

Discretization Methods

This chapter presents methods for reducing discretization error — the bias in Monte Carlo estimates that results from time-discretization of stochastic differential equations. Chapter 3 gives examples of continuous-time stochastic processes that can be simulated exactly at a finite set of dates, meaning that the joint distribution of the simulated values coincides with that of the continuous-time model at the simulated dates. But these examples are exceptional and most models arising in derivatives pricing can be simulated only approximately. The simplest approximation is the Euler scheme; this method is easy to implement and almost universally applicable, but it is not always sufficiently accurate. This chapter discusses methods for improving the Euler scheme and, as a prerequisite for this, discusses criteria for comparing discretization methods.

The issues addressed in this chapter are orthogonal to those in Chapters 4 and 5. Once a time-discretization method is fixed, applying a variance reduction technique or quasi-Monte Carlo method may improve precision in estimating an expectation at the fixed level of discretization, but it can do nothing to reduce discretization bias.

6.1 Introduction

We begin by discussing properties of the Euler scheme, the simplest method for approximate simulation of stochastic differential equations. We then undertake an expansion to refine the Euler scheme and present criteria for comparing methods.

6.1.1 The Euler Scheme and a First Refinement

We consider processes X satisfying a stochastic differential equation (SDE) of the form

$$dX(t) = a(X(t)) dt + b(X(t)) dW(t), \quad (6.1)$$

usually with $X(0)$ fixed. In the most general setting we consider, X takes values in \Re^d and W is an m -dimensional standard Brownian motion, in which case a takes values in \Re^d and b takes values in $\Re^{d \times m}$. Some of the methods in this chapter are most easily introduced in the simpler case of scalar X and W . The coefficient functions a and b are assumed to satisfy the conditions in Appendix B.2 for existence and uniqueness of a strong solution to the SDE (6.1); indeed, we will need to impose stronger conditions to reduce discretization error.

We use \hat{X} to denote a time-discretized approximation to X . The Euler (or Euler-Maruyama, after [254]) approximation on a time grid $0 = t_0 < t_1 < \dots < t_m$ is defined by $\hat{X}(0) = X(0)$ and, for $i = 0, \dots, m - 1$,

$$\hat{X}(t_{i+1}) = \hat{X}(t_i) + a(\hat{X}(t_i))[t_{i+1} - t_i] + b(\hat{X}(t_i))\sqrt{t_{i+1} - t_i}Z_{i+1},$$

with Z_1, Z_2, \dots independent, m -dimensional standard normal random vectors. To lighten notation, we restrict attention to a grid with a fixed spacing h , meaning that $t_i = ih$. Everything we discuss carries over to the more general case provided the largest of the increments $t_{i+1} - t_i$ decreases to zero. Adaptive methods, in which the time steps depend on the evolution of \hat{X} and are thus stochastic, require separate treatment; see, for example, Gaines and Lyons [133].

With a fixed time step $h > 0$, we may write $\hat{X}(ih)$ as $\hat{X}(i)$ and write the Euler scheme as

$$\hat{X}(i+1) = \hat{X}(i) + a(\hat{X}(i))h + b(\hat{X}(i))\sqrt{h}Z_{i+1}. \quad (6.2)$$

Implementation of this method is straightforward, at least if a and b are easy to evaluate. Can we do better? And in what sense is one approximation better than another? These are the questions we address.

In the numerical solution of *ordinary* differential equations, methods of higher-order accuracy often rely on Taylor expansions. If b were identically zero (and thus (6.1) non-stochastic), (6.2) would reduce to a linear approximation, and a natural strategy for improving accuracy would include higher-order terms in a Taylor expansion of $a(X(t))$. A similar strategy applies to stochastic differential equations, but it must be carried out consistent with the rules of Itô calculus rather than ordinary calculus.

A First Refinement

Inspection of the Euler scheme (6.2) from the perspective of Taylor expansion suggests a possible inconsistency: this approximation expands the drift to $O(h)$ but the diffusion term only to $O(\sqrt{h})$. The approximation to the diffusion term omits $O(h)$ contributions, so including a term of order h in the drift looks like spurious accuracy. This discrepancy also suggests that to refine the Euler scheme we may want to focus on the diffusion term.

We now carry out this proposal. We will see, however, that whether or not it produces an improvement compared to the Euler scheme depends on how we measure error.

We start with the scalar case $d = m = 1$. Recall that the SDE (6.1) abbreviates the relation

$$X(t) = X(0) + \int_0^t a(X(u)) du + \int_0^t b(X(u)) dW(u). \quad (6.3)$$

The Euler scheme results from the approximations

$$\int_t^{t+h} a(X(u)) du \approx a(X(t))h \quad (6.4)$$

and

$$\int_t^{t+h} b(X(u)) dW(u) \approx b(X(t))[W(t+h) - W(t)]. \quad (6.5)$$

In both cases, an integrand over $[t, t+h]$ is approximated by its value at t . To improve the approximation of the diffusion term, we need a better approximation of $b(X(u))$ over an interval $[t, t+h]$. We therefore examine the evolution of $b(X(u))$.

From Itô's formula we get

$$\begin{aligned} db(X(t)) &= b'(X(t)) dX(t) + \frac{1}{2} b''(X(t)) b^2(X(t)) dt \\ &= [b'(X(t))a(X(t)) + \frac{1}{2} b''(X(t)) b^2(X(t))] dt + b'(X(t))b(X(t)) dW(t) \\ &\equiv \mu_b(X(t)) dt + \sigma_b(X(t)) dW(t), \end{aligned}$$

where b' and b'' are the first and second derivatives of b . Applying the Euler approximation to the process $b(X(t))$ results in the approximation of $b(X(u))$, $t \leq u \leq t+h$ by

$$\begin{aligned} b(X(u)) &\approx b(X(t)) + \mu_b(X(t))[u-t] + \sigma_b(X(t))[W(u) - W(t)] \\ &= b(X(t)) + (b'(X(t))a(X(t)) + \frac{1}{2} b''(X(t)) b^2(X(t))) [u-t] \\ &\quad + b'(X(t))b(X(t))[W(u) - W(t)]. \end{aligned}$$

