MA50251

Applied SDEs

University of Bath

Lecture Notes

Alex Cox

Tony Shardlow

April 27, 2018

Chapter 1

Foundations[1]

[1] Alex Cox

Chapter 2

Methods[1]

2.1 Numerical analysis for SDEs

We consider Itô SDEs with drift $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and diffusion $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ given by

$$dX = \mu(X)\,dt + \sigma(X)\,dW(t),$$

where $W(t)$ is a vector of d **iid** Brownian motions. For simplicity in the notes, we assume that μ and σ are autonomous (independent of t), that $d = 1 = m$, and that μ and σ are smooth functions (i.e., globally Lipschitz continuous and sufficiently differentiable). For example, with an initial condition $X(0) = X_0$, this guarantees existence and uniqueness of a strong solution for $t > 0$. The results discussed and presented are available for much more general SDEs and these assumptions allow a simplicity of presentation.

We are interested in using SDEs for modelling in science and engineering. For general choice of μ and σ, exact solutions are unavailable and numerical approximations are required. One important exception is linear (even time-dependent linear) μ and σ, where solutions can be presented as a Gaussian process with explicit forms for the mean and covariance.

Here are some questions we'd like to ask of $X(t)$:

(i) Generate sample paths $X(t)$ corresponding to a fixed realisation of the Brownian

[1] Tony Shardlow

motion $W(t)$,

(ii) Sample the invariant measure (if it exists),

(iii) Compute average quantities $\mathbb{E}[\phi(X(T))]$ at time T, for some quantity of interest $\phi\colon \mathbb{R}^d \to \mathbb{R}$,

(iv) Compute exit times of the process $X(t)$ from a domain $D \subset \mathbb{R}^d$,

(v) Fit the SDE model to data, by finding μ and σ based on observations of $X(t)$.

Underlying many of the solutions to these questions is a time-stepping method.

2.1.1 Time-stepping methods

We develop time-stepping methods for the Itô SDE with $d = m = 1$. First, recall that the integral form is

$$X(t) = X_0 + \int_0^t \mu(X(s))\,ds + \int_0^t \sigma(X(s))\,dW(s), \qquad t > 0,$$

where the second integral is an Itô stochastic integral. Take the difference with $t = t_{n+1}$ and $t = t_n$,

$$X(t_{n+1}) = X(t_n) + \int_{t_n}^{t_{n+1}} \mu(X(s))\,ds + \int_{t_n}^{t_{n+1}} \sigma(X(s))\,dW(s). \tag{2.1}$$

Recall Itô's formula: for a smooth function $\phi\colon \mathbb{R} \to \mathbb{R}$,

$$\phi(X(t)) = \phi(X(s)) + \int_s^t \phi'(X(r))\,dX(r) + \frac{1}{2}\int_s^t \phi''(X(r))\sigma(X(r))^2\,dr.$$

Apply to $\phi = \mu$ and $\phi = \sigma$: for $s > t_n$,

$$\mu(X(s)) = \mu(X(t_n)) + \int_{t_n}^s \mu'(X(r))\,dX(r) + \frac{1}{2}\int_{t_n}^s \mu''(X(r))\sigma(X(r))^2\,dr$$

and similarly for σ. Substituting into (2.1), we have

$$X(t_{n+1}) = X(t_n) + \int_{t_n}^{t_{n+1}} \mu(X(t_n))\,ds + \int_{t_n}^{t_{n+1}} \sigma(X(t_n))\,dW(s) + \text{remainder}.$$

Let $\Delta W_n := W(t_{n+1}) - W(t_n)$. If we drop the remainder and substitute $X_n \approx X(t_n)$, we have

$$X_{n+1} = X_n + \mu(X_n)\Delta t + \sigma(X_n)\Delta W_n.$$

This is known as the Euler–Maruyama method. Notice how easy it is to generate samples of X_n given initial data X_0, using $\Delta W_n \sim \mathrm{N}(0, \Delta t)$ for $\Delta t = t_{n+1} - t_n$, and a Gaussian random-number generator to supply the Brownian increments.

It is similar to the explicit Euler method for ODEs with one extra term for the Brownian increment. We could also derive a method based on the **implicit** Euler method for ODEs and this would look like

$$X_{n+1} = X_n + \mu(X_{n+1})\Delta t + \sigma(X_n)\Delta W_n,$$

where the right-hand side is drift implicit (i.e., μ is evaluated at X_{n+1}) and is still diffusion explicit (i.e., σ is evalute at X_n). By its nature, when approximated on sub-intervals, the integrand of the Itô integral should be evaluated at the left-hand point and so the diffusion term for the implicit Euler–Maruyama method is unchanged.

By careful study of the remainder term, this method can be shown to converge in the mean-square sense; that is,

$$\mathbb{E}\left[\sup_{0 \le t_n \le T} \left| X_n - X(t_n) \right|^2 \right]^{1/2} = \mathcal{O}\left(\Delta t^{1/2} \right).$$

The rate of convergence is $\Delta t^{1/2}$, which is half that for the Euler method for an ODE. The reduced rate of convergence results from the irregularity of Brownian motion. To gain a higher-rate of convergence, we must refine the approximation procedure.

Consider the remainder term from our derivation of Euler–Maruyama: It consists of a sum of integrals of the form

$$\int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \text{integrand}\, dW(r)\, dW(s)$$

or similar integrals involving dr and ds instead of $dW(r)$ and $dW(s)$. On the interval $[t_n, t_{n+1}]$, the integral of dW has size $\Delta t^{1/2}$ and the integral of ds has size Δt. Hence, it is the double stochastic integral, which has the form

$$\int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \sigma'(X(r))\sigma(X(r))\, dW(r)\, dW(s),$$

that contributes most to the remainder. We apply the Itô formula with $\phi = \sigma'\sigma$ and find that, for $r > t_n$,

$$\sigma'(X(r))\sigma(X(r)) = \sigma'(X(t_n))\sigma(X(t_n)) + \int_{t_n}^{r} (\sigma'\sigma)'(X(p))\, dX(p) + \frac{1}{2} \int_{t_n}^{r} (\sigma'\sigma)''(X(p))\, dp.$$

Then, for $s > t_n$,

$$\sigma(X(s)) = \sigma(X(t_n)) + \sigma'(X(t_n))\sigma(X(t_n)) \int_{t_n}^{s} dW(r) + \text{neglected}.$$

The neglected term here is of size Δt . Now use this with (2.1) to see

$$\begin{aligned} X(t_{n+1}) = X(t_n) + \mu(X(t_n))\Delta t + \sigma(X(t_n))\Delta W_n \\ + \sigma'(X(t_n))\sigma(X(t_n)) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW(p)\, dW(s) + \text{neglected}, \end{aligned}$$

where the neglected term is of size $\Delta t^{3/2}$ (the lowest-order term consists of terms with three-times iterated stochastic integrals or with one stochastic/one deterministic integrals). Again throwing away the remainder and substituting $X_n \approx X(t_n)$, we have derived

$$X_{n+1} = X_n + \mu(X_n)\Delta t + \sigma(X_n)\Delta W_n + \sigma'(X_n)\sigma(X_n) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW(p)\, dW(s).$$

This is known as Milstein's method. Higher-order methods could also be derived by repeatedly applying Itô's formula to the remainder terms, to gain a so-called Itô–Taylor expansion and then truncating to a chosen number of terms.

Key to the applicability of Milstein's method is sampling the double stochastic integral. By the Itô formula,

$$\begin{aligned} \big(W(t) - W(s)\big)^2 &= 0^2 + \int_s^t 2\big(W(r) - W(s)\big)\, dW(r) + \frac{1}{2}\int_s^t 2\, dr \\ &= 2\int_s^t \int_s^r dW(p)\, dW(r) + (t-s), \end{aligned}$$

so that

$$2\int_s^t \int_s^r dW(p)\, dW(r) = \big(\, W(t) - W(s)\,\big)^2 - (t-s). \tag{2.2}$$

Then, Milstein's method becomes

$$X_{n+1} = X_n + \mu(X_n)\Delta t + \sigma(X_n)\Delta W_n + \sigma'(X_n)\sigma(X_n)\frac{1}{2}\big(\,\Delta W_n^2 - \Delta t\,\big).$$

This is again convenient to implement and depends only on knowing the Gaussian random variables ΔW_n at each step. The order of convergence is increased from $1/2$ to 1 and

$$\mathbb{E}\left[\sup_{0\le t_n\le T} \big|\, X_n - X(t_n)\,\big|^2\right]^{1/2} = \mathcal{O}\big(\,\Delta t^1\,\big)$$

(for sufficiently regular coefficients).

The difficulty with Milstein's method is the stochastic double integral. Our calculation is particular to one dimension and any attempt to apply Milstein's method in higher dimensions must simulate

$$\int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW^i(r)\, dW^j(s), \qquad i \neq j.$$

These integrals (or the related Levy areas) are not easy to simulate as their joint distribution does not decouple to allow use of (2.2). Some important cases do decouple, such as problems with a diagonal diffusion matrix, and Milstein's method can then be applied succesfully.

