MN480 Lecture Notes

Jamie Alcock and Nicholas Denman
Department of Mathematics
The University of Queensland

July 5, 2000

Contents

List of Figures

Chapter 1

Stochastic Differential Equations

In the first parts of this chapter, we will revise the primary concepts pertaining to numerical analysis of ODEs and an introduction to stochastic differential equations (SDEs). It is assumed that the reader has taken MS479 and so is familiar with these concepts. For a more detailed treatment of these topics, the reader is referred to either Mark Thompson's MS479 notes, or to (Øksendal 1998).

1.1 Review of ODEs

Recall the initial value problem (IVP) described by an ordinary differential equation (ODE) and an initial starting point

$$\begin{aligned} \dot{\mathbf{y}}(t) &= f(t, \mathbf{y}(t)), \qquad \mathbf{y} \in \mathbb{R}^m \\ \mathbf{y}(t_0) &= \mathbf{y}_0. \end{aligned} \tag{1.1}$$

If Equation (1.1) is a function of $\mathbf{y}(t)$ only, that is (1.1) can be written

$$\dot{\mathbf{y}}(t) = f(\mathbf{y}(t)),$$

then the ODE is said to be autonomous.

Example 1.1.1 *The IVP*

$$\begin{aligned} \dot{\mathbf{y}}(t) &= qy(t) \qquad \mathbf{y} \in \mathbb{R} \\ \mathbf{y}(0) &= \mathbf{y}_0, \end{aligned}$$

is an autonomous equation and has the analytic solution

$$\mathbf{y}_t = \mathbf{y}_0 e^{qt}. \tag{1.2}$$

Example 1.1.2 *The IVP*

$$\begin{aligned} \dot{\mathbf{y}}(t) &= q(t)y(t) \qquad \mathbf{y} \in \mathbb{R} \\ \mathbf{y}(t_0) &= \mathbf{y}_0, \end{aligned}$$

is non-autonomous, so the analytic solution is related to the fundamental solution

$$\phi(t, t_0) = \exp\left(\int_{t_0}^{t} q(s)ds\right). \tag{1.3}$$

We note that any non-autonomous m-dimensional ODE can be transformed into an autonomous $m+1$-dimensional ODE by the introduction of the equation

$$y'_{m+1}(t) = 1, \qquad y_{m+1}(t_0) = t_0. \tag{1.4}$$

When the ODE is linear, an analytic solution can easily be found. However when the ODE is non-linear, analytic solutions can be difficult to find. Consequently, numerical techniques are often used to find an approximate solution for non-linear ODEs. Many numerical schemes can be derived using Taylor series expansions.

Deterministic Taylor Series

Consider the solution of the ODE

$$\begin{aligned} \frac{dX_t}{dt} &= a(X_t), \qquad t \in [t_0, T] \\ X(t_0) &= x_{t_0}, \end{aligned}$$

which can be written in the integral form

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_s)ds.$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function and $a(X)$ be a sufficiently smooth function. Then by the chain rule,

$$\frac{d}{dt}f(X_t) = a(X_t)\frac{\partial}{\partial x}f(X_t). \tag{1.5}$$

At this point we introduce the deterministic operator $\mathcal{L}$

$$\mathcal{L}f = a\frac{\partial f}{\partial x},$$

so that $\mathcal{L}^2 f := \mathcal{L}(\mathcal{L}f)$. Then (1.5) can be expressed,

$$f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} \mathcal{L}f(X_s)ds. \quad t \in [t_0, T]. \tag{1.6}$$

We can use (1.6) recursively to define the deterministic Taylor series expansion of f. Let $f(x) = x$, then by (1.6),

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} \mathcal{L}f(X_s)ds \\ &= X_{t_0} + \int_{t_0}^{t} a(X_s)ds. \end{aligned} \tag{1.7}$$

Applying (1.6) to the integrand of (1.7) by letting $f(x) = a$ we obtain

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} \left[a(X_0) + \int_{t_0}^{s} \mathcal{L}a(X_r)dr \right] ds \\ &= X_{t_0} + a(X_0) \int_{t_0}^{t} ds + \int_{t_0}^{t} \int_{t_0}^{s} \mathcal{L}a(X_r)\, dr\, ds. \end{aligned}$$

We can derive the classical Taylor expansion of a general function $f \in \mathbb{C}^{n+1}$, in integral form by iterating in a like manner (applying (1.6) to the remaining integrands) n times,

$$\begin{aligned} f(X_t) = f(X_{t_0}) &+ \sum_{j=1}^{n} \frac{(t-t_0)^j}{j!} \mathcal{L}^j f(X_{t_0}) \\ &+ \int_{t_0}^{t} \cdots \int_{t_0}^{s_{n-1}} \mathcal{L}^{n+1} f(X_{s_1}) ds_1 \ldots ds_{n+1}. \end{aligned} \tag{1.8}$$

The last term in (1.8) is often referred to as *Taylor's Remainder term.*

Most readers will be familiar with the differential form of the Taylor expansion of a function $f(X_t, t) \in \mathbb{C}^{n+1}$, where X_t is not described by an ODE; that is

$$dX_t = dt.$$

In this case, $\mathcal{L}f = \frac{\partial f}{\partial x}$ and $\mathcal{L}^2 f = \frac{\partial^2 f}{\partial x^2}$ etc., giving

$$\begin{aligned} f(X_t) &= f(X_{t_0}) + \sum_{j=1}^{n} \frac{(t-t_0)^j}{j!} \mathcal{L}^j f(X_{t_0}) \\ &\quad + \int_{t_0}^{t} \cdots \int_{t_0}^{s_{n-1}} \mathcal{L}^{n+1} f(X_{s_1}) ds_1 \ldots ds_{n+1} \\ &= f(X_{t_0}) + \frac{(t-t_0)}{1!} f'(X_{t_0}) + \frac{(t-t_0)^2}{2!} f''(X_{t_0}) \\ &\quad + \frac{(t-t_0)^3}{3!} f'''(X_{t_0}) + \cdots + \frac{(t-t_0)^n}{n!} f^{(n)}(X_{t_0}) + \text{Remainder}. \end{aligned}$$

Taylor's formula (1.8) can be used to generate numerical schemes to solve ODEs by truncating the expansion at a particular term. The simplest such numerical technique is the *Euler method*.

The Euler Method

The Euler method is a discrete numerical method to approximate the analytic solution of the IVP

$$\begin{aligned} \dot{\mathbf{y}}(t) &= f(\mathbf{y}(t)), \qquad \mathbf{y} \in \mathbb{R}^m \\ \mathbf{y}(t_0) &= \mathbf{y}_0. \end{aligned}$$

As we have already stated, we can derive an entire family of discrete numerical methods (including the Euler method) by truncating the Taylor series and utilizing Taylor's Theorem.

First, we rewrite the Taylor series expansion (1.8) in differential form,

$$\begin{aligned} \mathbf{y}(t+h) &= \mathbf{y}(t) + \frac{h}{1!}\mathbf{y}'(t) + \frac{h^2}{2!}\mathbf{y}''(t) + \frac{h^3}{3!}\mathbf{y}'''(t) + \ldots \\ &= \mathbf{y}(t) + hf(\mathbf{y}(t)) + \frac{h^2}{2} f'(\mathbf{y}(t)) f(\mathbf{y}(t)) + \frac{h^3}{6} \left(f''(f,f) + f'(f'(f)) \right) + \ldots \end{aligned}$$

The Euler method is found by truncating the Taylor series at the first derivative, giving

$$\mathbf{y}(t+h) = \mathbf{y}(t) + hf(\mathbf{y}(t)) + e(o(h^2)),$$

where $e(o(h^2))$ is the error term - or Taylor's remainder term - which is of order h^2. So if we define $\mathbf{y}_n = \mathbf{y}(t)$ and $\mathbf{y}_{n+1} = \mathbf{y}(t+h)$ we obtain

$$\mathbf{y}_{n+1} = \mathbf{y}_n + hf(\mathbf{y}_n),$$

which has a local error (error at each step) of $o(h^2)$, giving a global error of $o(h)$.

Similarly if we truncate the Taylor series at the next term, we derive a numerical method with local error $o(h^3)$, and so on. However these higher order schemes require higher order derivatives to be calculated - which can be computationally expensive.

Stability, Stiffness and Implicit Methods

Numerical schemes for solving ODEs can be classified into either *explicit* or *implicit* methods. Explicit methods compute approximations that are dependent on previous approximations only, whereas implicit methods compute approximations which are dependent on previous and current approximations.

Example 1.1.3 (Euler Methods) *The Euler method presented earlier is also known as the explicit Euler method.*

The explicit Euler method is defined as

$$\mathbf{y}_{i+1} = \mathbf{y}_i + hf(\mathbf{y}_i),$$

where $\mathbf{y}_i = \mathbf{y}(t_i)$ *and* $\mathbf{y}_{i+1} = \mathbf{y}(t_i + h)$. *This method has a global error* $\approx o(h)$.

The implicit Euler method is defined as

$$\mathbf{y}_{i+1} = \mathbf{y}_i + hf(\mathbf{y}_{i+1}),$$

where $\mathbf{y}_i = \mathbf{y}(t_i)$ *and* $\mathbf{y}_{i+1} = \mathbf{y}(t_i + h)$. *This method also has a global error* $\approx o(h)$.

The implementation of the explicit method is trivial to calculate, whereas the implementation of the implicit Euler method involves solving a system of non-linear equations. If these two methods have the same global error, why does the implicit Euler method exist? The answer lies in the concepts of *stability* and *stiffness*.

We can introduce the concept of stability by way of an example.

Example 1.1.4 *Consider the IVP*

$$\begin{aligned} \dot{\mathbf{y}} &= \mathbf{q}\mathbf{y}, \qquad q < 0 \\ y(0) &= y_0. \end{aligned}$$

We know that the analytic solution is $\mathbf{y}(\mathbf{t}) = e^{\mathbf{q}t} y_0$. *The explicit Euler scheme gives*

$$\begin{aligned} \mathbf{y}_{i+1} &= \mathbf{y}_i + h\mathbf{q}\mathbf{y}_i \\ &= \mathbf{y}_i(1 + h\mathbf{q}). \end{aligned}$$

As the true solution decays, our numerical method must reflect this so we must have

$$|1 + h\mathbf{q}| \leq 1,$$

or

$$h \leq \frac{2}{|\mathbf{q}|}.$$

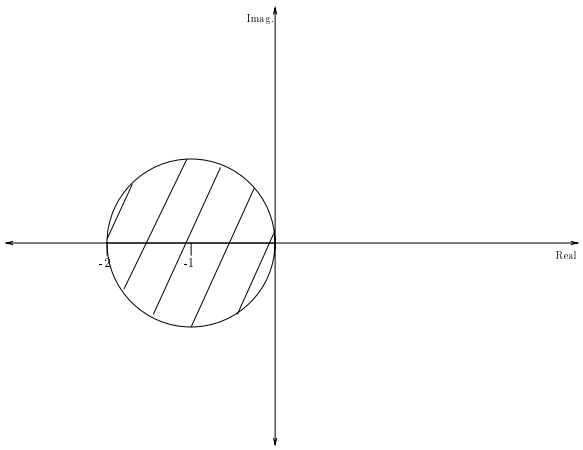


Figure 1.1: The region of stable $z = h\mathbf{q}$ values for the explicit Euler method

So if, for example, $\mathbf{q} = -2000000$ then $h \leq 10^{-6}$. That is, for stability, we have a very severe restriction on the step size.

In contrast, the same example using the implicit Euler method gives

$$\begin{aligned} \mathbf{y}_{i+1} &= \mathbf{y}_i + h\mathbf{q}\mathbf{y}_{i+1} \\ &= \frac{\mathbf{y}_i}{1 - h\mathbf{q}}. \end{aligned}$$

Again our numerical method must reflect that the true solution decays, so we must have

$$\left| \frac{1}{1 - h\mathbf{q}} \right| \leq 1$$

that is

$$|1 - h\mathbf{q}| = |1 - z| \geq 1. \tag{1.9}$$

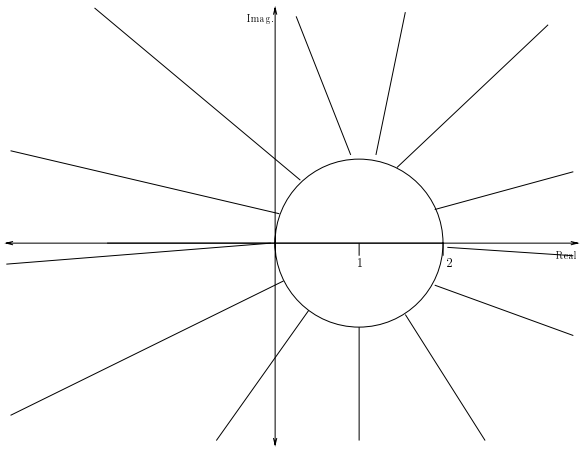


Figure 1.2: The region of stable $z = h\mathbf{q}$ values for the implicit Euler method

We can see the solution region for (1.9) in Figure 1.1.

So for $\mathbf{q} < 0$, $h\mathbf{q}$ will always be outside the circle in Figure 1.2 and so the implicit Euler method is stable for all h. Note that we will still have restrictions on the step size for reasons of accuracy - but not for stability.

For all discrete numerical methods, we can define a *stability region*, $\mathbf{S}$ as we have in the previous two examples. If the left half of the complex plane is a subset of $\mathbf{S}$, the method is said to be A-**stable**. So the implicit Euler method is A-stable but the explicit Euler method is not. A-stability is a highly desirable property for any numerical method to possess when solving *stiff* problems.

A stiff ODE results from processes with widely differing time scales. For example the general solution of a differential equation may involve sums of terms of the form ae^{ct} and be^{dt}, where both c and d are negative but c is much smaller than d. In such cases, using a larger value for the step size can introduce enough round-off errors to cause instability.

A stiff equation can be defined in terms of the eigenvalues of the coefficient matrix, A, of the system $\dot{y} = Ay$. If the eigenvalues have real parts that are all negative and differ widely in magnitude, the system is stiff. If the system is stiff, we should try to use an implicit A-stable method.

1.2 Analysis of Stochastic Differential Equations

1.2.1 Motivation of SDEs

Many systems studied in mathematics are described by a system of ODEs. A significant number of these ODEs are deterministic approximations to reality - chosen for mathematical convenience.

Consider the logistic equation, which arises in the study of population dynamics

$$\begin{aligned} \frac{dy}{dt} &= y(k-y) \\ y(0) &= y_0. \end{aligned} \tag{1.10}$$

The solutions of (1.10) are smooth deterministic paths, having steady states at $y = 0$ (unstable) and $y = k$ (stable).

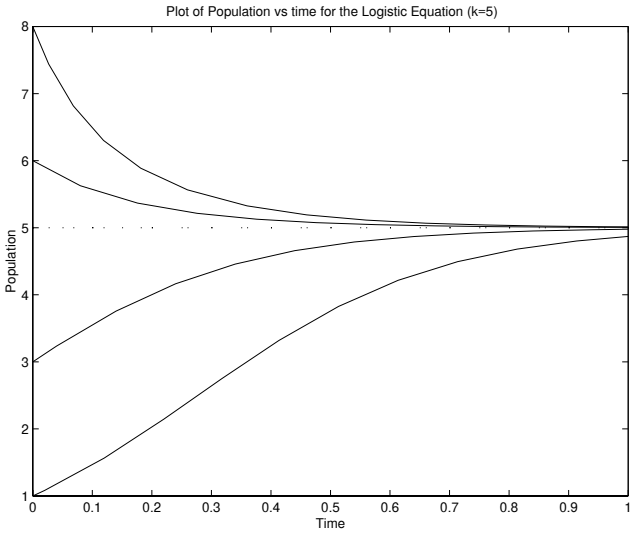


Figure 1.3: The Logistic Equation (k=5)

However the population of a species is far from deterministic. Many competing factors ensure that the population of a species may have an underlying trend that resembles Figure 1.3, but is nevertheless a random process about this path. An example of what a true population model might look like is given in Figure 1.4.

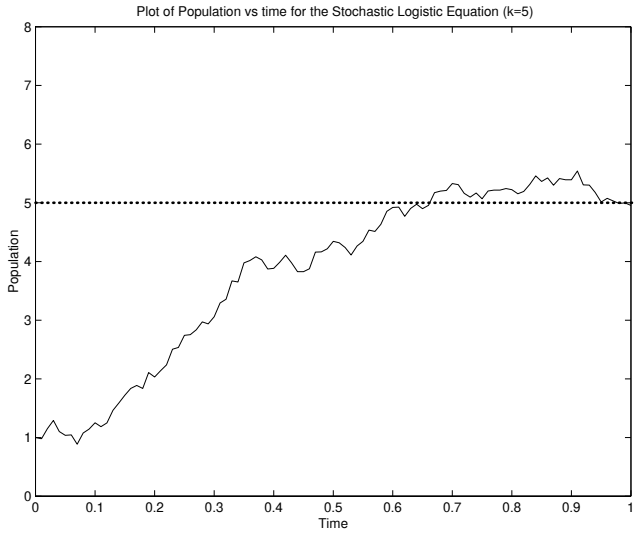


Figure 1.4: The Stochastic Logistic Equation - One Sample Path

Some population dynamicists may believe that modelling the underlying trend is adequate however there are reasons to study the more complete model other than simply mathematical correctness.

For example, we note in Figure 1.4 that the path is close to zero near $t = 0$. If the population suddenly decreased near $t = 0$, due to drought for example, and hit the time-axis [1], the path would be very different (see Figure 1.5). This possibility is not considered in the deterministic model.

Similarly environmental factors might change, implying that the carrying capacity of the model (the stable steady state, k) may vary randomly about the underlying mean, so that

$$k = k_0 + k_1 \Psi_t,$$

where Ψ_t is a white noise process. Equation (1.10) then becomes

$$\begin{aligned} dy &= y(k - y)dt \\ &= y(k_0 + k_1\Psi_t - y)dt \\ &= y(k_0 - y)dt + k_1 y \Psi_t dt \\ &= y(k_0 - y)dt + k_1 y dW_t \end{aligned} \tag{1.11}$$

where W_t refers to a Wiener process (aka Brownian motion). Note that (1.11)

[1] Indeed, we know that the event of a species eventually dying out occurs with probability 1.

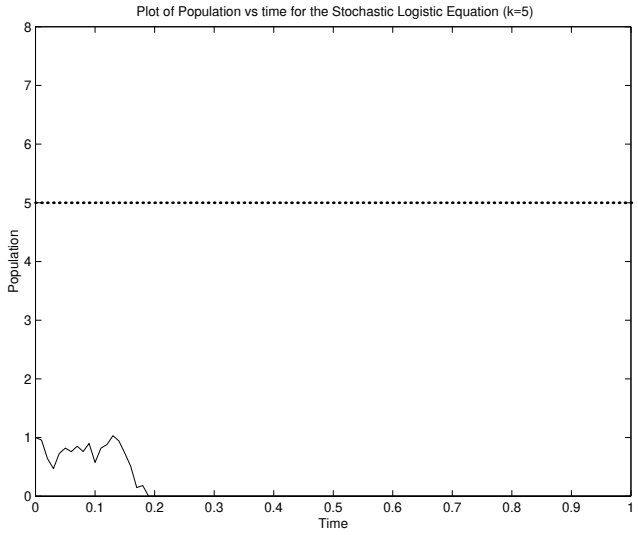


Figure 1.5: The Stochastic Logistic Equation - A Sample Path Resulting in Extinction

has both a deterministic part $[y(k_0 - y)dt]$ and a stochastic part $[k_1 y dW_t]$, which characterizes an SDE.

1.2.2 Stochastic Integration

Suppose we wish to describe the movement of a particle suspended in a moving liquid, subject to random molecular bombardments. Langevin showed it is quite easy to mathematically represent the motion of the particle by a stochastic differential equation of the form

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\xi_t \tag{1.12}$$

where $a(t, x) \in \mathbb{R}^3$ is the velocity of the particle and $\xi_t \in \mathbb{R}^3$ is 'white noise'. The Itô representation of the above equation is simply given by

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{1.13}$$

where W_t is 3-dimensional Brownian motion. For a stochastic differential equation of the form given in (1.13), we will call b the *drift coefficient* and σ the *diffusion coefficient*.
The solution, X_t, to (1.13) is given by

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{1.14}$$

In this section we will focus our attention on the existence and solution of

$$\text{“}\int_0^t \sigma(s, X_s) dW_s\text{”}. \tag{1.15}$$

Why does integration with respect to the Wiener process differ from ordinary integration? The problem with stochastic integration is that the paths of the Wiener process have, almost surely, an infinite variation. That is if we followed any segment of a Wiener path with a piece of string, we would almost surely need an infinitely long piece of string - even if we looked at a Wiener path over a very short time interval.

Riemann-Stieltjes integrals can only be defined when the integrator is of bounded (finite) variation. Stochastic integrals, on the other hand, rely on the integrator having bounded *quadratic variation*.

Definition 1.2.1 (Quadratic Variation) *The function h is said to have bounded* quadratic variation *if*

$$\sup_{\tau} \sum_{i=1}^{n} |h(t_i) - h(t_{i-1})|^2 < \infty,$$

where the supremum is taken over all partitions τ of $[0, T]$.

Without going into the fine detail of why the integrator needs to be defined in this way, we can make an intuitive connection by considering how we might construct the stochastic integral. It is natural to approximate the solution of (1.15) as

$$\int_0^T \sigma(s, X_s) dW_s \approx \lim_{n\to\infty} \sum_{j=0}^{n-1} \sigma(t_j^*, X_{t_j^*})(W_{t_{j+1}} - W_{t_j}), \quad t_j^* \in [t_j, t_{j+1}]. \tag{1.16}$$

As (1.15) is an integral with respect to a Wiener path and that the Wiener path can have more than one particular expression, then (1.15) is itself a random variable which can take on many different values (depending upon the particular Wiener path). Similarly our approximation (1.16) is also a random variable. So how do we prove that our intuitive approximation tends to the true integral?

In the Riemann-Stieltjes case we would prove that our approximation converges to the true integral as $n \to \infty$, however in the stochastic (integral) case we need to show that our approximation converges to the true integral *in the mean square sense*; that is we aim to show that

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\int_0^T \sigma(s, X_s) dW_s - \sum_{j=0}^{n-1} \sigma(t_j^*, X_{t_j^*})(W_{t_{j+1}} - W_{t_j})\right)^2\right] \to 0.$$

Example 1.2.1 *As an illustrative example, we examine the general case for the stochastic integral* $\int_a^b W_t dW_t$.

For the general stochastic integral, we can simplify the approximating sum

$$\begin{aligned}
\sum_{j=0}^{n-1} W_j^*(W_{j+1}-W_j) &= \sum_{j=0}^{n-1}(\theta W_j + (1-\theta)W_{j+1})(W_{j+1}-W_j) \\
&= \sum_{j=0}^{n-1}\left[\frac{1}{2}(W_{j+1}+W_j) - (\theta-\frac{1}{2})(W_{j+1}-W_j)\right](W_{j+1}-W_j) \qquad (1.17) \\
& \qquad\qquad (1.18) \\
&= \frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1}^2 - W_j^2) - (\theta-\frac{1}{2})\sum_{j=0}^{n-1}(W_{j+1}-W_j)^2 \\
&= \frac{1}{2}(W_n^2 - W_0^2) - (\theta-\frac{1}{2})\sum_{j=0}^{n-1}(\Delta W_j)^2. \qquad (1.19)
\end{aligned}$$

As $\sum_{j=1}^{n-1}(\Delta W_j)^2$ *converges in mean square to* $\sum_{j=1}^{n-1}\Delta t_j$, *that is*

$$\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{j=0}^{n-1}(\Delta W_j)^2 - \sum_{j=0}^{n-1}\Delta t_j\right)^2\right] = 0,$$

then (1.19), *and thus the approximating sum, converges in mean square to*

$$\frac{1}{2}(W_b^2 - W_a^2) - (\theta-\frac{1}{2})\sum_{j=0}^{n-1}\Delta t_j = \frac{1}{2}(W_n^2 - W_0^2) - (\theta-\frac{1}{2})(b-a).$$

That is,

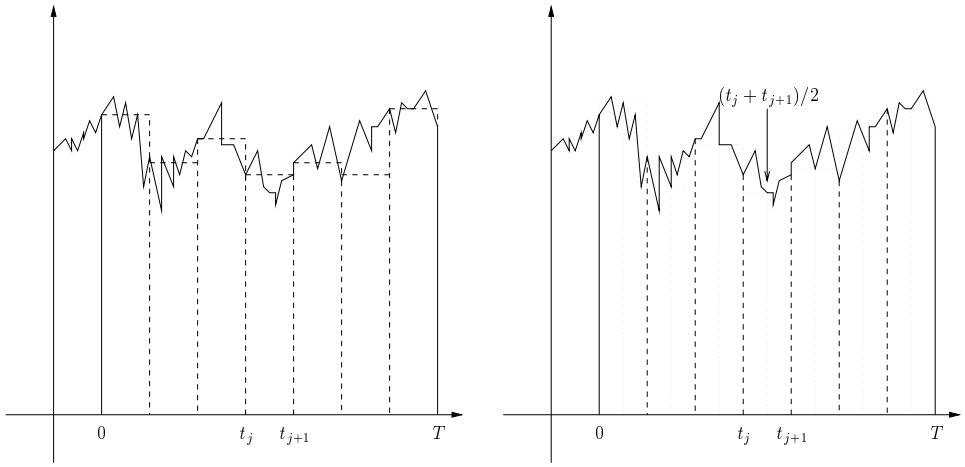
$$\lim_{n\to\infty}\mathbb{E}\left[\left(\int_a^b W_t^* dW_t - \left[\frac{1}{2}(W_n^2 - W_0^2) - (\theta-\frac{1}{2})(b-a)\right]\right)^2\right] = 0. \qquad (1.20)$$

When it comes to choosing t_j^* in (1.16) (or θ in (1.18)), two of the most common choices are:

- $t_j^* = t_j$ (or $\theta = 1$), which leads to the *Itô integral* , denoted by

$$\int_0^t \sigma(s, X_s) dW_s,$$

and



(a) Itô interpretation (b) Stratonovich interpretation

Figure 1.6: Plots representing the different partitions in the Itô and Stratonovich interpretation

- $t_j^* = (t_j + t_{j+1})/2$ (or $\theta = 0.5$), which leads to the *Stratonovich integral*, denoted by

$$\int_0^t \sigma(s, X_s) \circ dW_s.$$

The two different interpretations are shown in Figure 1.6. Unlike the Riemann case where it does not matter where the position of the t_j^* is, in the stochastic case it does matter.

We can see that we will obtain a different solution to our integral depending on which value we take for θ. If we choose $\theta = 1$ (Itô), we obtain[2]

$$\int_a^b W_t dW_t = \frac{W_b^2 - W_a^2 - (b-a)}{2},$$

however if we choose $\theta = \frac{1}{2}$ (Stratonovich), we obtain

$$\int_a^b W_t \circ dW_t = \frac{W_b^2 - W_a^2}{2}.$$

[2]From here on when we use the '=' sign in reference to a stochastic integral, we must remember that we are really saying that 'converges in the mean square sense'

We can also establish the relationship between the Itô and Stratonovich integrals, if we assume that

$$\int_0^T \mathbb{E}\left[f(W_t)\right]^2 dt < \infty \quad \text{and} \quad \int_0^T \mathbb{E}\left[f'(W_t)\right]^2 dt < \infty. \tag{1.21}$$

Theorem 1.2.1 (The Transformation Formula) *Assume the function f satisfies* (1.21)*. Then the transformation formula holds:*

$$\int_0^T f(W_t) \circ dW_t = \int_0^T f(W_t) dW_t + \frac{1}{2}\int_0^T f'(W_t) dt. \tag{1.22}$$

Sketch of Proof:

As for the Itô integral, we can represent the Stratonovich integral as an infinite series (in the mean square sense)

$$\int_0^T f(W_t) \circ dW_t = \lim_{n\to\infty} \sum_{i=1}^n f(W_{\tau_i}) \Delta W_t, \tag{1.23}$$

where $\tau_i = \frac{1}{2}(t_i + t_{i+1})$. We can expand the RHS of (1.23) using a Taylor series to obtain

$$\begin{aligned}
\sum_{i=1}^n f(W_{\tau_i})\Delta W_t &= \sum_{i=1}^n f(W_{t_{i-1}})\Delta W_t + \sum_{i=1}^n f'(W_{t_{i-1}})(W_{\tau_i} - W_{t_{i-1}})\Delta W_t + \dots \\
&= \sum_{i=1}^n f(W_{t_{i-1}})\Delta W_t + \sum_{i=1}^n f'(W_{t_{i-1}})(W_{\tau_i} - W_{t_{i-1}})^2 \\
&\qquad + \sum_{i=1}^n f'(W_{t_{i-1}})(W_{\tau_i} - W_{t_{i-1}})(W_{t_i} - W_{\tau_i}) + \dots
\end{aligned} \tag{1.24}$$

where we have neglected higher order terms. By definition of the Itô integral, the first term of (1.24) has the mean square limit $\int_0^T f(W_s) dW_s$. An application of condition (1.21) shows that the third term of (1.24) has the mean square limit zero,and it can also be shown that

$$\sum_{i=1}^n f'(W_{t_{i-1}})(W_{\tau_i} - W_{t_{i-1}})^2 \to \frac{1}{2}\int_0^T f'(W_t) dt$$

in mean square. Combining the mean square convergences of these three terms gives the required result. □

Throughout the rest of these notes, we will focus our attention on Itô integrals, although in some situations, the use of the Stratonovich integral may

be more appropriate. The reader may be wondering how do we choose which of the two interpretations to use. Both interpretations have their advantages and disadvantages. One of the nice features of the Itô integral is that is does not "look into the future", and since the Itô integral is a martingale this provides some nice computational advantages over the Stratonovich integral. For example

$$\begin{aligned}\mathbb{E}\left[\int_0^t g(X_s)dW_s\right] &= 0,\\ \mathbb{E}\left[\int_0^t g(X_s)dW_s\right]^2 &= \int_0^t \mathbb{E}\left[g(X_s)\right]^2 ds.\end{aligned}$$

However, the Stratonovich integral has the advantage of leading to ordinary chain rule formulas under transformation, making it natural to use in connection with stochastic differential equations on manifolds. Note that if $\sigma(t, X_t)$ does not depend on X_t, then the Itô interpretation and the Stratonovich interpretation are the same.

1.2.3 Analytic Methods for Solving SDEs

An important result due to Itô provides us with one method for determining the diffusion equation of $Y_t = g(t, X_t)$ given that X_t satisfies the diffusion equation given by (1.13).

Theorem 1.2.2 (Itô's formula) *Let X_t be an Itô process given by*

$$dX_t = u(t)dt + v(t)dW_t.$$

Let $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$, then

$$Y_t = g(t, X_t)$$

is again an Itô process, and

$$dY_t = \frac{\partial g(t,X_t)}{\partial t}dt + \frac{\partial g(t,X_t)}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g(t,X_t)}{dx^2}(dX_t)^2. \tag{1.25}$$

We will now give a sketch proof of Itô's formula.

Proof 1.2.1 *If we substitute dX_t into $g(t, X_t)$ we get*

$$\begin{aligned}g(t,X_t) &= g(0,X_0)\\ &\quad + \int_0^t \left(\frac{\partial g(s,X_s)}{\partial s} + u_s\frac{\partial g(s,X_s)}{\partial x} + u_s\frac{\partial g(s,X_s)}{\partial x} + \frac{1}{2}v_s^2\frac{\partial^2 g(\dot{s},X_s)}{\partial x^2}\right)ds\\ &\quad + \int_0^t v_s\frac{\partial g(\dot{s},X_s)}{\partial x}dW_s.\end{aligned}$$

By applying Taylor's theorem we get

$$\begin{aligned} g(t, X_t) &= g(0, X_0) + \sum_j \Delta g(t_j, X_{t_j}) \\ &= g(0, X_0) + \sum_j \frac{\partial g}{\partial t}\Delta t_j + \sum_j \frac{\partial g}{\partial x}\Delta X_j + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial t^2}(\Delta t_j)^2 \\ &\quad + \sum_j \frac{\partial^2 g}{\partial t \partial x}(\Delta t_j)(\Delta X_j) + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(\Delta X_j)^2 + \sum R_j \end{aligned}$$

where $R_j = o(|\Delta t_j|^2 + |\Delta X_j|^2)$ for all j. Thus, by ignoring terms of high order than $o(h)$ and noting that $(dW_t)^2 = dt$ we obtain Itô's formula. □

We will now focus our attention on methods to solve equations of the form given in (1.13). To illustrate one such method, we will look at a simple example making use of Itô's formula.

