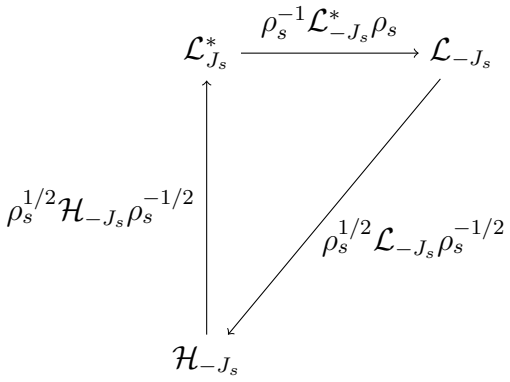


Figure 4.1: Transformation between the Fokker-Planck operator (4.124), the generator (4.123) and the Schrödinger operator (4.126).

4.10 Discussion and Bibliography

The proof of existence and uniqueness of classical solutions for the Fokker-Planck equation of a uniformly elliptic diffusion process with smooth drift and diffusion coefficients, Theorem 4.2, can be found in [24]. See also [96], in particular Theorem 1.1.9, for rigorous results on the backward and forward Kolmogorov equation for diffusion processes. Parabolic PDEs (in particular in bounded domains) are studied in detail in [21].

The condition that solutions to the Fokker-Planck equation do not grow too fast, see Definition 4.1, is necessary to ensure uniqueness. In fact, there are infinitely many solutions of

$$\begin{aligned} \frac{\partial p}{\partial t} &= \Delta p \quad \text{in } \mathbb{R}^d \times (0,T) \\ p(x,0) &= 0. \end{aligned}$$

Each of these solutions besides the trivial solution $p = 0$ grows very rapidly as $x \to +\infty$. More details can be found in [45, Ch. 7].

The Fokker-Planck equation is studied extensively in Risken's monograph [86]. See also [28, 38, 95, 102]. In these references several examples of diffusion processes whose Fokker-Planck equation can be solved analytically can be found. The connection between the Fokker-Planck equation and stochastic differential equations is presented in Chapter **??**. See also [5, 25, 26].

Diffusion processes in one dimension are studied in [65]. There is a complete classification of boundaries and boundary conditions in one dimension, the *Feller classification*: the boundaries can be *regular, exit, entrance* and *natural*. The Feller classification for one dimensional diffusion processes can be found in [47, 23]. We particularly recommend [47, Ch. 15] for a very detailed presentation of diffusion processes in one dimension. Several examples of Fokker-Planck operators in one dimension whose spectrum can be calculated analytically and whose eigenfunctions can be expressed in terms of orthogonal polynomials are presented in [13]. The study of the Fokker-Planck operator in one dimension is closely related to the study of Sturm-Liouville problems. More information on the Sturm-Liouville problem and one dimensional Schrödinger operators (that we obtain after the unitary transformation described in Section 4.9) can be found

in [100, Ch. 9].

Hermite polynomials appear very frequently in applications. We can prove that the Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}^d, \rho_\beta)$ without using the fact that they are the eigenfunctions of a symmetric operator with compact resolvent.[14] The proof of Proposition 4.3 can be found in [97, Lemma 2.3.4].

In Section 4.6 we studied convergence to equilibrium for reversible diffusions using a functional analytic approach and, in particular, the Poincaré inequality for the probability measure $Z^{-1}e^{-V}\,dx$. An alternative approach is the use of a *Lyapunov function* [63, 62, 52]: We will say that the function $U \in C^2(\mathbb{R}^d)$ is a Lyapunov function provided that

i. $U(x) \geqslant 0$ for all $x \in \mathbb{R}^d$;

ii. $\lim_{|x|\to+\infty} U(x) = +\infty$;

iii. there exist positive constants ρ and δ such that $U(x) \leqslant \rho e^{\delta|x|}$ and $|\nabla U(x)| \leqslant \rho e^{\delta|x|}$.

It is possible to show that the existence of a Lyapunov function satisfying

$$\mathcal{L}U(x) \leqslant -\alpha U(x) + \beta, \tag{4.138}$$

where $\alpha,\ \beta$ are positive constants, ensures convergence of the solution to the Fokker-Planck equation $p(x,t)$ to the unique steady state $p_s(x)$ (i.e. the solution of the stationary Fokker-Planck equation) for all initial conditions $p(x,0)$:

$$\lim_{t\to+\infty} p(t,x) = p_s(x), \tag{4.139}$$

the convergence being in $L^1(\mathbb{R}^d)$. The Fokker-Planck equation and the corresponding probability density function are called *g*lobally asymptotically stable [63]. Lyapunov function techniques for stochastic differential equations are studied in detail in [35]. A comparison between functional inequalities-based and Lyapunov function-based techniques for studying convergence to equilibrium for diffusion processes is presented in [7]. A systematic use of Lypunov functions in the study of the ergodic properties of Markov chains is presented in [69].

Dirichlet forms play an important role in the study of diffusion processes, both in finite and in infinite dimensions. Consider an ergodic diffusion process X_t with invariant measure $\mu(dx)$ and generator

$$\mathcal{L} = b(x)\cdot\nabla + \frac{1}{2}\mathbf{\Sigma}(x) : D^2.$$

The *opérateur carré du champ* , defined for example on $C^2(\mathbb{R}^d) \times C^2(\mathbb{R}^d)$, is

$$\Gamma(f,g) = \mathcal{L}(f\,g) - f\mathcal{L}g - g\mathcal{L}f. \tag{4.140}$$

In particular,

$$\Gamma(f,f) = \mathcal{L}f^2 - 2f\mathcal{L}f = \langle \mathbf{\Sigma}(x)\nabla f, \nabla f\rangle.$$

[14] In fact, Poincaré's inequality for Gaussian measures can be proved using the fact that that the Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}^d, \rho_\beta)$.

The Dirichlet form of the diffusion process X_t is then defined as

$$D_{\mathcal{L}}(f) = \int_{\mathbb{R}^d} \Gamma(f,f)\,\mu(dx). \tag{4.141}$$

Further information on Dirichlet forms and the study of diffusion processes can be found at [61].

Poincaré inequalities for probability measures is a vast subject with deep connections with the theory of Schrödinger operators and spectral theory. See [8] and the references therein. The proof of the Poincaré inequality under assumption 4.77, Theorem 4.8, can be found in [103, Thm. A.19]. As we saw in Section 4.5, Poincaré's inequality for the measure $\rho_\beta\,dx = \frac{1}{Z}e^{-\beta V}\,dx$ immediately implies exponentially fast convergence to equilibrium for the corresponding reversible diffusion process in the space $L^2(\mathbb{R}^d;\rho_\beta^{-1})$. However, theorem 4.9 is not very satisfactory since we are assuming that we are already close to equilibrium. Indeed, the assumption on the initial condition

$$\int_{\mathbb{R}^d} |\rho_0(x)|^2 \rho_\beta^{-1} < \infty$$

is very restrictive (think of the case where $V = \frac{1}{2}x^2$). The function space $L^2(\mathbb{R}^d;\rho_\beta^{-1})$ in which we prove convergence is not the right space to use. Since $\rho(\cdot,t) \in L^1$, ideally we would like to prove exponentially fast convergence in $L^1(\mathbb{R}^d)$. We can prove such a result assuming that the Gibbs density ρ_β satisfies a *logarithmic Sobolev inequality* (LSI) [33]. In fact, we can also prove convergence in relative entropy. The relative entropy norm controls the L^1 norm:

$$\|\rho_1 - \rho_2\|_{L^1}^2 \leqslant 2H(\rho_1|\rho_2)$$

This is the *Csiszar-Kullback* or *Pinsker* inequality. 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. We have the following result.