2.1.2 Modes of convergence

We already mentioned mean-square or $L^2(\Omega)$ convergence

$$\mathbb{E}\left[\sup_{0 \leq t_n \leq T} \left| X_n - X(t_n) \right|^2 \right]^{1/2} = \mathcal{O}\left(\Delta t^{\alpha} \right)$$

with $\alpha = 1/2$ for Euler–Maruyama and $\alpha = 1$ for Milstein. It is worth noting that Milstein's method with σ constant reduces to the Euler–Maruyama method and so Euler–Maruyama converges with rate $\alpha = 1$ for the important case of σ constant (also known as additive noise). This is proved using the Itô isometry to estimate the remainder terms and then applying Gronwall's inequality. A similar argument can also be used to show $L^p(\Omega)$ convergence; that is, for $p > 1$,

$$\mathbb{E}\left[\sup_{0 \leq t_n \leq T} \left| X_n - X(t_n) \right|^p \right]^{1/p} = \mathcal{O}\left(\Delta t^{\alpha} \right),$$

where the same rate α is expected to the $p = 2$ case (using the Burkholder–Gundy–Davis inequality in place of the Itô isometry). If $L^p(\Omega)$ convergence holds for all $p > 1$, a further extension can be made using the Borel–Cantelli lemma to show pathwise convergence; that is, for each $\epsilon > 0$, there exists a non-negative random variable K such that

$$\sup_{0 \leq t_n \leq T} \left| X_n - X(t_n) \right| \leq K \Delta t^{\alpha - \epsilon}.$$

Thus, we get approximation of individual sample paths with nearly the same rate, but with a constant K that depends on the particular Brownian sample path.

We now turn to averages, also known as weak convergence because we are not trying to approximate sample paths or couple $X(t_n)$ to X_n as in the cases of $L^p(\Omega)$ or

pathwise convergence. The target here is $\mathbb{E}[\phi(X(T))]$ for some quantity of interest $\phi\colon \mathbb{R} \to \mathbb{R}$ (for case $d = 1$). We measure the error by

$$\big|\, \mathbb{E}\big[\, \phi(X(T)) \,\big] - \mathbb{E}\big[\, \phi(X_N) \,\big] \,\big|, \qquad T = t_N,$$

and note that $X(T)$ and X_N may exist on completely different probability spaces. It is only their averages that are compared and $X(T)$ and its approximation X_N need not be coupled. The advantage here is that ΔW_n need only be sampled in distribution or may even be approximated, without loss of accuracy. Thus, we can replace ΔW_n by $\sqrt{\Delta t}\xi_n$ where ξ_n are **iid** $\mathsf{N}(0,1)$ random variables. Or we can replace ΔW_n by a two-point approximation to the Gaussian given by

$$\mathbb{P}\big(\, \xi_n = \pm 1 \,\big) = \frac{1}{2}.$$

This choice of ξ_n preserves the moments to order three, as it is easily verified that

$$\mathbb{E}\big[\, \xi_n^m \,\big] = \mathbb{E}\big[\, \Delta^m \,\big], \qquad \Delta \sim \mathsf{N}(0,1), \quad m = 0, 1, 2, 3.$$

The Euler–Maruyama method with either Gaussian random variables or the two-point approximation for ΔW_n has weak order of convergence one. Milstein's method also has weak convergence convergence of order one and discrete random variables can be introduced for the double stochastic integral (formula not given).

The proof of weak convergence is not given. It is usually proved by working with the backward Kolmogorov equation for $u(t, X_0) = \mathbb{E}[\phi(X(t)) \,|\, X(0) = X_0]$. This is a second-order PDE with initial condition $u(0, x) = \phi(x)$. The Euler–Maruaya method gives an approximation $u(t_n, X_0) \approx u_{\Delta t}(t_n, X_0) \coloneqq \mathbb{E}[\phi(X_n) \,|\, X_0]$ and PDE approximation theory can be used to estimate $u(t_n, X_0) - u_{\Delta t}(t_n, X_0)$.

For weak approximation, both bias error due to the choice of time-stepping method and statistical error due to approximating the expectation arise. The naive Monte Carlo method computes the sample average

$$\mathbb{E}\big[\, \phi(X_N) \,\big] \approx \frac{1}{M} \sum_{i=1}^{M} \phi(X_N^i),$$

where X_N^i are M **iid** samples of the Euler–Maruyama approximation X_N. Then, the right-hand side has variance

$$\mathsf{Var}\left(\frac{1}{M} \sum_{i=1}^{M} \phi(X_N^i) \right) = \frac{1}{M} \mathsf{Var}\big(\, \phi(X_N) \,\big),$$

as the samples are **iid**. Thus the variance of the Monte Carlo estimate with M samples is $\mathcal{O}(1/M)$ and we expect the statistical error proportional to $1/\sqrt{M}$.

It's best to have both errors (bias error of size $\mathcal{O}(\Delta t)$ and statistical error of size $\mathcal{O}(1/\sqrt{M})$) of the same size and to keep the ratio $\Delta t\sqrt{M}$ fixed as the computation is refined (for a weak first-order method). To achieve a tolerance of size ϵ, we require a time step $\Delta t = \mathcal{O}(\epsilon)$ and $M^{-1/2} = \mathcal{O}(\epsilon)$. The amount of work is the number of time steps $T/\Delta t$ times the number of samples, and hence this is at least $1/\epsilon^3$. Thus it is expensive to gain high accuracy. However, many variance-reduction techniques are available for Monte Carlo methods and these substantially improve efficiency (e.g., multilevel Monte Carlo reduces the workload to $1/\epsilon^2$).

Finally, for Stratonovich SDEs such as

$$dX = \mu(X)\,dt + \sigma(X) \circ dW(t),$$

different time-stepping methods are required. The simplest is called the Heun method and consists of

$$X_{n+1} = X_n + \mu(X_n)\Delta t + \frac{1}{2}\big(\,\sigma(X_n) + \sigma(\tilde{X}_{n+1})\,\big)\Delta W_n,$$

for $\tilde{X}_{n+1} = X_n + \sigma(X_n)\Delta W_n$. It converges with $L^2(\Omega)$-rate of convergence one.

2.2 Fokker–Planck equation

Again, we consider the initial-value problem for the autonomous Itô SDE

$$dX = \mu(X)\,dt + \sigma(X)\,dW(t), \qquad X(0) = X_0, \quad t > 0,$$

for given smooth functions $\mu, \sigma\colon \mathbb{R} \to \mathbb{R}$ and initial data $X_0 \in \mathbb{R}$. Here, we treat only the one-dimensional case with smooth coefficients, though the results presented extend easily to higher dimensions. The solution $X(t)$ at time $t > 0$ is a random variable and has a probability distribution $P(t, X_0, \cdot)$: that is, for Borel sets $\Gamma \subset \mathbb{R}$,

$$\mathbb{P}\big(\,X(t) \in \Gamma\,\big) = P(t, X_0, \Gamma).$$

Furthermore, we suppose the measure $P(t, X_0, \cdot)$ has a probability density function (pdf) $p(t, X_0, y)$, so that

$$P(t, X_0, \Gamma) = \int_\Gamma p(t, X_0, y)\,dy.$$

This is a non-trivial assumption and depends on the diffusion coefficient $\sigma(X)$ being non-degenerate, so the diffusion smooths out the initial distribution, which is a delta function centered at X_0.

The Chapman–Kolmogorov equation is a fundamental property of the family of measure $P(t, X_0, \cdot)$ that describes the autonomous (time-independent) nature of the SDE. It says, for $s, t \geq 0$, that

$$P(t+s, X_0, \Gamma) = \int_{\mathbb{R}} P(t, z, \Gamma)\, P(s, X_0, dz).$$

Thus, we integrate the time-t distribution with respect to the time-s distribution and get the time-$(t+s)$ distribution. In terms of the pdf, the Chapman–Kolmogorov equation becomes

$$p(t+s, X_0, y) = \int_{\mathbb{R}} p(t, z, y) p(s, X_0, z)\, dz.$$

The Chapman–Kolmogorov equation yields a semigroup. Define

$$\begin{aligned}(P_t f)(x) &:= \mathbb{E}\big[f(X(t)) \,|\, X_0 = x\big] = \int_{\mathbb{R}} f(y)\, P(t, x, dy) \\ &= \int_{\mathbb{R}} f(y)\, p(t, x, y)\, dy\end{aligned} \tag{2.3}$$

for smooth functions $f : \mathbb{R} \to \mathbb{R}$. Here we write the definition of $P_t f$ in three forms: using the expectation $\mathbb{E}[\cdot]$, the probability distribution $P(t, X_0, \cdot)$, and the pdf $p(t, X_0, \cdot)$. With this notation, the Chapman–Kolmogorov equation says that

$$P_{t+s} = P_t \circ P_s, \qquad s, t \geq 0.$$

We also have that $P_0 = I$ where I is the identity operator. These are the two conditions that define a semigroup. As P_t are also linear, we may now exploit some fundamental properties of semigroups of linear operators.

