Stochastic Calculus and so on

Thiago Ramos

April 15, 2026

Contents

1 Part 1: Basic Stuff **5**

1.1 Introduction 5

1.2 (Basic) Differential Equations 5

1.3 Stochastic Differential Equations and Itô Calculus 7

1.3.1 Integrating the Brownian Motion 10

1.3.2 Itô Chain Rule 11

1.3.3 Explicit Solutions to SDEs 13

1.4 The Backward and Forward Kolmogorov's Equations 14

1.4.1 Kolmogorov's Backward Equation 14

1.4.2 Fokker-Planck (-Kolmogorov ('s Forward)) Equation 16

1.5 Examples 19

1.5.1 Brownian Motion 19

1.5.2 Ornstein-Uhlenbeck 20

1.5.3 Langevin Dynamics 22

A Appendix **25**

A.1 SDE Cheat Sheet 25

Chapter 1

Part 1: Basic Stuff

1.1 Introduction

This note is based on Särkkä and Solin (2019) and Chewi (2023).

1.2 (Basic) Differential Equations

We begin by recalling how to solve ordinary differential equations in the deterministic setting. Our first example is the linear homogeneous equation

$$\frac{d}{dt}x(t) = Fx(t), \qquad x(0) \text{ given.}$$

To simplify notation, we will often write this simply as

$$\frac{d}{dt}x = Fx.$$

In undergraduate courses, one often solves this equation by an algebraic manipulation. In the scalar case, assuming $x(t) \neq 0$, we can write

$$\frac{dx}{dt} = Fx \quad \Longrightarrow \quad \frac{dx}{x} = F\,dt \quad \Longrightarrow \quad (\log x)' = F.$$

Integrating from 0 to t, we obtain

$$\log x(t) - \log x(0) = Ft,$$

and therefore

$$x(t) = e^{Ft}x(0).$$

This argument is correct in the scalar case, but there is a more general derivation that extends more naturally to other settings. Starting from the differential equation, we integrate in time:

$$\frac{d}{dt}x(t) = Fx(t) \quad \Longrightarrow \quad x(t) = x(0) + \int_0^t Fx(\tau_1)\,d\tau_1.$$

We now substitute this expression for $x(\tau_1)$ back into the integral:

$$x(t) = x(0) + F\int_0^t \left(x(0) + \int_0^{\tau_1} Fx(\tau_2)\,d\tau_2 \right) d\tau_1.$$

Expanding this gives

$$x(t) = x(0) + Ft\,x(0) + F^2 \int_0^t \int_0^{\tau_1} x(\tau_2)\,d\tau_2\,d\tau_1.$$

Repeating this procedure recursively, we obtain

$$x(t) = \left(1 + Ft + \frac{F^2t^2}{2!} + \frac{F^3t^3}{3!} + \cdots\right) x(0).$$

Hence,

$$x(t) = e^{Ft}x(0), \quad (\text{more generally } x(t) = e^{F(t-t_0)}x(t_0)).$$

This derivation is especially useful because it continues to make sense when F is a matrix (or, more generally, a linear operator), in which case the solution is still

$$x(t) = e^{Ft}x(0).$$

From now on, we will assume that we are in a multidimensional scenario, that is, x is a vector and F is a matrix.

Our next step is to consider the inhomogeneous case, that is,

$$\frac{d}{dt}x(t) = Fx(t) + Lw(t).$$

The idea for solving this problem comes from the fact that

$$\frac{d}{dt}\left(e^{-Ft}x(t)\right) = -Fe^{-Ft}x(t) + e^{-Ft}\frac{d}{dt}x(t) = e^{-Ft}\left(\frac{d}{dt}x(t) - Fx(t)\right),$$

therefore, the inhomogeneous equation can be written as

$$\begin{aligned} e^{-Ft}Lw(t) &= e^{-Ft}\left(\frac{d}{dt}x(t) - Fx(t)\right) \\ &= \frac{d}{dt}\left(e^{-Ft}x(t)\right). \end{aligned}$$

Integrating from 0 to t, we obtain

$$\int_0^t e^{-F\tau}Lw(\tau)\,d\tau = e^{-Ft}x(t) - x(0).$$

Therefore,

$$x(t) = e^{Ft}x(0) + \int_0^t e^{F(t-\tau)}Lw(\tau)\,d\tau.$$

To finish our ODE recap, we must learn how to solve a general linear differential equation, namely

$$\frac{d}{dt}x(t) = F(t)x(t) + L(t)w(t),$$

that is, in the general case, both F and L can depend on time.

In the first two cases, the solutions depend on the exponential function. For the general case, the matrix exponential no longer works, so we define a transition function $\Psi(t, t_0)$ with the analogous properties

$$\begin{aligned}
\frac{d}{d\tau}\Psi(\tau, t) &= F(\tau)\Psi(\tau, t),\\
\frac{d}{dt}\Psi(\tau, t) &= -\Psi(\tau, t)F(t),\\
\Psi(\tau, t) &= \Psi(\tau, s)\Psi(s, t),\\
\Psi(t, \tau) &= \Psi^{-1}(t, \tau),\\
\Psi(t, t) &= I.
\end{aligned}$$

We then define for the homogenous case:

$$x(t) = \Psi(t, t_0)x(t_0),$$

which recovers the homogeneous time-independent case when Ψ is the exponential function.

For example, when $L(t) = 0$, we have

$$\frac{d}{dt}x(t) = \frac{d}{dt}\Psi(t, t_0)x(t_0) = F(t)\Psi(t, t_0)x(t_0) = F(t)x(t).$$

When $L(t) \neq 0$, we use the same product-derivative idea. Note that

$$\frac{d}{dt}\left(\Psi(t_0, t)x(t)\right) = \frac{d}{dt}\Psi(t_0, t)x(t) + \Psi(t_0, t)\frac{d}{dt}x(t) = -\Psi(t_0, t)F(t)x(t) + \Psi(t_0, t)\frac{d}{dt}x(t),$$

hence

$$\frac{d}{dt}\left(\Psi(t_0, t)x(t)\right) = \Psi(t_0, t)\left(\frac{d}{dt}x(t) - F(t)x(t)\right) = \Psi(t_0, t)L(t)w(t).$$

Integrating from t_0 to t, we obtain

$$\int_{t_0}^{t} \Psi(t_0, \tau)L(\tau)w(\tau)\, d\tau = \Psi(t_0, t)x(t) - x(t_0).$$

Therefore, the **general solution is**:

$$x(t) = \Psi(t, t_0)x(t_0) + \int_{t_0}^{t} \Psi(t, \tau)L(\tau)w(\tau)\, d\tau.$$