Now $W(u) - W(t)$ is $O(\sqrt{u-t})$ (in probability) whereas the drift term in this approximation is $O(u-t)$ and thus of higher order. Dropping this higher-order term yields the simpler approximation

$$b(X(u)) \approx b(X(t)) + b'(X(t))b(X(t))[W(u) - W(t)], \quad u \in [t, t+h]. \quad (6.6)$$

Armed with this approximation, we return to the problem of refining (6.5). Instead of freezing $b(X(u))$ at $b(X(t))$ over the interval $[t, t+h]$, as in (6.5), we use the approximation (6.6). Thus, we replace (6.5) with

$$\begin{aligned}
& \int_t^{t+h} b(X(u)) dW(u) \\
& \approx \int_t^{t+h} (b(X(t)) + b'(X(t))b(X(t))[W(u) - W(t)]) dW(u) \\
& = b(X(t))[W(t+h) - W(t)] \\
& \quad + b'(X(t))b(X(t)) \left(\int_t^{t+h} [W(u) - W(t)] dW(u) \right). \tag{6.7}
\end{aligned}$$

The proposed refinement uses this expression in place of $b(\hat{X}(i))\sqrt{h}Z_{i+1}$ in the Euler scheme (6.2).

To make this practical, we need to simplify the remaining integral in (6.7). We can write this integral as

$$\begin{aligned}
& \int_t^{t+h} [W(u) - W(t)] dW(u) \\
& = \int_t^{t+h} W(u) dW(u) - W(t) \int_t^{t+h} dW(u) \\
& = Y(t+h) - Y(t) - W(t)[W(t+h) - W(t)] \tag{6.8}
\end{aligned}$$

with

$$Y(t) = \int_0^t W(u) dW(u);$$

i.e., $Y(0) = 0$ and

$$dY(t) = W(t) dW(t).$$

Itô's formula verifies that the solution to this SDE is

$$Y(t) = \frac{1}{2}W(t)^2 - \frac{1}{2}t.$$

Making this substitution in (6.8) and simplifying, we get

$$\int_t^{t+h} [W(u) - W(t)] dW(u) = \frac{1}{2}[W(t+h) - W(t)]^2 - \frac{1}{2}h. \tag{6.9}$$

Using this identity in (6.7), we get

$$\begin{aligned}
\int_t^{t+h} b(X(u)) dW(u) & \approx b(X(t))[W(t+h) - W(t)] \\
& \quad + \frac{1}{2}b'(X(t))b(X(t)) ([W(t+h) - W(t)]^2 - h).
\end{aligned}$$

Finally, we use this approximation to approximate $X(t+h)$. We refine the one-step Euler approximation

$$X(t+h) \approx X(t) + a(X(t))h + b(X(t))[W(t+h) - W(t)]$$

to

$$\begin{aligned} X(t+h) \approx X(t) + a(X(t))h + b(X(t))[W(t+h) - W(t)] \\ + \frac{1}{2}b'(X(t))b(X(t))([W(t+h) - W(t)]^2 - h). \end{aligned}$$

In a simulation algorithm, we apply this recursively at $h, 2h, \dots$, replacing the increments of W with $\sqrt{h}Z_{i+1}$; more explicitly, we have

$$\begin{aligned} \hat{X}(i+1) = \hat{X}(i) + a(\hat{X}(i))h + b(\hat{X}(i))\sqrt{h}Z_{i+1} \\ + \frac{1}{2}b'(\hat{X}(i))b(\hat{X}(i))h(Z_{i+1}^2 - 1). \end{aligned} \quad (6.10)$$

This algorithm was derived by Milstein [266] through an analysis of partial differential equations associated with the diffusion X . It is sometimes called the Milstein scheme, but this terminology is ambiguous because there are several important methods due to Milstein.

The approximation method in (6.10) adds a term to the Euler scheme. It expands both the drift and diffusion terms to $O(h)$. Observe that, conditional on $\hat{X}(i)$, the new term

$$\frac{1}{2}b'(\hat{X}(i))b(\hat{X}(i))h(Z_{i+1}^2 - 1)$$

has mean zero and is uncorrelated with the Euler terms because $Z_{i+1}^2 - 1$ and Z_{i+1} are uncorrelated. The question remains, however, whether and in what sense (6.10) is an improvement over the Euler scheme. We address this in Section 6.1.2, after discussing the case of vector-valued X and W .

The Multidimensional Case

Suppose, now, that $X(t) \in \Re^d$ and $W(t) \in \Re^m$. Write X_i , W_i , and a_i for the i th components of X , W , and a , and write b_{ij} for the ij -entry of b . Then

$$X_i(t+h) = X_i(t) + \int_t^{t+h} a_i(X(u)) du + \sum_{j=1}^m \int_t^{t+h} b_{ij}(X(u)) dW_j(u),$$

and we need to approximate the integrals on the right. As in the Euler scheme, we approximate the drift term using

$$\int_t^{t+h} a_i(X(u)) du \approx a_i(X(t))h.$$

The argument leading to (6.7) yields

$$\begin{aligned} \int_t^{t+h} b_{ij}(X(u)) dW_j(u) &\approx b_{ij}(X(t))[W_j(t+h) - W_j(t)] \\ &+ \sum_{\ell=1}^d \sum_{k=1}^m \frac{\partial b_{ij}}{\partial x_\ell}(X(t))b_{\ell k}(X(t)) \int_t^{t+h} [W_k(u) - W_k(t)] dW_j(u). \end{aligned} \quad (6.11)$$

For $k = j$, we can evaluate the integral in (6.11) as in the scalar case:

$$\int_t^{t+h} [W_j(u) - W_j(t)] dW_j(u) = \frac{1}{2}[W_j(t+h) - W_j(t)]^2 - \frac{1}{2}h.$$

However, there is no comparable expression for the off-diagonal terms

$$\int_t^{t+h} [W_k(u) - W_k(t)] dW_j(u), \quad k \neq j.$$

These mixed integrals (or more precisely their differences) are called *Lévy area* terms; see the explanation in Protter [300, p.82], for example. Generating samples from their distribution is a challenging simulation problem. Methods for doing so are developed in Gaines and Lyons [132] and Wiktorsson [356], but the difficulties involved limit the applicability of the expansion (6.11) in models driven by multidimensional Brownian motion. Fortunately, we will see that for the purpose of estimating an expectation it suffices to simulate rough approximations to these mixed Brownian integrals.