Theorem 4.17. *Let ρ denote the solution of the Fokker-Planck equation* (4.67) *where the potential is smooth and uniformly convex. Assume that the the initial conditions satisfy*

$$H(\rho_0|\rho_\beta) < \infty.$$

Then ρ converges to ρ_β exponentially fast in relative entropy:

$$H(\rho(\cdot,t)|\rho_\beta) \leqslant e^{-\lambda\beta^{-1}t} H(\rho_0|\rho_\beta).$$

Logarithmic Sobolev inequalities are studied in detail in [8]. The approach of using relative entropy and logarithmic Sobolev inequalities to study convergence to equilibrium for the Fokker-Planck equation is presented in [67, 4]. Similar ideas can be used for studying convergence to equilibrium for other types of parabolic PDEs and of kinetic equations, both linear and nonlinear. Further information can be found in [14, 15, 3]. The Bakry-Emery criterion (4.79) guarantees that the measure $e^{-V}\,dx$ satisfies a logarithmic Sobolev inequality with constant λ. This in turn implies that potentials of the form $V + v_0$ where $v_0 \in L^\infty(\mathbb{R}^d)$ also satisfy a logarithmic Sobolev inequality. This is the content of the Holley-Stroock perturbation lemma . See [67] and the references therein.

The connection between self-adjointness of the generator of an ergodic diffusion process in $L^2(\mu)$ and the gradient structure (existence of a potential function) of the drift is established in [71]. The equivalence between self-adjointness, the existence of a potential function, time-reversibility and zero entropy production for a diffusion process is studied in detail in [80, 44]. Conditions on the drift and diffusion coefficients that ensure that detailed balance holds are studied in [85]. Time reversal for diffusion processes with time dependent coefficients is studied in [36]; consider the stochastic equation

$$dX_t = b(X_t, t)\, dt + \sigma(X_t, t)\, dW_t, \tag{4.142}$$

in $\mathbb{R}^d$ and with $t \in (0, 1)$ where the drift and diffusion coefficients satisfy the assumptions of Theorem 3.6 (i.e. a unique strong solution exists) and assume, furthermore, that that probability density $p(t, x)$, the solution of the Fokker-Planck equation corresponding to (4.142) satisfies

$$\int_0^1 \int_{\mathcal{O}} \left[|p(x,t)|^2 + |\sigma(x,t) \cdot \nabla p(x,t)|^2 \right] dxdt < \infty, \tag{4.143}$$

for any open bounded set $\mathcal{O}$. Then the reversed process $\overline{X}_t = X_{1-t},\ t \in [0, 1]$ is a Markov diffusion process satisfying the SDE

$$d\overline{X}_t = \overline{b}(\overline{X}_t, t)\, dt + \overline{\sigma}(\overline{X}_t, t)\, dW_t, \tag{4.144}$$

with

$$\overline{b}(x,t) = -b(x, 1-t) + p(x, 1-t)^{-1} \nabla \cdot (\Sigma(x, 1-t)\, p(1-t, x)), \tag{4.145}$$

where $\Sigma = \sigma\sigma^T$ and

$$\overline{\sigma}(x,t) = \sigma(x, 1-t). \tag{4.146}$$

When the drift and diffusion coefficients in (4.142) are time independent and X_t is stationary with stationary distribution $p_s(x)$ then the formulas for the drift and the diffusion coefficients become

$$\overline{b}(x) = -b(x) + p_s(x)^{-1} \nabla \cdot (\Sigma(x)\, p_s(x)) \tag{4.147}$$

and

$$\overline{\sigma}(x) = \sigma(x). \tag{4.148}$$

Markov Chain Monte Carlo is the standard methodology for sampling from probability distributions in high dimensional spaces [16, 57]. Usually the stochastic dynamics is combined with an accept-reject (Metropolis-Hastings) step. When the Smoluchowski (overdamped Langevin) dynamics is combined with the Metropolis-Hastings step, the resulting algorithm is called the Metropolis adjusted Langevin algorithm (MALA) [89, 88, 87].

In Section 4.8 we saw that there are (infinitely many) different diffusion processes that can be used in order to sample from a given probability distribution $\pi(x)$. Choosing the diffusion process that converges the fastest to equilibrium leads to a computationally efficient algorithm. The Smoluchowski dynamics is not the optimal choice since it can lead to a slow convergence to the target distribution. The drift vector and/or the diffusion matrix have to be modified in order to accelerate convergence. It turns out that the addition of a non-reversible perturbation to the dynamics will in general speed up convergence to equilibrium; see [39, 40]. The optimal nonreversible perturbation can be calculated for diffusions with linear drift that can be used

in order to sample from Gaussian distributions. See [54]. Introducing a time dependent temperature can also accelerate convergence to the target distribution. This is related to the *simulated annealing* algorithm [37]. Another quantity of interest is the asymptotic variance σ_f^2 for an observable f. We can show that

$$\sigma_f^2 := \mathrm{Var}(f) = \langle(-\mathcal{L})^{-1} f, f\rangle_\pi, \tag{4.149}$$

where $\mathcal{L}$ denotes the generator of the dynamics, π the distribution from which we want to sample and $\langle\cdot,\cdot\rangle_\pi$ the inner product in $L^2(\mathbb{R}^d;\pi)$. It is possible to use techniques from the spectral theory of operators to study σ_f^2. See [70] and the references therein. A detailed analysis of algorithms for sampling from the Gibbs distribution $\frac{1}{Z}e^{-\beta V}$ can be found in [55].

Mapping a Fokker-Planck operator to a Schrödinger operator is very useful since Schrödinger operators are one of the most studied topics in mathematical physics, e.g. [81]. For example, the algebraic study of the spectrum of the generator of the Ornstein-Uhlenbeck process using creation and annihilation operators is a standard tool in quantum mechanics. See [100, Ch. 8]. In addition, semigroups generated by Schrödinger operators can be used in order to study properties of the corresponding Markov semigroup; see [93].

Conversely, it is possible to express the solution of the time dependent Schrödinger equation in terms of the solution to an appropriate Fokker-Planck equation. This is the basis for Nelson's *stochastic mechanics*. See [11, 73].

4.11 Exercises

1. Solve equation (4.18) by taking the Fourier transform, using the method of characteristics for first order PDEs and taking the inverse Fourier transform.

2. Use (4.26) to obtain formulas for the moments of the Ornstein-Uhlenbeck process. Prove, using these formulas, that the moments of the Ornstein-Uhlenbeck process converge to their equilibrium values exponentially fast.

3. Show that the autocorrelation function of the stationary Ornstein-Uhlenbeck is

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

   where $p_{OU}(x,t|x_0,0)$ denotes the transition probability function and $p_s(x)$ the invariant Gaussian distribution.

4. Let X_t be a one-dimensional diffusion process with drift and diffusion coefficients $a(y,t) = -a_0 - a_1 y$ and $b(y,t) = b_0 + b_1 y + b_2 y^2$ where $a_i,\ b_i \geqslant 0,\ i = 0,1,2$.

   (a) Write down the generator and the forward and backward Kolmogorov equations for X_t.

   (b) Assume that X_0 is a random variable with probability density $\rho_0(x)$ that has finite moments. Use the forward Kolmogorov equation to derive a system of differential equations for the moments of X_t.

(c) Find the first three moments $M_0,\ M_1,\ M_2$ in terms of the moments of the initial distribution $\rho_0(x)$.

(d) Under what conditions on the coefficients $a_i,\ b_i \geqslant 0,\ i = 0, 1, 2$ is M_2 finite for all times?

5. Consider a uniformly elliptic diffusion process in $\Omega \subset \mathbb{R}^d$ with reflecting boundary conditions and generator

$$\mathcal{L} = b(x) \cdot \nabla + \frac{1}{2}\Sigma : D^2. \tag{4.150}$$

Let $p(x, t)$ denote the probability density function, i.e. the solution of the Fokker-Planck equation, and $p_s(x)$ the stationary distribution. Show that the relative entropy

$$H(t) = \int_\Omega p(x, t) \ln\left(\frac{p(x, t)}{p_s(x)}\right) dx$$

is nonincreasing:

$$\frac{dH}{dt} \leqslant 0.$$

6. Let V be a confining potential in $\mathbb{R}^d$, $\beta > 0$ and let $\rho_\beta(x) = Z^{-1}e^{-\beta V(x)}$. Give the definition of the Sobolev space $H^k(\mathbb{R}^d; \rho_\beta)$ for k a positive integer and study some of its basic properties.

7. Let X_t be a multidimensional diffusion process on $[0, 1]^d$ with periodic boundary conditions. The drift vector is a periodic function $a(x)$ and the diffusion matrix is $2DI$, where $D > 0$ and I is the identity matrix.