Recall that the generator $\mathcal{L}$ is defined by

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

and the set of f where this limit is defined is called the domain $\mathcal{D}(\mathcal{L})$.

First, we recall the simplest example of a semigroup of linear operators.

Example 2.2.1 (semigroups of linear operators on $\mathbb{R}$)**.** Let $P_t f := \mathrm{e}^{\lambda t} f$ for $f \in \mathbb{R}$ and $\lambda, t > 0$. Then P_t defines a semigroup of linear operators on $\mathbb{R}$. Clearly $P_0 f = f = If$ and $P_t \circ P_s f = \mathrm{e}^{\lambda t}\, \mathrm{e}^{\lambda s}\, f = P_{s+t} f$. Let $u(t) = P_t u_0 = \mathrm{e}^{\lambda t}\, u_0$, then

$$\frac{du}{dt} = \frac{d\, \mathrm{e}^{\lambda t}\, u_0}{dt} = \lambda u, \qquad u(0) = u_0.$$

Thus, $u(t)$ satisfies a simple differential equation. The coefficient λ is called the generator of the semigroup P_t. Note that

$$\mathcal{L}f = \lim_{t \downarrow 0} \frac{\mathrm{e}^{\lambda t} f - f}{t} = \lambda f,$$

where the limit exists for any $f \in \mathbb{R}$.

This example illustrates some general facts about semigroups. They also apply to the semigroup P_t defined in (2.3). Thus, let

$$u(t,x) \coloneqq (P_t f)(x) = \mathbb{E}\big[\, f(X(t)) \,|\, X_0 = x \,\big]$$

then

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

where $\mathcal{L}$ is the generator of the semigroup P_t. This is known as the backward Kolmogorov equation. The general solution can be written as $u(t,x) = (\mathrm{e}^{\mathcal{L}t} f)(x)$, where the exponential $\mathrm{e}^{\mathcal{L}t}$ can be defined using spectral theory for example.

To calculate the generator $\mathcal{L}$, consider the integral form of the SDE:

$$X(t) = X_0 + \int_0^t \mu(X(s))\, ds + \int_0^t \sigma(X(s))\, dW(s).$$

Then, by Itô's formula,

$$\begin{aligned} f(X(t)) - f(X_0) = \int_0^t f'(X(s))\mu(X(s))\, ds + \int_0^t f'(X(s))\sigma(X(s))\, dW(s) \\ + \frac{1}{2}\int_0^t f''(X(s))\sigma(X(s))^2\, ds. \end{aligned}$$

Under the expectation, the Itô integral disappears and

$$\begin{aligned} \frac{1}{t}\big(\mathbb{E}\big[f(X(t))\big] - f(X_0)\big) = \frac{1}{t}\int_0^t f'(X(s))\mu(X(s))\, ds \\ + \frac{1}{2}\frac{1}{t}\int_0^t f''(X(s))\sigma(X(s))^2\, ds. \end{aligned}$$

Taking the limit as $t \downarrow 0$, for sufficiently smooth f, we get

$$\frac{1}{t}\big(\mathbb{E}\big[f(X(t))\big] - f(X_0)\big) = f'(X_0)\mu(X_0) + \frac{1}{2}f''(X_0)\sigma(X_0)^2.$$

Thus, in short hand,

$$\mathcal{L} = \mu\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}.$$

In this case, the domain $\mathcal{D}(\mathcal{L})$ is non-trivial as it depends on what limit is used to define the generator. In simple terms (where the limits are taken pointwise), it should contain functions that are two-times continuously differentiable.

We have derived the following:

Backward Kolmogorov equation The function $u(t,x) = \mathbb{E}[f(X(t)) \,|\, X(0) = x]$ satisfies the second-order PDE

$$\frac{\partial u}{\partial t} = \mathcal{L}u = \frac{1}{2}\sigma^2(x)u_{xx} + \mu(x)u_x,$$

with initial condition $u(0,x) = f(x)$.

If boundary conditions are applied, then the generator $\mathcal{L}$ has a different domain and the backward Kolmogorov equation inherits these boundary conditions. For example, for a domain $D \subset \mathbb{R}$, suppose $X(t)$ is killed when it hits the boundary of D. Then, if $X_0 \in \partial D$, $\mathbb{E}[f(X(t))] = 0$ for $t > 0$ (as $X(t)$ is killed instantly), and

$$\frac{1}{t}\big(\,\mathbb{E}\big[\,f(X(t))\,\big] - f(X_0)\,\big)$$

has a well-defined limit as $t \downarrow 0$ if and only $f(X_0) = 0$. Then $\mathcal{D}(\mathcal{L})$ consists of twice continuously differentiable functions that equal zero on the boundary. The backward Kolmogorov equation on $D = (a,b)$ is

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

with boundary conditions $u(t,a) = u(t,b) = 0$ for $t > 0$.

Similarly, if $X(t)$ is reflected at the boundary, homogeneous Neumann conditions $u_x(t,a) = u_x(t,b) = 0$ are applied to the generator and the backward Kolmogorov equation. To see why this is true, consider reflection at $X = 0$; this is like solving the extending SDE

$$d\tilde{X} = \tilde{\mu}(\tilde{X})\,dt + \tilde{\sigma}(\tilde{X})\,dW(t),$$

and considering $X(t) = |\tilde{X}(t)|$ for

$$\tilde{\mu}(X) = \begin{cases} \mu(X), & X \geq 0, \\ -\mu(|X|), & X < 0, \end{cases} \qquad \tilde{\sigma}(X) = \begin{cases} \sigma(X), & X \geq 0, \\ -\sigma(|X|), & X < 0. \end{cases}$$

The $X < 0$ half-space has the same dynamics as $X \geq 0$ when reflected. Now looking at the generator we need a well-defined limit for

$$\lim_{t\downarrow 0} \frac{\mathbb{E}\big[f(\tilde{X}(t))\big] - f(X_0)}{t} = \lim_{t\downarrow 0}\left[\,\frac{1}{t}\int_0^t f'(\tilde{X}(s))\,d\tilde{X}(s) + \frac{1}{2}\sigma^2(X_0)f''(X_0)\,\right]$$

This requires that $f'(0) = 0$ due to jump discontinuities in μ and/or σ.

2.2.1 The adjoint

There is a second important PDE that describe the behaviour of SDEs known as the Fokker–Planck equation (also known as the forward Kolmogorov equation). First, we need to review some general facts about the adjoint of a linear operator.

Consider a normed vector space $\mathcal{X}$ (such as a Banach or Hilbert space) and denote the set of continuous linear functions from $\mathcal{X}$ to $\mathbb{R}$ by $\mathcal{X}^*$ (the dual space of $\mathcal{X}$). Then $f \in \mathcal{X}$ and $g \in \mathcal{X}^*$ can be related by an inner product known as **the duality pairing**, which we denote $\langle f, g \rangle_{\text{pair}}$. By definition,

$$\langle f, g \rangle_{\text{pair}} = g(f)$$

where $g \colon \mathcal{X} \to \mathbb{R}$ and $f \in \mathcal{X}$. The duality pairing is a map $\mathcal{X} \times \mathcal{X}^* \to \mathbb{R}$.

Suppose we have a continuous linear function $\mathcal{L} \colon \mathcal{D}(\mathcal{L}) \subset \mathcal{X} \to \mathcal{X}$ (the domain of definition of $\mathcal{L}$ is $\mathcal{D}(\mathcal{L})$ and in general this will be a dense subset of $\mathcal{X}$ for the type of differential operator that concern us). Clearly, $\mathcal{L}f$ is well defined for $f \in \mathcal{D}(\mathcal{L})$. Further, for any $g \in \mathcal{X}^*$, $\langle \mathcal{L}f, g \rangle_{\text{pair}}$ is well defined. Finally, the adjoint is the linear map $\mathcal{L}^* \colon \mathcal{D}(\mathcal{L}^*) \subset \mathcal{X}^* \to \mathcal{X}^*$ such that

$$\langle \mathcal{L}f, g \rangle_{\text{pair}} = \langle f, \mathcal{L}^* g \rangle_{\text{pair}}, \qquad \forall f \in \mathcal{D}(\mathcal{L}), g \in \mathcal{D}(\mathcal{L}^*).$$

Here, $\mathcal{D}(\mathcal{L}^*)$ is the subset of $\mathcal{X}^*$ for which we can define $\mathcal{L}^* g$ in this way.