Example 1.2.2 (Geometric Brownian Motion) *Let X_t satisfy the stochastic differential equation*

$$\begin{aligned} dX_t &= \mu X_t dt + \sigma X_t dW_t \\ X(0) &= X_0. \end{aligned}$$

Hence,

$$\int_0^t \frac{dX_t}{X_t} = \mu t + \sigma W_t.$$

By applying Itô's formula to the left-hand side of the above equation, we obtain

$$\frac{dX_t}{X_t} = d(\ln X_t) + \frac{1}{2}\sigma^2 dt.$$

Hence,

$$\ln \frac{X_t}{X_0} = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t,$$

or

$$X_t = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right) X_0.$$

For comparison, the Stratonovich interpretation of

$$dX_t = \mu X_t dt + \sigma X_t \circ dW_t$$

is

$$X_t = \exp(\mu t + \sigma W_t) X_0.$$

Processes of this form are called geometric Brownian motions. *These models are very important as models for stochastic prices in economics.*

Here we have illustrated one of many methods for solving stochastic differential equations of the form given in (1.13). There exists methods of solution for stochastic differential equations that are linear in X_t, as well as methods that rely on Stratonovich calculus.

1.2.4 Fokker-Planck

Suppose we wish to determine the transitional probability distribution for a diffusion process, X_t, given by

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t.$$

That is, we are interested in finding an expression for determining the probability of being in a particular region in the future, given the current position, or

$$F(y, t|x, s) = P(W_t \leq y \mid W_s = x).$$

(Recall that the transitional density function $f(y, t \mid x, s) = \partial F(y, t|x, s)/\partial y$.)

When the drift and diffusion coefficients (a and σ resp.) of a diffusion process are moderately smooth functions, then its transition density, $f(y, t|x, s)$ satisfies other PDEs. These are the *Kolmogorov forward equation*

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial y}\left\{a(y, t)f\right\} - \frac{1}{2}\frac{\partial^2}{\partial y^2}\left\{\sigma^2(y, t)f\right\} = 0, \quad (x, s) \text{ fixed}, \tag{1.26}$$

and the *Kolmogorov backward equation*

$$\frac{\partial f}{\partial s} + a(y, t)\frac{\partial f}{\partial y} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial y^2} = 0, \quad (y, t) \text{ fixed}. \tag{1.27}$$

The first equation gives the forward evolution of the process from its starting point, whereas the second equation gives the backward evolution of the process from its end point. There is an enormous amount of theory surrounding these two equations and they can often be notoriously difficult to solve. The forward equation is often referred to in the physics and finance liturature as the *Fokker-Planck equation*, and it is the formal adjoint of the backward equation.

Exercise 1.2.1 *Verify that the transitional density function of the Wiener process,*

$$f(y,t|x,s) = [2\pi(t-s)]^{-1/2} \exp\left(-\frac{(y-x)^2}{2(t-s)}\right), \quad -\infty < y < \infty,$$

satisfies both the forward and backward equations and thus(subject to certain conditions) is the unique density function.

1.3 Numerical Simulation

We have seen one analytical method for solving SDEs. While other methods exist, SDEs are generally difficult to solve and so numerical simulations are often required.

Numerical simulations can help us approximate either a *strong* solution or a *weak* solution. A strong solution of an SDE is the trajectory of the solution, that is a sample path trajectory. Stong solutions are useful when we are interested in the dynamics of a system - whether there is a stationary solution or an attractor et cetera. A weak solution of an SDE refers to the statistics of the solution such as the expected value or the variance. To find a numerical approximation to the weak solution of an SDE, we compute many different strong solutions and calculate the relevant sample statistics.

So to compute either a weak or strong solution we need to compute a sample path, that is given an SDE

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dW_t,$$

we wish to plot a sample trajectory as in Figure 1.7.

For any time discrete approximation scheme that might be derived, we must define appropriate terminology to describe the accuracy of the numerical solutions - that is, to describe if the numerical scheme converges to the true solution of the SDE. We borrow some formal definitions from Kloeden & Platen (1992):

Definition 1.3.1 (Strong Convergence) *We say that a general time discrete approximation, Y_δ, with maximum time step-size δ* converges strongly *to X at time T if*

$$\lim_{\delta \downarrow 0} \mathbb{E}\left(|X_T - Y_\delta(T)|\right) = 0.$$

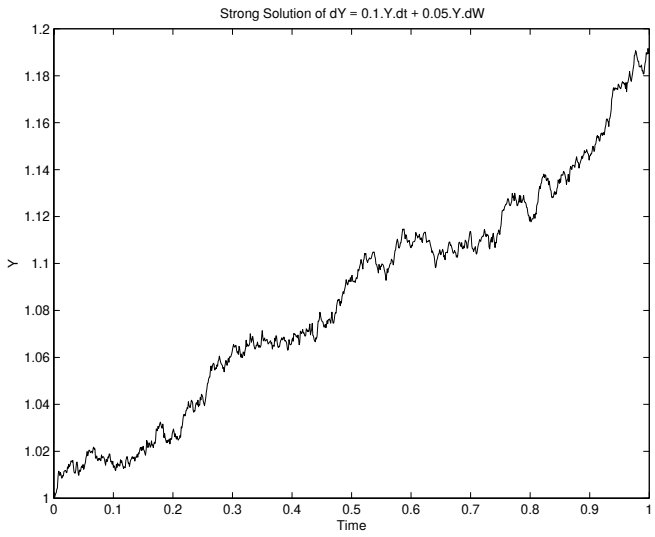


Figure 1.7: A Sample Trajectory - A Strong Solution

Definition 1.3.2 (Order of Strong Convergence) *We say that a general time discrete approximation,* Y_δ, converges strongly with order $\gamma > 0$ *at time* T *if there exists a positive constant* C, *which does not depend on* δ, *and a* $\delta_0 > 0$ *such that*

$$\mathbb{E}\left(|X_T - Y_\delta(T)|\right) \leq C\delta^\gamma,$$

for each $\delta \in (0, \delta_0)$.

Definition 1.3.3 (Weak Convergence) *We say that a general time discrete approximation,* Y_δ, *with maximum time step-size* δ *emphconverges weakly to* X *at time* T *as* $\delta \downarrow 0$ *with respect to a class* $\mathcal{C}$ *of test functions* $g : \mathbb{R}^d \to \mathbb{R}$ *if we have*

$$\lim_{\delta \downarrow 0} \left|\mathbb{E}\left(g(X_T)\right) - \mathbb{E}\left(g(Y_\delta(T))\right)\right| = 0,$$

for all $g \in \mathcal{C}$.

Definition 1.3.4 (Order of Weak Convergence) *We say that a general time discrete approximation,* Y_δ, converges weakly with order $\gamma > 0$ *at time* T *as* $\delta \downarrow 0$ *if for each* $g \in \mathcal{C}_p^{2(\beta+1)}(\mathbb{R}^d \to \mathbb{R})$ *there exists a positive constant* C, *which does not depend on* γ, *and a finite* $\delta_0 > 0$ *such that*

$$\left|\mathbb{E}\left(g(X_T)\right) - \mathbb{E}\left(g(Y_\delta(T))\right)\right| \leq C\delta^\beta,$$

for each $\delta \in (0, \delta_0)$.

1.3.1 Stochastic Taylor Series

As with the deterministic case, an entire family of numerical schemes for solving SDEs can be derived from a stochastic Taylor formula. There are several possibilities for a stochastic Taylor formula, most notably formulas using Itô or Stratonovich calculi. We shall introduce the Itô-Taylor expansion based on the iterated application of the Itô formula (1.25), as presented in Kloeden & Platen (1992). If we introduce the diffusion operators

$$\mathcal{A}f = a\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2} \quad and \quad \mathcal{B}f = \sigma\frac{\partial f}{\partial x},$$

then the Itô formula (1.25) can be re-written as

$$f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} \mathcal{A}f(X_s)ds + \int_{t_0}^{t} \mathcal{B}f(X_s)dW_s, \quad t \in [t_0, T]. \tag{1.28}$$

Clearly, for $f(x) = x$ we have $\mathcal{A}f = a$ and $\mathcal{B}f = \sigma$ giving the original Itô equation

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_s)ds + \int_{t_0}^{t} \sigma(X_s)dW_s. \tag{1.29}$$

Applying the Itô formula (1.28) to the function $f = a$ and $f = \sigma$ in (1.29), as we did in the deterministic case, yields

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} \left[a(X_{t_0}) + \int_{t_0}^{s} \mathcal{A}a(X_z)dz + \int_{t_0}^{s} \mathcal{B}a(X_z)dW_z \right] ds \\ &\quad + \int_{t_0}^{t} \left[\sigma(X_{t_0}) + \int_{t_0}^{s} \mathcal{A}\sigma(X_z)dz \int_{t_0}^{s} \mathcal{B}\sigma(X_z)dW_z \right] dW_s. \end{aligned} \tag{1.30}$$

We can continue, for instance, by applying the Itô formula to $f = \mathcal{B}b$ in (1.30) obtaining

$$X_t = X_{t_0} + a(X_{t_0}) \int_{t_0}^{t} ds + \sigma(X_{t_0}) \int_{t_0}^{t} dW_s + \mathcal{B}\sigma(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s} dW_z dW_s + R$$

where R is the Itô-Taylor remainder term. By continuing, we obtain the Itô-Taylor expansion of X_t. The Itô-Taylor expansion can, in a sense, be interpreted as a generalization of both the Itô formula and the deterministic Taylor expansion.

We can formulate the Itô-Taylor expansion for a general function f and for arbitrarily high order by employing a similar strategy. The reader is referred

to Chapter 5 of Kloeden & Platen (1992) for a more detailed treatment. However, we present here the Itô-Taylor expansion of $f(X_t)$, where X_t is given by (1.29), to the order which we will require in later sections:

$$\begin{aligned} f(X_t) &= f(X_{t_0}) + \int_{t_0}^{t} \mathcal{A}f(X_s)ds + \int_{t_0}^{t} \mathcal{B}f(X_s)dW_s + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{B}^2 f(X_z)dW_z.dW_s \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{B}\mathcal{A}f(X_z)dW_z ds + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{A}^2 f(X_z)dzds \\ &\quad + \text{Remainder.} \end{aligned} \tag{1.31}$$

For future reference, we also give the Itô-Taylor expansion for the function $f(x) = x$, in differential form. When $f(x) = x$, $\mathcal{A}f = a$ and $\mathcal{B}f = \sigma$ etc., giving

$$\begin{aligned} X_t &= X_{t_0} + a\int_{t_0}^{t} ds + \sigma\int_{t_0}^{t} dW_s \\ &\quad + \sigma'\sigma\int_{t_0}^{t}\int_{t_0}^{s} dW_z dW_s + a'\sigma\int_{t_0}^{t}\int_{t_0}^{s} dW_z ds \\ &\quad + (a'a + \frac{1}{2}a''\sigma^2)\int_{t_0}^{t}\int_{t_0}^{s} dzds + \text{Remainder} \\ &= X_{t_0} + a\Delta t + \sigma\Delta W_t + \frac{1}{2}\sigma'\sigma[(\Delta W_t)^2 - \Delta t] \\ &\quad + a'\sigma\Delta Z + (a'a + \frac{1}{2}a''\sigma^2)(\Delta t)^2 + \text{Remainder,} \end{aligned}$$

where ΔZ represents $\int_{t_0}^{t}\int_{t_0}^{s} dW_z ds$.

Kloeden & Platen (1992) also give proofs of convergence of truncated Itô-Taylor expansions - which allows us to utilize Itô-Taylor approximation schemes for solving SDEs numerically.

By using a similar methodology, we can also derive the Stratonovich-Taylor expansion of (1.29). The first few terms of this expansion are

$$\begin{aligned} X_t &= X_{t_0} + a\int_{t_0}^{t} ds + \sigma\int_{t_0}^{t} \circ dW_s \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{A}_S a(X_z)dzds + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{B}_S a(X_z) \circ dW_z ds \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{A}_S \sigma(X_z)dz \circ dW_s + \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{B}_S \sigma(X_z) \circ dW_z \circ dW_s \\ &\quad + \text{Remainder,} \end{aligned}$$

where the Stratonovich operators $\mathcal{A}_S f = a\frac{\partial f}{\partial x}$ and $\mathcal{B}_S f = \sigma\frac{\partial f}{\partial x}$ are used. While, for reasons already explained, we will not be referring to the Stratonovich-Taylor series anymore, some prefer it over the Itô-Taylor integral as it is easier to use.

1.3.2 Explicit Strong Taylor Approximations

Euler-Maruyama Method

All of the strong Taylor approximation schemes are derived by truncating the Itô-Taylor series at an appropriate term. The simplest of these methods is the *Euler-Maruyama* method, which is derived by truncating the Itô-Taylor expansion after one deterministic and one stochastic term. So given

$$\begin{aligned} dY_t &= a(t, Y_t)dt + \sigma(t, Y_t)dW_t, \\ Y(t_0) &= Y_0, \end{aligned} \tag{1.32}$$

we have the Euler-Maruyama approximation scheme

$$y_{n+1} = y_n + ha(t_n, y_n) + \Delta W_t \sigma(t_n, y_n), \tag{1.33}$$

where $\Delta W_t = (W_{\tau_{n+1}} - W_{\tau_n})$. We know that

$$\begin{aligned} W_{t_1} - W_{t_0} &\sim N(0, h) \\ \text{ie } \sqrt{h}\,(W_{t_1} - W_{t_0}) &\sim N(0, 1), \end{aligned}$$

so $\Delta W_t = \sqrt{h} \times R_n$, where $R_n \sim N(0, 1)$. Matlab code which implements the Euler-Maruyama method can be found in the appendix.

The Euler-Maruyama method can easily be extended to SDEs with more than one Brownian motion. Given the SDE

$$\begin{aligned} dY_t &= a(t, Y_t)dt + \sum_{j=1}^{d} \sigma_j(t, Y_t)dW_j, \\ Y(t_0) &= Y_0, \end{aligned} \tag{1.34}$$

the stochastic Euler method (1.33) becomes

$$y_{n+1} = y_n + ha(t_n, y_n) + \sum_{j=1}^{d} \Delta^{(j)} W_t \sigma_j(t_n, y_n). \tag{1.35}$$

In this case we will need to generate an $N \times d$ matrix of $i.i.d.N(0, h)$ random variables.

Kloeden & Platen (1992) show that as the increment size tends to zero, the numerical approximation tends to the true strong solution. They continue to prove that the global error rate is $o(\sqrt{h})$. That is, if the step length is reduced to a $h/4$, the global error reduces by a factor of $\left(\frac{1}{4}\right)^{0.5} = \frac{1}{2}$. This is not a terribly efficient algorithm. We can obtain more accurate methods by truncating the stochastic Taylor series at later terms.

Milstein Method

We can derive a numerical scheme which has a lower error rate by including more terms from the Itô-Taylor expansion. The Milstein scheme achieves a global error rate of $o(h)$ by adding just one more term to the Euler-Maruyama method.

The next term in the Itô-Taylor expansion (1.31) is $b'b \int_{t_0}^{t} \int_{t_0}^{s} dW_z dW_s$ which can be evaluated as

$$\begin{aligned} \int_0^t \int_0^s dW_z dW_s &= \int_0^t W_s dW_s \\ &= \frac{1}{2}\left(W_T^2 - T\right). \end{aligned} \tag{1.36}$$

Discretizing (1.36) as $\frac{1}{2}[(\Delta W)^2 - \Delta t]$, the Milstein scheme is

$$y_{n+1} = y_n + ha(y_n) + \Delta W_t \sigma(y_n) + \frac{1}{2}(\Delta W_t^2 - h)\sigma'(y_n)\sigma(y_n). \tag{1.37}$$

Matlab code which implements the Milstein method can be found in the appendix.

The Milstein scheme converges strongly as $h \to 0$ and has a global error rate $o(h)$.

We can continue to add extra terms from the Itô-Taylor expansion to our numerical schemes to obtain lower and lower error bounds.

Exercise 1.3.1 *Research the proof of error bound for the Euler-Maruyama scheme (Theorem* 10.2.2 *in Kloeden & Platen (1992)) and use a similar strategy to prove the error bound for the Milstein Scheme.*

A strong order 1.5 Taylor scheme has also been derived by Platen and Wagner by including even more terms from the Itô-Taylor expansion. In the

autonomous 1 dimensional case, the scheme is given by

$$\begin{aligned} y_{n+1} = y_n &+ ha + \Delta W_t b + \frac{1}{2}\left\{(\Delta W_t)^2 - h\right\}\sigma'\sigma + \Delta Z_t \sigma a' \\ &+ \frac{h^2}{2}\left(a'a + \frac{1}{2}a''\sigma^2\right) + \{h\Delta W_t - \Delta Z_t\}\left(\sigma' a + \frac{1}{2}\sigma''\sigma^2\right) \\ &+ \frac{1}{2}\left\{\frac{1}{3}(\Delta W_t)^2 - h\right\}\Delta W_t\left(\sigma''\sigma\sigma + (\sigma')^2\sigma\right) \end{aligned}$$

where ΔZ_t is used to represent the double integral

$$\Delta Z_t = \int_{t_0}^{t}\int_{t_0}^{s} dW_r ds.$$

It can be shown that ΔZ_t is normally distributed with mean $\mathbb{E}(\Delta Z_t) = 0$ and variance $\mathbb{E}((\Delta Z_t)^2) = h^3/3$ and covariance $\mathbb{E}(\Delta Z_t.\Delta W_t) = h^2/2$. We can simulate the pair of correlated normally distributed random variables, ΔZ_t and ΔW_t, by generating two standard normal random numbers, U_1 and U_2, by means of the transformation

$$\Delta W_t = U_1\sqrt{h}, \qquad \Delta Z_t = \frac{h^{3/2}}{2}\left(U_1 + \frac{U_2}{\sqrt{3}}\right).$$

1.3.3 Implicit Methods

As described earlier, explicit methods can suffer from stability issues. Implicit methods can be more stable and are necessary for the solution of stiff SDEs.

In Section 1.1 we saw that the implicit Euler method for solving ODEs was obtained by simply substituting the current iterate, y_{n+1}, for the previous iterate y_n. If we attempt to follow the same logic to derive an implicit Euler-Maruyama scheme for the SDE

$$dY_t = a(Y_t, t)dt + \sigma(Y_t, t)dW_t, \tag{1.38}$$

we obtain

$$\begin{aligned} y_{n+1} &= y_n + ha(y_{n+1}, \tau_{n+1}) + \Delta W_t \sigma(y_{n+1}, \tau_{n+1}) \\ &= y_n + hay_{n+1} + \Delta W_t \sigma y_{n+1} \\ &= \frac{y_n}{1 - ha - \Delta W_t \sigma}, \text{ in the case that (1.38) is linear.} \end{aligned} \tag{1.39}$$

However as W_t is unbounded, this scheme has the potential to be very unstable (in fact the first absolute moment, $\mathbb{E}(|Y_n|)$, does not exist).

An alternative to this is to introduce 'deterministically-implicit' schemes by using the current iterate, y_{n+1}, in the coefficients of the non-random terms and using the previous iterate, y_n, in the coefficients of the random terms. That is, we can define the (deterministically) implicit Euler-Maruyama scheme as

$$y_{n+1} = y_n + ha(y_{n+1}, \tau_{n+1}) + \Delta W_t \sigma(y_n, \tau_n). \tag{1.40}$$

In fact we can define a whole family of implicit Euler-Maruyama methods by introducing a factor, α, such that

$$y_{n+1} = y_n + h\left[\alpha a(y_{n+1}, \tau_{n+1}) + (1-\alpha)a(y_n, \tau_n)\right] + \Delta W_t \sigma(y_n, \tau_n).$$

Using this scheme with a value of $\alpha = 0$ describes the explicit Euler-Maruyama scheme.

Example 1.3.1 (Geometric Brownian Motion) *If we implement the implicit Euler-Maruyama scheme (with $\alpha = 1$) to solve* (1.38)*, we obtain*

$$\begin{aligned} y_{n+1} &= y_n + ha(y_{n+1}, \tau_{n+1}) + \Delta W_t \sigma(y_n, \tau_n) \\ &= y_n + hay_{n+1} + \Delta W_t \sigma y_n \\ &= \left(\frac{1 + \sigma \Delta W_t}{1 - ha}\right) y_n. \end{aligned}$$

This is a strongly converging scheme with errors of $o(h^{1/2})$ (Kloeden & Platen 1992).

If we implement the same implicit Euler-Maruyama scheme for general α to solve (1.38)*, we similarly obtain*

$$y_{n+1} = \left[\frac{1 + \sigma \Delta W_t + h(1-\alpha)}{1 - h\alpha a}\right] y_n. \tag{1.41}$$

This is also a strongly converging scheme with errors of $o(h^{1/2})$ (Kloeden & Platen 1992).

We can also derive a deterministically implicit Milstein scheme. Recall the explicit Milstein scheme for geometric Brownian motion is

$$y_{n+1} = y_n + hay_n + \Delta W_t \sigma y_n + \frac{1}{2}\sigma^2(\Delta W_t^2 - h)y_n.$$

We can analyze the term

$$\frac{1}{2}\sigma^2(\Delta W_t^2 - h)y_n = \frac{1}{2}\sigma^2 \Delta t^2 y_n^2 - \frac{h}{2}\sigma^2 y_n,$$

so we obtain the deteministically implicit Milstein scheme

$$\begin{aligned}\left(1 - ah + \frac{h}{2}\sigma^2\right) y_{n+1} &= \left(\sigma\Delta t + \frac{1}{2}\sigma^2\Delta t^2\right) y_n \\ \text{so } y_{n+1} &= \left(\frac{\sigma\Delta t + \frac{1}{2}\sigma^2\Delta t^2}{1 - ah + \frac{h}{2}\sigma^2}\right) y_n.\end{aligned}$$

Exercise 1.3.2 *Describe the family of (deterministically-) implicit Milstein schemes for numerically simulating an SDE and describe their convergence properties.*

The Balanced Method

Recently (1998), another ingenious alternative to fully implicit methods was introduced called the *Balanced method*. The Balanced method uses a technique already in use for solving ODEs called *splitting*. Let us examine the case of solving the SDE

$$dX_t = \sigma X_t dW_t. \tag{1.42}$$

As the stochastic term is "much larger" than the deterministic term, this can be considered stiff and so will require an implicit method to solve it efficiently. We might start with a non-stiff solver, such as the Euler-Maruyama scheme and try to 'implicitize' it using splitting. The Euler-Maruyama scheme for (1.42) is

$$y_{n+1} = y_n + \Delta W_n \sigma y_n. \tag{1.43}$$

Splitting involves adding and subtracting a variable in an advantageous manner, that is (1.43) becomes

$$\begin{aligned}y_{n+1} &= y_n + (\sigma y_n + V - V)\Delta W_n \\ y_{n+1} + V\Delta W_n &= y_n + \Delta W_n \sigma y_n + V\Delta W_n.\end{aligned} \tag{1.44}$$

The Balanced method utilizes splitting by setting

$$V = \sigma\frac{|\Delta W_n|}{\Delta W_n} y,$$

and by introducing an implicit term only on the LHS. Thus (1.44) becomes

$$\begin{aligned}y_{n+1} + \sigma\frac{|\Delta W_n|}{\Delta W_n}\Delta W_n y_{n+1} &= y_n + \sigma\Delta W_n y_n + \sigma\frac{|\Delta W_n|}{\Delta W_n}\Delta W_n y_n. \\ y_{n+1} &= y_n + \sigma\Delta W_n y_n + \sigma|\Delta W_n|(y_n - y_{n+1})\end{aligned} \tag{1.45}$$

One may ask why this method is any better than those already proposed. The reason that the Balanced method proves superior stems from the inherent difficulty with fully implicit methods. Recall that the fully implicit Euler-Maruyama method for (1.42) is

$$y_{n+1} = \frac{y_n}{1 - \sigma \Delta W_n}.$$

As we have already seen, if ΔW_n is close to $1/\sigma$ then the fully implicit method becomes unstable. However, this instability can only occur when ΔW_n is greater than zero (for positive σ). The beauty with the Balanced method is, as a little algebraic manipulation will show, when $\Delta W_n < 0$ (ie when $|\Delta W_n| = -\Delta W_n$) then (1.45) reduces to

$$y_{n+1} = \frac{y_n}{1 - \sigma \Delta W_n},$$

which is a stable fully implicit Euler method! When $\Delta W_n > 0$, the Balanced method gives a semi-implicit method:

$$y_{n+1} = \frac{y_n(1 + 2\sigma \Delta W_n)}{1 + \sigma \Delta W_n}.$$

That is, for 50% of the time the Balanced method is fully implicit, while for the other 50% of the time the Balanced method is semi-implicit!

For the general case of solving (1.32), we can define a family of Balanced methods as

$$y_{n+1} = y_n + ha(t_n, y_n) + \Delta W_n \sigma(t_n, y_n) + C_n(y_n - y_{n+1}),$$

where

$$C_n = hc^{(0)}(t_n, y_n) + |\Delta W_n| c^{(1)}(t_n, y_n).$$

In the case where X_t is a d-dimensional vector, then the functions $c^{(0)}$ and $c^{(1)}$ will ned to be $d \times d$ matrix functions. If $c^{(0)}$ and $c^{(1)}$ are positive definite and their components are uniformly bounded then the Balanced method converges with order 0.5.

1.3.4 Comparisons and Wiener Paths

Throughout the previous sections, we have introduced a number of numerical schemes to approximate a strong solution to an SDE. These methods have all required random variables to be simulated at each iteration of the scheme.

However, there is no requirement to generate these random numbers 'on the fly' - we can generate them in advance and use them sequentially, as needed. Once we have generated this sequence of random numbers and stored them in a vector, we can 'superimpose' the numerical scheme on top of them. This set of random numbers is called a Wiener path and an example of a Wiener path is given in Figure 1.8. As these random numbers are generated from

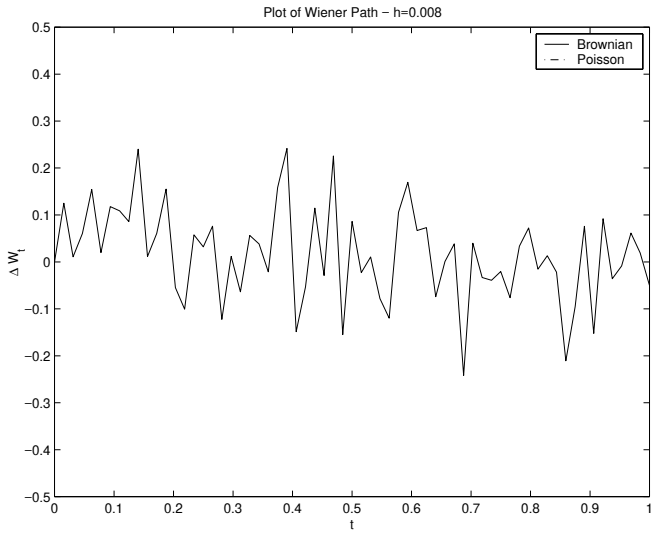


Figure 1.8: A Brownian Path (h=0.008)

a standard normal distribution, we expect that the asymptotic mean of the Wiener path be zero, and its asymptotic variance be one.

The benefit of implementing any numerical method in this way is that it allows us to compare different numerical methods by using the same Wiener path. For example, we can compare the Euler-Maruyama solution and Milstein solution to the exact solution of

$$\begin{aligned} dX_t &= 1.3X_t dt + 1.5X_t dW_t \\ X_0 &= 1, \end{aligned}$$

in Figure 1.9 . In Figure 1.9, we see that the two methods give a similar solution. However for a stiff SDE such as

$$\begin{aligned} dX_t &= \sigma X_t dW_t \\ X_0 &= 0, \end{aligned}$$

we can highlight the difference between the Euler-Maruyama method and a more stable method such as the Balanced method (see Figure 1.10). The

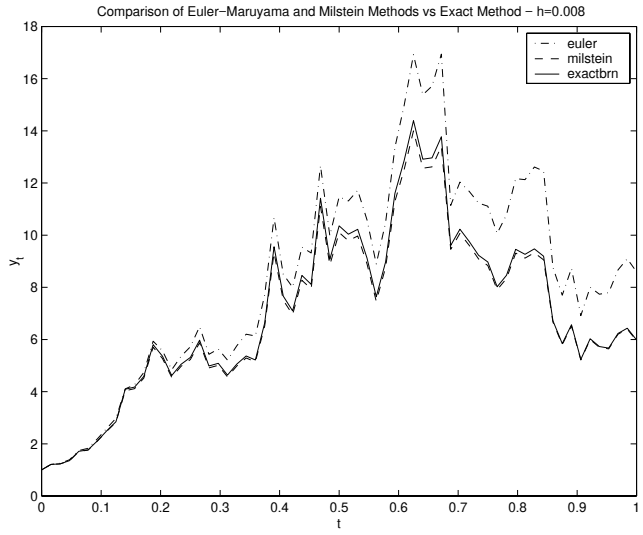


Figure 1.9: A Comparison of Euler-Maruyama and Milstein Methods for $dX_t = 1.3X_t dt + 1.5X_t dW_t$ using the Wiener path in Figure 1.8 (h=0.008)

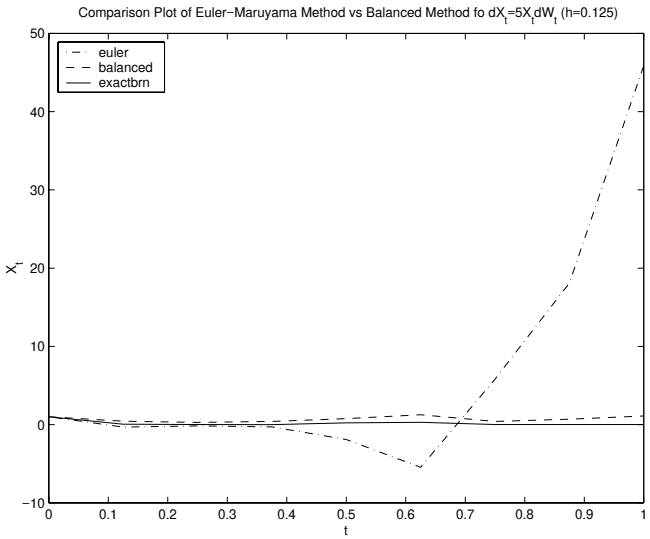


Figure 1.10: A Comparison of Euler-Maruyama and Balanced methods for $dX_t = 5X_t dW_t$ using the same Wiener path for all solutions ($h = 0.125$)

difference between the solutions in Figure 1.10 is quite stark (in some places the difference is greater than one!). This difference can get worse if we examine it closely (see Milstein, Platen & Schurz (1998) for more details).

Brownian Trees

Often we are also interested in noting the behaviour of a numerical method for decreasing step-sizes. In this case we need two Wiener paths, where one is embedded in the other. That is we wish to add more detail to an already existing Wiener path.

Given the Wiener path segment with step length h, we wish to create another Wiener path with step length $h/2$ that traverses the original Wiener path. This is done by generating further Wiener increments on subdivisions of the intervals, h. The set of Wiener paths that are constructed is called a Brownian tree.

