Lecture 9: Numerically solving SDEs

Readings

Recommended:
- Pavliotis [2014] Section 5.2
- Hingham [2001] – a short, simple introduction to numerically solving SDEs. It is recommended that you read this after first trying the homework – then you can read it for clues / more background.

Optional:
  See e.g. Ch. 5 (for the Itô-Taylor expansion), Ch 6.3 (for a version of stochastic stability) Ch. 8 (for a discussion of numerical methods for deterministic equations), Ch. 9 (for numerical discretization of SDEs, and convergence notions), Ch. 10+ (for a detailed discussion of specific schemes.)

Most SDEs don’t have explicit analytical solutions, like the ones we studied last class. One way to gain information about them is to simulate them numerically. This gives only an approximation to the “true” solutions, so the mathematical issue is to understand how close this approximation is, and to invent schemes that approximate it more closely, given finite computational resources.

This lecture will be an introduction to the major schemes and considerations in numerical SDEs. It certainly won’t be comprehensive. The bible on the topic is Kloeden and Platen [1999]. A short, basic review article (Hingham [2001]) has also been posted to the website. It is highly recommended that you read this – though it may be pedagogically better to try to homework first, and read it if/when you get stuck.

9.1 Basic schemes

We’d like to solve the scalar, time-homogeneous equation
\[ dX_t = b(X_t) dt + \sigma(X_t) dW_t. \] (1)

The first scheme to consider comes from approximating the Itô integral as a sum.

Definition (Euler-Maruyama Scheme).
\[ X_{n+1} = X_n + b(X_n) \Delta t + \sigma(X_n) \delta W_n, \] (2)

where \( \delta W_n \sim N(0, \Delta t) \) are i.i.d. random variables, \( \Delta t \) is the size of the timestep, and \( X_n \) is the value of the solution at the \( n \)th timestep.

Notes
- The random variable \( \delta W_n \) is sometimes replaced by a random variable with the same mean and variance, such as a uniform r.v., or a two-valued r.v. \( \xi = \pm \sqrt{\Delta t} \). (This only provides weak, not strong, solutions though.) This is advantageous when generating Gaussian r.v.s is expensive.
- This is the most widely-used method, but also the least accurate.
For better methods we need the stochastic Taylor expansion. Recall the *deterministic* Taylor expansion, for the solution of an ODE. Suppose \( X_t \) solves \( \frac{dX_t}{dt} = a(X_t) \). Let’s calculate its Taylor expansion about 0, in integral form.

First, we have
\[
X_t = X_0 + \int_0^t a(X_s) ds. \tag{3}
\]

To expand \( a(X_t) \), we need the following identity, which holds for a continuously differentiable function \( f \):
\[
\frac{d}{dt} f(X_t) = a(X_t) \frac{df(X_t)}{dx} \quad \implies \quad f(X_t) = f(X_0) + \int_0^t \mathcal{L} f(X_s) ds, \tag{4}
\]
where \( \mathcal{L} = a(x) \frac{\partial}{\partial x} \). Apply (4) to \( f = a \) in (3) to get
\[
X_t = X_0 + a(X_0) \int_0^t ds + \int_0^t \int_0^s \mathcal{L} a(X_u) duds. \tag{5}
\]

Apply (4) to \( f = \mathcal{L} a \) in (5) to get
\[
X_t = X_0 + a(X_0) \int_0^t ds + \int_0^t \int_0^s \mathcal{L} a(X_u) duds + R_3, \quad R_3 = \int_0^t \int_0^s \int_0^r \mathcal{L}^2 a(X_v) dv duds. \tag{6}
\]

In general, we have that
\[
f(X_t) = f(X_0) + \sum_{m=1}^r \frac{t^m}{m!} \mathcal{L}^m f(X_0) + R_{r+1}, \quad R_{r+1} = \int_0^t \int_0^s \int_0^r \mathcal{L}^{r+1} f(X_u) ds_1 \cdots ds_{r+1}. \tag{6}
\]

This is the Taylor formula in integral form.

Now let’s consider the *Stochastic Itô-Taylor expansion*. Consider the SDE
\[
X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s. \tag{7}
\]

For any function \( f \in C^2 \), the Itô formula tells us that
\[
f(X_t) = f(X_0) + \int_0^t \left( b(X_s) \frac{\partial f(X_s)}{\partial x} + \frac{1}{2} \sigma^2(X_s) \frac{\partial^2 f(X_s)}{\partial x^2} \right) ds + \int_0^t \sigma(X_s) \frac{\partial f(X_s)}{\partial x} dW_s.
\]