1.3 Stochastic Differential Equations and Itô Calculus

In this section, we study differential equations of the form

$$\frac{d}{dt}x(t) = f(x(t), t) + L(x(t), t)w(t),$$

but now, unlike in the previous section, we allow the input term $w(t)$ to be a Gaussian white noise, that is, $w(t) \sim \mathcal{N}(0, 1/dt)$ for any time t. The expression above is usually written formally as

$$dx(t) = f(x(t), t)\, dt + L(x(t), t)w(t)\, dt.$$

Since $w(t) \sim \mathcal{N}(0, 1/dt)$, we have that $w(t)\,dt \sim \mathcal{N}(0, dt)$[1]. We will write $w(t)\,dt = dB(t)$, where $B(t)$ is a Brownian motion:

Definition 1.3.1 (Brownian motion)**.** *Brownian motion $B(t) \in \mathbb{R}^s$ is a continuous stochastic process with the following properties:*

1. *Any increment*

$$\Delta B_k = B(t_{k+1}) - B(t_k)$$

is a zero-mean Gaussian random variable with covariance $Q\,\Delta t_k$, where Q is the diffusion matrix of the Brownian motion and

$$\Delta t_k = t_{k+1} - t_k.$$

2. *When the time intervals of the increments do not overlap, the increments are independent.*

3. *The process starts at the origin:*

$$B(0) = 0.$$

The measure-theoretic details behind this are not important (hehe). The main point to remember is that

$$dB(t) = B(t + dt) - B(t),$$

and, by the properties above, this increment has the same distribution as

$$B(dt) - B(0) = B(dt) = \sqrt{dt}\,\mathcal{N}(0, 1).$$

For example, if we want to simulate an SDE of the form

$$dx(t) = -x(t)dt + dB(t),$$

we can approximate it by

$$x(t + dt) - x(t) = -x(t)dt + \sqrt{dt}\mathcal{N}(0, 1).$$

Therefore,

$$x(t + dt) = x(t) - x(t)dt + \sqrt{dt}\mathcal{N}(0, 1).$$

So, starting from an initial condition $x(0)$, we can simulate the process by repeating the following steps:

1. Sample $z_k \sim \mathcal{N}(0, 1)$.

2. Update

$$x_{k+1} = x_k - x_k\,dt + \sqrt{dt}\,z_k.$$

[1] Formally, $w(t)$ is a Gaussian process with mean $\mu(t) = 0$ and covariance function $k(t, s) = \delta(t - s)$, where $\delta(t - s)$ is the Dirac delta. This encodes that values at different times are independent. The variance at each time is $\mathrm{Var}(w(t)) = k(t, t) = \delta(0) = \infty$. Heuristically, δ is the limit of a Gaussian with width $\epsilon \to 0$ and unit area, whose peak grows as $1/\epsilon$.

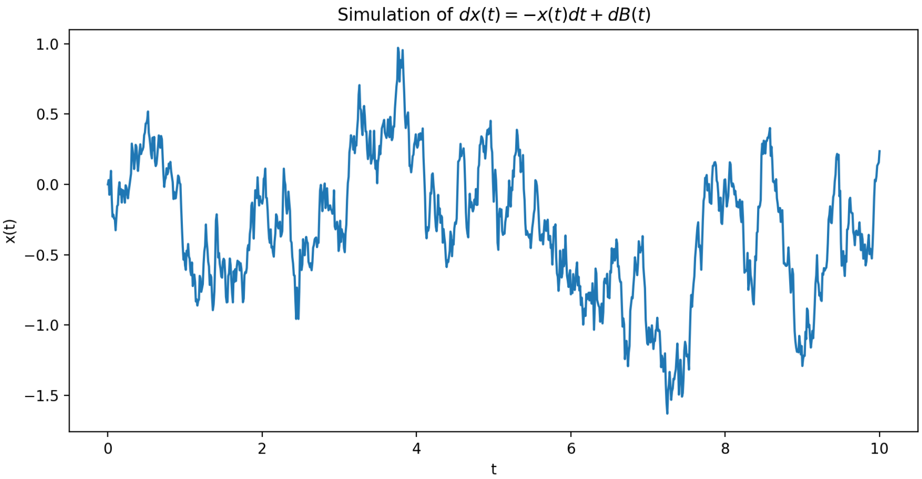

Figure 1.1: Sample path of the stochastic process $x(t)$ solving $dx(t) = -x(t)\,dt + dB(t)$, simulated with the Euler–Maruyama discretization.

3. Set $t_{k+1} = t_k + dt$.

Now, we need to develop tools to solve such differential equations, just as we did in the previous section. Note that, formally, we would expect the solution to be given simply by integration:

$$x(t) - x(0) = \int_0^t f(x(\tau), \tau)\,d\tau + \int_0^t L(x(\tau), \tau)\,dB(\tau).$$

The problem here is how to define the second integral, since it involves a random term. For example, we cannot simply define it as a Riemann–Stieltjes integral, because in a partition

$$0 = t_0 < t_1 < \cdots < t_n = t,$$

the natural Riemann–Stieltjes approximation would be

$$\sum_{k=0}^{n-1} L(x(\tau_k), \tau_k)\big(B(t_{k+1}) - B(t_k)\big), \qquad \tau_k \in [t_k, t_{k+1}].$$

The issue is that this sum depends strongly on the choice of the points τ_k. Since Brownian motion has random fluctuations at every scale, different choices of τ_k inside the same partition can lead to very different values, even when the partition is very fine. So the usual Riemann–Stieltjes construction does not work well in this setting.

The solution is to fix τ_k as either t_k or t_{k+1}. In Itô calculus, we take $\tau_k = t_k$. This choice will be justified later, but it is also possible to build a full theory by taking the other option, which leads to Stratonovich calculus. We then define the Itô integral as the limit, along a partition

$$0 = t_0 < t_1 < \cdots < t_n = t,$$

of the sums

$$\sum_{k=0}^{n-1} L(x(t_k), t_k)\big(B(t_{k+1}) - B(t_k)\big),$$

that is,

$$\int_0^t L(x(\tau),\tau)\,dB(\tau) := \lim_{n\to\infty}\sum_{k=0}^{n-1} L(x(t_k),t_k)\big(B(t_{k+1})-B(t_k)\big),$$

whenever this limit exists in the appropriate sense.

1.3.1 Integrating the Brownian Motion

Now that we have formalized how to interpret the stochastic integral in our SDE, we consider some simple examples. The first one is

$$\int_0^t B(s)\,dB(s).$$

This example is particularly interesting, as it reveals some distinctive features of Itô calculus.