A Brownian tree is made up of Wiener increments as follows:

$$\begin{array}{ccccccccccccccc} & & & & \Delta W_{1,1} & & & & & & \Delta W_{2,1} & & & & \\ & & & \swarrow & & \searrow & & & & \swarrow & & \searrow & & & \\ & & \Delta W_{1,2} & & & & \Delta W_{2,2} & & \Delta W_{3,2} & & & & \Delta W_{4,2} & & \\ & \swarrow & & \searrow & & \swarrow & & \searrow & & \swarrow & & \searrow & & \swarrow & & \searrow \\ \Delta W_{1,3} & & \Delta W_{2,3} & & \Delta W_{3,3} & & \Delta W_{4,3} & & \Delta W_{5,3} & & \Delta W_{6,3} & & \Delta W_{7,3} & & \Delta W_{8,3} \end{array} \tag{1.46}$$

where $\Delta W_{i,j}$ represents the i^{th} increment of the j^{th} level and where level one is the original Wiener path. The increments for level $j+1$ are computed for $j = 1, 2, \ldots,$ as

$$\begin{aligned} \Delta W_{2k-1,j+1} &= \frac{\Delta W_{k,j}}{2} + y_{k,j} \\ \Delta W_{2k,j+1} &= \Delta W_{k,j} - \Delta W_{2k-1,j+1}, \end{aligned}$$

where

$$y_{k,j} \sim N(0, \frac{h}{2^{2j}})$$

Example 1.3.2 *Consider the Wiener path, W of stepsize $h = 0.5$, which has the increments*

$$\Delta W_{1,1} = -0.3059$$

giving

$$W_{0,1} = 0 \quad W_{1,1} = -0.3059$$

To create a Wiener path of stepsize $h = 0.25$*, we generate* $n - 1 = 1$ *new random variables from a* $N(0, 0.125) = \sqrt{0.125} \times N(0, 1)$ *distribution. Suppose that the random number generated is* 0.7143*. Then the new set of Weiner increments is*

$$\begin{aligned} \Delta W_{1,2} &= \frac{\Delta W_{1,1}}{2} + \sqrt{0.125} \times 0.7143 \\ &= 0.0996 \\ \Delta W_{2,2} &= \Delta W_{1,1} - \Delta W_{1,2} \\ &= -0.3059 - 0.0996 \\ &= -0.4055 \end{aligned}$$

and the new Wiener path is

$$W_{0,2} = 0 \quad W_{1,2} = 0.0996 \quad W_{2,2} = -0.3059$$

The resulting Brownian tree is seen in Figure 1.11.

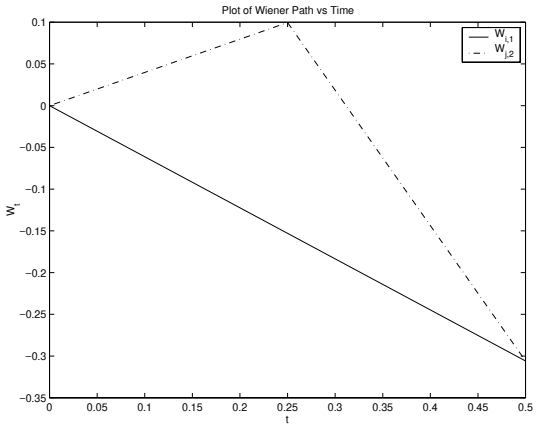


Figure 1.11: The Resultant Brownian Tree

If the same procedure is continued, the resulting Brownian tree is seen in Figure 1.12.

Once a Brownian tree has been generated, it can be used to estimate rates of convergence. By plotting the average absolute error for each stepsize on a log-log plot, an estimate of the scheme's rate of convergence can be obtained from the estimate of the slope of the log-log plot.

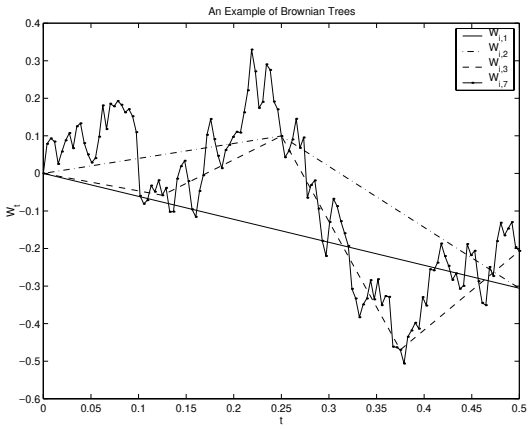


Figure 1.12: The Resultant Brownian Tree - 4 More Iterations

Exercise 1.3.3 *Formalize the above method of estimating the rate of convergence and implement this method to estimate the rate of convergence for both the Euler-Maruyama method and the Milstein method. Do these estimates correspond with theoretical rates of convergence obtained in Exercise 1.3.1?*

1.4 Summary

In this chapter, we have examined the theoretical concepts driving the development of numerical methods for ODEs and extended these concepts to the SDE case. We have also introduced various fixed stepsize numerical schemes to approximate a strong solution to SDEs and looked at different attempts to introduce implicit methods. Other relevant concepts which we have not discussed here include looking at variable stepsize implementation, non-Taylor-series based methods, more efficient weak approximation methods and other convergence issues (for example, could a fully implicit SDE solver ever converge to the solution of an Itô SDE?). Research in these areas is continuing and a lot of work is being done in this department. The reader is encouraged to investigate these issues further and discuss possible research topics in these areas with Prof. Burrage or Dr. Chandler.

Chapter 2

Black-Scholes Models

Black & Scholes (1973) derived an explicit solution to the problem of pricing and hedging a European put or call option on a non-dividend paying stock. Their paper was probably the most significant contribution to financial economics since the work of Markowitz on portfolio selection. The Black-Scholes formula can easily be used to derive formulas for option hedging parameters - delta, gamma, theta, vega and rho. Using simple arguments the Black-Scholes formula can be extended to options on futures and currencies as well as dividend paying stocks.

In this chapter we will derive the Black-Scholes model for European call and put options on non-dividend paying stock. We will also introduce the notion of an option (specifically European and American) as well as the idea of abritrage and how it can be applied to derive the Black-Scholes equation. We will then examine the 'Greeks' or option hedging parameters.

2.1 Options

An option gives its holder the right, but not the obligation, to buy or sell a certain amount of an asset by a certain date for a specific price, agreed upon before entering the contract. Before entering into the contract the writer must decide upon the type of option to be sold, the amount and type of underlying asset to be sold, the expiration date and the exercise price. Two of the most common types of options traded on the market include European and American options. A *European option* gives the holder the right, but not the obligation, to buy (a *call*) or sell (a *put*) the stock at some specified price K, on a certain date T. The time is often referred to as the *maturity date* or *expiry date*. An American option is similar to a European option except that it may be exercised at any time up to and including the expiry

date. American options will be discussed further in Chapter 4 as the problem of pricing an American option falls into the class of problems known as the optimal stopping problem.

Unlike the American option, the payoff for a European option only depends on the stock price at the expiry date, S_T. For a European call, if $S_T > K$ the holder would buy the stock at K and sell it for S_T and make a profit of $S_T - K$ dollars. If $S_T \leq K$, the option is worthless, and the holder would not wish to exercise the option. Thus, we can formally define the payoff of a European call option as

$$C(S,T) = \max(S_T - K, 0) =: (S_T - K)_+. \tag{2.1}$$

A similar argument leads to the formula for a European put, which is given by

$$P(S,T) = \max(K - S_T, 0) =: (K - S_T)_+. \tag{2.2}$$

The payoff diagrams for the European call and put are given in Figure 2.1 respectively. We will often abbreviate the notation for a European put and call to P_t and C_t respectively, suppressing the dependence on the underlying assets price.

In the following sections we will look at the questions of how to price and hedge the options correctly.

2.2 Arbitrage

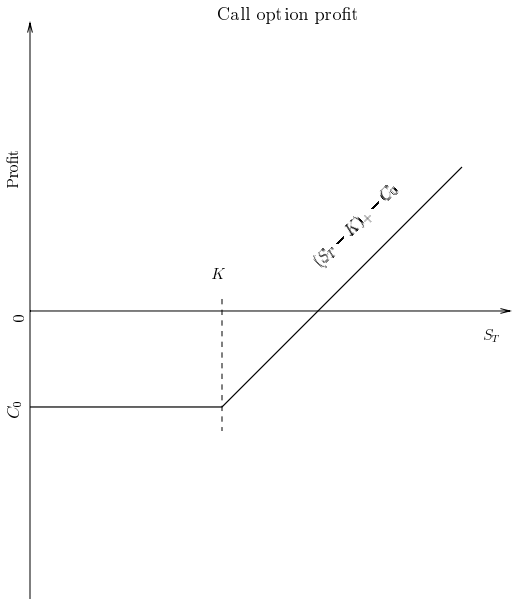
Arbitrage is the financial strategy whereby a riskless profit can be made. It could be as simple as buying from one trader at a lower price and selling to another trader at a higher price guaranteeing a risk-free profit.

To illustrate the idea of arbitrage it will be assumed that there exists a *risk-free interest rate*, r, at which traders can invest and borrow. If we invest in a government bond, worth B dollars, at the continuously compounding interest rate, it will be worth Be^{rT} dollars at time T.

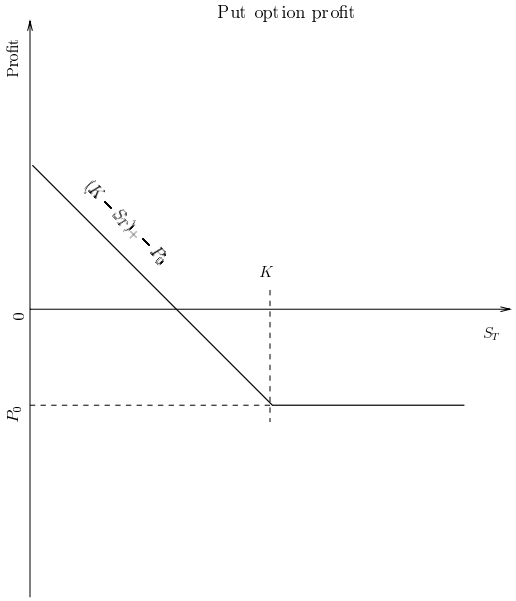
Suppose we are offered a forward contract to sell stock. What is the future strike price we should agree to?

One possible strategy is to borrow enough money to buy the stock now, wait until the date the contract matures, deliver the stock in exchange for the strike price as promised, and pay back the loan and interest accrued. Now if the strike price is greater than our borrowings we will have made a risk-free profit, and it is unlikely the buyer would have agreed to this.

A similar tactic can be used by the buyer. Sell the stock now and invest the cash at the risk-free interest rate. When the contract matures pay the



(a) Payoff diagram for a call



(b) Payoff diagram for a put

Figure 2.1: The Payoff diagrams for a European call and put option

strike price, receive the stock from the seller and pass it on to cover the original sale. If the cash plus interest earned is greater than the strike price, the buyer has an arbitrage opportunity to make money.

Thus the strike price for the forward contract must be equal to the cost of the stock initially plus the interest on this amount over the duration of the contract. The buyer and seller cannot agree on any other price.

2.3 The Black-Scholes Equation

Let Π denote the value of a portfolio of one long position and a short position in some quantity delta, Δ, of the underlying,

$$\Pi = V(S,t) - \Delta S. \tag{2.3}$$

We can write down the change in the value of the porfolio from time t to $t + dt$ as

$$d\Pi = dV - \Delta dS.$$

Assuming that the underlying asset follows a log-normal random walk and by applying Itô's formula, we can write down the change in the option value as

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt.$$

Thus, the change in the portfolio is

$$d\Pi = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt - \Delta dS. \tag{2.4}$$

In the above equation, if we choose

$$\Delta = \frac{\partial V}{\partial S}, \tag{2.5}$$

then (2.4) becomes purely deterministic as by choosing delta as above, we have reduced the randomness in the equation to zero. After choosing the quantity Δ as above, the change in the value of the portfolio can be written as

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt. \tag{2.6}$$

The above change in the value of the portfolio is completely riskless. If we have a risk-free change in the portfolio value, then it must be equal to the growth we would get if we put the equivalent amount of cash in a risk-free interest account. That is,

$$d\Pi = r\Pi dt. \tag{2.7}$$

The above is an example of the no arbitrage principle. Substituting equations (2.3), (2.5) and (2.6) into (2.7) and dividing by dt we get

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0. \tag{2.8}$$

This is the Black-Scholes equation. The Black-Scholes equation is a linear parabolic partial differential equation. Many of the partial differential equations arising in finance are linear and of a similar form to (2.8). They also tend to be parabolic, meaning they are related to heat and diffusion equations making them relatively easy to solve numerically.

2.4 Change of Measure

In this section we will show how to convert the discounted asset price into a martingale. This is done by finding an equivalent martingale measure $\mathbb{Q}$ by using Girsanov's Theorem.

Suppose we have

$$S_t = e^{Y_t} S_0, \quad t \in [0, \infty),$$

where Y_t is a Wiener process with probability measure $\mathbb{P}$. Under the probability measure $\mathbb{P}$ we know

$$Y_t \sim N(\mu t, \sigma^2 t).$$

Since the asset S_t is risky when discounted by the risk-free rate, S_t cannot be a martingale. That is, under the true probability measure $\mathbb{P}$ we cannot have

$$\mathbb{E}_{\mathbb{P}}(e^{-rt} S_t | S_u, u < t) = e^{-ru} S_u. \tag{2.9}$$

The problem is therefore to find a probability measure $\mathbb{Q}$ such that (2.9) holds. If we define the new probability measure $\mathbb{Q}$ by

$$N(\rho t, \sigma^2 t),$$

then the value of ρ that makes the discounted process S_t a martingale under the measure $\mathbb{Q}$ is

$$\rho = r - \frac{1}{2}\sigma^2$$

and

$$S_t = e^{(r-\frac{1}{2}\sigma^2)t+\sigma\sqrt{t}Z}S_0,$$

where $Z \sim N(0,1)$. Note that under the measure $\mathbb{Q}$, S_t depends on σ and r but not μ. Now that we have the probability distribution of S_t in terms of Z we can apply the risk-neutral valuation formula to calculate the time-zero value of a claim.

2.5 Black-Scholes Formula

The Black-Scholes formula gives the price of a call option, $C(S_t, t)$ when the following conditions apply

1. The underlying asset follows log-normal price dynamics.
2. The risk-free interest rate is a known function of time.
3. The underlying asset pays no dividends.
4. Continuous trading.
5. There are no transaction costs.
6. There exists no arbitrage opportunities.

In the paper by Black and Scholes they assume that the market consists of one risky asset and one riskless asset given by the following set of price dynamics

$$\begin{aligned} dB_t &= rB_t dt \\ dS_t &= \mu S_t dt + \sigma S_t dW_t \end{aligned}$$

where r, μ and σ are constants satisfying the conditions given by (4.2), and W_t is standard Brownian motion. For the European call option, with payoff function $X = (S_T - K)_+$, the Black-Scholes price is given by

$$C(S,t) = S\phi(d_1) - Ke^{-r(T-t)}\phi(d_2),$$

where

$$d_1 = \frac{\log(\frac{S}{K}) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \tag{2.10}$$

$$d_2 = \frac{\log(\frac{S}{K}) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \tag{2.11}$$

and $\phi(\cdot)$ is the standard normal cumulative distribution given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

One of the nice features of the Black-Scholes pricing formula is that it only depends on one unknown parameter, the volatility σ. We will derive the Black-Scholes formula for the call option from a measure theoretic point of view. This approach will avoid having to solve the Black-Scholes partial differential equation. Now,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}(B_T^{-1}X) &= \mathbb{E}_{\mathbb{Q}}(\exp(-rT)(S-K)_+) \\
&= \int_{-\infty}^{\infty} \exp(-rT)(S-K)_+ d\mathbb{Q} \\
&= \int_{-\infty}^{\infty} e^{-rT}(S_0 e^x - K)_+ \frac{1}{\sigma\sqrt{2\pi T}} e^{-\frac{(x-(r-\frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T}} dx.
\end{aligned}$$

To eliminate the max function, note that

$$S_0 e^x \geq K$$

implies that

$$x \geq \ln \frac{K}{S_0}$$

and hence

$$\begin{aligned}
C_0 &= \int_{\ln\frac{K}{S_0}}^{\infty} e^{-rT}(S_0 e^x - K) \frac{1}{\sigma\sqrt{2\pi T}} e^{-\frac{(x-(r-\frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T}} dx \\
&= S_0 \int_{\ln\frac{K}{S_0}}^{\infty} e^{-rT} e^x \frac{1}{\sigma\sqrt{2\pi T}} e^{-\frac{(x-(r-\frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T}} dx \\
&\quad -Ke^{-rT} \int_{\ln\frac{K}{S_0}}^{\infty} \frac{1}{\sigma\sqrt{2\pi T}} e^{-\frac{(x-(r-\frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T}} dx.
\end{aligned}$$

Letting $w = (x - (r - \frac{1}{2}\sigma^2)T)/\sigma\sqrt{T}$ and noting that $dx = \sigma\sqrt{T}dw$, we get

$$C_0 = \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln \frac{K}{S_0} - (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}}^{\infty} Se^{-\frac{1}{2}w^2 + \sigma\sqrt{T}w - \frac{1}{2}\sigma^2 T} - \frac{Ke^{-rT}}{\sqrt{2\pi}} \int_{\frac{\ln \frac{K}{S_0} - (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}}^{\infty} e^{-\frac{1}{2}w^2} dw.$$

Letting $-d_2 = -\frac{\ln \frac{S_0}{K} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$, we can rewrite the above equation as

$$C_0 = S_0 \frac{e^{-\frac{1}{2}\sigma^2 T}}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{\sigma\sqrt{T}w - \frac{1}{2}w^2} dw - \frac{Ke^{-rT}}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}w^2} dw.$$

Using the properties of the normal distribution,

$$C_0 = S_0 \frac{e^{-\frac{1}{2}\sigma^2 T}}{\sqrt{2\pi}} \int_{-\infty}^{d_2} e^{-\frac{1}{2}(w^2 + 2\sigma\sqrt{T}w)} dw - \frac{Ke^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{d_2} e^{-\frac{1}{2}w^2} dw.$$

Completing the square in the first integral, by adding and subtracting $e^{\frac{1}{2}\sigma^2 T}$, gives us

$$C_0 = \frac{S_0}{\sqrt{2\pi}} \int_{-\infty}^{d_2} e^{-\frac{1}{2}(w + \sigma\sqrt{T})^2} dw - \frac{Ke^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{d_2} e^{-\frac{1}{2}w^2} dw.$$

If we define $v = w + \sigma\sqrt{T}$ in the first integral, we get

$$\begin{aligned} C_0 &= \frac{S_0}{\sqrt{2\pi}} \int_{-\infty}^{d_2 + \sigma\sqrt{T}} e^{-\frac{1}{2}v^2} dv - \frac{Ke^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{d_2} e^{-\frac{1}{2}w^2} dw \\ &= S_0\phi(d_1) - Ke^{-rT}\phi(d_2), \end{aligned}$$

where $d_1 = d_2 + \sigma\sqrt{T}$. □

Theorem 2.5.1 (Put-Call Parity) *Let $C(S,t)$ and $P(S,t)$ be the price of a European call and put option respectively, then*

$$P(S,t) = C(S,t) - S + Ke^{-r(T-t)}$$

where K is the strike price and S is the stock price at time t.

By applying the Put-Call Parity, Theorem 2.5.1, to the Black-Scholes formula for a European call option, we get the formula for the European put option

$$P(S,t) = Ke^{-r(T-t)}\phi(-d_2) - S\phi(-d_1).$$

2.6 Greeks

Options and derivatives are traded either through an organised market (*exchange*) or in the *over the counter* market. The later requires direct contact between participants and the contract terms may be anything subject to negotiation.

When a financial institution sells an option or other derivative to a client it is faced with the problem of managing the associated risk. If the contract happens to be the same as one that is sold on the exchange, the financial institution can neutralise its exposure by buying the same option it holds on the exchange. Since it is often the case that the contract has been tailored to the needs of the client and does not correspond to a standardised product traded on the exchange, hedging the exposure becomes more difficult.

One strategy open to the writer of an option is to simply do nothing. Upon maturity the contract is either exercised or ignored. From the writer's point of view it is much better if the option finishes out of the money since their maximum profit will be realised. However, they risk a potentially large loss if the option finishes deep in the money.

An option without any offsetting position in an underlying asset is referred to as a *naked position*. Writing a call option while also owning the underlying stock is called a *covered position*.

The covered position is the opposite of a naked option. If the contract finishes in the money and the contract is exercised then this strategy performs well. If the contract finishes out of the money then the writer may be exposed to large losses. Neither of the above two strategies provides a satisfactory hedge.

Another basic scheme is known as the *stop loss strategy*. The idea behind the stop loss strategy is to hold a naked position whenever it is favourable, and a covered position otherwise. This strategy is achieved by buying or selling the stock each time the price passes the strike. Applying this strategy to hedge a call option would require the writer to be holding stock whenever the call is in the money and sell whenever the call is out of the money.

2.6.1 Delta hedging

The hedging strategies mentioned so far lack sophistification and traders usually use more sophisiticated schemes. This involves calculating hedging measures referred to as the 'Greeks'.

The *delta* of an option, Δ, is the rate of change of an option with respect to the price of the underlying asset. Mathematically, the delta of a call option

is the partial derivative of the call with respect to the underlying asset price

$$\Delta := \frac{\partial C}{\partial S}. \tag{2.12}$$

For a European call option on a non-dividend paying stock, the delta can be shown to be

$$\Delta_C = \phi(d_1), \tag{2.13}$$

where d_1 is defined in (2.10). Similarly the delta for a European put can be shown to be

$$\Delta_P = \phi(d_1) - 1. \tag{2.14}$$

Using delta hedging for a short position in a European call option involves keeping a long position of $\phi(d_1)$ shares at any given time. Similarly, a long position involves maintaining a short position of $\phi(d_1)$ shares.

For a European put option, a long position in a put option should be hedged with a long position in the underlying asset and a short position in a put option should be hedged with a short position in the underlying.

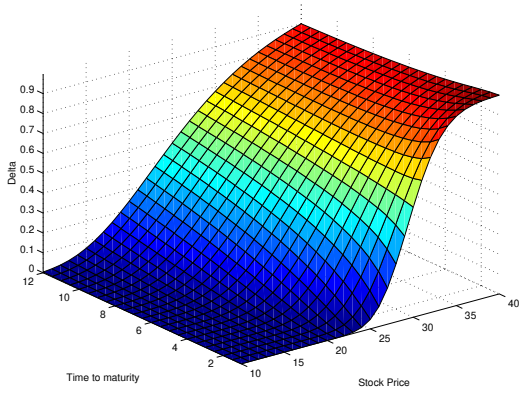


Figure 2.2: The delta of a European call with $K = 30$, $r = 0.1$ and $\sigma = 0.35$.

2.6.2 Gamma

The *gamma*, Γ, of a portfolio of derivatives on an underlying asset is the rate of change of the portfolio's delta with respect to the price of the underlying asset. It is the second partial derivative with respect to the underlying asset

$$\Gamma := \frac{\partial^2 C}{\partial S^2} = \frac{\partial^2 P}{\partial S^2} = \frac{e^{-d_1^2/2}}{S\sigma\sqrt{2\pi(T-t)}}.$$

If gamma is small, delta changes slowly and adjustments to keep the portfolio delta-neutral need only to be made infrequently. If gamma is large, the delta of an option is highly sensitive to the price of the underlying asset and the delta-neutral position should not remain unchanged for any length of time.

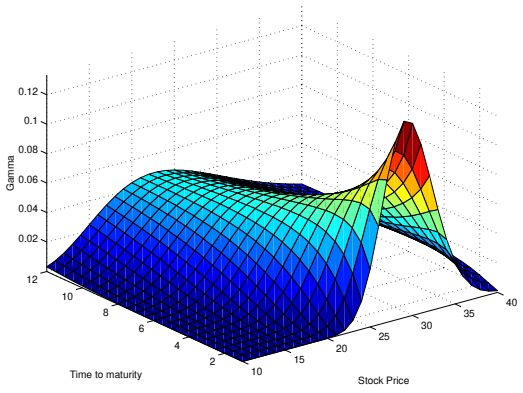


Figure 2.3: The gamma of a European call with $K = 30$, $r = 0.1$ and $\sigma = 0.35$.

2.6.3 Theta

The *theta* of an option, Θ, is the rate of change of the option value with respect to the time to maturity. It is also referred to as the *time decay* of the portfolio.

For a European call, it can be shown using Black-Scholes formula that

$$\Theta_C := \frac{\partial C}{\partial (T-t)} = -\frac{S\sigma e^{-d_1^2/2}}{2\sqrt{2\pi(T-t)}} - rKe^{-r(T-t)}\phi(d_2).$$

Similarly for a European put option

$$\Theta_P := \frac{\partial P}{\partial (T-t)} = -\frac{S\sigma e^{-d_1^2/2}}{2\sqrt{2\pi(T-t)}} + rKe^{-r(T-t)}\phi(-d_2).$$

For an option, theta is usually negative since as time to maturity decreases the option tends to be less valuable. Unlike delta hedging where it makes sense to hedge against changes in the underlying assets price it does not make sense to hedge against time to maturity, as time to maturity is not uncertain.

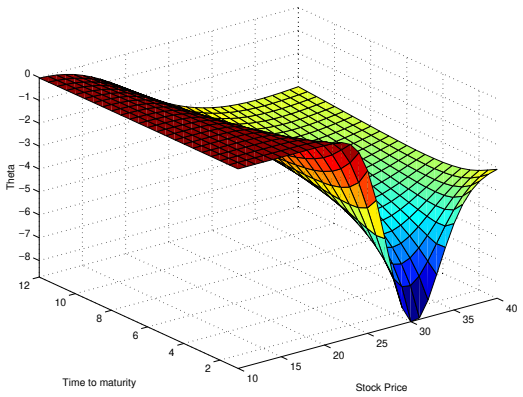


Figure 2.4: The theta of a European call with $K = 30$, $r = 0.1$ and $\sigma = 0.35$.

2.6.4 Vega

The *vega* of a portfolio of derivatives, v, is the rate of change of the value of the portfolio with respect to the volatility of the underlying asset[1]

$$v := \frac{\partial \Pi}{\partial \sigma} \tag{2.15}$$

If vega is high in absolute terms, the portfolio's value is very sensitive to small changes in volatility. If vega is low, volatility changes have little effect of the value of the portfolio.

For a European call or put option on a non-dividend paying stock, it can be shown using Black-Scholes formula that vega is given by

$$v := \frac{\partial C}{\partial \sigma} = \frac{S\sqrt{T-t}e^{-d_1^2/2}}{\sqrt{2\pi}}$$

2.6.5 Rho

The *rho* of a portfolio of derivatives is the rate of change of the portfolio value with respect to the interest rate

$$\rho = \frac{\partial \Pi}{\partial r}.$$

[1] Vega is considered to be one of the Greeks even though it is not one of the letters of the Greek alphabet

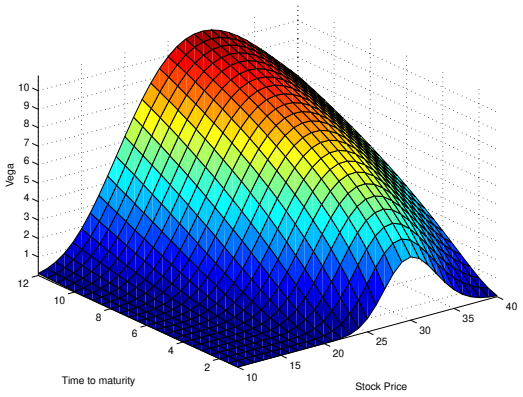


Figure 2.5: The vega of a European call with $K = 30$, $r = 0.1$ and $\sigma = 0.35$.

It measures the sensitivity of the portfolio value to interest rates. For a European call option

$$\rho_C := \frac{\partial C}{\partial r} = K(T-t)e^{-r(T-t)}\phi(d_2)$$

and for a European put option

$$\rho_P := \frac{\partial P}{\partial r} = -K(T-t)e^{-r(T-t)}\phi(d_2).$$

2.6.6 Implied Volatility

In calculating the price of a call option using the Black-Scholes formula, the only unknown parameter is the standard deviation σ of the underlying asset. A common problem in practice is to find the implied volatility given the observed market price. That is, we wish to solve the following equation for σ

$$c_0 = c(S, X, r, \sigma, T-t),$$

where c_0 is the observed market price and the function $c(\dot{})$ is the price under the Black-Scholes economy. However, there is no closed form solution for the implied volatility as a function of the option value, so we must solve the equation numerically to find σ. To calculate the value of σ we use the Newton-Rhapson method to find the root of an equation in a single variable. In our case the function we wish to find the root of is

$$f(\sigma) = c_0 - c_{BS}(\sigma)$$

and each new iteration will calculate a new value of σ via

$$\sigma_{i+1} = \sigma_i + \frac{c_0 - c_{BS}(\sigma_i)}{-\frac{\partial c_{BS}}{\partial \sigma}}$$

until

$$|f(\sigma_i)| < \epsilon,$$

where ϵ is the desired accuracy.

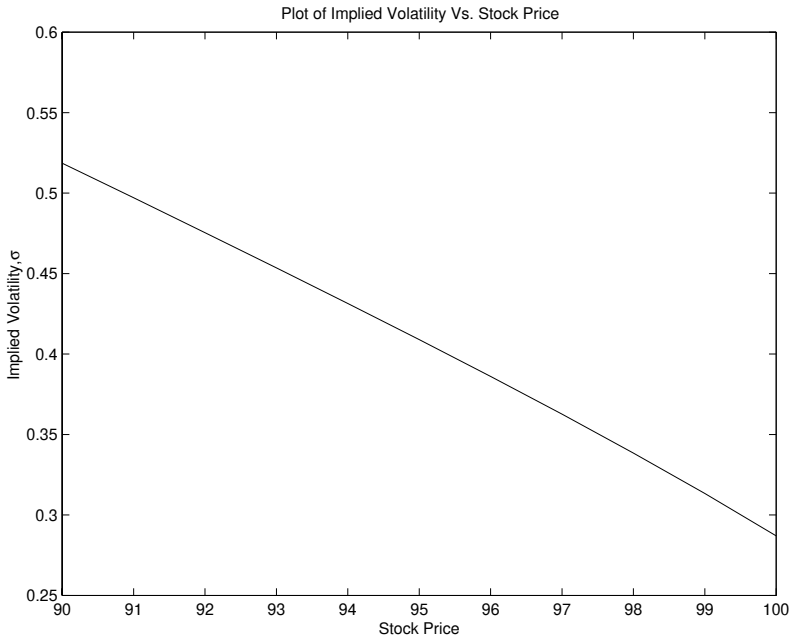


Figure 2.6: Plot of Implied Volatility versus Stock Price with $K = 95$, $r = 0.05$, $T - t = 0.5$ and $c_0 = 12$.

Chapter 3

Interest Rate Models

The correct modelling of the stochastic behaviour of interest rate models is important for the construction of realistic and reliable models for interest rate derivatives.

In this chapter we will devote our attention to looking at interest rate models. Interest rate models are used mainly to price and hedge bonds and bond options. In Chapter 2 when dealing with the Black-Scholes equation, we assumed that the interest rate was constant. Such an assumption may be acceptable when the life of the option is only six to nine months. However, the interest rate generally depends on both the date t of the loan emmision and on the maturity date of the loan T.

The valuation of interest rate derivatives depends on the level of the interest rates and the construction of valuation models for these securities depends on the stochastic movement of the rates.

Many approaches have been put forward for pricing interest rate derivatives, however, no definite consensus has been reached with regard to the best approach to the problem of pricing interest rate derivatives.

The main essence is to model the prices of the interest rate derivatives as a function of one or a few state variables such as the spot interest rate, long term interest rate or the spot forward rate.