6.1.2 Convergence Order

Equation (6.10) displays a refinement of the Euler scheme based on expanding the diffusion term to $O(h)$ rather than just $O(\sqrt{h})$. To discuss the extent and the sense in which this algorithm is an improvement over the Euler scheme, we need to establish a figure of merit for comparing discretizations.

Two broad categories of error of approximation are commonly used in measuring the quality of discretization methods: criteria based on the path-wise proximity of a discretized process to a continuous process, and criteria based on the proximity of the corresponding distributions. These are generally termed *strong* and *weak* criteria, respectively.

Let $\{\hat{X}(0), \hat{X}(h), \hat{X}(2h), \dots\}$ be any discrete-time approximation to a continuous-time process X . Fix a time T and let $n = \lfloor T/h \rfloor$. Typical strong error criteria are

$$\mathbb{E} [\|\hat{X}(nh) - X(T)\|], \quad \mathbb{E} [\|\hat{X}(nh) - X(T)\|^2],$$

and

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} \|\hat{X}(\lfloor t/h \rfloor h) - X(t)\| \right],$$

for some vector norm $\|\cdot\|$. Each of these expressions measures the deviation between the individual values of X and the approximation \hat{X} .

In contrast, a typical weak error criterion has the form

$$\left| \mathbb{E}[f(\hat{X}(nh))] - \mathbb{E}[f(X(T))] \right|, \quad (6.12)$$

with f ranging over functions from \Re^d to \Re typically satisfying some smoothness conditions. Requiring that an expression of the form (6.12) converge to zero as h decreases to zero imposes no constraint on the relation between the outcomes of $\hat{X}(nh)$ and $X(T)$; indeed, the two need not even be defined on the same probability space. Making the error criterion (6.12) small merely requires that the distributions of $\hat{X}(nh)$ and $X(T)$ be close.

For applications in derivatives pricing, weak error criteria are most relevant. We would like to ensure that prices (which are expectations) computed from \hat{X} are close to prices computed from X ; we are not otherwise concerned about the paths of the two processes. It is nevertheless useful to be aware of strong error criteria to appreciate the relative merits of alternative discretization methods.

Even after we fix an error criterion, it is rarely possible to ensure that the error using one discretization method will be smaller than the error using another in a specific problem. Instead, we compare methods based on their asymptotic performance for small h .

Under modest conditions, even the simple Euler scheme converges (with respect to both strong and weak criteria) as the time step h decreases to zero. We therefore compare discretization schemes based on the *rate* at which they converge. Following Kloeden and Platen [211], we say that a discretization \hat{X} has *strong order of convergence* $\beta > 0$ if

$$\mathbb{E} \left[\|\hat{X}(nh) - X(T)\| \right] \leq ch^\beta \quad (6.13)$$

for some constant c and all sufficiently small h . The discretization scheme has *weak order of convergence* β if

$$\left| \mathbb{E}[f(\hat{X}(nh))] - \mathbb{E}[f(X(T))] \right| \leq ch^\beta \quad (6.14)$$

for some constant c and all sufficiently small h , for all f in a set $C_P^{2\beta+2}$. The set $C_P^{2\beta+2}$ consists of functions from \Re^d to \Re whose derivatives of order $0, 1, \dots, 2\beta + 2$ are polynomially bounded. A function $g : \Re^d \rightarrow \Re$ is polynomially bounded if

$$|g(x)| \leq k(1 + \|x\|^q)$$

for some constants k and q and all $x \in \Re^d$. The constant c in (6.14) may depend on f .

In both (6.13) and (6.14), a larger value of β implies faster convergence to zero of the discretization error. The same scheme will often have a smaller strong order of convergence than its weak order of convergence. For example, the Euler scheme typically has a strong order of $1/2$, but it often achieves a weak order of 1 .

Convergence Order of the Euler Scheme

In more detail, the Euler scheme has strong order $1/2$ under conditions only slightly stronger than those in Theorem B.2.1 of Appendix B.2 for existence

and uniqueness of a (strong) solution to the SDE (6.1). We may generalize (6.1) by allowing the coefficient functions a and b to depend explicitly on time t as well as on $X(t)$. Because X is vector-valued, we could alternatively take t to be one of the components of $X(t)$; but that formulation leads to unnecessarily strong conditions for convergence because it requires that the coefficients be as smooth in t as they are in X . In addition to the conditions of Theorem B.2.1, suppose that

$$\mathbb{E} \left[\|X(0) - \hat{X}(0)\|^2 \right] \leq K\sqrt{h} \quad (6.15)$$

and

$$\|a(x, s) - a(x, t)\| + \|b(x, s) - b(x, t)\| \leq K(1 + \|x\|)\sqrt{|t - s|}, \quad (6.16)$$

for some constant K ; then the Euler scheme has strong order 1/2. (This is proved in Kloeden and Platen [211], pp.342–344. It is observed in Milstein [266] though without explicit hypotheses.) Condition (6.15) is trivially satisfied if $X(0)$ is known and we set $\hat{X}(0)$ equal to it.

Stronger conditions are required for the Euler scheme to have weak order 1. For example, Theorem 14.5.2 of Kloeden and Platen [211] requires that the functions a and b be four times continuously differentiable with polynomially bounded derivatives. More generally, the Euler scheme has weak order β if a and b are $2(\beta+1)$ times continuously differentiable with polynomially bounded derivatives; the condition (6.14) then applies only to functions f with the same degree of smoothness.

To see how smoothness can lead to a higher weak order than strong order, consider the following argument. Suppose, for simplicity, that $T = nh$ and that $X(0)$ is fixed so that $\mathbb{E}[f(X(0))]$ is known. By writing

$$\mathbb{E}[f(X(T))] = \mathbb{E}[f(X(0))] + \mathbb{E} \left[\sum_{i=0}^{n-1} \mathbb{E}[f(X((i+1)h)) - f(X(ih)) | X(ih)] \right],$$

we see that accurate estimation of $\mathbb{E}[f(X(T))]$ follows from accurate estimation of the conditional expectations $\mathbb{E}[f(X((i+1)h)) - f(X(ih)) | X(ih)]$. Applying a Taylor approximation to f (and taking X scalar for simplicity), we get

$$\begin{aligned} & \mathbb{E}[f(X((i+1)h)) - f(X(ih)) | X(ih)] \\ & \approx \sum_{j=0}^r \frac{f^{(j)}(X(ih))}{j!} \mathbb{E}[(X((i+1)h) - X(ih))^j | X(ih)]. \end{aligned} \quad (6.17)$$

Thus, if f is sufficiently smooth, then to achieve a high order of weak convergence a discretization scheme need only approximate conditional *moments* of the increments of the process X . With sufficient smoothness in the coefficient functions a and b , higher conditional moments are of increasingly high order

in h . Smoothness conditions on a , b , and f leading to a weak order of convergence β for the Euler scheme follow from careful accounting of the errors in expanding f and approximating the conditional moments; see Kloeden and Platen [211], Section 14.5, and Talay [340, 341].