Let us partition the interval $[0,t]$ into N subintervals of equal length t/N, and define

$$t_k = \frac{kt}{N}, \qquad k = 0,\dots,N.$$

Then, by the definition of the Itô integral,

$$\begin{aligned}\int_0^t B(s)\,dB(s) &= \lim_{N\to\infty}\sum_{k=0}^{N-1} B(t_k)\big(B(t_{k+1})-B(t_k)\big)\\ &= \lim_{N\to\infty}\sum_{k=0}^{N-1}\big(B(t_k)B(t_{k+1})-B(t_k)^2\big).\end{aligned}$$

Completing squares, we obtain

$$\int_0^t B(s)\,dB(s) = \lim_{N\to\infty}\left[-\frac{1}{2}\sum_{k=0}^{N-1}\big(B(t_{k+1})-B(t_k)\big)^2 + \frac{1}{2}\sum_{k=0}^{N-1}\big(B(t_{k+1})^2-B(t_k)^2\big)\right].$$

The second sum is telescoping, so

$$\frac{1}{2}\sum_{k=0}^{N-1}\big(B(t_{k+1})^2-B(t_k)^2\big) = \frac{1}{2}\big(B(t)^2-B(0)^2\big) = \frac{1}{2}B(t)^2,$$

since $B(0)=0$.

For the first sum, note that the increments of Brownian motion are independent and satisfy

$$B(t_{k+1})-B(t_k) \sim N\left(0,\frac{t}{N}\right).$$

Thus we may write

$$B(t_{k+1})-B(t_k) = \sqrt{\frac{t}{N}}\,Z_k,$$

where $Z_0,\dots,Z_{N-1}$ are i.i.d. standard normal random variables. Therefore,

$$\sum_{k=0}^{N-1}\big(B(t_{k+1})-B(t_k)\big)^2 = \frac{t}{N}\sum_{k=0}^{N-1} Z_k^2.$$

Since $\mathbb{E}[Z_k^2] = 1$, the strong law of large numbers gives

$$\frac{1}{N}\sum_{k=0}^{N-1} Z_k^2 \to 1 \qquad \text{almost surely,}$$

and hence

$$\sum_{k=0}^{N-1} \big(B(t_{k+1}) - B(t_k)\big)^2 \to t \qquad \text{almost surely.}$$

We conclude that

$$\int_0^t B(s)\, dB(s) = \frac{1}{2}B(t)^2 - \frac{1}{2}t. \tag{1.1}$$

This differs from what one might expect from ordinary calculus. Formally, one might guess that $\int_0^t B(s)\, dB(s) = \frac{1}{2}B(t)^2$, but this is not correct in the stochastic setting: an additional term, namely $\frac{1}{2}t$, appears.

1.3.2 Itô Chain Rule

In this section we derive the Itô chain rule, the central result of stochastic calculus. It will explain, among other things, why the solution of Eq. (1.1) takes the form it does.

We know that

$$dx(t) = f(x(t))dt + L(x(t))dB(t) \;\Rightarrow\; x(t + dt) = x(t) + f(x(t))dt + L(x(t))dB(t).$$

Let us start with a simple case where we want to calculate $d\phi(x(t))$ using Taylor expansion up to second order; for now, we assume that $x \in \mathbb{R}$:

$$\begin{aligned}\phi(x(t + dt)) &= \phi(x(t) + f(x(t))dt + L(x(t))dB(t)) \\ &= \phi(x(t)) + \partial_{x(t)}\phi(x(t))dx(t) + \frac{1}{2}\partial^2_{x(t)}\phi(x(t))(dx(t))^2 + o\big((dx(t))^2\big) \\ &= \phi(x(t)) + \partial_{x(t)}\phi(x(t))dx(t) \\ &\quad + \frac{1}{2}\partial^2_{x(t)}\phi(x(t))(dx(t))^2 + o\big((dx(t))^2\big).\end{aligned}$$

Now we need to know which term is not negligible when we divide by dt. Of course the first term is not negligible (think about why). Usually, in standard calculus, the second-order term vanishes when we divide by dt. However, the calculation in the previous section hints that maybe this is not the case for Itô calculus. Indeed,

$$(dx(t))^2 = f^2 dt^2 + L^2(dB(t))^2 + 2fL\, dB(t)\, dt.$$

When we divide by dt and take $dt \to 0$, the term with dt^2 vanishes. Moreover, using the fact that $dB(t) \sim \sqrt{dt}\mathcal{N}(0, 1)$, the term with $dB(t)\, dt \sim (dt)^{1.5}\mathcal{N}(0, 1)$, therefore this part also vanishes. Finally, using the same idea, the last term becomes

$$(dB(t))^2 = (B(t + dt) - B(t))^2 = \Big(\sqrt{dt}\mathcal{N}(0, 1)\Big)^2 = (\mathcal{N}(0, 1))^2\, dt.$$

So this part does not vanish when we divide by dt, differently from what happens in standard calculus! Also, we can simplify the remainder term to $o(dt)$. In the end, taking $dt \to 0$, our chain rule becomes

$$d\phi(x) = \partial_x \phi(x)\, dx + \frac{1}{2}\partial_x^2 \phi(x)(dx(t))^2.$$

In the high-dimensional case ($x(t) \in \mathbb{R}^d$), this becomes

$$d\phi(x) = \nabla_x \phi(x)^\top dx + \frac{1}{2} dx^\top \partial_{x,x}^2 \phi(x)\, dx$$

which can be written in a very cool and obfuscated form[2]:

$$d\phi(x) = \nabla_x \phi(x)^\top dx + \frac{1}{2}\mathrm{tr}\left\{ \left(\nabla \nabla^\top \phi(x) \right) dx\, dx^\top \right\}.$$

It is important to remember the following Itô algebra:

$$\begin{aligned} dB(t)dt &= 0, \\ dtdB(t) &= 0, \\ dB(t)dB(t)^T &= dt. \end{aligned}$$

Now, if we substitute $dx = f(x,t)\, dt + L(x,t)\, dB(t)$ in the expression above, we have

$$\begin{aligned} d\phi(x) &= \nabla_x \phi(x)^\top (f\, dt + L\, dB) + \frac{1}{2}\mathrm{tr}\left\{ \left(\nabla_x \nabla_x^\top \phi(x) \right) (f\, dt + L\, dB)\, (f\, dt + L\, dB)^\top \right\} \\ &= \nabla_x \phi(x)^\top f\, dt + \nabla_x \phi(x)^\top L\, dB \\ &+ \frac{1}{2}\mathrm{tr}\left\{ \left(\nabla_x \nabla_x^\top \phi(x) \right) \left(f f^\top dt^2 + L\, dB\, f^\top dt + f\, dB^\top L^\top dt + L\, dB\, dB^\top L^\top \right) \right\} \\ &= \nabla_x \phi(x)^\top f\, dt + \nabla_x \phi(x)^\top L\, dB + \frac{1}{2}\mathrm{tr}\left\{ \left(\nabla_x \nabla_x^\top \phi(x) \right) LL^\top \right\} dt \\ &= \left(\nabla_x \phi(x)^\top f + \frac{1}{2}\mathrm{tr}\left\{ \left(\nabla_x \nabla_x^\top \phi(x) \right) LL^\top \right\} \right) dt + \nabla_x \phi(x)^\top L\, dB. \end{aligned}$$

