8

Stochastic Ordinary Differential Equations

A stochastic ordinary differential equation (SODE) is an ordinary differential equation with a random forcing, usually given by a white noise $\zeta(t)$. White noise is chosen so that the random forces $\zeta(t)$ are uncorrelated at distinct times t. For example, adding noise to the ODE $du/dt = -\lambda u$, we consider the SODE

$$\frac{du}{dt} = -\lambda u + \sigma\zeta(t), \qquad u(0) = u_0 \tag{8.1}$$

for parameters $\lambda, \sigma > 0$ and an initial condition $u_0 \in \mathbb{R}$. As we saw in §6.3, $\zeta(t) = dW(t)/dt$ for a Brownian motion $W(t)$ and we rewrite (8.1) in terms of $W(t)$ by integrating over $[0,t]$. Consider

$$u(t) = u_0 - \lambda \int_0^t u(s)\, ds + \sigma W(t), \tag{8.2}$$

which is written in short as

$$du = -\lambda u\, dt + \sigma dW(t), \qquad u(0) = u_0.$$

The solution is a stochastic process $\{u(t)\colon t \geq 0\}$ such that (8.2) holds for $t \geq 0$ and is known as the *Ornstein–Uhlenbeck process*.

More generally, we introduce a vector-valued function $\boldsymbol{f}\colon \mathbb{R}^d \to \mathbb{R}^d$, known as the drift, and a matrix-valued function $G\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$, known as the diffusion, and consider SODEs of the form

$$\boldsymbol{u}(t) = \boldsymbol{u}_0 + \int_0^t \boldsymbol{f}(\boldsymbol{u}(s))\, ds + \int_0^t G(\boldsymbol{u}(s))\, d\boldsymbol{W}(s), \tag{8.3}$$

also written for brevity as

$$d\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{u})\, dt + G(\boldsymbol{u})\, d\boldsymbol{W}(t), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

where $\boldsymbol{u}_0 \in \mathbb{R}^d$ is the initial condition and $\boldsymbol{W}(t) = [W_1(t), \ldots, W_m(t)]^\mathsf{T}$ for *iid* Brownian motions $W_i(t)$. The last term in (8.3) is a *stochastic integral* and it needs careful definition. When $G(\boldsymbol{u})$ is independent of $\boldsymbol{u}$ (such as (8.3) for $G(u) = \sigma$), we say the SODE has *additive noise* and the stochastic integral is simply $\int_0^t G\, d\boldsymbol{W}(s) = G\, \boldsymbol{W}(t)$. The case where $G(\boldsymbol{u})$ varies with $\boldsymbol{u}$, such as Example 8.4, is called *multiplicative noise* and the integrand $G(\boldsymbol{u}(s))$ is random. The stochastic integrals in Chapter 6 are for deterministic integrands and do not include this case. The proper interpretation of the stochastic integral is a delicate issue and must be understood when modelling a physical situation, as the integral $\int_0^t G(\boldsymbol{u}(s))\, d\boldsymbol{W}(s)$ and hence the solution $\boldsymbol{u}(t)$ of the SODE depends on the interpretation.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

We prefer the Itô integral, which corresponds to modelling the Brownian increment $d\boldsymbol{W}(s)$ as independent of the current state $\boldsymbol{u}(s)$. The Itô integral is often used in mathematics and financial modelling as it leads to martingales (see Theorem 8.14). An alternative is to 'smooth' the differential equation, replacing $\boldsymbol{W}(t)$ by continuously differentiable approximations so that the integrals are defined by conventional calculus and a stochastic integral is defined in a limit. This approach leads to the *Stratonovich* integral, usually denoted

$$\int_0^t G(\boldsymbol{u}(s)) \circ d\boldsymbol{W}(s),$$

and to Stratonovich SODEs. Such an approach is often appropriate in physical modelling, where Brownian motion is used to model a driving noise process that is, in reality, smoother than Brownian motion.

We develop the Itô stochastic integral and calculus in §8.2. In §8.3, we introduce the Itô formula and study geometric Brownian motion, an example from financial modelling, and prove the existence and uniqueness of solutions to the initial-value problem for Itô SODEs subject to Lipschitz conditions on the drift and diffusion.

In §8.4 and §8.6, the numerical approximation by finite difference methods is addressed. The simplest such method is the *Euler–Maruyama* method, which is the stochastic version of the Euler method (3.12). We also look at a higher-order method called *Milstein's method*. We prove strong convergence for Milstein's method and examine stability of the Euler–Maruyama method. We show how to implement both methods in MATLAB and examine the $L^2(\Omega)$ approximation error numerically. For weak approximation, we describe the Monte Carlo method and a multilevel Monte Carlo method for variance reduction. A discussion of the corresponding ideas for Stratonovich SODEs is given in §8.7.

8.1 Examples of SODEs

We begin by presenting four examples of SODEs. The first three describe the dynamics of a particle in a mechanical system subject to noise. We denote the momentum by P, the position by Q, and the system energy of the particle by $H(Q,P)$. The fourth example, geometric Brownian motion, models the fluctuation of an asset price on the stock market.

Example 8.1 (Ornstein–Uhlenbeck process) Consider a particle of unit mass moving with momentum $P(t)$ at time t in a gas, subject to irregular bombardment by the ambient gas particles. The dynamics of the particle can be modelled by a dissipative force $-\lambda P(t)$, where $\lambda > 0$ is known as the dissipation constant, and a fluctuating force $\sigma\zeta(t)$, where $\zeta(t)$ is a white noise and $\sigma > 0$ is the diffusion constant. Newton's second law of motion gives the acceleration dP/dt as the sum of the two forces. We have

$$\frac{dP}{dt} = -\lambda P + \sigma\,\zeta(t), \tag{8.4}$$

which is the linear SODE (8.1) with additive noise $G = \sigma$. It is also written as

$$dP = -\lambda P\,dt + \sigma\,dW(t). \tag{8.5}$$

We usually consider the initial-value problem with $P(0) = P_0$ and the solution $P(t)$ is known as the *Ornstein–Uhlenbeck process*.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

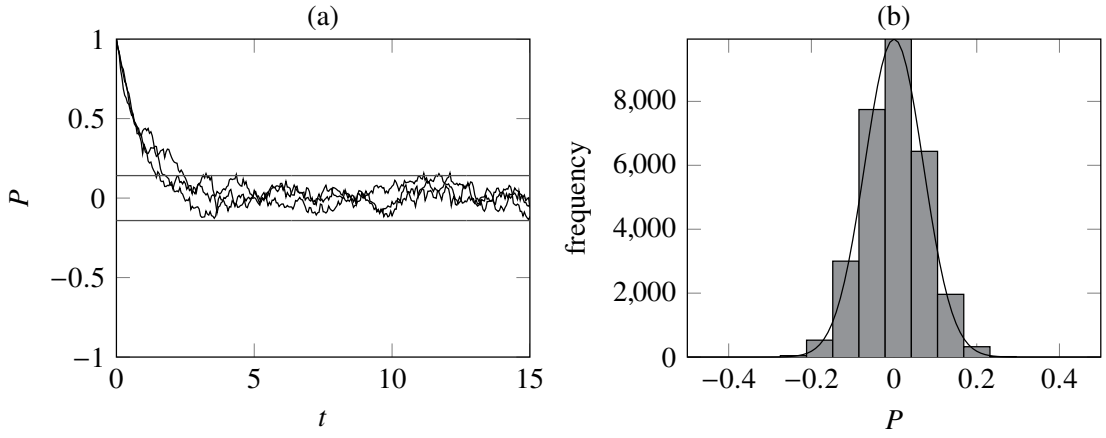

Figure 8.1 (a) Sample paths of the Euler–Maruyama approximation for (8.4) with $P_0 = 1$, $\lambda = 1$, and $\sigma = 1/2$ with $\Delta t = 0.05$. As $t \to \infty$, $P(t) \to \mathrm{N}(0, \sigma^2/2\lambda)$ in distribution and the horizontal lines mark the 90% confidence interval $\left[-2\sqrt{\sigma^2/2\lambda}, +2\sqrt{\sigma^2/2\lambda}\right]$. (b) A histogram of $P(t_n)$ for $t_n \in [10, 1000]$ and the shape is similar to the Gaussian pdf (also shown).

The variation of constants formula also applies for SODEs (see Example 8.21 and Exercise 8.7) and, in the case of (8.5), says that

$$P(t) = \mathrm{e}^{-\lambda t} P_0 + \sigma \int_0^t \mathrm{e}^{-\lambda(t-s)}\, dW(s), \tag{8.6}$$

where the second term is a stochastic integral (see Definition 6.22). We use this to derive an important relationship in statistical physics.

By definition, the expected kinetic energy per degree of freedom of a system at temperature T is given by $k_\mathrm{B}T/2$, where k_B denotes the Boltzmann constant. In particular, the temperature of a system of particles in thermal equilibrium can be determined from λ and σ. For the Ornstein–Uhlenbeck model, the expected kinetic energy $\mathbb{E}\big[P(t)^2/2\big]$ is easily calculated using Proposition 6.23:

$$\begin{aligned} \mathbb{E}\big[P(t)^2\big] &= \mathrm{e}^{-2\lambda t} P_0^2 + \sigma^2 \int_0^t \mathrm{e}^{-2\lambda(t-s)}\, ds \\ &= \mathrm{e}^{-2\lambda t} P_0^2 + \frac{\sigma^2}{2\lambda}\big(1 - \mathrm{e}^{-2\lambda t}\big) \to \frac{\sigma^2}{2\lambda} \qquad \text{as } t \to \infty. \end{aligned} \tag{8.7}$$

The expected kinetic energy $\mathbb{E}\big[\frac{1}{2}P(t)^2\big]$ converges to $\sigma^2/4\lambda$ as $t \to \infty$. Thus, $\sigma^2/4\lambda$ is the equilibrium kinetic energy and the equilibrium temperature is given by $k_\mathrm{B}T = \sigma^2/2\lambda$, which is known as the fluctuation–dissipation relation.

Taking this further, it can be shown that

$$P(t) \to \mathrm{N}(0, \sigma^2/2\lambda) \qquad \text{in distribution as } t \to \infty$$

(see Figure 8.1). In physics, the limiting distribution is known as the Gibbs canonical distribution and can be written $\mathrm{N}(0, k_\mathrm{B}T)$ with probability density function (pdf) $p(Q, P) = \mathrm{e}^{-\beta H(Q,P)}/Z$, where $\beta = 1/(k_\mathrm{B}T)$ is the inverse temperature, Z is a normalisation constant, and $H(Q, P)$ is the system energy for a particle with position Q and momentum P (in this case $H = P^2/2$).

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

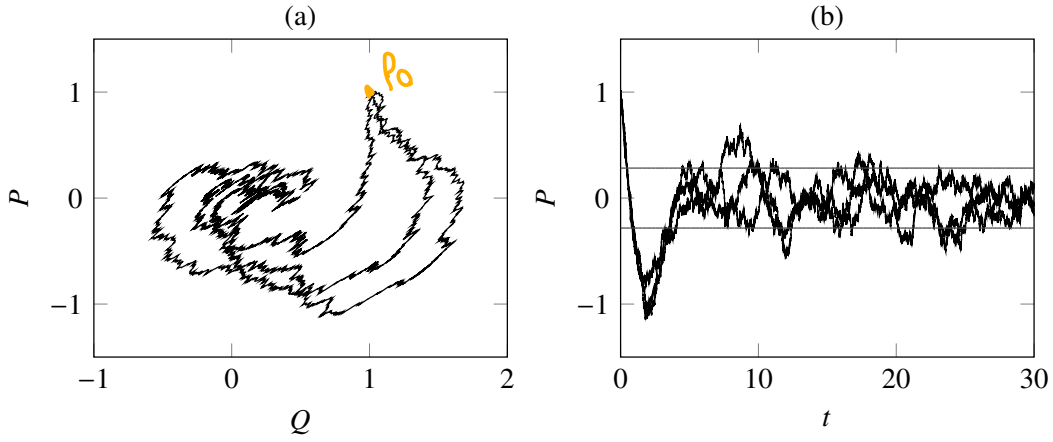

Figure 8.2 (a) Phase-plane plot of three approximate sample paths for the Langevin equation (8.8) with initial data $[Q(0),P(0)]^{\mathsf{T}} = [1,1]^{\mathsf{T}}$ and parameters $\lambda = 1$, $\sigma = 1/5$ in the case $V(Q) = Q^2/2$. (b) The same sample paths plotted as a time series $P(t)$ against t, with horizontal lines marking the 90% confidence interval $\left[-2\sqrt{\sigma^2/2\lambda}, +2\sqrt{\sigma^2/2\lambda}\right]$ of the distribution $\mathrm{N}(0,\sigma^2/2\lambda)$. The solutions were generated by the explicit Euler–Maruyama method with time step $\Delta t = 0.0025$.

Example 8.2 (Langevin equation) If the particle in Example 8.1 has potential energy $V(Q)$ at position $Q \in \mathbb{R}$, its dynamics are described by the following SODE:

$$\begin{aligned} dQ &= P\,dt \\ dP &= \left[-\lambda P - V'(Q)\right] dt + \sigma\,dW(t) \end{aligned} \tag{8.8}$$

for parameters $\lambda, \sigma > 0$. In the notation of (8.3), $d = 2$, $m = 1$, and

$$\boldsymbol{f}(\boldsymbol{u}) = \begin{pmatrix} P \\ -\lambda P - V'(Q) \end{pmatrix}, \qquad G(\boldsymbol{u}) = \begin{pmatrix} 0 \\ \sigma \end{pmatrix},$$

for $\boldsymbol{u} = [Q,P]^{\mathsf{T}}$. The noise is additive and acts directly on the momentum P only. This SODE is known as the Langevin equation and a phase-plane plot of three approximate sample paths is shown in Figure 8.2(a) for $V(Q) = Q^2/2$.

This time, the equilibrium distribution has pdf $p(Q,P) = \mathrm{e}^{-\beta H(Q,P)}/Z$ for $H(Q,P) = P^2/2 + V(Q)$, normalisation constant Z, and inverse temperature $\beta = 1/(k_{\mathrm{B}}T) = 2\lambda/\sigma^2$. As in Example 8.1, $P(t)$ converges to a Gaussian distribution and Figure 8.2(b) shows the long-time behaviour of three sample paths of $P(t)$.

Example 8.3 (Duffing–van der Pol equation) Consider the Duffing–van der Pol SODE

$$\begin{aligned} dQ &= P\,dt \\ dP &= \left[-P(\lambda + Q^2) + \left(\alpha Q - Q^3\right)\right] dt + \sigma Q\,dW(t) \end{aligned} \tag{8.9}$$

comprising a nonlinear dissipation with parameter λ and a conservative force due to the potential $V(Q) = Q^4/4 - \alpha Q^2/2$ with parameter α. We have the following SODE for $\boldsymbol{u}(t) = [Q(t),P(t)]^{\mathsf{T}}$:

$$d\boldsymbol{u} = \begin{pmatrix} P \\ -P(\lambda + Q^2) + \alpha Q - Q^3 \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma Q \end{pmatrix} dW(t). \tag{8.10}$$

The SODE has *multiplicative noise*, as the diffusion $G(\boldsymbol{u}) = [0,\sigma Q]^{\mathsf{T}}$ depends on $\boldsymbol{u} = [Q,P]^{\mathsf{T}}$. We explore the numerical solution of this equation in Exercise 8.9.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

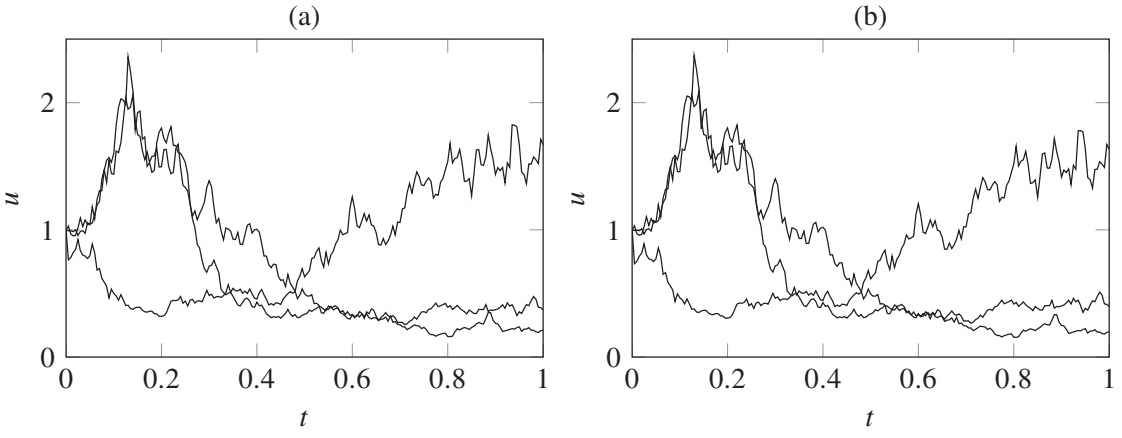

Figure 8.3 (a) Three sample paths of geometric Brownian motion, the solution of (8.11) with initial data $u_0 = 1$ and parameters $r = 1$, $\sigma = 5$ generated using the exact solution (8.34). (b) Approximations of the same sample paths given by the Euler–Maruyama method with $\Delta t = 0.005$. Geometric Brownian motion is non-negative with $u(t) \geq 0$ if $u_0 \geq 0$.

The solutions $\boldsymbol{u}(t)$ of these three examples are the same for the Itô and Stratonovich interpretations of the integral equation (8.3) (see Exercise 8.20).

Example 8.4 (geometric Brownian motion) In financial modelling, the price $u(t)$ at time t of a risk-free asset with interest rate r obeys a differential equation $du/dt = ru$. On the stock market, stock prices fluctuate rapidly and the fluctuations are modelled by replacing the risk-free interest rate r by a stochastic process $r + \sigma\zeta(t)$, for a parameter σ (known as the volatility) and white noise $\zeta(t)$. This gives the following SODE for the price $u(t)$:

$$du = r\,u\,dt + \sigma\,u\,dW(t), \qquad u(0) = u_0, \tag{8.11}$$

where $u_0 \geq 0$ is the price at time $t = 0$. In terms of (8.3), $d = m = 1$, the drift $f(u) = ru$, and the diffusion $G(u) = \sigma u$. As G depends on u, the SODE is said to have *multiplicative noise* and, for this example, the solution $u(t)$ depends on the interpretation of the integral $\int_0^t u(s)\,dW(s)$, as we show in Exercise 8.19. The Itô SODE is chosen to model stock prices, because the current price $u(t)$ of the stock is determined by past events and is independent of the current fluctuations $dW(t)$. The solution process $u(t)$ of (8.11) is called *geometric Brownian motion* and example paths are plotted in Figure 8.3(a). In Example 8.20, we find the exact solution $u(t)$ of (8.11) and hence show $u(t) \geq 0$. This is desirable in financial modelling, where stock prices are non-negative.

8.2 Itô integral

The purpose of this section is to develop the Itô integral

$$I(t) := \int_0^t X(s)\,dW(s), \tag{8.12}$$

for a class of stochastic processes $X(t)$. First, we introduce the technical conditions that ensure the integral is well defined. This subject is technical and many proofs are omitted or address only simple cases.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Stochastic processes model a time evolution and the idea of past and future is fundamental. For example, we can ask questions of a Brownian motion $W(t)$ from the point of view of a future or past time. Looking back from the future, $W(t)$ is known from our observations of the past. However, looking into the future, $W(t)$ is a random variable whose outcome is not yet known. The best we can say is the distribution of $W(t)$ at a future time t conditioned on our observations up to time $s < t$ is known to be Gaussian with mean $W(s)$ and variance $t-s$ (as the increments of Brownian motion are independent and have distribution $\mathrm{N}(0,t-s)$). Define the σ-algebra $\mathcal{F}_s := \sigma(W(s))$ (see Definition 4.18). Then,

$$\mathbb{E}\big[W(t) \,|\, \mathcal{F}_s\big] = W(s \wedge t)$$

where $s \wedge t := \min\{s,t\}$. This property of $W(t)$, known as the martingale property, extends to the stochastic integral $I(t)$, as we show in Theorem 8.14.

To handle the idea of past and future more generally, we use sub σ-algebras (see Definition 4.17) and introduce a sub σ-algebra $\mathcal{F}_t$ for each time t. Intuitively, the events in $\mathcal{F}_t$ are those observable by time t and, because we have more observations as time passes, the σ-algebras $\mathcal{F}_t$ contain more events as t increases. We make this precise with the notion of *filtration*.

Definition 8.5 (filtration) Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space.

(i) A *filtration* $\{\mathcal{F}_t : t \geq 0\}$ is a family of sub σ-algebras of $\mathcal{F}$ that are increasing; that is, $\mathcal{F}_s$ is a sub σ-algebra of $\mathcal{F}_t$ for $s \leq t$. Each $(\Omega,\mathcal{F}_t,\mathbb{P})$ is a measure space and we assume it is complete (see Definition 1.17).

(ii) A *filtered probability space* is a quadruple $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$, where $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space and $\{\mathcal{F}_t : t \geq 0\}$ is a filtration of $\mathcal{F}$.

Stochastic processes that conform to the notion of time described by the filtration $\mathcal{F}_t$ are known as *adapted* processes.

Definition 8.6 (adapted) Let $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$ be a filtered probability space. A stochastic process $\{X(t) : t \in [0,T]\}$ is $\mathcal{F}_t$-*adapted* if the random variable $X(t)$ is $\mathcal{F}_t$-measurable for all $t \in [0,T]$.

For example, if $X(t)$ is $\mathcal{F}_t$-adapted, the event $\{X(s) \leq a\}$ is observable at time s and should be included in $\mathcal{F}_t$ for each $t \geq s$. This is true because $\mathcal{F}_s$ is a sub σ-algebra of $\mathcal{F}_t$ and hence $X(s)$ is $\mathcal{F}_t$-measurable. Then, by definition of $\mathcal{F}_t$ measurable, the pullback set $X(s)^{-1}((-\infty,a]) = \{X(s) \leq a\}$ belongs to $\mathcal{F}_t$.

To stress the relationship between Brownian motion and a filtration, we give the following definition.

Definition 8.7 ($\mathcal{F}_t$-Brownian motion) A real-valued process $\{W(t) : t \geq 0\}$ is an $\mathcal{F}_t$-Brownian motion on a filtered probability space $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$ if

(i) $W(0) = 0$ a.s.,
(ii) $W(t)$ is continuous as a function of t,
(iii) $W(t)$ is $\mathcal{F}_t$-adapted and $W(t) - W(s)$ is independent of $\mathcal{F}_s$, $s < t$, and
(iv) $W(t) - W(s) \sim \mathrm{N}(0,t-s)$ for $0 \leq s \leq t$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

The $\mathcal{F}_t$-Brownian motion is a simple extension of Definition 5.11 (Brownian motion) and is easily constructed by defining an appropriate filtration. Just as each random variable X has an associated σ-algebra $\sigma(X)$ (see Definition 4.18), each process $\{X(t): t \geq 0\}$ has an associated filtration, known as the *natural filtration*.

Definition 8.8 (natural filtration) If $\{X(t): t \geq 0\}$ is a stochastic process, let $\mathcal{F}_t$ be the smallest σ-algebra such that $X(s)$ is measurable for all $s \leq t$. $\{\mathcal{F}_t : t \geq 0\}$ is called the *natural filtration* of $X(t)$.

Thus, $\mathcal{F}_t$-Brownian motions can be constructed from a Brownian motion $W(t)$ and its natural filtration.

Definition and properties

We turn to the definition of the Itô stochastic integral (8.12) of an $\mathcal{F}_t$-adapted process $X(t)$ with respect to an $\mathcal{F}_t$-Brownian motion $W(t)$. We start by assuming $X(t)$ has left-continuous sample paths; that is, $X(s_j,\omega) \to X(s,\omega)$ if $s_j \to s$ with $s_j \leq s$ for all $s \in [0,T]$ and $\omega \in \Omega$.

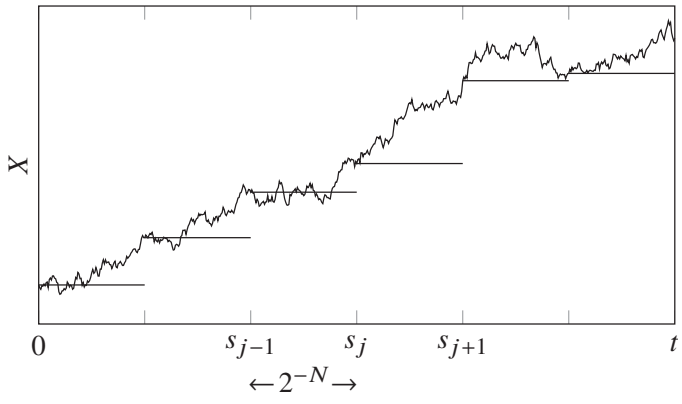

Figure 8.4 We approximate $X(t)$ on the interval $[s_j, s_{j+1})$ by $X(s_j)$, where $s_j = j2^{-N}$ for $j = 0,1,\ldots$. The interval $[0,t]$ is partitioned into $0 = s_0 < s_1 < s_2 < \cdots < s_k < t$, where $s_k = \max\{s_j : s_j < t\}$.

The idea is to introduce a grid of points $s_j = j\Delta t$ for $j = 0,1,2,\ldots$, where $\Delta t = 2^{-N}$, and approximate $X(s)$ for $s \in [s_j, s_{j+1})$ by $X(s_j)$ (see Figure 8.4) and $dW(s)$ by the increment $W(s_{j+1}) - W(s_j)$. We then define the Itô stochastic integral as

$$\int_0^t X(s)\, dW(s) := \lim_{N\to\infty} S_N^X(t), \tag{8.13}$$

where we refine the grid by taking $N \to \infty$ and

$$S_N^X(t) := \sum_{s_j < t} X(s_j)\big(W(s_{j+1} \wedge t) - W(s_j)\big). \tag{8.14}$$

All but the last increment in the sum are over an interval of length $\Delta t = 2^{-N}$ (which gives a nesting property convenient for the proof of Lemma 8.9).

Surprisingly, the limit in (8.13) and hence the stochastic integral depends on the choice of $X(s_j)$ to approximate $X(t)$ on $t \in [s_j, s_{j+1})$. For example, in the case that $X(t)$ is $\mathcal{F}_t$-adapted, $X(s_j)$ is $\mathcal{F}_{s_j}$-measurable and independent of the increment $W(s_{j+1} \wedge t) - W(s_j)$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

of an $\mathcal{F}_t$-Brownian motion $W(t)$. Hence, the product $X(s_j)(W(s_{j+1} \wedge t) - W(s_j))$ has mean zero. Moreover, the sum $S_N^X(t)$ and the Itô integral have mean zero. This is not true if $S_N^X(t)$ equals a sum of $X(r_j)(W(s_{j+1} \wedge t) - W(s_j))$ for some $r_j > s_j$.

The Itô integral is defined by the limit of $S_N^X(t)$ in $L^2(\Omega)$, which exists by the following lemma.

Lemma 8.9 *Let $W(t)$ be an $\mathcal{F}_t$-Brownian motion and let $\{X(t)\colon t \in [0,T]\}$ be an $\mathcal{F}_t$-adapted process with left-continuous sample paths.*

(i) *$\{S_N^X(t)\colon t \in [0,T]\}$ is $\mathcal{F}_t$-adapted and has continuous sample paths.*

(ii) *$\mathbb{E}[S_N^X(t) \,|\, \mathcal{F}_{s_j}] = S_N^X(s_j)$ for $0 \le s_j \le t$ a.s. In particular, $\mathbb{E}[S_N^X(t)] = 0$.*

(iii) *If $\{Y(t)\colon t \in [0,T]\}$ is also $\mathcal{F}_t$-adapted with left-continuous sample paths and $X, Y \in L^2(\Omega, L^2(0,T))$, then for $t_1, t_2 \in [0,T]$*

$$\mathbb{E}\left[S_N^X(t_1)\, S_M^Y(t_2)\right] \to \int_0^{t_1 \wedge t_2} \mathbb{E}[X(s)Y(s)]\, ds \qquad \text{as } N, M \to \infty. \tag{8.15}$$

In particular, $S_N^X(t)$ has an $L^2(\Omega)$ limit as $N \to \infty$.

Proof (i) Since $\mathcal{F}_{s_j}$ is a sub σ-algebra of $\mathcal{F}_t$ and $X(t)$ is $\mathcal{F}_t$-adapted, it holds that $X(s_j)$ is $\mathcal{F}_t$-measurable for $s_j \le t$. Similarly, $W(s)$ is $\mathcal{F}_t$-measurable for $s \le t$. Then, $S_N^X(t)$ is $\mathcal{F}_t$-measurable and the process $\{S_N^X(t)\colon t \in [0,T]\}$ is $\mathcal{F}_t$-adapted. Sample paths of $S_N^X(t)$ are continuous because $W(t)$ has continuous sample paths.