(a) Write down the generator and the forward and backward Kolmogorov equations for X_t.

(b) Assume that $a(x)$ is divergence-free ($\nabla \cdot a(x) = 0$). Show that X_t is ergodic and find the invariant distribution.

(c) Show that the probability density $p(x, t)$ (the solution of the forward Kolmogorov equation) converges to the invariant distribution exponentially fast in $L^2([0, 1]^d)$.

8. The Rayleigh process X_t is a diffusion process that takes values on $(0, +\infty)$ with drift and diffusion coefficients $a(x) = -ax + \frac{D}{x}$ and $b(x) = 2D$, respectively, where $a,\ D > 0$.

(a) Write down the generator the forward and backward Kolmogorov equations for X_t.

(b) Show that this process is ergodic and find its invariant distribution.

(c) Solve the forward Kolmogorov (Fokker-Planck) equation using separation of variables. (*Hint:* Use Laguerre polynomials).

9. Let $\mathbf{x}(t) = \{x(t),\ y(t)\}$ be the two-dimensional diffusion process on $[0,\ 2\pi]^2$ with periodic boundary conditions with drift vector $\mathbf{a}(x, y) = (\sin(y),\ \sin(x))$ and diffusion matrix $\mathbf{b}(x, y)$ with $b_{11} = b_{22} = 1,\ b_{12} = b_{21} = 0$.

(a) Write down the generator of the process $\{x(t),\ y(t)\}$ and the forward and backward Kolmogorov equations.

(b) Show that the constant function

$$\rho_s(x, y) = C$$

is the unique stationary distribution of the process $\{x(t),\, y(t)\}$ and calculate the normalization constant.

(c) Let $\mathbb{E}$ denote the expectation with respect to the invariant distribution $\rho_s(x, y)$. Calculate

$$\mathbb{E}\left(\cos(x) + \cos(y)\right) \quad \text{and} \quad \mathbb{E}(\sin(x)\sin(y)).$$

10. Let a, D be positive constants and let $X(t)$ be the diffusion process on $[0, 1]$ with periodic boundary conditions and with drift and diffusion coefficients $a(x) = a$ and $b(x) = 2D$, respectively. Assume that the process starts at x_0, $X(0) = x_0$.

(a) Write down the generator of the process $X(t)$ and the forward and backward Kolmogorov equations.

(b) Solve the initial/boundary value problem for the forward Kolmogorov equation to calculate the transition probability density $p(x, t|x_0, 0)$.

(c) Show that the process is ergodic and calculate the invariant distribution $p_s(x)$.

(d) Calculate the stationary autocorrelation function

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

11. Prove formulas (4.109), (4.112) and (4.113).

12. Let X_t be a reverisble diffusion process. Use the spectral analysis from Section 4.7 to obtain a spectral representation for an autocorrelation function of the form

$$\mathbb{E}(f(X_t)h(X_0)), \tag{4.151}$$

where f and h are arbitrary observables.

Appendices

Appendix A

Frequently Used Notation

In the following all summations are over indices from the set $\{1, 2, \ldots, d\}$, d being the dimension of the space. We use $\mathbb{R}^d$ to denote the $d-$dimensional Euclidean space. We denote by $\langle \cdot, \cdot \rangle$ the standard inner product on $\mathbb{R}^d$. We also use $\cdot$ to denote the inner product between two vectors, so that

$$\langle a, b \rangle = a \cdot b = \sum_i a_i b_i,$$

where $\{\xi_i\}_{i=1}^d$ are the components of a vector $\xi \in \mathbb{R}^d$ with respect to the standard basis $\{e_i\}_{i=1}^d$. The norm induced by this inner product is the Euclidean norm

$$|a| = \sqrt{a \cdot a}$$

and it follows that

$$|a|^2 = \sum_i a_i^2, \quad a \in \mathbb{R}^d.$$

The *inner product between matrices* is denoted by

$$A : B = \text{tr}(A^T B) = \sum_{ij} a_{ij} b_{ij}.$$

The norm induced by this inner product is the *Frobenius norm*

$$|A|_F = \sqrt{\text{tr}(A^T A)}. \tag{A.1}$$

Let ∇ and $\nabla \cdot$ denote *gradient* and *divergence* in $\mathbb{R}^d$. The gradient lifts a scalar (resp. vector) to a vector (resp. matrix) whilst the divergence contracts a vector (resp. matrix) to a scalar (resp. vector). The gradient acts on scalar valued functions $\phi(z)$, or vector valued functions $v(z)$, via

$$(\nabla \phi)_i = \frac{\partial \phi}{\partial x_i}, \qquad (\nabla v)_{ij} = \frac{\partial v_i}{\partial x_j}.$$

The divergence of a vector valued function $v(z)$ is

$$\nabla \cdot v = \text{Tr}(\nabla v) = \sum_i \frac{\partial v_i}{\partial z_i}.$$

The divergence and gradient operators do not commute:

$$\nabla\big(\nabla \cdot v\big) = \nabla \cdot \big((\nabla v)^T\big).$$

The divergence of a matrix valued function $A(x)$ is the unique vector field defined as

$$\nabla \cdot A(x) \cdot a = \nabla \cdot \left(A^T(x)a\right),$$

for all constant vectors $a \in \mathbb{R}^d$. Componentwise,

$$(\nabla \cdot A)_i = \sum_j \frac{\partial A_{ij}}{\partial z_j}, \quad i = 1, \dots d.$$

Given a vector valued function $v(x)$ and a matrix valued function $A(x)$ we have the following product rule

$$\nabla \cdot \big(A^T v\big) = \big(\nabla \cdot A\big) \cdot v + A : \nabla v.$$

For two matrix valued functions $A(x),\ B(x)$ we have the following product rule

$$\nabla \cdot \big(AB\big) \cdot a = (\nabla \cdot B) \cdot Aa + B : \nabla(Aa). \tag{A.2}$$

Given vector fields a, v we use the notation

$$a \cdot \nabla v := (\nabla v)a.$$

Thus we define the quantity by calculating $a \cdot \nabla v_k$ for each component of the vector v. Likewise we can extend to the notation

$$a \cdot \nabla \Theta,$$

where Θ is a matrix field, by using the above definition componentwise.

Since the gradient is defined for scalars and vectors we readily make sense of the expression

$$\nabla\nabla\phi$$

for any scalar ϕ; it is the *Hessian* matrix $D^2 f$ with entries $\frac{\partial^2 \phi}{\partial x_i \partial x_j}$. Similarly, we can also make sense of the expression

$$\nabla\nabla v$$

by applying $\nabla\nabla$ to each scalar component of the vector v, or indeed

$$\nabla\nabla\Theta,$$

again componentwise. We define the Laplacian of a scalar or vector field by

$$\Delta\phi = \nabla \cdot \nabla\phi; \quad \Delta v = \nabla \cdot \nabla v.$$

It follows that $\Delta\phi = I : \nabla\nabla\phi$. Applying this definition componentwise allows for the definition of $\Delta\Theta$. We also use the following notations

$$A : \nabla\nabla f = A : D^2 f = \mathrm{Tr}(AD^2)f = \sum_{i,j} A_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Appendix B

Elements of Probability Theory

In this appendix we put together some basic definitions and results from probability theory that we used. This is very standard material and can be found in all textbooks on probability theory and stochastic processes. In Section B.1 we give some basic definitions from the theory of probability. In Section B.2 we present some properties of random variables. In Section B.3 we introduce the concept of conditional expectation and in Section B.4 we define the characteristic function. A few calculations with Gaussian measures in finite dimensions and in separable Hilbert spaces are presented in Section B.5. Different types of convergence and the basic limit theorems of the theory of probability are discussed in Section B.6. Discussion and bibliographical comments are presented in Section B.7.

B.1 Basic Definitions from Probability Theory

In order to study stochastic processes we need to be able to describe the outcome of a random experiment and to calculate functions of this outcome. First we need to describe the set of all possible experiments.

Definition B.1. *The set of all possible outcomes of an experiment is called the* sample space *and is denoted by* Ω.

We define events to be subsets of the sample space. Of course, we would like the unions, intersections and complements of events to also be events. When the sample space Ω is uncountable, then technical difficulties arise. In particular, not all subsets of the sample space need to be events. A definition of the collection of subsets of events which is appropriate for finite additive probability is the following.