3.1 Introduction

A *bond* is a long-term contract under which the issuer (or borrower) promises to pay the bond holder periodic coupon interest payments plus principal on specific dates as specified in the bond agreement. A *zero-coupon bond* is a bond in which there are no coupon payments. The face value of the bond is often called the *par value* and the *maturity date* is the specified date on

which the par value of the bond must be repaid. The *premium* is the amount that should be paid by the bond holder at the beginning of the contract and is simply the present value of all future cash flows.

It is unrealistic to assume that the interest rate will remain constant throughout the life of the bond as it is usually 10 years or longer. Firstly we will assume that the interest rate is a deterministic function of time and derive the corresponding bond pricing formula. In the next section we will discuss various stochastic models for interest rates and associated bond pricing models.

3.1.1 Deterministic Interest Rates

Let $r(t)$ be the deterministic interest rate function, $t \in [0, T]$, where t is the time and T is the maturity date of the bond. Since the interest rate is a function of time and not an independent state variable, we can assume that the bond price is a function of time only. Let $P(t)$ and $c(t)$ denote the bond price and coupon rate respectively. The final condition is $P(T) = FV$, where FV is the par value. In time increment Δt, the change in the value of the bond and the coupon received must equal the riskless interest return by the no arbitrage principle. Hence,

$$\frac{dP}{dt} + c(t) = r(t)P, \quad t < T.$$

Together with the final condition, the solution to the first order linear ordinary differential equation is

$$P(t) = e^{-\int_t^T r(s)ds} \left[FV + \int_t^T c(\tau) e^{\int_t^T r(s)ds} d\tau \right].$$

A bond is called a *discount bond* if the price falls below the par value and a *premium bond* otherwise. Also the *pull-to-par phenomenon* says that the market value of a bond will always approach its par value as we get closer to maturity.

3.1.2 Term Structure of Interest Rates

Let $P(t, T)$ represent the price at time t of a zero-coupon bond maturing at T with par value $P(T, T) = 1$. The market values of $P(t, T)$ indicate the market expectation of the interest rate at future dates. We can define the *yield to maturity* $R(t, T)$ as

$$R(t, T) = \frac{1}{T - t} \ln P(t, T) \tag{3.1}$$

which gives the internal rate of return at time t of the bond. The *yield curve* is the plot of $R(t,T)$ against T. The dependence of the yield curve on the time to maturity $T-t$ is called the *term structure of interest rates*.

A *forward contract* is a contract whereby the holder agrees to purchase a zero-coupon bond at some future time T_1 with maturity date $T_2 (> T_1)$. The *forward price* is simply $P(t,T_2)/P(t,T_1)$ with the *forward rate* at time t, $T_1 < t < T_2$, related to the zero-coupon bond prices by

$$f(t,T_1,T_2) = -\frac{1}{T_2 - T_1} \ln \frac{P(t,T_2)}{P(t,T_1)}. \tag{3.2}$$

The *instantaneous forward rate* at time t for a bond with maturity date T is

$$F(t,T) = \lim_{\Delta T \to 0} \frac{\ln P(t,T) - \ln P(t,T+\Delta T)}{\Delta T} = -\frac{1}{P(t,T)} \frac{\partial P}{\partial T}(t,T). \tag{3.3}$$

$F(t,T)$ can be thought of as the marginal rate of return for committing a bond investment for an additional instant. Integrating (3.3) with respect to T, we can express the bond price in terms of the instantaneous forward rate as

$$P(t,T) = \exp\left(- \int_t^T F(t,u) du \right).$$

Combining (3.1) and (3.3) we can express the yield to maturity in terms of the forward rate as

$$R(t,T) = \frac{1}{T-t} \int_t^T F(t,u) du.$$

The above two equations show that the bond price/yield can be determined from the term structure of the forward rate. The *short interest rate* or *spot rate* $r(t)$ is simply

$$r(t) = \lim_{T \to t} R(t,T) = R(t,t) = F(t,t).$$

Term Structure Theories

Several theories of term structures have been proposed to explain the shape of a yield curve. Three of the most important term structure theories are:

Expectation Theory Long-term interest rates reflect expected future short-term rates.

Market Segmentation Theory Each borrower and lender has a preferred maturity so that the slope of the yield curve will depend on the supply and demand conditions for funds in the long-term market relative to the short-term market.

Liquidity Preference Theory Lenders prefer to make short-term loans rather than long-term loans since liquidity of capital is in general preferred.

3.2 One-factor Models

In this section we will derive the governing equation for the bond price using the arbitrage pricing approach. Let us assume that the spot rate $r(t)$ can be described by the following general stochastic differential equation

$$dr = u(r,t)dt + v(r,t)dW_t$$

where W_t is standard Brownian motion and $u(r,t)$ and $v(r,t)$ are the instantaneous drift and standard deviation of $r(t)$ respectively. Applying Itô's Lemma to $P(t)$ gives

$$dP = \left(\frac{\partial P}{\partial t} + u\frac{\partial P}{\partial r} + \frac{1}{2}v^2\frac{\partial^2 P}{\partial r^2}\right)dt + v\frac{\partial P}{\partial r}dW_t.$$

If we write the dynamics of the bond price as

$$\frac{dP}{P} = \mu(r,t)dt + \sigma(r,t)dW_t$$

then we can obtain the drift and variance of $P(r,t)$ to be

$$\mu(r,t) = \frac{1}{P}\left(\frac{\partial P}{\partial t} + u\frac{\partial P}{\partial r} + \frac{1}{2}v^2\frac{\partial^2 P}{\partial r^2}\right) \tag{3.4}$$

$$\sigma(r,t) = \frac{1}{P}v\frac{\partial P}{\partial r}, \tag{3.5}$$

respectively. Unlike the role of underlying assets in an equity option, interest rates cannot be used to hedge with bond as they themselves are not tradeable securities. Instead we try to hedge bonds of different maturities.

Suppose we construct a portfolio consisting of buying a bond worth V_1 with maturity T_1 and selling a bond worth V_2 at maturity T_2. The value of the portfolio is given by

$$\Pi = V_1 - V_2.$$

The change in the value of portfolio in time dt is

$$d\Pi = (V_1\mu(r,t,T_1) - V_2\mu(r,t,T_2))dt + (V_1\sigma(r,t,T_1) - V_2\sigma(r,t,T_2))dW_t. \tag{3.6}$$

If we choose the V_1 (and similarly V_2) such that

$$V_1 = \frac{\sigma(r,t,T_2)}{\sigma(r,t,T_2) - \sigma(r,t,T_1)}\Pi$$

then the stochastic part of (3.6) disappears and becomes

$$d\Pi = \Pi\frac{\mu(r,t,T_1)\sigma(r,t,T_2) - \mu(r,t,T_2)\sigma(r,t,T_1)}{\sigma(r,t,T_2) - \sigma(r,t,T_1)}dt.$$

Since the portfolio is instantaneously riskless, it must earn the riskless spot rate, that is, $d\Pi = r(t)\Pi dt$. Combining with the above result gives us

$$\frac{\mu(r,t,T_1) - r(t)}{\sigma(r,t,T_1)} = \frac{\mu(r,t,T_2) - r(t)}{\sigma(r,t,T_2)}.$$

The above equation is independent of the maturity dates T_1 and T_2 and thus the ratio $\frac{\mu(r,t)-r(t)}{\sigma(r,t)}$ is independent of the maturity date T. We will define the ratio by $\lambda(r,t)$, that is

$$\lambda(r,t) = \frac{\mu(r,t) - r(t)}{\sigma(r,t)}.$$

This quantity is referred to as the *market price of risk* as it gives the increase in expected instantaneous rate of return on a bond per an additional unit of risk.

When we substitute (3.4) and (3.5) into the equation for the market price of risk we obtain the partial differential equation for the price of a zero-coupon bond

$$\frac{\partial P}{\partial t} + \frac{1}{2}v^2\frac{\partial^2 P}{\partial r^2} + (u - \lambda v)\frac{\partial P}{\partial r} - rP = 0, \quad t < T, \tag{3.7}$$

with boundary condition $P(T,T) = 1$. Once we choose a suitable model for $r(t)$ and the market price of risk is specified, we can solve (3.7) to obtain the value of a zero-coupon bond.

3.2.1 Vasicek Model

Vasicek (1977) suggested that the spot interest rate be governed by the Ornstein-Uhlenbeck process

$$dr = a(b - r)dt + \sigma dW_t \tag{3.8}$$

where a,b and σ are non-negative constants. This process is often referred to as a mean reverting process as the instantaneous drift $a(b - r)$ pulls the process back to the long term mean b at a rate proportional to the distance from the mean. The mean and variance of the spot rate at time T conditional on the current spot rate are given by

$$\begin{aligned} \mathbb{E}[r(T)|r(t)] &= b + (r(t) - b)e^{-a(T-t)} \\ var[r(T)|r(t)] &= \frac{\sigma^2}{2a}\left[1 - e^{-2a(T-t)}\right]. \end{aligned}$$

If we set $r_t = U_t V_t$, we can solve for U_t and V_t to give the solution to the Vasicek interest rate model as,

$$r_t = r_0 e^{-at} + b(1 - e^{-at}) + e^{-at}\int_0^t \sigma e^{-at} dW_s.$$

For simplicity we assume the market price of risk λ to be constant, independent of r and t. For the Vasicek model given in (3.8) the pde for the price of a bond becomes

$$\frac{\partial P}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial r^2} + (a(b - r) - \lambda\sigma)\frac{\partial P}{\partial r} - rP = 0. \tag{3.9}$$

Let us assume that the solution be of the form

$$P(r, t : T) = A(t, T)e^{B(t,T)r}. \tag{3.10}$$

Substituting (3.10) into (3.9) gives us the following set of odes to be solved

$$\begin{aligned} \frac{dA}{dt} + (\lambda\sigma - ab)AB + \frac{1}{2}\sigma^2 AB^2 &= 0, \quad t < T \\ \frac{dB}{dt} - aB + 1 &= 0, \quad t < T \end{aligned}$$

with boundary conditions $A(T, T) = 1$ and $B(T, T) = 0$. Upon solving the above set of equations and substituting the solution into (3.10), the solution is with $t < T$

$$P(r, t : T) = \exp(\frac{1}{a}[1 - e^{-a(T-t)}](R_\infty - r) - R_\infty(T - t) - \frac{\sigma^2}{4a^3}[1 - e^{-a(T-t)}]^2),$$

where $R_\infty = b - \frac{\sigma\lambda}{a} - \frac{\sigma^2}{2a^2}$. The yield to maturity or the term structure of interest rates can easily be found to be (Vasicek 1977)

$$R(t,T) = R_\infty + \frac{[r(t) - R_\infty][1 - e^{-a(T-t)}]}{a(T-t)} + \frac{\sigma^2}{4a^3(T-t)}[1 - e^{-a(T-t)}]^2$$

where $R_\infty = \lim_{T\to\infty} R(t,T)$.

3.2.2 The Cox-Ingersoll-Ross model

Cox, Ingersoll & Ross (1985) suggested modelling the instantaneous interest rate by the following stochastic differential equation,

$$dr = a(b - r)dt + \sigma\sqrt{r}dW_t.$$

Unlike the model proposed by Vasicek, the interest rates may never go negative given an initial non-negative interest rate. This is also a mean reverting process however, the volatility is no longer constant and is dependent upon the interest rate. It is easy to see that if $\sigma^2 > 2ab$ then $r(t)$ may reach zero while if $2ab \geq \sigma^2$ the interest rate may never hit zero. The mean and variance of $r(T)$ are

$$\begin{aligned} \mathbb{E}[r(T)|r(t)] &= r(t)e^{-a(T-t)} + b[1 - e^{-a(T-t)}], \\ var[r(T)|r(t)] &= r(t)\frac{\sigma^2}{a}[e^{-a(T-t)} - e^{-2a(t-t)}] + \frac{b\sigma^2}{2a}[1 - e^{-a(T-t)}]^2. \end{aligned}$$

Some interesting properties to note are

- as $a \to \infty$, the mean tends to b and the variance tends to zero,
- as $a \to 0^+$, the mean tends to $r(t)$ and the variance tends to $\sigma^2(T-t)r(t)$.

To solve for the price of a zero-coupon bond we assume the same form for the solution as in (3.10). For the model proposed by Cox et al. (1985) the system of odes to be solved now becomes

$$\begin{aligned} \frac{dA}{dt} - abAB &= 0, \quad t < T \\ \frac{dB}{dt} - (a + \lambda\sigma)B - \frac{1}{2}\sigma^2 B^2 + 1 &= 0, \quad t < T, \end{aligned}$$

with boundary conditions $A(T,T) = 1$ and $B(T,T) = 0$. The market price of risk is taken to be $\lambda\sqrt{r(t)}$ where λ is a constant.

The solution to the above set of equations is (Cox et al. 1985)

$$P(r,t:T) = A(t,T)e^{-B(t,T)r}$$

where

$$\begin{aligned} A(t,T) &= \left[\frac{2\gamma e^{[(\kappa+\lambda+\gamma)(T-t)]/2}}{(\gamma+\kappa+\lambda)(e^{\gamma(T-t)}-1)+2\gamma}\right]^{\frac{2\kappa\theta}{\sigma^2}} \\ B(t,T) &= \frac{2(e^{\gamma(T-t)}-1)}{(\gamma+\kappa+\lambda)(e^{\gamma(T-t)}-1)+2\gamma} \end{aligned}$$

and γ is given by $\gamma = ((\kappa+\lambda)^2+2\sigma^2)^{\frac{1}{2}}$. Note that the market price of risk λ only appears in γ in the above solution.

3.3 Two-factor Models

In general the one-factor interest rate models discussed in the previous section cannot capture the rich structure found in practice. Most of these models offer the analytic tractability where closed form solutions of the term structures and bond prices can be found. However, this approach tends to oversimplify the true behaviour of interest rate movements.

In this section we will focus our attention on multi-factor models that involve the short rate together with some other state variable. Generally the second state variable in a two-factor model is either the long rate (Brennan & Schwartz 1982) or the variance of the short rate (Fong & Vasicek). Due to the higher degrees of freedom used to model the term structure of interest rates we usually need to resort to some numerical scheme to value the bond prices.

3.3.1 Brennan and Schwartz

Brennan & Schwartz (1982) chose the two stochastic factors to be the short-term interest rate r and the long-term interest rate l. The general form of the model is

$$\begin{aligned} dr &= a_r(r,l,t)dt + b_r(r,l,t)dW_t^1 \\ dl &= a_l(r,l,t)dt + b_l(r,l,t)dW_t^2 \end{aligned}$$

and let ρ denote the correlation between dW_t^1 and dW_t^2. By Itô's Lemma, the stochastic differential equation for the price of a zero-coupon bond is given

by

$$\frac{dP}{P} = \mu(r,l,t)dt + \sigma_r(r,l,t)dW_t^1 + \sigma_l(r,l,t)dW_t^2$$

where

$$\begin{aligned}
\mu(r,l,t) &= \frac{1}{P}\left[\frac{\partial P}{\partial t} + a_r\frac{\partial P}{\partial r} + a_l\frac{\partial P}{\partial l} + \frac{1}{2}b_r^2\frac{\partial^2 P}{\partial r^2} + \frac{1}{2}b_l^2\frac{\partial^2 P}{\partial l^2} + \rho b_r b_l\frac{\partial^2 P}{\partial r \partial l}\right] \\
\sigma_r(r,l,t) &= \frac{1}{P}b_r\frac{\partial P}{\partial r} \\
\sigma_l(r,l,t) &= \frac{1}{P}b_l\frac{\partial P}{\partial l}.
\end{aligned}$$

Let us construct a portfolio containing V_1, V_2 and V_3 units of bonds with maturity dates T_1, T_2 and T_3 respectively. Let Π denote the value of the portfolio, then the rate of return on the portfolio in time dt is

$$\begin{aligned}
d\Pi &= [V_1\mu(T_1) + V_2\mu(T_2) + V_3\mu(T_3)]dt \\
&\quad +[V_1\sigma_r(T_1) + V_2\sigma_r(T_2) + V_3\sigma_r(T_3)]dW_t^1 \\
&\quad +[V_1\sigma_l(T_1) + V_2\sigma_l(T_2) + V_3\sigma_l(T_3)]dW_t^2.
\end{aligned}$$

As last time, we shall select the values of V_1, V_2 and V_3 such that the stochastic terms in the above equation vanish. This leads to the following set of equations for V_1, V_2 and V_3

$$\begin{aligned}
V_1\sigma_r(T_1) + V_2\sigma_r(T_2) + V_3\sigma_r(T_3) &= 0 \\
V_1\sigma_l(T_1) + V_2\sigma_l(T_2) + V_3\sigma_l(T_3) &= 0.
\end{aligned}$$

Since the portfolio is instantaneously riskless, it must earn the spot interest rate to avoid arbitrage, that is,

$$d\Pi = [V_1\mu(T_1) + V_2\mu(T_2) + V_3\mu(T_3)]dt = r(V_1 + V_2 + V_3)dt$$

so that

$$V_1[\mu(T_1) - r] + V_2[\mu(T_2) - r] + V_3[\mu(T_3) - r] = 0.$$

Thus we have obtained a system of equations for V_1, V_2 and V_3

$$\begin{pmatrix} \sigma_r(T_1) & \sigma_r(T_2) & \sigma_r(T_3) \\ \sigma_l(T_1) & \sigma_l(T_2) & \sigma_l(T_3) \\ \mu(T_1) - r & \mu(T_2) - r & \mu(T_3) - r \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The solution to the above system will be non-trivial provided that

$$\begin{vmatrix} \sigma_r(T_1) & \sigma_r(T_2) & \sigma_r(T_3) \\ \sigma_l(T_1) & \sigma_l(T_2) & \sigma_l(T_3) \\ \mu(T_1)-r & \mu(T_2)-r & \mu(T_3)-r \end{vmatrix} = 0.$$

The above condition is valid for arbitrary values of T_1, T_2 and T_3 provided that

$$\mu(t) - r = \lambda_r(r,l,t)\sigma_r(t) + \lambda_l(r,l,t)\sigma_l(t), \tag{3.11}$$

where λ_r and λ_l are the market prices of risk for the short-term and long-term interest rates respectively. Substituting the expressions for $\mu(t)$, $\sigma_r(t)$ and $\sigma_l(t)$ into (3.11) gives us the following partial differential equation

$$\begin{aligned} 0 &= \frac{\partial P}{\partial t} + \frac{1}{2}b_r^2\frac{\partial^2 P}{\partial r^2} + \rho b_r b_l \frac{\partial^2 P}{\partial r \partial l} + \frac{1}{2}b_l^2\frac{\partial^2 P}{\partial l^2} \\ &\quad + (a_r - \lambda_r b_r)\frac{\partial P}{\partial r} + (a_l - \lambda_l b_l)\frac{\partial P}{\partial l} - rP. \end{aligned}$$

The specific form of the equations for the short-term and long-term rates postulated by Brennan & Schwartz (1982) has the form

$$\begin{aligned} dr &= (a_1 + b_1(l-r))dt + \sigma_1 r dW_t^1, \\ dl &= l(a_2 - b_2 r + c_2 l)dt + \sigma_2 l dW_t^2. \end{aligned}$$

Both the long and short rates display locally log-normal behaviour. In addition, the drift term of the short rate model is mean-reverting, reverting to the long rate at a speed b_1. This is consistent with theories of the term structure based on expectations (Brennan & Schwartz 1982). That is, theories which assume long rates to be based upon expectation of future short rates.

Substituting the coefficients in the specific model into the partial differential equation makes us need to use a numerical finite differences method to solve it. In the numerical solution we need the full set of boundary conditions for the bond price equation. It is expected that the bond price would tend to zero and either one of the interest rates goes to infinity. However, the boundary condition at the limiting case of zero can be quite difficult to implement.

3.3.2 Fong and Vasicek

Many other multi-factor models have been proposed in the literature. Fong and Vasicek, (Fong & Vasicek 1991), chose the variance of the short term

rate to be the second state variable to more accurately describe the term structure of interest rates. They proposed that the stochastic differential equations for the short rate and the variance of the short rate should be

$$\begin{aligned} dr &= a(\bar{r} - r)dt + \sqrt{\sigma}dW_t^1 \\ d\sigma &= b(\bar{\sigma} - \sigma)dt + \xi\sqrt{\sigma}dW_t^2 \end{aligned}$$

where $\bar{r}$ and $\bar{\sigma}$ are the long term means of r and σ respectively. The market prices of risk are assumed to be proportional to $\sqrt{\sigma}$. Using the free arbitrage argument the partial differential equation for the price of a discount bond is

$$\begin{aligned} \frac{\partial P}{\partial r} &= \frac{1}{2}\sigma\frac{\partial^2 P}{\partial r^2} + \rho\xi\sigma\frac{\partial^2 P}{\partial r \partial \sigma} + \frac{1}{2}\xi^2\sigma\frac{\partial^2 P}{\partial \sigma^2} \\ &\quad + (a\bar{r} - ar + \lambda\sigma)\frac{\partial P}{\partial r} + (b\bar{\sigma} - (b + \xi\eta)\sigma)\frac{\partial P}{\partial \sigma} - rP \end{aligned}$$

where $\lambda\sqrt{\sigma}$ and $\eta\sqrt{\sigma}$ are assumed to be the market prices of risk for r and σ respectively and ρ is the correlation between W_t^1 and W_t^2. Although very difficult, the closed form solution can be found[1].

3.3.3 Longstaff and Schwartz

Longstaff & Schwartz consider the following model for the risk-adjusted variables,

$$\begin{aligned} dx &= a(\bar{x} - x)dt + \sqrt{x}dW_t^1, \\ dy &= b(\bar{y} - y)dt + \sqrt{y}dW_t^2, \end{aligned}$$

where the spot interest rate is given by

$$r = cx + dy.$$

The simple form of the terms in the above equations results in the explicit solution for simple interest rate products.

3.4 No-Arbitrage Models

In this section we will take market data, such as the current term structure of interest rates, and develop the no arbitrage yield curve model so that is is consistent with the current market data. These type of models are called *no*

[1] The closed form solution is in terms of confluent hypergeometric functions with complex arguments (Selby & Strickland 1995).

arbitrage models. These models allow for time dependent functions within the models, with the functions being determined in such a way that the bond prices match the observed market prices. No arbitrage models can be developed using either the bond prices, forward rates or the short rates. In the rest of this section we will explore some of the popular no arbitrage models.

3.4.1 Ho & Lee Model

This is the first no arbitrage model proposed (Ho & Lee 1986), where the initial model was developed in the form of a binomial tree. The continuous time limit of this process is given by

$$dr = a(t)dt + \sqrt{\sigma}dX$$

where r is the short rate and σ is the instantaneous standard deviation of the short rate. The function $a(t)$ is chosen such that the model fits the initial term structure. The value of a zero-coupon bond is given by

$$P(r, t : T) = e^{A(t,T) - rB(t,T)}$$

where

$$\begin{aligned} B(t,T) &= T - t \\ A(t,T) &= -\int_t^T a(\tau)(T-\tau)d\tau + \frac{1}{6}b(T-t)^3. \end{aligned}$$

3.4.2 Hull & White Model

In their paper, Hull & White (1990) extend the one-factor models of Vasicek (1977) and Cox et al. (1985) to models which are consistent with the current term structure of interest rates and either the current volatilities of all the spot interest rates or the volatilities of all the forward interest rates.

The extended model of Vasicek takes the form

$$dr = [\theta(t) + a(t)(b - r)]dt + \sigma(t)dW_t.$$

This implies that any contingent claim, $f(r)$, must satisfy

$$f_t + [\phi(t) - a(t)r]f_r + \frac{1}{2}\sigma(t)^2 f_{rr} - rf = 0, \tag{3.12}$$

where

$$\phi(t) = a(t)b + \theta(t) - \lambda(t)\sigma(t).$$

The solution to (3.12) with boundary condition $f(t) = 1$ at maturity T is the price of a zero-coupon bond of the form

$$f(t,T) = A(t,T)e^{-B(t,T)r}. \tag{3.13}$$

This function satisfies (3.12) and the boundary condition when

$$\begin{aligned} A_t - \phi(t)AB + \frac{1}{2}\sigma(t)^2 AB^2 &= 0 & (3.14)\\ B_t - a(t)B + 1 &= 0 & (3.15) \end{aligned}$$

with

$$A(T,T) = 1, \quad B(T,T) = 0. \tag{3.16}$$

Solving (3.14) and (3.15) subject to the boundary conditions given by (3.16), Equation (3.13) gives the price of a zero-coupon bond maturing at time T. If we assume that $a(t)$, $\phi(t)$ and $\sigma(t)$ are constants then the solution is that given by Vasicek,

$$B(t,T) = \frac{B(0,T) - B(0,t)}{\partial B(0,t)/\partial t},$$

$$\begin{aligned} \log A(t,T) &= \log A(0,T) - \log A(0,t) - B(t,T)\frac{\partial \log A(0,t)}{\partial t} \\ &\quad -\frac{1}{2}\left[B(t,T)\frac{\partial B(0,t)}{\partial t}\right]^2 \int_0^t \left[\frac{\sigma(\tau)}{\partial B(0,\tau)/\partial \tau}\right]^2 d\tau \end{aligned}$$

we obtain

$$a(t) = -\frac{\partial^2 B(0,t)/\partial t^2}{\partial B(0,t)/\partial t}.$$

3.4.3 Black-Derman-Toy Model

The Black-Derman-Toy model, (Black, Derman & Toy 1990), as in the Ho-Lee model, is formulated as a binomial tree. The continuous time limit of the model is

$$d\ln r = \left[\theta(t) - \frac{\sigma'(t)}{\sigma(t)}\ln r\right] dt + \sigma(t)dW_t.$$

This process models the changes in the short rate as log-normally distributed random variables with the interest rates always being non-negative. Similar to the Ho-Lee model, the function $\theta(t)$ is chosen so that the model fits the term structure of short rates and $\sigma(t)$ is chosen to fit the term structure of short rate volatilities. If $\sigma(t)$ is constant, this model reduces to a log-normal version of the Ho-Lee model.

3.5 Heath-Jarrow-Morton

The HJM model, (Heath, Jarrow & Morton 1992), attempts to construct a family of continuous time stochastic processes for the term structure, consistent with the observed initial term structure data. The driving state variable of the model is chosen to be $F(t,T)$, the forward rate at time t for instantaneous borrowing at a later time T. The stochastic process is assumed to be

$$dF(t,T) = a_F(t,T)dt + \sum_{i=1}^{n} \sigma_F^i(t,T)dW_t^i, \quad 0 \le t \le T,$$

where $a_F(t,T)$ is the instantaneous forward rate's drift, σ_F is the vector of volatilities of the forward rates and W_t is an n-dimensional vector of independent Wiener processes.

The Heath, Jarrow and Morton approach is to specify the volatilities of all instantaneous forward rates for all future times, often referred to as the volatility structure. By defining the risk-neutral process $P(t,T)$ as

$$dP(t,T) = r(t)P(t,T)dt + v(t,T)P(t,T)dW_t. \tag{3.17}$$

The expected return of the risk-neutral process is $r(t)$. The volatility, $v(t,T)$, can be any well-behaved function, but since a bond's price volatility declines to zero at maturity,

$$v(t,t) = 0,$$

which is equivalent to the assumption that all discount bonds have finite drifts at all times. Applying Itô's Formula, Theorem 1.2.2, to $\ln P(t,T)$ and using (3.17)

$$d\ln P(t,T) = \left(r(t) - \frac{v(t,T)^2}{2}\right)dt + v(t,T)dW_t$$

and using the definition of the forward rate given in (3.2) we get

$$df(t,T_1,T_2) = \frac{v(t,T_2)^2 - v(t,T_1)^2}{2(T_2-T_1)}dt + \frac{v(t,T_1) - v(t,T_2}{T_2-T_1}dW_t.$$

The above equation shows us that the risk-neutral process for $f(t,T_1,T_2)$ depends only r and P to the extent that v can be a function of r and P. If we let $T_1 = T$ and $T_2 = T + \Delta T$ and take the limit as $\Delta T \to 0$ we

will obtain the stochastic differential equation for $F(t,T)$, the instantaneous forward rate.

$$dF(t,T) = v(t,T)v_T(t,T)dt - v_T(t,T)dW_t$$

where $v_T(t,T)$ denotes the partial derivative of $v(t,T)$ with respect to T. Once we know the function $v(t,T)$ for all values of t and T the process $F(t,T)$ is fully known.

Next we will derive the risk-neutral process for $r(t)$ from the bond price volatilities and the initial term structure. We know that

$$F(t,t) = F(0,t) + \int_0^t dF(\tau,t)d\tau$$

and since $r(t) = F(t,t)$ it is given by

$$r(t) = F(0,t) + \int_0^t v(\tau,t)v_t(\tau,t)dt + \int_0^t v_t(\tau,t)dW_t.$$

3.6 Affine Models

In the previous sections we have looked at a number of different models, characterised by the dynamics of their state variables. Given a stochastic differential equation for the short rate of the form

$$dr = \mu_r(r,t)dt + \sigma_r(r,t)dW_t$$

then the absence of arbitrage requires that for any asset V there should correspond a parabolic pde

$$\frac{\partial V(r,t)}{\partial t} + \mu_r\frac{\partial V(r,t)}{\partial r} + \frac{1}{2}\sigma_r^2\frac{\partial^2 V(r,t)}{\partial r^2} = r(t)V(r,t).$$

In the above description, the equations are completely general. However, if we place restrictions on the form of μ_r and σ_r, we can obtain useful results about the form of the discounted bond price. If the functions μ_r and σ_r are affine or linear functions of the form

$$\mu_r = a_1(t) + a_2(t)r$$

and

$$\sigma_r^2 = b_1(t) + b_2(t)r,$$

then the solution of the partial differential equation for the discounted bond price $P(t,T)$, with boundary condition $P(T,T)=1$, is of the form

$$P(r,t,T) = \exp[A(t,T) - B(t,T)r(t)].$$

As you can see, many of the models discussed in the previous sections belong to this category. Models such as those of Vasicek, Hull & White and even Cox, Ingersoll & Ross fall into this group. However, any log-normal model such as Black-Derman-Toy do not belong to this class.

In general, the ordinary differential equations for $A(t,T)$ and $B(t,T)$ must be solved numerically.

Chapter 4

Optimal Stopping

4.1 More on Itô Diffusions

Markov Property

Definition 4.1.1 *A (time-homogeneous) Itô diffusion is a stochastic process* $X_t(\omega) = X(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}^m$ *satisfying a stochastic differential equation of the form*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq s; \quad X_s = x \tag{4.1}$$

where W_t *is m-dimensional Brownian motion and* $b : \mathbb{R}^m \to \mathbb{R}^m$, $\sigma : \mathbb{R}^m \to \mathbb{R}^{m\times d}$ *satisfy*

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|, \quad x, y \in \mathbb{R}^m \tag{4.2}$$

where $|\sigma|^2 = \sum |\sigma_{ij}|^2$.

We can think of condition (4.2) as guaranteeing that the solution to (4.1) is *unique*. In other words, if there exists two processes, $X_1(t,\omega)$ and $X_2(t,\omega)$ satisfying (4.1), then

$$X_1(t,\omega) = X_2(t,\omega), \quad \forall t \leq T, a.s. \tag{4.3}$$

We will denote this unique solution to (4.1) by $X_t = X_t^{s,x}$, $t \geq s$.

If for the process, X_t, the future behaviour given the history up to time t, is the same as starting the process at X_t, we say that X_t satisfies the *Markov property*. More formally, this can be written as

Theorem 4.1.1 (The Markov Property for Itô diffusions) *Let* f *be a bounded Borel function from* $\mathbb{R}^m \to \mathbb{R}$. *Then, for* $t,h \geq 0$

$$\mathbb{E}^x[f(X_{t+h}) \mid \mathcal{F}_t](\omega) = \mathbb{E}^{X_t(\omega)}[f(X_h)]. \tag{4.4}$$

Generator

If we can associate a second order partial differential operator A to an Itô diffusion X_t, we refer to A as the *generator* of the process X_t.