(ii) We assume for simplicity that t is integer valued, in which case

$$S_N^X(t) = \sum_{s_j < t} X(s_j)\big(W(s_{j+1}) - W(s_j)\big).$$

Now, $S_N^X(t) = S_N^X(s_k) + \big(S_N^X(t) - S_N^X(s_k)\big)$ and

$$\begin{aligned}\mathbb{E}\left[S_N^X(t) \,\middle|\, \mathcal{F}_{s_k}\right] &= \mathbb{E}\left[S_N^X(s_k) \,\middle|\, \mathcal{F}_{s_k}\right] + \sum_{s_k \le s_j < t} \mathbb{E}\left[X(s_j)\big(W(s_{j+1}) - W(s_j)\big) \,\middle|\, \mathcal{F}_{s_k}\right] \\ &= S_N^X(s_k) + \sum_{s_k \le s_j < t} \mathbb{E}\left[X(s_j)\, \mathbb{E}\left[\big(W(s_{j+1}) - W(s_j)\big) \,\middle|\, \mathcal{F}_{s_j}\right] \,\middle|\, \mathcal{F}_{s_k}\right],\end{aligned}$$

where we use the 'taking out what is known' and 'tower property' (see Theorem 4.52) of conditional expectation. Finally,

$$\mathbb{E}\left[S_N^X(t) \,\middle|\, \mathcal{F}_{s_k}\right] = S_N^X(s_k), \qquad \text{a.s.}$$

because the increment $W(s_{j+1}) - W(s_j)$ is independent of $\mathcal{F}_{s_j}$ and hence

$$\mathbb{E}\left[\big(W(s_{j+1}) - W(s_j)\big) \,\middle|\, \mathcal{F}_{s_j}\right] = \mathbb{E}[W(s_{j+1}) - W(s_j)] = 0, \qquad a.s.$$

(iii) Let $s_j = j2^{-N}$ and $r_k = k2^{-M}$ for $j, k = 0, 1, \ldots$ and for simplicity assume that t_1, t_2 are

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

integers. Write $X_j = X(s_j)$ and $Y_k = Y(r_k)$. Now,

$$\begin{aligned}\mathbb{E}\big[S_N^X(t_1)S_M^Y(t_2)\big] &= \mathbb{E}\bigg[\sum_{s_j<t_1}\sum_{r_k<t_2} X_jY_k\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\bigg]\\ &= \mathbb{E}\bigg[\mathbb{E}\bigg[\sum_{0\le r_k<s_j<t_1} X_jY_k\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\,\Big|\,\mathcal{F}_{s_j}\bigg]\bigg]\\ &\quad+\mathbb{E}\bigg[\mathbb{E}\bigg[\sum_{0\le s_j\le r_k<t_2} X_jY_k\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\,\Big|\,\mathcal{F}_{r_k}\bigg]\bigg].\end{aligned}$$

By the increment property of Definition 8.7, we have (see Exercise 8.2) almost surely

$$\mathbb{E}\Big[\big(W(r)-W(p)\big)\big(W(t)-W(s)\big)\,\Big|\,\mathcal{F}_s\Big] = \begin{cases} r-s, & p\le s\le r\le t,\\ 0, & p<r\le s<t.\end{cases} \tag{8.16}$$

We assume that $M > N$, so that $[r_k, r_{k+1})$ is either a subset of or has empty intersection with $[s_j, s_{j+1})$. For $s_j \le r_k$, both X_j and Y_k are $\mathcal{F}_{r_k}$-measurable and almost surely

$$\begin{aligned}&\mathbb{E}\Big[X_jY_k\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\,\Big|\,\mathcal{F}_{r_k}\Big]\\ &\quad= X_jY_k\mathbb{E}\Big[\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\,\Big|\,\mathcal{F}_{r_k}\Big]\\ &\quad= \begin{cases}0, & s_{j+1}\le r_k,\\ X_jY_k(r_{k+1}-r_k), & s_j\le r_k<s_{j+1}.\end{cases}\end{aligned}$$

Similarly, for $r_k < s_j$, we have $r_{k+1} \le s_j$ and

$$\begin{aligned}&\mathbb{E}\Big[X_jY_k\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\,\Big|\,\mathcal{F}_{s_j}\Big]\\ &\quad= X_jY_k\mathbb{E}\Big[\big(W(s_{j+1})-W(s_j)\big)\big(W(r_{k+1})-W(r_k)\big)\,\Big|\,\mathcal{F}_{s_j}\Big] = 0, \qquad \text{a.s.}\end{aligned}$$

Let $\phi(r) := X_jY_k$ if $r \in [s_j, s_{j+1}) \cap [r_k, r_{k+1})$. By left continuity of the sample paths, $\phi(r) = X(s_j)Y(r_k) \to X(r)Y(r)$ as $N, M \to \infty$ with $M > N$. Then,

$$\mathbb{E}\Big[S_N^X(t_1)\,S_M^Y(t_2)\Big] = \sum_{0\le r_k<t_1\wedge t_2}\mathbb{E}\big[\phi(r_k)\big](r_{k+1}-r_k),$$

which converges to $\int_0^{t_1\wedge t_2}\mathbb{E}[X(r)Y(r)]\,dr$ as $N, M \to \infty$ with $M > N$ and hence we have (8.15). If $X \in L^2(\Omega, L^2(0,T))$, then

$$\mathbb{E}\Big[S_N^X(t)S_M^X(t)\Big] \to \|X\|^2_{L^2(\Omega,L^2(0,t))} < \infty$$

and, by Lemma 5.34, $S_N^X(t)$ has a well-defined limit in $L^2(\Omega)$. □

We have shown the Itô integral $I(t) = \int_0^t X(s)\,dW(s)$ is well defined by the $L^2(\Omega)$ limit of $S_N^X(t)$ as in (8.13), for adapted processes $X(t)$ with left-continuous sample paths. Further from (8.15), we have the so-called *Itô isometry*:

$$\bigg\|\int_0^t X(s)\,dW(s)\bigg\|^2_{L^2(\Omega)} = \mathbb{E}\bigg[\bigg|\int_0^t X(s)\,dW(s)\bigg|^2\bigg] = \int_0^t \mathbb{E}\big[|X(s)|^2\big]\,ds, \tag{8.17}$$

In order to conveniently apply limit theorems in Banach spaces, it is important to extend the definition to the predictable processes.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Definition 8.10 (predictable) A stochastic process $\{X(t)\colon t \in [0,T]\}$ is *predictable* if there exists $\mathcal{F}_t$-adapted and left-continuous processes $\{X_n(t)\colon t \in [0,T]\}$ such that $X_n(t) \to X(t)$ as $n \to \infty$ for $t \in [0,T]$.

Proposition 8.11 (Banach space $\mathcal{L}_2^T$) *Let $\mathcal{L}_2^T$ denote the space of predictable real-valued processes $\{X(t)\colon t \in [0,T]\}$ with*

$$\|X\|_{\mathcal{L}_2^T} := \mathbb{E}\left[\int_0^T |X(s)|^2\, ds\right]^{1/2} < \infty.$$

Then, $\mathcal{L}_2^T$ is a Banach space with norm $\|\cdot\|_{\mathcal{L}_2^T}$.

We define the Itô integral for any integrand in $\mathcal{L}_2^T$.

Definition 8.12 (Itô integral) Let $W(t)$ be an $\mathcal{F}_t$-Brownian motion and let $X \in \mathcal{L}_2^T$. By definition, X equals the limit in $\mathcal{L}_2^T$ of a sequence of left-continuous and $\mathcal{F}_t$-adapted processes $X_n(t)$, $n \in \mathbb{N}$. The stochastic integral $I_n(t) = \int_0^t X_n(s)\, dW(s)$ is well defined by Lemma 8.9 and, from (8.17) with $X = X_n - X_m$,

$$\|I_n(t) - I_m(t)\|_{L^2(\Omega)} = \mathbb{E}\left[|I_n(t) - I_m(t)|^2\right] = \int_0^t \mathbb{E}\left[|X_n(s) - X_m(s)|^2\right] ds. \tag{8.18}$$

Then, $I_n(t)$ is a Cauchy sequence in $L^2(\Omega)$ and the Itô integral

$$\int_0^t X(s)\, dW(s) := \lim_{n\to\infty} \int_0^t X_n(s)\, dW(s) \qquad \text{in } L^2(\Omega).$$

Definition 8.13 (vector-valued Itô integral) Let $\boldsymbol{W} = [W_1, \ldots, W_m]^\mathsf{T}$, for *iid* Brownian motions $W_j(t)$. Denote by $\mathcal{L}_2^T(\mathbb{R}^{d\times m})$ the set of $\mathbb{R}^{d\times m}$-valued processes where each component belongs to $\mathcal{L}_2^T$ and let $X \in \mathcal{L}_2^T(\mathbb{R}^{d\times m})$. The Itô integral $\int_0^t X(s)\, d\boldsymbol{W}(s)$ is the random variable in $L^2(\Omega, \mathbb{R}^d)$ with ith component

$$\sum_{j=1}^m \int_0^t X_{ij}(s)\, dW_j(s). \tag{8.19}$$

We give some useful properties of the Itô integral.

Theorem 8.14 (Itô integral process) *Let $X \in \mathcal{L}_2^T(\mathbb{R}^{d\times m})$. The following properties hold:*

(i) $\left\{\int_0^t X(s)\, d\boldsymbol{W}(s)\colon t \in [0,T]\right\}$ *is a predictable process.*

(ii) *The martingale property holds: for $0 \le r \le t \le T$,*

$$\mathbb{E}\left[\int_0^t X(s)\, d\boldsymbol{W}(s)\,\middle|\,\mathcal{F}_r\right] = \int_0^r X(s)\, d\boldsymbol{W}(s), \qquad \textit{almost surely,} \tag{8.20}$$

and in particular the integral has mean zero : $\mathbb{E}\left[\int_0^t X(s)\, d\boldsymbol{W}(s)\right] = 0$.

(iii) *For $X, Y \in \mathcal{L}_2^T(\mathbb{R}^{d\times m})$,*

$$\mathbb{E}\left[\left\langle \int_0^{t_1} X(s)\, d\boldsymbol{W}(s), \int_0^{t_2} Y(s)\, d\boldsymbol{W}(s)\right\rangle\right] = \int_0^{t_1\wedge t_2} \sum_{i=1}^m \mathbb{E}\left[\langle \boldsymbol{X}_i(s), \boldsymbol{Y}_i(s)\rangle\right] ds, \tag{8.21}$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

where $\boldsymbol{X}_i, \boldsymbol{Y}_i$ denote the columns of X and Y and $\langle \cdot, \cdot \rangle$ is the $\mathbb{R}^d$ inner product. In particular, we have the Itô isometry:

$$\mathbb{E}\left[\left\| \int_0^t X(s)\, d\boldsymbol{W}(s) \right\|_2^2\right] = \int_0^t \mathbb{E}\left[\|X(s)\|_{\mathrm{F}}^2\right] ds, \qquad t \in [0,T]. \tag{8.22}$$

Proof It is clear from (8.18) that I_n is also a Cauchy sequence in $\mathcal{L}_2^T$. As $\mathcal{L}_2^T$ is Banach space, there exists a limit $I \in \mathcal{L}_2^T$ and the limit is predictable. For $d = 1 = m$, (ii)–(iii) follow from Lemma 8.9(ii)–(iii). For general d, m, the identities follow by writing out the inner products in component forms. □

Example 8.15 Consider the stochastic integral

$$I(t) = \int_0^t W(s)\, dW(s), \qquad t \geq 0.$$

The martingale property gives $\mathbb{E}[I(t)] = 0$. The Itô isometry (8.22) gives

$$\mathbb{E}\big[I(t)^2\big] = \int_0^t \mathbb{E}\big[W(s)^2\big]\, ds = \int_0^t s\, ds = \frac{1}{2}t^2 \tag{8.23}$$

and (8.21) gives

$$\mathbb{E}\big[I(t)I(s)\big] = \int_0^{s\wedge t} \mathbb{E}\big[W(r)^2\big]\, dr = \int_0^{s\wedge t} r\, dr = \frac{1}{2}(s \wedge t)^2.$$

As Itô integrals have mean zero, $\mathrm{Cov}(I(s), I(t)) = (s \wedge t)^2/2$.

8.3 Itô SODEs

Having described the Itô integral, the next step is to examine the initial-value problem for Itô SODEs of the form (8.3). First, we study the existence and uniqueness of solutions and then, using the one-dimensional Itô formula, we obtain explicit solutions for some particular SODEs in the case $d = m = 1$.

Existence and uniqueness of solutions

To show existence and uniqueness of solutions, we use the contraction mapping theorem (Theorem 1.10) on the following Banach space.

Definition 8.16 Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space. $\mathcal{H}_{2,T}$ is the set of $\mathbb{R}^d$-valued predictable processes $\{\boldsymbol{u}(t) \colon t \in [0,T]\}$ such that

$$\|\boldsymbol{u}\|_{\mathcal{H}_{2,T}} := \sup_{t\in[0,T]} \|\boldsymbol{u}(t)\|_{L^2(\Omega,\mathbb{R}^d)} = \sup_{t\in[0,T]} \mathbb{E}\left[\|\boldsymbol{u}(t)\|_2^2\right]^{1/2} < \infty.$$

This is a Banach space with the norm $\|\boldsymbol{u}\|_{\mathcal{H}_{2,T}}$ (strictly speaking it is equivalence classes of almost sure equal processes that form the Banach space).

We require the following assumption on the drift and diffusion, similar to the condition for ODEs used in Theorem 3.2.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Assumption 8.17 (linear growth/Lipschitz condition) There exists a constant $L > 0$ such that the linear growth condition holds:

$$\begin{aligned} \|\boldsymbol{f}(\boldsymbol{u})\|_2^2 &\le L\big(1 + \|\boldsymbol{u}\|_2^2\big), \\ \|G(\boldsymbol{u})\|_{\mathrm{F}}^2 &\le L\big(1 + \|\boldsymbol{u}\|_2^2\big), \qquad \forall \boldsymbol{u} \in \mathbb{R}^d, \end{aligned} \tag{8.24}$$

and the global Lipschitz condition holds:

$$\begin{aligned} \|\boldsymbol{f}(\boldsymbol{u}_1) - \boldsymbol{f}(\boldsymbol{u}_2)\|_2 &\le L\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_2, \\ \|G(\boldsymbol{u}_1) - G(\boldsymbol{u}_2)\|_{\mathrm{F}} &\le L\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_2, \qquad \forall \boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^d. \end{aligned} \tag{8.25}$$

Theorem 8.18 (existence and uniqueness for SODEs) *Suppose that Assumption 8.17 holds and that $W(t)$ is an $\mathcal{F}_t$-Brownian motion on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. For each $T > 0$ and $\boldsymbol{u}_0 \in \mathbb{R}^d$, there exists a unique $\boldsymbol{u} \in \mathcal{H}_{2,T}$ such that for $t \in [0,T]$*

$$\boldsymbol{u}(t) = \boldsymbol{u}_0 + \int_0^t \boldsymbol{f}(\boldsymbol{u}(s))\, ds + \int_0^t G(\boldsymbol{u}(s))\, d\boldsymbol{W}(s). \tag{8.26}$$

Proof Consider a random variable $\boldsymbol{X} \in L^2(\Omega, \mathbb{R}^d)$ that is independent of $\mathcal{F}_0$ and hence of the process $\boldsymbol{W}(t)$. For $\boldsymbol{u} \in \mathcal{H}_{2,T}$, let

$$\mathcal{J}(\boldsymbol{u})(t) := \boldsymbol{X} + \int_0^t \boldsymbol{f}(\boldsymbol{u}(s))\, ds + \int_0^t G(\boldsymbol{u}(s))\, d\boldsymbol{W}(s), \qquad t \in [0,T].$$

If $\boldsymbol{u} \in \mathcal{H}_{2,T}$ is a fixed point of $\mathcal{J}$, then it satisfies the integral equation (8.26) if $\boldsymbol{X} = \boldsymbol{u}_0$. We show the existence of a unique fixed point by applying the contraction mapping theorem (Theorem 1.10) on the Banach space $\mathcal{H}_{2,T}$. To extend the result to any time interval, we may reapply the argument with an initial condition $\boldsymbol{X} = \boldsymbol{u}(kT)$ on an interval $[kT, kT + T]$ for $k = 1, 2, \ldots$.

We demonstrate the two conditions of the contraction mapping theorem and first show $\mathcal{J}$ maps into $\mathcal{H}_{2,T}$. The process $\mathcal{J}(\boldsymbol{u})(t)$ is predictable, because of Theorem 8.14(i). It remains to show the $\mathcal{H}_{2,T}$ norm of $\mathcal{J}(\boldsymbol{u})(t)$ is finite. By Jensen's inequality (A.9),

$$\mathbb{E}\left[\left\|\int_0^t \boldsymbol{f}(\boldsymbol{u}(s))\, ds\right\|_2^2\right] \le \mathbb{E}\left[t \int_0^t \|\boldsymbol{f}(\boldsymbol{u}(s))\|_2^2\, ds\right] \tag{8.27}$$

and, by the Itô isometry (8.22),

$$\mathbb{E}\left[\left\|\int_0^t G(\boldsymbol{u}(s))\, d\boldsymbol{W}(s)\right\|_2^2\right] = \int_0^t \mathbb{E}\left[\|G(\boldsymbol{u}(s))\|_{\mathrm{F}}^2\right] ds. \tag{8.28}$$

Using (A.10), (8.27), and (8.28),

$$\begin{aligned} \mathbb{E}\left[\|\mathcal{J}(\boldsymbol{u})(t)\|_2^2\right] &\le 3\mathbb{E}\left[\|\boldsymbol{X}\|_2^2\right] + 3\mathbb{E}\left[\left\|\int_0^t \boldsymbol{f}(\boldsymbol{u}(s))\, ds\right\|_2^2\right] + 3\mathbb{E}\left[\left\|\int_0^t G(\boldsymbol{u}(s))\, d\boldsymbol{W}(s)\, ds\right\|_2^2\right] \\ &\le 3\mathbb{E}\left[\|\boldsymbol{X}\|_2^2\right] + 3\mathbb{E}\left[t \int_0^t \|\boldsymbol{f}(\boldsymbol{u}(s))\|_2^2\, ds\right] + 3\int_0^t \mathbb{E}\left[\|G(\boldsymbol{u}(s))\|_{\mathrm{F}}^2\right] ds. \end{aligned}$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Assumption 8.17 gives

$$\mathbb{E}\left[\|\mathcal{J}(\boldsymbol{u})(t)\|_2^2\right] \le 3\mathbb{E}\left[\|\boldsymbol{X}\|_2^2\right] + 3t\int_0^t L^2\mathbb{E}\left[1+\|\boldsymbol{u}(s)\|_2^2\right]ds + 3\int_0^t L^2\mathbb{E}\left[1+\|\boldsymbol{u}(s)\|_2^2\right]ds.$$

Finally, we take the supremum over $t \in [0,T]$ in the last two terms:

$$\mathbb{E}\left[\|\mathcal{J}(\boldsymbol{u})(t)\|_2^2\right] \le 3\mathbb{E}\left[\|\boldsymbol{X}\|_2^2\right] + 3L^2(t+1)t\left(1+\sup_{t\in[0,T]}\mathbb{E}\left[\|\boldsymbol{u}(t)\|_2^2\right]\right). \tag{8.29}$$

Hence, $\|\mathcal{J}(\boldsymbol{u})\|_{\mathcal{H}_{2,T}} < \infty$ and $\mathcal{J}$ maps $\mathcal{H}_{2,T}$ into $\mathcal{H}_{2,T}$.

To show $\mathcal{J}$ is a contraction, we argue similarly using the Lipschitz condition in place of the linear growth condition:

$$\begin{aligned}
&\mathbb{E}\left[\|\mathcal{J}(\boldsymbol{u}_1)(t)-\mathcal{J}(\boldsymbol{u}_2)(t)\|_2^2\right]\\
&\quad\le 2t\,\mathbb{E}\left[\int_0^t \|\boldsymbol{f}(\boldsymbol{u}_1(s))-\boldsymbol{f}(\boldsymbol{u}_2(s))\|_2^2\,ds\right] + 2\mathbb{E}\left[\left\|\int_0^t \left(G(\boldsymbol{u}_1(s))-G(\boldsymbol{u}_2(s))\right)d\boldsymbol{W}(s)\right\|_2^2\right]\\
&\quad\le 2L^2t\int_0^t \mathbb{E}\left[\|\boldsymbol{u}_1(s)-\boldsymbol{u}_2(s)\|_2^2\right]ds + 2L^2\int_0^t \mathbb{E}\left[\|\boldsymbol{u}_1(s)-\boldsymbol{u}_2(s)\|_2^2\right]ds\\
&\quad\le 2L^2t(t+1)\sup_{t\in[0,T]}\mathbb{E}\left[\|\boldsymbol{u}_1(t)-\boldsymbol{u}_2(t)\|_2^2\right].
\end{aligned}$$

Choose $T = \min\{1, 1/8L^2\}$. Then,

$$\sup_{t\in[0,T]}\mathbb{E}\left[\|\mathcal{J}(\boldsymbol{u}_1)(t)-\mathcal{J}(\boldsymbol{u}_2)(t)\|_2^2\right] \le \frac{1}{2}\sup_{t\in[0,T]}\mathbb{E}\left[\|\boldsymbol{u}_1(t)-\boldsymbol{u}_2(t)\|_2^2\right]$$

and $\|\mathcal{J}(\boldsymbol{u}_1)-\mathcal{J}(\boldsymbol{u}_2)\|_{\mathcal{H}_{2,T}} \le \frac{1}{2}\|\boldsymbol{u}_1-\boldsymbol{u}_2\|_{\mathcal{H}_{2,T}}$. We see that $\mathcal{J}$ is a contraction on $\mathcal{H}_{2,T}$ and the contraction mapping theorem applies. Hence, there exists a unique fixed point of $\mathcal{J}$ and unique solution to (8.26). □

One-dimensional Itô formula and exact solutions

The chain rule of ordinary calculus changes in the Itô calculus to the so-called Itô formula. We discuss the Itô formula in one dimension, giving applications to geometric Brownian motion and the mean reverting Ornstein–Uhlenbeck process. The general Itô formula is given later in Lemma 8.42.

Lemma 8.19 (one-dimensional Itô formula) *Let $\Phi\colon [0,T]\times\mathbb{R}\to\mathbb{R}$ have continuous partial derivatives $\frac{\partial\Phi}{\partial t}$, $\frac{\partial\Phi}{\partial u}$ and $\frac{\partial^2\Phi}{\partial u^2}$. Let u satisfy*

$$du = f(u)\,dt + g(u)\,dW(t), \qquad u(0) = u_0$$

and suppose Assumption 8.17 holds for $d = m = 1$. Then, almost surely,

$$d\Phi = \frac{\partial\Phi}{\partial t}\,dt + \frac{\partial\Phi}{\partial u}\,du + \frac{1}{2}\frac{\partial^2\Phi}{\partial u^2}g^2\,dt \tag{8.30}$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

or written in full

$$\Phi(t,u(t)) = \Phi(0,u_0) + \int_0^t \frac{\partial\Phi}{\partial t}(s,u(s)) + \frac{\partial\Phi}{\partial u}(s,u(s))f(u(s))\,ds + \frac{1}{2}\int_0^t \frac{\partial^2\Phi}{\partial u^2}(s,u(s))g(u(s))^2\,ds + \int_0^t \frac{\partial\Phi}{\partial u}(s,u(s))g(u(s))\,dW(s).$$

Proof We consider the special case $\Phi(t,u) = \phi(u)$ for a quadratic polynomial $\phi(u)$, so that (8.30) reduces to

$$d\phi(u(t)) = \phi'(u(t))\,du + \frac{1}{2}\phi''(u(t))g^2\,dt. \tag{8.31}$$

We also take $f, g \in \mathbb{R}$ to be constants, so that $u(t) = u_0 + ft + gW(t)$.

Let $s_j = j\Delta t$, where $\Delta t = 2^{-N}$ and $j = 0,1,2,\ldots$. Consider the telescoping sum

$$\phi(u(t)) - \phi(u(0)) = \sum_{s_j<t} \big(\phi(u(s_{j+1}\wedge t)) - \phi(u(s_j))\big).$$

We assume t is an integer so that $s_{j+1}\wedge t = s_{j+1}$. As ϕ is quadratic, the second-order Taylor series of $\phi(u)$ is exact and

$$\phi(u(s_{j+1})) = \phi(u(s_j)) + \phi'(u(s_j))\big(u(s_{j+1}) - u(s_j)\big) + \tfrac{1}{2}\phi''(u(s_j))\big(u(s_{j+1}) - u(s_j)\big)^2.$$

Then, $\phi(u(t)) - \phi(u(0)) = A_N + B_N$ for

$$\begin{aligned} A_N &:= \sum_{s_j<t} \phi'(u(s_j))\big(u(s_{j+1}) - u(s_j)\big), \\ B_N &:= \frac{1}{2}\sum_{s_j<t} \phi''(u(s_j))\big(u(s_{j+1}) - u(s_j)\big)^2. \end{aligned} \tag{8.32}$$

Substituting $u(t) - u(s) = (f\,t - f\,s) + (g\,W(t) - g\,W(s))$,

$$\begin{aligned} A_N &= \sum_{s_j<t} \phi'(u(s_j))\big(f\,s_{j+1} - f\,s_j\big) + \sum_{s_j<t} \phi'(u(s_j))\big(g\,W(s_{j+1}) - g\,W(s_j)\big) \\ &\to \int_0^t \phi'(u(s))\,f\,ds + \int_0^t \phi'(u(s))\,g\,dW(s) \qquad \text{in } L^2(\Omega) \text{ as } N\to\infty \end{aligned}$$

by definition of the Itô integral. As $\phi''(u)$ is constant, from (8.32),

$$B_N = \frac{1}{2}\phi'' \sum_{s_j<t} \big(u(s_{j+1}) - u(s_j)\big)^2 = \frac{1}{2}\phi'' \sum_{s_j<t} C_j\Delta t$$

for $C_j := (u(s_{j+1}) - u(s_j))^2/\Delta t$. Note that $C_j = (f\,s_{j+1} - f\,s_j + g\,W(s_{j+1}) - g\,W(s_j))^2/\Delta t$, which are *iid* random variables whose variance is bounded uniformly over Δt. Then,

$$\mathrm{Var}(B_N) = \frac{(\phi'')^2}{4}\sum_{s_j<t} \mathrm{Var}(C_j)\Delta t^2 \to 0 \qquad \text{as } N\to\infty.$$

Hence, $B_N \to \mathbb{E}[B_N]$ in $L^2(\Omega)$ as $N\to\infty$. Now, $\mathbb{E}[C_j] = f^2\Delta t + g^2$ and

$$\mathbb{E}\big[B_N\big] = \frac{\phi''}{2}\sum_{s_j<t} f^2\Delta t^2 + g^2\Delta t \to \frac{1}{2}\int_0^t \phi''(u(s))g\,ds, \qquad \text{as } N\to\infty.$$

Finally, $\phi(u(t)) - \phi(u(0)) = \lim_{N\to\infty}(A_N + B_N)$ and this gives (8.31). □

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Consider the solution $u(t)$ of the deterministic ODE $\frac{du}{dt} = f$ and let $\phi(u) = u^2$. The standard chain rule implies that

$$\frac{d\phi(u)}{dt} = 2u\,\frac{du}{dt}. \tag{8.33}$$

However, if $u(t)$ satisfies $du = f\,dt + g\,dW(t)$, the Itô formula (8.30) says that

$$d\phi(u) = \big(2uf\,dt + 2ug\,dW(t)\big) + g^2\,dt = 2u\,du + g^2\,dt$$

and we pick up an unexpected extra term $g^2\,dt$. Since the Brownian increment $W(s_{j+1}) - W(s_j)$ is of size $|s_{j+1} - s_j|^{1/2}$, the terms arising from $(u(s_{j+1}) - u(s_j))^2$ (in B_N in the proof) that lead to $\mathbb{E}\big[(W(s_{j+1}) - W(s_j))^2\big]$ are important and the familiar rule $du^2 = 2u\,du$ does not apply in the Itô calculus.

We look at two applications of the Itô formula to solve SODEs.

Example 8.20 (geometric Brownian motion) We show the solution of the geometric Brownian motion SODE (8.11) is given by

$$u(t) = \exp\big((r - \sigma^2/2)\,t + \sigma W(t)\big)u_0. \tag{8.34}$$

For $u(0) = 0$, $u(t) = 0$ is clearly the solution to (8.11). For $u > 0$, let $\phi(u) = \log u$, so that $\phi'(u) = u^{-1}$ and $\phi''(u) = -u^{-2}$. By the Itô formula (8.31) with $\Phi(t,u) = \phi(u)$,

$$d\phi(u) = r\,dt + \sigma\,dW(t) - \frac{1}{2}\sigma^2\,dt.$$

Hence,

$$\phi(u(t)) = \phi(u_0) + \int_0^t \left(r - \frac{1}{2}\sigma^2\right) ds + \int_0^t \sigma\,dW(s)$$

and $\log u(t) = \log(u_0) + (r - \frac{1}{2}\sigma^2)t + \sigma W(t)$. Taking the exponential, we find (8.34). It is clear that, when $u_0 \geq 0$, the solution $u(t) \geq 0$ for all $t \geq 0$.

In the following example, we use the Itô formula (8.30) to examine a generalisation of the Ornstein–Uhlenbeck process (8.5) and obtain the variation of constants solution (8.6).

Example 8.21 (mean-reverting Ornstein–Uhlenbeck process) We consider the following generalisation of (8.5):

$$du = \lambda(\mu - u)dt + \sigma dW(t), \qquad u(0) = u_0,$$

for $\lambda, \mu, \sigma \in \mathbb{R}$. We solve this SODE by using the Itô formula (8.30) with $\Phi(t,u) = \mathrm{e}^{\lambda t}\,u$. Then,

$$d\Phi(t,u) = \lambda\,\mathrm{e}^{\lambda t}u\,dt + \mathrm{e}^{\lambda t}\big(\lambda(\mu - u)dt + \sigma dW(t)\big) + 0$$

and

$$\Phi(t,u(t)) - \Phi(0,u_0) = \mathrm{e}^{\lambda t}u(t) - u_0 = \lambda\mu \int_0^t \mathrm{e}^{\lambda s}\,ds + \sigma \int_0^t \mathrm{e}^{\lambda s}\,dW(s).$$

After evaluating the deterministic integral, we find

$$u(t) = \mathrm{e}^{-\lambda t}u_0 + \mu\big(1 - \mathrm{e}^{-\lambda t}\big) + \sigma \int_0^t \mathrm{e}^{\lambda(s-t)}dW(s)$$

and this is known as the *variation of constants* solution.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Notice that $u(t)$ is a Gaussian process and can also be specified (see Corollary 5.19) by its mean $\mu(t) = \mathbb{E}[u(t)]$ and covariance $C(s,t) = \mathrm{Cov}(u(s),u(t))$. Using the mean-zero property of the Itô integral (Theorem 8.14), the mean

$$\mu(t) = \mathbb{E}\big[u(t)\big] = \mathrm{e}^{-\lambda t}u(0) + \mu\big(1 - \mathrm{e}^{-\lambda t}\big)$$

so that $\mu(t) \to \mu$ as $t \to \infty$ and the process is mean reverting. For the covariance,

$$\begin{aligned} C(s,t) = \mathrm{Cov}(u(t),u(s)) &= \mathbb{E}\Big[\big(u(s) - \mathbb{E}\big[u(s)\big]\big)\big(u(t) - \mathbb{E}\big[u(t)\big]\big)\Big] \\ &= \mathbb{E}\left[\int_0^s \sigma \mathrm{e}^{\lambda(r-s)}\,dW(r) \int_0^t \sigma \mathrm{e}^{\lambda(r-t)}\,dW(r)\right] \\ &= \sigma^2 \mathrm{e}^{-\lambda(s+t)}\mathbb{E}\left[\int_0^s \mathrm{e}^{\lambda r}\,dW(r) \int_0^t \mathrm{e}^{\lambda r}\,dW(r)\right]. \end{aligned}$$

Then, using the Itô isometry (8.21),

$$C(s,t) = \frac{\sigma^2}{2\lambda}\mathrm{e}^{-\lambda(s+t)}\big(\mathrm{e}^{2\lambda(s\wedge t)} - 1\big).$$

In particular, the variance $\mathrm{Var}(u(t)) = \sigma^2(1 - \mathrm{e}^{-2\lambda t})/2\lambda$ so that $\mathrm{Var}(u(t)) \to \sigma^2/2\lambda$ and $u(t) \to \mathrm{N}(\mu, \sigma^2/2\lambda)$ in distribution as $t \to \infty$.