Therefore we have the identity
\[
f(X_t) = f(X_0) + \int_0^t \mathcal{L}_0 f(X_s) ds + \int_0^t \mathcal{L}_1 f(X_s) dW_s, \tag{8}
\]
where \( \mathcal{L}_0 = b \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2} \), \( \mathcal{L}_1 = \sigma \frac{\partial}{\partial x} \). Apply (8) to \( f = b \) and \( f = \sigma \) in (7):
\[
X_t = X_0 + b(X_0) \int_0^t ds + \sigma(X_0) \int_0^t dW_s + R_1, \tag{9}
\]
with
\[ R_1 = \int_0^t \int_0^s L_0 b(X_z) dz ds + \int_0^t \int_0^s L_1 b(X_z) dW_z ds + \int_0^t \int_0^s L_0 \sigma(X_z) dz dW_s + \int_0^t \int_0^s L_1 \sigma(X_z) dW_z dW_s. \]

We can continue. Apply (8) to \( f = L_1 \sigma \) in \( R_1 \), to get
\[ X_t = X_0 + b(X_0) \int_0^t ds + \sigma(X_0) \int_0^t dW_s + L_1 \sigma(X_0) \int_0^t dW_z dW_s + R_2, \] (10)

where \( R_2 \) is a remainder term that can be worked out.

From these expansions we derive approximation schemes, by computing the integrals exactly, and ignoring the remainder terms.

- The Euler-Maruyama scheme comes from (9).
- The Milstein scheme comes from (10).

**Definition** (Milstein scheme).
\[ X_{n+1} = X_n + b(X_n) \Delta t + \sigma(X_n) \Delta W_n + \frac{1}{2} \sigma \sigma' (X_n) (\Delta W_n^2 - \Delta t). \] (11)

**Notes**
- This comes from calculating the multiple stochastic integral
\[ \int_0^t \int_0^s dW_z dW_s = \int_0^t W_s dW_s = \frac{1}{2} (W_t^2 - t). \]
- For higher-order approximations, we have to calculate integrals of the form \( \int_0^t dW_{i_1}^l \int_0^{s_1} dW_{i_2}^l \cdots \int_0^{s_{k-1}} dW_{i_k}^l \), where \( i_j \in \{0, 1\} \), and \( W_0^l \equiv t, W_1^l \equiv W_t \).
- For the scalar case, these integrals can be computed recursively (see HW).
- For the vector case, the Brownian motions are different, and the multiple integrals are not known analytically. They can be approximated, e.g. using the Karhunen-Loeve representation, but usually E-M (or variants of it) is the best method.

The Milstein scheme requires computing \( \sigma' \). When this is hard, we can replace it with a numerical, Runge-Kutta style approximation.

**Definition** (Runge-Kutta scheme).
\[ \hat{X}_n = X_n \sigma(X_n) \sqrt{\Delta t} \]
\[ X_{n+1} = X_n + b(X_n) \Delta t + \sigma(X_n) \Delta W_n + \frac{1}{2} \frac{1}{\sqrt{\Delta t}} (\sigma(\hat{X}_n) - \sigma(X_n)) (\Delta W_n^2 - \Delta t). \]

### 9.2 Strong and weak convergence

How can we judge a scheme’s quality? Commonly-used notions are
– consistency
– convergence
– stability

Consistency asks whether the local discretization error converges to 0 with order at least 1 with time step. Convergence asks whether the global error converges to zero in some sense with time step. Stability asks (roughly) whether the numerical method converges to the same long-time states as the exact solution.

For each of these, there is a strong and a weak form.

– strong forms deal with pathwise results.
– weak forms deal with probability distributions.

In the following, let $Y_{n}^{\Delta t}$ be a discrete approximation to $X_t$, at times $\{t_n\}$, with maximum increment $\Delta t$.

**Definition.** $Y_{n}^{\Delta t}$ converges strongly (in mean-square) to $X_t$ on $[0,T]$ with order $\alpha$ if there exist constants $C > 0$, $\delta_0 > 0$, independent of $\Delta t$, such that

$$\max_{0 \leq t \leq T} \mathbb{E} |Y_{n}^{\Delta t} - X_{t_n}|^2 \leq C(\Delta t)^{2\alpha} \quad \forall 0 < \Delta t < \delta_0,$$

**Definition.** $Y_{n}^{\Delta t}$ converges weakly to $X_t$ on $[0,T]$ with order $\beta$ if there exist constants $C_f > 0$, $\delta_0 > 0$, independent of $\Delta t$, such that

$$\max_{0 \leq t \leq T} |\mathbb{E} f(Y_{n}^{\Delta t}) - \mathbb{E} f(X_{t_n})| \leq C_f(\Delta t)^{\beta} \quad \forall 0 < \Delta t < \delta_0,$$

for all $f \in C_\infty^b(\mathbb{R}^d) = \text{bounded, infinitely differentiable functions with bounded derivatives}$. The constant $C_f$ can depend on $f$ but not on $\Delta t$, and the above holds