We summarize the previous results in the following theorem:

Theorem 1.3.1 (Itô's Lemma)**.** *Let $x(t) \in \mathbb{R}^d$ satisfy the SDE*

$$dx(t) = f(x(t),t)\, dt + L(x(t),t)\, dB(t),$$

and let $\phi : \mathbb{R}^d \to \mathbb{R}$ be twice continuously differentiable. Then

$$d\phi(x) = \nabla_x \phi(x)^\top dx + \frac{1}{2} tr\left\{ \left(\nabla_x \nabla_x^\top \phi(x) \right) dx\, dx^\top \right\},$$

or equivalently, substituting dx,

$$d\phi(x) = \left(\nabla_x \phi(x)^\top f(x,t) + \frac{1}{2}\, tr\left\{ \left(\nabla_x \nabla_x^\top \phi(x) \right) L(x,t)L(x,t)^\top \right\} \right) dt + \nabla_x \phi(x)^\top L(x,t)\, dB(t).$$

[2] Recall that $\langle A, B\rangle_F = \mathrm{tr}\left\{A^T B\right\} = \sum_{i,j} A_{ij} B_{ij}$.

Example 1.3.1. *Let* $\phi(x) = \frac{1}{2}x^2$ *with* $x(t) = B(t)$*. Lets apply the chain rule. We know that* $(\phi(x))'' = (x)' = 1$*, then:*

$$\begin{aligned} d\phi(x) &= xdx + \frac{1}{2}(dx)^2 \\ &= B(t)dB(t) + \frac{1}{2}(dB(t))^2 \\ &= B(t)dB(t) + \frac{1}{2}dt. \end{aligned}$$

That is,

$$\frac{1}{2}\, d(B(t))^2 = B(t)dB(t) + \frac{1}{2}dt \Rightarrow \int B(s)dB(s) = \frac{1}{2}B(t)^2 - \frac{1}{2}t,$$

just as in Eq. (1.1).

Example 1.3.2. *Suppose that* $dx = f(x)dt + dB(t)$ *where* $B(t)$ *is a Brownian Motion with constant* q*. Lets apply the chain rule for* $\phi(x) = \sin(\omega x)$*. We know that* $\phi(x)'' = (\omega\cos(\omega x))' = -\omega^2\sin(\omega x)$*. Therefore:*

$$\begin{aligned} d\phi(x) &= \omega\cos(\omega x)dx - \frac{1}{2}\omega^2\sin(\omega x)(dx)^2 \\ &= \omega\cos(\omega x)(f(x)dt + dB(t)) - \frac{1}{2}\omega^2\sin(\omega x)(f(x)dt + dB(t))^2 \\ &= \omega\cos(\omega x)(f(x)dt + dB(t)) - \frac{1}{2}\omega^2\sin(\omega x)qdt \end{aligned}$$

1.3.3 Explicit Solutions to SDEs

Suppose we have an SDE like this, with L and F constant, then

$$dx = Fxdt + LdB(t),$$

following the same idea as Section 1.2, we can multiply everything by e^{-Ft}, and then:

$$d(e^{-Ft}x) = e^{-Ft}dx - e^{-Ft}Fxdt = e^{-Ft}LdB(t)$$

therefore,

$$x(t) = e^{F(t-t_0)}x(0) + \int_{t_0}^{t} e^{F(t-\tau)}L\, dB(\tau).$$

In the more general case, with $F(t)$ and $L(t)$ depending on time, we do the same thing but using the transition matrix $\Psi(t, t_0)$, and the general solution is:

$$x(t) = \Psi(t,t_0)x(t_0) + \int_{t_0}^{t} \Psi(t,\tau)L(\tau)dB(\tau).$$

We can use this expression to simulate our SDE. To do so, note that if we split the time interval $[t_0, t]$ into N pieces, with $\Delta t = (t - t_0)/N$, then

$$\begin{aligned} \int_{t_0}^{t} e^{F(t-\tau)}dB(\tau) &= \lim_{N\to\infty}\sum_{i=0}^{N-1} e^{F(t-t_i)}\big(B(t_{i+1}) - B(t_i)\big) \\ &= \lim_{N\to\infty}\sum_{i=0}^{N-1} e^{F(t-t_i)}\sqrt{\Delta t}\, Z_i, \quad Z_i \sim \mathcal{N}(0,1)\,. \end{aligned}$$

Example 1.3.3. *The scalar SDE*

$$dx = -\lambda x\, dt + dB, \qquad x(0) = x_0,$$

where $\lambda > 0$ is a given constant and $B(t)$ is a Brownian motion, is known as the Ornstein–Uhlenbeck (OU) process. Its complete solution is

$$x(t) = e^{-\lambda t} x_0 + \int_0^t e^{-\lambda(t-\tau)}\, dB(\tau).$$

Let us choose the parameter values $\lambda = 1/2$ and $x_0 = 4$. In this case,

$$x(t) = 4e^{-t/2} + \int_0^t e^{-(t-\tau)/2}\, dB(\tau).$$

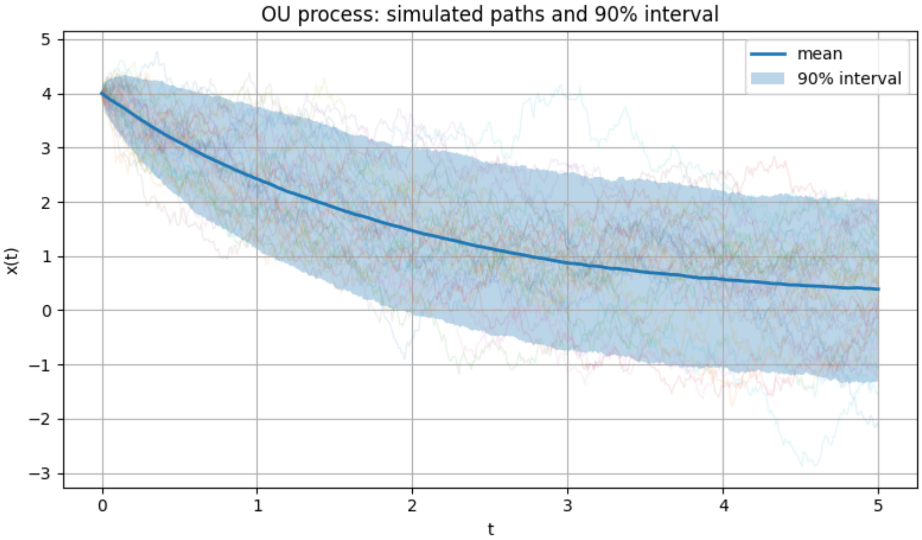

Figure 1.2: OU process simualted 2000 times.