Definition B.2. *A collection* $\mathcal{F}$ *of* Ω *is called a field on* Ω *if*

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

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

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

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

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

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

Definition B.3 (σ-algebra)**.** *A collection $\mathcal{F}$ of Ω is called a σ-field or σ-algebra on Ω if*

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

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

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

A σ-algebra is closed under the operation of taking countable intersections. Standard examples of a σ-algebra are $\mathcal{F} = \{\emptyset,\ \Omega\}$, $\mathcal{F} = \{\emptyset,\ A,\ A^c,\ \Omega\}$ where A is a subset of Ω and the power set of Ω, denoted by $\{0,1\}^{\Omega}$ which contains all subsets of Ω.

Let now $\mathcal{F}$ be a collection of subsets of Ω. It can be extended to a σ-algebra (take for example the power set of Ω). Consider all the σ-algebras that contain $\mathcal{F}$ and take their intersection, denoted by $\sigma(\mathcal{F})$, i.e. $A \subset \Omega$ if and only if it is in every σ-algebra containing $\mathcal{F}$. It is a standard exercise to show that $\sigma(\mathcal{F})$ is a σ-algebra. It is the smallest algebra containing $\mathcal{F}$ and it is called the σ-algebra generated by $\mathcal{F}$.

Example B.4. *Let $\Omega = \mathbb{R}^n$. The σ-algebra generated by the open subsets of $\mathbb{R}^n$ (or, equivalently, by the open balls of $\mathbb{R}^n$) is called the Borel σ-algebra of $\mathbb{R}^n$ and is denoted by $\mathcal{B}(\mathbb{R}^n)$.*

Let X be a closed subset of $\mathbb{R}^n$. Similarly, we can define the Borel σ-algebra of X, denoted by $\mathcal{B}(X)$. A sub-σ-algebra is a collection of subsets of a σ-algebra which satisfies the axioms of a σ-algebra. The $\sigma-$field $\mathcal{F}$ of a sample space Ω contains all possible outcomes of the experiment that we want to study. Intuitively, the σ-field contains all the information that is available to use about the random experiment that we are performing.

Now we want to assign probabilities to the possible outcomes of an experiment.

Definition B.5 (Probability measure)**.** *A probability measure $\mathbb{P}$ on the measurable space $(\Omega,\ \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \mapsto [0,1]$ satisfying*

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

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

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

Definition B.6. *The triple $\big(\Omega,\ \mathcal{F},\ \mathbb{P}\big)$ comprising a set Ω, a σ-algebra $\mathcal{F}$ of subsets of Ω and a probability measure $\mathbb{P}$ on $(\Omega,\ \mathcal{F})$ is a called a probability space.*

A standard example is that of $\Omega = [0,1],\ \mathcal{F} = \mathcal{B}([0,1]),\ \mathbb{P} = \text{Leb}([0,1])$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

B.2 Random Variables

We are usually interested in the consequences of the outcome of an experiment, rather than the experiment itself. The function of the outcome of an experiment is a *random variable*, that is, a map from Ω to $\mathbb{R}$.

Definition B.7. *A sample space Ω equipped with a $\sigma-$field of subsets $\mathcal{F}$ is called a measurable space.*

Definition B.8. *Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces. A function $X : \Omega \to E$ such that the event*

$$\{\omega \in \Omega : X(\omega) \in A\} =: \{X \in A\} \tag{B.1}$$

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

When E is $\mathbb{R}$ equipped with its Borel σ-algebra, then (B.1) can by replaced with

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

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

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

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

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

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

$$\mu_X(B) = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(E). \tag{B.2}$$

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

Example B.9. *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: $\sum_{i\in\mathcal{I}} \rho_{0,i} = 1$.*

Consider the case where $E = \mathbb{R}$ equipped with the Borel $\sigma-$algebra. In this case a random variable is defined to be a function $X : \Omega \to \mathbb{R}$ such that

$$\{\omega \in \Omega : X(\omega) \leqslant x\} \subset \mathcal{F} \quad \forall x \in \mathbb{R}.$$

We can now define the probability distribution function of X, $F_X : \mathbb{R} \to [0, 1]$ as

$$F_X(x) = \mathbb{P}\big(\{\omega \in \Omega \big| X(\omega) \leqslant x\}\big) =: \mathbb{P}(X \leqslant x). \tag{B.3}$$

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

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

Definition B.10. *A random variable X with values on $\mathbb{R}$ is called discrete if it takes values in some countable subset $\{x_0, x_1, x_2, \dots\}$ of $\mathbb{R}$. i.e.: $\mathbb{P}(X = x) \neq x$ only for $x = x_0, x_1, \dots$.*

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

Example B.11. *The Poisson random variable is the nonnegative integer valued random variable with probability mass function*

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

where $\lambda > 0$.

Example B.12. *The binomial random variable is the nonnegative integer valued random variable with probability mass function*

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

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

Definition B.13. *A random variable X with values on $\mathbb{R}$ is called continuous if $\mathbb{P}(X = x) = 0\, \forall x \in \mathbb{R}$.*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable with distribution F_X. This is a probability measure on $\mathcal{B}(\mathbb{R})$. We will assume that it is absolutely continuous with respect to the Lebesgue measure with density ρ_X: $F_X(dx) = \rho(x)\, dx$. We will call the density $\rho(x)$ the probability density function (PDF) of the random variable X.

Example B.14. *i. The exponential random variable has PDF*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x < 0, \end{cases}$$

with $\lambda > 0$.

ii. The uniform random variable has PDF

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & x \notin (a, b), \end{cases}$$

with $a < b$.

Definition B.15. *Two random variables X and Y are independent if the events $\{\omega \in \Omega \,|\, X(\omega) \leqslant x\}$ and $\{\omega \in \Omega \,|\, Y(\omega) \leqslant y\}$ are independent for all $x, y \in \mathbb{R}$.*

Let X, Y be two continuous random variables. We can view them as a random vector, i.e. a random variable from Ω to $\mathbb{R}^2$. We can then define the joint distribution function

$$F(x, y) = \mathbb{P}(X \leqslant x,\, Y \leqslant y).$$

The mixed derivative of the distribution function $f_{X,Y}(x,y) := \frac{\partial^2 F}{\partial x \partial y}(x,y)$, if it exists, is called the joint PDF of the random vector $\{X,\, Y\}$:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)\,dxdy.$$

If the random variables X and Y are independent, then

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

and

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

The joint distribution function has the properties

$$\begin{aligned} F_{X,Y}(x,y) &= F_{Y,X}(y,x), \\ F_{X,Y}(+\infty,y) &= F_Y(y), \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx. \end{aligned}$$

We can extend the above definition to random vectors of arbitrary finite dimensions. Let X be a random variable from $(\Omega,\mathcal{F},\mu)$ to $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$. The (joint) distribution function $F_X\mathbb{R}^d \to [0,1]$ is defined as

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

Let X be a random variable in $\mathbb{R}^d$ with distribution function $f(x_N)$ where $x_N = \{x_1,\ldots x_N\}$. We define the marginal or reduced distribution function $f^{N-1}(x_{N-1})$ by

$$f^{N-1}(x_{N-1}) = \int_{\mathbb{R}} f^N(x_N)\,dx_N.$$

We can define other reduced distribution functions:

$$f^{N-2}(x_{N-2}) = \int_{\mathbb{R}} f^{N-1}(x_{N-1})\,dx_{N-1} = \int_{\mathbb{R}}\int_{\mathbb{R}} f(x_N)\,dx_{N-1}dx_N.$$

Expectation of Random Variables

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

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\,dF_X(x) \tag{B.4}$$

and

$$\mathbb{P}[X \in G] = \int_G dF_X(x), \quad G \in \mathcal{B}(E). \tag{B.5}$$

The above formulas apply to both discrete and continuous random variables, provided that we define the integrals in (B.4) and (B.5) appropriately.