**Notes**

- Strong convergence requires the paths to be close. Therefore the same Brownian motion must be used for all approximations.
- Weak convergence requires only the probability distributions to converge. Therefore a different Brownian motion can be used for each numerical approximation, or even a random process that is not Brownian motion, but has increments with the same mean, variance.
- Some books use the $L^1$ norm to define strong convergence, as $\max_{0 \leq t \leq T} \mathbb{E} |Y_{n}^{\Delta t} - X_{t_n}| \leq C(\Delta t)^{\alpha}$

**Theorem.** $\beta \geq \alpha$ (Weak order $\geq$ Strong order)

**Proof.** Suppose $|f'| \leq K$. Then

$$|\mathbb{E} f(Y_{n}^{\Delta t}) - \mathbb{E} f(X_{t_n})| \leq \mathbb{E} |f(Y_{n}^{\Delta t}) - f(X_{t_n})| \leq K \mathbb{E} |Y_{n}^{\Delta t} - X_{t_n}| \quad \text{(by Mean Value Theorem)}$$

$$\leq K \left( \mathbb{E} |Y_{n}^{\Delta t} - X_{t_n}|^2 \right)^{1/2} \quad \text{(Cauchy-Schwartz)}$$

\[\square\]
Theorem.

<table>
<thead>
<tr>
<th></th>
<th>Strong Order</th>
<th>Weak Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-M</td>
<td>1/2</td>
<td>1</td>
</tr>
<tr>
<td>Milstein</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Runge-Kutta</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Proof of strong convergence, E-M. (from [E et al. 2014]. See also [Kloeden and Platen 1999], Theorem 10.2.2 p.342.) We show this for $\sigma = 1$, and where $b$ satisfies a global Lipschitz condition with constant $L$.

We have

$$X_{n+1} = X_n + \int_{t_n}^{t_{n+1}} b(X_t)dt + \delta W_n \quad \text{(exact solution)}$$

$$Y_{n+1} = Y_n + \int_{t_n}^{t_{n+1}} b(Y_t)dt + \delta W_n \quad \text{(approximate solution)}$$

Let $e_n = X_n - Y_n$ be the error. Then

$$e_{n+1} = e_n + \int_{t_n}^{t_{n+1}} (b(X_t) - b(Y_t))dt.$$ 

Square this, use the inequality $2ab \leq a^2 \Delta t + b^2/\Delta t$, to get

$$|e_{n+1}|^2 \leq |e_n|^2 + \left( \int_{t_n}^{t_{n+1}} (b(X_t) - b(Y_t))dt \right)^2 + 2e_n \int_{t_n}^{t_{n+1}} (b(X_t) - b(Y_t))dt$$

$$\leq |e_n|^2 (1 + \Delta t) + \left( 1 + \frac{1}{\Delta t} \right) \left( \int_{t_n}^{t_{n+1}} (b(X_t) - b(Y_t))dt \right)^2$$

$$\leq |e_n|^2 (1 + \Delta t) + L^2(1 + \Delta t) \int_{t_n}^{t_{n+1}} |X_t - Y_t|^2 dt$$

Note that $|X_t - Y_t|^2 \leq 2|X_t - X_n|^2 + 2|X_n - Y_n|^2$. The first term on the RHS is $\leq K_2 \Delta t$, where $K_2$ is a constant that depends on $L, T, E(X_0^2, \Delta t)$ (ELFS, or Bonus Q on HW), and the second term is $2e_n^2$. Therefore, letting $L_1 = 1 + 2L^2(1 + T), L_2 = 2L^2(1 + \Delta t)K_2$, we have

$$\mathbb{E}|e_{n+1}|^2 \leq \mathbb{E}|e_n|^2 (1 + L_1 \Delta t) + L_2(\Delta t)^2,$$

$$\Rightarrow \mathbb{E}|e_n|^2 \leq \frac{L_2}{L_1}(e^{L_1T} - 1)\Delta t.$$

For proofs of convergence of other schemes, see [Kloeden and Platen 1999].

Notes

- To prove weak convergence requires the PDE version of the theory, which we will save for later.
- For additive noise ($\sigma = \text{const}$), the E-M and Milstein schemes are equivalent, since $\sigma \sigma' = 0$. Therefore E-M will converge with strong and weak order 1.
- How to demonstrate the order of convergence numerically?
– Use true solution, and compare to this, if known.

– Use a very high-resolution simulation to compare to. (Problem: this is very expensive.)

– Look at the difference between solutions as you double the resolution. We expect

\[ \frac{\|Y_{\Delta t} - Y_{\Delta t/2}\|}{\|Y_{\Delta t/2} - Y_{\Delta t/4}\|} = 2^\gamma + O(\Delta t), \]

where \( \gamma \) is the order and \( \| \cdot \| \) is some norm.