8.4 Numerical methods for Itô SODEs

The geometric Brownian motion SODE (8.11) is exceptional in that the solution (8.34) is an explicit function of Brownian motion. In general, for nonlinear drift or diffusion functions, the explicit solution is not available and numerical techniques are necessary. The solution $\{\boldsymbol{u}(t)\colon t \geq 0\}$ of an SODE is a stochastic process. In this section, we examine the numerical approximation of $\boldsymbol{u}(t_n)$ by random variables $\boldsymbol{u}_n$ where $t_n = n\Delta t$, $n = 0,1,2,\ldots$.

One simple possibility is the Euler–Maruyama method and we sketch its derivation. From (8.3),

$$\boldsymbol{u}(t_{n+1}) = \boldsymbol{u}_0 + \int_0^{t_{n+1}} \boldsymbol{f}(\boldsymbol{u}(s))\,ds + \int_0^{t_{n+1}} G(\boldsymbol{u}(s))\,d\boldsymbol{W}(s),$$

and

$$\boldsymbol{u}(t_n) = \boldsymbol{u}_0 + \int_0^{t_n} \boldsymbol{f}(\boldsymbol{u}(s))\,ds + \int_0^{t_n} G(\boldsymbol{u}(s))\,d\boldsymbol{W}(s).$$

Subtracting, we see

$$\boldsymbol{u}(t_{n+1}) = \boldsymbol{u}(t_n) + \int_{t_n}^{t_{n+1}} \boldsymbol{f}(\boldsymbol{u}(s))\,ds + \int_{t_n}^{t_{n+1}} G(\boldsymbol{u}(s))\,d\boldsymbol{W}(s).$$

Taking both $\boldsymbol{f}$ and G constant over $[t_n,t_{n+1})$, we obtain the Euler–Maruyama method

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \boldsymbol{f}(\boldsymbol{u}_n)\,\Delta t + G(\boldsymbol{u}_n)\,\Delta \boldsymbol{W}_n, \qquad \Delta \boldsymbol{W}_n := \boldsymbol{W}(t_{n+1}) - \boldsymbol{W}(t_n). \tag{8.35}$$

Since on each subinterval G is evaluated at the left-hand end point, (8.35) is consistent with our definition of the Itô integral (8.13).

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

We derive the Euler–Maruyama method as well as a higher-order method, known as the Milstein method, using Taylor series. In preparation for proving convergence of these methods in §8.5, we carefully identify the remainder terms. Assume that the drift $\boldsymbol{f} \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$. Taylor's theorem (Theorem A.1) gives

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{u}(r)) = \boldsymbol{f}(\boldsymbol{u}(s)) + D\boldsymbol{f}(\boldsymbol{u}(s))\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big) \\ + \int_0^1 (1-h) D^2\boldsymbol{f}\big(\boldsymbol{u}(s) + h(\boldsymbol{u}(r) - \boldsymbol{u}(s))\big)\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big)^2\, dh,\end{aligned}$$

which we may write as

$$\boldsymbol{f}(\boldsymbol{u}(r)) = \boldsymbol{f}(\boldsymbol{u}(s)) + \boldsymbol{R}_f(r; s, \boldsymbol{u}(s)), \qquad 0 \le s \le r,$$

for a remainder $\boldsymbol{R}_f \in \mathbb{R}^d$ given by

$$\begin{aligned}\boldsymbol{R}_f(r; s, \boldsymbol{u}(s)) \coloneqq D\boldsymbol{f}(\boldsymbol{u}(s))\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big) \\ + \int_0^1 (1-h) D^2\boldsymbol{f}\big(\boldsymbol{u}(s) + h(\boldsymbol{u}(r) - \boldsymbol{u}(s))\big)\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big)^2\, dh.\end{aligned} \tag{8.36}$$

Similarly, if the diffusion $G \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$,

$$G(\boldsymbol{u}(r)) = G(\boldsymbol{u}(s)) + DG(\boldsymbol{u}(s))\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big) + R_G(r; s, \boldsymbol{u}(s)), \tag{8.37}$$

where we note that $DG(\boldsymbol{u}(s)) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^{d\times m})$ and the remainder R_G is a $d \times m$ matrix defined by

$$R_G(r; s, \boldsymbol{u}(s)) \coloneqq \int_0^1 (1-h) D^2 G\big(\boldsymbol{u}(s) + h(\boldsymbol{u}(r) - \boldsymbol{u}(s))\big)\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big)^2\, dh. \tag{8.38}$$

Substitute the Taylor expansions above for $\boldsymbol{f}$ and G into the integral equation

$$\boldsymbol{u}(t) = \boldsymbol{u}(s) + \int_s^t \boldsymbol{f}(\boldsymbol{u}(r))\, dr + \int_s^t G(\boldsymbol{u}(r))\, d\boldsymbol{W}(r), \tag{8.39}$$

to find

$$\boldsymbol{u}(t) = \boldsymbol{u}(s) + \boldsymbol{f}(\boldsymbol{u}(s))(t-s) + G(\boldsymbol{u}(s)) \int_s^t d\boldsymbol{W}(r) + \boldsymbol{R}_E(t; s, \boldsymbol{u}(s)), \tag{8.40}$$

where the remainder

$$\begin{aligned}\boldsymbol{R}_E(t; s, \boldsymbol{u}(s)) \coloneqq \int_s^t \boldsymbol{R}_f(r; s, \boldsymbol{u}(s))\, dr + \int_s^t R_G(r; s, \boldsymbol{u}(s))\, d\boldsymbol{W}(r) \\ + \int_s^t DG(\boldsymbol{u}(s))\big(\boldsymbol{u}(r) - \boldsymbol{u}(s)\big)\, d\boldsymbol{W}(r).\end{aligned} \tag{8.41}$$

The Euler–Maruyama method is found by dropping the remainder from (8.40).

Definition 8.22 (Euler–Maruyama method) For a time step $\Delta t > 0$ and initial condition $\boldsymbol{u}_0 \in \mathbb{R}^d$, the Euler–Maruyama approximation $\boldsymbol{u}_n$ to the solution $\boldsymbol{u}(t_n)$ of (8.3) for $t_n = n\Delta t$ is defined by

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \boldsymbol{f}(\boldsymbol{u}_n)\Delta t + G(\boldsymbol{u}_n)\,\Delta \boldsymbol{W}_n, \qquad \Delta \boldsymbol{W}_n \coloneqq \int_{t_n}^{t_{n+1}} d\boldsymbol{W}(r) \tag{8.42}$$

Note also that $\Delta \boldsymbol{W}_n = \boldsymbol{W}(t_{n+1}) - \boldsymbol{W}(t_n)$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

The increments $\Delta \boldsymbol{W}_n \sim \mathrm{N}(\boldsymbol{0}, \Delta t I_m)$ *iid* and are easy to sample and the update rule (8.42) is simple to apply. The approximations $\boldsymbol{u}_n$ are random variables and, in computations, we typically generate a sample path $\boldsymbol{u}_n(\omega)$ for $n = 0, 1, \ldots, N$ that approximates a sample path $\boldsymbol{u}(\cdot, \omega)$ of the solution process $\boldsymbol{u}(t)$. Algorithm 8.1 generates sample paths $\boldsymbol{u}_n(\omega)$.

Algorithm 8.1 Code to find a sample path of the Euler–Maruyama approximation $\boldsymbol{u}_n$ for (8.3). The inputs are the initial data `u0`, final time `T`, the number of subintervals `N`, dimensions `d` and `m`, and handles `fhandle, ghandle` to the drift and diffusion functions. The outputs are a vector $[0, \Delta t, \ldots, T]^\mathsf{T}$ of times `t` and a matrix `u` with columns $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_N$.

```
function [t, u]=EulerMaruyama(u0,T,N,d,m,fhandle,ghandle)
Dt=T/N; u=zeros(d,N+1); t=[0:Dt:T]'; sqrtDt=sqrt(Dt);
u(:,1)=u0; u_n=u0; % initial data
for n=1:N, % time loop
    dW=sqrtDt*randn(m,1); % Brownian increment
    u_new=u_n+Dt*fhandle(u_n)+ghandle(u_n)*dW;
    u(:,n+1)=u_new; u_n=u_new;
end
```

Example 8.23 (Duffing–van der Pol) The Duffing–van der Pol SODE (8.10) has drift and diffusion

$$\boldsymbol{f}(\boldsymbol{u}) = \begin{pmatrix} P \\ -P(\lambda + Q^2) + \alpha Q - Q^3 \end{pmatrix}, \qquad G(\boldsymbol{u}) = \begin{pmatrix} 0 \\ \sigma Q \end{pmatrix}$$

where $\boldsymbol{u} = [u_1, u_2]^\mathsf{T} = [Q, P]^\mathsf{T}$, for parameters λ, α, and σ. In the case $\lambda = \alpha = \sigma = 1$ and $\boldsymbol{u}_0 = [1/2, 0]^\mathsf{T}$, we find approximate sample paths using the Euler–Maruyama method with $\Delta t = 0.01$ on the time interval $[0, 10]$ by calling Algorithm 8.2 with the following commands:

```
>> lambda=1; alpha=1; sigma=1;
>> u0=[0.5;0]; T=10; N=1000;
>> [t, u]=vdp(u0, T, N, alpha, lambda, sigma);
```

Figure 8.5 gives plots of the resulting approximation. Exercise 8.9 investigates the Duffing–van der Pol equation as the parameters α and λ are varied.

Algorithm 8.2 Code to find a sample path of the Euler–Maruyama approximation to the Duffing–van der Pol SODE (8.10). The inputs are the initial data `u0`, final time `T`, number of time steps `N`, and parameters `alpha`, `lambda`, `sigma` for α, λ, σ. The outputs are as in Algorithm 8.1.

```
function [t,u]=vdp(u0,T,N,alpha,lambda,sigma)
[t, u]=EulerMaruyama(u0, T, N, 2, 1, @(u) vdp_f(u,lambda, alpha),...
                                     @(u) vdp_g(u,sigma));
function f=vdp_f(u, lambda, alpha) % define drift
f=[u(2); -u(2)*(lambda+u(1)^2)+alpha*u(1)-u(1)^3];
function g=vdp_g(u,sigma) % define diffusion
g=[0; sigma*u(1)];
```

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

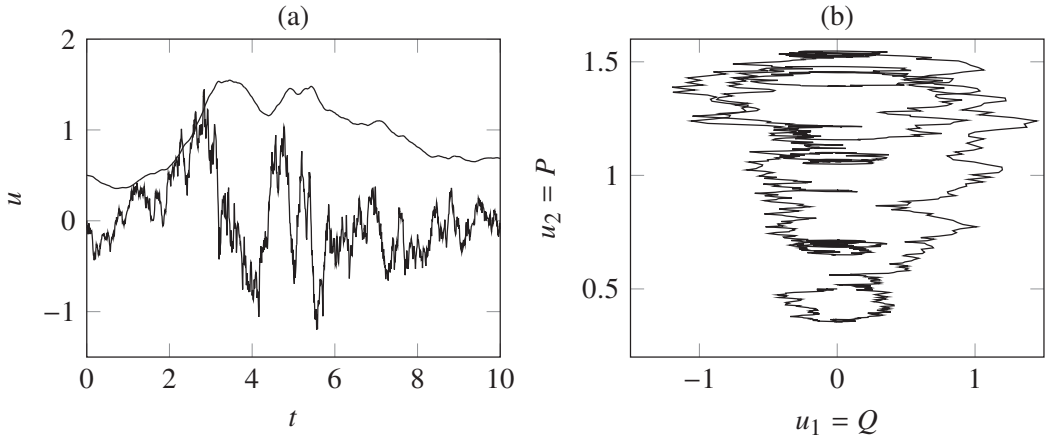

Figure 8.5 (a) A sample path of the Euler–Maruyama approximation to the Duffing–van der Pol SODE (8.10) with $\alpha = \lambda = \sigma = 1$ and initial data $\boldsymbol{u}_0 = [1/2, 0]^{\mathsf{T}}$. Note that $Q(t) = 1/2 + \int_0^t P(s)\,ds$ and is smoother that $P(t)$, which oscillates rapidly due to the Brownian motion forcing. (b) A phase-plane plot of the same sample path.

Milstein method

We develop $\boldsymbol{u}(t)$ further to derive a second numerical method, which has a smaller remainder and a better rate of convergence. From (8.40), we know

$$\boldsymbol{u}(t) = \boldsymbol{u}(s) + G(\boldsymbol{u}(s)) \int_s^t d\boldsymbol{W}(r) + \boldsymbol{R}_1(t; s, \boldsymbol{u}(s)), \tag{8.43}$$

where the remainder is

$$\boldsymbol{R}_1(t; s, \boldsymbol{u}(s)) := \boldsymbol{f}(\boldsymbol{u}(s))(t - s) + \boldsymbol{R}_E(t; s, \boldsymbol{u}(s)). \tag{8.44}$$

Using (8.43) with (8.37),

$$\begin{aligned} G(\boldsymbol{u}(r)) = G(\boldsymbol{u}(s)) &+ DG(\boldsymbol{u}(s))\left(G(\boldsymbol{u}(s)) \int_s^r d\boldsymbol{W}(p)\right) \\ &+ DG(\boldsymbol{u}(s))\boldsymbol{R}_1(t, s; \boldsymbol{u}(s)) + R_G(r; s, \boldsymbol{u}(s)) \end{aligned}$$

and then (8.39) gives

$$\begin{aligned} \boldsymbol{u}(t) = \boldsymbol{u}(s) &+ \boldsymbol{f}(\boldsymbol{u}(s))\,(t - s) + G(\boldsymbol{u}(s)) \int_s^t d\boldsymbol{W}(p) \\ &+ \int_s^t DG(\boldsymbol{u}(s))\left(G(\boldsymbol{u}(s)) \int_s^r d\boldsymbol{W}(p)\right) d\boldsymbol{W}(r) + \boldsymbol{R}_M(t; s, \boldsymbol{u}(s)), \end{aligned} \tag{8.45}$$

where the remainder $\boldsymbol{R}_M$ is given by

$$\begin{aligned} \boldsymbol{R}_M(t; s, \boldsymbol{u}(s)) := &\int_s^t \boldsymbol{R}_f(r; s, \boldsymbol{u}(s))\,dr + \int_s^t R_G(r; s, \boldsymbol{u}(s))\,d\boldsymbol{W}(r) \\ &+ \int_s^t DG(\boldsymbol{u}(s))\boldsymbol{R}_1(r; s, \boldsymbol{u}(s))\,d\boldsymbol{W}(r). \end{aligned} \tag{8.46}$$

The Milstein method is found by dropping the remainder $\boldsymbol{R}_M$ from (8.45).

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

To simplify the presentation of Milstein's method, let $I_i(s,t) \coloneqq \int_s^t dW_i(p)$ and exploit the identities

$$\int_s^t \int_s^r dW_i(p)\, dW_j(r) + \int_s^t \int_s^r dW_j(p)\, dW_i(r) = I_i(s,t)\, I_j(s,t), \qquad i \neq j,$$

and

$$\int_s^t \int_s^r dW_i(p)\, dW_i(r) = \frac{1}{2}\big(I_i(s,t)^2 - (t-s)\big).$$

See Exercises 8.4 and 8.5. Further, define

$$A_{ij}(s,t) \coloneqq \int_s^t \int_s^r dW_i(p)\, dW_j(r) - \int_s^t \int_s^r dW_j(p)\, dW_i(r). \tag{8.47}$$

These are known as *Lévy areas* and the double integrals in (8.45) are written as

$$\begin{aligned} I_i I_j + A_{ij} &= 2\int_s^t \int_s^r dW_i(p)\, dW_j(r), \\ I_i I_j - A_{ij} &= 2\int_s^t \int_s^r dW_j(p)\, dW_i(r). \end{aligned} \tag{8.48}$$

Using $\boldsymbol{u}$ to denote $\boldsymbol{u}(s)$, the fourth term in (8.45) is

$$\int_s^t DG(\boldsymbol{u})\left(G(\boldsymbol{u}) \int_s^r d\boldsymbol{W}(p)\right) d\boldsymbol{W}(r)$$

and has kth component (see Exercise 8.10)

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^m \sum_{\ell=1}^d \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{u}) g_{\ell i}(\boldsymbol{u})\big(I_i(s,t)^2 - \Delta t\big) \\ &+ \frac{1}{2}\sum_{i<j=1}^m \sum_{\ell=1}^d \left(\frac{\partial g_{kj}}{\partial u_\ell}(\boldsymbol{u}) g_{\ell i}(\boldsymbol{u}) + \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{u}) g_{\ell j}(\boldsymbol{u})\right) I_i(s,t)\, I_j(s,t) \\ &+ \frac{1}{2}\sum_{i<j=1}^m \sum_{\ell=1}^d \left(\frac{\partial g_{kj}}{\partial u_\ell}(\boldsymbol{u}) g_{\ell i}(\boldsymbol{u}) - \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{u}) g_{\ell j}(\boldsymbol{u})\right) A_{ij}(s,t). \end{aligned}$$

Definition 8.24 (Milstein method) Consider a time step $\Delta t > 0$ and initial condition $\boldsymbol{u}_0 \in \mathbb{R}^d$, the Milstein approximation $\boldsymbol{u}_n$ to $\boldsymbol{u}(t_n)$ for $t_n = n\Delta t$ is defined by

$$\begin{aligned} u_{k,n+1} = u_{kn} &+ f_k(\boldsymbol{u}_n)\,\Delta t + \sum_{j=1}^m g_{kj}(\boldsymbol{u}_n)\,\Delta W_{jn} + \frac{1}{2}\sum_{i=1}^m \sum_{\ell=1}^d \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{u}_n) g_{\ell i}(\boldsymbol{u}_n)\big(\Delta W_{in}^2 - \Delta t\big) \\ &+ \frac{1}{2}\sum_{i<j=1}^m \sum_{\ell=1}^d \left(\frac{\partial g_{kj}}{\partial u_\ell}(\boldsymbol{u}_n) g_{\ell i}(\boldsymbol{u}_n) + \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{u}_n) g_{\ell j}(\boldsymbol{u}_n)\right)\Delta W_{in}\Delta W_{jn} \\ &+ \frac{1}{2}\sum_{i<j=1}^m \sum_{\ell=1}^d \left(\frac{\partial g_{kj}}{\partial u_\ell}(\boldsymbol{u}_n) g_{\ell i}(\boldsymbol{u}_n) - \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{u}_n) g_{\ell j}(\boldsymbol{u}_n)\right) A_{ij,n}, \end{aligned}$$

where $\Delta W_{in} \coloneqq \int_{t_n}^{t_{n+1}} dW_i(r)$, $A_{ij,n} \coloneqq A_{ij}(t_n, t_{n+1})$ and u_{kn} is the kth component of $\boldsymbol{u}_n$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

If the noise is additive then the Milstein method of Definition 8.24 and the Euler–Maruyama method of Definition 8.22 are equivalent since the derivatives of g_{kj} are zero.

A simple method for sampling the Lévy areas A_{ij} is considered in Exercise 5.3. In many cases however, the A_{ij} are not needed. When the noise is diagonal (i.e., G is a diagonal matrix), each component u_k is affected by W_k only, as $g_{kj} = 0$ when $j \neq k$, and Milstein's method reduces to

$$u_{k,n+1} = u_{kn} + f_k(\boldsymbol{u}_n)\,\Delta t + g_{kk}(\boldsymbol{u}_n)\,\Delta W_{kn} + \frac{1}{2}\frac{\partial g_{kk}}{\partial u_k}(\boldsymbol{u}_n)g_{kk}(\boldsymbol{u}_n)\big(\Delta W_{kn}^2 - \Delta t\big). \quad (8.49)$$

We implement the Milstein method for SODEs with diagonal noise in Algorithm 8.3. More generally, we say the noise is commutative when $\frac{\partial g_{kj}}{\partial u_\ell} g_{\ell i} = \frac{\partial g_{ki}}{\partial u_\ell} g_{\ell j}$ and again we do not need to sample Lévy areas.

Algorithm 8.3 Code to find a sample path of the Milstein approximation to (8.3) for a diagonal diffusion matrix G ($d = m$). Inputs and outputs are similar to those in Algorithm 8.1. However, `ghandle` specifies a vector in $\mathbb{R}^d$ for the diagonal entries of G. An additional input handle is required, `dghandle`, to specify the derivatives $\frac{\partial g_{kk}}{\partial u_k}$, $k = 1, \cdots, m$. This is also given as a vector in $\mathbb{R}^d$.

```
function [t, u]=MilsteinDiag(u0,T,N,d,m,fhandle,ghandle,dghandle)
Dt=T/N; u=zeros(d,N+1); t=[0:Dt:T]'; sqrtDt=sqrt(Dt);
u(:,1)=u0; u_n=u0; % initial data
for n=1:N, % time loop
  dW=sqrtDt*randn(m,1); gu_n=ghandle(u_n);
  u_new=u_n+Dt*fhandle(u_n)+gu_n.*dW ...
            +0.5*(dghandle(u_n).*gu_n).*(dW.^2-Dt);
  u(:,n+1)=u_new; u_n=u_new;
end
```

Example 8.25 Consider the SODE for $d = m = 2$

$$d\boldsymbol{u} = A\boldsymbol{u}\,dt + B(\boldsymbol{u})\,d\boldsymbol{W}(t), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0, \quad (8.50)$$

where A and $B(\boldsymbol{u})$ are 2×2 matrices. If $B(\boldsymbol{u})$ is diagonal, then we may apply Milstein's method using Algorithm 8.3. For the case,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad B(\boldsymbol{u}) = \begin{pmatrix} u_1 & 0 \\ 0 & 2u_2 \end{pmatrix}, \quad (8.51)$$

we have two decoupled geometric Brownian motions and, instead of forming the matrix B, we simply use the diagonal. We call Algorithm 8.3 with the following commands:

```
>> d=2; m=2; A=[1 0 ;0 -1]; sigma=[1; 2];
>> T=1; N=5000; u0=[1;1];
>> [t,u]=MilsteinDiag(u0, T, N, d, m, @(u) A*u, @(u) sigma.*u, @(u) sigma);
```

Figure 8.6 shows the resulting approximate sample path alongside an approximation by the Euler–Maruyama method.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

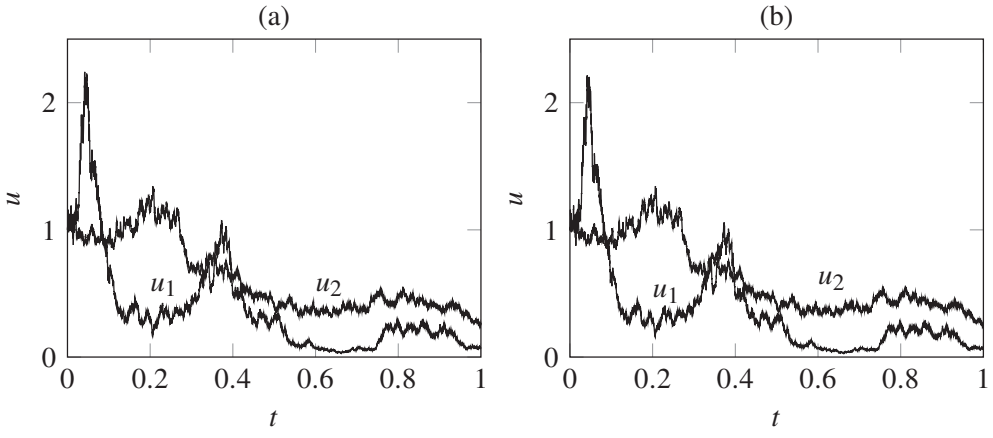

Figure 8.6 For the SODE (8.50), we plot (a) the Milstein approximation using Algorithm 8.3 and (b) the Euler–Maruyama approximation using Algorithm 8.1 to the same sample path of the solution $\boldsymbol{u}(t)$. In both cases, $\Delta t = 2 \times 10^{-4}$. Visually these two approximations appear identical.

Stability and implicit methods

To understand the stability of numerical methods for ODEs, we investigated the linear test equation in Example 3.9 and found a time step restriction is necessary for explicit methods to mimic dissipative behaviour. We now perform a stability analysis for SODEs. Instead of the linear test equation, we use the SODE (with $d = m = 1$) for geometric Brownian motion

$$du = ru\,dt + \sigma u\,dW(t), \qquad u(0) = u_0,$$

which has multiplicative noise. The solution from Example 8.20 is

$$u(t) = \exp\big((r - \sigma^2/2)\,t + \sigma\,W(t)\big)u_0$$

and, using (4.9), we see $\mathbb{E}\big[u(t)^2\big] = \mathrm{e}^{(2r+\sigma^2)t}u_0^2$. For illustration, we consider geometric Brownian motion with parameters $r = -8$ and $\sigma = 3$. Approximate sample paths generated by the Euler–Maruyama method (using Algorithm 8.1) are shown in Figure 8.7 alongside exact sample paths. Two time steps are chosen and we see an instability for $\Delta t = 0.25$ where the numerical approximation becomes very large, while for $\Delta t = 0.0125$ the sample paths stay closer to the true solution. Stability is an issue for the numerical solution of SODEs, just as it was for the ODEs of §3.1.

Consider the case $u_0 \neq 0$. Clearly, $\mathbb{E}\big[u(t)^2\big]$ converges to zero as $t \to \infty$ if and only if $r + \sigma^2/2 < 0$. For what range of parameters does a numerical approximation u_n satisfy $\mathbb{E}\big[u_n^2\big] \to 0$ in the limit $n \to \infty$? We study this question for the Euler–Maruyama method.

Example 8.26 (Euler–Maruyama method) The Euler–Maruyama approximation u_n with time step Δt for the geometric Brownian motion SODE (8.34) is given by

$$u_{n+1} = u_n + ru_n\,\Delta t + \sigma u_n\,\Delta W_n.$$

Then, u_n is written explicitly in terms of u_0 as

$$u_n = \prod_{j=0}^{n-1}\big(1 + r\Delta t + \sigma\Delta W_j\big)\,u_0.$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

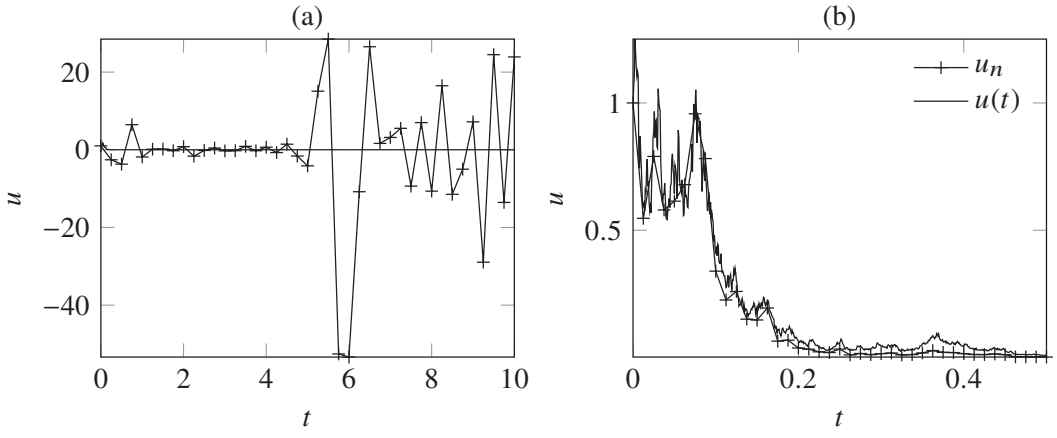

Figure 8.7 A sample path of geometric Brownian motion (8.34) with $u_0 = 1$, $r = -8$, $\sigma = 3$ plotted alongside the numerical approximation by the Euler–Maruyama method for (8.11) with (a) $\Delta t = 0.25$ and (b) $\Delta t = 0.0125$. The time step is smaller in (b) and the approximate sample path is more accurate. In (a), with the larger time step, the approximate path is unstable over a long time period and far from the exact solution.