When $E = \mathbb{R}^d$ and a PDF exists, $dF_X(x) = f_X(x)\,dx$, we have

$$F_X(x) := \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_d} f_X(x)\,dx..$$

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

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

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

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

Let $X,\ Y$ be random variables we want to know whether they are correlated and, if they are, to calculate how correlated they are. We define the covariance of the two random variables as

$$\text{cov}(X,Y) = \mathbb{E}\big[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\big] = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y.$$

The correlation coefficient is

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(X)}} \tag{B.6}$$

The Cauchy-Schwarz inequality yields that $\rho(X,Y) \in [-1,1]$. We will say that two random variables X and Y are uncorrelated provided that $\rho(X,Y) = 0$. It is not true in general that two uncorrelated random variables are independent. This is true, however, for Gaussian random variables.

Example B.16.

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

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

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

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

and the variance is

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

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

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

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

$$\mathbb{E}(X) = b \tag{B.7}$$

and the covariance matrix is

$$\mathbb{E}\Big((X-b)\otimes(X-b)\Big) = \Sigma. \tag{B.8}$$

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 $\mathcal{N}(0,1)$. Similarly, since the mean and covariance matrix completely specify a Gaussian random variable on $\mathbb{R}^d$, the Gaussian is commonly denoted by $\mathcal{N}(m,\Sigma)$.

Some analytical calculations for Gaussian random variables will be presented in Section B.5.

B.3 Conditional Expecation

One of the most important concepts in probability is that of the dependence between events.

Definition B.17. *A family $\{A_i : i \in I\}$ of events is called independent if*

$$\mathbb{P}\big(\cap_{j\in J} A_j\big) = \Pi_{j\in J}\mathbb{P}(A_j)$$

for all finite subsets J of I.

When two events $A,\ B$ are dependent it is important to know the probability that the event A will occur, given that B has already happened. We define this to be *conditional probability*, denoted by $\mathbb{P}(A|B)$. We know from elementary probability that

$$P(A|B) = \frac{P(A\cap B)}{\mathbb{P}(B)}.$$

A very useful result is that of the *law of total probability*.

Definition B.18. *A family of events $\{B_i : i \in I\}$ is called a partition of Ω if*

$$B_i \cap B_j = \emptyset,\ \ i \neq j \quad \textit{and} \quad \cup_{i\in I} B_i = \Omega.$$

Proposition B.19. *Law of total probability. For any event A and any partition $\{B_i : i \in I\}$ we have*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

The proof of this result is left as an exercise. In many cases the calculation of the probability of an event is simplified by choosing an appropriate partition of Ω and using the law of total probability.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and fix $B \in \mathcal{F}$. Then $\mathbb{P}(\cdot|B)$ defines a probability measure on $\mathcal{F}$. Indeed, we have that

$$\mathbb{P}(\emptyset|B) = 0, \quad \mathbb{P}(\Omega|B) = 1$$

and (since $A_i \cap A_j = \emptyset$ implies that $(A_i \cap B) \cap (A_j \cap B) = \emptyset$)

$$P(\cup_{j=1}^{\infty} A_i|B) = \sum_{j=1}^{\infty} \mathbb{P}(A_i|B),$$

for a countable family of pairwise disjoint sets $\{A_j\}_{j=1}^{+\infty}$. Consequently, $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|B))$ is a probability space for every $B \in \mathcal{F}$.

Assume that $X \in L^1(\Omega, \mathcal{F}, \mu)$ 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 (random variable) $\mathbb{E}[X|\mathcal{G}] : \Omega \mapsto E$ 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.

We list some of the most important properties of conditional expectation.

Proposition B.20. *[Properties of Conditional Expectation]. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{F}$.*

(a) If X is $\mathcal{G}-$measurable and integrable then $\mathbb{E}(X|\mathcal{G}) = X$.

(b) (Linearity) If $X_1,\ X_2$ are integrable and $c_1,\ c_2$ constants, then

$$\mathbb{E}(c_1X_1 + c_2X_2|\mathcal{G}) = c_1\mathbb{E}(X_1|\mathcal{G}) + c_2\mathbb{E}(X_2|\mathcal{G}).$$

(c) (Order) If $X_1,\ X_2$ are integrable and $X_1 \leqslant X_2$ a.s., then $\mathbb{E}(X_1|\mathcal{G}) \leqslant \mathbb{E}(X_2|\mathcal{G})$ a.s.

(d) If Y and XY are integrable, and X is $\mathcal{G}-$measurable then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.

(e) (Successive smoothing) If $\mathcal{D}$ is a sub–σ–algebra of $\mathcal{F}$, $\mathcal{D} \subset \mathcal{G}$ and X is integrable, then $\mathbb{E}(X|\mathcal{D}) = \mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{D}] = \mathbb{E}[\mathbb{E}(X|\mathcal{D})|\mathcal{G}]$.

(f) (Convergence) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables such that, for all n, $|X_n| \leqslant Z$ where Z is integrable. If $X_n \to X$ a.s., then $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ a.s. and in L^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 (see Theorem B.20)

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1). \tag{B.9}$$

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

$$\mathbb{E}(f(X)|\mathcal{G}) = \int_{\mathbb{R}} f(x)P_X(dx|\mathcal{G}). \tag{B.10}$$

B.4 The Characteristic Function

Many of the properties of (sums of) random variables can be studied using the Fourier transform of the distribution function. Let $F(\lambda)$ be the distribution function of a (discrete or continuous) random variable X. The characteristic function of X is defined to be the Fourier transform of the distribution function

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda}\, dF(\lambda) = \mathbb{E}(e^{itX}). \tag{B.11}$$

For a continuous random variable for which the distribution function F has a density, $dF(\lambda) = p(\lambda)d\lambda$, (B.11) gives

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda}p(\lambda)\, d\lambda.$$

For a discrete random variable for which $\mathbb{P}(X = \lambda_k) = \alpha_k$, (B.11) gives

$$\phi(t) = \sum_{k=0}^{\infty} e^{it\lambda_k}a_k.$$

From the properties of the Fourier transform we conclude that 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 $\phi(t)$. Furthermore, in the exercises at the end of the chapter the reader is asked to prove the following two results.

Lemma B.21. *Let $\{X_1, X_2, \dots X_n\}$ be independent random variables with characteristic functions $\phi_j(t)$, $j = 1, \dots n$ 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 B.22. *Let X be a random variable with characteristic function $\phi(t)$ and assume that it has finite moments. Then*

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

B.5 Gaussian Random Variables

In this section we present some useful calculations for Gaussian random variables. In particular, we calculate the normalization constant, the mean and variance and the characteristic function of multidimensional Gaussian random variables.

Theorem B.23. *Let $\mathbf{b} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ a symmetric and positive definite matrix. Let $\mathbf{X}$ be the multivariate Gaussian random variable with probability density function*

$$\gamma_{\Sigma,b}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{1}{2}\langle\Sigma^{-1}(\mathbf{x}-\mathbf{b}), \mathbf{x}-\mathbf{b}\rangle\right).$$

Then

i. The normalization constant is

$$Z = (2\pi)^{d/2}\sqrt{det(\Sigma)}.$$

ii. The mean vector and covariance matrix of $\mathbf{X}$ are given by

$$\mathbb{E}\mathbf{X} = \mathbf{b}$$

and

$$\mathbb{E}((\mathbf{X} - \mathbb{E}\mathbf{X}) \otimes (\mathbf{X} - \mathbb{E}\mathbf{X})) = \Sigma.$$

iii. The characteristic function of $\mathbf{X}$ is

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

Proof. i. From the spectral theorem for symmetric positive definite matrices we have that there exists a diagonal matrix Λ with positive entries and an orthogonal matrix B such that

$$\Sigma^{-1} = B^T\Lambda^{-1}B.$$

Let $\mathbf{z} = \mathbf{x} - \mathbf{b}$ and $\mathbf{y} = B\mathbf{z}$. We have

$$\begin{aligned}\langle\Sigma^{-1}\mathbf{z}, \mathbf{z}\rangle &= \langle B^T\Lambda^{-1}B\mathbf{z}, \mathbf{z}\rangle \\ &= \langle\Lambda^{-1}B\mathbf{z}, B\mathbf{z}\rangle = \langle\Lambda^{-1}\mathbf{y}, \mathbf{y}\rangle \\ &= \sum_{i=1}^d \lambda_i^{-1}y_i^2.\end{aligned}$$