The latter two show the solution converges to something. However, we have no guarantee that the solution it converges to is the correct one.

– The numerically-measured error will also contain sampling error, since you have a finite number of samples. Therefore, if you wish to accurately measure the order of convergence, the number of samples must be large enough to make this smaller than the error due to discretization. You can also construct error bars on your estimate of the order – see Kloeden and Platen [1999] section 9.3.

– There will also be error due to bias in the random number generator, and rounding error due to machine precision – interestingly it is the former that becomes important first, when \( \Delta t \) is small enough and the number of samples is large enough! See Hingham [2001].

9.3 Stochastic stability

Convergence bounds the error for time intervals \([0, T]\) using a constant \( C(T) \). But typically, \( C(T) \searrow \infty \) as \( T \searrow \infty \). In many situations (e.g. first-passage problems, long-time behaviour, etc) we need to simulate the equation indefinitely. In these cases, we may ask that the numerics reproduce the correct qualitative behaviour. One way to do this is to consider the notion of stability. Typically, we pick a particular class of equations to study this concept. One common choice is to look at the behaviour near fixed points of linear equations.

Recall the concept of linear stability for deterministic ODEs:

– We typically study the behaviour of \( \frac{dX}{dt} = \lambda X \), where \( \lambda \) is a complex number.
– The fixed point \( X = 0 \) is asymptotically stable if \( \lim_{t \to \infty} X(t) = 0 \). This happens for the equation above when \( \text{Re}\{\lambda\} < 0 \).
– If we discretize, we ask when the numerical scheme reproduces this same behaviour: that \( X_n \to 0 \) as \( n \to \infty \). This will typically be true only for some step sizes \( \Delta t \). The set of values of \( \lambda \Delta t \) for which the numerical scheme also converges to zero forms the domain of linear stability of the scheme.

For SDEs, one common option is to study the equation

\[ dX_t = \lambda X_t dt + \mu X_t dW_t, \quad \lambda, \mu \in \mathbb{C}. \]  

(12)

This is a Geometric Brownian Motion that often comes from linearizing a nonlinear equation about a “stable” point. Since we are now dealing with random variables, which are infinite-dimensional objects, norms are not equivalent in general so there are different notions of stability. Two common ones are:

**Definition.** The solution \( X_t = 0 \) is mean-square stable (for a given pair \( \lambda, \mu \)) if \( \lim_{t \to \infty} E X_t^2 = 0 \).
**Definition.** The solution $X_t = 0$ is **asymptotically stable** if $P(\lim_{t \to \infty} |X_t| = 0) = 1$.

**Notes**

- GBM is mean-square stable $\iff$ $\text{Re}\{\lambda\} + \frac{1}{2}|\mu|^2 < 0$.
- GBM is asymptotically stable $\iff$ $\text{Re}\{\lambda - \frac{1}{2}\mu^2\} < 0$.
- For GBM, mean-square stability $\Rightarrow$ asymptotic stability, but not the reverse.

When does a numerical method give the same type of stability? Consider EM:

$$X_{n+1} = X_n(1 + \lambda \Delta t + \mu \Delta W) \Rightarrow \mathbb{E}X_{n+1}^2 = (1 + \lambda \Delta t)^2 + |\mu|^2 \Delta t \mathbb{E}X_n^2 \Rightarrow \lim_{n \to \infty} \mathbb{E}X_n^2 = 0 \iff |1 + \lambda \Delta t|^2 + |\mu|^2 \Delta t < 1.$$  

Suppose $\lambda, \mu \in \mathbb{R}$. Let $y = \mu^2 \Delta t$, $x = \lambda \Delta t$. The numerical method gives mean-square stability when $(1 + x)^2 + y < 1 \iff y < -x(2 - x)$.

The true solution gives mean-square stability when $y < -2x$.

To reproduce the correct behaviour, we need a small enough timestep:

$$\Delta t < \frac{-(\mu^2 + 2\lambda)}{2\lambda^2}.$$  

It can be shown (using the Strong Law of Large Numbers and the Law of the Iterated Logarithm) that asymptotic stability for the EM method requires

$$\mathbb{E} \log |1 + \Delta t \lambda + \sqrt{\Delta t} \mu N(0, 1)| < 0.$$
Notes

- What about implicit methods? These have much better stability properties for deterministic ODEs. For SDEs, however, they are hard to implement because there is no control on the size of the noise, so one cannot rule out dividing by 0 (see [Kloeden and Platen 1999], p. 336). Therefore implicit methods are generally only used for the deterministic parts of the equations. See [Pavliotis 2014], 5.2.1 for a brief discussion.

- Note that numerical stability is another concept, related to but different from that discussed above, that asks whether two nearby trajectories tend to stay together or to diverge. See [Kloeden and Platen 1999], section 9.8.

References