The second moment of u_n is

$$\mathbb{E}[u_n^2] = \mathbb{E}\left[\prod_{j=0}^{n-1}(1 + r\Delta t + \sigma\Delta W_j)\right]^2 u_0^2 = \prod_{j=0}^{n-1}\mathbb{E}\left[(1 + r\Delta t + \sigma\Delta W_j)^2\right]u_0^2$$
$$= \prod_{j=0}^{n-1}\left((1 + r\Delta t)^2 + \sigma^2\Delta t\right)u_0^2,$$

where we exploit $\Delta W_j \sim \mathrm{N}(0, \Delta t)$ *iid*. Thus, $\mathbb{E}[u_n^2] \to 0$ as $n \to \infty$ if and only if

$$\left|(1 + r\,\Delta t)^2 + \sigma^2\Delta t\right| = 1 + 2\Delta t\left(r + \sigma^2/2 + \Delta t\, r^2/2\right) < 1.$$

Then, the Euler–Maruyama approximation u_n obeys $\mathbb{E}[u_n^2] \to 0$ as $n \to \infty$ if and only if

$$r + \sigma^2/2 + \Delta t\, r^2/2 < 0. \tag{8.52}$$

This is more restrictive than the stability condition $r + \sigma^2/2 < 0$ for the true solution $u(t)$. To achieve stability of the Euler–Maruyama approximation, we must choose the time step so that

$$0 < \Delta t < -2\left(r + \sigma^2/2\right)/r^2.$$

As with ODEs, time step restrictions are often too limiting in practical situations and we turn to implicit methods. We introduce an implicit version of the Euler–Maruyama method known as the θ-Euler–Maruyama method. Exercise 8.16 introduces an alternative for semilinear SODEs.

Definition 8.27 (θ-Euler–Maruyama) Consider a time step $\Delta t > 0$ and initial condition $\boldsymbol{u}_0 \in \mathbb{R}^d$, the θ-Euler–Maruyama approximation $\boldsymbol{u}_n$ to $\boldsymbol{u}(t_n)$ is defined by

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \left[(1-\theta)\boldsymbol{f}(\boldsymbol{u}_n) + \theta\boldsymbol{f}(\boldsymbol{u}_{n+1})\right]\Delta t + G(\boldsymbol{u}_n)\Delta\boldsymbol{W}_n, \tag{8.53}$$

where θ is a parameter in $[0, 1]$ that controls the degree of implicitness in the drift term.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

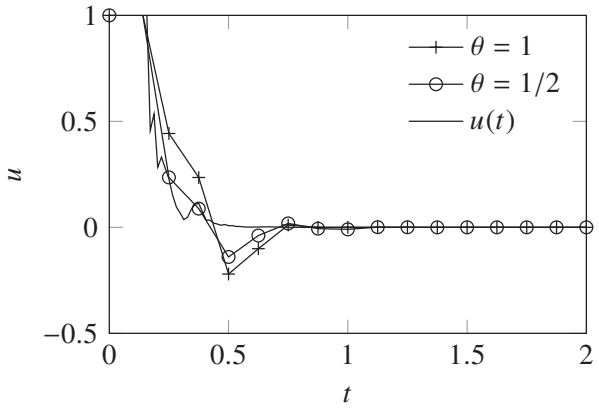

Figure 8.8 A sample path of geometric Brownian motion (8.34) with $u_0 = 1$, $r = -8$, $\sigma = 3$ and an approximation u_n of the sample path using the θ-Euler–Maruyama method on (8.11) with $\theta = 1$ and $\theta = 1/2$ with $\Delta t = 0.25$. Compare to Figure 8.7, where the Euler–Maruyama method was unstable.

The diffusion is treated explicitly, because the point of evaluation of G mimics the definition of the Itô integral.

Example 8.28 Reconsider geometric Brownian motion with parameters $r = -8$ and $\sigma = 3$. Approximate sample paths are generated by the θ-Euler–Maruyama method with $\theta = 1$ and $\theta = 1/2$ with $\Delta t = 1/4$ and are plotted in Figure 8.8(a) alongside the exact sample path. Compare to Figure 8.7(a), where the Euler–Maruyama method is unstable for the same value of Δt (i.e., $r + \sigma^2/2 + \Delta t\, r^2/2 = 4.5 > 0$). The approximation u_n is computed by the θ-Euler–Maruyama method for $\theta = 1$ and $\theta = 1/2$ by calling Algorithm 8.4 with the following commands:

```
>> u0=1; T=10; N=40; d=1; m=1; theta=1;
>> r=-8; sigma=3;
>> [t,u]=EulerMaruyamaTheta(u0,T,N,d,m,@(u)(r*u),@(u)(sigma*u),theta);
```

In both cases, the correct large n behaviour of the sample paths is observed. See Figure 8.8 for the cases $\theta = 1/2$ and $\theta = 1$. Even though the large n behaviour is correct, the numerical approximations can take negative values, which does not happen for geometric Brownian motion with a starting value $u_0 \geq 0$.

The stability calculation can be repeated to see that $\mathbb{E}[u_n^2] \to 0$ as $n \to \infty$ for the θ-Euler–Maruyama approximation u_n to geometric Brownian motion, when

$$r + \sigma^2/2 + \Delta t(1 - 2\theta)r^2/2 < 0.$$

In particular, if $\theta = 1/2$, the stability condition is $r + \sigma^2/2 < 0$, which is independent of the choice of time step and identical to the stability condition for geometric Brownian motion. See Exercise 8.13.

8.5 Strong approximation

For a given time step Δt, both the Euler–Maruyama and Milstein methods generate random variables $\boldsymbol{u}_n$ that approximate the random variables $\boldsymbol{u}(t_n)$ given by the exact solution of the SODE at $t_n = n\Delta t$. We now establish precisely how $\boldsymbol{u}_n$ converges to $\boldsymbol{u}(t_n)$ as $\Delta t \to 0$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Algorithm 8.4 Code to find a sample path of the θ-Euler–Maruyama approximation to (8.3). The inputs and outputs are similar to those in Algorithm 8.1. An additional input `theta` is required, to specify the parameter θ.

```
function [t, u]=EulerMaruyamaTheta(u0,T,N,d,m,fhandle,ghandle,theta)
Dt=T/N; u=zeros(d,N+1); t=[0:Dt:N*Dt]'; sqrtDt=sqrt(Dt);
options=optimset('Display','Off');
u(:,1)=u0; u_n=u0; % initial data
for n=2:N+1, % time loop
    dW=sqrtDt*randn(m,1); % Brownian increment
    u_explicit=u_n+Dt*fhandle(u_n)+ghandle(u_n)*dW;
    if (theta>0) % solve nonlinear eqns for update
                 % u_explicit is initial guess for nonlinear solve
        v=u_n+(1-theta)*Dt*fhandle(u_n)+ghandle(u_n)*dW;
        u_new=fzero(@(u) -u+v +theta*fhandle(u)*Dt,...
                  u_explicit, options);
    else % explicit case
        u_new=u_explicit;
    end
    u(:,n)=u_new; u_n=u_new;
end
```

As we saw in §4.3, there are many notions of convergence for random variables and the convergence of numerical methods for SODEs has to be approached with care. First, we consider *strong convergence*; that is, we care about approximating the sample path $\boldsymbol{u}(\cdot,\omega)$ for an $\omega \in \Omega$. In contrast in §8.6, we look at the *weak convergence* where only the distributions matter.

To make the notion of strong convergence definite, we use the root-mean-square or $L^2(\Omega,\mathbb{R}^d)$ error and show that

$$\sup_{0\le t_n\le T} \|\boldsymbol{u}(t_n)-\boldsymbol{u}_n\|_{L^2(\Omega,\mathbb{R}^d)} = \sup_{0\le t_n\le T} \mathbb{E}\left[\|\boldsymbol{u}(t_n)-\boldsymbol{u}_n\|_2^2\right]^{1/2} \tag{8.54}$$

is $\mathcal{O}(\Delta t)$ for the Milstein method in Theorem 8.32 and $\mathcal{O}(\Delta t^{1/2})$ for the Euler–Maruyama method in Exercise 8.11. Strong convergence is most closely linked to the numerical analysis of deterministic differential equations and makes a natural starting point. When the error (8.54) is small, sample paths $\boldsymbol{u}(\cdot,\omega)$ of the solution are well approximated (on finite time intervals) by a discrete sample path $\boldsymbol{u}_n(\omega)$ and this is useful in examining SODEs as a dynamical system. For example, the Duffing–van der Pol SODE exhibits bifurcations that can be investigated using approximate sample paths; see Exercise 8.9.

Upper bounds for the remainder terms

To prove strong convergence of the Milstein method, we find upper bounds for the remainder terms $\boldsymbol{R}_f$ in (8.36), R_G in (8.38), $\boldsymbol{R}_1$ in (8.44) and $\boldsymbol{R}_M$ in (8.46) under the following assumption.

Assumption 8.29 Let Assumption 8.17 hold, and suppose that f and G are twice continuously differentiable and the second derivatives are uniformly bounded.

First we prove a form of time regularity for the solution of the SODE (8.3).

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Lemma 8.30 *Let Assumption 8.29 hold and $\boldsymbol{u}(t)$ be the solution of* (8.3). *For each $T > 0$ and $\boldsymbol{u}_0 \in \mathbb{R}^d$, there exists $K > 0$ such that for $0 \le s,t \le T$*

$$\|\boldsymbol{u}(t) - \boldsymbol{u}(s)\|_{L^2(\Omega,\mathbb{R}^d)} \le K|t-s|^{1/2}. \tag{8.55}$$

Proof From (8.3),

$$\boldsymbol{u}(t) - \boldsymbol{u}(s) = \int_s^t \boldsymbol{f}(\boldsymbol{u}(r))\,dr + \int_s^t G(\boldsymbol{u}(r))\,d\boldsymbol{W}(r).$$

Then, using (A.10),

$$\mathbb{E}\Big[\|\boldsymbol{u}(t) - \boldsymbol{u}(s)\|_2^2\Big] \le 2\mathbb{E}\left[\left\|\int_s^t \boldsymbol{f}(\boldsymbol{u}(r))\,dr\right\|_2^2\right] + 2\mathbb{E}\left[\left\|\int_s^t G(\boldsymbol{u}(r))\,d\boldsymbol{W}(r)\right\|_2^2\right].$$

Following the argument for deriving (8.29), we find

$$\mathbb{E}\Big[\|\boldsymbol{u}(t) - \boldsymbol{u}(s)\|_2^2\Big] \le 2L^2\big((t-s)+1\big)(t-s)\left(1 + \sup_{t\in[0,T]}\mathbb{E}\Big[\|\boldsymbol{u}(t)\|_2^2\Big]\right).$$

As $\boldsymbol{u} \in \mathcal{H}_{2,T}$, this implies (8.55). □

The following proposition gives the required estimates on the remainders.

Proposition 8.31 *Let Assumption 8.29 hold and $\boldsymbol{u}(t)$ be the solution of* (8.3). *For each $T > 0$ and $\boldsymbol{u}_0 \in \mathbb{R}^d$, there exists $K > 0$ such that the following hold for $0 \le s,t \le T$:*

(i) *For $\boldsymbol{R}_f$ and R_G given by* (8.36) *and* (8.38) *respectively,*

$$\begin{aligned}\mathbb{E}\Big[\|\boldsymbol{R}_f(t;s,\boldsymbol{u}(s))\|_2^2\Big] &\le K|t-s|,\\ \mathbb{E}\Big[\|R_G(t;s,\boldsymbol{u}(s))\|_{\mathrm{F}}^2\Big] &\le K|t-s|^2.\end{aligned} \tag{8.56}$$

(ii) *For $\boldsymbol{R}_1$ given by* (8.44),

$$\mathbb{E}\Big[\|\boldsymbol{R}_1(t;s,\boldsymbol{u}(s))\|_2^2\Big] \le K|t-s|^2. \tag{8.57}$$

(iii) *For $\boldsymbol{R}_M$ given by* (8.46) *and $t_j = j\Delta t$ for $j = 0,1,2\ldots$,*

$$\mathbb{E}\left[\left\|\sum_{t_j<t} \boldsymbol{R}_M(t_{j+1}\wedge t; t_j, \boldsymbol{u}(t_j))\right\|_2^2\right] \le K\Delta t^2.$$

Proof (i) Under Assumption 8.29, the estimates follow from the definitions of $\boldsymbol{R}_f, R_G$ in (8.36)–(8.38) and the bound on $\|\boldsymbol{u}(t) - \boldsymbol{u}(s)\|_{L^2(\Omega,\mathbb{R}^d)}$ in (8.55).

(ii) From (8.41) and (A.10),

$$\begin{aligned}\mathbb{E}\Big[\|\boldsymbol{R}_E(t;s,\boldsymbol{u}(s))\|_2^2\Big] &\le 3\mathbb{E}\left[\left\|\int_s^t \boldsymbol{R}_{\boldsymbol{f}}(r;s,\boldsymbol{u}(s))\,dr\right\|_2^2\right]\\ &\quad + 3\mathbb{E}\left[\left\|\int_s^t R_G(r;s,\boldsymbol{u}(s))\,d\boldsymbol{W}(r)\right\|_2^2\right] + 3\mathbb{E}\left[\left\|\int_s^t DG(\boldsymbol{u}(s))(\boldsymbol{u}(r)-\boldsymbol{u}(s))\,d\boldsymbol{W}(r)\right\|_2^2\right].\end{aligned}$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

By Jensen's inequality (A.9) and the Itô isometry (8.22),

$$\mathbb{E}\left[\left\|\boldsymbol{R}_E(t;s,\boldsymbol{u}(s))\right\|_2^2\right] \le 3\mathbb{E}\left[\left\|\boldsymbol{R}_{\boldsymbol{f}}(t;s,\boldsymbol{u}(s))\right\|_2^2\right]|t-s|^2 + 3\int_s^t \mathbb{E}\left[\left\|R_G(r;s,\boldsymbol{u}(r))\right\|_{\mathrm{F}}^2\right] dr + 3\int_s^t \mathbb{E}\left[\left\|DG(\boldsymbol{u}(s))(\boldsymbol{u}(r)-\boldsymbol{u}(s))\right\|_{\mathrm{F}}^2\right] dr.$$

The first two terms are $\mathcal{O}(|t-s|^3)$ by (8.56). Because $DG(\boldsymbol{u})$ is a bounded linear operator uniformly over $\boldsymbol{u} \in \mathbb{R}^d$,

$$\left\|DG(\boldsymbol{u}(s))\big(\boldsymbol{u}(r)-\boldsymbol{u}(s)\big)\right\|_{\mathrm{F}} \le K\left\|\boldsymbol{u}(r)-\boldsymbol{u}(s)\right\|_2$$

for some constant K and Lemma 8.30 says that the third term is $\mathcal{O}(|t-s|^2)$. Finally, taken together, we have $\|\boldsymbol{R}_E(t;s,\boldsymbol{u}(s))\|_{L^2(\Omega,\mathbb{R}^d)} = \mathcal{O}(|t-s|)$. (8.57) follows easily from the definition of $\boldsymbol{R}_1$ in (8.44).

(iii) For simplicity, assume that t is an integer multiple of Δt. Using (8.46) with (A.10) and (8.22), we have

$$\mathbb{E}\left[\left\|\sum_{t_j<t} \boldsymbol{R}_M(t_{j+1}\wedge t;t_j,\boldsymbol{u}(t_j))\right\|_2^2\right] \le 3\mathbb{E}\left[\left\|\sum_{t_j<t}\int_{t_j}^{t_{j+1}} \boldsymbol{R}_f(r;t_j,\boldsymbol{u}(t_j))\,dr\right\|_2^2\right] + 3\sum_{t_j<t}\int_{t_j}^{t_{j+1}}\left(\mathbb{E}\left[\left\|R_G(r;t_j,\boldsymbol{u}(t_j))\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\left\|DG(\boldsymbol{u}(s))\boldsymbol{R}_1(s,\boldsymbol{u}(s))\right\|_{\mathrm{F}}^2\right]\right) dr.$$

Proposition 8.31(i) provides an $\mathcal{O}(\Delta t^2)$ estimate for the second term. For the first term, recall the definition of $\boldsymbol{R}_f$ from (8.36).

$$\begin{aligned}\boldsymbol{R}_f(r;s,\boldsymbol{u}(s)) &= D\boldsymbol{f}(\boldsymbol{u}(s))\big(\boldsymbol{u}(r)-\boldsymbol{u}(s)\big) \\ &\quad + \int_0^1 (1-h)D^2\boldsymbol{f}\big(\boldsymbol{u}(s)+h(\boldsymbol{u}(r)-\boldsymbol{u}(s))\big)\big(\boldsymbol{u}(r)-\boldsymbol{u}(s)\big)^2\,dh \\ &= D\boldsymbol{f}(\boldsymbol{u}(s))\left(G(\boldsymbol{u}(s))\int_s^r d\boldsymbol{W}(p)\right) + D\boldsymbol{f}(\boldsymbol{u}(s))\,\boldsymbol{R}_1(s,\boldsymbol{u}(s)) \\ &\quad + \int_0^1 (1-h)D^2\boldsymbol{f}\big(\boldsymbol{u}(s)+h(\boldsymbol{u}(r)-\boldsymbol{u}(s))\big)\big(\boldsymbol{u}(r)-\boldsymbol{u}(s)\big)^2\,dh.\end{aligned}$$

using (8.43) gives for $r > s$. Let

$$\boldsymbol{\Theta}_j := \int_{t_j}^{t_{j+1}} D\boldsymbol{f}(\boldsymbol{u}(t_j))\left(G(\boldsymbol{u}(t_j))\int_{t_j}^r d\boldsymbol{W}(p)\right) dr$$

so that

$$\begin{aligned}\mathbb{E}\left[\left\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \boldsymbol{R}_f(r;t_j,\boldsymbol{u}(t_j))\,dr\right\|_2^2\right] &\le 2\mathbb{E}\left[\left\|\sum_{j=0}^{k-1}\boldsymbol{\Theta}_j\right\|_2^2\right] \\ &\quad + 2\mathbb{E}\left[\left\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} D\boldsymbol{f}(\boldsymbol{u}(t_j))\,\boldsymbol{R}_1(t_j,\boldsymbol{u}(t_j))\right.\right. \\ &\qquad \left.\left. + \int_0^1 (1-h)D^2\boldsymbol{f}\big(\boldsymbol{u}(t_j)+h(\boldsymbol{u}(r)-\boldsymbol{u}(t_j))\big)\big(\boldsymbol{u}(r)-\boldsymbol{u}(t_j)\big)^2\,dh\,dr\right\|_2^2\right].\end{aligned}$$

The second term is $\mathcal{O}(\Delta t^2)$ by (8.57) and (8.55). For the first term, expand the sum in

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

terms of $\langle \boldsymbol{\Theta}_j, \boldsymbol{\Theta}_k \rangle$ using the definition of the 2-norm. For $k > j$, we have $\mathbb{E}[\langle \boldsymbol{\Theta}_j, \boldsymbol{\Theta}_k \rangle] = \mathbb{E}[\langle \boldsymbol{\Theta}_j, \mathbb{E}[\boldsymbol{\Theta}_k \mid \mathcal{F}_{t_k}] \rangle] = 0$ a.s. Thus

$$\mathbb{E}\left[\left\|\sum_{j=0}^{k-1} \boldsymbol{\Theta}_j\right\|_2^2\right] = \sum_{j=0}^{k-1} \mathbb{E}\left[\left\|\int_{t_j}^{t_{j+1}} D\boldsymbol{f}(\boldsymbol{u}(t_j))\left(G(\boldsymbol{u}(t_j)) \int_{t_j}^{r} d\boldsymbol{W}(p)\right) dr\right\|_2^2\right].$$

As $\int_{t_j}^{t_{j+1}} \int_{t_j}^{r} d\boldsymbol{W}(p)\, dr \sim \mathrm{N}(\boldsymbol{0}, \frac{1}{3}\Delta t^3 I_m)$ (see Exercise 8.3), each term in the sum is $\mathcal{O}(\Delta t^3)$ and the sum is $\mathcal{O}(\Delta t^2)$ overall. □

Strong convergence of the Milstein method

We prove strong convergence of the Milstein method. For the proof, we define a continuous time process $\boldsymbol{u}_{\Delta t}(t)$ that agrees with the approximation $\boldsymbol{u}_n$ at $t = t_n$. To do this, introduce the variable $\hat{t} := t_n$ for $t_n \le t < t_{n+1}$ and let

$$\begin{aligned}\boldsymbol{u}_{\Delta t}(t) = \boldsymbol{u}_{\Delta t}(\hat{t}) + \boldsymbol{f}(\boldsymbol{u}_{\Delta t}(\hat{t})) \int_{\hat{t}}^{t} ds + G(\boldsymbol{u}_{\Delta t}(\hat{t})) \int_{\hat{t}}^{t} d\boldsymbol{W}(r) \\ + \int_{\hat{t}}^{t} DG(\boldsymbol{u}_{\Delta t}(\hat{t}))\left(G(\boldsymbol{u}_{\Delta t}(\hat{t})) \int_{\hat{t}}^{r} d\boldsymbol{W}(p)\right) d\boldsymbol{W}(r).\end{aligned}$$

Notice that $\boldsymbol{u}_{\Delta t}(\hat{t}) = \boldsymbol{u}_n$ for $\hat{t} = t_n$. Furthermore, using the definition of $\hat{t}$,

$$\begin{aligned}\boldsymbol{u}_{\Delta t}(t) = \boldsymbol{u}_{\Delta t}(t_0) + \int_{t_0}^{t} \boldsymbol{f}(\boldsymbol{u}_{\Delta t}(\hat{s}))ds + \int_{t_0}^{t} G(\boldsymbol{u}_{\Delta t}(\hat{s}))d\boldsymbol{W}(s) \\ + \int_{t_0}^{t} DG(\boldsymbol{u}_{\Delta t}(\hat{s}))\left(G(\boldsymbol{u}_{\Delta t}(\hat{s})) \int_{\hat{s}}^{s} d\boldsymbol{W}(r)\right) d\boldsymbol{W}(s).\end{aligned} \tag{8.58}$$

In the following, we assume in (8.59) that $DG(\boldsymbol{u})G(\boldsymbol{u})$ is globally Lipschitz. This occurs if $G(\boldsymbol{u})$ is a linear function of $\boldsymbol{u}$ (as with geometric Brownian motion) or if $G(\boldsymbol{u})$, $DG(\boldsymbol{u})$, and $D^2G(\boldsymbol{u})$ are all uniformly bounded over $\boldsymbol{u} \in \mathbb{R}^d$.

Theorem 8.32 (convergence of Milstein method) *Let Assumption 8.29 hold. Further assume, for some $L_2 > 0$, that*

$$\left\|DG(\boldsymbol{u}_1)(G(\boldsymbol{u}_1)\boldsymbol{\xi}) - DG(\boldsymbol{u}_2)(G(\boldsymbol{u}_2)\boldsymbol{\xi})\right\|_{\mathrm{F}} \le L_2 \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_2 \|\boldsymbol{\xi}\|_2 \tag{8.59}$$

for all $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^d$ and $\boldsymbol{\xi} \in \mathbb{R}^m$. Let $\boldsymbol{u}_n$ denote the Milstein approximation to the solution $\boldsymbol{u}(t)$ of (8.3). *For $T \ge 0$, there exists $K > 0$ such that*

$$\sup_{0 \le t_n \le T} \left\|\boldsymbol{u}(t_n) - \boldsymbol{u}_n\right\|_{L^2(\Omega, \mathbb{R}^d)} \le K\Delta t. \tag{8.60}$$

Proof To simplify notation, let $\boldsymbol{D}(s) := \boldsymbol{f}(\boldsymbol{u}_{\Delta t}(\hat{s})) - \boldsymbol{f}(\boldsymbol{u}(\hat{s}))$ and

$$\begin{aligned}M(s) := &\left[G(\boldsymbol{u}_{\Delta t}(\hat{s})) - G(\boldsymbol{u}(\hat{s}))\right] \\ &+ \left[DG(\boldsymbol{u}_{\Delta t}(\hat{s}))\left(G(\boldsymbol{u}_{\Delta t}(\hat{s})) \int_{\hat{s}}^{s} d\boldsymbol{W}(r)\right) - DG(\boldsymbol{u}(\hat{s}))\left(G(\boldsymbol{u}(\hat{s})) \int_{\hat{s}}^{s} d\boldsymbol{W}(r)\right)\right].\end{aligned}$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Then, subtracting (8.45) from (8.58),

$$\begin{aligned} \boldsymbol{u}_{\Delta t}(t) - \boldsymbol{u}(t) = &\int_0^t \boldsymbol{D}(s)\,ds + \int_0^t M(s)\,d\boldsymbol{W}(s) \\ &+ \sum_{t_j<t} \boldsymbol{R}_M(t_{j+1}\wedge t; t_j, \boldsymbol{u}(t_j)) + \boldsymbol{R}_M(t; t_n, \boldsymbol{u}(t_n)). \end{aligned} \tag{8.61}$$

Let $\boldsymbol{e}(\hat{t}\,) = \boldsymbol{u}_{\Delta t}(\hat{t}\,) - \boldsymbol{u}(\hat{t}\,)$. Then,

$$\begin{aligned} \mathbb{E}\left[\|\boldsymbol{e}(\hat{t}\,)\|_2^2\right] \le\ & 3\mathbb{E}\left[\left\|\int_0^{\hat{t}} \boldsymbol{D}(s)\,ds\right\|_2^2\right] + 3\mathbb{E}\left[\left\|\int_0^{\hat{t}} M(s)\,d\boldsymbol{W}(s)\right\|_2^2\right] \\ &+ 3\mathbb{E}\left[\left\|\sum_{t_j<\hat{t}} \boldsymbol{R}_M(t_{j+1}\wedge\hat{t}; t_j, \boldsymbol{u}(t_j))\right\|_2^2\right]. \end{aligned} \tag{8.62}$$

For the first term, the Lipschitz condition (8.25) on $\boldsymbol{f}$ and the Cauchy–Schwarz inequality give

$$\mathbb{E}\left[\left\|\int_0^{\hat{t}} \boldsymbol{D}(s)\,ds\right\|_2^2\right] \le\ \hat{t}\int_0^{\hat{t}} \mathbb{E}\left[\|\boldsymbol{D}(s)\|_2^2\right] ds \le \hat{t}\int_0^{\hat{t}} L^2\mathbb{E}\left[\|\boldsymbol{e}(s)\|^2\right] ds.$$

For the second term of (8.62), we have, by (8.22),

$$\mathbb{E}\left[\left\|\int_0^{\hat{t}} M(s)\,d\boldsymbol{W}(s)\right\|_2^2\right] = \int_0^{\hat{t}} \mathbb{E}\left[\|M(s)\|_{\mathrm{F}}^2\right] ds.$$

Under (8.24) and (8.59),

$$\begin{aligned} \mathbb{E}\left[\|M(s)\|_{\mathrm{F}}^2\right] \le\ & 2\mathbb{E}\left[\|G(\boldsymbol{u}_{\Delta t}(\hat{s})) - G(\boldsymbol{u}(\hat{s}))\|_{\mathrm{F}}^2\right] \\ + 2\mathbb{E}&\left[\left\|DG(\boldsymbol{u}_{\Delta t}(\hat{s}))\left(G(\boldsymbol{u}_{\Delta t}(\hat{s}))\int_{\hat{s}}^s d\boldsymbol{W}(r)\right) - DG(\boldsymbol{u}(\hat{s}))\left(G(\boldsymbol{u}(\hat{s}))\int_{\hat{s}}^s d\boldsymbol{W}(r)\right)\right\|_{\mathrm{F}}^2\right] \\ &\le 2(L+L_2)\,\mathbb{E}\left[\|\boldsymbol{e}(s)\|_2^2\right]. \end{aligned}$$

Consequently, for some $K_3 > 0$ independent of Δt,

$$\mathbb{E}\left[\left\|\int_0^{\hat{t}} M(s)\,d\boldsymbol{W}(s)\right\|_2^2\right] \le K_3\int_0^{\hat{t}} \mathbb{E}\left[\|\boldsymbol{e}(s)\|_2^2\right] ds.$$

For the third term of (8.62), there exists $K_2 > 0$ by Proposition 8.31(iii) again independent of Δt such that

$$\mathbb{E}\left[\left\|\sum_{j=0}^{n-1} \boldsymbol{R}_M(t_{j+1}\wedge\hat{t}; t_j, \boldsymbol{u}(t_j))\right\|_2^2\right] \le K_2\Delta t^2.$$

Putting the estimates together with (8.62), we have for $t \in [0,T]$

$$\mathbb{E}\left[\|\boldsymbol{e}(\hat{t}\,)\|_2^2\right] \le 3(TL^2 + K_3)\int_0^{\hat{t}} \mathbb{E}\left[\|\boldsymbol{e}(s)\|_2^2\right] ds + 3K_2\Delta t^2.$$

Gronwall's inequality (Exercise 3.4) completes the proof. □

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Theorem 8.32 holds for both multiplicative noise (where G depends on $\boldsymbol{u}$) and additive noise (where G is independent of $\boldsymbol{u}$). For the special case of additive noise, the Euler–Maruyama method (recall Definition 8.22) and the Milstein method are the same and Theorem 8.32 shows convergence of the Euler–Maruyama method with $\mathcal{O}(\Delta t)$. However, in general, the Euler–Maruyama method converges with $\mathcal{O}(\Delta t^{1/2})$. See Exercise 8.11.

To apply the Milstein method in dimension $d > 1$, the Lévy areas (8.47) are usually required. They are difficult to sample and, in general, must be approximated. This introduces an extra error term into the remainder defined in (8.46). To achieve the same $L^2(\Omega, \mathbb{R}^d)$ order of approximation, the error in approximating the Lévy areas must be consistent with Proposition 8.31(iii) and hence must be approximated with an error of size $\mathcal{O}(\Delta t^{3/2})$. Exercise 8.12 gives a method for approximating the Lévy areas based on applying the Euler–Maruyama method to a specific SODE. The method is slow and often the cost of approximating the Lévy areas outweighs the advantages of the higher convergence rate of the Milstein method.