Furthermore, we have that $\det(\Sigma^{-1}) = \Pi_{i=1}^d \lambda_i^{-1}$, that $\det(\Sigma) = \Pi_{i=1}^d \lambda_i$ and that the Jacobian of an orthogonal transformation is $J = \det(B) = 1$. Hence,

$$\begin{aligned}\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle\Sigma^{-1}(\mathbf{x}-\mathbf{b}),\mathbf{x}-\mathbf{b}\rangle\right)\,d\mathbf{x} &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle\Sigma^{-1}\mathbf{z},\mathbf{z}\rangle\right)\,d\mathbf{z}\\ &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\sum_{i=1}^d \lambda_i^{-1}y_i^2\right)|J|\,d\mathbf{y}\\ &= \prod_{i=1}^d \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\lambda_i^{-1}y_i^2\right)\,dy_i\\ &= (2\pi)^{d/2}\Pi_{i=1}^n\lambda_i^{1/2} = (2\pi)^{d/2}\sqrt{\det(\Sigma)},\end{aligned}$$

from which we get that

$$Z = (2\pi)^{d/2}\sqrt{\det(\Sigma)}.$$

In the above calculation we have used the elementary calculus identity

$$\int_{\mathbb{R}} e^{-\alpha\frac{x^2}{2}}\,dx = \sqrt{\frac{2\pi}{\alpha}}.$$

ii. From the above calculation we have that

$$\begin{aligned}\gamma_{\Sigma,b}(\mathbf{x})\,d\mathbf{x} &= \gamma_{\Sigma,b}(B^T\mathbf{y}+\mathbf{b})\,d\mathbf{y}\\ &= \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}}\prod_{i=1}^d \exp\left(-\frac{1}{2}\lambda_i y_i^2\right)\,dy_i.\end{aligned}$$

Consequently

$$\begin{aligned}\mathbb{E}\mathbf{X} &= \int_{\mathbb{R}^d}\mathbf{x}\gamma_{\Sigma,b}(\mathbf{x})\,d\mathbf{x}\\ &= \int_{\mathbb{R}^d}(B^T\mathbf{y}+\mathbf{b})\gamma_{\Sigma,b}(B^T\mathbf{y}+\mathbf{b})\,d\mathbf{y}\\ &= \mathbf{b}\int_{\mathbb{R}^d}\gamma_{\Sigma,b}(B^T\mathbf{y}+\mathbf{b})\,d\mathbf{y} = \mathbf{b}.\end{aligned}$$

We note that, since $\Sigma^{-1} = B^T\Lambda^{-1}B$, we have that $\Sigma = B^T\Lambda B$. Furthermore, $\mathbf{z} = B^T\mathbf{y}$. We calculate

$$\begin{aligned}&\mathbb{E}((X_i-b_i)(X_j-b_j)) = \int_{\mathbb{R}^d} z_i z_j \gamma_{\Sigma,b}(\mathbf{z}+\mathbf{b})\,d\mathbf{z}\\ =\ & \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}}\int_{\mathbb{R}^d}\sum_k B_{ki}y_k\sum_m B_{mi}y_m\exp\left(-\frac{1}{2}\sum_\ell\lambda_\ell^{-1}y_\ell^2\right)\,d\mathbf{y}\\ =\ & \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}}\sum_{k,m}B_{ki}B_{mj}\int_{\mathbb{R}^d} y_k y_m\exp\left(-\frac{1}{2}\sum_\ell\lambda_\ell^{-1}y_\ell^2\right)\,d\mathbf{y}\\ =\ & \sum_{k,m}B_{ki}B_{mj}\lambda_k\delta_{km}\\ =\ & \Sigma_{ij}.\end{aligned}$$

iii. Let $\mathbf{y}$ be a multivariate Gaussian random variable with mean $\mathbf{0}$ and covariance I. Let also $C = B\sqrt{\Lambda}$. We have that $\Sigma = CC^T = C^TC$. We have that

$$\mathbf{X} = C\mathbf{Y} + \mathbf{b}.$$

To see this, we first note that $\mathbf{X}$ is Gaussian since it is given through a linear transformation of a Gaussian random variable. Furthermore,

$$\mathbb{E}\mathbf{X} = \mathbf{b} \quad \text{and} \quad \mathbb{E}((X_i - b_i)(X_j - b_j)) = \Sigma_{ij}.$$

Now we have:

$$\begin{aligned}
\phi(\mathbf{t}) &= \mathbb{E}e^{i\langle \mathbf{X},\mathbf{t}\rangle} = e^{i\langle \mathbf{b},\mathbf{t}\rangle}\mathbb{E}e^{i\langle C\mathbf{Y},\mathbf{t}\rangle} \\
&= e^{i\langle \mathbf{b},\mathbf{t}\rangle}\mathbb{E}e^{i\langle \mathbf{Y},C^T\mathbf{t}\rangle} = e^{i\langle \mathbf{b},\mathbf{t}\rangle}\mathbb{E}e^{i\sum_j(\sum_k C_{jk}t_k)y_j} \\
&= e^{i\langle \mathbf{b},\mathbf{t}\rangle}e^{-\frac{1}{2}\sum_j\left|\sum_k C_{jk}t_k\right|^2} = e^{i\langle \mathbf{b},\mathbf{t}\rangle}e^{-\frac{1}{2}\langle C\mathbf{t},C\mathbf{t}\rangle} \\
&= e^{i\langle \mathbf{b},\mathbf{t}\rangle}e^{-\frac{1}{2}\langle \mathbf{t},C^TC\mathbf{t}\rangle} = e^{i\langle \mathbf{b},\mathbf{t}\rangle}e^{-\frac{1}{2}\langle \mathbf{t},\Sigma\mathbf{t}\rangle}.
\end{aligned}$$

Consequently,

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

□

B.5.1 Gaussian Measures in Hilbert Spaces

In the following we let H to be a separable Hilbert space and we let $\mathcal{B}(H)$ to be the Borel σ-algebra on H. We start with the definition of a Gaussian measure.

Definition B.24. *A probability measure μ on $(H,\ \mathcal{B}(H))$ is called Gaussian if for all $h \in H$ there exists an $m \in \mathbb{R}$ such that*

$$\mu\left(x \in H; \langle h, x\rangle \in A\right) = \mathcal{N}(A),\ A \in \mathcal{B}(\mathbb{R}). \tag{B.12}$$

Let μ be a Gaussian measure. We define the following continuous functionals:

$$H \to \mathbb{R}\ \ h \to \int_H \langle h, x\rangle \mu(dx) \tag{B.13a}$$

$$H \times H \to \mathbb{R}\ \ (h_1, h_2) \to \int_H \langle h_1, x\rangle\langle h_2, x\rangle \mu(dx) \tag{B.13b}$$

The functional in (B.13b) is symmetric. We can use the Riesz representation theorem on H and $H \times H$ to conclude:

Theorem B.25. *There exists an $m \in H$, and a symmetric, nonnegative continuous operator Q such that*

$$\int_H \langle h, x\rangle \mu(dx) = \langle m, h\rangle\ \ \forall h \in H$$

and

$$\int_H \langle h_1, x\rangle\langle h_2, x\rangle \mu(dx) - \langle m, h_1\rangle\langle m, h_2\rangle = \langle Qh_1, h_2\rangle\ \ \forall h_1,\ h_2 \in H.$$

We will call m the mean and Q the covariance operator of the measure μ.

A Gaussian measure μ on H with mean m and covariance Q has the following characteristic function:

$$\mu(\lambda) = \int e^{i\langle \lambda \cdot x\rangle} \mu(dx) = e^{i\langle \lambda \cdot m\rangle - \frac{1}{2}\langle Q\lambda, \lambda\rangle}. \tag{B.14}$$

Consequently, a Gaussian measure is uniquely determined by m and Q. Using the characteristic function of μ one can prove that Q is a trace class operator.

Let now $\{e_k\}$ and $\{\lambda_k\}$ be the eigenfunctions and eigenvalues of Q, respectively. Since Q is symmetric and bounded, $\{e_k\}$ forms a complete orthonormal basis on H. Further, let $x_k = \langle x, e_k\rangle$, $k \in \mathbb{N}$. In the sequel we will set $m = 0$.