To simulate this process, we split the interval $[0, t]$ into N pieces, with $\Delta t = t/N$, and approximate the stochastic integral by

$$\int_0^t e^{-(t-\tau)/2}\, dB(\tau) \approx \sum_{i=0}^{N-1} e^{-(t-t_i)/2}\big(B(t_{i+1}) - B(t_i)\big).$$

Since

$$B(t_{i+1}) - B(t_i) = \sqrt{\Delta t}\, Z_i, \qquad Z_i \sim \mathcal{N}(0, 1),$$

we get

$$x(t) \approx 4e^{-t/2} + \sum_{i=0}^{N-1} e^{-(t-t_i)/2} \sqrt{\Delta t}\, Z_i.$$

1.4 The Backward and Forward Kolmogorov's Equations

1.4.1 Kolmogorov's Backward Equation

The Markov semigroup associated to the SDE $dx = f(x,t)\, dt + L(x,t)\, dB(t)$ is the family of operators $(P_t)_{t\geq 0}$ acting on functions $\phi : \mathbb{R}^d \to \mathbb{R}$ by

$$P_t \phi(x) = \mathbb{E}[\phi(x(t)) \mid x(0) = x].$$

Intuitively, $P_t\phi(x)$ can be understood as follows: we start at x at time 0, run the SDE many times for t seconds, evaluate ϕ at each endpoint, and take the average. One can show that $P_0 = \mathrm{id}$ and $P_sP_t = P_tP_s = P_{s+t}$.

A very useful related concept is the infinitesimal generator. For a general time-dependent process, it is convenient to define the generalized infinitesimal generator as

$$\mathcal{A}_t\phi(x,t) = \lim_{s\to 0^+} \frac{\mathbb{E}[\phi(x(t+s), t+s) \mid x(t) = x] - \phi(x,t)}{s}.$$

If ϕ does not depend explicitly on time, this becomes

$$\mathcal{A}_t\phi(x) = \lim_{s\to 0^+} \frac{\mathbb{E}[\phi(x(t+s)) \mid x(t) = x] - \phi(x)}{s}.$$

Note that by Itô's chain rule,

$$\begin{aligned}
d\phi(x(t),t) &= \partial_t\phi(x(t),t)dt + \nabla_x\phi(x(t),t)^\top dx(t) + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top\phi(x(t),t)\, dx(t)\, dx(t)^\top\right\} \\
&= \partial_t\phi(x(t),t)dt + \nabla_x\phi(x(t),t)^\top \big(f(x(t),t)dt + L(x(t),t)dB(t)\big) \\
&\quad + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top\phi(x(t),t)\, L(x(t),t)QL(x(t),t)^\top\right\} dt \\
&= \left(\partial_t\phi(x(t),t) + \nabla_x\phi(x(t),t)^\top f(x(t),t) + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top\phi(x(t),t)\, L(x(t),t)QL(x(t),t)^\top\right\}\right) dt \\
&\quad + \nabla_x\phi(x(t),t)^\top L(x(t),t)dB(t).
\end{aligned}$$

Taking the conditional expectation given $x(t) = x$ and using that $\mathbb{E}[dB(t) \mid x(t) = x] = 0$[3], we obtain

$$\mathbb{E}[d\phi(x(t),t) \mid x(t) = x] = \left(\partial_t\phi(x,t) + \nabla_x\phi(x,t)^\top f(x,t) + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top\phi(x,t)\, L(x,t)QL(x,t)^\top\right\}\right) dt.$$

Therefore,

$$\mathcal{A}_t(\,\cdot\,) = \partial_t(\,\cdot\,) + \nabla_x(\,\cdot\,)^\top f(x,t) + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top(\,\cdot\,)\, L(x,t)QL(x,t)^\top\right\}.$$

If the process is time-invariant, that is, if f and L do not depend explicitly on time, then the generator does not depend on t. In this case, for time-independent functions ϕ, we recover the classical infinitesimal generator

$$\mathcal{A}\phi(x) = \lim_{s\to 0^+} \frac{P_s\phi(x) - \phi(x)}{s} = \lim_{s\to 0^+} \frac{\mathbb{E}[\phi(x(s)) \mid x(0) = x] - \phi(x)}{s},$$

with

$$\mathcal{A}(\,\cdot\,) = \nabla_x(\,\cdot\,)^\top f(x) + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top(\,\cdot\,)\, L(x)QL(x)^\top\right\}.$$

In a somewhat informal sense, in the time-independent setting, we have that

$$\mathcal{A}\phi(x) = \left.\frac{dP_s\phi(x)}{ds}\right|_{s=0} = \frac{P_{ds}\phi(x) - P_0\phi(x)}{ds} = \frac{P_{ds}\phi(x) - \phi(x)}{ds}.$$

[3] $dB(t) = B(t+dt) - B(t)$. Since $dB(t)$ is the Brownian increment over $[t, t+dt]$, it has zero conditional mean given the information available at time t.

Thinking of $\mathcal{A}$ in this way is very useful. For example, for a fixed ϕ,

$$\begin{aligned} dP_s(P_t\phi(x)) &= P_{ds}P_t\phi(x) - P_t\phi(x) \\ &= P_{t+ds}\phi(x) - P_t\phi(x) \\ &= d(P_t\phi(x)), \end{aligned}$$

which gives $\mathcal{A}P_t\phi = \frac{d}{dt}P_t\phi$. Moreover,

$$dP_s(P_t\phi(x)) = P_t\big(P_{ds}\phi(x) - \phi(x)\big),$$

which gives $\mathcal{A}P_t\phi = P_t\mathcal{A}\phi$. These two identities are known as Kolmogorov's Backward Equation:

Lemma 1.4.1 (Kolmogorov's Backward Equation (KBE)). *Let $(P_t)_{t\geq 0}$ be the Markov semigroup with generator $\mathcal{A}$. Then for any sufficiently regular ϕ,*

$$\frac{d}{dt}P_t\phi = \mathcal{A}P_t\phi = P_t\mathcal{A}\phi.$$

1.4.2 Fokker-Planck (-Kolmogorov ('s Forward)) Equation

In the previous section, we derived Kolmogorov's Backward Equation, which gives an expression for the evolution of the expected value of ϕ at time t, given that the process started at x. Since x_t is a random variable, instead of studying $\mathbb{E}\left[\phi(x(t))|x(0) = x\right]$, one may be interested in understanding the evolution of $p_t = p(\,\cdot\,, t)$, that is, the distribution of $x(t)$. In this section, we will derive the corresponding expression for the evolution of p_t.

Given a process $dx = f(x,t)dt + L(x,t)dB$, we have that for a small time interval dt:

$$\begin{aligned} x(t+dt) &= x(t) + f(x,t)dt + L(x,t)(B(t+dt) - B(t)) \\ &= x(t) + f(x,t)dt + L(x,t)\sqrt{dt}\mathcal{N}\left(0,1\right), \end{aligned}$$

that is, dx is random due to the Brownian motion factor. So, there exists a density function such that $x(t) \sim p(x,t)$. Our goal in this section is to describe how such density function varies in time.