Definition 4.1.2 *Let $\{X_t\}$ be a (time-homogeneous) Itô diffusion in $\mathbb{R}^m$. The (infinitesimal) generator A of X_t is defined by*

$$Af(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}; \qquad x \in \mathbb{R}^m. \tag{4.5}$$

The set of functions $f : \mathbb{R}^m \to \mathbb{R}$ such that the limit exists at x is denoted by $\mathcal{D}_A(x)$, while $\mathcal{D}_A$ denotes the set of functions for which the limit exists for all $x \in \mathbb{R}^m$.

The generator A of the diffusion X_t can be thought of as the infinitesimal difference as t tends to zero between the future expected value of the function of X_t and the function at the starting point, x.

We will now give the relationship between the generator A and the coefficients b and σ in the stochastic differential equation (4.1). To be able to determine the relationship between A and the coefficients of (4.1), we will need to make use of the following lemma.

Lemma 4.1.1 *Let $Y_t = Y_t^x$ be an Itô process in $\mathbb{R}^m$ of the form*

$$Y_t^x(\omega) = x + \int_0^t u(s,\omega)ds + \int_0^t v(s,\omega)dW_s. \tag{4.6}$$

Let $f \in C_0^2(\mathbb{R}^m)$ and let τ be a stopping time with respect to $\{\mathcal{F}_t\}$, and assume that $\mathbb{E}^x[\tau] < \infty$. Then

$$\mathbb{E}^x[f(Y_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau \left(\sum_i u_i(s,\omega)\frac{\partial f(Y_s)}{\partial x_i} + \frac{1}{2}\sum_{i,j}(vv^T)_{i,j}(s,\omega)\frac{\partial^2 f(Y_s)}{\partial x_i \partial x_j}\right) ds\right] \tag{4.7}$$

That is, given that Y_t is a process of the form (4.6), and the expected value of the stopping time is finite, then the expected value of $f(Y_\tau)$ is given by (4.7).

Theorem 4.1.2 *Let X_t be the Itô diffusion*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t. \tag{4.8}$$

If $f \in C_0^2(\mathbb{R}^m)$, then $f \in \mathcal{D}_A$ and

$$Af(x) = \sum_i b_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{4.9}$$

In other words, if f is twice continuously differentiable on $\mathbb{R}^m$ and has compact support, i.e. closed and bounded, then for an Itô diffusion of the form (4.8) its generator A is given by (4.9). The proof of Theorem 4.1.2 is given in (Øksendal 1998) using Lemma 4.1.1 with $\tau = t$ and the definition of the generator A.

Example 4.1.1 *Consider the Ornstein-Uhlenbeck process given by*

$$dX_t = \mu X_t dt + \sigma dW_t. \tag{4.10}$$

Then the generator A of X_t is

$$Af(x) = \mu x \frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2}, \quad f \in C_0^2(\mathbb{R}^m). \tag{4.11}$$

Example 4.1.2 *For the generator A of an Itô diffusion given by*

$$Af(x) = \frac{\partial f}{\partial x} + \frac{\partial^2 f}{\partial x^2}, \quad f \in C_0^2(\mathbb{R}^m), \tag{4.12}$$

it is easy to verify that the stochastic differential equation of X_t is given by

$$dX_t = dt + \sqrt{2} dW_t. \tag{4.13}$$

The Dynkin Formula

If we combine Theorem 4.1.2 with Lemma 4.1.1 , it is easy to check that we get:

Theorem 4.1.3 (Dynkin's formula) *Let $f \in C_0^2(\mathbb{R}^m)$. Suppose τ is a stopping time, $\mathbb{E}^x[\tau] < \infty$. Then*

$$\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}\left[\int_0^\tau Af(X_s) ds\right]. \tag{4.14}$$

Dynkin's formula provides us with a way of evaluating the expected value of a function of X_t at a stopping time τ in terms of the function at our starting position, x and the expected value of the integral of the generator of X_t.

Characteristic Operator

We will now introduce an operator which is slightly more suited for some problems, but is still closely related to the generator A.

Definition 4.1.3 *Let $\{X_t\}$ be an Itô diffusion. The characteristic operator $\mathcal{A} = \mathcal{A}_X$ of $\{X_t\}$ is defined by*

$$\mathcal{A}f(x) = \lim_{U \downarrow x} \frac{\mathbb{E}^x[f(X_{\tau_U})] - f(x)}{\mathbb{E}^x[\tau_U]} \tag{4.15}$$

where the U's are open sets U_k decreasing to the point x, in the sense that $U_{k+1} \subset U_k$ and $\bigcap_k U_k = \{x\}$, and $\tau_U = \inf\{t > 0; X_t \notin U\}$ is the first exit time from U for X_t. The set of functions f such that the limit (4.15) exists for all $x \in \mathbb{R}^m$ (and all $\{U_k\}$) is denoted by $\mathcal{D}_{\mathcal{A}}$. If $\mathbb{E}^x[\tau_U] = \infty$ for all open $U \ni x$, we define $\mathcal{A}f(x) = 0$.

It turns out that

$$\mathcal{D}_A \subseteq \mathcal{D}_{\mathcal{A}}$$

and that

$$Af = \mathcal{A}f$$

for all $f \in \mathcal{D}_A$. It also turns out that if f is a twice continuously differentiable function, then the generator of X_t and the characteristic operator of X_t coincide. So for all functions $f \in C^2$, $Af = \mathcal{A}f$.

As for the generator A of the process X, we can establish a relationship between the characteristic operator, $\mathcal{A}$ and the coefficients of the Itô process given in (4.1).

Theorem 4.1.4 *Let $f \in C^2$. Then $f \in \mathcal{D}_{\mathcal{A}}$ and*

$$\mathcal{A}f = \sum_i b_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma\sigma^T)_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{4.16}$$

We will now demonstrate how we can apply Theorem 4.1.4 to an Itô process, X_t to determine the characteristic operator $\mathcal{A}$ of X_t.

Example 4.1.3 (Brownian motion on the unit circle) *Define the process $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ by the set of stochastic differential equations*

$$\begin{aligned} dX_1 &= -\frac{1}{2} X_1 dt - X_2 dW_t \\ dX_2 &= -\frac{1}{2} X_2 dt + X_1 dW_t. \end{aligned}$$

We can rewrite the above equations as

$$dX = b(X)dt + \sigma(X)dW_t$$

where

$$b(x) = \begin{pmatrix} -\frac{1}{2}x_1 \\ -\frac{1}{2}x_2 \end{pmatrix}, \quad \sigma(x) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}.$$

Therefore,

$$\sigma\sigma^T = \begin{pmatrix} x_2^2 & -x_1x_2 \\ -x_1x_2 & x_1^2 \end{pmatrix}$$

and the characteristic operator $\mathcal{A}$ of X is

$$\mathcal{A}f(x) = \frac{1}{2}\left(x_2^2\frac{\partial^2 f}{\partial x_1^2} - 2x_1x_2\frac{\partial^2 f}{\partial x_1 \partial x_2} + x_1^2\frac{\partial^2 f}{\partial x_2^2} - x_1\frac{\partial f}{\partial x_1} - x_2\frac{\partial f}{\partial x_2}\right).$$

Cameron-Martin-Girsanov Theorem

An important result that builds upon the results given in subsection A.1.2 is Girsanov's Theorem. Girsanov's Theorem is a fundamental result in the general theory of stochastic analysis, and has many important applications, for example, in finance. Suppose we wish to model the movement of stock price by a stochastic process of the form

$$S_t = S_0 \exp(\mu t + \sigma W_t)$$

and where the time value of money is

$$B_t = e^{rt}.$$

We wish to find the probability measure $\mathbb{Q}$ such that the discounted stock process $B_t^{-1}S_t$ is a martingale. Now

$$\begin{aligned} B_t^{-1}S_t &= S_0 \exp(\mu t + \sigma W_t)e^{-rt} && (4.17) \\ &= S_0 \exp((\mu - r)t + \sigma w_t). && (4.18) \end{aligned}$$

Then, the expected value under $\mathbb{P}$ is

$$\begin{aligned} \mathbb{E}(B_t^{-1}S_t) &= S_0 \exp((\mu - r)t)\mathbb{E}(e^{\sigma W_t}) && (4.19) \\ &= S_0 \exp\left(\left(\mu - r + \frac{1}{2}\sigma^2\right)t\right). && (4.20) \end{aligned}$$

If $\mu - r + \frac{1}{2}\sigma^2 = 0$, then $\mathbb{Q}$ is the risk-neutral measure. We can rewrite (4.17) as

$$B_t^{-1} S_t = S_0 \exp\left(\sigma \tilde{W}_t - \frac{1}{2}\sigma^2 t\right),$$

where

$$\tilde{W}_t = W_t + \frac{\mu - r + \frac{1}{2}\sigma^2}{\sigma} t.$$

Thus, if $\tilde{W}$ is Brownian motion under the new measure $\mathbb{Q}$, then

$$\mathbb{E}(B_t^{-1} S_t) = S_0 \mathbb{E}_{\mathbb{Q}}\left(\sigma \tilde{W}_t - \frac{1}{2}\sigma^2 t\right) = S_0$$

and $\mathbb{Q}$ must be the martingale measure for the discounted stock price process. Girsanov's Theorem tells us the conditions required for $\mathbb{Q}$ to be a martingale measure for the process and the form of the Radon-Nikodym derivative for the change of measure.

Theorem 4.1.5 (Girsanov's Theorem) *Let W_t be Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose γ_t is a process adapted to the accompanying filtration $\mathcal{F}_t$. Define*

$$\tilde{W}_t = W_t + \int_0^t \gamma_s ds \tag{4.21}$$

and

$$\Lambda = \exp\left(-\int_0^T \gamma_u dW_u - \frac{1}{2}\int_0^T \gamma_u^2 du\right),$$

and define a new probability measure $\mathbb{Q}$ by

$$\mathbb{Q}(A) = \int_A \Lambda d\mathbb{P},$$

then, provided $\mathbb{E}_{\mathbb{P}}\left(\exp\left(\frac{1}{2}\int_0^T \gamma_t^2 dt\right)\right) < \infty$, $\tilde{W}_t$ is Brownian motion under the measure $\mathbb{Q}$ where Λ is the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$.

We may also note that the converse to Girsanov's Theorem is also true, and that changing the measure by applying Girsanov's Theorem only changes the mean. The variance as well as the volatilities, quadratic variation and covariations remain unaffected.

4.2 The Optimal Stopping Problem

In this section we formulate the optimal stopping problem and present various methods for its solution. We start this chapter by examining the fundamental concepts used in formulating the optimal stopping problem. We will conclude this section with some examples.

To begin, we illustrate the fundamental concepts involved in the optimal stopping problem by way of an example.

Example 4.2.1 (An illustrative example) *Suppose that we are playing a game of shells. There is a line of shells in front of us, and under each of them is a random amount of money. In this game we are allowed to look under one shell at a time and after looking under the shell, we have two options. We can either keep the money under the shell OR we can look under the next shell.*

So after each shell we have to decide whether to stop or not. If we take the money, we could be losing a lot more in the future, yet we could also currently be at the maximum.

When, or at what event, do we stop and take the money?

This is an example of the *optimal stopping problem*. That is, we wish to determine a stopping time, of some stochastic process, that will return the greatest *reward* (the reward in this case being cash!).

A Stopping Time

One of the most common misunderstandings in optimal stopping is the definition of a stopping time. Initially, one may think that a stopping time is a time that, once we have solved the problem, remains fixed. In reality a stopping time is in reference to an event that is fixed, and the actual stopping time is a random variable dependent on this event occurring. So once we have fixed the event (by solving the problem), the actual time that we stop will be different for each observation of the underlying process.

There is another subtlety of the definition of a stopping time which is important to understand. For τ to be a stopping time, it must be possible to decide whether or not to stop on the basis of the knowledge of the history of the process up to and including the current time t.

Definition 4.2.1 (A Stopping Time) *Let $\{\sigma_t\}$ be an increasing family of σ-algebras (of subsets of Ω). A function $\tau : \Omega \to [O, \infty]$ is called a (strict) stopping time w.r.t. $\{\sigma_t\}$ if*

$$\{\omega; \tau(\omega) \leq t\} \in \sigma_t, \quad \textit{for all } t \geq 0.$$

In other words, it should be possible to decide whether or not $\tau \leq t$ has occurred on the basis of the knowledge of the history of our process up to and including our current time, that is on the basis of the knowledge of σ_t.

Example 4.2.2 (Stopping Time for a Martingale) *A fair coin is tossed repeatedly; let T be the time of the first head. Writing X_i for the number of heads on the ith toss, we have that*

$$\{T = n\} = \{X_n = 1, X_j = 0 \text{ for } 1 \leq j < n\} \in \mathcal{F}_n$$

where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. That is, it is possible to decide whether or not our total number of heads is 1 by considering only the results of our coin tossing up to and including the current toss. Therefore T is a stopping time. In this case T is finite almost surely.

In addition, it is important to understand that we cannot think about an optimal stopping time for a particular future path as we do not, nor by Definition 4.2.1 should we need to know, the individual future path. Instead, we wish to find a stopping time that will give the largest *expected* reward. That is we must start thinking about the stopping time which will return the best reward over the long run.

The Reward Function

The principle aim in optimal stopping problems is to maximise some expected *reward* function. It seems logical therefore that before we look at a formal definition of optimal stopping we must, at least heuristically, decide what is meant by a reward function.

Clearly, a reward function will be some function of the underlying stochastic process. For example if the underlying stochastic process is the movement of our stock price, then a possible reward function could be the net discounted profit made if we sold our stocks now

$$g(S_t, t) = e^{-r(T-t)}(S_t - S_0). \tag{4.22}$$

Some reward functions do not involve time as a variable. These reward functions only depend on the state of the underlying stochastic process. When the reward function does not involve time, we refer to the problem as the *time homogeneous* optimal stopping problem.

Some reward functions, such as (4.22), do depend on time. In this case the solution to the optimal stopping problem will depend not only on the underlying stochastic process, but also in some way on the timing of the process

as well. So a state of the process at one time may have a different effect on our decision to stop than the same state at a different time. When the reward function involves time, we refer to the problem as the *time inhomogeneous* optimal stopping problem.

4.2.1 The Optimal Stopping Problem

We are now in a position to look at the formal specification of the optimal stopping problem.

Definition 4.2.2 (The Optimal Stopping Problem) *Let X_t be an Itô diffusion on $\mathbb{R}^m$ and let g (the reward function) be a given function on $\mathbb{R}^m$, satisfying*

a) $g(\xi) \geq 0$ for all $\xi \in \mathbb{R}^m$, and

b) g is continuous.

We wish to find a stopping time $\tau^* = \tau^*(x, \omega)$ (called on optimal stopping time) for $\{X_t\}$ such that

$$\mathbb{E}^x[g(X_{\tau^*})] = \sup_{\tau} \mathbb{E}^x[g(X_\tau)] \quad \text{for all } x \in \mathbb{R}^m, \tag{4.23}$$

the supremum being taken over all stopping times τ for $\{X_t\}$. We also want to find the corresponding optimal expected reward

$$g^*(x) = \mathbb{E}^x[g(X_{\tau^*})]. \tag{4.24}$$

That is, we wish to find a stopping time such that the reward gained by stopping is greater than the expected reward if we did not stop.

We will see later that the first condition on g may be relaxed in certain situations.

To solve this problem, we will examine the time homogeneous case first. We will then outline modifications to this method of solution for the time inhomogeneous case.

Time homogeneity

Throughout the rest of this chapter we will examine what happens in the time homogeneous case. At first this may seem a little dismissive however this is not such a terrible thing. If we encounter a problem that is not time homogeneous (where the reward function does indeed depend on time), we can define a new diffusion Y_t that transforms our time inhomogeneous

problem into a homogeneous one. From there we can carry on as with the homogeneous case.

If, in our time inhomogeneous problem, we have an Itô diffusion X_t defined as

$$X_t = b(r, X_t)dt + \sigma(\alpha, X_t)dW_t,$$

and a reward function

$$g^*(t,x) = \sup_\tau \mathbb{E}^{(t_0,x)}\left[g(\tau, X_\tau)\right] = \mathbb{E}^{(t_0,x)}\left[g(\tau^*, X_{\tau^*})\right],$$

then we can define a new Itô diffusion $Y_t = Y_t^{(s,x)} \in \mathbb{R}^{m+1}$,

$$Y_t = \begin{bmatrix} s+t \\ X_t^x \end{bmatrix} \quad ; \quad t \geq 0.$$

That is, $Y_t - Y_t^{(s,x)}$ is the graph of $X_t - X_t^x$ shifted to start at (s,x). Now as the new Itô diffusion takes time into account, the original problem as in Definition 4.2.2 can be written in terms of Y_t,

$$g^*(s,x) = \sup_\tau \mathbb{E}^{(s,x)}\left[g(Y_\tau)\right] = \mathbb{E}^{(s,x)}\left[g(Y_{\tau^*})\right],$$

and the problem can then be solved using the same methods as for the time homogeneous problem.

In the rare case that something different is required for the time inhomogeneous case, we will make a note. Otherwise, when we have a time inhomogeneous problem, we will replace in the notes X_t with Y_t and $g(x)$ with $g(s,x)$.

4.2.2 The Least Super-harmonic Majorant

The first step of our solution involves finding a special function called the *least super-harmonic majorant* of our reward function, g. This concept can take a little getting used to so we will present it, as Øksendal (Øksendal 1998) does, in stages.

Definition 4.2.3 (A Super-harmonic Function) *A continuous function $f : \mathbb{R}^m \to [0, \infty]$ is called* super-harmonic *(w.r.t. the Itô diffusion, X_t) if*

$$f(x) \geq \mathbb{E}^x[f(X_\tau)] \tag{4.25}$$

for all stopping times τ and all $x \in \mathbb{R}^m$.

That is a function is super-harmonic if, given our starting point x, the expected value of the function is always less than or equal to the value of the function at our starting point. So if f is a reward function which is super-harmonic, our expected future reward will always be less than or equal to our current reward.

At first glance, the definition of super-harmonic may resemble the definition of a super-martingale, however the two are very different. The definition of super-harmonic refers to a property of a function (of an Itô diffusion), whereas the definition of a super-martingale refers to a property of the underlying stochastic process.

Another way of defining or determining super-harmonicity can be found by considering a rearrangement for equation (4.25) (and considering for all t, not just stopping times),

$$\begin{aligned} & \mathbb{E}^x[f(X_t)] \leq f(x) && (4.26)\\ \Rightarrow \quad & \mathbb{E}^x[f(X_t)] - f(x) \leq 0 && (4.27)\\ \Rightarrow \quad & \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} \leq 0 && (4.28) \end{aligned}$$

which we should recognise as the generator of f. Reconsidering this last equation in terms of stopping times gives us the following corollary.

Corollary 4.2.1 *Let $f : \mathbb{R}^m \to [0, \infty]$ be twice continuously differentiable such that Af is bounded, where A is the generator of the Itô diffusion, X_t. Then f is called super-harmonic if*

$$\mathbb{E}^x \left[\int_0^\tau Af(X_s) ds \right] \leq 0, \tag{4.29}$$

for all stopping times, τ.

Proof: *By Dynkin's formula.*

Definition 4.2.4 (Least Super-harmonic Majorant) *Let h be a real function on $\mathbb{R}^m$. If f is a super-harmonic function and $f \geq h$ we say that f is a* super-harmonic majorant *of h (w.r.t X_t). The function*

$$\overline{h}(x) = \inf_f f(x); \quad x \in \mathbb{R}^m, \tag{4.30}$$

the infimum being taken over all super-harmonic majorants f of h, is called the least super-harmonic majorant *of h.*

The construction of the least super-harmonic majorant of a function h is not a trivial task. For this reason the main result of this section, the method of solution for the optimal stopping problem, will be presented first followed by the method of construction of the least super-harmonic majorant.

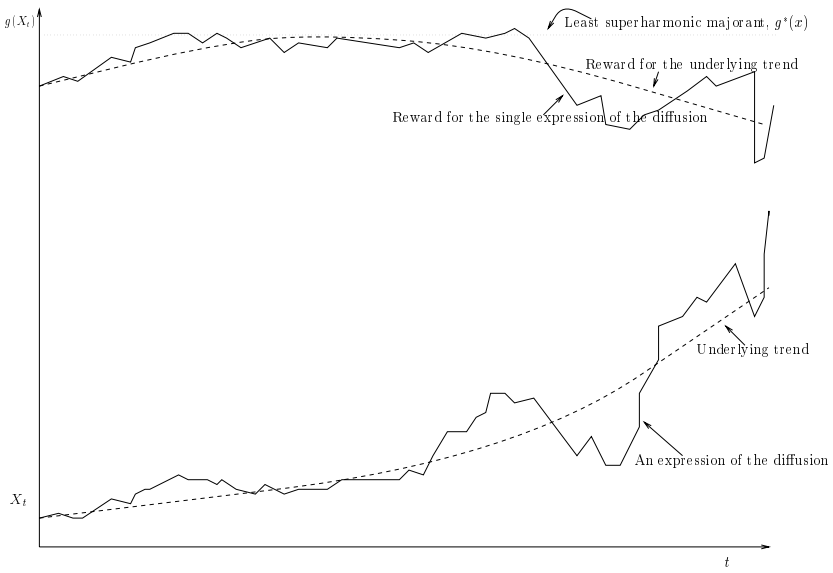


Figure 4.1: An example of the concept of the least super-harmonic majorant, g^*, of the reward function, g

4.2.3 Formulation of Some Problem Applications

We now include the formulation of two application problems using the notation used above. The first of these is a problem that is very close to our hearts, the 'When is the best time to sell my stock' problem.

Example 4.2.3 (When is the best time to sell stock?) *Suppose we currently own a stock (or for that matter, any asset - house, business, etc) whose price, X_t varies according to geometric Brownian motion, that is*

$$dX_t = rX_t dt + \alpha X_t dW_t,$$

where r is the underlying trend of the stock and α is the variance term.

Let us also suppose that the reward is the return on this stock which, if we sell it, is a discounted price process, that is,

$$g(t, X_t) = e^{-\rho t}(X_t - a),$$

where ρ is the riskless rate of interest and a is a some constant representing transaction costs. So we are wanting to know when to sell the stock so that we receive the largest reward, after taking the transaction cost into account.

As the reward function does depend on time, the problem is a time-inhomogeneous one. Consequently we define a new diffusion process, Y_t as

$$Y_t^{(s,x)} = \begin{pmatrix} s+t \\ X_t^x \end{pmatrix},$$

that is,

$$dY_t = \begin{bmatrix} 1 \\ rX_t \end{bmatrix} dt + \begin{bmatrix} 0 \\ \alpha X_t \end{bmatrix} dW_t.$$

Our reward function $g(t,x)$ then becomes $g(s,x) = e^{-\rho s}(x-a)$.

This problem is an interesting one, so we will use it as our illustrative example throughout the next section.

Example 4.2.4 (American Put Option) *An American option is similar to a European option in that the owner has the right but not the obligation to exercise the option. However, unlike an European option, the American option does not have to be exercised only on the maturity date but at any time up to and including the maturity date, T. The decision to exercise the option at any time t only depends on the history up to time t. More formally,*

$$\{\omega; \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

We can define the payoff function for the American option as

$$F(\tau(\omega), \omega) = \sup_{\tau \leq T} (S_\tau - K, 0)_+$$

where K is the strike price for the option. That is, the supremum at time τ less than T of the maximum of the stock price minus the strike price or zero. Let us define the Itô diffusion $X_t = (X_0(t), X_1(t))$ which consists of a riskless asset

$$dX_0(t) = \rho(t, X_0(t))X_0(t)dt$$

and a risky asset

$$dX_1(t) = \mu(t, X_1(t))dt + \sigma(t, X_1(t))dW_t$$

where $\rho(t, X_0(t))$, $\mu(t, X_1(t))$ and $\sigma(t, X_1(t))$ satisfy the conditions given in (4.2). If we define a new probability measure $\mathbb{Q}$ on $\mathcal{F}_t$, then by applying a change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = M(T, \omega)$$

where $M(t,\omega)$ is given by

$$M(t,\omega) = \exp\left(- \int_0^t u(s, X_s) dW_s - \frac{1}{2} \int_0^t u^2(s, X_s) ds \right)$$

then X_t is a martingale under the new probability measure $\mathbb{Q}$. If the payoff function for the American option is of the form

$$F(t,\omega) = g(t, X_t)$$

then if we assume that the market is complete, the price of the contingent claim is given by

$$\begin{aligned} p_A(F) &= \sup_{\tau \leq T} \mathbb{E}_{\mathbb{Q}}[\xi(\tau) g(\tau, X_\tau)] \\ &= \sup_{\tau \leq T} \mathbb{E}_{\mathbb{P}}[M(\tau)\xi(\tau) g(\tau, X_\tau)] \end{aligned}$$

where $\xi(\tau)$ is the discount process. If we define the function $K(t)$ by

$$\begin{aligned} K(t) &= M(t)\xi(t) \\ &= \exp\left(-\int_0^t u(s, X_s) dW_s - \int_0^t \left(\frac{1}{2} u^2(s, X_s) + \rho(s, X_s) \right) ds \right) \end{aligned}$$

then the corresponding diffusion equation for $K(t)$ is given by

$$dK(t) = -\rho(t, X_t) K(t) dt - u(t, X_t) K(t) dW_t.$$

If we define the 4-dimensional Itô diffusion Y_t by

$$dY_t = \begin{bmatrix} dt \\ dK(t) \\ dX_t \end{bmatrix} = \begin{bmatrix} 1 \\ -\rho K \\ \rho X_0 \\ \mu \end{bmatrix} dt + \begin{bmatrix} 0 \\ -uK \\ 0 \\ \sigma \end{bmatrix} dW_t,$$

then we can see that the price of the contingent claim is given by

$$p_A(F) = \sup_{\tau \leq T} \mathbb{E}[G(Y_\tau],$$

where $G(y) = G(s,k,x) = kg(s,x)$ and $y = (s,k,x) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2$. Thus, the price of the American contingent claim, $p_A(F)$, can be thought of as the solution to the optimal stopping problem with Itô diffusion Y_t given above.

Unfortunately, even in the most simplistic case of the American Put option problem, an explicit solution has not yet been found, however, some interesting partial results can be found in (Øksendal 1998) with further references given within.

4.3 Methods of Solution

We are now ready for our main results on the optimal stopping problem.

4.3.1 Solution Using First Principles

Theorem 4.3.1 (Existence Theorem for Optimal Stopping) *Let g be a continuous, non-negative reward function which has a least super-harmonic majorant, $\hat{g}$. Let g^* denote the optimal reward.*
a) Then

$$g^*(x) = \hat{g}(x). \tag{4.31}$$

b) If we define the continuation region, D

$$D = \{x; g(x) < g^*(x)\}, \tag{4.32}$$

and for $N = 1, 2, \ldots$ define $g_N = \min(g, N), D_N = \{x; g_N(x) < \hat{g_N}(x)\}$ and $\sigma_N = \tau_{D_N}$.
Then $D_N \subset D_{N+1}$ and $D = \bigcup_N D_N$.
If $\sigma_N < \infty$ a.s. w.r.t Q^x (the probability measure of X_t) for all N then

$$g^*(x) = \lim_{N \to \infty} \mathbb{E}^x \left[g(X_{\sigma_N}) \right]. \tag{4.33}$$

c) In particular, if $\tau_D < \infty$ a.s. w.r.t. Q^x and the family $\{g(X_{\sigma_N})\}_N$ is uniformly integrable w.r.t. Q^x, then

$$g^*(x) = \mathbb{E}^x \left[g(X_{\tau_D}) \right] \tag{4.34}$$

and $\tau^* = \tau_D$ is an optimal stopping time.

The proof for this Theorem can be found in Øksendal (Øksendal 1998), page 201.

This theorem gives a sufficient condition for the existence of an optimal stopping time τ^*. Unfortunately, it does not prove that an optimal stopping time necessarily exists in general.

Corollary 4.3.1 *Suppose there exists a Borel set H such that*

$$\tilde{g}_H(x) := \mathbb{E}^x \left[g(X_{\tau_H}) \right] \tag{4.35}$$

is a super-harmonic majorant of g. Then

$$g^*(x) = \tilde{g}_H(x), \tag{4.36}$$

so $\tau^* = \tau_H$ is optimal.

Corollary 4.3.2 *Let*

$$D = \{x; g(x) < \hat{g}(x)\} \tag{4.37}$$

and put

$$\tilde{g}(x) = \tilde{g}_D(x) = \mathbb{E}^x \left[g(X_{\tau_D}) \right]. \tag{4.38}$$

If $\tilde{g} \geq g$ *then* $\tilde{g} = g^*$.

Theorem 4.3.2 (Uniqueness Theorem for Optimal Stopping) *Define as before*

$$D = \{x; g(x) < g^*(x)\} \subset \mathbb{R}^m.$$

Suppose there exists an optimal stopping time $\tau^* = \tau^*(x, \omega)$ *for the problem 4.23 for all x. Then*

$$\tau^* \geq \tau_D \quad \text{for all } x \in D \tag{4.39}$$

and

$$g^*(x) = \mathbb{E}^x \left[g(X_{\tau_D}) \right] \quad \text{for all } x \in \mathbb{R}^m. \tag{4.40}$$

Hence τ_D *is an optimal stopping time for the problem 4.23.*

The proof for this Theorem can be found in Øksendal (Øksendal 1998), page 205.

These existence and uniqueness theorems give us a two step method for solving the optimal stopping problem:

Step 1. Identify g^* with the least super-harmonic majorant $\hat{g}$ of g with respect to X_t.

Step 2. Define the continuation region $D \subset \mathbb{R}^m$ by

$$D = \{x; g(x) < g^*(x)\}.$$

Then (under certain conditions) the first exit time $\tau^* = \tau_D$ for D for X_t solves the optimal stopping problem (4.23), and hence the optimal reward (4.24) is given by

$$g^*(x) = \mathbb{E}^x \left[g(X_{\tau_D}) \right].$$

So our method of solution tries to identify a function, $\hat{g}$, that has the following properties:

- The new function, $\hat{g}$, is greater than our reward function,
- The new function, $\hat{g}$, has a future expectation which is less than the current value, and
- The new function, $\hat{g}$, is the least of all such functions.

Once we have found this function $\hat{g}$, then we can define our continuation region as specified earlier. It is left as an exercise to show that if D is time invariant then it must have the form

$$D = \{x : 0 < x < x_0\}.$$

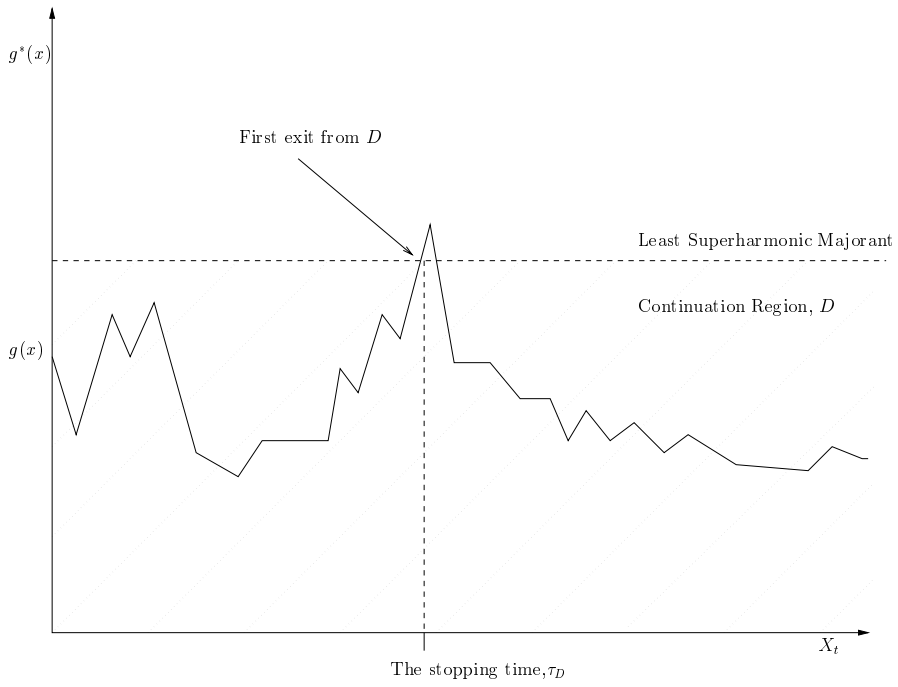


Figure 4.2: An example of the concept of the least super-harmonic majorant, g^*, and its role in the determination of the continuation region, D.