Lemma B.26. *The random variables $(x_1, \ldots, x_n)$ are independent.*

Proof. We compute:

$$\begin{aligned} \int_H x_i x_i \, \mu(dx) &= \int_H \langle x, e_i\rangle \langle x, e_j\rangle \, \mu(dx) \\ &= \langle Qe_i, e_j\rangle \\ &= \lambda_i \delta_{ij}. \end{aligned} \tag{B.15}$$

□

Now we have the following:

Proposition B.27. *Let $\mu \in \mathcal{N}(0, Q)$ on H. Let*

$$S_Q = \inf_{\lambda \in \sigma(Q)} \frac{1}{2\lambda} = \frac{1}{2\|Q\|} \tag{B.16}$$

where $\sigma(Q)$ is the spectrum of Q. Then, $\forall\, s \in [0, S_Q)$ we have:

$$\begin{aligned} \int_H e^{s|x|^2} \mu(dx) &= \exp\left[-\frac{1}{2} Tr\left(\log(I - 2sQ)\right)\right] \\ &= \exp\left[\frac{1}{2}\sum_{k=1}^{\infty} \frac{(2s)^k}{k} Tr(Q^k)\right] \end{aligned} \tag{B.17}$$

Proof. 1. First we observe that since Q is a bounded operator we have $S_Q > 0$. Now, for $s \in [0, S_Q)$ we have:

$$\log(I - 2sQ) = \sum_{k=1}^{\infty} \frac{(2s)^k}{k}, \tag{B.18}$$

the series being absolutely convergent in $L(H)$. Consequently, the operator $\log(I - 2sQ)$ is also trace class.

2. We fix $s \in [0, S_Q)$ and we consider finite dimensional truncations of the integral that appears on the left hand side of equation (B.17):

$$\begin{aligned} I_n &= \int_H e^{s\sum_{i=1}^n x_i^2}\mu(dx) \\ &= \prod_{i=1}^n \int_H e^{sx_i^2}\mu(dx) \quad (\{x_i\} \text{ are independent}) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\lambda_i}} \int_{-\infty}^{\infty} e^{\left(s\xi^2 - \frac{x_i^2}{2\lambda_i}\right)}\mu(dx) \quad (x_i \in \mathcal{N}(0,\lambda_i)) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{1-2\lambda_i s}} = e^{\left(-\frac{1}{2}\sum_{i=1}^n \log(1-2\lambda_i s)\right)} \\ &= e^{\left(-\frac{1}{2}\mathrm{Tr}\log(I-2sQ_n)\right)} \end{aligned} \tag{B.19}$$

with

$$Q_n x = \sum_{i=1}^n \lambda_i \langle x, e_i\rangle e_i, \ x \in H. \tag{B.20}$$

Now we let $n \to \infty$ and use the fact that $\log(I - 2sQ_n)$ is trace class to obtain (B.17). □

From the above proposition we immediately obtain the following corollary:

Corollary B.28. *For arbitrary $p \in \mathbb{N}$ there exists a constant C_p such that*

$$\int_H |x|^{2p}\,\mu(dx) \leqslant C_p\,[Tr(Q)]^p \tag{B.21}$$

for arbitrary $\mu \in \mathcal{N}(0, Q)$.

Proof. Differentiate equation (B.17) p times and set $s = 0$. □Now we make a few remarks on the above proposition and corollary. First, C_p is a combinatorial constant and grows in p. Moreover, we have

$$\int_H |x|^2\,\mu(dx) = \mathrm{Tr}(Q). \tag{B.22}$$

Let now X be a Gaussian variable on H with distribution $\mu(dx)$. Then we have:

$$\mathbb{E}|X|^{2p} = \int_H |x|^{2p}\,\mu(dx) \leqslant C_p\,\left(\mathbb{E}|X|^2\right)^p. \tag{B.23}$$

We will use the notation $\mathbb{E}|X|^{2p} := \|X\|^{2p}_{L^{2p}}$. Let X_t be a stationary stochastic process on H with distribution $\mu(dx)$, $\mu \in \mathcal{N}(0, Q)$. Then, using the above corollary we can bound the L^{2p} norm of X_t:

$$\|X_t\|_{L^{2p}} \leqslant C_p\,\|X_t\|_{L^2}. \tag{B.24}$$

B.6 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. We list the most important types of convergence below.

Definition B.29. *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 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

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

for all $\lambda \in \mathbb{R}$ at which F is continuous.

Recall that the distribution function F_X of a random variable from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}$ induces a probability measure on $\mathbb{R}$ and that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X)$ is a probability space. We can show that the convergence in distribution is equivalent to the weak convergence of the probability measures induced by the distribution functions.

Definition B.30. *Let (E, d) be a metric space, $\mathcal{B}(E)$ the $\sigma-$algebra of its Borel sets, P_n a sequence of probability measures on $(E, \mathcal{B}(E))$ and let $C_b(E)$ denote the space of bounded continuous functions on E. We will say that the sequence of P_n converges weakly to the probability measure P if, for each $f \in C_b(E)$,*

$$\lim_{n\to+\infty} \int_E f(x)\, dP_n(x) = \int_E f(x)\, dP(x).$$

Theorem B.31. *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 and only if, for all $g \in C_b(\mathbb{R})$*

$$\lim_{n\to+\infty} \int_X g(x)\, dF_n(x) = \int_X g(x)\, dF(x). \tag{B.25}$$

Notice that (B.25) is equivalent to

$$\lim_{n\to+\infty} \mathbb{E}_n g(X_n) = \mathbb{E} g(X),$$

where E_n and E denote the expectations with respect to F_n and F, respectively.

When the sequence of random variables whose convergence we are interested in takes values in $\mathbb{R}^d$ or, more generally, a metric space space (E, d) then we can use weak convergence of the sequence of probability measures induced by the sequence of random variables to define convergence in distribution.

Definition B.32. *A sequence of real valued random variables X_n defined on a probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and taking values on a metric space (E, d) is said to converge in distribution if the indued measures $F_n(B) = P_n(X_n \in B)$ for $B \in \mathcal{B}(E)$ converge weakly to a probability measure P.*

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:

$$\mathbb{P}\Big(\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} X_n = V \Big) = 1. \tag{B.26}$$

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:

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

This is the *central limit theorem*.

A useful result is Slutksy's theorem.

Theorem B.33. (Slutsky) *Let $\{X_n\}_{n=1}^{+\infty}$, $\{Y_n\}_{n=1}^{+\infty}$ be sequences of random variables such that X_n converges in distribution to a random variable X and Y_n converges in probability to a constant $c \neq 0$. Then*

$$\lim_{n\to+\infty} Y_n^{-1} X_n = c^{-1} X,$$

in distribution.

B.7 Discussion and Bibliography

The material of this appendix is very standard and can be found in many books on probability theory and stochastic processes. See, for example [10, 22, 23, 58, 59, 49, 97].

The connection between conditional expectation and orthogonal projections is discussed in [12].

The reduced distribution functions defined in Section B.2 are used extensively in statistical mechanics. A different normalization is usually used in physics textbooks. See for instance [9, Sec. 4.2].

The calculations presented in Section B.5 are essentially an exercise in linear algebra. See [53, Sec. 10.2]. Section B.5 is based on [78, Sec. 2.3] where additional information on probability measures in infinite dimensional spaces can be found.

Limit theorems for stochastic processes are studied in detail in [43].

Index

Bibliography

[1] R. J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, volume 12 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1990.

[2] Y. Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, 36(2):906–937, 2008.

[3] A. Arnold, J. A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederman, P. A. Markowich, G. Toscani, and C. Villani. Entropies and equilibria of many-particle systems: an essay on recent research. *Monatsh. Math.*, 142(1-2):35–43, 2004.

[4] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2):43–100, 2001.

[5] L. Arnold. *Stochastic differential equations: theory and applications*. Wiley-Interscience [John Wiley & Sons], New York, 1974. Translated from the German.

[6] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[7] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

[8] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[9] R. Balescu. *Statistical dynamics. Matter out of equilibrium*. Imperial College Press, London, 1997.

[10] L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.

[11] E. A. Carlen. Conservative diffusions. *Comm. Math. Phys.*, 94(3):293–315, 1984.