Algorithm 8.5 Code to find M *iid* sample paths of the Euler–Maruyama approximation $\boldsymbol{u}_n$ to the solution $\boldsymbol{u}(t)$ of (8.3). Inputs and outputs are similar to those in Algorithm 8.1. In addition, `M` specifies the number of samples and `kappa0`$=\kappa$ determines the size of the Brownian increments used in (8.64) for Δt_{ref} =`T/(N*kappa0)`. `u` is a three-dimensional array and `u(:,j,n+1)` $= \boldsymbol{u}_n^j$ for $n = 0, \ldots, N$ and $j = 1, \ldots, M$.

```
function [t, u]=EMpath(u0,T,N,d,m,fhandle,ghandle,kappa0,M)
Dtref=T/N;  % small step
Dt=kappa0*Dtref; % large step
NN=N/kappa0;  u=zeros(d,M,NN+1); t=zeros(NN+1,1);
gdW=zeros(d,M); sqrtDtref=sqrt(Dtref); u_n=u0;
for n=1:NN+1
  t(n)=(n-1)*Dt; u(:,:,n)=u_n;
  dW=sqrtDtref*squeeze(sum(randn(m,M,kappa0),3));
  for mm=1:M
    gdW(:,mm)=ghandle(u_n(:,mm))*dW(:,mm);
  end
  u_new=u_n+Dt*fhandle(u_n)+gdW; u_n=u_new;
end
```

Calculating the L^2 error numerically

We consider strong convergence numerically by estimating the $L^2(\Omega, \mathbb{R}^d)$ norm using a sample average. That is,

$$\|\boldsymbol{u}(T) - \boldsymbol{u}_N\|_{L^2(\Omega,\mathbb{R}^d)} \approx \left(\frac{1}{M} \sum_{j=1}^{M} \left\| \boldsymbol{u}^j(T) - \boldsymbol{u}_N^j \right\|_2^2 \right)^{1/2}. \tag{8.63}$$

Here, $\boldsymbol{u}^j(T) = \boldsymbol{u}(T, \omega_j)$ are independent samples of the exact solution $\boldsymbol{u}(T)$ and $\boldsymbol{u}_N^j = \boldsymbol{u}_N(\omega_j)$ is the numerical approximation to $\boldsymbol{u}(T, \omega_j)$ for $T = N\Delta t$. There are a number of different numerical issues to consider. First, we need to simulate $\boldsymbol{u}_N^j$ using the same Brownian path as $\boldsymbol{u}^j(T)$. To do this, we choose independent sample paths $\boldsymbol{W}(\cdot, \omega_j)$ of the Brownian motion $\boldsymbol{W}(t)$ and make sure the same sample path is used for both samples.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Second, it is unusual to have explicit solutions and, in practice, $\boldsymbol{u}^j(T)$ are approximated by accurate reference solutions $\boldsymbol{u}^j_{N_{\text{ref}}}$, computed with a small time step $\Delta t_{\text{ref}} = T/N_{\text{ref}}$. To keep the same sample path, the increments $W((n+1)\Delta t_{\text{ref}}) - W(n\Delta t_{\text{ref}})$ for the reference sample paths are used to generate the required increments over a time step $\Delta t = \kappa \Delta t_{\text{ref}}$ by

$$\boldsymbol{W}((n+1)\Delta t) - \boldsymbol{W}(n\Delta t) = \sum_{j=n\kappa}^{(n+1)\kappa-1} \big(\boldsymbol{W}((j+1)\Delta t_{\text{ref}}) - \boldsymbol{W}(j\Delta t_{\text{ref}})\big). \tag{8.64}$$

Third, we need to generate a large number of samples M for (8.63) to be a good approximation. Finally, we need to do this over a range of different time steps Δt to see the error in (8.63) decay and hence observe convergence.

To start with, we compute approximations to $\boldsymbol{u}(t)$ with two different time steps using the same sample path of $\boldsymbol{W}(t)$ (taking $M = 1$). This is implemented in Algorithm 8.5 using the Euler–Maruyama method (compare to Algorithm 8.1). This algorithm uses two time steps Δt and Δt_{ref} such that $\Delta t = \kappa \Delta t_{\text{ref}}$ for some $\kappa \in \mathbb{N}$. Then, $T = N_{\text{ref}} \Delta t_{\text{ref}} = N \Delta t$ and $N_{\text{ref}} = N\kappa$. We use $\boldsymbol{u}_n$ to denote the approximation to $\boldsymbol{u}(t_n)$ with time step Δt and $\boldsymbol{u}_{n\kappa}$ to denote the approximation to $\boldsymbol{u}(t_n)$ with step Δt_{ref}. The Brownian increments for the larger step Δt are computed using steps of size Δt_{ref} using (8.64).

Example 8.33 (one sample path) Consider the two-dimensional SODE in Example 8.25 (which has independent geometric Brownian motions in each component). We examine a single sample path ($M = 1$) and generate approximations with time steps Δt and Δt_{ref} using the same Brownian path. For the time step Δt_{ref}, call Algorithm 8.5 with `N=Nref` and `kappa0=1` as follows:

```
>> d=2; m=2; T=1; Nref=5000; M=1; A=[1 0 ; 0 -1]; u0=[1;1];
>> defaultStream = RandStream.getGlobalStream;
>> savedState = defaultStream.State;
>> [tref, uref]=EMpath(u0,T,Nref,d,m,...
                        @(u) A*u,@(u)[u(1) 0 ; 0 2*u(2)],1,M);
```

To approximate the same sample path with time step $\Delta t = \kappa \Delta t_{\text{ref}}$, we reset the random number generator so the next set of Brownian increments uses the *same* random numbers as the first call. This ensures we approximate the same sample path of $\boldsymbol{u}(t)$. We then call Algorithm 8.5 with `kappa0=kappa`.

```
>> defaultStream.State = savedState; kappa=100;
>> [t, u]=EMpath(u0,T,Nref,d,m,@(u)A*u,@(u)[u(1) 0 ;0 2*u(2)],kappa,M);
>> err=squeeze(u-uref(:,:,1:kappa:end));
```

In Figure 8.9(a), we plot [`tref`, `uref`] and [`t`, `u`] for the same sample path shown in Figure 8.6. In this case, the time steps are $\Delta t_{\text{ref}} = 2 \times 10^{-4}$ and $\Delta t = \kappa \Delta t_{\text{ref}} = 0.02$ for $\kappa = 100$. The matrix `err` has columns `err(:,n)` corresponding to $\boldsymbol{u}_n - \boldsymbol{u}_{n\kappa}$ and, using this data, we plot the error $\|\boldsymbol{u}_n - \boldsymbol{u}_{n\kappa}\|_2$ in Figure 8.9(b).

We consider how to approximate $\|\boldsymbol{u}(T) - \boldsymbol{u}_N\|_{L^2(\Omega,\mathbb{R}^d)}$. We take M independent samples $\boldsymbol{u}^j(T)$ of the solution, their approximations $\boldsymbol{u}^j_N$, and estimate the error for each. Rather than use an analytical solution, we approximate $\boldsymbol{u}^j(T)$ by a reference solution $\boldsymbol{u}^j_{N_{\text{ref}}}$ obtained with $\Delta t_{\text{ref}} = T/N_{\text{ref}}$ and $N_{\text{ref}} = \kappa N$, some $\kappa > 1$. As the time step $\Delta t_{\text{ref}} = T/\kappa N$ is smaller

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

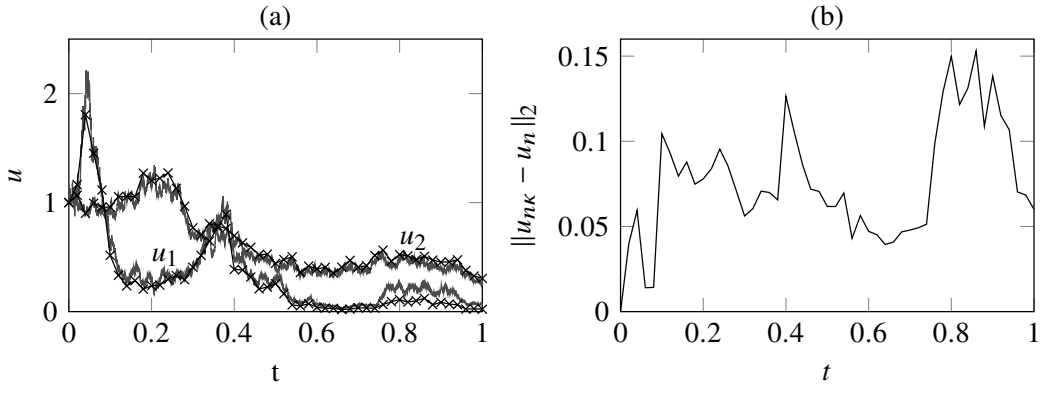

Figure 8.9 (a) Approximation to a sample path of (8.50) by the Euler–Maruyama method with $\Delta t = 0.04$ (marked by $\times$) and reference approximation computed with $\Delta t_{\text{ref}} = 2 \times 10^{-4}$ (solid line). (b) Plot of the error $\|\boldsymbol{u}_{n\kappa} - \boldsymbol{u}_n\|_2$ against t_n for $n = 0, 1, \ldots, N$.

than Δt, $\boldsymbol{u}^j_{N_{\text{ref}}}$ should be more accurate than $\boldsymbol{u}^j_N$. Then, we approximate

$$\|\boldsymbol{u}(T) - \boldsymbol{u}_N\|_{L^2(\Omega,\mathbb{R}^d)} \approx \left(\frac{1}{M} \sum_{j=1}^{M} \left\| \boldsymbol{u}^j_{N_{\text{ref}}} - \boldsymbol{u}^j_N \right\|_2^2 \right)^{1/2} \tag{8.65}$$

where we have replaced the sample $\boldsymbol{u}^j(T)$ in (8.63) by the reference solution $\boldsymbol{u}^j_{N_{\text{ref}}}$. This is different to Example 8.33 where we have one sample ($M = 1$) and we plot the error on $[0, T]$. For $M > 1$, Algorithm 8.5 computes M *iid* approximate sample paths simultaneously using the Euler–Maruyama method. Algorithm 8.6 adapts the MATLAB commands in Example 8.33 to evaluate (8.65). To deal with large M, the samples are split into blocks of size at most `Mstep`. For each block, Algorithm 8.5 is called twice and in each case Brownian paths are computed using the small time step Δt_{ref} (we reset the seed between calls to Algorithm 8.5). These blocks may be run in parallel, although here they are run in serial.

Example 8.34 (M sample paths) Continuing with Example 8.33, we estimate numerically rates of convergence of the Euler–Maruyama method using Algorithm 8.6. Although we have an exact solution of (8.50), we estimate the error using a reference solution and compute (8.65). Using Algorithm 8.6, the following commands compute (8.65) with $M = 5000$ samples and $N_{\text{ref}} = 5000$ for the six time steps $\Delta t = \kappa \Delta t_{\text{ref}}$ given by $\kappa = 5, 10, 50, 100, 500, 1000$.

```
>> d=2; m=2; T=1; N=5000; Dtref=T/N; M=5000;
>> A=[1 0 ; 0 -1]; u0=repmat([1;1],[1,M]);
>> kappa=[5,10,50,100,500,1000];
>> for k=1:length(kappa)
     [rmsErr(k)]=runEMpath(u0,T,N,d,m,@(u) A*u,...
                           @(u)[u(1) 0 ; 0 2*u(2)],kappa(k),M);
   end
```

Using this data, Figure 8.10 illustrates the rates of convergence for the Euler–Maruyama method. The figure also shows results for the Milstein method, found using a code similar to Algorithm 8.5 based on Algorithm 8.3. The rates of convergence found numerically are in agreement with the strong convergence theory of Theorem 8.32 and Exercise 8.11.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Algorithm 8.6 Code to approximate the $L^2(\Omega, \mathbb{R}^d)$ error for the Euler–Maruyama method using (8.65). Inputs are as in Algorithm 8.5, except this time `kappa` denotes κ. The output `rmsErr` gives (8.65). For the last block of samples, we also output the results `[tref, uref]` and `[t, u]` of calling Algorithm 8.5 with `kappa0=1` and `kappa0=kappa` respectively. Crucially, `u(:,j,:)` and `uref(:,j,:)` approximate the same sample path $\boldsymbol{u}^j(t)$.

```
function [rmsErr,t,u,tref,uref]=runEMpath(u0,T,Nref,d,m,...
                                          fhandle,ghandle,kappa,M)
S=0; Mstep=1000; m0=1;
for mm=1:Mstep:M
  MM=min(Mstep,M-mm+1); u00=u0(:,mm:m0+MM-1);
  defaultStream = RandStream.getGlobalStream;
  savedState = defaultStream.State;
  [tref, uref]=EMpath(u00, T, Nref, d, m, fhandle, ghandle,1,MM);
  defaultStream.State = savedState;
  [t, u]=EMpath(u00, T, Nref, d, m,fhandle, ghandle,kappa,MM);
  err=u(:,:,end)-uref(:,:,end);
  S=S+sum(sum(err.*err)); m0=m0+MM;
end
rmsErr=sqrt(S./M);
```

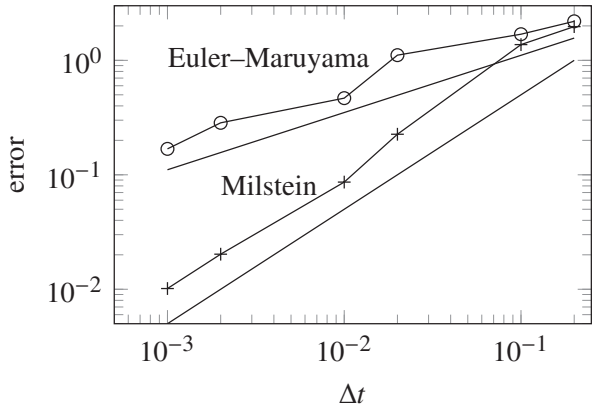

Figure 8.10 A log-log plot of the approximation (8.65) to $\|u(T) - u_N\|_{L^2(\Omega,\mathbb{R}^2)}$ at $t_N = T = 1$ against Δt, showing convergence of the Euler–Maruyama and Milstein methods for (8.50). Here $M = 5000$ samples and $\Delta t_{\text{ref}} = 2 \times 10^{-4}$. Also shown are reference lines of slope 1/2 and 1 indicating the convergence rates of the Euler–Maruyama and Milstein methods are $\mathcal{O}(\Delta t^{1/2})$ and $\mathcal{O}(\Delta t)$, respectively, in agreement with the theory.

8.6 Weak approximation

In many situations, individual sample paths are not of interest and it is average quantities, such as the temperature $\mathbb{E}\left[\frac{1}{2}P(T)^2\right]$ in Example 8.1, that are needed. The approximation of $\mathbb{E}[\phi(\boldsymbol{u}(T))]$ for a given quantity of interest or test function $\phi\colon \mathbb{R}^d \to \mathbb{R}$ is known as *weak approximation* and we now use the Monte Carlo and Euler–Maruyama methods to compute weak approximations. In other words, we generate M independent samples $\boldsymbol{u}_N^j$ for $j = 1, \ldots, M$ of the Euler–Maruyama approximation $\boldsymbol{u}_N$ to $\boldsymbol{u}(T)$ for $t_N = T$ and

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

approximate $\mathbb{E}[\phi(\boldsymbol{u}(T))]$ by the sample average

$$\mu_M \coloneqq \frac{1}{M}\sum_{j=1}^{M} \phi(\boldsymbol{u}_N^j). \tag{8.66}$$

As we saw in Example 4.72, the error can be divided as the sum of the weak discretisation error due to approximating $\boldsymbol{u}(T)$ by $\boldsymbol{u}_N$ and the Monte Carlo error due to taking M samples,

$$\mathbb{E}\big[\phi(\boldsymbol{u}(T))\big] - \mu_M = \underbrace{\Big[\mathbb{E}\big[\phi(\boldsymbol{u}(T))\big] - \mathbb{E}\big[\phi(\boldsymbol{u}_N)\big]\Big]}_{\text{weak discretization error}} + \underbrace{\Big[\mathbb{E}\big[\phi(\boldsymbol{u}_N)\big] - \mu_M\Big]}_{\text{Monte Carlo error}}. \tag{8.67}$$

From (4.34), the Monte Carlo error satisfies

$$\mathbb{P}\left(\big|\mathbb{E}\big[\phi(\boldsymbol{u}_N)\big] - \mu_M\big| < \frac{2\sigma}{\sqrt{M}}\right) > 0.95 + \mathcal{O}(M^{-1/2}), \qquad \sigma^2 \coloneqq \operatorname{Var}(\phi(\boldsymbol{u}_N)).$$

By Theorem 4.58(ii) (with $r = 1/2$, $H = \mathbb{R}^d$, $X_j = \boldsymbol{u}_N$, $\Delta t = T/j$), the weak discretisation error $\mathbb{E}[\phi(\boldsymbol{u}(T))] - \mathbb{E}[\phi(\boldsymbol{u}_N)]$ for globally Lipschitz test functions $\phi\colon \mathbb{R}^d \to \mathbb{R}$ is controlled by the root-mean-square error $\|\boldsymbol{u}(T) - \boldsymbol{u}_N\|_{L^2(\Omega,\mathbb{R}^d)}$. In particular, due to Exercise 8.11, there exists $K > 0$ such that

$$\Big|\mathbb{E}\big[\phi(\boldsymbol{u}_N)\big] - \mathbb{E}\big[\phi(\boldsymbol{u}(T))\big]\Big| \le K\Delta t^{1/2}, \qquad t_N = T. \tag{8.68}$$

Consequently, the weak error obeys

$$\mathbb{P}\left(\big|\mathbb{E}\big[\phi(\boldsymbol{u}(T))\big] - \mu_M\big| < K\Delta t^{1/2} + \frac{2\sigma}{\sqrt{M}}\right) > 0.95 + \mathcal{O}(M^{-1/2}),$$

or informally we write $\mu_M = \mathbb{E}[\phi(\boldsymbol{u}(T))] + \mathcal{O}(\Delta t^{1/2}) + \mathcal{O}(M^{-1/2})$. In this section, we investigate weak approximation in more detail. We first look at the weak discretisation error and show, in Theorem 8.45, that the error for the Euler–Maruyama method is in fact $\mathcal{O}(\Delta t)$ when ϕ is sufficiently regular. This is smaller than the $\mathcal{O}(\Delta t^{1/2})$ error predicted by (8.68). We then turn to the Monte Carlo error and introduce the multilevel Monte Carlo method.

Markov property

We require two theoretical concepts to study weak approximation: the Markov property and the backward Kolmogorov equation. We start with the Markov property and this is easiest to describe for discrete time, such as the random variables $\boldsymbol{u}_n$ given by the Euler–Maruyama or Milstein methods. In both cases, we determine $\boldsymbol{u}_{n+1}$ from $\boldsymbol{u}_n$ and do not use $\boldsymbol{u}_{n-1}, \boldsymbol{u}_{n-2}, \ldots$. In other words, the conditional distribution of the future (the random variable $\boldsymbol{u}_{n+1}$) given the present (the random variable $\boldsymbol{u}_n$) is independent of the past (the random variables $\boldsymbol{u}_{n-k}$ for $k = 1, \ldots, n$). In terms of condition expectation (see Definition 4.47), we have for any bounded measurable function $\phi\colon \mathbb{R}^d \to \mathbb{R}$ that

$$\mathbb{E}\Big[\phi(\boldsymbol{u}_{n+1}) \,\Big|\, \boldsymbol{u}_n\Big] = \mathbb{E}\Big[\phi(\boldsymbol{u}_{n+1}) \,\Big|\, \boldsymbol{u}_n, \boldsymbol{u}_{n-1}, \ldots, \boldsymbol{u}_0\Big].$$

This is known as the *Markov property* for a sequence $\boldsymbol{u}_n$ of random variables.

Moving to continuous time, we are primarily interested in stochastic processes with the Markov property, so-called *Markov processes*, such as the solution $\{\boldsymbol{u}(t)\colon t \ge 0\}$ of (8.3). We give the formal definition.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Definition 8.35 (transition function, Markov process) Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space. Let $B_b(\mathbb{R}^d)$ denote the set of bounded Borel measurable functions $\phi\colon \mathbb{R}^d \to \mathbb{R}$. For an $\mathbb{R}^d$-valued $\mathcal{F}_t$-adapted stochastic process $\{\boldsymbol{u}(t)\colon t \geq 0\}$, the *transition functions* are defined by

$$P_{s,t}\phi(\boldsymbol{x}) := \mathbb{E}\Big[\phi(\boldsymbol{u}(t))\,\Big|\,\boldsymbol{u}(s) = \boldsymbol{x}\Big] \qquad \text{for } 0 \leq s \leq t \text{ and } \phi \in B_b(\mathbb{R}^d).$$

We say $\{\boldsymbol{u}(t)\colon t \geq 0\}$ is a *Markov process* if a.s.

$$\mathbb{E}\Big[\phi(\boldsymbol{u}(t+h))\,\Big|\,\mathcal{F}_t\Big] = P_{t,t+h}\phi(\boldsymbol{u}(t)) \qquad \text{for } \phi \in B_b(\mathbb{R}^d) \text{ and } h,t \geq 0. \tag{8.69}$$

Further, we say the Markov process is *time homogeneous* if $P_{t,t+h}$ is independent of t and write $P_h = P_{0,h}$.

The drift $\boldsymbol{f}$ and diffusion G are functions of $\boldsymbol{u} \in \mathbb{R}^d$ and independent of time t. Moreover, the increments of Brownian motion $W(t)$ for $t > s$ are independent of $\mathcal{F}_s$. From the integral equation (8.3), we see that the conditional distribution of the solution $\boldsymbol{u}(t+h)$ of (8.3) given $\mathcal{F}_t$ depends only on $\boldsymbol{u}(t)$ and $h > 0$. As such, (8.69) holds and we see the solution $\boldsymbol{u}(t)$ of (8.3) is a time-homogeneous Markov processes.

Theorem 8.36 (Markov processes for SODEs) *Under Assumption 8.17, the solution* $\boldsymbol{u}(t)$ *of* (8.3) *is a time-homogeneous Markov process.*

Time-homogeneous Markov processes have an important semigroup property (see Definition 3.10) that we now introduce.

Lemma 8.37 (Chapman–Kolmogorov) *If* $\boldsymbol{u}(t)$ *is a* $\mathbb{R}^d$*-valued time-homogeneous Markov process, then*

$$P_{s+t}\,\phi = P_s P_t\,\phi, \qquad \forall s,t \geq 0, \quad \phi \in B_b(\mathbb{R}^d).$$

This is known as the Chapman–Kolmogorov *equation.*

Proof By the tower property of conditional expectation, a.s.,

$$\begin{aligned} P_{s+t}\phi(\boldsymbol{x}) &= \mathbb{E}\Big[\phi(\boldsymbol{u}(s+t))\,\Big|\,\boldsymbol{u}(0) = \boldsymbol{x}\Big] = \mathbb{E}\Big[\mathbb{E}\big[\phi(\boldsymbol{u}(s+t))\,\big|\,\mathcal{F}_s\big]\,\Big|\,\boldsymbol{u}(0) = \boldsymbol{x}\Big] \\ &= \mathbb{E}\Big[P_{s,s+t}\,\phi(\boldsymbol{u}(s))\,\Big|\,\boldsymbol{u}(0) = \boldsymbol{x}\Big] = \mathbb{E}\Big[P_t\,\phi(\boldsymbol{u}(s))\,\Big|\,\boldsymbol{u}(0) = \boldsymbol{x}\Big]. \end{aligned}$$

If $\psi(\boldsymbol{x}) := P_t\phi(\boldsymbol{x})$, then $P_{s+t}\phi(\boldsymbol{x}) = \mathbb{E}\big[\psi(\boldsymbol{u}(s))\,\big|\,\boldsymbol{u}(0) = \boldsymbol{x}\big] = P_s\psi(\boldsymbol{x})$ and hence $P_{s+t}\phi = P_s P_t \phi$. □

Backward Kolmogorov equation

To understand the approximation of $\mathbb{E}[\phi(\boldsymbol{u}(t))]$ for the solution $\boldsymbol{u}(t)$ of (8.3) with initial condition $\boldsymbol{u}(0) = \boldsymbol{u}_0$, the function $\Phi(t,\boldsymbol{u}_0) := \mathbb{E}[\phi(\boldsymbol{u}(t))]$ plays a key role. Because $\Phi(t,\boldsymbol{u}_0) = P_t\phi(\boldsymbol{u}_0)$ (see Definition 8.35), the Chapman–Kolmogorov equation applies and we can analyse $\Phi(t,\boldsymbol{u}_0)$ via a PDE, known as the *backward Kolmogorov equation*. We start by differentiating $\boldsymbol{u}(t)$ with respect to the initial data $\boldsymbol{u}_0$. Because we are looking *back* from time $t > 0$ to see how $\boldsymbol{u}(t)$ varies with respect to the initial data $\boldsymbol{u}(0) = \boldsymbol{u}_0$, we have the name *backward* Kolmogorov equation.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Assumption 8.38 Suppose that $\boldsymbol{f} \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$, $G \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^{d\times m})$, and all derivatives of $\boldsymbol{f}$ and G of order one and higher are bounded.

Assumption 8.38 implies the Lipschitz condition and hence also Assumption 8.17.

Lemma 8.39 (dependence on initial data) *Let Assumption 8.38 hold. The nth derivative $D^n\boldsymbol{u}(t)$ of $\boldsymbol{u}(t)$ with respect to the initial data $\boldsymbol{u}_0$ is a bounded linear operator for any $n \in \mathbb{N}$. Further, for each $T > 0$, there exists $K > 0$ such that for $0 \le t \le T$*

$$\mathbb{E}\left[\left\|D^n\boldsymbol{u}(t)(\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_n)\right\|_2^2\right] \le K\left\|\boldsymbol{\xi}_1\right\|_2 \cdots \left\|\boldsymbol{\xi}_n\right\|_2, \qquad \forall \boldsymbol{u}_0 \in \mathbb{R}^d,\ \boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_n \in \mathbb{R}^d.$$

Proof We prove this for $n = 1$; the general case is similar. Denote the matrix of first-order partial derivatives (the Jacobian) of $\boldsymbol{u}(t)$ with respect to $\boldsymbol{u}_0$ by $J(t) \coloneqq d\boldsymbol{u}(t)/d\boldsymbol{u}_0$. By taking the derivative of the integral equation (8.3) in the direction $\boldsymbol{\xi} \in \mathbb{R}^d$,

$$J(t)\boldsymbol{\xi} = \boldsymbol{\xi} + \int_0^t D\boldsymbol{f}(\boldsymbol{u}(s))J(s)\boldsymbol{\xi}\,ds + \int_0^t DG(\boldsymbol{u}(s))J(s)\boldsymbol{\xi}\,d\boldsymbol{W}(s). \tag{8.70}$$

Applying the Itô isometry (8.22) to the last term,

$$\begin{aligned}\mathbb{E}\left[\left\|J(t)\boldsymbol{\xi}\right\|_2^2\right] \le 3\left\|\boldsymbol{\xi}\right\|_2^2 &+ 3t\int_0^t \mathbb{E}\left[\left\|D\boldsymbol{f}(\boldsymbol{u}(s))J(s)\boldsymbol{\xi}\right\|_2^2\right] ds \\ &+ 3\int_0^t \mathbb{E}\left[\left\|DG(\boldsymbol{u}(s))J(s)\boldsymbol{\xi}\right\|_{\mathrm{F}}^2\right] ds.\end{aligned}$$

Because the derivatives are bounded, for a constant $K > 0$,

$$\mathbb{E}\left[\left\|J(t)\boldsymbol{\xi}\right\|_2^2\right] \le 3\left\|\boldsymbol{\xi}\right\|_2^2 + 3Kt\int_0^t \mathbb{E}\left[\left\|J(s)\boldsymbol{\xi}\right\|_2^2\right] ds + 3K\int_0^t \mathbb{E}\left[\left\|J(s)\boldsymbol{\xi}\right\|_2^2\right] ds.$$

Gronwall's inequality completes the proof. □

Definition 8.40 (polynomial growth) A function $\phi\colon \mathbb{R}^d \to \mathbb{R}$ has *polynomial growth* if $\sup_{\boldsymbol{u}_0\in\mathbb{R}^d} |\phi(\boldsymbol{u}_0)|/(1 + \|\boldsymbol{u}_0\|_2^p) < \infty$, for some $p \in \mathbb{N}$. Let $\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d)$ denote the set of $\phi \in \mathrm{C}^\infty(\mathbb{R}^d)$ such that ϕ and all its derivatives have polynomial growth. Let $C^\infty_{\mathrm{poly}}([0,T]\times\mathbb{R}^d)$ denote the set of $\Phi \in \mathrm{C}^\infty([0,T]\times\mathbb{R}^d)$ such that

$$\sup_{t\in[0,T]}\left|\frac{\partial^{|\boldsymbol{\alpha}|}}{\partial^{\alpha_1}u_1\dots\partial^{\alpha_d}u_d}\Phi(t,\boldsymbol{u}_0)\right|$$

has polynomial growth for any multi-index $\boldsymbol{\alpha}$.

We know from Exercise 4.10 that every moment of a Gaussian random variable is well defined. In the next result, we show that the moments $\mathbb{E}\left[\|\boldsymbol{u}(t)\|_2^p\right]$ of solutions $\boldsymbol{u}(t)$ of SODEs are finite and this ensures $\mathbb{E}[\phi(\boldsymbol{u}(t))]$ is finite when ϕ has polynomial growth. For this reason, polynomial growing functions are a natural choice for test functions in weak approximation.