Example 2.2.2 (finite dimensions)**.** For finite-dimensional vector spaces $\mathcal{X} = \mathbb{R}^d$, it is well known that the set of linear mappings from $\boldsymbol{x} \in \mathbb{R}^d \to \mathbb{R}$ is given by the inner product $\boldsymbol{g}^{\mathsf{T}}\boldsymbol{x}$ for some $\boldsymbol{g} \in \mathbb{R}^d$ (the Riesz representation theorem). Thus, any element in $\mathcal{X}^*$ can be represented by a $\boldsymbol{g} \in \mathbb{R}^d$ and $\langle \boldsymbol{f}, \boldsymbol{g} \rangle_{\text{pair}} = \boldsymbol{g}^{\mathsf{T}}\boldsymbol{f}$. Similarly, any linear operator $\mathbb{R}^d \to \mathbb{R}^d$ can be represented by a $d \times d$ matrix A. Then,

$$\langle A\boldsymbol{f}, \boldsymbol{g} \rangle_{\text{pair}} = (A\boldsymbol{f})^{\mathsf{T}}\boldsymbol{g} = \boldsymbol{f}^{\mathsf{T}}A^{\mathsf{T}}\boldsymbol{g} = \langle \boldsymbol{f}, A^{\mathsf{T}}\boldsymbol{g} \rangle_{\text{pair}},$$

so that the adjoint of A is A^{T}.

2.2.2 Fokker–Planck equation

Dual space, linear functionals, and adjoints on finite-dimensional spaces are easy. We would now like to apply these ideas for the semigroup of linear operators P_t with generator $\mathcal{L}$.

Let $C_b(\mathbb{R})$ denote the space of bounded continuous functions from $\mathbb{R}$ to $\mathbb{R}$ with the supremum norm. This is a Banach space. The dual space of $C_b(\mathbb{R})$ is the set of signed finite measures (with a finiteness condition on the measure) and the duality pairing is given by

$$\langle f, \nu \rangle_{\text{pair}} := \int_{\mathbb{R}} f(y)\, \nu(dy),$$

where ν is a measure on $\mathbb{R}$. The proof of this is difficult, but we can use this in our calculations without much trouble.

Now, P_t is a well-defined family of linear operators $C_b(\mathbb{R}) \to C_b(\mathbb{R})$. Further, P_t^*, the adjoint of P_t, is a linear operator from $\mathcal{X}^* \to \mathcal{X}^*$, where $\mathcal{X}^*$ is the set of measures on $\mathbb{R}$. To interpret the duality pairing, note that

$$\langle P_t f, \nu \rangle_{\text{pair}} = \int_{\mathbb{R}} P_t f(y)\, \nu(dy) = \int_{\mathbb{R}} f(y)\, P_t^* \nu(dy),$$

where $P_t^* \nu$ is the measure found by applying the adjoint semigroup P_t^* to ν. In terms of the pdf, for $\nu = \delta(y - X_0)$,

$$\int_{\mathbb{R}} P_t f(y)\, d\nu(y) = \int_{\mathbb{R}} f(y) p(t, X_0, y)\, dy = \int_{\mathbb{R}} f(y)\, P_t^* \nu(dy).$$

Hence, $p(t, X_0, \cdot) = P_t^* \nu$. We find the probability distribution at time t by applying the adjoint of the linear operator P_t to the initial distribution ν. The generator of the adjoint semigroup is the adjoint of the original generator. Again applying semigroup theory: the pdf $p(t, X_0, y)$ satisfies

$$\frac{dp}{dt} = \mathcal{L}^* p, \qquad p(0, y) = \delta(y - X_0),$$

where $\mathcal{L}^*$ is the generator of P_t^*, also equal to the adjoint of $\mathcal{L}$.

To find $\mathcal{L}^*$, we use the usual $L^2(\mathbb{R})$-duality pairing given by

$$\langle f, g \rangle_{L^2(\mathbb{R})} = \int_{\mathbb{R}} f(x) g(x)\, dx, \qquad f, g \in L^2(\mathbb{R}).$$

Then,

$$\begin{aligned} \langle \mathcal{L} f, g \rangle_{L^2(\mathbb{R})} &= \int_{\mathbb{R}} \mathcal{L} f(x) g(x)\, dx \\ &= \int_{\mathbb{R}} \left(\mu f' + \frac{1}{2} \sigma^2 f'' \right) g(x)\, dx = \int_{\mathbb{R}} \left(-(\mu g)' f + \frac{1}{2} (\sigma^2 g)'' f \right) dx, \end{aligned}$$

by applying integration by parts twice, and assuming rapid decay of f at infinity. Hence,

$$\mathcal{L}^* g = -\frac{\partial}{\partial x}(\mu g) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(\sigma^2 g \right).$$

We have derived the following

Fokker–Planck equation The pdf $p(t, X_0, x)$ of $X(t)$ with initial data $X(0) = X_0$ satisfies the second-order PDE

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p = \frac{1}{2}(\sigma^2(x)p)_{xx} - (\mu(x)p)_x,$$

with initial condition $p(0, x) = \delta(x - X_0)$.

Clearly, $\mathcal{L}$ and its adjoint $\mathcal{L}^*$ are different (in technical terms, $\mathcal{L}$ is not self-adjoint for $\mu \neq 0$), as a sign change occurs in the first-order term. The Fokker–Planck equation inherits boundary conditions from $\mathcal{L}^*$. For absorbing boundary conditions (where $X(t)$ is killed on the boundary of some $D \subset \mathbb{R}$), the boundary conditions are guessed easily: $p(t, X_0, a) = p(t, X_0, b) = 0$ for $t > 0$ if $D = (a, b)$. This says the pdf is zero on the boundary, which is natural as $X(t)$ is killed on the boundary.

For reflecting boundary conditions, the appropriate boundary conditions is less obvious. Write the Fokker–Planck equation as

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} J(p), \qquad J(p) \coloneqq -\frac{1}{2}\big(\, \sigma^2(x)p \,\big)_x + \mu(x)p,$$

where J is known as the probability flux. Thus, the rate of change of p at any point is given by the gradient of $J(p)$. By integrating over $D = (a, b)$,

$$\frac{d}{dt}\int_D p(t, X_0, y)\, dy = -\big[\, J(p) \,\big]_a^b.$$

For reflecting boundary conditions, the total probability $\int_D p(t, X_0, y)\, dy = 1$ (as particles are never killed), so the right-hand side is zero. We could take $J(p)(b) = J(p)(a)$, but this gives the dynamics for periodic boundary conditions. It is the choice $J(p)(a) = J(p)(b) = 0$ that describes reflecting boundary conditions.

In summary, the Fokker–Planck for the pdf of $X(t)$ is derived using duality arguments, starting from the backward Kolmogorov equation and the generator. Because the generator is not self-adjoint, the theory has many subtleties. In some cases (e.g., in one dimension where the invariant measure is known), a change of variable can be introduced so that a self-adjoint version of the Fokker–Planck equations is available. Then, explicit solutions to the Fokker–Planck equation can be written down in terms of eigenfunctions of a self-adjoint operator.

2.2.3 Examples with the Fokker–Planck equation

Why are we interested in the Fokker–Planck equation?

Example 2.2.3 (invariant measures)**.** The steady-state behaviours of a stochastic process are described by invariant measures. We say the distribution is invariant if the distribution at time t given by $P(t, X_0, \cdot)$ does not change with respect to t. Many processes have a unique invariant measure (this is known as ergodicity). It is of interest to identify systems that are ergodic and find their invariant measures.

From the Fokker–Planck equation, we know that $p(t, X_0, \cdot) = \mathrm{e}^{t\mathcal{L}^*}\,\delta(\cdot - X_0)$. If the system is ergodic, as $t \to \infty$, $p(t, X_0, \cdot) \to p_\infty$, a function independent of t and X_0. We must have $\mathcal{L}^* p_\infty = 0$ so that $\mathrm{e}^{\mathcal{L}^* t}\, p_\infty = \mathrm{e}^0\, p_\infty = p_\infty$, by looking at the Fokker–Planck equation. That means, we can identify invariant measures by solving the steady-state Fokker–Planck equation

$$\mathcal{L}^* p_\infty = 0.$$

Example 2.2.4 (Brownian dynamics)**.** An important example in statistical physics is the SDE

$$dX = -V'(X)\,dt + \sigma\,dW(t), \qquad X(0) = X_0, \tag{2.4}$$

where $V\colon \mathbb{R} \to \mathbb{R}$ is a potential energy function and σ is a constant. Suppose that $V(x) \to \infty$ as $|x| \to \infty$ and that

$$\int_{\mathbb{R}} \mathrm{e}^{-\beta V(x)}\,dx < \infty, \qquad \beta > 0, \tag{2.5}$$

(so the potential grows sufficiently rapidly that it confines the dynamics of $X(t)$). (2.4) describes the behaviour of a mass-less particle in a potential connected to a heat bath. Let's calculate the invariant measure. First, the generator

$$\mathcal{L} = -V'(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}.$$

The adjoint is

$$\mathcal{L}^* f = \frac{\partial}{\partial x}\big(\,V'(x) f\,\big) + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}.$$