To this end, note that for any function $\phi(x(t))$, using the same calculations we did for $\mathcal{A}\phi$:

$$\begin{aligned} d\phi(x) = &\nabla_x\phi(x)^\top f(x,t)dt \\ &+ \nabla_x\phi(x)^\top L(x,t)dB(t) \\ &+ \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top\phi(x)\, L(x,t)QL(x,t)^\top\right\} dt. \end{aligned}$$

Now, taking the expectation, the second term vanishes, because

$$\begin{aligned} \mathbb{E}\left[\nabla_x\phi(x)L(x,t)dB(t)\right] &= \mathbb{E}\left[\mathbb{E}\left[\nabla_x\phi(x)L(x,t)dB(t) \mid \mathcal{F}_t\right]\right] \\ &= \mathbb{E}\left[\nabla_x\phi(x)L(x,t)\mathbb{E}\left[dB(t) \mid \mathcal{F}_t\right]\right] \\ &= 0, \end{aligned}$$

since $\nabla_x\phi(x)L(x,t)$ is adapted and $dB(t)$ has zero conditional mean.

Therefore, via integration by parts [4]:

$$\begin{aligned}
\mathbb{E}\left[\frac{d}{dt}\phi(x)\right] &= \mathbb{E}\left[\nabla_x \phi(x)^\top f(x,t)\right] + \frac{1}{2}\mathbb{E}\left[\mathrm{tr}\left\{\left[\nabla\nabla^\top \phi(x)\right]\, L(x,t)QL(x,t)^\top\right\}\right] \\
&= \int \nabla_x \phi(x)^\top f(x,t)\, p(x,t)dx + \frac{1}{2}\int \mathrm{tr}\left\{\left[\nabla\nabla^\top \phi(x)\right]\, L(x,t)QL(x,t)^\top\right\} p(x,t)dx \\
&= -\int \phi(x)\nabla_x^\top \left[f(x,t)\, p(x,t)\right] dx - \frac{1}{2}\int \mathrm{tr}\left\{\left[\nabla\phi(x)\right]\, \nabla^\top \left[L(x,t)QL(x,t)^\top p(x,t)\right]\right\} dx \\
&= -\int \phi(x)\nabla_x^\top \left[f(x,t)\, p(x,t)\right] dx + \frac{1}{2}\int \phi(x)\mathrm{tr}\left\{\nabla\nabla^\top \left[L(x,t)QL(x,t)^\top p(x,t)\right]\right\} dx.
\end{aligned}$$

Note that for fixed time t, $X = x(t)$ is a random variable with density $p(x,t)$, therefore:

$$\mathbb{E}\left[\phi(X)\right] = \int \phi(x)p(x,t)dx,$$

hence,

$$\frac{d}{dt}\mathbb{E}\left[\phi(x)\right] = \frac{d}{dt}\int \phi(x)p(x,t)dx = \int \phi(x)\frac{d}{dt}p(x,t)dx,$$

and since this is true for any ϕ, we deduce that:

Proposition 1.4.1 (Fokker-Planck-Kolmogorov Equation)**.** *The density $p(x,t)$ of the process $x(t)$ satisfies*

$$\frac{d}{dt}p(x,t) = -\nabla_x^\top \left[f(x,t)\, p(x,t)\right] + \frac{1}{2}tr\left\{\nabla\nabla^\top \left[L(x,t)QL(x,t)^\top p(x,t)\right]\right\}.$$

This last equation is know as Fokker-Planck-Kolmogorov Equation.

Dual Formulation of FPK

Note that if $X \sim p$, then

$$\mathbb{E}_p\left[\phi(X)\right] = \int \phi(x)\, p(x)\, dx = \langle \phi, p\rangle_{L^2_X},$$

where the inner product is taken over the space of square-integrable functions. **TODO**: falar sobre representacao de riez

Thus, we can reinterpret the calculations from the previous section as

$$\langle \phi, \frac{d}{dt}p\rangle_{L^2_X} = \frac{d}{dt}\langle \phi, p\rangle_{L^2_X} = \langle \mathcal{A}\phi, p\rangle_{L^2_X} = \langle \phi, \mathcal{A}^* p\rangle_{L^2_X},$$

where

$$\mathcal{A}^*(\,\cdot\,) = -\nabla_x^\top \left[f(x,t)\,(\,\cdot\,)\right] + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top \left[L(x,t)QL(x,t)^\top(\,\cdot\,)\right]\right\}.$$

Since ϕ is arbitrary, it follows that

$$\frac{d}{dt}p = \mathcal{A}^* p.$$

Here, $\mathcal{A}^*$ denotes the adjoint operator of $\mathcal{A}$. As the reader may verify from the previous calculations, the operator $\mathcal{A}^*$ arises from the integration-by-parts argument used to isolate the function ϕ. Therefore, once we have access to the infinitesimal generator $\mathcal{A}$, it is straightforward to derive its adjoint by using the following facts.

[4]We assume that the mass at infinity vanishes.

- The operation of multiplication by a function $f(x)$ is self-adjoint.
- The differentiation operator satisfies

$$\left(\frac{\partial}{\partial x}\right)^* = -\frac{\partial}{\partial x},$$

  and hence the second derivative operator is self-adjoint:

$$\left(\frac{\partial^2}{\partial x^2}\right)^* = \frac{\partial^2}{\partial x^2}.$$

- The adjoint of a sum is

$$(A_1 + A_2)^* = A_1^* + A_2^*.$$

- The adjoint of a product of operators is

$$(A_1 A_2)^* = A_2^* A_1^*.$$

At this point, it is convenient to introduce the adjoint semigroup P_t^*, defined by the duality relation

$$\langle P_t \phi, p_0 \rangle = \langle \phi, P_t^* p_0 \rangle,$$

for every sufficiently regular test function ϕ and density p_0.