The accuracy of a discretization scheme in estimating an expression of the form $E[f(X(T))]$ does not necessarily extend to the simulation of other quantities associated with the same process. In Section 6.4 we discuss difficulties arising in simulating the maximum of a diffusion, for example. Talay and Zheng [344] analyze discretization error in estimating quantiles of the distribution of a component of $X(T)$. They provide very general conditions under which the bias in a quantile estimate computed from an Euler approximation is $O(h)$; but they also show that the implicit constant in this $O(h)$ error is large — especially in the tails of the distribution — and that this makes accurate quantile estimation difficult.

Convergence Order of the Refined Scheme

Theorem 10.3.5 of Kloeden and Platen [211] and Theorem 2-2 of Talay [340] provide conditions under which Milstein's refinement (6.10) and its multidimensional generalization based on (6.11) have strong order 1. The conditions required extend the linear growth, Lipschitz condition, and (6.16) to derivatives of the coefficient functions a and b . Thus, under these relatively modest additional conditions, expanding the diffusion term to $O(h)$ instead of just $O(\sqrt{h})$ through the derivation in Section 6.1.1 increases the order of strong convergence.

But the weak order of convergence of the refined scheme (6.10) is also 1, as it is for the Euler scheme. In this respect, including additional terms — as in (6.10) and (6.11) — does not result in greater accuracy. This should not be viewed as a deficiency of Milstein's method; rather, the Euler scheme is better than it "should" be, achieving order-1 weak convergence without expanding all terms to $O(h)$. This is in fact just the simplest example of a broader pattern of results on the number of terms required to achieve strong or weak convergence of a given order (to which we return in Section 6.3.1). In order to achieve a weak order greater than that of the Euler scheme, we need to expand dt -integrals to order h^2 and stochastic integrals to order h . We carry this out in the next section to arrive at a method with a higher weak order of convergence.

It is reassuring to know that a discretization scheme has a high order of convergence, but before venturing into our next derivation we should take note of the fact that good accuracy on smooth functions may not be directly relevant to our intended applications: option payoffs are typically nondifferentiable. Bally and Talay [34] show that the weak order of the Euler scheme holds for very general f and Yan [357] analyzes SDEs with irregular coefficients, but most of the literature requires significant smoothness assumptions.

When applying higher-order discretization methods, it is essential to test the methods numerically.

6.2 Second-Order Methods

We now proceed to further refine the Euler scheme to arrive at a method with weak order 2. The derivation follows the approach used in Section 6.1.1, expanding the integrals of $a(X(t))$ and $b(X(t))$ to refine the Euler approximations in (6.4) and (6.5), but now we keep more terms in the expansions. We begin by assuming that in the SDE (6.1) both X and W are scalar.

6.2.1 The Scalar Case

To keep the notation manageable, we adopt some convenient shorthand. With the scalar SDE (6.1) defining X , we associate the operators

$$\mathcal{L}^0 = a \frac{d}{dx} + \frac{1}{2} b^2 \frac{d^2}{dx^2} \quad (6.18)$$

and

$$\mathcal{L}^1 = b \frac{d}{dx}, \quad (6.19)$$

meaning that for any twice differentiable f , we have

$$\mathcal{L}^0 f(x) = a(x)f'(x) + \frac{1}{2}b^2(x)f''(x)$$

and

$$\mathcal{L}^1 f(x) = b(x)f'(x).$$

This allows us to write Itô's formula as

$$df(X(t)) = \mathcal{L}^0 f(X(t)) dt + \mathcal{L}^1 f(X(t)) dW(t). \quad (6.20)$$

To accommodate functions $f(t, X(t))$ that depend explicitly on time, we would generalize (6.18) to

$$\mathcal{L}^0 = \frac{\partial}{\partial t} + a \frac{\partial}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2}{\partial x^2}.$$

As in Section 6.1.1, the key to deriving a discretization scheme lies in approximating the evolution of X over an interval $[t, t+h]$. We start from the representation

$$X(t+h) = X(t) + \int_t^{t+h} a(X(u)) du + \int_t^{t+h} b(X(u)) dW(u), \quad (6.21)$$

and approximate each of the two integrals on the right.

The Euler scheme approximates the first integral using the approximation $a(X(u)) \approx a(X(t))$ for $u \in [t, t+h]$. To derive a better approximation for $a(X(u))$, we start from the exact representation

$$a(X(u)) = a(X(t)) + \int_t^u \mathcal{L}^0 a(X(s)) ds + \int_t^u \mathcal{L}^1 a(X(s)) dW(s);$$

this is Itô's formula applied to $a(X(u))$. Next we apply the Euler approximation to each of the two integrals appearing in this representation; in other words, we set $\mathcal{L}^0 a(X(s)) \approx \mathcal{L}^0 a(X(t))$ and $\mathcal{L}^1 a(X(s)) \approx \mathcal{L}^1 a(X(t))$ for $s \in [t, u]$ to get

$$a(X(u)) \approx a(X(t)) + \mathcal{L}^0 a(X(t)) \int_t^u ds + \mathcal{L}^1 a(X(t)) \int_t^u dW(s).$$

Now we use this approximation in the first integral in (6.21) to get

$$\begin{aligned} & \int_t^{t+h} a(X(u)) du \\ & \approx a(X(t))h + \mathcal{L}^0 a(X(t)) \int_t^{t+h} \int_t^u ds du + \mathcal{L}^1 a(X(t)) \int_t^{t+h} \int_t^u dW(s) du \\ & \equiv a(X(t))h + \mathcal{L}^0 a(X(t))I_{(0,0)} + \mathcal{L}^1 a(X(t))I_{(1,0)}, \end{aligned} \quad (6.22)$$

with $I_{(0,0)}$ and $I_{(1,0)}$ denoting the indicated double integrals. This gives us our approximation to the first term in integral in (6.21).