We wish to solve $\mathcal{L}^* p = 0$. Using an integrating factor,

$$\frac{\partial}{\partial x}\big(\,V' p\,\big) + \frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{1}{2}\sigma^2\,\mathrm{e}^{-2V/\sigma^2}\Big(\,\mathrm{e}^{2V/\sigma^2}\,p\Big)'\right).$$

For a constant of integration c_1,

$$\frac{1}{2}\sigma^2\,\mathrm{e}^{-2V/\sigma^2}\Big(\,\mathrm{e}^{2V/\sigma^2}\,p\Big)' = c_1.$$

Rearrange

$$\Big(\,\mathrm{e}^{2V/\sigma^2}\,p\Big)' = c_1\frac{2}{\sigma^2}\,\mathrm{e}^{2V/\sigma^2}$$

and integrate again, with a second constant of integration c_2,

$$p(x) = c_1\frac{2}{\sigma^2}\,\mathrm{e}^{-2V(x)/\sigma^2}\int_{-\infty}^{x}\mathrm{e}^{2V(y)/\sigma^2}\,dy + c_2\,\mathrm{e}^{-2V(x)/\sigma^2}$$

(here we use that $p(-\infty) = 0$ for a well-defined density). To have the first term finite in the case $V(x) \to \infty$ as $|x| \to \infty$, we must take $c_1 = 0$, so that

$$p(x) = c_2 \, \mathrm{e}^{-2V(x)/\sigma^2}.$$

Notice that $\int_{\mathbb{R}} p(x)\, dx$ is finite under our assumption on V and we can choose c_2 to find the density of a probability measure. This is usually written in the form

$$p(x) = \frac{1}{Z} \mathrm{e}^{-\beta H(x)},$$

where Z is a normalisation constant (Z for Zustand or system in German), $H(x) = V(x)$ is the Hamiltonian, and $\beta = 2/\sigma^2$ is the inverse temperature.

2.3 Exit times and Kramers' escape time

Consider the initial-value problem for the ODE

$$\frac{dX}{dt} = -V'(X), \qquad X(0) = X_0.$$

Clearly, if $V'(X(t)) > 0$, then $X(t)$ decreases. Similarly, if $V' < 0$, then X increases and $X(t)$ moves towards a critical point of V. This is an example of a gradient system and it can be shown $X(t) \to X_c$, where X_c is a critical point of V and X_c is usually a minimum. This is the basis of gradient-descent methods used in numerical optimisation.

For SDEs, the story is more interesting. Consider

$$dX = -V'(X)\, dt + \sigma \, dW(t), \qquad X(0) = X_0, \tag{2.6}$$

and suppose $V(X)$ is a double-well potential (e.g., $V(X) = (X^2 - 1)^2$). For $\sigma = 0$ and $X_0 \neq 0$, the solution $X(t) \to \pm 1$ depending on the sign of X_0. In fact, the regions $\pm X > 0$ are known as basins of attraction for the equilibria $X_c = \pm 1$. For $\sigma > 0$, there is **always** a possibility of switching between different basins of attraction.

For $\sigma = \epsilon$ small, we know the SDE stays close to the ODE in the following sense.

Theorem 2.3.1. *Suppose that $V' \colon \mathbb{R} \to \mathbb{R}$ is globally Lipschitz continuous. Then, for $T > 0$, there exists $K > 0$ such that*

$$\sup_{0 \leq t \leq T} \mathbb{E}\Big[\, \big| \, X(t) - X_\epsilon(t) \, \big|^2 \, \Big]^{1/2} \leq K \big| \, \epsilon \, \big|,$$

where $X(t)$ (resp., $X_\epsilon(t)$) is the solution of (2.6) *with $\sigma = 0$ (resp., $\sigma = \epsilon$).*

Proof. This can be proved using Gronwall's inequality. From the integral form of (2.6), for $f(X) = -V'(X)$,

$$X(t) - X_\epsilon(t) = \int_0^t \big(f(X(s)) - f(X_\epsilon(s)) \big)\, ds - \epsilon W(t).$$

Then, as $(a+b)^2 \le 2(a^2+b^2)$ and $\big(\int_0^t \phi(s)\, ds\big)^2 \le t \int_0^t \phi(s)^2\, ds$ by Jensen's inequality,

$$\begin{aligned}\big|X(t) - X_\epsilon(t)\big|^2 &\le 2t \int_0^t \big|f(X(s)) - f(X_\epsilon(s))\big|^2\, ds + 2\epsilon^2 \big|W(t)\big|^2 \\ &\le 2tL^2 \int_0^t \big|X(s) - X_\epsilon(s)\big|^2\, ds + 2\epsilon^2 \big|W(t)\big|^2,\end{aligned}$$

where L denotes the Lipschitz constant of f. Hence,

$$\mathbb{E}\Big[\, \big|\, X(t) - X_\epsilon(t)\, \big|^2 \,\Big] \le 2tL^2 \int_0^t \mathbb{E}\Big[\, \big|\, X(s) - X_\epsilon(s)\, \big|^2 \,\Big]\, ds + 2\epsilon^2 t.$$

Gronwall's inequality completes the proof. □

Doob's maximal inequality for Brownian motion implies that

$$\mathbb{E}\Big[\sup_{0 \le t \le T} W(t)^2 \Big] \le 4\mathbb{E}\Big[\, \big|\, W(T)\, \big|^2 \,\Big] = 4T.$$

This inequality allows the proof of Theorem 2.3.1 to be modified, so that Gronwall applies to

$$\mathbb{E}\Big[\sup_{0 \le t \le T} \big|\, X(T) - X_\epsilon(T)\, \big|^2 \Big]$$

and a stronger result is proved (with the $\sup$ inside the expectation).

On short time scales, the SDE stays close to the ODE, where noise is small and the Brownian motion is insignificant so the system is stable and remains close to the deterministic equilibrium behaviours. However, if we wait long enough, the estimate breaks down and the Brownian motion will kick the system into a new equilibrium. On long time scales, **large deviations** occur and equilibria are said to be metastable or stable on a certain time horizon.

Metastability is important in applications. For example, proteins, which are long strings of amino acids whose interactions can be modelled by a potential energy, fold to take on different configurations or conformations corresponding to minima of the potential energy. Due to noise and thermal fluctuations in the environment, the protein can switch to a different conformation and it is interesting to understand the time scale on which the switch happens.

Our task then is to understand the time scale for switching and we proceed as follows: 1) develop a PDE for the mean first-exit time (MFET) and 2) for a simple case, use asymptotics in the small-noise limit $\epsilon \to 0$, to analyse the MFET.

2.3.1 A PDE for MFET

Consider the SDE

$$dX = \mu(X)\,dt + \sigma(X)\,dW(t), \qquad X(0) = X_0,$$

and a bounded domain D. Let $\tau(X_0) \coloneqq \inf\{t \geq 0 \colon X(t) \notin D\}$. The MFET $\bar{\tau}(X_0) \coloneqq \mathbb{E}[\tau(X_0)]$. It is neatly characterised by the following boundary-value problem for a second-order PDE:

Theorem 2.3.2.

$$-\mathcal{L}\bar{\tau} = 1, \qquad x \in D$$

with boundary conditions $\bar{\tau} = 0$ *on* ∂D*, where* $\mathcal{L}$ *is the generator of* $X(t)$*.*

If $\partial D = \partial D_1 \cup \partial D_2$ *and instead of stopping on* ∂D_2*, we reflect and only stop the process when it reaches* ∂D_1*, then the boundary condition becomes* $\bar{\tau}(x) = 0$ *for* $x \in \partial D_1$ *and* $\bar{\tau}_x(x) = 0$ *for* $x \in \partial D_2$*.*

Proof. For the boundary conditions, $\bar{\tau}(X_0) = 0$ if $X_0 \in \partial D_1$ as the process hits the boundary immediately. Roughly, if $X_0 \in \partial D_2$, then $\bar{\tau}(X_0 + \delta)$ and $\bar{\tau}(X_0 - \delta)$ are the same as $\delta \to 0$, as the process $X(t)$ is reflected to the same point for initial data $X_0 \pm \delta$ for $t > 0$. Hence $\bar{\tau}_x(X_0) = 0$.