In other words, while P_t acts on observables, the adjoint operator P_t^* acts on probability densities. If p_0 is the density of the initial condition, then $P_t^* p_0$ is precisely the density at time t. If the transition law admits a density $p(y, t|x, 0)$, then

$$P_t \phi(x) = \int \phi(y)\, p(y, t|x, 0)\, dy,$$

and consequently

$$P_t^* p_0(y) = \int p(y, t|x, 0)\, p_0(x)\, dx.$$

Using these facts, together with Kolmogorov's Backward Equation (Lemma 1.4.1), we obtain its dual formulation:

$$\begin{aligned}
\frac{d}{dt} \langle P_t \phi(x), p(\,\cdot\,, 0) \rangle &= \frac{d}{dt} \langle \phi(x), P_t^* p(\,\cdot\,, 0) \rangle, \\
\langle \mathcal{A} P_t \phi(x), p(\,\cdot\,, 0) \rangle &= \langle \phi(x), P_t^* \mathcal{A}^* p(\,\cdot\,, 0) \rangle, \\
\langle P_t \mathcal{A} \phi(x), p(\,\cdot\,, 0) \rangle &= \langle \phi(x), \mathcal{A}^* P_t^* p(\,\cdot\,, 0) \rangle.
\end{aligned}$$

That is:

Lemma 1.4.2 (Kolmogorov's Forward Equation (KFE)). *Let $(P_t)_{t \geq 0}$ be the Markov semigroup with generator $\mathcal{A}$ and let $p_0 = p(\,\cdot\,, 0)$. Then*

$$\frac{d}{dt} P_t^* p_0 = \mathcal{A}^* P_t^* p_0 = P_t^* \mathcal{A}^* p_0.$$

Note that, just as Kolmogorov's Backward Equation gave us an expression for the evolution of $P_t\phi$, the Forward Equation gives us an expression for the corresponding evolution of the probability distribution over time.

The connection between the semigroup and the transition density is the following. If the law of $x(t)$ conditioned on $x(0) = x$ admits a density, denoted by $p(y, t|x, 0)$, then

$$P_t\phi(x) = \mathbb{E}\left[\phi(x(t))|x(0) = x\right] = \int \phi(y)\, p(y, t|x, 0)\, dy.$$

That is, the transition density is the integral kernel associated with the semigroup P_t.

Moreover, if the initial condition has density p_0, then the density at time t is given by

$$p_t(y) = P_t^* p_0(y) = \int p(y, t|x, 0)\, p_0(x)\, dx.$$

In this sense, P_t evolves observables, while P_t^* evolves probability densities.

1.5 Examples

In this section, we will apply the concepts learned in this chapter to practical examples.

1.5.1 Brownian Motion

Our first example is the simple case

$$dx(t) = dB(t),$$

for $B(t)$ a Brownian motion. That is,

$$x(t) = x(t_0) + \int_{t_0}^{t} dB(s).$$

Tts mean is $\mathbb{E}\left[x(t)\right] = x(t_0)$, moreover, if $t < s$, using the fact that $B(s) = B(t) - (B(t) - B(s))$:

$$\begin{aligned}
\operatorname{Cov}\left(x(t), x(s)\right) &= \mathbb{E}\left[B(t)(B(t) - (B(t) - B(s)))\right] \\
&= \mathbb{E}\left[B(t)B(t)\right] - \mathbb{E}\left[B(t)(B(t) - B(s))\right] \\
&= \mathbb{E}\left[B(t)B(t)\right] \\
&= t = t \wedge s.
\end{aligned}$$

This implies that

$$p(z, t|x, 0) = \mathcal{N}\left(z;\, x, t\right)$$

For any $\phi : \mathbb{R}^d \to \mathbb{R}$, by the Chain Rule (Theorem 1.3.1),

$$\begin{aligned}
d\phi(x) &= \nabla\phi(x)^\top dx + \frac{1}{2}\operatorname{tr}\left\{\nabla\nabla^\top\phi(x)\, dx\, dx^\top\right\} \\
&= \nabla\phi(x)^\top dB(t) + \frac{1}{2}\operatorname{tr}\left\{\nabla\nabla^\top\phi(x)\right\} dt.
\end{aligned}$$

For example, if $\phi(x) = x^2/2$, then

$$d\frac{B(t)^2}{2} = B(t)dB(t) + \frac{1}{2}dt,$$

as we have already shown in Equation (1.1).

By the above calculations, it follows that

$$\mathcal{A}(\,\cdot\,) = \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top(\,\cdot\,)\right\} = \frac{1}{2}\sum_i \partial_i^2(\,\cdot\,) = \frac{1}{2}\Delta(\,\cdot\,),$$

and by the rules of adjoint calculation,

$$\mathcal{A}^*(\,\cdot\,) = \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top(\,\cdot\,)\right\} = \frac{1}{2}\sum_i \partial_i^2(\,\cdot\,) = \frac{1}{2}\Delta(\,\cdot\,),$$

meaning that $\mathcal{A}$ is self-adjoint.

By the FPK Equation (Proposition 1.4.1), the distribution p_t of $x(t)$ satisfies:

$$\frac{d}{dt}p_t = \mathcal{A}^* p_t \Rightarrow \frac{d}{dt}p_t = \frac{1}{2}\Delta p_t.$$

This is the Heat Equation, and it reveals an important physical interpretation of Brownian motion: its distribution evolves according to the heat equation.

Now, lets calculate the process semigroup. We know that

$$\begin{aligned}
P_t\phi(x) &= \mathbb{E}\left[\phi(x(t))|x(0)=x\right] \\
&= \mathbb{E}\left[\phi(x(t))|x(0)=x\right] \\
&= \mathbb{E}\left[\phi(B(t)+x)\right] \\
&= \frac{1}{(\sqrt{2t\pi})^d}\int \phi(z+x)e^{\frac{-\|z\|^2}{2t}}dz \\
&= \frac{1}{(\sqrt{2t\pi})^d}\int \phi(z)e^{\frac{-\|z-x\|^2}{2t}}dz,
\end{aligned}$$

In particular, this implies that, as we've shown before:

$$p(z,t|x,0) = \frac{1}{(\sqrt{2\pi t})^d}e^{-\frac{\|z-x\|^2}{2t}},$$

which also shows that this function solves the heat equation, as discussed above.

1.5.2 Ornstein-Uhlenbeck

For our second example, we consider the OU process, given by:

$$dx(t) = -\lambda x(t)dt + dB(t).$$

It is easy to show that:

$$x(t) = e^{-\lambda(t-t_0)}x(t_0) + \int_{t_0}^t e^{-\lambda(t-s)}dB(s),$$

hence its mean is:

$$\mathbb{E}\left[x(t)\right] = e^{-\lambda(t-t_0)}x(t_0) \to 0.$$

Moreover, if $T < S$:

$$\begin{aligned}
\mathrm{Cov}\left(x(T), x(S)\right) &= \mathbb{E}\left[\int_0^T e^{-\lambda(T-t)} dB(t) \int_0^S e^{-\lambda(S-s)} dB(s)\right] \\
&\approx \mathbb{E}\left[\frac{1}{\sqrt{N}} \sum_{i=0}^{TN} e^{-\lambda(T-i/N)} Z_i \frac{1}{\sqrt{N}} \sum_{j=0}^{SN} e^{-\lambda(S-j/N)} Z_j\right] \\
&= \frac{1}{N} \sum_{i=0}^{TN} e^{-\lambda(T-i/N)} e^{-\lambda(S-i/N)} \\
&= \frac{1}{N} e^{-\lambda(T+S)} \sum_{i=0}^{TN} e^{\lambda 2i/N} \\
&\approx e^{-\lambda(T+S)} \int_0^T e^{2x\lambda} dx \\
&= e^{-\lambda(T+S)} \frac{1}{2\lambda}\left(e^{2\lambda T} - 1\right) \\
&= \frac{1}{2\lambda}\left(e^{-\lambda(S-T)} - e^{-\lambda(T+S)}\right),
\end{aligned}$$

therefore,

$$p(z, t|x, 0) = \mathcal{N}\left(z;\, e^{-\lambda t} x, \frac{1}{2\lambda}\left(1 - e^{-2\lambda t}\right)\right)$$