Corollary 8.41 *Let Assumption 8.38 hold. Suppose that $\phi \in C^\infty_{\mathrm{poly}}(\mathbb{R}^d)$ and let $\Phi(t,\boldsymbol{u}_0) = \mathbb{E}[\phi(\boldsymbol{u}(t))]$ where $\boldsymbol{u}(t)$ satisfies* (8.3). *Then $\Phi \in C^\infty_{\mathrm{poly}}([0,T]\times\mathbb{R}^d)$.*

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Proof As we showed in Lemma 8.39, $J(t) = d\boldsymbol{u}(t)/d\boldsymbol{u}_0$ is uniformly bounded on $[0,T]$ by a constant K. Then, for $\boldsymbol{\xi} \in \mathbb{R}^d$ and $t \in [0,T]$,

$$\begin{aligned}\left|\nabla\Phi(t,\boldsymbol{u}_0)^\mathsf{T}\boldsymbol{\xi}\right| &= \left|\mathbb{E}\left[\nabla\phi(\boldsymbol{u}(t))^\mathsf{T} J(t)\boldsymbol{\xi}\right]\right| \\ &\le \left\|\nabla\phi(\boldsymbol{u}(t))\right\|_{L^2(\Omega,\mathbb{R}^d)}\left\|J(t)\boldsymbol{\xi}\right\|_{L^2(\Omega,\mathbb{R}^d)} \le K\left\|\nabla\phi(\boldsymbol{u}(t))\right\|_{L^2(\Omega,\mathbb{R}^d)}\|\boldsymbol{\xi}\|_2.\end{aligned}$$

As $\phi \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d)$, we have for some integer $p \ge 2$ that

$$\sup_{0\le t\le T}\left|\nabla\Phi(t,\boldsymbol{u}_0)^\mathsf{T}\boldsymbol{\xi}\right| \le \sup_{0\le t\le T} K\left(1+\left\|\boldsymbol{u}(t)\right\|^p_{L^2(\Omega,\mathbb{R}^d)}\right)\|\boldsymbol{\xi}\|_2.$$

This is finite for $p = 2$ by Theorem 8.18 and, for $p > 2$, by Exercise 8.18. This argument shows that the first order derivatives of $\Phi(t,\boldsymbol{u}_0)$ have polynomial growth in the sense of Definition 8.40. The argument can be extended to higher-order derivatives to show $\Phi \in \mathrm{C}^\infty_{\mathrm{poly}}([0,T],\mathbb{R}^d)$. □

The final tool for deriving the backward Kolmogorov equation is the Itô formula, now for $d, m > 1$. Denote by $\mathrm{C}^{1,2}([0,T]\times\mathbb{R}^d)$ the set of functions $\Phi(t,\boldsymbol{u}_0)$ that are continuously differentiable in $t \in [0,T]$ and twice continuously differentiable in $\boldsymbol{u}_0 \in \mathbb{R}^d$.

Lemma 8.42 (Itô formula) *If $\Phi \in \mathrm{C}^{1,2}([0,T]\times\mathbb{R}^d)$ and $\boldsymbol{u}(t)$ denotes the solution of* (8.3) *under Assumption 8.17, then*

$$\Phi(t,\boldsymbol{u}(t)) = \Phi(0,\boldsymbol{u}_0) + \int_0^t \left(\frac{\partial}{\partial t} + \mathcal{L}\right)\Phi(s,\boldsymbol{u}(s))\,ds + \sum_{k=1}^m \int_0^t \mathcal{L}^k\Phi(s,\boldsymbol{u}(s))\,dW_k(s)$$

where, for $\boldsymbol{x} \in \mathbb{R}^d$ and $t > 0$,

$$\mathcal{L}\Phi(t,\boldsymbol{x}) \coloneqq \boldsymbol{f}(\boldsymbol{x})^\mathsf{T}\nabla\Phi(t,\boldsymbol{x}) + \frac{1}{2}\sum_{k=1}^m \boldsymbol{g}_k(\boldsymbol{x})^\mathsf{T}\nabla^2\Phi\,(t,\boldsymbol{x})\boldsymbol{g}_k(\boldsymbol{x}), \tag{8.71}$$

$$\mathcal{L}^k\Phi(t,\boldsymbol{x}) \coloneqq \nabla\Phi(t,\boldsymbol{x})^\mathsf{T}\boldsymbol{g}_k(\boldsymbol{x}). \tag{8.72}$$

Here $\boldsymbol{g}_k$ denotes the kth column of the diffusion matrix G. $\nabla\Phi$ is the gradient and $\nabla^2\Phi$ the Hessian matrix of second partial derivatives of $\Phi(t,\boldsymbol{x})$ with respect to $\boldsymbol{x}$.

Proposition 8.43 (backward Kolmogorov equation) *Suppose that Assumption 8.38 holds and $\phi \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d)$. Let $\Phi(t,\boldsymbol{u}_0) = \mathbb{E}[\phi(\boldsymbol{u}(t))]$. Then $\Phi \in \mathrm{C}^{1,2}([0,T]\times\mathbb{R}^d)$ and*

$$\frac{\partial}{\partial t}\Phi = \mathcal{L}\Phi, \qquad \Phi(0,\boldsymbol{u}_0) = \phi(\boldsymbol{u}_0), \tag{8.73}$$

where $\mathcal{L}$ is known as the generator and is defined by (8.71).

Proof $\boldsymbol{u}(t)$ is a Markov process by Theorem 8.36 and $\Phi(t,\boldsymbol{u}_0) = P_t\phi(\boldsymbol{u}_0)$ using the transition function of Definition 8.35. By the Chapman–Kolmogorov equation of Lemma 8.37,

$$\frac{\Phi(t+h,\boldsymbol{u}_0)-\Phi(t,\boldsymbol{u}_0)}{h} = \frac{P_{t+h}\phi(\boldsymbol{u}_0)-P_t\phi(\boldsymbol{u}_0)}{h} = \frac{P_h\psi(\boldsymbol{u}_0)-\psi(\boldsymbol{u}_0)}{h}$$

for $\psi(\boldsymbol{u}_0) \coloneqq P_t\phi(\boldsymbol{u}_0)$. If the limit exists as $h \to 0$,

$$\frac{\partial}{\partial t}\Phi = \lim_{h\to 0}\frac{P_h\psi-\psi}{h}.$$

To show (8.73), we prove the limit is well defined and equals $\mathcal{L}\Phi$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Note that $\psi(\boldsymbol{u}_0) = \Phi(t, \boldsymbol{u}_0)$ has two derivatives by Corollary 8.41 and hence the Itô formula applies. Then,

$$\psi(\boldsymbol{u}(h)) = \psi(\boldsymbol{u}_0) + \int_0^h \mathcal{L}\psi(\boldsymbol{u}(s))\,ds + \sum_{k=1}^m \int_0^h \mathcal{L}^k\psi(\boldsymbol{u}(s;\boldsymbol{u}_0))\,dW_k(s).$$

Rearranging,

$$\frac{\psi(\boldsymbol{u}(h)) - \psi(\boldsymbol{u}_0)}{h} = \frac{1}{h}\int_0^h \mathcal{L}\psi(\boldsymbol{u}(s))\,ds + \text{an Itô integral}.$$

Taking expectations and the limit $h \to 0$, the Itô integral vanishes and, using $\mathbb{E}[\psi(\boldsymbol{u}(h))] = P_h\psi(\boldsymbol{u}_0)$, we have

$$\lim_{h\to 0} \frac{P_h\psi(\boldsymbol{u}_0) - \psi(\boldsymbol{u}_0)}{h} = \mathbb{E}\big[\mathcal{L}\psi(\boldsymbol{u}(0;\boldsymbol{u}_0))\big] = \mathcal{L}\psi(\boldsymbol{u}_0), \tag{8.74}$$

as required. □

Example 8.44 (heat equation) Fix $T > 0$ and let

$$X^x(t) \coloneqq x + W(T-t), \qquad \Psi(T,x) \coloneqq \mathbb{E}\big[\phi(X^x(0))\big].$$

As we saw in Example 5.14, Ψ obeys the heat equation

$$\Psi_t = \tfrac{1}{2}\Psi_{xx} \tag{8.75}$$

with initial condition $\Psi(0,x) = \phi(x)$. In terms of Proposition 8.43, $\Psi(T,x) = \mathbb{E}[\phi(u(T))]$ for $u(T) = x + W(T)$ and $u(t)$ solves the SODE (8.3) with $f = 0$ and $G = 1$. Then, the generator is $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$ and the backward Kolmogorov equation (8.73) is the same as (8.75) with $u_0 = x$.

Weak convergence of Euler–Maruyama

With the help of the backward Kolmogorov equation, we prove the weak discretisation error $\mathbb{E}[\phi(\boldsymbol{u}(T))] - \mathbb{E}[\phi(\boldsymbol{u}_N)]$ for the Euler–Maruyama approximation $\boldsymbol{u}_N$ to $\boldsymbol{u}(T)$ is $\mathcal{O}(\Delta t)$ for infinitely differentiable functions ϕ with polynomial growth.

Theorem 8.45 (Euler–Maruyama convergence) *Let Assumption 8.38 hold. For all $\phi \in C^\infty_{\text{poly}}(\mathbb{R}^d)$, $T \geq 0$, and $\boldsymbol{u}_0 \in \mathbb{R}^d$, there exists $K > 0$ independent of Δt such that*

$$\Big|\mathbb{E}\big[\phi(\boldsymbol{u}(T))\big] - \mathbb{E}\big[\phi(\boldsymbol{u}_N)\big]\Big| \leq K\Delta t, \qquad t_N = T, \tag{8.76}$$

where $\boldsymbol{u}_n$ is the Euler–Maruyama approximation to the solution $\boldsymbol{u}(t_n)$ of (8.3).

Proof Recall from §8.5 that $\hat{t} \coloneqq t_n$ if $t \in [t_n, t_{n+1})$. Let $\boldsymbol{u}_{\Delta t}(t)$ denote the solution of

$$d\boldsymbol{u}_{\Delta t}(t) = \boldsymbol{f}(\boldsymbol{u}_{\Delta t}(\hat{t}))\,dt + G(\boldsymbol{u}_{\Delta t}(\hat{t}))\,d\boldsymbol{W}(t), \qquad \boldsymbol{u}_{\Delta t}(0) = \boldsymbol{u}_0. \tag{8.77}$$

Because the drift and diffusion are constant on intervals $[t_n, t_{n+1})$, $\boldsymbol{u}_{\Delta t}(t_n) = \boldsymbol{u}_n$ for $n \in \mathbb{N}$ and $\boldsymbol{u}_{\Delta t}(t)$ interpolates the numerical method.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

By applying the Itô formula to (8.77) on intervals $[t_n, t_{n+1})$, we see that

$$d\Phi(t, \boldsymbol{u}_{\Delta t}(t)) = \left(\frac{\partial}{\partial t} + \widehat{\mathcal{L}}(t)\right)\Phi(t, \boldsymbol{u}_{\Delta t}(t))\, dt + \sum_{k=1}^{m} \widehat{\mathcal{L}}^k(t)\Phi(t, \boldsymbol{u}_{\Delta t}(t))\, dW_k(t), \tag{8.78}$$

where

$$\widehat{\mathcal{L}}(t)\Phi(t,\boldsymbol{x}) \coloneqq \boldsymbol{f}(\boldsymbol{u}(\hat{t}))^{\mathsf{T}}\nabla\Phi(t,\boldsymbol{x}) + \frac{1}{2}\sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{u}(\hat{t}))^{\mathsf{T}}\nabla^2\Phi(t,\boldsymbol{x})\boldsymbol{g}_k(\boldsymbol{u}(\hat{t})),$$

$$\widehat{\mathcal{L}}^k(t)\Phi(t,\boldsymbol{x}) \coloneqq \nabla\Phi(t,\boldsymbol{x})^{\mathsf{T}}\boldsymbol{g}_k(\boldsymbol{u}(\hat{t})), \qquad k = 1,\dots,m.$$

Let $e \coloneqq \mathbb{E}[\phi(\boldsymbol{u}(T)) - \phi(\boldsymbol{u}_{\Delta t}(T))]$ and $\Phi(t,\boldsymbol{u}_0) \coloneqq \mathbb{E}[\phi(\boldsymbol{u}(t))]$. Then,

$$e = \Phi(T, \boldsymbol{u}_0) - \Phi(0, \boldsymbol{u}_{\Delta t}(T))$$

and we must show $e = \mathcal{O}(\Delta t)$. By Proposition 8.43, $\Phi \in \mathrm{C}^{1,2}([0,T]\times\mathbb{R}^d)$ and the Itô formula (8.78) gives

$$\begin{aligned} e &= \Phi(T, \boldsymbol{u}_{\Delta t}(0)) - \Phi(0, \boldsymbol{u}_{\Delta t}(T)) \\ &= \mathbb{E}\left[\int_0^T \frac{\partial}{\partial t}\Phi(T-s, \boldsymbol{u}_{\Delta t}(s)) + \widehat{\mathcal{L}}(s)\Phi(T-s, \boldsymbol{u}_{\Delta t}(s))\, ds\right]. \end{aligned} \tag{8.79}$$

Changing $t \mapsto T - t$ in the backward Kolmogorov equation (8.73),

$$-\frac{\partial}{\partial t}\Phi(T-t,\boldsymbol{x}) = \mathcal{L}\Phi(T-t,\boldsymbol{x}). \tag{8.80}$$

Let $\boldsymbol{x} = \boldsymbol{u}_{\Delta t}(s)$ and substitute (8.80) into (8.79), to give

$$e = \mathbb{E}\left[\int_0^T \left(\widehat{\mathcal{L}}(s)\Phi(T-s, \boldsymbol{u}_{\Delta t}(s)) - \mathcal{L}\Phi(T-s, \boldsymbol{u}_{\Delta t}(s))\right) ds\right].$$

For simplicity, we treat only the case $\boldsymbol{f} = \boldsymbol{0}$. Then,

$$\mathcal{L}\phi = \frac{1}{2}\sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x})^{\mathsf{T}}\nabla^2\phi\, \boldsymbol{g}_k(\boldsymbol{x}),$$

$$\mathcal{L}(s)\phi = \frac{1}{2}\sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{u}(\hat{s}))^{\mathsf{T}}\nabla^2\phi\, \boldsymbol{g}_k(\boldsymbol{u}(\hat{s})).$$

We can write

$$\begin{aligned} \widehat{\mathcal{L}}(s)\Phi(T-s, \boldsymbol{u}_{\Delta t}(s)) - \mathcal{L}\Phi(T-s, \boldsymbol{u}_{\Delta t}(s)) &= \Theta_1(s, \boldsymbol{u}_{\Delta t}(s)) - \Theta_1(\hat{s}, \boldsymbol{u}_{\Delta t}(\hat{s})) \\ &\quad + \Theta_2(\hat{s}, \boldsymbol{u}_{\Delta t}(\hat{s})) - \Theta_2(s, \boldsymbol{u}_{\Delta t}(s)) \end{aligned}$$

where

$$\Theta_1(s,\boldsymbol{x}) \coloneqq \sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{u}_{\Delta t}(\hat{s}))^{\mathsf{T}}\nabla^2\Phi(T-s,\boldsymbol{x})\boldsymbol{g}_k(\boldsymbol{u}_{\Delta t}(\hat{s})), \tag{8.81}$$

$$\Theta_2(s,\boldsymbol{x}) \coloneqq \sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x})^{\mathsf{T}}\nabla^2\Phi(T-s,\boldsymbol{x})\boldsymbol{g}_k(\boldsymbol{x}). \tag{8.82}$$

By the Itô formula (8.78), for $i = 1, 2$,

$$\Theta_i(s, \boldsymbol{u}_{\Delta t}(s)) = \Theta_i(\hat{s}, \boldsymbol{u}_{\Delta t}(\hat{s})) + \int_{\hat{s}}^{s} \left(\frac{\partial}{\partial t} + \widehat{\mathcal{L}}(r)\right)\Theta_i(r, \boldsymbol{u}_{\Delta t}(r))\, dr + \text{Itô integral}.$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Let

$$Q_i(r,\boldsymbol{u}_0) \coloneqq \left(\frac{\partial}{\partial t} + \widehat{\mathcal{L}}(r)\right)\Theta_i(r,\boldsymbol{u}_0) \tag{8.83}$$

so that

$$\mathbb{E}\big[\Theta_i(s,\boldsymbol{u}_{\Delta t}(s))\big] = \mathbb{E}\big[\Theta_i(\hat{s},\boldsymbol{u}_{\Delta t}(\hat{s}))\big] + \int_{\hat{s}}^{s} \mathbb{E}\big[Q_i(r,\boldsymbol{u}_{\Delta t}(r))\big]\,dr.$$

From (8.73),

$$\frac{\partial}{\partial s}\nabla^2\Phi(T-s,\boldsymbol{x}) = \nabla^2\frac{\partial}{\partial s}\Phi(T-s,\boldsymbol{x}) = -\nabla^2\mathcal{L}\Phi(T-s,\boldsymbol{x})$$

and hence, referring to (8.81) and (8.82), the time derivative in (8.83) can be changed to a spatial derivative. Therefore, as $\Theta_i \in \mathrm{C}^\infty_{\mathrm{poly}}([0,T]\times\mathbb{R}^d)$, so does Q_i and

$$\mathbb{E}\big[Q_i(r,\boldsymbol{u}_{\Delta t})\big] \le K\mathbb{E}\big[1 + \|\boldsymbol{u}_{\Delta t}(r)\|_2^p\big].$$

for some K,p independent of Δt. For $p=2$, this is bounded uniformly for $r\in[0,T]$ because of Exercise 8.11. It also generalises to $p>2$ as we saw in Exercise 8.18. Finally, because

$$e = \int_0^T \int_{\hat{s}}^{s} \mathbb{E}\big[Q_1(r,\boldsymbol{u}_{\Delta t}(r))\big] + \mathbb{E}\big[Q_2(r,\boldsymbol{u}_{\Delta t}(r))\big]\,dr\,ds, \tag{8.84}$$

and $|s-\hat{s}|\le\Delta t$, we gain (8.76) as required. □

Multilevel Monte Carlo

Consider the cost of computing the Monte Carlo and Euler–Maruyama approximation μ_M in (8.66). To approximate $\mathbb{E}[\phi(\boldsymbol{u}(T))]$ to an accuracy of ϵ, both the Euler–Maruyama weak discretisation error and the Monte Carlo sampling error should be $\mathcal{O}(\epsilon)$ in (8.67). Under the assumptions of Theorem 8.45, the weak discretisation error is $\mathcal{O}(\Delta t)$ so we require $\Delta t = \mathcal{O}(\epsilon)$. From (4.34), we also know the Monte Carlo error is $\mathcal{O}(1/\sqrt{M})$ so we require $M^{-1} = \mathcal{O}(\epsilon^2)$. We measure the computational cost by counting the total number of steps taken by the Euler–Maruyama method. Finding one sample of $\boldsymbol{u}_N$ requires $T/\Delta t$ steps and finding M samples requires $MT/\Delta t$ steps. Thus, by taking M proportional to ϵ^{-2} and Δt proportional to ϵ, we obtain accuracy ϵ with total cost

$$\mathrm{cost}(\mu_M) = M\frac{T}{\Delta t} = \mathcal{O}(\epsilon^{-3}). \tag{8.85}$$

A more efficient method, in many cases, is the multilevel Monte Carlo method. The idea is to use a hierarchy of time steps Δt_ℓ at levels $\ell = \ell_0,\dots,L$ rather than a fixed time step Δt. To be specific, we choose $\Delta t_\ell = \kappa^{-\ell}T$ for some $\kappa\in\{2,3,\dots\}$. At each level, we compute the Euler–Maruyama approximation $\boldsymbol{u}_{\kappa^\ell}$ to $\boldsymbol{u}(T)$ using the time step Δt_ℓ. To simplify notation, we write $\boldsymbol{U}_\ell$ for $\boldsymbol{u}_{\kappa^\ell}$. The multilevel Monte Carlo method exploits the telescoping sum

$$\mathbb{E}\big[\phi(\boldsymbol{U}_L)\big] = \mathbb{E}\Big[\phi(\boldsymbol{U}_{\ell_0})\Big] + \sum_{\ell=\ell_0+1}^{L} \mathbb{E}\big[\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})\big]. \tag{8.86}$$

The expectation on level L (with the smallest step Δt_L) is written as the expectation on level

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

ℓ_0 (with the largest step Δt_{ℓ_0}) plus a sum of corrections. Each correction is the difference of two approximations to $\boldsymbol{u}(T)$, one with time step Δt_ℓ (known as the *fine* step) and the other with time step $\Delta t_{\ell-1}$ (known as the *coarse* step). The idea is to estimate each of the expectations on the right-hand side of (8.86) separately. To estimate $\mathbb{E}[\phi(\boldsymbol{U}_{\ell_0})]$ at level ℓ_0, we compute the sample average

$$\mu_{\ell_0} \coloneqq \frac{1}{M_{\ell_0}} \sum_{j=1}^{M_{\ell_0}} \phi(\boldsymbol{U}_{\ell_0}^j) \tag{8.87}$$

using M_{ℓ_0} *iid* samples $\boldsymbol{U}_{\ell_0}^j$ of $\boldsymbol{U}_{\ell_0}$. To estimate the correction at level ℓ, we compute

$$\mu_\ell \coloneqq \frac{1}{M_\ell} \sum_{j=1}^{M_\ell} \Big(\phi\Big(\boldsymbol{U}_\ell^{j,\text{fine}}\Big) - \phi\Big(\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}\Big)\Big), \qquad \ell = \ell_0 + 1, \ldots, L, \tag{8.88}$$

where $\boldsymbol{U}_\ell^{j,\text{fine}}$ and $\boldsymbol{U}_\ell^{j,\text{coarse}}$ are *iid* samples of $\boldsymbol{U}_\ell$. The success of the method depends on coupling $\boldsymbol{U}_\ell^{j,\text{fine}}$ and $\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}$ so that $\big\|\boldsymbol{U}_\ell^{j,\text{fine}} - \boldsymbol{U}_{\ell-1}^{j,\text{coarse}}\big\|_{L^2(\Omega,\mathbb{R}^d)}$ is small. This is achieved by computing $\boldsymbol{U}_\ell^{j,\text{fine}}$ and $\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}$ with time steps Δt_ℓ and $\Delta t_{\ell-1}$ respectively and increments from the same Brownian sample path. The increments needed for different j and μ_ℓ are independent. Finally, $\mathbb{E}[\phi(\boldsymbol{u}(T))]$ is estimated by

$$\tilde{\mu}_M \coloneqq \sum_{\ell=\ell_0}^{L} \mu_\ell. \tag{8.89}$$

For the multilevel method, the error can be divided as

$$\mathbb{E}\big[\phi(\boldsymbol{u}(T))\big] - \tilde{\mu}_M = \underbrace{\Big(\mathbb{E}\big[\phi(\boldsymbol{u}(T))\big] - \mathbb{E}\big[\phi(\boldsymbol{U}_L)\big]\Big)}_{\text{weak discretization error}} + \underbrace{\Big(\mathbb{E}\big[\phi(\boldsymbol{U}_L)\big] - \tilde{\mu}_M\Big)}_{\text{multilevel Monte Carlo error}},$$

and the multilevel Monte Carlo error can be further decomposed as

$$\Big(\mathbb{E}\Big[\phi(\boldsymbol{U}_{\ell_0})\Big] - \mu_{\ell_0}\Big) + \sum_{\ell=\ell_0+1}^{L} \Big(\mathbb{E}\big[\big(\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})\big)\big] - \mu_\ell\Big).$$

The multilevel estimator (8.89) is more efficient than (8.66), because we achieve a Monte Carlo error of $\mathcal{O}(\epsilon)$ by computing μ_{ℓ_0} with a large number M_{ℓ_0} of samples (which are cheap to compute due to the larger time step) and by computing the correction terms with a smaller number of samples (which are more costly to compute due to the smaller time step). The error in estimating $\mathbb{E}[\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})]$ by μ_ℓ is proportional to the standard deviation of $\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})$. Since the Euler–Maruyama approximations $\boldsymbol{U}_\ell$ converge to $\boldsymbol{u}(T)$ in $L^2(\Omega,\mathbb{R}^d)$, the standard deviation of $\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})$ decreases as ℓ increases and the time step decreases. Thus, fewer samples M_ℓ are needed to compute μ_ℓ to a given tolerance as ℓ increases. In contrast, a standard Monte Carlo simulation needs to generate a large number of samples with (fixed) smallest time step Δt_L.

In order that the smallest time step $\Delta t_L = \kappa^{-L} T \le \epsilon/2$, we put

$$L = \left\lceil \frac{\log(2T/\epsilon)}{\log(\kappa)} \right\rceil. \tag{8.90}$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

With this choice of L, under the assumptions of Theorem 8.45, the approximation $\boldsymbol{U}_L$ to $\boldsymbol{u}(T)$ has a weak discretisation error of $\mathcal{O}(\epsilon)$. We choose the number of samples M_ℓ at each level to reduce the variance of $\tilde{\mu}_M$, our estimate of $\mathbb{E}[\phi(\boldsymbol{u}(T))]$, to $\mathcal{O}(\epsilon^2)$. This guarantees that the multilevel Monte Carlo error is $\mathcal{O}(\epsilon)$ and the same as the Euler–Maruyama weak discretisation error.

Lemma 8.46 *Let Assumption 8.29 hold. For a globally Lipschitz function $\phi\colon \mathbb{R}^d \to \mathbb{R}$, consider the estimator $\tilde{\mu}_M$ of $\mathbb{E}[\phi(\boldsymbol{u}(T))]$ defined by (8.89). If $M_{\ell_0} = \lceil \epsilon^{-2} \rceil$ and $M_\ell = \lceil \epsilon^{-2}(L-\ell_0)\Delta t_\ell \rceil$ for $\ell = \ell_0 + 1, \dots, L$, then $\operatorname{Var}(\tilde{\mu}_M) = \mathcal{O}(\epsilon^2)$.*

Proof To simplify notation, let $Y := \phi(\boldsymbol{u}(T))$ and $Y_\ell := \phi(\boldsymbol{U}_\ell)$ for $\ell = \ell_0, \dots, L$. To compute $\boldsymbol{U}_\ell^{j,\text{fine}}$ and $\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}$, we use increments from the same Brownain path and hence $\phi(\boldsymbol{U}_\ell^{j,\text{fine}}) - \phi(\boldsymbol{U}_{\ell-1}^{j,\text{coarse}})$ has the same distribution as $Y_\ell - Y_{\ell-1}$. A straightforward upper bound on the variance of a random variable $X \in L^2(\Omega)$ is given by

$$\operatorname{Var}(X) = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}[X]\big)^2 = \|X\|_{L^2(\Omega)}^2 - \big(\mathbb{E}[X]\big)^2 \leq \|X\|_{L^2(\Omega)}^2. \tag{8.91}$$

Using (8.91) and the triangle inequality in $L^2(\Omega)$, we find

$$\begin{aligned}\operatorname{Var}(Y_\ell - Y_{\ell-1}) &\leq \|Y_\ell - Y_{\ell-1}\|_{L^2(\Omega)}^2 = \|Y_\ell - Y + Y - Y_{\ell-1}\|_{L^2(\Omega)}^2 \\ &\leq \big(\|Y_\ell - Y\|_{L^2(\Omega)} + \|Y - Y_{\ell-1}\|_{L^2(\Omega)}\big)^2.\end{aligned}$$

Let L_ϕ denote the Lipschitz constant of ϕ. The $L^2(\Omega, \mathbb{R}^d)$ error for the Euler–Maruyama method is $\mathcal{O}(\Delta t^{1/2})$ (see Exercise 8.11) and hence

$$\|Y_\ell - Y\|_{L^2(\Omega)}^2 \leq L_\phi^2 \|\boldsymbol{U}_\ell - \boldsymbol{u}(T)\|_{L^2(\Omega,\mathbb{R}^d)}^2 \leq K\Delta t_\ell, \tag{8.92}$$

for some constant $K > 0$. Since $\Delta t_{\ell-1} = \kappa \Delta t_\ell$, this gives

$$\operatorname{Var}(Y_\ell - Y_{\ell-1}) \leq K\big(1 + \sqrt{\kappa}\big)^2 \Delta t_\ell.$$

By independence of the samples, we have

$$\begin{aligned}\operatorname{Var}(\mu_\ell) &= \frac{1}{M_\ell^2} \sum_{j=1}^{M_\ell} \operatorname{Var}(Y_\ell - Y_{\ell-1}) = \frac{1}{M_\ell} \operatorname{Var}(Y_\ell - Y_{\ell-1}) \\ &\leq K\big(1 + \sqrt{\kappa}\big)^2 \frac{\Delta t_\ell}{M_\ell} \leq K\big(1 + \sqrt{\kappa}\big)^2 \frac{\epsilon^2}{L - \ell_0}\end{aligned} \tag{8.93}$$

by the choice of M_ℓ. Notice also that $\operatorname{Var}(\mu_{\ell_0}) = \operatorname{Var}(Y_{\ell_0})/M_{\ell_0} \leq \operatorname{Var}(Y_{\ell_0})\epsilon^2$ by (4.31) and the choice of M_{ℓ_0}. Since the samples are independent over different levels,

$$\begin{aligned}\operatorname{Var}(\tilde{\mu}_M) &= \operatorname{Var}(\mu_{\ell_0}) + \sum_{\ell=\ell_0+1}^{L} \operatorname{Var}(\mu_\ell) \\ &\leq \operatorname{Var}(Y_{\ell_0})\epsilon^2 + K\big(1 + \sqrt{\kappa}\big)^2 \epsilon^2\end{aligned} \tag{8.94}$$

and we see that $\operatorname{Var}(\tilde{\mu}_M) = \mathcal{O}(\epsilon^2)$. □

The question now is: what is the computational cost of finding $\tilde{\mu}_M$ using the M_ℓ in the preceding lemma? Again, we measure the cost as the total number of steps the numerical method takes to compute $\tilde{\mu}_M$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Lemma 8.47 *If* $M_{\ell_0} = \lceil \epsilon^{-2} \rceil$, $M_\ell = \lceil \epsilon^{-2}(L - \ell_0)\Delta t_\ell \rceil$ *for* $\ell = \ell_0 + 1, \ldots, L$, *and* L *is given by* (8.90), *the computational cost of finding* $\tilde{\mu}_M$ *in* (8.89) *is* $\mathcal{O}(\epsilon^{-2}|\log \epsilon|^2)$.