Example 4.3.1 (Øksendal (Øksendal 1998) page. 206) *Let $X_t = B_t$ be Brownian motion in $\mathcal{R}^3$ and let the reward function be*

$$g(x) = \begin{cases} |x|^{-1} & for\ |x| \geq 1 \\ 1 & for\ |x| < 1 \end{cases} \quad ; \quad x \in \mathcal{R}^3.$$

Then g is super-harmonic in $\mathcal{R}^3$, so $g^* = g$ everywhere and the best policy is to stop immediately, no matter where the starting point is.

This example illustrates the principle that if the reward function itself is super-harmonic, then it is always optimal to stop immediately. This is because each point in the function does escape the continuation region, or at least lies on its boundary. If the next step in the diffusion is a negative step (that is, the process dips slightly), then a new continuation region is set and again, each point lies on its boundary.

Construction of the Least Super-harmonic Majorant

Earlier in the chapter, the reader was asked to suspend belief for a moment and to assume that the least super-harmonic majorant of the reward function could be readily found. If we assumed this, then understanding the method of solution for the optimal stopping problem became easier. Now that the method of solution is understood, it is time to explore the construction of the least super-harmonic majorant.

The method of construction presented by Øksendal (Øksendal 1998) is an iterative procedure, starting with the reward function g.

Theorem 4.3.3 (Construction of the Least Super-harmonic Majorant) *Let $h_0 = g$, where g is the reward function, be a non-negative continuous function of $\mathbb{R}^m$ and define inductively*

$$h_n(x) = \sup_{t \geq 0} \mathbb{E}^x \left[h_{n-1}(X_t) \right]; \quad n = 1, 2, \ldots \tag{4.41}$$

Then $h_n \uparrow \hat{g}$, where $\hat{g}$ is the least super-harmonic majorant.

This is not a simple task. The authors have had difficulty finding a simple example by which to demonstrate this method. We will see later that it can be done reasonably using numerical methods.

4.3.2 Approximate Solution Using Characteristic Operators

While solution using first principles tries to define a continuation region by finding the least super-harmonic majorant, this method tries to approximate the continuation directly using characteristic operators. If we can find the continuation region directly, we can eliminate the need to find the least super-harmonic majorant.
Now the definition of our continuation region D,

$$\{x : g(x) < g^*(x)\},$$

where $g^*(x) = \mathbb{E}^x [g(X_{\tau^*})]$, can be rewritten as

$$\{x : \mathbb{E}^x [g(X_{\tau^*})] - g(x) > 0\}.$$

We can use a similar argument to that used for Corollary 4.2.1 (but considering only stopping times) to intuitively show Lemma 4.3.1 below.

Lemma 4.3.1 *Let $\mathcal{A}$ be the characteristic operator of X_t. Assume $g \in \mathcal{C}^2(\mathbb{R}^m)$. Define*

$$U = \{x; \mathcal{A}g(x) > 0\}. \tag{4.42}$$

Then, with D as before,

$$U \subset D. \tag{4.43}$$

This observation has two implications. The first is that it is never optimal to stop the process before it exits from U. So if $U = D$ (for example if $U = \mathbb{R}^m$, then $U = D$), we have another tool for calculating our continuation region. But if $U \neq D$ (the usual case), then it is optimal to *proceed* beyond U before stopping.

Example 4.3.2 (Stock Selling Problem) *The characteristic operator of the process $Y_t = (s + t, X_t)$ where X_t is geometric Brownian motion is given by*

$$\mathcal{A}f(s, x) = \frac{\partial f}{\partial s} + rx\frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2}; \quad f \in C^2(\mathbb{R}^m).$$

So for $g(s,x) = e^{-\rho t}(x-a)$, $\mathcal{A}g(s,x)$ *is given by*

$$\begin{aligned} \mathcal{A}g(s,x) &= -\rho e^{-\rho s}(x-a) + rxe^{-\rho s} \\ &= e^{-\rho s}((r-\rho)x + \rho\alpha). \end{aligned}$$

and solving $\mathcal{A}g(x) > 0$ *gives*

$$\begin{array}{ll} x < \frac{\alpha\rho}{\rho - r} & \text{if } r < \rho \ , \\ x \in \mathcal{R} \times \mathcal{R}_+ & \text{if } r \geq \rho \ . \end{array} \tag{4.44}$$

For $r \geq \rho$, *this means that* U *is defined as*

$$U = \{(s,x) : \mathcal{A}g(s,x) > 0\} = \{\mathcal{R} \times \mathcal{R}_+\} \, ,$$

and as $U \subseteq D$, *then the continuation region* D *must be*

$$D = \{\mathcal{R} \times \mathcal{R}_+\} \, .$$

Because the continuation region is infinite, the stock price process will never leave, so an optimal stopping time does not exist. That is, if $r \geq \rho$ *we never sell our stock! This is hardly surprising when we realise that* $r \geq \rho$ *means that that the average rate of return of the stock process is greater than the riskless rate of return.*

While this example may seem as if it is proving the obvious, it does illustrate an important feature of using characteristic operators - that sometimes it is simple to show that U is unbounded, so D is unbounded and thus no optimal stopping time will exist.

In the above example, for $r < \rho$, D is defined for some region where the least super-harmonic majorant is greater than $x = \frac{\alpha\rho}{\rho - r}$. This can be done using characteristic operators, but is reasonably difficult. The next section shows a much easier way of doing the same thing.

4.3.3 Solution Using Variational Inequalities

Once we have determined the conditions for which a stopping time exists (see the previous section), we now look for a simple method by which we can determine the continuation region, D, precisely. This can be done using variational inequalities.

This theorem is given for a time inhomogeneous problem. For a time homogeneous problem, we make the usual adjustments and the theorem still holds.

Theorem 4.3.4 (Variational Inequalities for Optimal Stopping) *Suppose we can find a function $\phi : V \to \mathcal{R}$, where V is a closed set, such that*

1. *ϕ is continuous on the boundary of V, and continuously differentiable on the remainder of V,*

2. *$\phi \geq g$ on V and $\phi = g$ on the boundary of V, where g is the reward function,*

3. *$\mathcal{A}\phi = 0$ on D, where D is defined as*

$$D = \{x \in V : \phi(x) > g(x)\}.$$

   and $\mathcal{A}$ is the characteristic operator as before,

4. *$\phi(x)$ is sufficiently 'nice' in other ways,*

5. *τ_D is the time of the first exit from D.*

Then

$$\phi(y) = \sup_{\tau \leq T} \mathbb{E}^y \left[g(Y_\tau) \right]; \qquad y \in V \tag{4.45}$$

and

$$\tau^* = \tau_D$$

is an optimal stopping time for this problem.

The proof for this theorem is quite lengthy, so it will not be given here, but it can be found in Øksendal (Øksendal 1998).

Although this theorem seems a little daunting, it gives us a four step program for solution success!

Step 1. Using characteristic operators as in section 4.3.2 to determine the constraints on when an optimal stopping time exists
Step 2. Assume that the continuation region D, is of the form

$$D = \{(s,x) : 0 < x < x_0\}, \text{ for some constant } x_0, \text{ which is to be found}$$

Step 3. Find $\phi(s,x)$, the bounded solution of the free boundary problem

$$\begin{aligned} \mathcal{A}\phi &= 0, \qquad 0 < x < x_0 \\ \phi(s,x_0) &= g(s,x_0). \end{aligned}$$

Step 4. Find the characterisation of the continuation region, x_0 by solving

$$\phi'(s,x) = g'(s,x) \quad \text{at } x = x_0 \tag{4.46}$$

We can illustrate this method by solving the selling stock problem.

Example 4.3.3 (Selling Stock cont') *Given the stock price process follows geometric Brownian motion,*

$$dX_t = rX_t dt + \alpha X_t dW_t,$$

where r is the underlying trend of the stock and α is the variance term, and also given the reward function,

$$g(t,X_t) = e^{-\rho t}(X_t - a),$$

where ρ is the riskless rate of interest and a is a some constant representing transaction costs.
Step 1. *In section 4.3.2, we showed that this problem does not have an optimal stopping time unless $r < \rho$.*
Step 2. *We assume the continuation region is time invariant, that is the continuation region has the form*

$$D = \{(s,x) : 0 < x < x_0\},$$

Step 3. *We find a bounded solution to the free boundary problem*

$$\begin{aligned} \mathcal{A}\phi &= 0 \\ \phi(s,x_0) &= g(s,x_0). \end{aligned}$$

That is, we look for a solution of

$$\frac{\partial \phi}{\partial s} + rx\frac{\partial \phi}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 \phi}{\partial x^2} = 0 \tag{4.47}$$

$$\phi(s, x_0) = g(s, x_0). \tag{4.48}$$

We can solve this problem using separation of variables. If we assume $\phi(s,x)$ *is of the form* $\phi(s,x) = e^{-\rho s}\psi(x)$*, then equation* (4.48) *becomes*

$$\begin{aligned} 0 &= \frac{\partial \phi}{\partial s} + rx\frac{\partial \phi}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 \phi}{\partial x^2} \\ &= -\rho e^{-\rho s}\psi(x) + rxe^{-\rho s}\psi'(x)\frac{1}{2}\alpha^2 x^2 e^{-\rho s}\psi''(x) \\ &= -\rho\psi(x) + rx\psi'(x)\frac{1}{2}\alpha^2 x^2 \psi''(x) \end{aligned}$$

and so we can reduce the PDE problem to a free boundary ODE problem,

$$-\rho\psi(x) + rx\psi'(x)\frac{1}{2}\alpha^2 x^2 \psi''(x) = 0 \tag{4.49}$$

$$\psi(x_0) = x_0 - a. \tag{4.50}$$

The general solution to this ODE is

$$\psi(x) = C_1 x^{\gamma_1} + C_2 x^{\gamma_2},$$

where C_1, C_2 *are arbitrary constants and* γ_i*,* $i = 1, 2$ *are the solution to the characteristic equation*

$$\begin{aligned} 0 &= \frac{1}{2}\alpha^2\gamma(\gamma-1) + r\gamma - \rho \\ &= \frac{1}{2}\alpha^2\gamma^2 + (r - \frac{1}{2}\alpha^2)\gamma - \rho. \end{aligned}$$

That is,

$$\gamma_i = \frac{1}{2} - \alpha^{-2}\left[r \pm \sqrt{(r - \frac{1}{2}\alpha^2)^2 + 2\rho\alpha^2}\right], \quad i = 1, 2.$$

And we note that $\gamma_2 < 0 < \gamma_1$*.*

Now since $\psi(x)$ *is bounded as* $x \to 0$ *we must have* $C_2 = 0$ *and the boundary requirement* $\psi(x_0) = x_0 - a$ *gives* $C_1 = x_0^{-\gamma_1}(x_0 - a)$*. Thus the bounded solution* $\phi(s,x)$ *of* (4.48) *is*

$$\phi(s,x) = e^{-\rho s}(x_0 - a)\left(\frac{x}{x_0}\right)^{\gamma_1}, \quad 0 < x < x_0. \tag{4.51}$$

Step 4. *We find x_0 by solving*

$$\begin{aligned} \phi'(s,x) &= g'(s,x) \\ \textit{that is} \quad \gamma_1(x_0-a)x_0^{-1} &= 1. \end{aligned}$$

Solving this gives us

$$x_0 = \frac{a\gamma_1}{\gamma_1 - 1}.$$

And so for $r < \rho$, our continuation region D is

$$D = \left\{(s,x) : 0 < x < \frac{a\gamma_1}{\gamma_1 - 1}\right\},$$

and our optimal stopping time is the time that the stock price first exits this continuation region. That is, we sell out stock the first time the stock value reaches $\frac{a\gamma_1}{\gamma_1 - 1}$ in value.

If we take look at a simulation of the selling stock problem for $r < \rho$, we can gain a better appreciation for this result.

Suppose the stock price process follows the SDE

$$dX_t = 0.07X_t dt + 0.12X_t dW_t,$$

and we have a reward function

$$g(\xi, s) = e^{-0.1s}(\xi - 5),$$

where the transaction cost is \$5. Then we obtain $\gamma_1 = 1.375$ and thus $x_{max} = \$18.32$. If we plot the boundary of the continuation region and overlay a simulation of the stock price, as in Figure 4.3, then we can see why selling the stock at this price does return the optimal reward (the bottom line is the reward corresponding to the stock price).

If we then contrast this by looking at the same style of plot for the $r > \rho$ case (see Figure 4.4), we can see that we should never sell our stock - as the reward function continues increasing.

4.4 Numerical Implementation

As it stands we have two methods for solving the optimal stopping problem, one using first principles and another using variational inequalities. Analytically we found that variational inequalities provided a straight forward method of solving the optimal stopping stopping problem whereas first principles, whilst making fewer assumptions, proved to be a difficult skill to master. In this section we examine the feasibility of numerically implementing both of these methods.

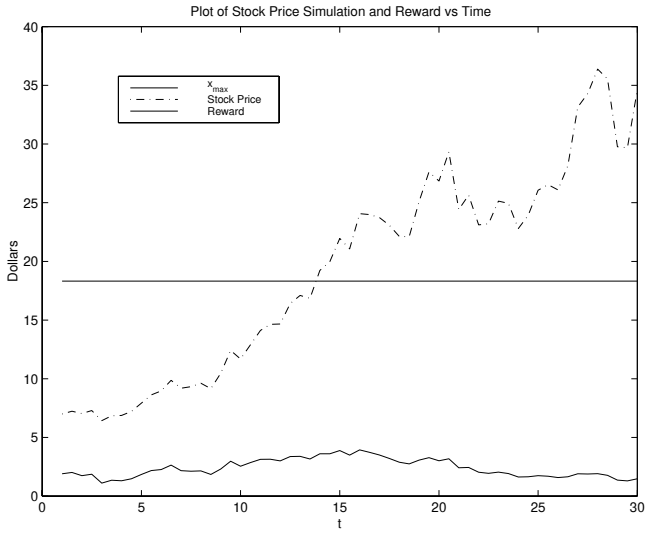


Figure 4.3: A simulation of the stock process shown against the continuation region using 100 time steps.

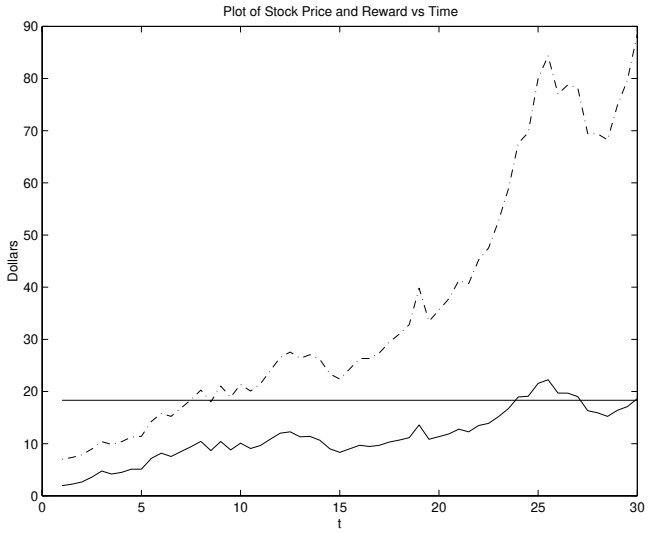


Figure 4.4: A simulation of the stock process shown against the continuation Region (for $r > \rho$ case using 100 time steps.) .

4.4.1 First Principles

We saw in Theorem 4.3.3 that we could construct the least super-harmonic majorant iteratively by defining $h_0(x) = g(x)$, where $g(x)$ is the reward function, and

$$h_n(x) = \sup_{t \geq 0} \mathbb{E}^x \left[h_{n-1}(X_t) \right]; \quad n = 1, 2, \ldots \tag{4.52}$$

then $h_n \uparrow \hat{g}$, where $\hat{g}$ is the least super-harmonic majorant.

We also noted that, analytically, this was not a trivial task. We can however, attempt to emulate this numerically.

If we look closely at the iterative formula (4.52), we will note that h_n is a function of x, that is each new iterate is a function of the starting value x. This means that to simulate the iterative definition, we will need to discretize the set of starting values (x's), so that we will obtain a function over this set.

Once we have discretized the set of starting values, we will look at each of these in turn to arrive at some value for that particular x_i, so that we obtain a set of points which we can interpolate h_n with. This concept is seen in Figure 4.5.

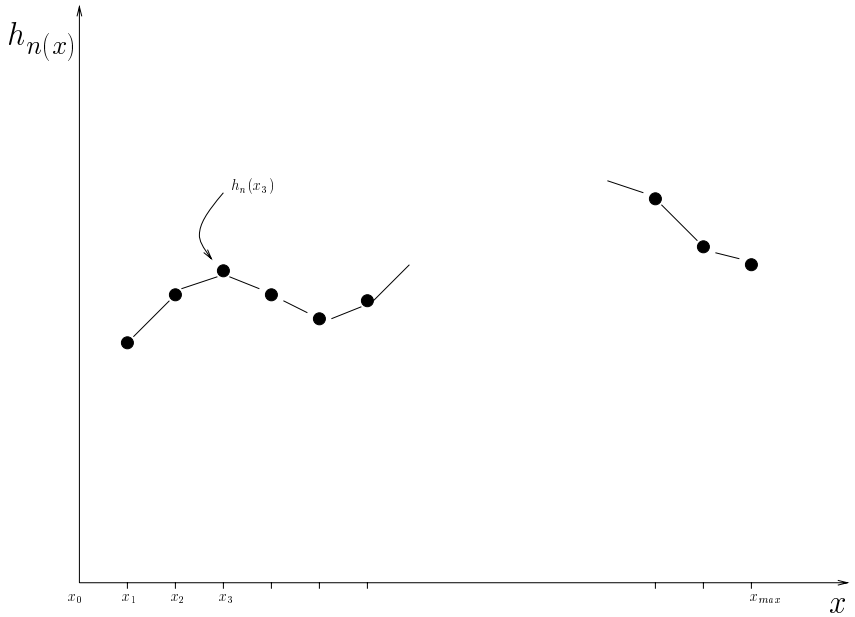


Figure 4.5: An example of how the set of starting points, x, must be discretized so that we can generate $h_n(x)$.

Now for each of these starting points x_i, we will need to find a numerical

approximation for

$$h_n(x_i) = \sup_{t \geq 0} \mathbb{E}\left[h_{n-1}(X_t)|x_i\right]. \tag{4.53}$$

We can approximate equation (4.53) with

$$h_n(x_i) = \max_t \left(\frac{1}{n} \sum_{i=1}^{n} h_{n-1}(X_t^i) \right), \tag{4.54}$$

where X_t^i is the i^{th} simulation of the diffusion process.

That is for each starting value x_i, we simulate the diffusion process (conditional on x_i) a number of times and take the average of $h_{n-1}(X_t^i)$ - this gives an approximation for the expected value. We then take the maximum value over time of these averages (for each starting value, x_i). This maximum will then form the point $(x_i, h_n(x_i))$. A graphical understanding of what we are trying to achieve is given in Figure 4.6.

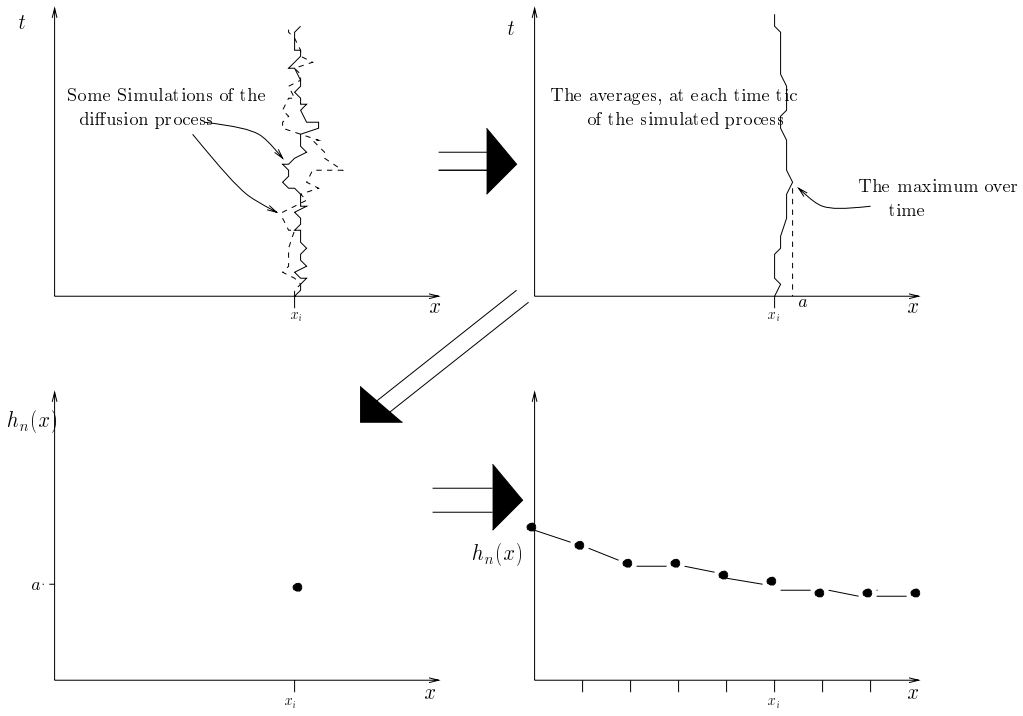


Figure 4.6: An illustration of the steps involved in the BAD method.

So our algorithm is

- Discretize x $\eta = \{x_{min} : x_{step} : x_{max}\}$
- Discretize t $\xi = \{t_0 : t_{step} : T\}$
- Repeat
  - *for each x_i in η, repeat
    - -simulate a diffusion process through time
      (using, say, the stochastic Euler method)
    - -substitute this process to give $h_{n-1}(X_t^i)$
    - -update the mean $\frac{1}{n}\sum_{i=1} nh_{n-1}(X_t^i)$
  - *until the mean converges
  - *find the maximum value over time of the means
  - *set $h_n(x_i) = max_t \bar{h_{n-1}}(X_t^i)$
- until $|h_n(x) - h_{n-1}(x)| < \epsilon$

Working matlab code that implements this algorithm can be found in the appendix.

Results

This algorithm was implemented on the stochastic logistic equation,

$$dX_t = X_t(1 - X_t)dt + X_t dW_t. \tag{4.55}$$

This problem was used because it is a familiar problem. The deterministic version of this equation is a popular model for population studies. The solution to the deterministic logistic equation has a stable steady state solution, as seen in Figure 4.7.

So we were interested in observing if adding a noise term changed this solution dramatically.

When we implemented this as yet unnamed algorithm - let us call it the *Burrage-Alcock-Denman* or *BAD* method - on the stochastic logistic equation, we obtained the plot in Figure 4.8.

So we see that a steady state has evolved, and a least super-harmonic majorant is forming. This plot does highlight a shortfall of the BAD method, that while we can clearly see that the LSM is a straight line - the BAD method's solution will not quite get there (at least not for a long time).

If we assume that the BAD method eventually arrives at the horizontal line $y = 1$ for our LSM, then we can observe a simulated stopping time in Figure 4.9

When we implemented the BAD method to the selling stock problem when we chose $r > \rho$, another shortcoming is highlighted. If we take our

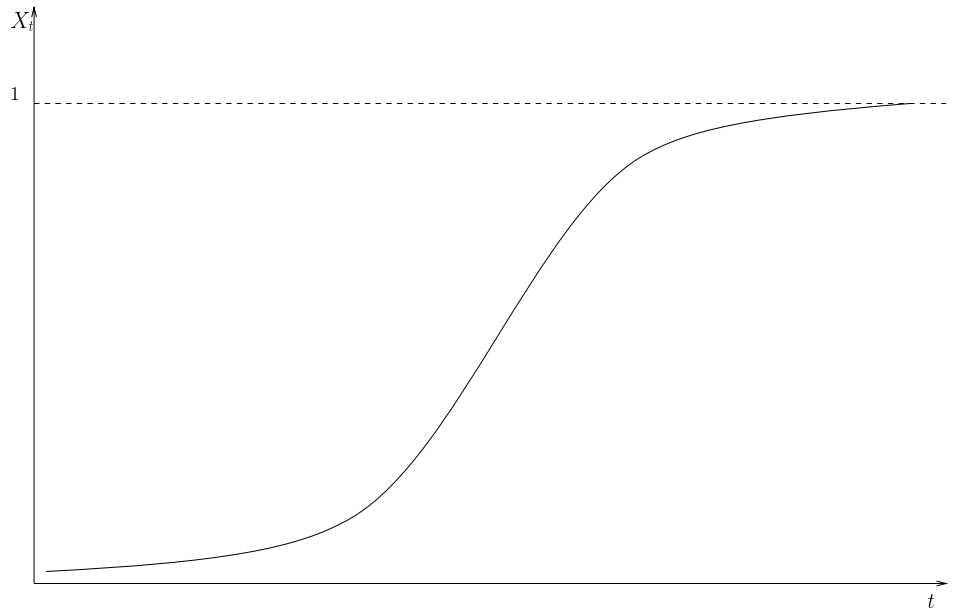


Figure 4.7: The stable solution to the deterministic logistic equation.

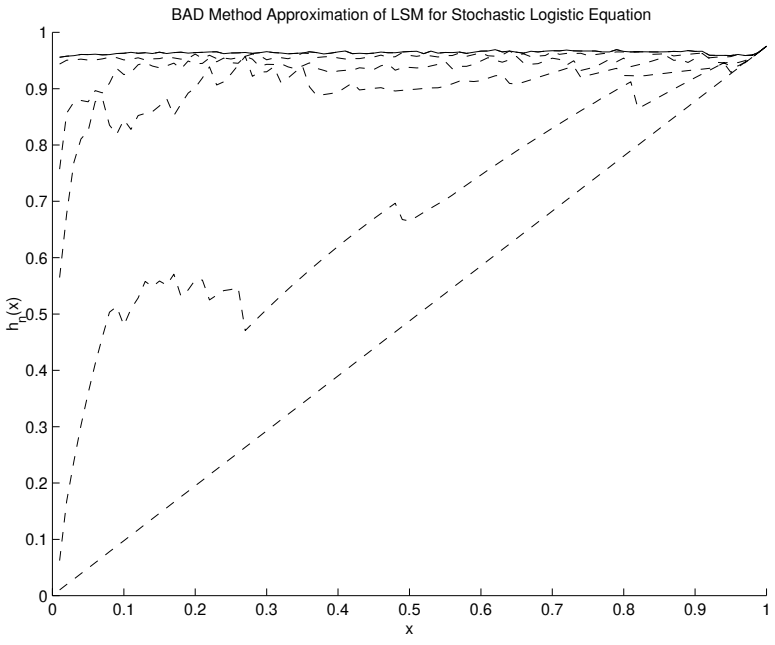


Figure 4.8: The 6 iterations of the BAD method required to obtain the LSM.

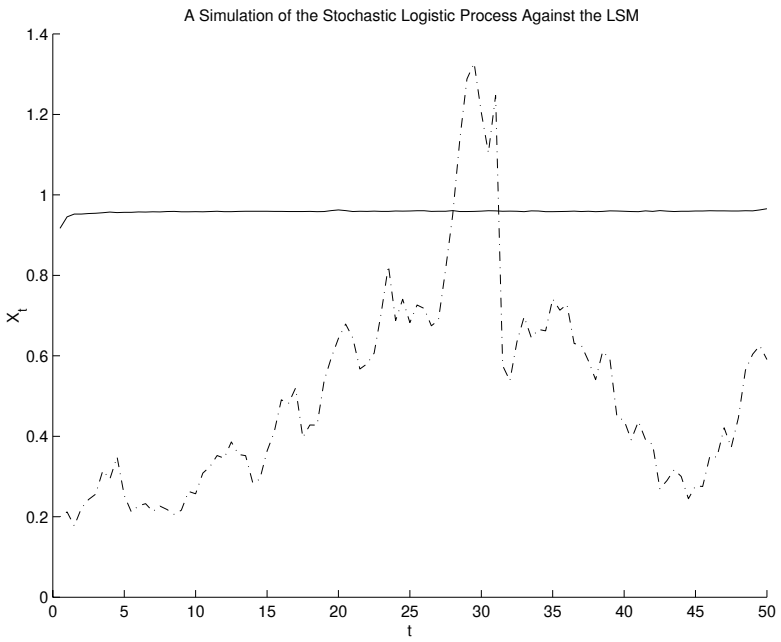


Figure 4.9: A simulation of the stochastic logistic equation (using a stepsize of $h = 0.5$) compared to the LSM approximated in Figure 4.9.

set of initial values to be $[0, 20]$ we obtain one LSM. If we use another set of initial values, say $[0, 40]$, we obtain a different LSM.

This happens because, as we saw in part 3, there is no least superharmonic majorant for this problem if $r > \rho$. Thus the continuation region is infinite and so when we take a subset of the starting values, the LSM obtained by the BAD method depends on x_{max}.

4.4.2 Variational Inequalities

For the problem of finding the optimal reward function, g^*, using the variational inequalities it is not always possible to find an analytic solution. Recall the problem is to find a solution to the free boundary problem

$$\begin{aligned} \mathcal{A}\phi(x) &= 0 \\ \phi(s, x_0) &= g(s, x_0) \\ \phi'(s, x_0) &= g'(s, x_0), \end{aligned}$$

where $\mathcal{A}$ is the characteristic operator of X_t. When an analytic solution is very difficult to find, it is necessary to implement a numerical technique to

approximate the solution. One such method method for solving the above style of problem (PDE's) numerically is the method of lines (MOL). However, to use such a numerical method we need a set of initial values (both initial value and initial derivative values). As yet we have not been able to determine the set of initial conditions necessary to solve this problem numerically.

4.4.3 Further Research

We have introduced the BAD method for approximating the least superharmonic majorant, and implemented it on the stochastic logistic equation. There are however, many questions which still need to answered. Some of the more interesting questions about the BAD method are:

• As this algorithm resembles the Picard iteration for differential equations, and seems to be some fixed point process, can we derive some results about the speed of convergence? In our simple experiments, we noted that the number of iterations to convergence was far fewer in the cases when the ratio $\frac{r}{\alpha}$, where r and α are the drift and diffusion coefficients respectively, was large and far more when this ratio was small. There may be some relationship to 'stiff' ordinary differential equations and their respective convergence results.

• Given that we can derive some convergence results, can this rate of convergence be accelerated? Can we use some Newton-like acceleration or perhaps use windowing principles?

Numerically implementing the result due to variational inequalities appears difficult at this stage due to the lack of initial conditions, but there may be some windows of research opportunity. We might be able to either guess the set of initial conditions and check the validity of the solutions, or to further delve into undertanding the problem and try to derive the correct set of initial conditions.

These are just some of the many questions that need to be answered and certainly warrant future research efforts.

Chapter 5

Stochastic Control

5.1 Deterministic Control

5.1.1 The Problem

In control theory, we are interested in governing the *state* of a *system* by using *controls*. The most common problem is that of finding the control which steers a system from a prescribed initial state to a final state so as to optimise some performance index. We can best illustrate these concepts by way of an example.