We use corresponding steps for the second integral in (6.21). We approximate the integrand $b(X(u))$, $u \in [t, t+h]$ using

$$\begin{aligned} b(X(u)) &= b(X(t)) + \int_t^u \mathcal{L}^0 b(X(s)) ds + \int_t^u \mathcal{L}^1 b(X(s)) dW(s) \\ &\approx b(X(t)) + \mathcal{L}^0 b(X(t)) \int_t^u ds + \mathcal{L}^1 b(X(t)) \int_t^u dW(s) \end{aligned}$$

and thus approximate the integral as

$$\begin{aligned} & \int_t^{t+h} b(X(u)) dW(u) \\ & \approx b(X(t))[W(t+h) - W(t)] + \mathcal{L}^0 b(X(t)) \int_t^{t+h} \int_t^u ds dW(u) \\ & \quad + \mathcal{L}^1 b(X(t)) \int_t^{t+h} \int_t^u dW(s) dW(u) \\ & \equiv b(X(t))[W(t+h) - W(t)] + \mathcal{L}^0 b(X(t))I_{(0,1)} + \mathcal{L}^1 b(X(t))I_{(1,1)}. \end{aligned} \quad (6.23)$$

Once again, the $I_{(i,j)}$ denote the indicated double integrals.

If we combine (6.22) and (6.23) and make explicit the application of the operators \mathcal{L}^0 and \mathcal{L}^1 to a and b , we arrive at the approximation

$$\begin{aligned} X(t+h) \approx X(t) + ah + b\Delta W + (aa' + \frac{1}{2}b^2a'')I_{(0,0)} \\ + (ab' + \frac{1}{2}b^2b'')I_{(0,1)} + ba'I_{(1,0)} + bb'I_{(1,1)}, \end{aligned} \quad (6.24)$$

with $\Delta W = W(t+h) - W(t)$, and the functions a, b and their derivatives all evaluated at $X(t)$.

The Discretization Scheme

To turn the approximation in (6.24) into an implementable algorithm, we need to be able to simulate the double integrals $I_{(i,j)}$. Clearly,

$$I_{(0,0)} = \int_t^{t+h} \int_t^u ds du = \frac{1}{2}h^2.$$

From (6.9) we know that

$$I_{(1,1)} = \int_t^{t+h} [W(u) - W(t)] dW(u) = \frac{1}{2}[(\Delta W)^2 - h].$$

The term $I_{(0,1)}$ is

$$I_{(0,1)} = \int_t^{t+h} \int_t^u ds dW(u) = \int_t^{t+h} (u-t) dW(u).$$

Applying integration by parts (which can be justified by applying Itô's formula to $tW(t)$), we get

$$\begin{aligned} I_{(0,1)} &= hW(t+h) - \int_t^{t+h} W(u) du \\ &= h[W(t+h) - W(t)] - \int_t^{t+h} [W(u) - W(t)] du \\ &= h\Delta W - I_{(1,0)}. \end{aligned} \quad (6.25)$$

So, it only remains to examine

$$I_{(1,0)} = \int_t^{t+h} [W(u) - W(t)] du.$$

Given $W(t)$, the area $I_{(1,0)}$ and the increment $\Delta W = W(t+h) - W(t)$ are jointly normal. Each has conditional mean 0; the conditional variance of ΔW is h and that of $I_{(1,0)}$ is $h^3/3$ (see (3.48)). For their covariance, notice first that

$$\mathbb{E}[I_{(1,0)}|W(t), \Delta W] = \frac{1}{2}h\Delta W \quad (6.26)$$

(as illustrated in Figure 6.1), so $\mathbb{E}[I_{(1,0)}\Delta W] = \frac{1}{2}h^2$. We may therefore simulate $W(t+h) - W(t)$ and $I_{(1,0)}$ as

$$\begin{pmatrix} \Delta W \\ \Delta I \end{pmatrix} \sim N \left(0, \begin{pmatrix} h & \frac{1}{2}h^2 \\ \frac{1}{2}h^2 & \frac{1}{3}h^3 \end{pmatrix} \right). \quad (6.27)$$

This leads to the following second-order scheme:

$$\begin{aligned} \hat{X}((i+1)h) = & \hat{X}(ih) + ah + b\Delta W + (ab' + \frac{1}{2}b^2b'')[\Delta Wh - \Delta I] \\ & + a'b\Delta I + \frac{1}{2}bb'[\Delta W^2 - h] \\ & + (aa' + \frac{1}{2}b^2a'')\frac{1}{2}h^2, \end{aligned} \quad (6.28)$$

with the functions a , b and their derivatives all evaluated at $\hat{X}(ih)$.

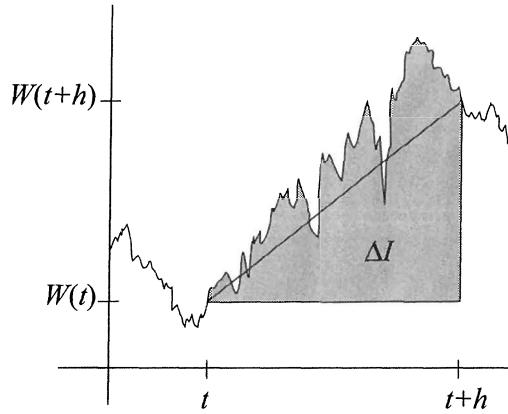


Fig. 6.1. The shaded area is ΔI . Given $W(t)$ and $W(t+h)$, the conditional expectation of W at any intermediate time lies on the straight line connecting these endpoints. The conditional expectation of ΔI is given by the area of the triangle with base h and height $\Delta W = W(t+h) - W(t)$.

This method was introduced by Milstein [267] in a slightly different form. Talay [341] shows that Milstein's scheme has weak order 2 under conditions on the coefficient functions a and b . These conditions include the requirement that the functions a and b be six times continuously differentiable with uniformly bounded derivatives. The result continues to hold if ΔI is replaced by its conditional expectation $\Delta Wh/2$; this type of simplification becomes essential in the vector case, as we explain in the next section.