To derive the PDE, let $p(t, X_0, y)$ denote the pdf of $X(t)$ with initial data $X(0) = X_0$, where we kill $X(t)$ after it exits the domain D. This means the total probability $\int_D p(t, X_0, y)\,dy$ decreases as t increases. Denote the probability that $X(t)$ is inside D at time t by

$$S(t, X_0) = \mathbb{P}\big(\, X(s) \in D \colon \forall s \leq t \,\big) = \int_D p(t, X_0, y)\,dy$$

and note that $S(t, X_0)$ is a decreasing function of $t > 0$. In the following, we assume $S(\infty, X_0) = 0$ (i.e., $X(t)$ reaches the boundary almost surely). To relate S to τ, note

that

$$S(t, X_0) = \mathbb{P}\big(\tau \geq t\big) = \int_t^\infty \underbrace{\mathbb{P}\big(\tau(X_0) = s\big)}_{\text{pdf of } \tau(X_0)}\, ds.$$

Hence, as $S(t, X_0) = -\int_t^\infty \frac{\partial S}{\partial t}(s, X_0)\, ds$ assuming $S(\infty, X_0) = 0$, $-\frac{\partial S}{\partial t}(s, X_0)$ is the pdf for $\tau(X_0)$.

Consequently, integrating by parts,

$$\begin{aligned}
\bar{\tau}(X_0) = \mathbb{E}\big[\tau(X_0)\big] &= \int_0^\infty -\frac{\partial S}{\partial t}(s, X_0) s\, ds \\
&= \big[-S(s, X_0)s\big]_0^\infty + \int_0^\infty S(s, X_0)\, ds \\
&= \int_0^\infty \int_D p(s, X_0, y)\, dy\, ds.
\end{aligned}$$

We know p satisfies the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \qquad p(0, X_0, y) = \delta(X_0 - y)$$

with appropriate boundary conditions on D (e.g., $p(t, X_0, y) = 0$ for $y \in D_1$). Now, the solution of the Fokker–Planck equation $p(t, X_0, y) = \mathrm{e}^{\mathcal{L}^* t}\, \delta(X_0 - y)$ so

$$\int_D p(t, X_0, y)\, dy = \int_D \mathrm{e}^{\mathcal{L}^* t}\, \delta(X_0 - y)\, dy = \int_D \delta(X_0 - y)\big(\, (\mathrm{e}^{\mathcal{L} t} 1)(y)\,\big)\, dy,$$

by definition of the adjoint. Hence,

$$\bar{\tau}(X_0) = \int_0^\infty \int_D \delta(X_0 - y)\big(\, \mathrm{e}^{\mathcal{L} t} 1\,\big)(y)\, dy\, dt = \int_0^\infty (\mathrm{e}^{\mathcal{L} t} 1)(X_0)\, dt.$$

Recall that $\frac{\partial u}{\partial t} = \mathcal{L} u$ for $u = \mathrm{e}^{\mathcal{L} t} u_0$, so that

$$\begin{aligned}
\mathcal{L}\bar{\tau} &= \int_0^\infty \mathcal{L}(\mathrm{e}^{\mathcal{L} t} 1)(X_0)\, dt \\
&= \int_0^\infty \frac{d}{dt}(\mathrm{e}^{\mathcal{L} t} 1)(X_0)\, dt \\
&= \big[(\mathrm{e}^{\mathcal{L} t} 1)(X_0)\big]_0^\infty \\
&= -1,
\end{aligned}$$

assuming that $\mathrm{e}^{\mathcal{L} t} 1 \to 0$ as $t \to \infty$. □

Example 2.3.3. Suppose that $D = [a, b]$ and $X(t) = W(t)$. Then $\mathcal{L} = \frac{1}{2}\frac{\partial^2}{\partial x^2}$ and

$$-\frac{1}{2}\bar{\tau}_{xx} = 1, \qquad \tau(a) = \tau(b) = 0.$$

This can be integrated twice and the two constants of integration identified using the boundary conditions. We find $\tau(X_0) = -X_0^2 + (a + b)X_0 - ab$.

2.3.2 Kramers' escape time

We now look at the case that $\sigma = \epsilon \to 0$, the so-called small-noise limit. We work with the SDE

$$dX = -V'(X)\,dt + \epsilon\,dW(t), \qquad X(0) = X_0,$$

for a potential energy V. Assume that V has a single minimum at $x = a$ and maximum at $x = b$ and that $V(x) \to \infty$ as $x \to -\infty$. We assume that $X_0 \approx a$ and study the exit time from the interval $[-\infty, b]$. There is no possibility of exit from $X = -\infty$ due to the confining potential (see (2.5)) and, for mathematical simplicity, we impose reflecting boundary conditions at $X = -\infty$. Thus, the MFET $\bar{\tau}$ satisfies the PDE

$$\mathcal{L}\bar{\tau} = -1, \qquad \tau(b) = 0, \quad \tau_x(-\infty) = 0.$$

Here, the generator

$$\mathcal{L} = \frac{1}{2}\epsilon^2 \frac{\partial^2}{\partial x^2} - V'(x)\frac{\partial}{\partial x}.$$

It is convenient to write $\mathcal{L}$ using the integrating factor $\mathrm{e}^{-2V(x)/\epsilon^2}$. You can verify that

$$\mathcal{L} = \mathrm{e}^{2V(x)/\epsilon^2}\,\frac{\epsilon^2}{2}\frac{\partial}{\partial x}\left(\,\mathrm{e}^{-2V(x)/\epsilon^2}\,\frac{\partial}{\partial x}\,\right).$$

Then, the MFET $\bar{\tau}$ satisfies

$$\frac{\partial}{\partial x}\left(\,\mathrm{e}^{-2V(x)/\epsilon^2}\,\frac{\partial}{\partial x}\,\right)\bar{\tau} = \frac{-2}{\epsilon^2}\,\mathrm{e}^{-2V(x)/\epsilon^2}.$$

Now we perform some simple calculus to find an expression for $\bar{\tau}$. First, integrate using the boundary condition $\tau_x(-\infty) = 0$,

$$\mathrm{e}^{-2V(x)/\epsilon^2}\,\bar{\tau}_x(x) = \frac{-2}{\epsilon^2}\int_{-\infty}^{x} \mathrm{e}^{-2V(y)/\epsilon^2}\,dy.$$

Rearrange,

$$\bar{\tau}_x(x) = \frac{-2}{\epsilon^2}\,\mathrm{e}^{2V(x)/\epsilon^2}\int_{-\infty}^{x} \mathrm{e}^{-2V(y)/\epsilon^2}\,dy,$$

and integrate again using $\bar{\tau}(b) = 0$,

$$\bar{\tau}(x) = \frac{-2}{\epsilon^2}\int_{b}^{x} \mathrm{e}^{2V(z)/\epsilon^2}\int_{-\infty}^{z} \mathrm{e}^{-2V(y)/\epsilon^2}\,dy\,dz.$$

We are interesting in the limiting behaviour as $\epsilon \to 0$. Note that $\mathrm{e}^{-2V(y)/\epsilon^2}$ is very much smaller that $\mathrm{e}^{-2V(a)/\epsilon^2}$ when ϵ is small and y is far from the minimum a. As in a Laplace approximation, we write the Taylor expansion

$$V(y) = V(a) + 0 + \frac{1}{2}(y-a)^2 V''(a) + \text{neglect},$$

where we use that $V'(a) = 0$ (as a minimum of V). We assume the minimum is non-degenerate so that $V''(a) > 0$ and will neglect the higher-order terms. With this simplification,

$$\int_{-\infty}^{z} \mathrm{e}^{-2V(y)/\epsilon^2}\, dy = \int_{-\infty}^{z} \mathrm{e}^{-2V(a)/\epsilon^2}\, \mathrm{e}^{-(y-a)^2 V''(a)/\epsilon^2}\, dy \approx \int_{-\infty}^{\infty} \mathrm{e}^{-2V(a)/\epsilon^2}\, \mathrm{e}^{-(y-a)^2 V''(a)/\epsilon^2}\, dy.$$

Here, we can extend the range of integration because the integral is dominated by the region $y \approx a$. Recall the Gaussian integral

$$\int_{\mathbb{R}} \mathrm{e}^{-x^2/2\sigma^2}\, dx = \sqrt{2\pi\sigma^2}.$$

Then, with $\sigma^2 = \epsilon^2/2V''(a)$,

$$\int_{-\infty}^{z} \mathrm{e}^{-2V(y)/\epsilon^2}\, dy \approx \mathrm{e}^{-2V(a)/\epsilon^2}\, \sqrt{\pi\epsilon^2/V''(a)}.$$

We must also study the integral

$$\int_{x}^{b} \mathrm{e}^{2V(z)/\epsilon^2}\, dz.$$

This time, as b is a maximum of V, write $V(z) = V(b) + 0 + \frac{1}{2}(z-b)^2 V''(b) +$ neglected terms for $z \in (a, b)$. In this case, we assume the maximum is non-degenerate so that $V''(b) < 0$. Then,

$$\begin{aligned}
\int_{x}^{b} \mathrm{e}^{2V(z)/\epsilon^2}\, dz &= \int_{x}^{b} \mathrm{e}^{2\left(V(b)+\frac{1}{2}(z-b)^2 V''(b)\right)/\epsilon^2}\, dz \\
&= \mathrm{e}^{2V(b)/\epsilon^2} \int_{x}^{b} \mathrm{e}^{-(z-b)^2|V''(b)|/\epsilon^2}\, dz \\
&\approx \mathrm{e}^{2V(b)/\epsilon^2}\, \frac{1}{2} \int_{-\infty}^{\infty} \mathrm{e}^{-(z-b)^2|V''(b)|/\epsilon^2}\, dz
\end{aligned}$$

using the Gaussian integral with $\sigma^2 = \epsilon^2/2|V''(v)|$

$$= \mathrm{e}^{2V(b)/\epsilon^2}\, \frac{1}{2} \sqrt{\pi\epsilon^2/\left|V''(b)\right|}.$$