Proof The cost of computing one sample on level ℓ_0 is $T\Delta t_{\ell_0}^{-1}$ and we compute M_{ℓ_0} samples, so $\text{cost}(\mu_{\ell_0}) = M_{\ell_0} T \Delta t_{\ell_0}^{-1}$. For μ_ℓ, we need M_ℓ samples and compute both $\boldsymbol{U}_\ell^{j,\text{fine}}$ (needing $T\Delta t_\ell^{-1}$ steps) and $\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}$ (needing $T\Delta t_{\ell-1}^{-1}$ steps). Hence,

$$\text{cost}(\mu_\ell) = M_\ell T(\Delta t_\ell^{-1} + \Delta t_{\ell-1}^{-1}) = M_\ell T \Delta t_\ell^{-1}(1 + \kappa^{-1}),$$

where we have used that $\Delta t_{\ell-1} = \kappa \Delta t_\ell$. Thus, the total cost of finding $\tilde{\mu}_M$ is

$$\text{cost}(\tilde{\mu}_M) = M_{\ell_0} T \Delta t_{\ell_0}^{-1} + \sum_{\ell=\ell_0+1}^{L} M_\ell T \Delta t_\ell^{-1}(1 + \kappa^{-1}). \tag{8.95}$$

Using the definition of M_ℓ, we see $\text{cost}(\tilde{\mu}_M) = \mathcal{O}(\epsilon^{-2}) + \mathcal{O}(L^2\epsilon^{-2})$. Since $L = \mathcal{O}(|\log \epsilon|)$ by (8.90), $\text{cost}(\tilde{\mu}_M) = \mathcal{O}(\epsilon^{-2}|\log \epsilon|^2)$. □

The cost of the multilevel Monte Carlo method is $\mathcal{O}(\epsilon^{-2}|\log \epsilon|^2)$ and is much lower than the cost for the naive Monte Carlo method of $\mathcal{O}(\epsilon^{-3})$ found in (8.85). For example if $\epsilon = 0.001$, we have one order of magnitude in difference in cost between the two methods.

Finally, we show how to choose M_ℓ using the variances

$$V_\ell = \begin{cases} \text{Var}(\phi(\boldsymbol{U}_{\ell_0})), & \ell = \ell_0, \\ \text{Var}(\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})), & \ell = \ell_0 + 1, \ldots, L, \end{cases} \tag{8.96}$$

to achieve the least Monte Carlo error (as measured by $\text{Var}(\tilde{\mu}_M)$) for a given computational cost. The variances V_ℓ may be approximated by sample variances during computations and allow the algorithm to adaptively choose M_ℓ.

Lemma 8.48 *For a fixed computational cost* C, *there exists a* $K_C > 0$ *depending on* C *so that the variance* $\text{Var}(\tilde{\mu}_M)$ *is minimised by taking*

$$M_{\ell_0} = \left\lceil K_C \sqrt{V_{\ell_0} \Delta t_{\ell_0}} \right\rceil, \qquad M_\ell = \left\lceil K_C \sqrt{V_\ell \Delta t_\ell / (1 + \kappa^{-1})} \right\rceil, \tag{8.97}$$

where V_ℓ *are defined by* (8.96). *To achieve* $\text{Var}(\tilde{\mu}_M) = \epsilon^2/2$ *for a tolerance* ϵ, *set*

$$K_C = 2\epsilon^{-2} \left(\sqrt{V_{\ell_0} \Delta t_{\ell_0}^{-1}} + \sum_{\ell=\ell_0+1}^{L} \sqrt{V_\ell \Delta t_\ell^{-1}(1 + \kappa^{-1})} \right). \tag{8.98}$$

Proof By (8.94), $\text{Var}(\tilde{\mu}_M) = \sum_{\ell=\ell_0}^{L} V_\ell / M_\ell$. From (8.95) with the notation $\delta_{\ell_0} = \Delta t_{\ell_0}$ and $\delta_\ell = \Delta t_\ell / (1 + \kappa^{-1})$ for $\ell = \ell_0 + 1, \ldots, L$, we have $\text{cost}(\tilde{\mu}_M) = \sum_{\ell=\ell_0}^{L} T M_\ell / \delta_\ell$. We wish to minimise the following over $M_{\ell_0}, \ldots, M_L \in \mathbb{N}$:

$$\text{Var}(\tilde{\mu}_M) = \sum_{\ell=\ell_0}^{L} \frac{V_\ell}{M_\ell} \qquad \text{subject to} \qquad \text{cost}(\tilde{\mu}_M) = \sum_{\ell=\ell_0}^{L} T \frac{M_\ell}{\delta_\ell} = C.$$

We introduce a Lagrange multiplier λ and treat M_ℓ as real-valued variables, so that the optimality condition is

$$\frac{\partial}{\partial M_\ell} \left(\sum_{k=\ell_0}^{L} \frac{V_k}{M_k} + \lambda \left(\sum_{k=\ell_0}^{L} \frac{M_k}{\delta_k} - \frac{C}{T} \right) \right) = 0, \qquad \ell = \ell_0, \ldots, L.$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Thus, $-M_\ell^{-2}V_\ell + \lambda/\delta_\ell = 0$ and

$$M_\ell = \sqrt{V_\ell \delta_\ell/\lambda}, \qquad \ell = \ell_0, \ldots, L. \tag{8.99}$$

We take $K_C = 1/\sqrt{\lambda}$ to find (8.97). To achieve $\mathrm{Var}(\tilde{\mu}_M) = \epsilon^2/2$, we need

$$\frac{1}{2}\epsilon^2 = \mathrm{Var}(\tilde{\mu}_M) = \sum_{\ell=\ell_0}^{L} \frac{V_\ell}{M_\ell} = \sqrt{\lambda} \sum_{\ell=\ell_0}^{L} \sqrt{V_\ell \delta_\ell^{-1}}$$

and hence $K_C = \frac{1}{\sqrt{\lambda}} = 2\epsilon^{-2} \sum_{\ell=\ell_0}^{L} \sqrt{V_\ell \delta_\ell^{-1}}$. □

The multilevel Monte Carlo method is implemented in Algorithms 8.7 and 8.8 for the Euler–Maruyama method. The user calls Algorithm 8.7 and specifies the underlying initial value problem, final time T, test function ϕ, multilevel parameter κ, maximum time step $\Delta t_{\max}$, and the desired accuracy ϵ. The algorithm then determines the number of levels L by (8.90) and sets the initial level ℓ_0, so that $\Delta t_{\ell_0} \le \Delta t_{\max} < \kappa \Delta t_{\ell_0}$. At each level, 10 sample paths are computed (i.e., $M_\ell = 10$) initially. Then, M_ℓ is increased if required according to (8.98) as sample variances become available to use in place of V_ℓ. The final estimate of

Algorithm 8.7 Code to implement the multilevel Monte Carlo method. Inputs are similar to those in Algorithm 8.1 with additional parameters `kappa`=κ, `epsilon`=ϵ and `DTMX`=$\Delta t_{\max}$. Outputs are the estimate `EPu` of $\mathbb{E}[\phi(\boldsymbol{u}(T))]$ and a vector `M` containing the number of samples taken on each level.

```
function [EPu, M]=mlmc(u0,T,d,m,fhandle,ghandle,kappa,epsilon,DTMX)
Levels=ceil(log(2*T/epsilon)/log(kappa))+1; % (=L+1)
DT=T*kappa.^(-(0:Levels-1))'; % time steps
L0=find(DT<=DTMX, 1); % coarsest level (=ell_0+1)
M=10*ones(Levels,1); % initial samples
Ckappa=(1+1./kappa); S1=zeros(Levels,1); S2=S1; ML=S1; VL=S1;
for j=L0:Levels
  N=kappa^j;
  % get samples for level j, initial pass
  [S1,S2]=getlevel(u0,T,N,d,m,fhandle,ghandle,kappa,M(j),j,L0,S1,S2);
  % estimate variance
  VL(L0:j)=S2(L0:j)./M(L0:j)-(S1(L0:j)./M(L0:j)).^2;
  KC=(2/epsilon^2)*sqrt(VL(L0)/DT(L0)); % KC coarse level only
  if j>L0, % esimate samples required (corrections)
    KC=KC+(2/epsilon^2)*sum(sqrt(VL(L0+1:j)./DT(L0+1:j)*Ckappa));
    ML(L0+1:j)=ceil(KC*sqrt(VL(L0+1:j).*DT(L0+1:j)/Ckappa));
  else % estimate sample required (coarsest level)
    ML(L0)=ceil(KC*sqrt(VL(L0)*DT(L0)));
  end
  for l=L0:j,
    dM=ML(l)-M(l);
    if dM>0 % extra samples needed
      N=kappa^l;
      M(l)=M(l)+dM; % get dM extra samples
      [S1,S2]=getlevel(u0,T,N,d,m,fhandle,ghandle,kappa,dM,l,L0,S1,S2);
    end
  end
end
EPu=sum(S1(L0:Levels)./M(L0:Levels));
```

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Algorithm 8.8 Code to compute the multilevel Monte Carlo contributions at level ℓ. Inputs are as in Algorithm 8.1 with, in addition, parameters `K` $= \kappa$, the number of sample paths `MS`, and the levels `L` $= \ell$ and `L0` $= \ell_0$. `S1` contains $M_\ell \mu_\ell$ and `S2` contains $\sum_{j=1}^{M_\ell} \left(\phi(\boldsymbol{U}_\ell^{j,\text{fine}}) - \phi(\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}) \right)^2$. The outputs are the new values of `S1` and `S2`.

```
function [S1,S2]=getlevel(u00,T,N,d,m,fhandle,ghandle,...
                          kappa,MS,L,L0,S1,S2)
S(1)=0; S(2)=0;
Mstep=10000; % compute samples in blocks
for M=1:Mstep:MS
  MM=min(Mstep,MS-M+1);
  u0=bsxfun(@times,u00,ones(d,MM));
  if L==L0
    % compute Euler-Maruyama samples on the coarsest level
    [t,u]=EMpath(u0, T, N, d, m,fhandle,ghandle,1,MM);
    u=squeeze(u(:,:,end));
    S(1)=S(1)+sum(phi(u));
    S(2)=S(2)+sum(phi(u).^2);
  else % fine levels
    defaultStream = RandStream.getGlobalStream;
    % save state of random number generator
    savedState = defaultStream.State;
    % compute Euler-Maruyama samples
    [t,uu]=EMpath(u0, T, N, d, m,fhandle,ghandle,1,MM);
    uref=squeeze(uu(:,:,end));
    % reset random number generator
    defaultStream.State = savedState;
    % recompute the same samples with large time step
    [t,uu]=EMpath(u0, T, N, d, m,fhandle,ghandle,kappa,MM);
    u=squeeze(uu(:,:,end));
    X=(phi(uref)-phi(u));
    S(1)= S(1)+sum(X);
    S(2)= S(2)+sum(X.^2);
  end
end
S1(L)=S1(L)+S(1);
S2(L)=S2(L)+S(2);

% define the quantity of interest phi
function phiv=phi(v)
phiv=v(end,:);
```

$\tilde{\mu}_M$, given by the output `EPu` of Algorithm 8.7, is found from (8.89). Algorithm 8.7 calls Algorithm 8.8 in order to evaluate $M_\ell \mu_\ell$ and $\sum_{j=1}^{M_\ell} \left(\phi(\boldsymbol{U}_\ell^{j,\text{fine}}) - \phi(\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}) \right)^2$. For $\ell > \ell_0$, Algorithm 8.6 is used to compute the Euler–Maruyama approximations $\boldsymbol{U}_\ell^{j,\text{fine}}$ and $\boldsymbol{U}_{\ell-1}^{j,\text{coarse}}$ of the same sample of $\boldsymbol{u}(T)$ with time steps Δt_ℓ and $\Delta t_{\ell-1}$. For level ℓ_0, Algorithm 8.5 is used to compute the necessary Euler–Maruyama approximations. Notice that the computations in Algorithm 8.8 are split into blocks of size `Mstep`, taken here to be 10,000.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

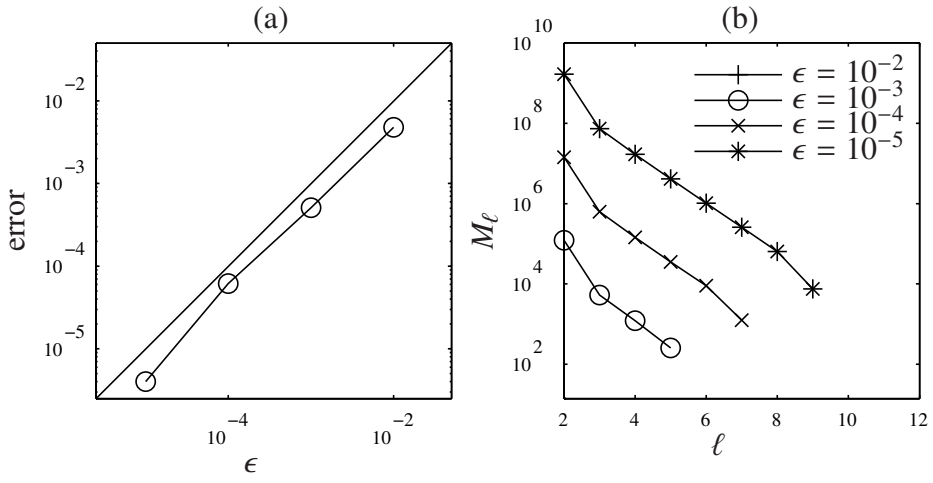

Figure 8.11 The multilevel Monte Carlo method is applied to (8.11) to estimate $\mathbb{E}[u(1)]$ for an accuracy $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}$ and 10^{-5}. In (a), we plot the root-mean-square error in $\mathbb{E}[u(1)]$, averaged over 10 calls of Algorithm 8.7. In (b), the number of samples M_ℓ is plotted against ℓ (for a single call of Algorithm 8.7).

Example 8.49 Consider the geometric Brownian motion SODE (8.11) with $r = -1$ and $\sigma = 0.5$ and suppose we want to estimate $\mathbb{E}[u(T)]$ at $T = 1$. We use the multilevel Monte Carlo and Euler–Maruyama methods with $\kappa = 4$ and $\Delta t_{\max} = 0.2$. The following commands call Algorithm 8.7 and estimate $\mathbb{E}[u(1)]$ with an accuracy of $\epsilon = 10^{-3}$:

```
>> d=1; m=1; r=-1; sigma=0.5; u0=1;
>> kappa=4; T=1; DTMX=0.2; epsilon=1e-3;
>> [EPu, M]=mlmc(u0,T,d,m,@(u) r*u,@(u) sigma*u,kappa,epsilon,DTMX);
```

On a single execution of these commands, we recorded an estimate of $\mathbb{E}[u(1)] \approx 0.3680$ (returned as `EPu`). From (8.34) and (4.9), the true value of $\mathbb{E}[u(1)] = \mathrm{e}^{-1} \approx 0.3679$ and so the error in the estimate is approximately 1.5×10^{-4}. Note, however, that each run of the multilevel Monte Carlo method produces a different estimate. With the stated choices for κ, ϵ, and $\Delta t_{\max}$, we have $L = 5$ and $\ell_0 = 2$. On level 2, the time step is $\Delta t_2 = 0.0625$ and 113,089 samples of the Euler–Maruyama approximation were computed. In contrast, on level 5, where the time step is much smaller, only 96 samples were taken.

In Figure 8.11(a), we examine the root-mean-square error in the multilevel Monte Carlo approximation for accuracies $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, by averaging over 10 sample runs of Algorithm 8.7. We observe that the estimate is more accurate than the desired accuracy ϵ in each case, as the computed errors all lie below the solid diagonal line. In Figure 8.11(b), we plot the required numbers of samples M_ℓ against the level ℓ, for $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$; in each case, fewer samples are required as ℓ is increased. Now, for the standard Monte Carlo method, we would expect to have to take $M = 10^4, 10^6, 10^8, 10^{10}$ sample paths, respectively. Thus, we observe roughly one order of magnitude fewer sample paths are required using the multilevel method.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

8.7 Stratonovich integrals and SODEs

In many physical applications, Brownian motion is too irregular a process and the modelling situation is better served by approximating $W(t)$ by a process $W_J(t)$ with a well-defined time derivative $W_J'(t) := dW_J(t)/dt$. We develop this idea for the case $d = 1 = m$ corresponding to the Itô SODE

$$du = f(u)\,dt + g(u)dW(t), \qquad u(0) = u_0, \tag{8.100}$$

for $f, g : \mathbb{R} \to \mathbb{R}$ and a one-dimensional Brownian motion $W(t)$. We replace the differential $dW(t)$ by $W_J'(t)\,dt$ and consider the ODE

$$\frac{dv_J}{dt} = f(v_J) + g(v_J)W_J'(t), \qquad v_J(0) = v_0. \tag{8.101}$$

Notice that $u(t)$ denotes the solution of the Itô SODE (8.100) and $v_J(t)$ denotes the solution of the ODE (8.101). To make our discussion definite, we work on the time interval $[0,1]$ and choose the truncated Karhunen–Loève expansion

$$W_J(t) := \sum_{j=0}^{J} \frac{\sqrt{2}}{(j+\frac{1}{2})\pi}\xi_j \sin((j+\tfrac{1}{2})\pi t), \qquad \xi_j \sim \mathrm{N}(0,1)\ iid, \tag{8.102}$$

for the smooth approximation $W_J(t)$ to $W(t)$ (see Exercise 5.12). For fixed J, the time derivative of any sample path is given by

$$W_J'(t) = \sum_{j=0}^{J} \sqrt{2}\,\xi_j \cos((j+\tfrac{1}{2})\pi t). \tag{8.103}$$

Example 8.50 We must be careful with quantities that depend on $W_J'(t)$ as their limits as $J \to \infty$ may not exist or not converge to the obvious thing. As an example, consider the limit of

$$\int_0^1 W(t)W_J'(t)\,dt \qquad \text{as } J \to \infty.$$

Substituting $W_J'(t)\,dt = \frac{dW_J(t)}{dt}\,dt = dW_J(t)$, it is tempting to believe that $\int_0^1 W(t)\,dW_J(t)$ converges to the Itô integral $\int_0^1 W(t)\,dW(t) = \frac{1}{2}\big(W(1)^2 - 1\big)$ (see Exercise 8.4). However, we show that

$$\int_0^1 W(t)W_J'(t)\,dt \to \frac{1}{2}W(1)^2 \qquad \text{as } J \to \infty.$$

First note from (8.102) and (8.103) that

$$\lim_{J\to\infty} \int_0^1 W(s)W_J'(s)\,ds = 2\sum_{j,k=0}^{\infty} \int_0^1 \frac{\sin((k+\frac{1}{2})\pi t)\xi_k}{(k+\frac{1}{2})\pi}\xi_j \cos((j+\tfrac{1}{2})\pi t)\,dt$$

$$= \sum_{j,k=0}^{\infty} \int_0^1 \frac{\sin((k+\frac{1}{2})\pi t)\xi_k}{(k+\frac{1}{2})\pi}\xi_j \cos((j+\tfrac{1}{2})\pi t) + \frac{\sin((j+\frac{1}{2})\pi t)\xi_j}{(j+\frac{1}{2})\pi}\xi_k \cos((k+\tfrac{1}{2})\pi t)\,dt.$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Integrating by parts,

$$\int_0^1 \frac{\sin((k+\frac{1}{2})\pi t)}{(k+\frac{1}{2})\pi}\cos((j+\tfrac{1}{2})\pi t)+\frac{\sin((j+\frac{1}{2})\pi t)}{(j+\frac{1}{2})\pi}\cos((k+\tfrac{1}{2})\pi t)\,dt = \frac{\sin((k+\frac{1}{2})\pi)\sin((j+\frac{1}{2})\pi)}{(k+\frac{1}{2})(j+\frac{1}{2})\pi^2}.$$

Since

$$\sum_{j,k=0}^{\infty} \frac{\sqrt{2}\sin((j+\frac{1}{2})\pi t)\xi_j}{(j+\frac{1}{2})\pi}\,\frac{\sqrt{2}\sin((k+\frac{1}{2})\pi t)\xi_k}{(k+\frac{1}{2})\pi} = W(1)^2,$$

we have

$$\lim_{J\to\infty}\int_0^1 W(s)W_J'(s)\,ds = \frac{1}{2}W(1)^2.$$

The message here is that limits involving $W_J'(t)$ must be carefully examined and we do not expect the solution $v_J(t)$ of (8.101) to converge to the solution of the Itô SODE (8.100). We now identify the limit of $v_J(t)$ as $J\to\infty$ as the solution of an Itô SODE with a modified drift $\tilde{f}$.

Theorem 8.51 (Wong–Zakai) *Suppose that $f,g\colon\mathbb{R}\to\mathbb{R}$ and Assumption 8.29 holds. Further suppose the function $u\mapsto f(u)/g(u)$ is globally Lipschitz continuous and $g(u)\geq g_0>0$ for all $u\in\mathbb{R}$. Let $v_J(t)$ be the solution of* (8.101) *and $v(t)$ be the solution of*

$$dv = \tilde{f}(v)\,dt + g(v)\,dW(t), \qquad v(0)=v_0 \tag{8.104}$$

for the modified drift $\tilde{f}(v) := f(v)+\frac{1}{2}g(v)g'(v)$. Then,

$$\sup_{0\leq t\leq 1}|v(t)-v_J(t)|\to 0 \qquad \text{as } J\to\infty.$$

Proof Let

$$\phi(x) := \int_0^x \frac{1}{g(y)}\,dy$$

and note that $\phi'(x)=1/g(x)$ and $\phi''(x)=-g'(x)/g(x)^2$. By Itô's formula (Lemma 8.19),

$$d\phi(v) = \left[\phi'(v)\tilde{f}(v)+\frac{1}{2}\phi''(v)g(v)^2\right]dt+\phi'(v)g(v)\,dW(t) = \frac{f(v)}{g(v)}\,dt+\,dW(t),$$

as $\tilde{f}/g+(-g'/g^2)(g^2/2)=f/g$. By the chain rule,

$$\frac{d\phi(v_J)}{dt} = \phi'(v_J)\frac{dv_J}{dt} = \frac{f(v_J)}{g(v_J)}+W_J'(t).$$

In integral form, we have

$$\phi(v(t))-\phi(v_0) = \int_0^t \frac{f(v(s))}{g(v(s))}\,ds+W(t),$$
$$\phi(v_J(t))-\phi(v_0) = \int_0^t \frac{f(v_J(s))}{g(v_J(s))}\,ds+W_J(t).$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

By assumption, $f(u)/g(u)$ is Lipschitz continuous and

$$\left|\phi(x) - \phi(y)\right| = \left|\int_x^y \frac{1}{g(s)}\, ds\right| \geq \frac{1}{g_0}|x - y|.$$

For some constant L,

$$\left|v(t) - v_J(t)\right| \leq g_0\left|\phi(v(t)) - \phi(v_J(t))\right| \leq \int_0^t L\left|v(s) - v_J(s)\right| ds + \left|W(t) - W_J(t)\right|.$$

Gronwall's inequality (see Exercise 3.4) implies that

$$\left|v(t) - v_J(t)\right| \leq \sup_{0\leq s\leq 1}\left|W(s) - W_J(s)\right| \mathrm{e}^{Lt}. \tag{8.105}$$

Finally, the Karhunen–Loève expansion of Brownian motion converges uniformly, because

$$\mathbb{E}\left[\left|\sum_{j=0}^{\infty} \frac{\xi_j}{(j+1/2)\pi}\right|^2\right] = \sum_{j=0}^{\infty} \frac{1}{(j+1/2)^2\pi^2} < \infty$$

and hence $\sum_{j=0}^{\infty} \frac{\xi_j}{(j+1/2)\pi} < \infty$ a.s. This implies

$$\begin{aligned}\sup_{0\leq t\leq 1}\left|W_J(t) - W(t)\right| &= \sup_{0\leq t\leq 1}\left|\sum_{j=J+1}^{\infty} \frac{\xi_j}{(j+1/2)\pi}\sin((j+1/2)\pi t)\right| \\ &\leq \left|\sum_{j=J+1}^{\infty} \frac{\xi_j}{(j+1/2)\pi}\right| \to 0, \qquad J\to\infty.\end{aligned}$$

With (8.105), this completes the proof. □

Stratonovich SODEs for d = 1 = m

Rather than using $W_J(t)$ and $v_J(t)$ to define a limiting process $v(t)$, it is usual to define the *Stratonovich integral* and consider $v(t)$ as the solution of a so-called *Stratonovich SODE*.

Definition 8.52 (Stratonovich integral) Let $W(t)$ be an $\mathcal{F}_t$-Brownian motion (see Definition 8.7) and let $X \in \mathcal{L}_2^T$ (see Definition 8.10). For $t \in [0,T]$, the *Stratonovich integral* $\int_0^t X(s) \circ dW(s)$ is defined by

$$\int_0^t X(s) \circ dW(s) := \int_0^t X(s)\, dW(s) + \frac{1}{2}[X,W]_t$$

where on the right we have an Itô integral and the quadratic variation $[X,W]_t$, defined by the $L^2(\Omega)$ limit

$$[X,W]_t := \lim_{\Delta t\to 0} \sum_{0<t_{n+1}\leq t} \big(X(t_{n+1}) - X(t_n)\big)\big(W(t_{n+1}) - W(t_n)\big), \qquad t_n := n\Delta t.$$

To see the connection with Theorem 8.51, we compute the quadratic variation when $X(t) = \phi(u(t))$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Proposition 8.53 *If $\phi \in \mathrm{C}^2(\mathbb{R})$ and $u(t)$ is the solution of* (8.100),

$$[\phi(u), W]_t = \int_0^t \phi'(u(s))g(u(s))\,ds$$

and

$$\int_0^t \phi(u(s)) \circ dW(s) = \int_0^t \phi(u(s))\,dW(s) + \frac{1}{2}\int_0^t \phi'(u(s))g(u(s))\,ds. \tag{8.106}$$

Proof This is a calculation using Taylor expansions of $\phi(u(t_{n+1}) - \phi(u(t_n))$, similar to the calculations in Lemma 8.30 and Proposition 8.31. See Exercise 8.22. □

We know $v(t)$ satisfies the Itô SODE (8.104), so that

$$v(t) = v_0 + \int_0^t \tilde{f}(v(s))\,ds + \int_0^t g(v(s))\,dW(s).$$

Replacing the Itô integral with a Stratonovich integral using (8.106), we have

$$v(t) = v_0 + \int_0^t \left[\tilde{f}(v(s)) - \tfrac{1}{2}g'(v(s))g(v(s))\right] ds + \int_0^t g(v(s)) \circ dW(s).$$

Since $\tilde{f} = f + g'g/2$, we find

$$v(t) = v_0 + \int_0^t f(v(s))\,ds + \int_0^t g(v(s)) \circ W(s). \tag{8.107}$$

That is, $v(t)$ is the solution of an SODE with the original drift f and diffusion g *if* we treat the stochastic integral in the Stratonovich sense. Equation (8.107) is known as a Stratonovich SODE and also written as

$$dv = f(v)\,dt + g(v) \circ dW(t), \qquad v(0) = v_0. \tag{8.108}$$

To numerically approximate the Stratonovich SODE (8.108), we can transform the equation to an Itô SODE (8.104) with a modified drift $\tilde{f}$ and apply the Euler–Maruyama and Milstein methods derived in §8.4. It is also possible to approximate solutions of Stratonovich SODEs directly and we discuss two basic methods, known as the Stratonovich–Milstein method and Heun's method.

The Stratonovich–Milstein method is derived in the same way as the Milstein method of Definition 8.24, except we interpret the stochastic integrals in the Stratonovich sense. Alternatively, we may write down the Milstein method for the Itô SODE (8.104) with modified drift $\tilde{f}$. For example, the $d = 1$ Milstein method for the Itô SODE (8.104) is

$$v_{n+1} = v_n + \tilde{f}(v_n)\,\Delta t + g(v_n)\,\Delta W_n + \frac{1}{2}g'(v_n)g(v_n)\big(\Delta W_n^2 - \Delta t\big),$$

where $\Delta W_n \sim \mathrm{N}(0, \Delta t)$ *iid*. Substituting $\tilde{f} = f + \frac{1}{2}g'g$ gives the Stratonovich–Milstein method in terms of the coefficients of the Stratonovich SODE (8.108). That is,

$$v_{n+1} = v_n + f(v_n)\,\Delta t + g(v_n)\,\Delta W_n + \frac{1}{2}g'(v_n)g(v_n)\,\Delta W_n^2. \tag{8.109}$$

The approximation v_n converges to $v(t_n)$ for $t_n = T$ fixed and the $L^2(\Omega)$ error is $\mathcal{O}(\Delta t)$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Heun's method is the generalisation of the Euler–Maruyama method to Stratonovich SODEs. It is found by examining the last two terms of the Milstein method. First notice that, for $\delta \in \mathbb{R}$,

$$g(v) + g'(v)g(v)\delta = g\big(v + g(v)\delta\big) - R$$

where the remainder $R = \frac{1}{2}g''(v)\xi^2 g(v)^2\delta^2$, for some $\xi \in [0,1]$. Let $\tilde{v}_{n+1} = v_n + g(v_n)\,\Delta W_n$; then (8.109) becomes

$$v_{n+1} = v_n + f(v_n)\,\Delta t + g(v_n)\,\Delta W_n + \frac{1}{2}\big(g(\tilde{v}_{n+1}) - g(v_n)\big)\,\Delta W_n + R.$$

If we drop the remainder, we have Heun's method

$$v_{n+1} = v_n + f(v_n)\,\Delta t + \frac{1}{2}\big(g(v_n) + g(\tilde{v}_{n+1})\big)\,\Delta W_n,$$

where again $\tilde{v}_{n+1} = v_n + g(v_n)\,\Delta W_n$. It does not depend on the derivative of g and is easier to apply than the Stratonovich–Milstein method. The $L^2(\Omega)$ error is only $\mathcal{O}(\Delta t^{1/2})$ in general.