Example 5.1.1 (A National Economy) *The economy of a typical capatilistic nation is a* ***system*** *made up in part of the population (as consumers and producers), companies, material goods, production facilities, cash and credit available, and so on. The* ***state*** *of the system can be thought of as a massive collection of data: wages and salaries, profits, losses, sales of goods and services, investment, unemployment, welfare costs, the inflation rate et cetera. The federal government can influence the state of this system by using several* ***controls****, notably the prime interest rate, taxation policy, and persuasion regarding wage and price settlements.*

Most readers will be familiar with classical optimisation problems and constrained optimisation problems. In *optimal control problems*, constraints are induced by the *dynamics* of the system. So, an optimal control problem is characterised by an optimisation problem subject to constraints expressed as differential equations.

We will also be interested in the set of allowable controls (in the National Economy example, we cannot allow the prime interest rate to be negative).

We can state the deterministic optimal control problem more formally

$$\min_{\mathbf{u}\in U} \int_{T_0}^{T_f} g(\mathbf{x}, \mathbf{u}, t)dt$$

subject to

$$\begin{aligned} \dot{\mathbf{x}} &= f(\mathbf{x}, \mathbf{u}) \\ \mathbf{x}(0) &= \mathbf{x}_0 \end{aligned}$$

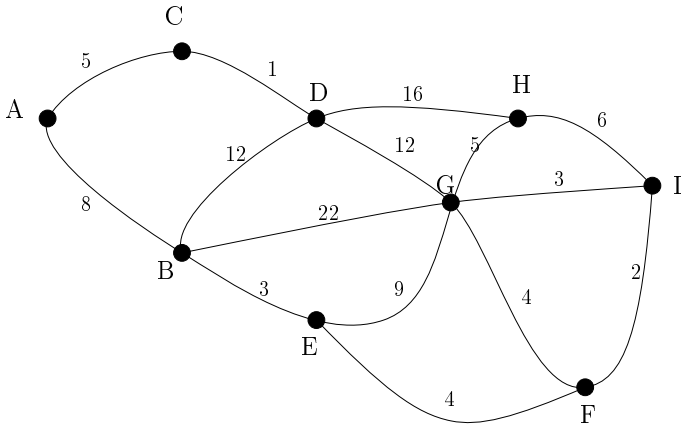
where $\mathbf{u}$ are the controls, U is the set of valid controls and $g(\mathbf{x}, \mathbf{u}, t)$ the objective function (or performance measure).

5.1.2 Solution using Dynamic Programming

Dynamic Programming

One method of solving the deterministic optimal control problem is by utilising dynamic programming principles. Many readers will be aware of dynamic programming, it is a computational technique often used to sequentially make decisions which together define an optimal policy. A typical application of dynamic programming is the routing problem.

Example 5.1.2 (The Routing Problem) *Consider a motorist wishing to travel from town* A *to town* I *in the shortest posible time, where the times required to travel between towns is given in Figure 5.1.2.*



We could find the optimal path by calculating the time required for every possible path, but this would be very inefficient (especially as the number of possible paths increases). We want to be able to identify a solution much more efficiently.

An optimal policy can be found by employing *Bellman's Principle of Optimality.*

Definition 5.1.1 (Bellman's Principle of Optimality) *An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*

So, in the routing problem, for $A-B-E-F-I$ to be the optimal path from town A to town I, $B-E-F-I$ must be the optimal path from town B to town I (which requires $E-F-I$ to be the optimal path from town E to town I, etc.).

Bellman's principle of optimality is the fundamental concept in dynamic programming.

The Hamilton-Jacobi-Bellman Equation

We can use Bellman's principle of optimaliy in the optimal control problem to derive a non-linear p.d.e. called the Hamilton-Jacobi-Bellman (HJB) equation.

Recall that we are trying to control the system described by the state equations

$$\dot{\mathbf{x}}(t) = \mathbf{a}(\mathbf{x}(t), \mathbf{u}(t), t) \tag{5.1}$$

to minimise the objective function

$$J = h(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} g(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau, \tag{5.2}$$

where h and g are specified functions, t_0 and t_f are fixed and τ is a dummy variable of integration.

To allow for a more general solution, we will let the initial time be a parameter, t, rather than it being fixed at t_0. So we will now look to minimise

$$J(\mathbf{x}(t), t, \mathbf{u}_{t\leq\tau\leq t_f}(\tau)) = h(\mathbf{x}(t_f), t_f) + \int_{t}^{t_f} g(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau. \tag{5.3}$$

The minimum performance measure, $J^*(\mathbf{x}(t), t)$, over the set of valid controls is then

$$J^*(\mathbf{x}(t), t) = \min_{\mathbf{u}(\tau), t\leq\tau\leq t_f} \left\{ \int_{t}^{t_f} g(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau + h(\mathbf{x}(t_f), t_f) \right\}. \tag{5.4}$$

By subdividing the interval, we obtain

$$J^*(\mathbf{x}(t),t) = \min_{\mathbf{u}(\tau),t\le\tau\le t_f} \left\{ \int_t^{t+\Delta t} g d\tau + \int_{t+\Delta t}^{t_f} g d\tau + h(\mathbf{x}(t_f),t_f) \right\}. \tag{5.5}$$

The principle of optimality requires that

$$J^*(\mathbf{x}(t),t) = \min_{\mathbf{u}(\tau),t\le\tau\le t_f} \left\{ \int_t^{t+\Delta t} g d\tau + J^*(\mathbf{x}(t+\Delta t),t+\Delta t) \right\}. \tag{5.6}$$

where $J^*(\mathbf{x}(t+\Delta t),t+\Delta t)$ is the minimum cost for the process for the time interval $t+\Delta t \le \tau \le t_f$ with initial state $\mathbf{x}(t+\Delta t)$.

Assuming that the second partial derivatives of J^* exist and are bounded, we can expand $J^*(\mathbf{x}(t+\Delta t),t+\Delta t)$ about the point $(\mathbf{x}(t+\Delta t),t)$ using Taylor series to obtain

$$\begin{aligned} J^*(\mathbf{x}(t),t) \;=\; & \min_{\mathbf{u}(\tau),t\le\tau\le t+\Delta t} \left\{ \int_t^{t+\Delta t} g d\tau + J^*(\mathbf{x}(t),t) + \left[\frac{\partial J^*}{\partial t}(\mathbf{x}(t),t)\right] \Delta t \right. \\ & \left. + \left[\frac{\partial J^*}{\partial \mathbf{x}}(\mathbf{x}(t),t)\right]^T [\mathbf{x}(t+\Delta t)-\mathbf{x}(t)] + \text{terms of higher order} \right\}. \end{aligned} \tag{5.7}$$

Now for small Δt

$$\begin{aligned} J^*(\mathbf{x}(t),t) \;=\; & \min_{\mathbf{u}(\tau)} \{ g(\mathbf{x}(t),\mathbf{u}(t),t)\Delta t + J^*(\mathbf{x}(t),t) + J_t^*(\mathbf{x}(t),t)\Delta t \\ & + J_{\mathbf{x}}^*(\mathbf{x}(t),t)\left[\mathbf{a}(\mathbf{x}(t),\mathbf{u}(t),t)\right]\Delta t + o(\Delta t) \}. \end{aligned} \tag{5.8}$$

As J^* and J_t^* do not depend on our controls, $\mathbf{u}(t)$, we can eliminate them from the minimization. Then by dividing by Δt, we obtain

$$0 = J_t^*(\mathbf{x}(t),t) + \min_{\mathbf{u}(t)} \left\{ g(\mathbf{x}(t),\mathbf{u}(t),t) + J_{\mathbf{x}}^*(\mathbf{x}(t),t)\left[\mathbf{a}(\mathbf{x}(t),\mathbf{u}(t),t)\right] + \frac{o(\Delta t)}{\Delta t} \right\}. \tag{5.9}$$

Finally, we take the limit as $\Delta t \to 0$, we have

$$0 = J_t^* + \min_{\mathbf{u}\in U} \{ g(\mathbf{x},\mathbf{u}) + J_{\mathbf{x}}^* \mathbf{a}(\mathbf{x},\mathbf{u}) \}, \tag{5.10}$$

$$J^*(\mathbf{x}(t_f),t_f) = h(\mathbf{x}(t_f),t_f). \tag{5.11}$$

Equation (5.10) (along with its boundary condition (5.11)) is known as the Hamilton-Jacobi-Bellman equation.

Example 5.1.3 (Optimal Investment and Consumption - Kohn (1999)) *Suppose an investor can choose between two different investment opportunities:*

1. *a money market account, paying constant interest, r, and*

2. *a high-yield account, paying constant interest $R > r$.*

The investor can move money between the two accounts, but in doing so incurs a transaction fee proportional to the amount of money moved: the transaction cost is μX, where X is the amount of money moved. Thus if X dollars was transferred out of the money-market account, $(1-\mu)X$ would be moved into the high-yield account.

Consumption is restricted to the removal of money from the money-market account - there is no transaction fee associated with consumption. The investor can take short positions in either account, however we impose a 'solvency constraint' to avoid the obvious arbitrage. That is, liquidation into money market should not leave the investor in debt. When X is the money-market position and Y is the high-yield position, the solvency condition says:

1. *if $Y \geq 0$ then $X + (1-\mu)Y \geq 0$,*

2. *if $Y \leq 0$ then $X + Y/(1-\mu) \geq 0$.*

The investors goal is to maximise the discounted value of his/her total future consumption.

The state *is an $\mathbb{R}^2$-values function of time, (X_t, Y_t), where X_t is the money-market position at time t and Y_t is the high-yield position at time t. The solvency condition is a* state constraint. *The* control *is an $\mathbb{R}^3$-values function of time, $(\alpha(t), \beta(t), \gamma(t))$, where $\alpha(t)$ is the rate at which money is being moved from money-market to high-yield at time t, $\beta(t)$ is the rate at which money is being moved from high-yield to money-market at time t and $\gamma(t)$ is the consumption rate at time t.*

We also have

$$\begin{aligned} \frac{dX}{dt} &= rX - \alpha + (1-\mu)\beta - \gamma \\ \frac{dY}{dt} &= RY + (1-\mu)\alpha - \beta \end{aligned}$$

with initial conditions $X(0) = X_0$ and $Y(0) = Y_0$. Using a power law to calculate the future vaue of consumption, the investors goal is to maximise

$$\max_{\alpha,\beta,\gamma} \int_0^T e^{-s}\gamma^p(s)ds. \tag{5.12}$$

So our problem is formulated as the solution to the optimal control problem:

Maximise

$$J(x,t) = \int_0^T e^{-s}\gamma^p(s)ds \tag{5.13}$$

over controls restricted by $\alpha(t) \in A$, subject to

$$\dot{\mathbf{X}} = \begin{bmatrix} r \\ R \end{bmatrix} \mathbf{X} + \begin{bmatrix} -1 \\ 1-\mu \end{bmatrix} \alpha + \begin{bmatrix} 1-\mu \\ -1 \end{bmatrix} \beta - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \gamma.$$

The solution, $J^*(x,t)$, of (5.13) is the solution to the HJB eqn:

$$\begin{aligned} 0 \;=\; & J_t^* + \\ & \max_{\alpha \in A} \Bigg\{ e^{-s}\gamma^p(s) + \\ & \left(\begin{bmatrix} r \\ R \end{bmatrix} \mathbf{X} + \begin{bmatrix} -1 \\ 1-\mu \end{bmatrix} \alpha + \begin{bmatrix} 1-\mu \\ -1 \end{bmatrix} \beta - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \gamma \right) \cdot \nabla J^* \Bigg\}. \end{aligned}$$

The HJB equation clearly serves as a necessary condition for optimality; that is the minimum objective function value $J^*(\mathbf{x}(t_f), t_f)$ must satisfy (5.10).

It is also true that if there is a objective function value $\hat{J}((\mathbf{x}(t_f), t_f)$ that satisfies (5.10), then $\hat{J}$ is the minimum objective function value, that is

$$\hat{J}((\mathbf{x}(t_f), t_f) = J^*(\mathbf{x}(t_f), t_f).$$

So the HJB equation also serves as a sufficient condition for optimality.

However, while the HJB equation serves as both necessary and sufficient condition for an optimal performance measure, it is usually quite difficult to solve analytically and so numerical techniques are often used to find its solution.

5.2 Stochastic Control

5.2.1 Restatement of the Problem

In the previous section, we derived the HJB equation for the deterministic optimal control problem, where the state of the system was defined by a set of ODEs. Sometimes, the state of a system is more accurately described by a set of stochastic differential equations (SDEs). This type of problem is known as the stochastic optimal control problem.

Given a probability space $(\Omega, \mathcal{F}, P)$ we wish to control the system described by

$$d\mathbf{x}(t) = a(\mathbf{x}(t), \mathbf{u}(t))dt + b(\mathbf{x}(t), \mathbf{u}(t))dB_t, \tag{5.14}$$

so as to minimise the objective function

$$J = \min_{\mathbf{u}\in U} \mathbb{E}_t \left\{ \int_t^{t_f} g(\mathbf{x}(t), \mathbf{u}(t), t)d\tau \right\}, \tag{5.15}$$

where $a(\mathbf{x}(t), \mathbf{u}(t))$ is the drift coefficient of the SDE (5.14) and $b(\mathbf{x}(t), \mathbf{u}(t))$ is its volatility coefficient. The Weiner process B_t can also be replaced with another stochastic process such as the Poisson process.

We note that, as with all SDEs, the control variable $\mathbf{u}$ must satisfy some additional properties for it to be admissible. $\mathbf{u}(t)$ is said to be admissible if

$$\begin{aligned} \mathbf{u}(t) &\in L^2_{\mathcal{F}}(0,T), \\ \text{where } L^2_{\mathcal{F}}(0,T) &:= \left\{ \mathbf{u}(t) \,\middle|\, \mathbb{E}\int_0^T |\mathbf{u}(t)|^2 ds < \infty \right\}. \end{aligned} \tag{5.16}$$

That is, $\mathbf{u}(t)$ is admissible if it is Lesbesgue square integrable on the filtration $\mathcal{F}_t$ of states of the system.

5.2.2 Analytic Methods of Solution for the Stochastic Case

If we follow the same method as in the deterministic case, we may be able to obtain a meaningful result.

$$\begin{aligned} J^*(\mathbf{x}(t), t) &= \min_{\mathbf{u}\in U} \mathbb{E}_t \left\{ \int_t^{t_f} g(\mathbf{x}(t), \mathbf{u}(t), t)d\tau \right\} \\ &= \min_{\mathbf{u}\in U} \mathbb{E}_t \left\{ \int_t^{t+\Delta t} g\, d\tau \right\} + \min_{\mathbf{u}\in U} \mathbb{E}_{t+\Delta t} \left\{ \int_{t+\Delta t}^{t_f} g\, d\tau \right\} \\ &= \min_{\mathbf{u}\in U} \mathbb{E}_t \left\{ \int_t^{t+\Delta t} g\, d\tau \right\} + J^*(\mathbf{x}(t+\Delta t), t+\Delta t). \end{aligned} \tag{5.17}$$

At this point in the deterministic derivation of the HJB equation, we expanded $J^*(\mathbf{x}(t),t)$ using a Taylor series. When we examine any Taylor series expansion we usually wish to include terms which are meaningful and ignore terms which will be negligible. Thus, in the deterministic case (if we will eventually use limits) we include first order terms such as Δt and ignore higher order terms such as $(\Delta t)^2$ (if Δt is very small, $(\Delta t)^2$ will be negligible).

However when our terms include stochastic process, we cannot simply ignore second order terms. Intuitively, the variance of any stochastic process is a second order term and by ignoring it we are enforcing the variance to be zero - thus returning the process to a deterministic setting.

So

$$\begin{aligned} J^*(\mathbf{x}(t+\Delta t),t+\Delta t) &= J^*(\mathbf{x}(t),t) + \frac{\partial J^*}{\partial \mathbf{x}}\Delta \mathbf{x} + \frac{\partial J^*}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 J^*}{\partial \mathbf{x}^2}(\Delta \mathbf{x})^2 \\ &\quad + \frac{\partial^2 J^*}{\partial \mathbf{x}\partial t}\Delta \mathbf{x}\Delta t + + \frac{1}{2}\frac{\partial^2 J^*}{\partial t^2}(\Delta t)^2 + o(\Delta t) \end{aligned}$$

Recalling that

$$\Delta \mathbf{x} = a(\mathbf{x}(t),\mathbf{u}(t))\Delta t + b(\mathbf{x}(t),\mathbf{u}(t))\Delta B_t + o(\Delta t),$$

and that we treat

$$(\Delta t)^2 = 0, \quad \Delta t \Delta B_t = 0 \ \text{ and } \ (\Delta B_t)^2 = \Delta t,$$

then (5.17) becomes

$$\begin{aligned} J^*(\mathbf{x}(t),t) &= \min_{\mathbf{u}\in U} \mathbb{E}_t \left\{ \int_t^{t+\Delta t} g \, d\tau + J^*(\mathbf{x}(t),t) \right. \\ &\quad + \left[J^*_{\mathbf{x}} + J^*_{\mathbf{x}t}\right]\left[a(\mathbf{x},\mathbf{u})\Delta t + b(\mathbf{x},\mathbf{u})\Delta B_t\right] + J^*_t \Delta t \\ &\quad \left. + \frac{1}{2} J^*_{\mathbf{xx}}\left[a(\mathbf{x},\mathbf{u})\Delta t + b(\mathbf{x},\mathbf{u})\Delta B_t\right]^2 + o(\Delta t) \right\} \\ &= \min_{\mathbf{u}\in U} \mathbb{E}_t \left\{ \int_t^{t+\Delta t} g \, d\tau + J^*(\mathbf{x}(t),t) \right. \\ &\quad + \left[J^*_{\mathbf{x}} a(\mathbf{x},\mathbf{u}) + J^*_t + \frac{1}{2} + J^*_{\mathbf{xx}} b^2(\mathbf{x},\mathbf{u}) \right] \Delta t \\ &\quad \left. + J^*_{\mathbf{x}} b(\mathbf{x},\mathbf{u})\Delta B_t + o(\Delta t) \right\}. \end{aligned} \tag{5.18}$$

Now as $\mathbb{E}(\Delta B_t) = 0$,

$$\begin{aligned} J^*(\mathbf{x}(t),t) &= \min_{\mathbf{u}\in U}\Bigg[\mathbb{E}_t\left\{\int_t^{t+\Delta t} g\ d\tau + J^*(\mathbf{x}(t),t)\right. \\ &\quad \left. + \left(J^*_{\mathbf{x}}a(\mathbf{x},\mathbf{u}) + J^*_t + \frac{1}{2} + J^*_{\mathbf{xx}}b^2(\mathbf{x},\mathbf{u})\right)\Delta t\right\} \\ &\quad + \mathbb{E}\left(o(\Delta t)\right)\Bigg]. \end{aligned}$$

If we now take the expectation operator inside the brackets, divide by Δt and take the limit as $\Delta t \to 0$, we obtain the stochastic HJB equation:

$$0 = J^*_t + \min_{\mathbf{u}\in U}\left\{g(\mathbf{x},\mathbf{u}) + J^*_{\mathbf{x}}a(\mathbf{x},\mathbf{u}) + \frac{1}{2}J^*_{\mathbf{xx}}b^2(\mathbf{x},\mathbf{u})\right\}. \tag{5.19}$$

The stochastic HJB equation is simply the deterministic HJB equation with a $J^*_{\mathbf{xx}}$ term added.

The astute reader may also be able to recognise another familiar construct in Equation (5.19), the differential generator $\mathcal{A}$. Recall from Chapter 4, the diffusion generator for the Itô diffusion given by

$$d\mathbf{X}_t = \mathbf{a}(t,X_t)dt + \mathbf{b}(t,X_t)d\mathbf{B}_t,$$

is given by

$$\mathcal{A}f(t,x) = \frac{\partial f}{\partial t} + \sum_i a_i\frac{\partial f}{\partial x} + \frac{1}{2}\sum_{i,j}(bb^T)_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}, \quad f \in C^2(\mathbb{R}).$$

However, when the objective function involves an integral, the characteristic operator becomes (see Øksendal (1998), page 212),

$$\mathcal{A}J(t,\mathbf{x}) = \mathcal{A}f(t,x) + g(\mathbf{x},\mathbf{u}).$$

So Equation (5.19) becomes

$$\boxed{\min_{\mathbf{u}\in U}\{\mathcal{A}J(t,\mathbf{x})\} = 0,} \tag{5.20}$$

with the boundary conditions being determined by the individual problem.

Using (5.20) instead of (5.19) gives a method for solving optimal stochastic control problems where the state equations are described by an Itô diffusion.

5.2.3 Example

Example 5.2.1 (Optimal Porftolio Selection and Consumption) *From Kohn (1999). This is the simplest of a class of problems solved by Robert Merton in his paper "Optimal consumption and portfolio rules in a continuous-time model", Merton (1971). Consider a world with one risky asset and one risk-free asset. The risk free asset grows at a constant risk-free rate r, that is its price per share satisfies $dp_1/dt = p_1 r$. The risky asset executes a geometric Brownian motion with constant drift $\mu > r$ and volatility σ, that is its price per share solves the SDE $dp_2 = \mu p_2 dt + \sigma p_2 dW$.*

The control problem is this: an investor starts with initial wealth x at time t. His control variables are

$$\begin{aligned} \alpha_1(s) &= \text{fraction of total wealth invested in risky asset at time } s \\ \alpha_2(s) &= \text{rate of consumption at time } s. \end{aligned}$$

It is natural to restrict these controls by $0 \leq \alpha_1(s) \leq 1$ and $\alpha_2(s) \geq 0$. We ignore transaction costs. The state is the investors total wealth y as a function of time; it solves

$$dy = (1-\alpha_1) y r dt + \alpha_1 y(\mu dt + \sigma dW) - \alpha_2 dt$$

as long as $y(s) > 0$. We denote by τ the first time $y(s) = 0$ if this occurs before time T, or $\tau = T$ (a fixed time horizon) otherwise. The investor seeks to maximise the discounted total utility of his consumption. We therefore consider the value function

$$u(x,t) = \max_{\alpha_1, \alpha_2} \mathbb{E}_{y(t)} \int_t^\tau e^{-\rho s} h[\alpha_2(s)] ds$$

where $h(\cdot)$ is a specified utility function (monotone increasing and concave, with $h(0) = 0$). We shall specialize below to $h(\alpha_2) = \alpha_2^\gamma$ with $0 < \gamma < 1$.

By (5.20), the HJB equation is then

$$\max_{\alpha_1, \alpha_2} \left\{ u_t + e^{-\rho s} h(\alpha_2) + \left[(1-\alpha_1) x r + \alpha_1 x \mu - \alpha_2\right] u_x + \frac{1}{2} x^2 \alpha_1^2 \sigma^2 u_{xx} \right\} = 0. \tag{5.21}$$

Solution *- Let us assume that $u_x > 0$ and $u_{xx} < 0$ (reflecting the concavity of the utility function). Then optimal α_1 (ignoring the constraint $0 \leq \alpha_1 \leq 1$) is*

$$\alpha_1^* = -\frac{(\mu - r) u_x}{\sigma^2 x u_{xx}},$$

which is positive. We proceed, postponing till later the verification that $\alpha_1^* \leq 1$. *The optimal* α_2 *satisfies*

$$h'(\alpha_2^*) = e^{\rho t} u_x;$$

we can be sure this α_2^* *is positive by assuming that* $h'(0) = \infty$.

When $h(\alpha_2) = \alpha_2^\gamma$ *with* $0 < \gamma < 1$ *we can get an explicit solution. Indeed, let us look for a solution of the form*

$$u(x,t) = g(t)x^\gamma.$$

The associated α_1^* *and* α_2^* *are*

$$\begin{aligned} \alpha_1^* &= \frac{(\mu - r)}{\sigma^2(1-\gamma)}, \\ \alpha_2^* &= \left[e^{\rho t} g(t)\right]^{1/(\gamma-1)} x. \end{aligned}$$

We assume henceforth that $\mu - r < \sigma^2(1-\gamma)$ *so that* $\alpha_1^* < 1$. *Substituting these values into the HJB equation gives, after some analysis,*

$$\frac{dg}{dt} + \nu\gamma g + (1-\gamma) g (e^{\rho t} g)^{1/(\gamma-1)} = 0$$

where

$$\nu = r + \frac{(\mu - r)^2}{2\sigma^2(1-\gamma)}.$$

We must solve this with $g(T) = 0$. *The substitution* $f = (e^{\rho t} g)^{1/(\gamma-1)}$ *leads to a linear differential equation for* $f(s)$. *It is readily solved to give*

$$g(t) = e^{-\rho t} \left[\frac{1-\gamma}{\rho - \nu\gamma} \left(1 - e^{\frac{(\rho - \nu\gamma)(T-t)}{1-\gamma}} \right) \right]^{1-\gamma},$$

which is only meaningful if $\rho > \nu\gamma$.

Thus we have solved the HJB equation.

5.3 Numerical Implementation

One method, developed by Harold Kushner, to numerically solve the SOC problem uses a discrete-time, discrete-state Markov chain to simulate the diffusion process. First we choose a Markov Chain whose local properties closely resemble those of the diffusion process we are trying to control, then

by forming a discrete approximation to the objective function we can develop a discretization of the HJB equation to solve numerically.

While it is important for the Markov chain to be locally consistent with the diffusion process, any numerical method we use must also converge as the discretization grid becomes finer. The method presented in the next few sections has been shown to converge in such a manner, however the proofs will not be given here - they can be found in Kushner & Dupuis (1992).

Local Consistency

Let $\Psi_n^h, n < \infty$ denote a controlled discrete parameter Markov chain on a discrete state space $S_h \in \mathbb{R}^r$ with transition probabilities denoted by $p^h(x, y|u)$. Here h is used to denote the stepsize although another discretization parameter, Δt_h will also be needed. The second discretization parameter will need to be related to the first by

$$(h \to 0) \quad \Rightarrow \quad \sup \Delta t_h \to 0.$$

The control variable is denoted by u and the control action is denoted by α. We can define the change in state of the Markov chain as $\Delta\Psi_n^h := \Psi_{n+1}^h - \Psi_n^h$.

As we will be using a discrete-time Markov chain to approximate the diffusion process, the Markov chain properties must be locally consistent with the properties of the diffusion process, that is that the mean and variance of the chain are the same as the mean and variance of the diffusion process

$$d\mathbf{X}_t = \mathbf{a}(t, X_t)dt + \mathbf{b}(t, X_t)d\mathbf{B}_t.$$

That is

$$\mathbb{E}(\Delta\Psi_n^h) = \mathbf{a}(t, X_t)\Delta t_h + o(h),$$

and

$$\mathbb{E}\left[\Delta\Psi_n^h - \mathbb{E}(\Delta\Psi_n^h)\right]\left[\Delta\Psi_n^h - \mathbb{E}(\Delta\Psi_n^h)\right]^T = \mathbf{b}(t, X_t)\mathbf{b}^T(t, X_t)\Delta t_h + o(h).$$

For a control to be *admissible* under such an approximation, the Markov property must hold for the chain, that is

$$P\{\Psi_{n+1}^h = y|\Psi_i^h, u_i^h, i \leq n\} = p^h(\Psi_n^h, y|u_n^h).$$

The difficulty with developing such a Markov chain is in determining the transition probabilities that ensure the local properties hold. However, this can be done using a finite difference approximation. While using finite diferences is good for introducing the elementary concepts of numerical stochastic

control, problem-specific alterations are often required and the reader is encouraged to investigate a good text (such as Kushner & Dupuis (1992)) for more details.

Suppose we wish to control a one-dimensional diffusion defined by

$$dx = a(x,u)dt + b(x,u)dW_t,$$

using a feedback control $u(\cdot) \in \mathcal{U}$, so as to optimize the objective function

$$W(x,u) = \mathbb{E}_x^u \left[\int_0^\tau k(x(s), u(x(s)))ds + g(x(\tau)) \right]. \tag{5.22}$$

If $W(\cdot)$ is smooth enough then $W(\cdot)$ satisfies Itô's Lemma:

$$\begin{aligned} \mathcal{L}W_x(x) + k(x) &= 0, \quad x \in (0,B), \\ W(0) = W(B) &= 0, \end{aligned} \tag{5.23}$$

where $\mathcal{L}$ is the differential operator as defined in Chapter 4[1].

The transition probabilities can be obtained by trying to solve equation (5.23) numerically using finite differences. The finite difference aproximation for the second derivative is, as seen before,

$$f_{xx}(x) \approx \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}.$$

However, if the finite difference approximation is to have an interpretation in terms of a Markov chain (ie the transition probabilities are all non-negative), then the first derivative approximation must be

$$f_x(x) \approx \frac{f(x+h) - f(x)}{h} f^+(x) - \frac{f(x) - f(x-h)}{h} f^-(x),$$

where $f^+(x) = \max[f(x), 0]$, and $f^-(x) = \min[-f(x), 0]$ [2].

When both these terms are substituted into (5.23), we obtain (after simplification)

$$\begin{aligned} W^h(x) = \quad & \frac{\sigma^2(x)/2 + hb^+(x)}{\sigma^2(x) + h|b(x)|} W^h(x+h) + \frac{\sigma^2(x)/2 + hb^-(x)}{\sigma^2(x) + h|b(x)|} W^h(x-h) \\ & + k(x) \frac{h^2}{\sigma^2(x) + h|b(x)|}. \end{aligned} \tag{5.24}$$

[1] It is now more obvious that the optimal control would be the infimum over all admissible controls of $\mathcal{L}J = 0$

[2] Note that $f^+(x) + f^-(x) = |f(x)|$

It is (5.24) that delivers the transition probabilities. By rewriting (5.24) as

$$W^h(x) = p^h(x, x+h)W^h(x+h) + p^h(x, x-h)W^h(x-h) + k(x)\Delta t^h(x), \tag{5.25}$$

and setting $p^h(x, y) = 0$ for $y \neq x \pm h$, we have determined a complete set of transition probabilities which, it can be shown, give the approximating Markov chain the 'local consistency' properties discussed above.

Objective Function

Once the approximating Markov chain is defined, an approximation for the objective function can be derived. If (5.25) has a unique solution, then it is the cost associated with the controlled chain, that is

$$W^h(x, u) = \mathbb{E}_x^u \left[\sum_0^{N_h - 1} k(\Psi_n^h) \Delta t^h(\Psi_n^h, u(\Psi_n^h)) \right].$$

As already discussed, the optimal control is the control which optimises the performance measure. In the discrete approximation, following (5.20),

$$\begin{aligned} V^h(x) &= \min_{\alpha \in \mathcal{U}} W^h(x, u) \\ &= \min_{\alpha \in \mathcal{U}} \mathbb{E}_x^u \left[\sum_0^{N_h - 1} k(\Psi_n^h) \Delta t^h(\Psi_n^h, u(\Psi_n^h)) \right] \\ &= \min_{\alpha \in \mathcal{U}} \left[\sum_y p^h(x, y|\alpha) V^h(y) + k(x, \alpha) \Delta t^h(x, \alpha) \right]. \end{aligned} \tag{5.26}$$

Once the dynamic programming equation, (5.26), is solved, the SOC problem is solved.

Solving the Dynamic Programming Equation

Once the transition probabilities have been derived and the dynamic programming equations have been written, the final step is to solve the DPEs. That is, we need a numerical procedure for solving (5.26), or in vector form

$$W^h(u) = R^h(u)W^h(u) + C^h(u) \tag{5.27}$$

$$V^h = \min_{u(x) \in \mathcal{U}} \left[R^h(u)V^h(u) + C^h(u) \right], \tag{5.28}$$

where $R^h(u)$ is a matrix of probabilities.

Two main methods exist for the solution of the DPEs. The first, called *'approximation in policy space'*, attempts to iteratively define the optimal control policy.