By the Itô's Chain rule:

$$\begin{aligned}
d\phi(x) &= \nabla\phi(x)^\top dx + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top \phi(x) dx dx^\top\right\} \\
&= \nabla\phi(x)^\top \left(-\lambda x dt + dB(t)\right) + \frac{1}{2}\Delta\phi(x) dt.
\end{aligned}$$

Therefore,

$$\mathcal{A}\phi(x) = -\lambda \nabla\phi(x)^\top x + \frac{1}{2}\Delta\phi(x),$$

and its adjoint is given by:

$$\mathcal{A}^* p_t(x) = \lambda \nabla^\top (x p_t(x)) + \frac{1}{2}\Delta p_t(x).$$

Therefore, by FPK equation:

$$\frac{d}{dt} p_t = \lambda \nabla^\top (x p_t) + \frac{1}{2}\Delta p_t.$$

We already now the solution for this PDE, namely:

$$p(z, t|x, 0) = \mathcal{N}\left(z;\, e^{-\lambda t} x, \frac{1}{2\lambda}\left(1 - e^{-2\lambda t}\right)\right).$$

Moreover, note that in this case we also can check what happens if we are in the stationary state, that is $\frac{d}{dt} p_t = 0$. Supposing that $x \in \mathbb{R}$,

$$\lambda x p_t(x) = -\frac{1}{2}\nabla p_t(x) \Rightarrow \nabla \log p_t(x) = -2\lambda x.$$

Therefore,

$$\log p(x) = -\lambda x^2 + C \Rightarrow p(x) \propto e^{-\lambda x^2} = e^{-x^2/(2/(2\lambda))},$$

that is, $\mathcal{N}\left(0, 1/(2\lambda)\right)$, which is compatible with the general solution taking $t \to \infty$.

1.5.3 Langevin Dynamics

Example 1.5.1. *When the SDE is time independent, that is,*

$$dx = f(x)dt + L(x)dB(t),$$

the solution to the FPK equation often converges to a stationary distribution satisfying $\frac{d}{dt}p(x,t) = 0$. *In some cases, it is possible to explicitly find the expression for such stationary distribution.*

Let

$$dx = -\frac{1}{2}\nabla v(x)dt + \sqrt{2D}\,dB(t).$$

By the FPK equation,

$$\frac{d}{dt}p(x,t) = -\nabla_x^\top\left(-\frac{1}{2}\nabla v(x)p(x,t)\right) + D\,tr\left\{\nabla\nabla^\top(p(x,t))\right\}.$$

Thus, if $t \gg 1$, *we may assume that we are at the stationary distribution, and therefore*

$$-\nabla^\top\left(-\frac{1}{2}\nabla v(x)p(x)\right) + D\,tr\left\{\nabla\nabla^\top(p(x))\right\} = 0.$$

Therefore

$$\nabla^\top\left(\nabla v(x)\,p(x) + 2D\nabla p(x)\right) = 0.$$

A solution to this equation is obtained by imposing

$$2D\,\nabla p(x) = -\nabla v(x)\,p(x).$$

Dividing both sides by $p(x)$, *we get*

$$2D\,\frac{\nabla p(x)}{p(x)} = -\nabla v(x),$$

that is,

$$2D\,\nabla\log p(x) = -\nabla v(x).$$

Hence,

$$\log p(x) = -\frac{v(x)}{2D} + c,$$

for some constant c, *and therefore*

$$p(x) = \frac{1}{Z}e^{-v(x)/(2D)},$$

where Z *is a normalization constant.*

Example 1.5.2. *Let*

$$dx = -\frac{\lambda}{2}xdt + \sqrt{2D}\,dB(t)$$

be an OU process. Taking

$$v(x) = \frac{\lambda}{2}x^2,$$

we have that the stationary distribution of such process is

$$p(x) = \frac{1}{Z}e^{-\lambda x^2/(4D)}.$$

Bibliography

Chewi, S. (2023). Log-concave sampling. *Book draft available at https://chewisinho. github. io*, 9:17–18. 5

Särkkä, S. and Solin, A. (2019). *Applied Stochastic Differential Equations*. 5

Appendix A

Appendix

A.1 SDE Cheat Sheet

Let $dx = f(x)dt + L(x)dB(t)$ with $x, B(t) \in \mathbb{R}^d$ and a function $\phi(x) \in \mathbb{R}$.

Chain rule (Thm. 1.3.1). By a Second Order Taylor expansion:

$$\begin{aligned} d\phi(x) &= \nabla_x \phi(x)^\top dx + \frac{1}{2}\mathrm{tr}\left\{\left(\nabla\nabla^\top \phi(x)\right) dx\, dx^\top\right\} \\ &= \left(\nabla_x \phi(x)^\top f(x,t) + \frac{1}{2}\,\mathrm{tr}\left\{\left(\nabla_x \nabla_x^\top \phi(x)\right) L(x,t)L(x,t)^\top\right\}\right) dt + \nabla_x \phi(x)^\top L(x,t)\, dB(t). \end{aligned}$$

Markov Semigroup. The Markov Semigroup P_t is defined as:

$$P_t \phi(x) = \mathbb{E}\left[\phi(x(t)) | x(0) = x\right].$$

Infinitesimal Generator. We define the infinitesimal generator $\mathcal{A}$ as:

$$\mathcal{A}\phi(x) = \frac{d}{ds} P_s \phi(x)\bigg|_{s=0} = \frac{P_{ds}\phi(x) - P_0\phi(x)}{ds} = \frac{P_{ds}\phi(x) - \phi(x)}{ds}$$

Fokker-Plank-Kolmogorov By the chain rule, integration by parts and the fact that for a time t, letting $X = x(t)$:

$$\mathbb{E}\left[\phi(x)\right] = \int \phi(x) p(x,t) dx \Rightarrow \frac{d}{dt}\mathbb{E}\left[\phi(x)\right] = \int \phi(x) \frac{d}{dt} p(x,t) dx$$

we deduce that:

$$\frac{d}{dt} p(x,t) = -\nabla_x^\top \left[f(x,t)\, p(x,t)\right] + \frac{1}{2}\mathrm{tr}\left\{\nabla\nabla^\top \left[L(x,t) Q L(x,t)^\top p(x,t)\right]\right\}.$$