Implementation of (6.28) and similar methods requires calculation of the derivatives of the coefficient functions of a diffusion. Methods that use difference approximations to avoid derivative calculations without a loss in convergence order are developed in Milstein [267] and Talay [341]. These types of approximations are called Runge-Kutta methods in analogy with methods used in the numerical solution of ordinary differential equations.

6.2.2 The Vector Case

We now extend the scheme in (6.28) to d -dimensional X driven by m -dimensional W . Much as in the scalar case, we start from the representation

$$X_i(t+h) = X_i(t) + \int_t^{t+h} a_i(u) du + \sum_{k=1}^m \int_t^{t+h} b_{ik}(u) dW_k(u), \quad i = 1, \dots, d,$$

and approximate each of the integrals on the right. In this setting, the relevant operators are

$$\mathcal{L}^0 = \frac{\partial}{\partial t} + \sum_{i=1}^d a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^m b_{ik} b_{jk} \frac{\partial^2}{\partial x_i \partial x_j} \quad (6.29)$$

and

$$\mathcal{L}^k = \sum_{i=1}^d b_{ik} \frac{\partial}{\partial x_i}, \quad k = 1, \dots, m. \quad (6.30)$$

The multidimensional Itô formula for twice continuously differentiable $f : \mathbb{R}^d \rightarrow \mathbb{R}$ becomes

$$df(X(t)) = \mathcal{L}^0 f(X(t)) dt + \sum_{k=1}^m \mathcal{L}^k f(X(t)) dW_k(t). \quad (6.31)$$

Applying (6.31) to a_i , we get

$$a_i(X(u)) = a_i(X(t)) + \int_t^u \mathcal{L}^0 a_i(X(s)) ds + \sum_{k=1}^m \int_t^u \mathcal{L}^k a_i(X(s)) dW_k(s).$$

The same steps leading to the approximation (6.22) in the scalar case now yield the approximation

$$\int_t^{t+h} a_i(X(u)) du \approx a_i(X(t))h + \mathcal{L}^0 a_i(X(t)) I_{(0,0)} + \sum_{k=1}^m \mathcal{L}^k a_i(X(t)) I_{(k,0)},$$

with

$$I_{(k,0)} = \int_t^{t+h} \int_t^u dW_k(s) du, \quad k = 1, \dots, m.$$

Similarly, the representation

$$b_{ik}(X(u)) = b_{ik}(X(t)) + \int_t^u \mathcal{L}^0 b_{ik}(X(s)) ds + \sum_{j=1}^m \int_t^u \mathcal{L}^j b_{ik}(X(s)) dW_j(s),$$

leads to the approximation

$$\begin{aligned} & \int_t^{t+h} b_{ik}(X(u)) dW_k(u) \\ & \approx b_{ik}(X(t))h + \mathcal{L}^0 b_{ik}(X(t)) I_{(0,k)} + \sum_{j=1}^m \mathcal{L}^j b_{ik}(X(t)) I_{(j,k)}, \end{aligned}$$

with

$$I_{(0,k)} = \int_t^{t+h} \int_t^u ds dW_k(u), \quad k = 1, \dots, m,$$

and

$$I_{(j,k)} = \int_t^{t+h} \int_t^u dW_j(u) dW_k(u), \quad j, k = 1, \dots, m.$$

The notational convention for these integrals should be evident: in $I_{(j,k)}$ we integrate first over W_j and then over W_k . This interpretation extends to $j = 0$ if we set $W_0(t) \equiv t$.

By combining the expansions above for the integrals of a_i and b_{ik} , we arrive at the discretization

$$\begin{aligned} \hat{X}_i(t+h) &= \hat{X}_i(t) + a_i(\hat{X}(t))h + \sum_{k=1}^m b_{ik}(\hat{X}(t))\Delta W_k \\ &\quad + \frac{1}{2}\mathcal{L}^0 a_i(\hat{X}(t))h^2 + \sum_{k=1}^m \mathcal{L}^k a_i(\hat{X}(t))I_{(k,0)} \\ &\quad + \sum_{k=1}^m \left(\mathcal{L}^0 b_{ik}(\hat{X}(t))I_{(0,k)} + \sum_{j=1}^m \mathcal{L}^j b_{ik}(\hat{X}(t))I_{(j,k)} \right), \end{aligned} \quad (6.32)$$

for each $i = 1, \dots, d$. Here we have substituted $h^2/2$ for $I_{(0,0)}$ and abbreviated $W_k(t+h) - W_k(t)$ as ΔW_k . The application of each of the operators \mathcal{L}^j to any of the coefficient functions a_i , b_{ik} produces a polynomial in the coefficient functions and their derivatives; these expressions can be made explicit using (6.29) and (6.30). Using the identity

$$I_{(0,j)} + I_{(j,0)} = \Delta W_j h,$$

which follows from (6.25), we could rewrite all terms involving $I_{(0,j)}$ as multiples of $(\Delta W_j h - I_{(j,0)})$ instead. Thus, to implement (6.32) we need to sample, for each $j = 1, \dots, m$, the Brownian increments ΔW_j together with the integrals $I_{(j,0)}$ and $I_{(j,k)}$, $k = 1, \dots, m$. We address this issue next.

Commutativity Condition

As noted in Section 6.1.1, the mixed Brownian integrals $I_{(j,k)}$ with $j \neq k$ are difficult to simulate, so (6.32) does not provide a practical algorithm without further simplification. Simulation of the mixed integrals is obviated in models satisfying the *commutativity condition*

$$\mathcal{L}^k b_{ij} = \mathcal{L}^j b_{ik} \quad (6.33)$$

for all $i = 1, \dots, d$. This is a rather artificial condition and is not often satisfied in practice, but it provides an interesting simplification of the second-order approximation.