Theorem 5.3.1 (Approximation in Policy Space) *Assume that $r(x, y|\alpha)$ and $C(x, \alpha)$ are continuous functions of α for each x and y in the state space S. Also assume that there is at least one admissible feedback control $u_0(\cdot)$ such that $R(u_0)$ is a contraction, and the infima of the costs over all admissible controls is bounded from below and that $R(u)$ is a contraction for any feedback control $u(\cdot)$ for which the associated cost is bounded.*

Then there is a unique solution to (5.28)*, and it is the infimum of the cost functions over all time independant feedback controls. Let $u_0(\cdot)$ be an admissible feesback control such that the cost $W(u_0)$ is bounded. For $n \geq 1$, define the sequence of feedback controls $u_n(\cdot)$ and costs $W(u_n)$ recursively by* (5.27) *together with the formula*

$$u_{n+1}(x) = \arg\min_{\alpha \in \mathcal{U}} \left[\sum_y r(x, y|\alpha) W(y, u_n) + C(x, \alpha) \right].$$

Then $W(u_n) \to V$.

Under the initial condition that, if the cost associated with the use of the feedback controls $u^1(\cdot), \ldots, u^n(\cdot), \ldots$ in sequence, is bounded, then

$$R(u^1) \ldots R(u^n) \xrightarrow{n} 0,$$

then V is the infimum of the costs over all admissible control sequences.

The second method, called *'approximation in value space'*, can be interpreted as a fixed point iteration method.

Theorem 5.3.2 (Approximation in Value Space - The Jacobi Iteration) *Let $u(\cdot)$ be an admissible feedback control such that $R(u)$ is a contraction. Then for any initial vector W_0, the sequence W_n defined by*

$$W_{n+1}(x, u) = \sum_y r(x, y|u(x)) W_n(y, u) + C(x, u(x))$$

converges to $W(u)$, the unique solution to (5.27)*. Assume that similar conditions to those in Theorem 5.3.1 hold. Then for any vector V_0, the sequence recursively defined by*

$$V_{n+1} = \min_{u(x) \in \mathcal{U}} \left[R(u) V_n + C(u) \right] \tag{5.29}$$

converges to V*, the unique solution to* (5.28)*. In detail,* (5.29) *is*

$$V_{n+1} = \min_{\alpha \in \mathcal{U}} \left[\sum_{y} r(x, y|\alpha) V_n(y) + C(x, \alpha) \right].$$

V_n *is the minimal cost for an* $n-$*step problem with terminal cost vector* V_0*.*

It must be pointed out that the *'approximation in policy space'* method and the *'approximation in value space'* method can be combined in a 'predictor-corrector' style. More details about combining these methods, as well as the proofs and error bounds for Theorems 5.3.1 and 5.3.2 can be found in Kushner & Dupuis (1992).

Appendix A

Mathematical Preliminaries

A.1 Probability Theory

The purpose of this introductory chapter is to give a brief overview of the theoretical background. To fully appreciate this report, the reader is advised to persist with this chapter until it is fully understood. For readers who are familiar with the measure theoretic presentation of probability, it is advised to briefly peruse this chapter to become refamiliarised with the relevant topics.

A.1.1 Introductory Measure Theory

In a random experiment, the outcome is not known in advance. The set of all possible outcomes is denoted by Ω. An *event*, A, is any set of possible outcomes - so that A is a subset of Ω. Sometimes we will be interested in when one of two possible events occurs (union) and other times we will be interested in when both of these two possible events occur (intersection).

Sometimes it is useful to identify all the events of interest pertaining to a particular random experiment. Thus, we usually denote by $\mathcal{F}$ the family of all **events of interest**. This family, known as a σ-algebra (pronounced sigma algebra), has the following definition.

Definition A.1.1 *A* σ-algebra over Ω *is a collection of subsets of* Ω *which satisfies*
- $\Omega \in \mathcal{F}$,
- *if* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$, *and*
- *if* $A_1, A_2, \cdots \in \mathcal{F}$ *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ *is known as a measurable space.*

We need to introduce a definition for random processes through time. This concept is embodied in *filtrations*.

Definition A.1.2 *A* filtration *(on $(\Omega, \mathcal{F})$) is a family of sigma algebras $\mathcal{M} = \left\{\mathcal{M}_t^H\right\}_{t\geq 0}$, $\mathcal{M}_t \subset \mathcal{F}$ such that*

$$0 \leq s < t \Rightarrow \mathcal{M}_s \subset \mathcal{M}_t,$$

that is $\mathcal{M}_t$ is increasing.

So for our purposes, we can think of a filtration as the 'history' of a process. Further, if a function is 'contained' in the history up until time t, then the function is said to be *adapted* to the filtration $\mathcal{F}_t$. For example the function $h(t,\omega) = B_{\frac{t}{2}}(\omega)$ is $\mathcal{F}_t$-adapted, while $h(t,\omega) = B_{2t}(\omega)$ is not $\mathcal{F}_t$-adapted.

Now, in order to assign probabilities, we wish to define a function $P : \mathcal{F} \to [0,1]$, such that $P(A), A \in \mathcal{F}$ can be interpreted as "the probability that A occurs". For such a concept to exist it must at least satisfy the properties of probability that we are familiar with.

Definition A.1.3 *A* probability measure, *P,* on *$(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to \mathbb{R}$ satisfying*
- *$P(A) \geq 0$ if $A \in \mathcal{F}$,*
- *$P(\Omega) = 1$, and*
- *P is countably additive.*

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

A.1.2 Change of Measure

We cannot refer to random processes without referring to measure. It is in fact slightly ambiguous to refer to a process without specifying a probability measure for the process. We can intuitively think of a probability measure as something that assigns relative likelihoods to the possible sample paths that a process may take. A probability measure, $\mathbb{P}$ on (Ω, X) is a mapping of a sample space X into the interval $[0,1]$, i.e.

$$\mathbb{P} : X \to [0,1]$$

subject to

(a) $P(A) \geq 0$, if $A \in X$,

(b) $P(\Omega) = 1$,

(c) P is countably additive, i.e.

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

if $A_1, A_2, \cdots \in X$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint.

Before we state the general Radon-Nikodym result, which is used in conditional expectation and conditional probability, we first need to introduce the definition of absolute continuity of measures.

Definition A.1.4 *The probability measure $\mathbb{Q}$ is called absolutely continuous with respect to $\mathbb{P}$, defined as $\mathbb{Q} << \mathbb{P}$, if every event with zero probability under $\mathbb{P}$ also has zero probability under $\mathbb{Q}$.*

Note that if $\mathbb{Q} << \mathbb{P}$ and $\mathbb{P} << \mathbb{Q}$, that is $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ and $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$, then $\mathbb{P}$ and $\mathbb{Q}$ satisfy the definition of *equivalent measures*.

Theorem A.1.1 (Radon-Nikodym) *Let $\mathbb{Q} << \mathbb{P}$, then there exists a random variable $\Lambda \geq 0$ such that $\mathbb{E}_{\mathbb{P}}(\Lambda) = 1$, and*

$$\mathbb{Q}(A) = \int_A \Lambda d\mathbb{P},$$

for any measurable set A. Λ is $\mathbb{P}$-a.s. unique. Note that the converse also holds.

In other words, if there exists two measures, $\mathbb{Q}$ and $\mathbb{P}$ such that $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ and there exists a random variable Λ such that the expected value under the measure $\mathbb{P}$ is one, then

$$\mathbb{Q}(A) = \int_A \Lambda d\mathbb{P}.$$

To illustrate the ideas of a change of measure, we will given an example from Mark Thompson's MS479 lecture notes.

Example A.1.1 *Let $\Omega = \{HH, HT, TH, TT\}$, be the set of coin toss sequences of length two. Let $\mathbb{P}$ correspond to the probability $\frac{1}{3}$ for H and $\frac{2}{3}$ for T, and let $\mathbb{Q}$ correspond to the probability $\frac{1}{2}$ for H and $\frac{1}{2}$ for T. Then the Radon-Nikodym derivative is*

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(HH) = \frac{9}{4}, \ \frac{d\mathbb{Q}}{d\mathbb{P}}(HT) = \frac{9}{8}, \ \frac{d\mathbb{Q}}{d\mathbb{P}}(TH) = \frac{9}{8}, \ \frac{d\mathbb{Q}}{d\mathbb{P}}(TT) = \frac{9}{16}.$$

In Chapter 2, we introduced a theorem due to Cameron, Martin and Girsanov which gives us the conditions necessary for the new measure $\mathbb{Q}$ to be a martingale measure and the form of the Radon-Nikodym derivative under this change of measure.

A.1.3 Expectation in a Measurable Space

Given a measure $\mathbb{P}$ as defined in section A.1.2 we can now define the expected value of a random variable X under this measure. If $\int_\Omega |X| d\mathbb{P} < \infty$, then

$$\mathbb{E}_{\mathbb{P}}(X) := \int_\Omega X d\mathbb{P} \tag{A.1}$$

is called the *expectation* of X with respect to $\mathbb{P}$. Throughout the rest of this section, for notational convenience, we will use $\mathbb{E}(X)$ instead of $\mathbb{E}_{\mathbb{P}}(X)$, unless explicitly stated otherwise. We can think of the expected value of X as the weighted average of the random variable X. If we have a function $g : \mathbb{R} \to \mathbb{R}$, then

$$\mathbb{E}(g(X)) = \int_\Omega g(X) d\mathbb{P}. \tag{A.2}$$

To conclude our very brief introduction to expectation in a measurable space, we will briefly describe some of the basic properties of expectation.

(a) If X, Y are independent, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

(b) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, if $a, b \in \mathbb{R}$.

(c) If $X(\omega) \geq Y(\omega)$, $\forall \omega \in \Omega$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

A.1.4 Martingales

We now introduce the reader to the class of random processes known as martingales. Martingales are a broad class of random variables which rely on very few, but very powerful assumptions.

Definition A.1.5 (A Martingale) *Let $\mathcal{F}$ be a filtration of the probability space $(\Omega, \mathcal{F}, P)$ and let Y be a sequence of random variables adapted to $\mathcal{F}$. We call the pair $(Y, \mathcal{F}) = \{(Y_n, \mathcal{F}_n); n \geq 0\}$ a **martingale** if, for all $n \geq 0$*

$$\mathbb{E}_P |Y_n| < \infty, \tag{A.3}$$

$$\mathbb{E}_P(Y_{n+1} \mid \mathcal{F}_n) = Y_n. \tag{A.4}$$

So a martingale is simply a random variable whose expected value in the next step, given the history up to the current step, is the current value. For example, if a stock price process is known to be a martingale then the expected value of the price in the next time step, given the price history, is just the current price of the stock.

This definition can be extended to allow for an inequality in equation (A.4), giving the concepts of sub-martingale and super-martingale.

Definition A.1.6 (Semi-Martingales) *Let $\mathcal{F}$ be a filtration of the probability space $(\Omega, \mathcal{F}, P)$ and let Y be a sequence of random variables adapted to $\mathcal{F}$. We call the the pair $(Y, \mathcal{F})$ a* ***sub-martingale*** *if, for all $n \geq 0$,*

$$\begin{aligned} \mathbb{E}_P(Y_n^+) &< \infty, & \text{(A.5)} \\ \mathbb{E}_P(Y_{n+1} \mid \mathcal{F}_n) &\geq Y_n, & \text{(A.6)} \end{aligned}$$

or a ***super-martingale*** *if, for all $n \geq 0$,*

$$\begin{aligned} \mathbb{E}_P(Y_n^-) &< \infty, & \text{(A.7)} \\ \mathbb{E}_P(Y_{n+1} \mid \mathcal{F}_n) &\leq Y_n, & \text{(A.8)} \end{aligned}$$

where $Y^+ = \max\{0, Y\}$, and $Y^- = -\min\{0, Y\}$.

Clearly for $(Y, \mathcal{F})$ to be a martingale it must also be both a sub and super-martingale. Also if $(Y, \mathcal{F})$ is a sub-martingale, then $(-Y, \mathcal{F})$ is a super-martingale. This allows us to appreciate a famous result on the decomposition of any sub-martingale by Doob.

Theorem A.1.2 (Doob Decomposition Theorem) *A sub-martingale $(Y, \mathcal{F})$ with finite mean may be expressed in the form*

$$Y_n = M_n + S_n, \qquad \text{(A.9)}$$

where $(M, \mathcal{F})$ is a martingale, and $(S, \mathcal{F})$ is an increasing predictable process. This decomposition is unique.

If the stock price process that we are modelling is a martingale, then the analysis becomes trivial. This analysis becomes trivial in the sense that if we have a stock price model, S_t and a model for the time value of money B_t and if the discounted stock price process $B_t^{-1}S_t$ is a martingale, then we can apply the risk neutral valuation formula

$$\mathbb{E}_{\mathbb{Q}}\left(B_T^{-1}X\right)$$

to price relatively arbitrary claims with payoff X. However if the stock process is not a martingale we might be able to change the probability measure P, of the process so that it becomes a martingale. We will see in the next section the conditions required to perform this change of measure.

Appendix B

Markov Chains

The random walk and branching process are two simple examples of random or stochastic processes. Typically a random process is a family of random variables $\{X_t; t \in T\}$ which may be either a discrete-time, $T = \{0, 1, 2, \dots\}$, or continuous-time, $T = \mathbb{R}$ or $\mathbb{R}^+$, process. Both the random walk and branching process share an interesting property. That is, conditional upon their current value, their future value does not depend upon their previous values. In this section we will focus our attention on the general theory of process with this property.

B.1 Basic description

Definition B.1.1 *The process X is a Markov chain if it satisfies the Markov condition.*

$$P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

for all $n \geq 1$ and all $x_0, x_1, \dots, x_n \in S$.

The path that the process follows is described by its transition probabilities $P(X_{n+1} = j | X_n = i)$. We shall look at the case where the evolution of the chain does not depend upon n.

Definition B.1.2 *X is called homogeneous if*

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)$$

for all n, i, j.

The transition matrix $P = (p_{ij})$ is an $|S| \times |S|$ matrix of transition probabilities

$$p_{ij} = P(X_{n+1} = j | X_n = i).$$

Theorem B.1.1 *P is a stochastic matrix, i.e.*

(a) P has non-negative entries, $p_{ij} \geq 0$

(b) P has row sums equal to one, $\sum_j p_{ij} = 1$.

We can define the n-step transition matrix $P_n = (p_{ij}(n))$ as the matrix of n-step transition probabilities

$$p_{ij}(n) = P(X_{m+n} = j | X_m = i).$$

An important theorem is the Chapman-Kolmogorov equations which tell us how the long term behaviour depends on the short term behaviour and how X_n depends on the initial variable X_0.

Theorem B.1.2 (Chapman-Kolmogorov equations)

$$p_{ij}(m+n) = \sum_k p_{ik}(m) p_{kj}(n).$$

Hence $P_{m+n} = P_m P_n$ and so $P_n = P^n$.

Proof:

$$\begin{aligned}
p_{ij}(m+n) &= P(X_{m+n} = j | X_0 = i) \\
&= \sum_k P(X_{m+n} = j, X_m = k | X_0 = i) \\
&= \sum_k P(X_{m+n} = j | X_m = k, X_0 = i) P(X_m = k | X_0 = i) \\
&= \sum_k P(X_{m+n} = j | X_m = k) P(X_m = k | X_0 = i) \\
&= \sum_k p_{ik}(m) p_{kj}(n).
\end{aligned}$$

B.2 Classification of states and chains

Definition B.2.1 *State i is called persistent (or recurrent) if*

$$P(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1.$$

If this probability is strictly less than 1, state i is called transient.

Corollary B.2.1 *(a) j is persistent if $\sum_n p_{jj}(n) = \infty$, and if this holds then $\sum_n p_{ij}(n) = \infty$ for all i such that there is a positive probability that the chain ever visits state j starting from i.*

(b) j is transient if $\sum_n p_{jj}(n) < \infty$, and if this holds then $\sum_n p_{ij}(n) < \infty$ for all i.

A result that follows immediately from the above corollary is that if j is transient then $p_{ij}(n) \to 0$ as $n \to \infty$ for all i.

From the above it is clear that each state is either persistent or transient and the number of times, $N(i)$, which the chain visits its starting point i satisfies

$$P(N(i) = \infty) = \begin{cases} 1, & \text{if } i \text{ is persistent} \\ 0, & \text{if } i \text{ is transient.} \end{cases}$$

Let

$$T_j = \min\{n \geq 1; X_n = j\}$$

be the time of the first visit to j, with the convention that $T_j = \infty$ if this visit never occurs. Then $P(T_i = \infty | X_0 = i) > 0$ if and only if i is transient, and $\mathbb{E}(T_i | X_0 = i) = \infty$.

Definition B.2.2 *The persistent state i is called*

(a) null if $\mu_i = \infty$

(b) positive (or non-null) if $\mu_i < \infty$.

There is a simple criterion for nullity in terms of the transition probabilities.

Theorem B.2.1 *A persistent state is null if and only if $p_{ii}(n) \to 0$ as $n \to 0$. If this holds, then $p_{ji}(n) \to 0$ for all j.*

We will also give the definitions of the period of a state and ergodicity.

Definition B.2.3 *The period $d(i)$ of a state i is defined by*

$$d(i) = \gcd\{n; p_{ii}(n) > 0\}.$$

We call i periodic if $d(i) > 0$ and aperiodic if $d(i) = 1$. That is to say, $p_{ii}(n) = 0$ unless n is a multiple of $d(i)$.

Definition B.2.4 *A state is called ergodic if it is positive recurrent and aperiodic.*

We will now consider the way in which the states of a Markov chain are related to each other.

Definition B.2.5 *For a Markov chain X*

(a) i communicates with j, $i \to j$, if the chain may ever visit state j with positive probability, starting from i. That is, $p_{ij} > 0$ for some $n \geq 0$.

(b) i and j intercommunicate if $i \to j$ and $j \to i$, written as $i \leftrightarrow j$.

The state space S may be partitioned into the equivalence classes of $\leftrightarrow$. Within each class, all states are of the same type.

Theorem B.2.2 *If $i \leftrightarrow j$*

(a) i and j have the same period

(b) i is transient iff j is transient

(c) i is null persistent iff j is null persistent.

B.3 Stationary distributions

Definition B.3.1 *The vector π is called a stationary distribution of the chain if π has entries $(\pi_j; j \in S)$ such that*

(a) $\pi_j \geq 0$ for all j, and $\sum_j \pi_j = 1$

(b) $\pi = \pi \mathbf{P}$, which is to say that $\pi_j = \sum_i \pi_i p_{ij}$ for all j.

Theorem B.3.1 *An irreducible chain has a stationary distribution π if and only if all the states are positive recurrent. If so, π is the unique stationary distribution and is given by $\pi_i = \mu_i^{-1}$ for each $i \in S$ where μ_i is the mean recurrence time of i.*

Theorem B.3.2 *For an irreducible aperiodic chain, we have that*

$$p_{ij}(n) \to \frac{1}{\mu_j}$$

as $n \to \infty$ *for all* i *and* j.

Theorem B.3.3 *For any aperiodic state* j *of a Markov chain,*

$$p_{jj}(n) \to \frac{1}{\mu_j}$$

as $n \to \infty$. *Furthermore, if* i *is any other state then*

$$p_{ij}(n) \to \frac{1}{\mu_j} f_{ij}$$

as $n \to \infty$.

Appendix C

Numerical Methods for PDEs

Partial differential equations (PDEs) form the basis of mathematical models for many real world phenomena and have more recently been applied to economics and finance. Often it is necessary to approximate the solution of these PDEs numerically and in this chapter we shall be concerned with the numerical solution of parabolic PDEs. The problem of solving a hyperbolic or eliptic PDE is much more difficult and numerical methods to solve them can be found in (Morton & Mayers 1994) and (Ames 1992)

C.1 Classification of Equations

Consider the linear second-order PDE of the form

$$a\frac{\partial^2 z}{\partial x^2} + b\frac{\partial^2 z}{\partial x \partial y} + c\frac{\partial^2 z}{\partial y^2} + d\frac{\partial z}{\partial x} + e\frac{\partial z}{\partial y} + fz = g, \tag{C.1}$$

where the functions $a, \ldots, g$ depend only on x and y. It turns out that the terms involving the second order derivatives of z are of greatest significance. If we look at the discriminant of the above equation, we say that the equation is *hyperbolic* if $b^2 - 4ac > 0$, *elliptic* if $b^2 - 4ac < 0$ and *parabolic* when $b^2 - 4ac = 0$.

C.2 Method of Lines

The method of lines (MOL) reduces a partial differential equation into a system of ordinary differential equations by descretizing in the space-like independent variable. If the original PDE is an initial value problem (IVP), the resulting system of ODEs forms an IVP. The same is also true for boundary value problems.

Consider the uniform grid $x_i = ih$, $h = 1/N$ where we want to compute our approximation $u_i(t)$ to $z(x_i, t)$. Recalling elementary difference approximations we can approximate the derivatives of z by

$$\begin{aligned} z_x(x_{i-1/2}, t) &\approx (z(x_i, t) - z(x_{i-1}, t))/h, && \text{(C.2)} \\ z_{xx}(x_i, t) &\approx (z_x(x_{i+1/2}, t) - z_x(x_{i-1/2}, t))/h \\ &\approx (z(x_{i+1}, t) - 2z(x_i, t) + z(x_{i-1}, t))/h^2. && \text{(C.3)} \end{aligned}$$

Example C.2.1 *Consider*

$$u_t + uu_x = u_{xx}, \quad 0 < x < 1,\ 0 < t, \tag{C.4}$$

with boundary conditions

$$u(0, t) = u(1, t) = 0, \quad u(x, 0) = \sin \pi x. \tag{C.5}$$

Using our central difference scheme, our approximation $u_i(t)$ should satisfy

$$\begin{aligned} \partial u_i(t) &= -\frac{u_i}{2h}(u_{i+1}(t) - u_{i-1}(t)) + \\ &\quad \frac{1}{h^2}(u_{i+1}(t) - 2u_i(t) + u_{i-1}(t)), \quad i = 1, \ldots, N-1, \\ u_i(0) &= \sin \pi i h, \quad i = 1, \ldots, N-1, \\ u_0(t) &= u_N(t) = 0. \end{aligned}$$

The solution to the above problem is shown in Figure C.1.

The Matlab code to solve the above problem is

```
N = 100;
x = (1:N-1)'/N;
h = 1/N;

V = sin(pi*x);
t = 0:0.01:1;
[ts,us] = ode15s('F',t,V);
surf(x,ts,us);
xlabel('x');
ylabel('time')
title('u_t + uu_x = u_{xx}');
```

where the equation to be solved is given by

```
function du = F(t,u)

N = length(u) + 1;
iN = (2:N)';
h = 1/N;

uu = [0; u; 0];
du = -uu(iN)/(2*h).*(uu(iN+1)-uu(iN-1)) +
(uu(iN+1)-2*uu(iN)+uu(iN-1))/h^2;
```

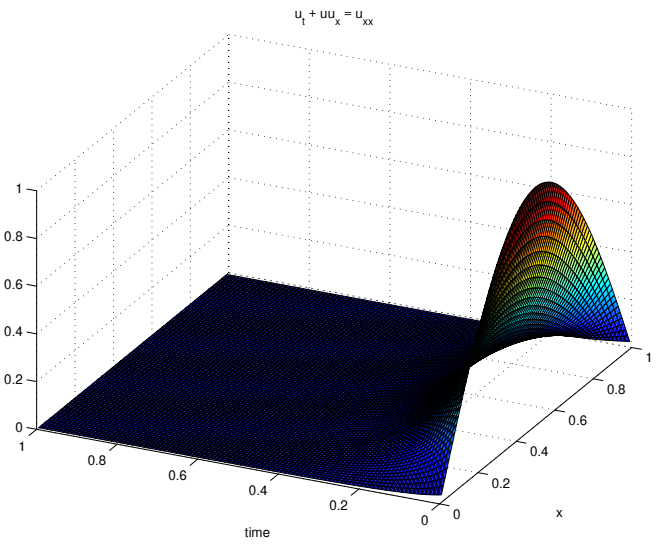


Figure C.1: Plot of the numerical solution of $u_t + uu_x = u_{xx}$ with $N = 100$ and $t \in [0, 1]$.

Appendix D

Source Code

The source code for the fixed stepsize Euler-Maruyama method for Geometric Brownian motion:

```
function y=euler(pars)
%
% This function numerically simulates a strong solution
% of the stochastic differential equation using the Euler method
%
%       dY_{t} = aY_{t}dt + bY_{t}dW_{t}
%
% a and b must be scalars.
% Use the function call
%
%       y=euler(pars)
%
% where pars is a struct including
%        st = initial value
%      time = the time discretization
%       brp = the Brownian increment path
%        pa = other parameters, including
%                  a = drift coefficient
%                  b = diffusion coefficient
%
format long
y0=pars.st;
t=pars.time;
brp=pars.brp;
other=pars.pa;
```

```
a=other(1);
b=other(2);

n=length(t);
h=(t(n)-t(1))/(n-1);
y=zeros(size(t));
y(1)=y0;
yp=y0;

% generate the full path
for count = 2:n
    y(count)= yp + yp*a*h + yp*b*brp(count);
    yp=y(count);
end
```

The source code for the fixed stepsize Milstein method for GBM:

```
function y=milstein(pars)
%
% This function numerically simulates a strong solution
% of the stochastic differential equation using the Milsein method
%
%       dY_{t} = aY_{t}dt + bY_{t}dW_{t}
%
% a and b must be scalars (so b'(X,t) = b).
% Use the function call
%
%       y=milstein(pars)
%
% where pars is a struct including
%        st = initial value
%        pa = vector containing
%             a = drift coefficient
%             b = diffusion coefficient
%       brp = brownian path
%      time = time vector
%

format long
y0=pars.st;
t=pars.time;
```

```
brp=pars.brp;
other=pars.pa;
a=other(1);
b=other(2);

n=length(t);
h=(t(n)-t(1))/(n-1);
y=zeros(size(t));
y(1)=y0;
yp=y0;

% generate the full path
for count = 2:n
    y(count)= yp + a*yp*h + b*yp*brp(count) + ...
                0.5*b*b*yp*((brp(count))^2 - h);
    yp=y(count);
end
```

The source code for the fixed stepsize Balanced method for GBM:

```
function y=bal(par)
%
% This function numerically simulates a strong solution
% of the stochastic differential equation using the balanced method
%
%       dY_{t} = aY_{t}dt + bY_{t}dW_{t}.
%
% a and b must be scalars.
% Use the function call
%
%       y=bal(par)
% where pars is a struct including
%        st = initial value
%        pa = vector containing
%             a = drift coefficient
%             b = diffusion coefficient
%       brp = brownian path
%      time = time vector
%

format long
```

```
y0=par.st;
t=par.time;
brp=par.brp;
param=par.pa;
a=param(1);
b=param(2);

n=length(t);
h=(t(n)-t(1))/(n-1);
y=zeros(size(t));
y(1)=y0;
yp=y0;

% generate the full path
for count = 2:n
    t1=h*count;
    y(count)=yp*(1+(2*a*h)+(b*(brp(count)+abs(brp(count))))) ...
              /(1+(a*h)+(b*abs(brp(count))));
    yp=y(count);
end
```

Bibliography

Ames, W. (1992), *Numerical Methods for Partial Differential Equations*, Computer Science and Scientific Computing, Academic Press.

Artemiev, S. & Averina, T. (1997), *Numerical Analysis of Systems of Ordinary and Stochastic Differential Equations*, VSP BV.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3), 637–654.

Black, F., Derman, E. & Toy, W. (1990), 'A one-factor model of interest rates and its application to Treasury bond options', *Financial Analysts Journal* pp. 33–39.

Boyce, W. & DiPrima, R. (1992), *Elementary Differential Equations and Boundary Value Problems*, 5th edn, WIley.

Brennan, M. & Schwartz, E. S. (1982), 'An equilibrium model of bond pricing and a test of market efficiency', *Journal of Financial and Quantitative Analysis* **17**, 301–329.

Burrage, K. (1995), *Parallel and Sequential Methods for Ordinary Differential Equations*, Oxford University Press.

Burrage, K. & Tian, T. (1999), 'A note on the stability properties of the euler methods for solving stochastic differential equations', *The special issue on the retirement of Professor John Butcher.*

Chriss, N. A. (1997), *Black-Scholes and beyond: option pricing models*, Irwin Professional Publishing.

Cox, J., Ingersoll, J. & Ross, S. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Fong, H. G. & Vasicek, O. A. (1991), 'Fixed-income volatility management', *Journal of Portfolio Management* pp. 41–46.

G.R.Grimmett & D.R.Stirzaker (1992), *Probability and Random Processes*, Oxford University Press.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology', *Econometrica* **60**(1), 77–105.

Ho, T. & Lee, S. (1986), 'Term structure movements and pricing interest rate contingent claims', *Journal of Finance* **41**, 1011–1029.

Hull, J. & White, A. (1990), 'Pricing interest rate derivative securities', *Review of Financial Studies* **3**, 573–592.

Hull, J. C. (2000), *Options, Futures & Other Derivatives*, 4th edn, Prentice Hall.

Jarrow, R. & Turnbull, S. (1996), *Derivative Securities*, South-Western College Publishing.

Khasminskii, Krylov & Moshchuk (1999), 'On the estimation of parameters for linear stochastic differential equations', *Probability theory and related fields*.

Kloeden, P. & Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag.

Kohn, R. (1999), *Partial Differential Equations for Finance*, New York University.

Kushner, H. (1967), *Stochastic stability and control*, Academic press.

Kushner, H. (1971), *Introduction to Stochastic Control*, Holt, Rinehart and Winston.

Kushner, H. (1977), *Probability Models for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press.

Kushner, H. & Dupuis, P. (1992), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Spinger-Verlag.

Lamberton, D. & Lapeyre, B. (1996), *Introduction to stochastic calculus applied to finance*, Chapman & Hall.

Longstaff, F. A. & Schwartz, E. S. (1992), 'A two-factor interest rate model and contingent claims valuation', *Journal of Fixed Income* **3**, 16–23.

Lythe, G. D. (1994), Stochastic slow-fast dynamics, PhD thesis, Trinity College, Cambridge.

Macki, J. & Strauss, A. (1982), *Introduction to Optimal Control Theory*, Springer-Verlag.

McCausland, I. (1969), *Introduction to Optimal Control*, Wiley.

Merton, R. (1971), 'Optimal consumtion and portfolio rules in a continuous-time model', *Journal of economic theory*.

Mikosch, T. (1998), *Elementary Stochastic Calculus*, World Scientific.

Milstein, G., Platen, E. & Schurz, H. (1998), 'Balanced implicit methods for stiff stochastic systems.', *SIAM Journal of Numerical Analysis*.

Morton, K. & Mayers, D. (1994), *Numerical Solution of Partial Differential Equations*, Cambridge Univeristy Press.

Musiela, M. & Rutkowski, M. (1997), *Martingale Methods in Financial Modelling*, Springer.

Neftci, S. (1996), *An Introduction to the Mathematics of Financial Derivatives*, Academic Press.

Odegaard, B. (1998), Financial numerical recipes, Technical report, Norwegian School of Management.

Øksendal, B. (1998), *Stochastic Differential Equations - An Introduction with Applications*, 5th edn, Springer.

Petersen, W. (1998), 'A general implicit splitting for stabilizing numerical simulations of ito stochastic differential equations.', *SIAM Journal of Numerical Analysis*.

Pollett, P. (1999), 'MS308 Lecture Notes'.

Rebonato, R. (1998), *Interst-Rate Option Models*, 2nd edn, John-Wiley.

Selby, M. J. P. & Strickland, C. (1995), 'Computing the Fong and Vasicek pure discount bond price formula', *Journal of Fixed Income* pp. 177–188.

Shoji, I. & Ozaki, T. (1997), 'Comparitive study of estimation methods for continuous time stochastic processes', *Journal of time series analysis*.

Shreve, S. & Soner, H. (1991), 'Optimal imvestment and consumption with two bonds and transaction costs.', *Mathematical Finance*.

Thompson, M. (1999), 'MS479 Lecture Notes'.

Tian, T., Burrage, K. & Volker, R. (1999), The stochastic model and numerical simulations for solute transport in porous media, To appear.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**, 177–188.

Wilmott, P. (1998), *Derivatives:The Theory and Practice of Financial Engineering*, Wiley.