[12] A.J. Chorin and O.H. Hald. *Stochastic tools in mathematics and science*, volume 1 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2006.

[13] R. I. Cukier, K. Lakatos-Lindenberg, and K. E. Shuler. Orthogonal polynomial solutions of the Fokker-Planck equation. *J. Stat. Phys.*, 9(2): 137–144 , 1973.

[14] L. Desvillettes and C. Villani. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation. *Comm. Pure Appl. Math.*, 54(1):1–42, 2001.

[15] L. Desvillettes and C. Villani. On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. *Invent. Math.*, 159(2):245–316, 2005.

[16] P. Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009.

[17] N. Wax (editor). *Selected Papers on Noise and Stochastic Processes*. Dover, New York, 1954.

[18] A. Einstein. *Investigations on the theory of the Brownian movement*. Dover Publications Inc., New York, 1956. Edited with notes by R. Fürth, Translated by A. D. Cowper.

[19] S.N. Ethier and T.G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[20] L. C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, Providence, RI, 2013.

[21] L.C. Evans. *Partial Differential Equations*. AMS, Providence, Rhode Island, 1998.

[22] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.

[23] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.

[24] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[25] A. Friedman. *Stochastic differential equations and applications. Vol. 1*. Academic Press, New York, 1975. Probability and Mathematical Statistics, Vol. 28.

[26] A. Friedman. *Stochastic differential equations and applications. Vol. 2*. Academic Press, New York, 1976. Probability and Mathematical Statistics, Vol. 28.

[27] T. C. Gard. *Introduction to stochastic differential equations*, volume 114 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 1988.

[28] C. W. Gardiner. *Handbook of stochastic methods*. Springer-Verlag, Berlin, second edition, 1985. For physics, chemistry and the natural sciences.

[29] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[30] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes*. Dover Publications Inc., Mineola, NY, 1996.

[31] J. Glimm and A. Jaffe. *Quantum physics*. Springer-Verlag, New York, second edition, 1987. A functional integral point of view.

[32] G.R. Grimmett and D.R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, 2001.

[33] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[34] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[35] R. Z. Has′minskiĭ. *Stochastic stability of differential equations*, volume 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980.

[36] U. G. Haussmann and É. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.

[37] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[38] W. Horsthemke and R. Lefever. *Noise-induced transitions*, volume 15 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1984. Theory and applications in physics, chemistry, and biology.

[39] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating Gaussian diffusions. *Ann. Appl. Probab.*, 3(3):897–913, 1993.

[40] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating diffusions. *Ann. Appl. Probab.*, 15(2):1433–1444, 2005.

[41] S. M. Iacus. *Simulation and inference for stochastic differential equations*. Springer Series in Statistics. Springer, New York, 2008. With ßfR examples.

[42] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. North-Holland Publishing Co., Amsterdam, second edition, 1989.

[43] J. Jacod and A.N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2003.

[44] D.-Q. Jiang, M. Qian, and M.-P. Qian. *Mathematical theory of nonequilibrium steady states*, volume 1833 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2004.

[45] F. John. *Partial differential equations*, volume 1 of *Applied Mathematical Sciences*. Springer-Verlag, New York, fourth edition, 1991.

[46] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[47] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic Press Inc., New York, 1981.

[48] S. Karlin and H.M. Taylor. *A first course in stochastic processes*. Academic Press, New York-London, 1975.

[49] L. B. Koralov and Y. G. Sinai. *Theory of probability and random processes*. Universitext. Springer, Berlin, second edition, 2007.

[50] N. V. Krylov. *Introduction to the theory of random processes*, volume 43 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[51] Y.A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2004.

[52] A. Lasota and M. C. Mackey. *Chaos, fractals, and noise*, volume 97 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1994.

[53] P. D. Lax. *Linear algebra and its applications*. Pure and Applied Mathematics (Hoboken). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2007.

[54] T. Lelievre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.*, 152(2): 237–274 , 2013.

[55] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations*. Imperial College Press, London, 2010. A mathematical perspective.

[56] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition, 2001.

[57] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.

[58] M. Loève. *Probability theory. I*. Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[59] M. Loève. *Probability theory. II*. Springer-Verlag, New York, fourth edition, 1978. Graduate Texts in Mathematics, Vol. 46.

[60] L. Lorenzi and M. Bertoldi. *Analytical Methods for Markov Semigroups*. CRC Press, New York, 2006.

[61] Z.M. Ma and M. Röckner. *Introduction to the theory of (nonsymmetric) Dirichlet forms*. Universitext. Springer-Verlag, Berlin, 1992.

[62] M. C. Mackey. *Time's arrow*. Dover Publications Inc., Mineola, NY, 2003. The origins of thermodynamic behavior, Reprint of the 1992 original [Springer, New York; MR1140408].

[63] M.C. Mackey, A. Longtin, and A. Lasota. Noise-induced global asymptotic stability. *J. Statist. Phys.*, 60(5-6):735–751, 1990.

[64] Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.

[65] P. Mandl. *Analytical treatment of one-dimensional Markov processes*. Die Grundlehren der mathematischen Wissenschaften, Band 151. Academia Publishing House of the Czechoslovak Academy of Sciences, Prague, 1968.

[66] X. Mao. *Stochastic differential equations and their applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997.

[67] P. A. Markowich and C. Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp.*, 19:1–29, 2000.

[68] R.M. Mazo. *Brownian motion*, volume 112 of *International Series of Monographs on Physics*. Oxford University Press, New York, 2002.

[69] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[70] A. Mira and C. J. Geyer. On non-reversible Markov chains. In *Monte Carlo methods (Toronto, ON, 1998)*, volume 26 of *Fields Inst. Commun.*, pages 95–110. Amer. Math. Soc., Providence, RI, 2000.

[71] E. Nelson. The adjoint Markoff process. *Duke Math. J.*, 25:671–690, 1958.

[72] E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, N.J., 1967.

[73] E. Nelson. *Quantum fluctuations*. Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1985.

[74] J.R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

[75] B. Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, 2003.

[76] G.A. Pavliotis and A.M. Stuart. *Multiscale methods*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[77] R. F. Pawula. Approximation of the linear Boltzmann equation by the Fokker–Planck equation. *Phys. Rev*, 162(1):186–188, 1967.

[78] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1992.

[79] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[80] H. Qian, M. Qian, and X. Tang. Thermodynamics of the general duffusion process: time–reversibility and entropy production. *J. Stat. Phys.*, 107(5/6):1129–1141, 2002.

[81] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press, New York, 1978.

[82] M. Reed and B. Simon. *Methods of modern mathematical physics. I.* Academic Press Inc., New York, second edition, 1980. Functional analysis.

[83] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[84] F. Riesz and B. Sz.-Nagy. *Functional analysis*. Dover Publications Inc., New York, 1990. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.

[85] H. Risken. Solutions of F-Planck equation in detailed balance. *Z. Physik*, 251(3):231, 1972.

[86] H. Risken. *The Fokker-Planck equation*, volume 18 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1989.

[87] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001.

[88] G. O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodol. Comput. Appl. Probab.*, 4(4):337–357 (2003), 2002. International Workshop in Applied Probability (Caracas, 2002).

[89] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[90] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

[91] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

[92] C. Schwab and R.A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[93] B. Simon. Schrödinger semigroups. *Bull. Amer. Math. Soc. (N.S.)*, 7(3):447–526, 1982.

[94] B. Simon. *Functional integration and quantum physics*. AMS Chelsea Publishing, Providence, RI, second edition, 2005.

[95] C. Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17 of *Series on Advances in Mathematics for Applied Sciences*. World Scientific Publishing Co. Inc., River Edge, NJ, 1994.

[96] D. W. Stroock. *Partial differential equations for probabilists*, volume 112 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008.

[97] D.W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993.

[98] D.W. Stroock. *An introduction to Markov processes*, volume 230 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 2005.

[99] G. I. Taylor. Diffusion by continuous movements. *London Math. Soc.*, 20:196, 1921.

[100] G. Teschl. *Mathematical methods in quantum mechanics*, volume 99 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009. With applications to Schrödinger operators.

[101] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36(5):823–841, Sep 1930.

[102] N. G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland Publishing Co., Amsterdam, 1981. Lecture Notes in Mathematics, 888.

[103] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.