When (6.33) holds, we may group terms in (6.32) involving mixed integrals $I_{(j,k)}$, $j, k \geq 1$, and write them as

$$\sum_{j=1}^m \sum_{k=1}^m \mathcal{L}^j b_{ik} I_{(j,k)} = \sum_{j=1}^m \mathcal{L}^j b_{ij} I_{(j,j)} + \sum_{j=1}^m \sum_{k=j+1}^m \mathcal{L}^j b_{ik} (I_{(j,k)} + I_{(k,j)}).$$

As in the scalar case (6.9), the diagonal term $I_{(j,j)}$ evaluates to $(\Delta W_j^2 - h)/2$ and is thus easy to simulate. The utility of the commutativity condition lies in the observation that even though each $I_{(j,k)}$, $j \neq k$, is difficult to simulate, the required sums simplify to

$$I_{(j,k)} + I_{(k,j)} = \Delta W_j \Delta W_k. \quad (6.34)$$

This follows from applying Itô's formula to $W_j(t)W_k(t)$ to get

$$W_j(t+h)W_k(t+h) - W_j(t)W_k(t) = \int_t^{t+h} W_k(u) dW_j(u) + \int_t^{t+h} W_j(u) dW_k(u)$$

and then subtracting $W_k(t)\Delta W_j + W_j(t)\Delta W_k$ from both sides.

When the commutativity condition is satisfied, the discretization scheme (6.32) thus simplifies to

$$\begin{aligned} \hat{X}_i(t+h) &= \hat{X}_i(t) + a_i(\hat{X}(t))h + \sum_{k=1}^m b_{ik}(\hat{X}(t))\Delta W_k + \frac{1}{2}\mathcal{L}^0 a_i(\hat{X}(t))h^2 \\ &+ \sum_{k=1}^m \left([\mathcal{L}^k a_i(\hat{X}(t)) - \mathcal{L}^0 b_{ik}(\hat{X}(t))] \Delta I_k + \mathcal{L}^0 b_{ik}(\hat{X}(t))\Delta W_k h \right) \\ &+ \sum_{j=1}^m \left(\mathcal{L}^j b_{ij}(\hat{X}(t)) \frac{1}{2}(\Delta W_j^2 - h) + \sum_{k=j+1}^m \mathcal{L}^j b_{ik}(\hat{X}(t))\Delta W_j \Delta W_k \right), \end{aligned} \quad (6.35)$$

with $\Delta I_k = I_{(k,0)}$. Because the components of W are independent of each other, the pairs $(\Delta W_k, \Delta I_k)$, $k = 1, \dots, m$, are independent of each other. Each such pair has the bivariate normal distribution identified in (6.27) and is thus easy to simulate.

Example 6.2.1 LIBOR Market Model. As an illustration of the commutativity condition (6.33), we consider the LIBOR market model of Section 3.7. Thus, take X_i to be the i th forward rate L_i in the spot measure dynamics in (3.112). This specifies that the evolution of L_i is governed by an SDE of the form

$$dL_i(t) = L_i(t)\mu_i(L(t), t) dt + L_i(t)\sigma_i(t)^\top dW(t),$$

with, for example, σ_i a deterministic function of time. In the notation of this section, $b_{ij} = L_i \sigma_{ij}$. The commutativity condition (6.33) requires

$$\sum_{r=1}^d b_{rk} \frac{\partial b_{ij}}{\partial x_r} = \sum_{r=1}^d b_{rj} \frac{\partial b_{ik}}{\partial x_r},$$

and this is satisfied because both sides evaluate to $\sigma_{ij}\sigma_{ik}L_i$. More generally, the commutativity condition is satisfied whenever $b_{ij}(X(t))$ factors as the product of a function of $X_i(t)$ and a deterministic function of time.

If we set $X_i(t) = \log L_i(t)$ then X solves an SDE of the form

$$dX_i(t) = (\mu_i(X(t), t) - \frac{1}{2}\|\sigma_i(t)\|^2) dt + \sigma_i(t)^\top dW(t).$$

In this case, $b_{ij} = \sigma_{ij}$ does not depend on X at all so the commutativity condition is automatically satisfied. \square

A Simplified Scheme

Even when the commutativity condition fails, the discretization method (6.32) can be simplified for practical implementation. Talay [340] and Kloeden and Platen [211, p.465] show that the scheme continues to have weak order 2 if each ΔI_j is replaced with $\frac{1}{2}\Delta W_j h$. (Related simplifications are used in Milstein [267] and Talay [341].) Observe from (6.26) that this amounts to replacing ΔI_j with its conditional expectation given ΔW_j . As a consequence, $\frac{1}{2}\Delta W_j h$ has the same covariance with ΔW_j as ΔI_j does:

$$\mathbb{E}[\Delta W_j \cdot \frac{1}{2}\Delta W_j h] = \frac{1}{2}h\mathbb{E}[\Delta W_j^2] = \frac{1}{2}h^2.$$

It also has the same mean as ΔI_j but variance $h^3/4$ rather than $h^3/3$, an error of $O(h^3)$. This turns out to be close enough to preserve the order of convergence. In the scalar case (6.28), the simplified scheme is

$$\begin{aligned} \hat{X}(n+1) &= \hat{X}(n) + ah + b\Delta W \\ &+ \frac{1}{2}(a'b + ab' + \frac{1}{2}b^2b'')\Delta Wh + \frac{1}{2}bb'[\Delta W^2 - h] \\ &+ (aa' + \frac{1}{2}b^2a'')\frac{1}{2}h^2, \end{aligned} \tag{6.36}$$

with a , b , and their derivatives evaluated at $\hat{X}(n)$.

In the vector case, the simplified scheme replaces the double integrals in (6.32) with simpler random variables. As in the scalar case, $I_{(0,k)}$ and $I_{(k,0)}$ are approximated by $\Delta W_k h/2$. Each $I_{(j,j)}$, $j \neq 0$, evaluates to $(\Delta W_j^2 - h)/2$. For j, k different from zero and from each other, $I_{(j,k)}$ is approximated by (Talay [341], Kloeden and Platen [211], Section 14.2)

$$\frac{1}{2}(\Delta W_j \Delta W_k - V_{jk}), \tag{6.37}$$

with $V_{kj} = -V_{jk}$, and the V_{jk} , $j < k$, independent random variables taking values h and $-h$ each with probability $1/2$. Let $V_{jj} = h$. The resulting approximation is, for each coordinate $i = 1, \dots, d$,

$$\begin{aligned}
\hat{X}_i(n+1) = & \\
& \hat{X}_i(n) + a_i h + \sum_{k=1}^m b_{ik} \Delta W_k + \frac{1}{2} \mathcal{L}^0 a_i h^2 + \frac{1}{2} \sum_{k=1}^m (\mathcal{L}^k a_i + \mathcal{L}^0 b_{ik}) \Delta W_k h \\
& + \frac{1}{2} \sum_{k=1}^m \sum_{j=1}^m \mathcal{L}^j b_{ik} (\Delta W_j \Delta W_k - V_{jk}),
\end{aligned} \tag{6.38}$$

with all a_i , b_{ij} , and their derivatives evaluated at $\hat{X}(n)$.