Putting everything together, we have shown that

$$\bar{\tau}(X_0) \approx \frac{2}{\epsilon^2}\, \mathrm{e}^{2(V(b)-V(a))/\epsilon^2}\, \frac{1}{2} \sqrt{\frac{\pi^2\epsilon^4}{V''(a)|V''(b)|}} = \mathrm{e}^{\beta\Delta V}\, \frac{\pi}{\omega_a \omega_b},$$

for $\beta = 2/\epsilon^2$ and $\omega_x = \sqrt{|V''(x)|}$ and $\Delta V = V(b) - V(a)$. Notice this is independent of initial data $X_0 \approx a$ and that the time is very large when the potential drop ΔV is much large than the temperature β^{-1}. It also grows like $\mathrm{e}^{2\Delta V/\epsilon^2}$ in the small-noise limit $\epsilon \downarrow 0$, so extremely rapidly! In numerical simulations, it is impossible to probe such time scales accurately and different questions must be asked, such as what is the lowest ΔV or potential barrier to a different metastable state?

2.4 Parameter estimation

Suppose we are given an SDE model

$$dX = \mu(X)\,dt + \sigma(X)\,dW(t), \qquad X(0) = X_0,$$

and wish to determine $\mu(X)$ and $\sigma(X)$ from observations of $X(t)$. We study this problem under the following simplifying assumptions: 1) $X(t)$ is one-dimensional and 2) $\mu(X)$ and $\sigma(X)$ are known up to some parameters. For example, we may assume that $X(t)$ is an Ornstein–Uhlenbeck (OU) process and that $\mu(X) = \lambda X$ and $\sigma(X) = \sigma$ (a constant) and need to determine $\lambda, \sigma \in \mathbb{R}$ to fit the data $X(t)$.

Our strategy is 1) replace the SDE by one with a constant diffusion coefficient by a change of variables (known as the Lamperti transform), 2) estimate the constant diffusion coefficient, and 3) estimate parameters in the (transformed) drift.

2.4.1 Lamperti transform

The Lamperti transform is a change of variables that allows one-dimensional SDEs to be rewritten as an SDE with additive noise. Let $Y = h(X)$ for some function h to be determined. Then, by Itô's formula,

$$dY = \mathcal{L}h(X)\,dt + h'(X)\,\sigma(X)\,dW(t)$$

and, to achieve constant diffusion, we require $h'(X)\sigma(X) = \sigma_0$, for some convenient constant σ_0. Hence, set

$$h(x) = \int_0^x \frac{\sigma_0}{\sigma(y)}\,dy.$$

Note that

$$\mathcal{L}h = \left[\mu\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}\right]h = \mu\frac{\sigma_0}{\sigma} + \frac{1}{2}\frac{\sigma^2(-\sigma'\sigma_0)}{\sigma^2} = \frac{\sigma_0\mu}{\sigma} - \frac{\sigma'\sigma_0}{2}.$$

Hence, Y satisfies the SDE with additive noise given by

$$dY = \bar{\mu}(Y)\,dt + \sigma_0\,dW(t),$$

for the transformed drift

$$\bar{\mu}(Y) := \frac{\sigma_0\mu(h^{-1}(Y)}{\sigma(h^{-1}(Y))} - \frac{1}{2}\sigma_0\sigma'(h^{-1}(Y)).$$

Example 2.4.1 (Cox–Ingersol–Ross)**.** One of the SDE models in mathematical finance is

$$dX = (a_0 - a_1 X)\,dt + \sigma_0\sqrt{X}\,dW(t)$$

for some parameters a_0, a_1, σ_0. We can eliminate the $\sqrt{X}$ term in the diffusion, using a change of variable $h(x) = \int x^{-1/2}\,dx = 2\sqrt{x}$ (constants of integration do not matter). Then $h^{-1}(y) = (y/2)^2$ and

$$\begin{aligned} dY &= \left[\frac{a_0 - a_1(Y/2)^2}{Y/2} - \frac{1}{2}\sigma_0^2\frac{1}{2(Y/2)}\right]dt + \sigma_0\,dW(t) \\ &= \left[\frac{2a_0}{Y} - \frac{a_1 Y}{2} - \frac{\sigma_0^2}{2Y}\right]dt + \sigma_0\,dW(t). \end{aligned}$$

This transformation allows us to work with

$$dX = \mu(X)\,dt + \sigma\,dW(t),$$

where σ is a constant. We could also scale $\sigma = 1$, but that puts another unknown coefficient into the drift μ. We prefer to work out the diffusion coefficient with the following method.

2.4.2 Diffusion coefficient

Theorem 2.4.2. *Suppose that μ is bounded. Let*

$$\hat{\sigma}_J^2 := \frac{1}{J\Delta t}\sum_{j=0}^{J-1}(X_{j+1} - X_j)^2, \qquad X_j := X(j\Delta t).$$

Then, $\mathbb{E}[|\hat{\sigma}_J^2 - \sigma^2|] \le C(\Delta t + \Delta t^{1/2})$ for some constant C. Hence, $\hat{\sigma}_J^2 \to \sigma^2$ in $L^1(\Omega)$ as $J \to \infty$ with $J\Delta t$ fixed.

Proof. From the integral form of the SDE,

$$X_{j+1} - X_j = \int_{t_j}^{t_{j+1}} \mu(X(s))\,ds + \sigma\Delta W_j = I_j + \sigma\Delta W_j,$$

for $\Delta W_j = W(t_{j+1}) - W(t_j)$ and $I_j = \int_{t_j}^{t_{j+1}} \mu(X(s))\,ds$. Hence,

$$\hat{\sigma}_J^2 = \frac{1}{J\Delta t}\sigma^2\sum_{j=0}^{J-1}\Delta W_j^2 + \frac{1}{J\Delta t}\sigma\sum_{j=0}^{J-1}2I_j\,\Delta W_j + \frac{1}{J\Delta t}\sum_{j=0}^{J-1}I_j^2.$$

Now,

$$\begin{aligned}\hat{\sigma}_J^2 - \sigma^2 &= \frac{1}{J\Delta t}\sigma^2 \sum_{j=0}^{J-1}\left(\Delta W_j^2 - \Delta t\right) + \frac{1}{J\Delta t}\sigma \sum_{j=0}^{J-1} 2I_j\,\Delta W_j + \frac{1}{J\Delta t}\sum_{j=0}^{J-1} I_j^2 \\ &= \frac{1}{J}\sigma^2 \sum_{j=0}^{J-1} A_j + \frac{1}{J}\sigma \sum_{j=0}^{J-1} B_j + \frac{1}{J}\sum_{j=0}^{J-1} C_j,\end{aligned}$$

for $A_j \coloneqq (\Delta W_j^2 - \Delta t)/\Delta t$, $B_j \coloneqq 2I_j\Delta W_j/\Delta t$, and $C_j \coloneqq I_j^2/\Delta t$.

As μ is bounded, we have $C \coloneqq \|\mu\|_\infty < \infty$. Note that

(i) A_j are **iid** with mean zero and variance 2 (i.e., $\mathbb{E}\left[A_j^2\right] = 2$).

(ii) $I_j \le C\Delta t$ so that $C_j \le C^2\Delta t$.

(iii) Using $ab \le (a^2 + b^2)/2$, $B_j \le \Delta t^{-3/2} I_j^2 + \Delta t^{-1/2}\Delta W_j^2 \le C^2\Delta t^{1/2} + \Delta t^{-1/2}\Delta W_j^2$.