Stratonovich SODEs for $d, m > 1$

To end this chapter, we briefly review Stratonovich SODEs in dimension $d > 1$ with forcing by m Brownian motions. Consider the initial-value problem

$$d\boldsymbol{v} = \boldsymbol{f}(\boldsymbol{v})\,dt + G(\boldsymbol{v}) \circ d\boldsymbol{W}(t), \qquad \boldsymbol{v}(0) = \boldsymbol{v}_0, \tag{8.110}$$

for $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $G\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$, and $\boldsymbol{W} = [W_1, \ldots, W_m]^\mathsf{T}$ for independent Brownian motions $W_i(t)$. Then, $\boldsymbol{v}(t) = [v_1(t), \ldots, v_d(t)]^\mathsf{T}$ is a solution to (8.110) if the components $v_i(t)$ for $i = 1, \ldots, d$ obey

$$v_i(t) = v_{0,i} + \int_0^t f_i(\boldsymbol{v}(s))\,ds + \sum_{j=1}^m \int_0^t g_{ij}(\boldsymbol{v}(s)) \circ dW_j(s), \qquad t > 0,$$

and the last term is the Stratonovich integral of Definition 8.52. The Milstein method for the Stratonovich SODE (8.110) can be derived by applying the Milstein method of Definition 8.24 to the Itô SODE with an appropriate drift term. In this general case, define $\tilde{\boldsymbol{f}}(\boldsymbol{v}) := \big[\tilde{f}_1(\boldsymbol{v}), \ldots, \tilde{f}_d(\boldsymbol{v})\big]^\mathsf{T}$, where

$$\tilde{f}_i(\boldsymbol{v}) = f_i(\boldsymbol{v}) + \frac{1}{2}\sum_{k=1}^d \sum_{j=1}^m g_{kj}(\boldsymbol{v}) \frac{\partial g_{ij}(\boldsymbol{v})}{\partial u_k}, \qquad i = 1, \ldots, d, \quad \boldsymbol{v} \in \mathbb{R}^d. \tag{8.111}$$

Then the solution $\boldsymbol{v}(t)$ of (8.110) satisfies the Itô SODE

$$d\boldsymbol{v} = \tilde{\boldsymbol{f}}(\boldsymbol{v})\,dt + G(\boldsymbol{v})\,d\boldsymbol{W}(t), \qquad \boldsymbol{v}(0) = \boldsymbol{v}_0.$$

This leads to the following version of Milstein's method for Stratonovich SODEs. We also define Heun's method.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Definition 8.54 (numerical methods for Stratonovich SODEs) For a time step $\Delta t > 0$, the *Stratonovich–Milstein approximation* $\boldsymbol{v}_n = [v_{1n}, \dots, v_{dn}]^\mathsf{T}$ to $\boldsymbol{v}(t_n)$ for $t_n = n\Delta t$ is defined by

$$\begin{aligned} v_{k,n+1} = v_{kn} + f_k(\boldsymbol{v}_n)\,\Delta t + \sum_{j=1}^{m} g_{kj}(\boldsymbol{v}_n)\,\Delta W_{jn} + \frac{1}{2}\sum_{i=1}^{m}\sum_{\ell=1}^{d} \frac{\partial g_{ki}}{\partial v_\ell}(\boldsymbol{v}_n) g_{\ell i}(\boldsymbol{v}_n)\,\Delta W_{in}^2 \\ + \frac{1}{2}\sum_{i<j=1}^{m}\sum_{\ell=1}^{d}\left(\frac{\partial g_{kj}}{\partial u_\ell}(\boldsymbol{v}_n) g_{\ell i}(\boldsymbol{v}_n) + \frac{\partial g_{ki}}{\partial u_\ell}(\boldsymbol{v}_n) g_{\ell j}(\boldsymbol{v}_n)\right)\Delta W_{in}\,\Delta W_{jn} \\ + \frac{1}{2}\sum_{i<j=1}^{m}\sum_{\ell=1}^{d}\left(\frac{\partial g_{kj}}{\partial v_\ell}(\boldsymbol{v}_n) g_{\ell i}(\boldsymbol{v}_n) - \frac{\partial g_{ki}}{\partial v_\ell}(\boldsymbol{v}_n) g_{\ell j}(\boldsymbol{v}_n)\right) A_{ij,n}, \end{aligned}$$

where $\Delta W_{in} \coloneqq \int_{t_n}^{t_{n+1}} dW_i(r)$ and $A_{ij,n} \coloneqq A_{ij}(t_n, t_{n+1})$ are defined by (8.47) (and are independent of the interpretation of the integral as Stratonovich or Itô). The *Heun approximation* $\boldsymbol{v}_n$ to $\boldsymbol{v}(t_n)$ for $t_n = n\Delta t$ is defined by

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + \boldsymbol{f}(\boldsymbol{v}_n)\,\Delta t + \tfrac{1}{2}\big(G(\boldsymbol{v}_n) + G(\tilde{\boldsymbol{v}}_{n+1})\big)\,\Delta \boldsymbol{W}_n$$

and $\tilde{\boldsymbol{v}}_{n+1} = \boldsymbol{v}_n + G(\boldsymbol{v}_n)\,\Delta \boldsymbol{W}_n$.

Finally, the following lemma corresponds to the Itô formula (Lemma 8.42) and shows that the chain rule for Stratonovich calculus resembles the classical chain rule.

Lemma 8.55 (Stratonovich rule) *Suppose that* $\Phi \in \mathrm{C}^{1,2}([0,T] \times \mathbb{R}^d)$ *and* $\boldsymbol{v}(t)$ *is the solution of* (8.110)*. Then*

$$\Phi(t, \boldsymbol{v}(t)) = \Phi(0, \boldsymbol{v}_0) + \int_0^t \left(\frac{\partial}{\partial t} + \mathcal{L}_{\text{strat}}\right)\Phi(s, \boldsymbol{v}(s))\,ds + \sum_{k=1}^{m}\int_0^t \mathcal{L}^k \Phi(s, \boldsymbol{v}(s)) \circ dW_k(s),$$

where $\mathcal{L}_{\text{strat}}\Phi \coloneqq \boldsymbol{f}^\mathsf{T}\nabla\Phi$ *and* $\mathcal{L}^k\Phi \coloneqq \nabla\Phi^\mathsf{T}\boldsymbol{g}_k$ *(cf.* (8.71)*).*

8.8 Notes

SODEs and their mathematical theory are described in many texts, including (Gīhman and Skorohod, 1972; Karatzas and Shreve, 1991; Mao, 2008; Øksendal, 2003; Protter, 2005) and from a physical perspective in (Gardiner, 2009; Nelson, 1967; Öttinger, 1996; van Kampen, 1997). Numerical methods and their analysis are covered in (Graham and Talay, 2013; Kloeden and Platen, 1992; Milstein and Tretyakov, 2004).

Filtered probability spaces are often considered under the *usual conditions*, which is important for working with jump processes. Formally, a filtration $\{\mathcal{F}_t : t \geq 0\}$ is said to be *right continuous* if $\mathcal{F}_t = \mathcal{F}_{t+}$, for $\mathcal{F}_{t+} \coloneqq \{F \in \mathcal{F}_0 : F \in \mathcal{F}_s \text{ for all } s > t\}$. Then, a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfies the *usual conditions* if $(\Omega, \mathcal{F}, \mathbb{P})$ is a *complete* measure space (see Definition 1.17), the filtration $\{\mathcal{F}_t : t \geq 0\}$ is right continuous, and $\mathcal{F}_0$ contains all sets $F \in \mathcal{F}$ with $\mathbb{P}(F) = 0$. A filtration is easily enlarged to satisfy the usual conditions.

In Definition 8.12, the stochastic integral is defined for square-integrable predictable processes, where predictable processes are the limits of left-continuous $\mathcal{F}_t$-adapted processes.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

Equivalently, we may define predictable processes as functions $[0,T]\times\Omega\to\mathbb{R}$ that are measurable with respect to the so-called predictable σ-algebra $\mathcal{P}_T$, which is the smallest σ-algebra containing the sets

$$\{0\}\times F, \text{ for } F\in\mathcal{F}_0, \qquad (s,t]\times F \text{ for } 0\le s<t<T \text{ and } F\in\mathcal{F}_s.$$

Note also $\mathcal{L}_2^T = L^2([0,T]\times\Omega,\mathcal{P}_T,\text{Leb}\times\mathbb{P})$. The definition of a $\mathbb{R}^{d\times m}$-valued predictable process is given in terms of the components, which is equivalent to the process being $\mathcal{P}_T$-measurable as a function $[0,T]\times\Omega\to\mathbb{R}^{d\times m}$. $\mathcal{P}_T$ is a sub σ-algebra of $\mathcal{B}([0,T])\times\mathcal{F}$ and it is clear that predictable processes are jointly measurable as functions of (t,ω). In fact, the stochastic integral may also be developed under the assumption that the process is $\mathcal{F}_t$-adapted and jointly measurable (e.g., Kloeden and Platen, 1992; Øksendal, 2003). It is enough to consider the smaller class of predictable processes in this book.

We have shown convergence of the Euler–Maruyama and Milstein methods in $L^2(\Omega,\mathbb{R}^d)$ subject to smoothness conditions on the drift $\boldsymbol{f}$ and diffusion G (in particular, the global Lipschitz condition). There are a number of extensions to this result. First, the error in the $L^p(\Omega,\mathbb{R}^d)$ norm for any $p\ge 2$ converges to zero with the same rate. The key additional tool required is the Burkholder–Davis–Gundy inequality (Mao, 2008, Theorem 7.3), which says for $p\ge 2$ and a constant $C_p>0$ that, for any $X\in\mathcal{L}_2^T(\mathbb{R}^{d\times m})$,

$$\left\|\sup_{0\le t\le T}\int_0^t X(s)\,d\boldsymbol{W}(s)\right\|^2_{L^p(\Omega,\mathbb{R}^d)} \le C_p\left\|\int_0^T \|X(s)\|_{\text{F}}^2\,ds\right\|_{L^{p/2}(\Omega)} \tag{8.112}$$

Theorem 4.58(iii) then implies convergence in probability and, for $\epsilon>0$, there exists a random variable $K(\omega)>0$ independent of Δt such that

$$\left|\boldsymbol{u}(T,\omega)-\boldsymbol{u}_N(\omega)\right|\le K(\omega)\Delta t^{r-\epsilon}, \qquad T=t_N,$$

where $\boldsymbol{u}_n$ is the Euler–Maruyama (respectively, Milstein) approximation and $r=1/2$ (resp. $r=1$). These results may be proved with fewer assumptions on the drift $\boldsymbol{f}$ and diffusion function G and, in particular, the global Lipschitz conditions can be relaxed. See (Gyöngy, 1998a; Higham, Mao, and Stuart, 2002; Jentzen, Kloeden, and Neuenkirch, 2009a,b; Kloeden and Neuenkirch, 2007). For further work on the stability of numerical methods, see (Buckwar and Sickenberger, 2011; Burrage, Burrage, and Mitsui, 2000; Higham, 2000; Higham, Mao, and Stuart, 2003).

The Euler–Maruyama and Milstein methods are basic time-stepping methods and limited in many situations. In particular, their accuracy is poor due to the order one convergence and the Euler–Maruyama method may diverge for SODEs with non-globally Lipschitz coefficients (Hutzenthaler, Jentzen, and Kloeden, 2010). More sophisticated solution strategies include, for example, Runge–Kutta methods, linear multi-step methods, variable step size control methods, and asymptotically efficient methods; see (Buckwar and Winkler, 2006; Müller-Gronbach and Ritter, 2008; Rößler, 2010). The development of higher-order methods is limited by sampling methods for the iterated integrals. Even for the Milstein methods, the Lévy area is difficult to sample and the existing methods are not quick enough to make the Milstein method efficient in all cases. This is especially true in higher dimensions. For forcing by $m=2$ Brownian motions, (Gaines and Lyons, 1994; Ryden and Wiktorsson, 2001) provide very efficient methods, but correlations between pairs of Brownian motions are significant and Lévy areas cannot be generated pairwise for $m>2$. General methods

using the Karhunen–Loève expansion are developed by Kloeden, Platen, and Wright (1992) and an interesting method for improving the accuracy is given in Wiktorsson (2001).

Weak convergence is reviewed in Talay (1996) and the extrapolation method developed in Talay and Tubaro (1990). Long-time numerical approximation is not usually accurate in the sense of strong approximation. However, weak approximation errors over long times are small in many problems where the underlying SODE is ergodic. See for example (Mattingly, Stuart, and Tretyakov, 2010; Shardlow and Stuart, 2000; Talay, 2002; Talay and Tubaro, 1990).

The multilevel Monte Carlo method for SODEs was first popularised in Giles (2008b). In our presentation, we used time steps $\Delta t_\ell = \kappa^{-\ell} T$ for $\ell = \ell_0, \ldots, L$. The smallest level ℓ_o can be chosen to avoid stability issues arising from large step sizes and the maximum number of levels L is fixed. In Giles (2008b), a value of $\kappa = 4$ is suggested as a compromise between an estimate of the optimal $\kappa \approx 7$ and the decrease in the number of levels resulting increasing κ in (8.90). In other presentations, often $\ell_0 = 0$ (so $\Delta t_0 = T$) and L is not fixed. Numerical estimates of $\mathbb{E}[\phi(\boldsymbol{U}_\ell) - \phi(\boldsymbol{U}_{\ell-1})]$ can be combined with Richardson extrapolation to reduce the error (as in Giles (2008b)). The multilevel Monte Carlo method has received a great deal of interest recently. It can readily be extended to Milstein or other methods and also to other noise processes (Dereich, 2011; Giles, 2008a). It has been examined for non-globally Lipschitz functions ϕ (Giles, Higham, and Mao, 2009) and combined with other variance reduction techniques (Giles and Waterhouse, 2009).

The Wong–Zakai theorem is described in (Twardowska, 1996; Wong and Zakai, 1965). Definition 8.52 of the Stratonovich (also called the Fisk–Stratonovich) integral is given in (Karatzas and Shreve, 1991; Protter, 2005). In analogue to (8.13), the Stratonovich integral may also be defined by

$$\int_0^t X(s) \circ dW(s) := \lim S_N^X(t) \qquad \text{as } N \to \infty \tag{8.113}$$

for $S_N^X(t) := \sum_{s_{j+1/2} \le t} X(s_{j+1/2})(W(s_{j+1} \wedge t) - W(s_j))$, $s_j = j2^{-N}$, and $s_{j+1/2} := (s_j + s_{j+1})/2$. That is, we choose the midpoint of the interval $[s_j, s_{j+1}]$ for the Stratonovich integral and the left-hand point for the Itô integral.

Many SODE models are posed on a bounded domain $D \subset \mathbb{R}^d$ and boundary conditions specified on ∂D. For example, the process may be killed or reflected at the boundary and such problems are intimately connected to elliptic boundary-value problems. For a discussion of numerical solution of SODEs with boundary conditions, see (Gobet, 2000, 2001; Milstein and Tretyakov, 1999, 2004).

In place of a Brownian motion $\boldsymbol{W}(t)$, other types of stochastic process can be used to force the SODE and this is currently an active research topic. For example, we may consider forcing by a fractional Brownian motion $B_H(t)$ in place of $W(t)$. See, for example, Mishura (2008) and numerical methods for such equations in Neuenkirch (2008). More generally, Lyon's theory of rough paths (Davie, 2007; Friz and Victoir, 2010) allows very general driving terms to be considered.

Exercises

8.1 Show that an $\mathcal{F}_t$-Brownian motion $\{W(t)\colon t \geq 0\}$ has mean function $\mu(t) = 0$ and covariance function $C(s,t) = \min\{s,t\}$ and hence satisfies the conditions of Definition 5.11.

8.2 Let $W(t)$ be an $\mathcal{F}_t$-Brownian motion. Prove that (8.16) holds.

8.3 Show that the distribution of both $\int_0^t W(s)\,ds$ and $\int_0^t s\,dW(s)$ is $\mathrm{N}(0, t^3/3)$.

8.4 Using the Itô formula for $W(t)^2$, show that

$$\int_0^t W(s)\,dW(s) = \frac{1}{2}W(t)^2 - \frac{1}{2}t. \tag{8.114}$$

8.5 For *iid* Brownian motions $W(t), W_1(t), W_2(t)$, prove that

$$\int_0^t s\,dW(s) + \int_0^t W(s)\,ds = t\,W(t)$$
$$\int_0^t W_1(s)\,dW_2(s) + \int_0^t W_2(s)\,dW_1(s) = W_1(t)W_2(t).$$

8.6 For $T > 0$, let $s_j := j2^{-N}$ for $j = 0, 1, \ldots$ and

$$\widetilde{S}_N(t) := \sum_{s_j < t} X\left(\frac{s_j + s_{j+1}}{2}\right)\big(W(s_{j+1} \wedge t) - W(s_j)\big), \qquad t \in [0,T],$$

for $X \in \mathcal{L}_T^2$. In the case $X = W$, prove that $\mathbb{E}\big[\widetilde{S}_N(t)\big] \to t/2$ as $N \to \infty$.

8.7 Consider the semilinear Itô SODE

$$du = \big[-\lambda u + f(u)\big]\,dt + G(u)\,dW(t), \qquad u(0) = u_0 \in \mathbb{R} \tag{8.115}$$

for $f \in \mathrm{C}^2(\mathbb{R},\mathbb{R})$, $G \in \mathrm{C}^2(\mathbb{R},\mathbb{R})$, and $\lambda > 0$. Using the Itô formula (8.30), show that the variation of constants formula holds: that is,

$$u(t) = \mathrm{e}^{-\lambda t}u_0 + \int_0^t \mathrm{e}^{-\lambda(t-s)}f(u(s))\,ds + \int_0^t \mathrm{e}^{-\lambda(t-s)}G(u(s))\,dW(s). \tag{8.116}$$

8.8 a. Let u_n denote the Euler–Maruyama approximation (8.42) of (8.2) for $\lambda, \sigma > 0$. In the case that $u_0 = 0$ and $T > 0$, prove that there exists $K > 0$ such that

$$\big|\mathbb{E}\big[\phi(u_n)\big] - \mathbb{E}\big[\phi(u(T))\big]\big| \leq K\|\phi\|_{L^2(\mathbb{R})}\Delta t,$$

for $t_n = T$ and $\phi \in L^2(\mathbb{R})$. Further, show that

$$\mathbb{E}\big[u_n^2\big] \to \frac{\sigma^2}{2\lambda + \lambda^2\Delta t} \qquad \text{as } n \to \infty.$$

b. Let u_n satisfy the $\theta = 1/2$ Euler–Maruyma rule

$$u_{n+1} = u_n - \frac{1}{2}\lambda(u_n + u_{n+1})\Delta t + \sigma\Delta W_n.$$

Prove that u_n^2 has the correct long term average as defined by (8.7):

$$\mathbb{E}\big[u_n^2\big] \to \frac{\sigma^2}{2\lambda} \qquad \text{as } n \to \infty.$$

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

8.9 Consider the Duffing–van der Pol SODE (8.10) and, for the following range of parameters, approximate and plot a sample path of the solution $\boldsymbol{u}(t)$ on the (u_1, u_2) phase-plane and describe how the dynamics change.

a. For $\lambda = 1$ and $\sigma = 0.5$, vary α from $\alpha = -0.1$ up to $\alpha = 0.2$.

b. For $\alpha = -1$ and $\sigma = 0.3$, vary λ from $\lambda = -0.1$ up to $\lambda = 0.1$.

8.10 a. From (8.45), show that, with $d = m = 1$, the Milstein method for (8.2) can be written as

$$u_{n+1} = u_n + f(u_n)\,\Delta t + g(u_n)\,\Delta W_n + \frac{1}{2}\frac{\partial g}{\partial u}(u_n) g(u_u)\big(\Delta W_n^2 - \Delta t\big).$$

b. Show that with $m = 1$ and $d \in \mathbb{N}$, the Milstein method can be written as

$$u_{k,n+1} = u_{k,n} + f_k(\boldsymbol{u}_n)\,\Delta t + g_k(\boldsymbol{u}_n)\,\Delta W_{k,n} + \frac{1}{2}\sum_{\ell=1}^{d}\frac{\partial g_k}{\partial u_\ell}(\boldsymbol{u}_n) g_\ell(\boldsymbol{u}_n)\big(\Delta W_n^2 - \Delta t\big).$$

c. Derive the form for the Milstein method given in Definition 8.24 with $m = d = 2$.

8.11 Let $\boldsymbol{u}_n$ denote the Euler–Maruyama approximation with time step Δt to (8.3) under Assumption 8.29. Prove the following: for $T > 0$, there exists $K_1, K_2 > 0$ independent of $\Delta t > 0$ such that for $0 \le t_n = n\Delta t \le T$

$$\big\|\boldsymbol{u}_n - \boldsymbol{u}(t_n)\big\|_{L^2(\Omega,\mathbb{R}^d)} \le K_1(\mathrm{e}^{K_2 t_n} - 1)^{1/2}\Delta t^{1/2}. \tag{8.117}$$

8.12 Consider the following Itô SODE with $d = 3$ and $m = 2$

$$d\boldsymbol{u} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -u_2 & u_1 \end{pmatrix} d\boldsymbol{W}(t), \qquad \boldsymbol{u}(0) = \boldsymbol{0} \in \mathbb{R}^3.$$

Show that $u_3(t)$ equals the Lévy area $A_{12}(0,t)$ defined in (8.47). Show how to approximate $u_3(t)$ when $t \ll 1$ by an Euler–Maruyama approximation $u_{3,n}$ and determine a condition on Δt to achieve $\|u_3(t) - u_{3,n}\|_{L^2(\Omega)} = \mathcal{O}(t^{3/2})$ for $t = n\Delta t$.

8.13 Write down the θ-Milstein method for (8.11) and show that the mean-square stability condition is

$$\big(r + \tfrac{1}{2}\sigma^2\big) + \Big(\big(\tfrac{1}{2} - \theta\big)r^2 + \frac{1}{4}\sigma^4\Big)\Delta t < 0.$$

8.14 Show that the Itô formula (Lemma 8.42) can be rewritten as follows: In component form, where $\boldsymbol{f} = [f_1, \ldots, f_d]^{\mathsf{T}}$ and G has entries g_{ij},

$$d\Phi = \left(\frac{\partial \Phi}{\partial t} + \sum_{i=1}^{d}\frac{\partial \Phi}{\partial u_i} f_i + \frac{1}{2}\sum_{i,j=1}^{d}\sum_{k=1}^{m}\frac{\partial^2 \Phi}{\partial u_i \partial u_j} g_{ik}\, g_{jk}\right) dt + \sum_{i=1}^{d}\sum_{j=1}^{m}\frac{\partial \Phi}{\partial u_i} g_{ij}\, dW_j(t).$$

and, in the notation of vector calculus,

$$d\Phi = \left(\frac{\partial \Phi}{\partial t} + \nabla\Phi \cdot \boldsymbol{f} + \tfrac{1}{2}\operatorname{Tr}\nabla^2\Phi\, G\, G^{\mathsf{T}}\right) dt + \nabla\Phi \cdot G\, d\boldsymbol{W}(t),$$

where $\operatorname{Tr}\nabla^2\Phi\, G\, G^{\mathsf{T}}$ may be equivalently written as $\operatorname{Tr} G\, G^{\mathsf{T}}\, \nabla^2\Phi$ or $\operatorname{Tr} G^{\mathsf{T}}\, \nabla^2\Phi\, G$.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

8.15 Consider the Itô SODE

$$d\boldsymbol{u} = A\boldsymbol{u}\,dt + \sum_{i=1}^{m} B_i\boldsymbol{u}\,dW_i(t), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0 \in \mathbb{R}^d, \tag{8.118}$$

where $A, B_i \in \mathbb{R}^{d\times d}$ and $W_i(t)$ are *iid* Brownian motions for $i = 1,\ldots,m$. If the matrices A, B_i all commute (so that $AB_i = B_iA$ and $B_iB_j = B_jB_i$), show that

$$\boldsymbol{u}(t) = \exp\left(\left(A - \frac{1}{2}\sum_{i=1}^{m} B_i^2\right)t + \sum_{i=1}^{m} B_iW_i(t)\right)\boldsymbol{u}_0, \tag{8.119}$$

where $\exp(A)$ is the matrix exponential defined by $\exp(A) := \sum_{k=0}^{\infty} A^k/k!$.

8.16 Consider the semilinear SODE

$$d\boldsymbol{u} = \left[-A\boldsymbol{u} + \boldsymbol{f}(\boldsymbol{u})\right]dt + G(\boldsymbol{u})\,d\boldsymbol{W}(t), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0, \tag{8.120}$$

where $A \in \mathbb{R}^{d\times d}$, $\boldsymbol{f}\colon \mathbb{R}^d \to \mathbb{R}^d$, and $G\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$, and $\boldsymbol{W}(t)$ is a vector of m *iid* Brownian motions. The semi-implicit Euler–Maruyama method to approximate the solution $\boldsymbol{u}(t_n)$ of (8.120) at $t_n = n\Delta t$ is given by

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \Delta t\left(-A\boldsymbol{u}_{n+1} + \boldsymbol{f}(\boldsymbol{u}_n)\right)dt + G(\boldsymbol{u}_n)\,\Delta\boldsymbol{W}_n \tag{8.121}$$

where the increments are $\Delta\boldsymbol{W}_n = \boldsymbol{W}(t_{n+1}) - \boldsymbol{W}(t_n)$.

a. For $d = m = 1$. Let $f(u) = \frac{\sigma}{2}u - u^3$ and $G(u) = \sqrt{\sigma}u$. Implement the semi-implicit Euler–Maruyama method with $\sigma = 2$, $M = -1$, $u_0 = 1$, $t \in [0,1]$ and plot a sample path with $\Delta t = 0.001$. The exact solution (Kloeden and Platen, 1992, p.125) is given by

$$u(t) = \frac{u_0 \mathrm{e}^{-tA+\sqrt{\sigma}W(t)}}{\sqrt{1 + 2u_0^2\int_0^t \mathrm{e}^{-2sA+2\sqrt{\sigma}W(s)}ds}}. \tag{8.122}$$

Compare the numerical solution with the exact solution using the trapezium rule to approximate the integral in (8.122).

b. For $d = m = 4$, let $\boldsymbol{f}(\boldsymbol{u}) = \boldsymbol{u} - \boldsymbol{u}^3$ (so $f_j = u_j - u_j^3$), $G(\boldsymbol{u}) = \mathrm{diag}(\boldsymbol{u})$ and

$$A = r\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

Implement the semi-implicit Euler–Maruyama method for this system. Choose $r = 4$ and investigate strong convergence numerically. Compare the stability properties of the semi-implicit to the explicit Euler–Maruyama method when r is large. Choose $r = 40$ and consider time steps $\Delta t = 0.002$ and $\Delta t = 0.02$.

8.17 Consider the semilinear SODE (8.115). Using the variation of constants formula (8.116), derive the numerical method

$$u_{n+1} = \mathrm{e}^{-\Delta t\lambda}u_n + \phi_1(-\Delta t\lambda)f(u_n)\,\Delta t + \mathrm{e}^{-\Delta t\lambda}G(u_n)\,\Delta W_n, \tag{8.123}$$

where $\phi_1(z) := z^{-1}(\mathrm{e}^z - 1)$. This is an example of an *exponential integrator*.

Implement (8.123) for (8.115) with $\lambda = 1$, $f(u) = u - u^3$ and $G(u) = \sqrt{2}u$ with

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press

$u(0) = 1$. Test your sample path against the solution found using the trapezium rule to approximate the integral in (8.122) and take $T = 1$, $\Delta t = 0.001$. Compare the approximate sample to path to that found by the Euler–Maruyama method.

8.18 Show that the solution $\boldsymbol{u}(t)$ of (8.3) under Assumption 8.17 satisfies the following: for any $\boldsymbol{u}_0 \in \mathbb{R}^d$, $p \geq 2$, $T > 0$, there exists $K > 0$ such that

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[\left\|\boldsymbol{u}(t)\right\|_2^p\right] \leq K.$$

Use Gronwall's inequality and (8.112).

8.19 For $r, \sigma > 0$, let $u(t), v(t)$ be the Itô and Stratonovich geometric Brownian motions, defined by the SODEs

$$\begin{aligned} du &= ru\,dt + \sigma u\,dW(t), & u(0) &= u_0, \\ dv &= rv\,dt + \sigma v \circ dW(t), & v(0) &= u_0. \end{aligned}$$

By giving the exact form for $u(t)$ and $v(t)$ in terms of $W(t)$, show that $u(t,\omega) \leq v(t,\omega)$ for any $t \geq 0$ for almost all $\omega \in \Omega$. Give an example of parameters r, σ where $u(t,\omega) \to 0$ and $v(t,\omega) \to \infty$ as $t \to \infty$ a.s.

8.20 Consider $\lambda, \sigma > 0$ and $f : \mathbb{R} \to \mathbb{R}$. Using the modified drift (8.111), find the Itô form of the following Stratonovich SODEs:

$$\begin{aligned} d\begin{pmatrix} q \\ p \end{pmatrix} &= \begin{pmatrix} p \\ -p\lambda - f(q) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma q \end{pmatrix} \circ dW(t) \\ d\begin{pmatrix} q \\ p \end{pmatrix} &= \begin{pmatrix} p \\ -p\lambda - f(q) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma p \end{pmatrix} \circ dW(t). \end{aligned}$$

8.21 Show that if $W_1(t)$, $W_2(t)$ are *iid* Brownian motions, then the quadratic variation $[W_1, W_2]_t = 0$. Hence, show the Lévy areas are equal for the Itô and Stratonovich interpretation of the integral.

8.22 Let $u(t)$ be the solution of the initial value problem (8.3) in the case $d = m = 1$. For $\phi : \mathbb{R} \to \mathbb{R}$, show that the quadratic variation

$$\left[\phi(u), W\right]_t = \int_0^t \phi'(u(s))G(u(s))\,ds.$$

State any smoothness conditions you need on f, G, and ϕ. Explain why the quadratic variation is unchanged if the Stratonovich interpretation of the integral is used to define the process $u(t)$.

8.23 Convert the following Stratonovich SODE to the equivalent Itô SODE.

$$du = \left(u - u^3\right) dt + \sigma u \circ dW, \qquad u(0) = 1.$$

Implement Heun's method and compute an approximate sample path for the SODE with $\sigma = 2$, $T = 2$, and $\Delta t = 0.002$. Compare this approximate sample path to that found by the Euler–Maruyama method for the equivalent Itô SODE.

https://doi.org/10.1017/CBO9781139017329.009 Published online by Cambridge University Press