In these simplified schemes, the ΔW can be replaced with other random variables $\widehat{\Delta W}$ with moments up to order 5 that are within $O(h^3)$ of those of ΔW . (See the discussion following (6.17) and, for precise results Kloeden and Platen [211, p.465] and Talay [341, 342].) This includes the three-point distributions

$$P(\widehat{\Delta W} = \pm \sqrt{3h}) = \frac{1}{6}, \quad P(\widehat{\Delta W} = 0) = \frac{2}{3}.$$

These are faster to generate, but using normally distributed ΔW will generally result in smaller bias. The justification for using (6.37) also lies in the fact that these simpler random variables have moments up to order five that are within $O(h^3)$ of those of the $I_{(j,k)}$; see Section 5.12 of Kloeden and Platen [211, p.465], Section 1.6 of Talay [341], or Section 5 of Talay [342]. Talay [341, 342] calls these “Monte Carlo equivalent” families of random variables.

Example 6.2.2 Stochastic volatility model. In Section 3.4, we noted that the square-root diffusion is sometimes used to model stochastic volatility. Heston’s [179] model is

$$\begin{aligned}
dS(t) &= rS(t) dt + \sqrt{V(t)} S(t) dW_1(t) \\
dV(t) &= \kappa(\theta - V(t)) dt + \sqrt{V(t)} (\sigma_1 dW_1(t) + \sigma_2 dW_2(t)),
\end{aligned}$$

with S interpreted as, e.g., a stock price. The Brownian motions W_1 and W_2 are independent of each other. Heston [179] derives a formula for option prices in this setting using Fourier transform inversion. This provides a benchmark against which to compare simulation methods.

The simplified second-order scheme (6.38) for this model is as follows:

$$\begin{aligned}
\hat{S}(i+1) = & \hat{S}(i)(1 + rh + \sqrt{\hat{V}(i)} \Delta W_1) + \frac{1}{2} r^2 \hat{S}(i) h^2 \\
& + \left(\left[r + \frac{\sigma_1 - \kappa}{4} \right] \hat{S}(i) \sqrt{\hat{V}(i)} + \left[\frac{\kappa\theta}{4} - \frac{\sigma^2}{16} \right] \frac{\hat{S}(i)}{\sqrt{\hat{V}(i)}} \right) \Delta W_1 h \\
& + \frac{1}{2} \hat{S}(i) (\hat{V}(i) + \frac{\sigma_1}{2}) (\Delta W_1^2 - h) + \frac{1}{4} \sigma_2 \hat{S}(i) (\Delta W_2 \Delta W_1 + \xi)
\end{aligned}$$

and

$$\begin{aligned}
\hat{V}(i+1) = & \\
& \kappa\theta h + (1 - \kappa h)\hat{V}(i) + \sqrt{\hat{V}(i)}(\sigma_1 \Delta W_1 + \sigma_2 \Delta W_2) - \frac{1}{2}\kappa^2(\theta - \hat{V}(i))h^2 \\
& + \left(\left[\frac{\kappa\theta}{4} - \frac{\sigma^2}{16} \right] \frac{1}{\sqrt{\hat{V}(i)}} - \frac{3\kappa}{2}\sqrt{\hat{V}(i)} \right) (\sigma_1 \Delta W_1 + \sigma_2 \Delta W_2)h \\
& + \frac{1}{4}\sigma_1^2(\Delta W_1^2 - h) + \frac{1}{4}\sigma_2^2(\Delta W_2^2 - h) + \frac{1}{2}\sigma_1\sigma_2\Delta W_1\Delta W_2,
\end{aligned}$$

with $\sigma^2 = \sigma_1^2 + \sigma_2^2$ and ξ taking the values h and $-h$ with probability $1/2$ independent of the Brownian increments. To avoid taking the square root of a negative number or dividing by zero, we replace $\hat{V}(i)$ by its absolute value before advancing these recursions.

Figure 6.2 displays numerical results using this scheme and a simple Euler approximation. We use parameters $S(0) = 100$, $V(0) = 0.04$, $r = 5\%$, $\kappa = 1.2$, $\theta = 0.04$, $\sigma = 0.30$, and $\sigma_1 = \rho\sigma$ with $\rho = -0.5$. Using Heston's [179] formula, the expectation $E[e^{-rT}(S(T) - K)^+]$ with $T = 1$ and $K = 100$ evaluates to 10.3009. We compare our simulation results against this value to estimate bias. We use simulation time step $h = T/n$, with $n = 3, 6, 12, 25$, and 100 and run 2–4 million replications at each n for each method.

Figure 6.2 plots the estimated log absolute bias against $\log n$. The bias in the Euler scheme for this example falls below 0.01 at $n = 25$ steps per year, whereas the second-order method has a bias this small even at $n = 3$ steps per year. As n increases, the results for the Euler scheme look roughly consistent with first-order convergence; the second-order method produces smaller estimated biases but its convergence is much more erratic. In fact our use of (6.38) for this problem lacks theoretical support because the square-root functions in the model dynamics and the kink in the call option payoff violate the smoothness conditions required to ensure second-order convergence. The more regular convergence displayed by the Euler scheme in this example lends itself to the extrapolation method in Section 6.2.4.

6.2.3 Incorporating Path-Dependence

The error criterion in (6.14) applies to expectations of the form $E[f(X(T))]$ with T fixed. Accurate estimation of $E[f(X(T))]$ requires accurate approximation only of the distribution of $X(T)$. In many pricing problems, however, we are interested not only in the terminal state of an underlying process, but also in the path by which the terminal state is reached. The error criterion (6.14) does not appear to offer any guarantees on the approximation error in simulating functions of the path, raising the question of whether properties of the Euler and higher-order schemes extend to such functions.

One way to extend the framework of the previous sections to path-dependent quantities is to transform dependence on the past into dependence on supplementary state variables. This section illustrates this idea.