Now,

$$\mathbb{E}\left[\left|\hat{\sigma}_J^2 - \sigma^2\right|\right] \le \frac{1}{J}\left[\sigma^2\mathbb{E}\left[\left|\sum_{j=0}^{J-1} A_j\right|\right] + \sigma\mathbb{E}\left[\left|\sum_{j=0}^{J-1} B_j\right|\right] + \mathbb{E}\left[\left|\sum_{j=0}^{J-1} C_j\right|\right]\right]$$

(as $\mathbb{E}[X] \le (\mathbb{E}[X^2])^{1/2}$ Jensen's inequality)

$$= \frac{1}{J}\left[\sigma^2\left(\mathbb{E}\left[\left|\sum_{j=0}^{J-1} A_j\right|^2\right]\right)^{1/2} + \sigma\mathbb{E}\left[\left|\sum_{j=0}^{J-1} B_j\right|\right] + \mathbb{E}\left[\left|\sum_{j=0}^{J-1} C_j\right|\right]\right]$$

(cross terms A_jA_k have mean zero, triangle inequality, and (ii) above)

$$\le \frac{1}{J}\left[\sigma^2\left(\mathbb{E}\left[\sum_{j=0}^{J-1} A_j^2\right]\right)^{1/2} + \sigma\sum_{j=0}^{J-1}\mathbb{E}\left[\left|B_j\right|\right] + \sum_{j=0}^{J-1} C^2\Delta t\right]$$

(using (i) and (iii) above with expectation)

$$\le \sigma^2\left(2\Delta t\right)^{1/2} + \sigma\left(C^2\Delta t^{1/2} + \Delta t^{1/2}\right) + C^2\Delta t.$$

Hence the result is proved for a possibly larger constant C. □

This is the quadratic variation introduced earlier. Notice that $\hat{\sigma}_J^2$ is easily computed given samples X_j.

2.4.3 Parameter estimation for the drift

To determine the drift $\mu(X)$, we employ a maximum-likelihood estimator and optimise a measure

$$L(\text{observations} \,|\, \text{parameters})$$

of the likelihood of the observed data over choice of parameters. In the case of Gaussian random variables $X \sim \mathrm{N}(\mu, \sigma^2)$, the likelihood of single piece of data x is given by the pdf,

$$L(x \,|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

and, for multiple data points $x_1, \ldots, x_N$ assumed to be independent,

$$L(x_1, \ldots, x_N \,|\, \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{N/2}} \prod_{i=1}^{N} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right).$$

Now, to determine μ and σ, maximise the likelihood by solving

$$\frac{\partial L}{\partial \mu} = \frac{\partial L}{\partial \sigma} = 0.$$

It is simple to show in this case that

$$\mu = \frac{x_1 + \cdots + x_N}{N}, \qquad \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2.$$

How does this scenario generalise for SDEs?

2.4.4 Likelihood function for SDEs

The path $X(t)$ generated by the SDE is a continuous function $[0, T] \to \mathbb{R}^d$ or, in other words, $X(t)$ gives a $C([0, T], \mathbb{R}^d)$-valued random variable. To use the maximum-likelihood method with data given by observations of paths, we need to define a likelihood function on $C([0, T], \mathbb{R}^d)$. Random variables on this space have no pdf in the classical sense, as the space is too infinite dimensional. Instead, we look for a pdf with respect to a well-defined measure on the path space. This is given to us by the famous Girsanov formula.

Let $\mathbb{P}_W$ be the measure on $C([0, T], \mathbb{R}^d)$ given by Brownian motion. We use $\mathbb{P}_W$ as a reference measure. The distribution of $X(t)$, which we denote $\mathbb{P}_X$ has a density with

respect to $\mathbb{P}_W$ and this generalised pdf can be used as a likelihood function. Girsanov's formula gives that

$$\frac{d\mathbb{P}_X}{d\mathbb{P}_W}(X) = \exp\left[\int_0^T \mu(X(s))\,dX(s) - \frac{1}{2}\int_0^T \mu(X(s))^2\,ds\right].$$

The left-hand side is the derivative in the sense of Radon–Nikodyn. This means

$$\mathbb{P}_X(\Gamma) = \int_\Gamma \exp\left[\int_0^T \mu(X(s))\,dX(s) - \frac{1}{2}\int_0^T \mu(X(s))^2\,ds\right] d\mathbb{P}_W(X),$$

for a measurable set $\Gamma \subset C([0,T],\mathbb{R}^d)$. We use $d\mathbb{P}_X/d\mathbb{P}_W$ for the likelihood $L(\{X(t)\colon 0 \le t \le T\}\,|\,\mu(X))$. Let's do an example, before worrying about where this formula comes from.

Example 2.4.3 (OU process)**.** Consider $X(t)$ such that

$$dX = -\lambda X\,dt + \sigma\,dW(t).$$

Suppose that σ is known and we wish to determine λ from observations of the sample path $X(t)$. The log likelihood

$$\begin{aligned}\log L &= \int_0^T \mu(X(s))\,dX(s) - \frac{1}{2}\int_0^T \mu(X(s))^2\,ds \\ &= -\lambda\int_0^T X(s)\,dX(s) - \frac{1}{2}\lambda^2\int_0^T X(s)^2\,ds.\end{aligned}$$

Now solve $\frac{d\log L}{d\lambda} = 0$ (maximising L is equivalent to maximising $\log L$), which means

$$-\int_0^T X(s)\,dX(s) - \lambda\int_0^T X^2(s)\,ds = 0$$

and $\lambda = -B_1/M_2$, if

$$B_n := \int_0^T X(s)^n\,dX(s), \qquad M_n := \int_0^T X(s)^n\,ds.$$

These can be evaluated given a sample path $X(s)$ via quadrature.

Example 2.4.4. Consider the SDE

$$dX = (a_0X - a_1X^3)\,dt + dW(t),$$

for some parameters a_0, a_1 to be determined. The log likelihood

$$\log L = \int_0^T (a_0X(s) - a_1X(s)^3)\,dX(s) - \frac{1}{2}\int_0^T (a_0X(s) - a_1X(s)^3)^2\,ds.$$

To find the maximum-likelihood estimates for a_0, a_1, put

$$\frac{\partial \log L}{\partial a_0} = \frac{\partial \log L}{\partial a_1} = 0.$$

Hence,

$$B_1 - \frac{2}{2}a_0 M_2 + a_1 M_4 = 0$$
$$-B_3 + a_0 M_4 - a_1 M_6 = 0,$$

which can be rephrased as a linear system of equations

$$\begin{bmatrix} -M_2 & M_4 \\ M_4 & -M_6 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} -B_1 \\ B_3 \end{bmatrix}.$$

The matrix and right-hand side can be estimated using quadrature given samples of $X(t)$. Then, the linear system can be solved to determine a_0 and a_1.

2.4.5 Girsanov's theorem

It remains to derive Girsanov's change-of-variable formula. We take a very simple case and suppose $\sigma = 1$, so that the Euler–Maruyama approximation says that

$$\Delta X_n \coloneqq X_{n+1} - X_n = \mu(X_n)\Delta t + \Delta W_n.$$

The pdf of the increments $\Delta W_i \sim \mathrm{N}(0, \Delta t)$ is

$$p_W(\Delta W_i) \coloneqq \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\Delta W_i)^2 \right).$$

We have the conditional distribution $X_{n+1} - X_n | X_n \sim \mathrm{N}(\mu(X_n)\Delta t, \Delta t)$ and hence the vector of increments ΔX_i has pdf

$$\begin{aligned} p_X(\Delta X_i) &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\Delta X_i - \mu(X_i)\Delta t)^2 \right) \\ &= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\Delta X_i^2 - 2\mu(X_i)\Delta t \Delta X_i + \mu(X_i)^2 \Delta t^2) \right). \end{aligned}$$

Taking the ratio,

$$\frac{p_X(\Delta X_i)}{p_W(\Delta X_i)} = \exp\left(\sum_{i=0}^{N-1} \mu(X_i)\Delta X_i - \frac{1}{2}\mu(X_i)^2 \Delta t \right).$$

Now take the limit as $\Delta t \to 0$,

$$\frac{p_X(\Delta X_i)}{p_W(\Delta X_i)} \to \exp\left(\int_0^T \mu(X(s))\, dX(s) - \frac{1}{2}\int_0^T \mu(X(s))^2\, ds \right).$$

This gives the form for Girsanov's formula.

2.5 Notes

Further details of the material discussed can be found in

(i) Numerical methods [3, 2].

(ii) Fokker–Planck equation and exit-time problems [6, 1]. The rigorous theory of semigroups of linear operators is described in [7].

(iii) Girsanov formula and parameter estimation [4, 6].

Appendix A

Gronwall's inequality: Suppose that $z(t)$ satisfies

$$0 \leq z(t) \leq a + \int_0^t b(s)z(s)\,ds, \qquad t \geq 0,$$

for $a \geq 0$ and a non-negative and integrable function $b(s)$. Then,

$$z(t) \leq a\,\mathrm{e}^{B(t)}, \qquad \text{where } B(t) \coloneqq \int_0^t b(s)\,ds.$$

Bibliography

[1] Mark Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260. Springer, 2012.

[2] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[3] Gabriel Lord, Catherine Powell, and Tony Shardlow. *An Introduction to Computational Stochastic PDEs*. CUP, 2014.

[4] Bernt Øksendal. *Stochastic Differential Equations*. Springer, 2003.

[5] Grigorios Pavliotis and Andrew Stuart. *Multiscale Methods*, volume 53 of *Texts in Applied Mathematics*. Springer, 2008.

[6] Grigorios A. Pavliotis. *Stochastic Processes and Applications*. Springer, 2014.

[7] Amnon Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44. Springer, 2012.