MATHEMATICAL AND NUMERICAL METHODS

BRAD BAXTER

Version: 202303131253

ABSTRACT. These notes contain all examinable theoretical material for the Methods course in 2022–2023, although I shall add further examples during the year.

CONTENTS

1. Introduction

You can access these notes, and other material, via my office machine:

`http://econ109.econ.bbk.ac.uk/brad/Methods/`

My main lecture notes are available here:

`http://econ109.econ.bbk.ac.uk/brad/Methods/new_methods_notes.pdf }`

These notes are fairly stable, having evolved while teaching three different MSc programmes: MSc Mathematical Finance at Imperial College, London, MSc Financial Engineering here, and now MSc Mathematical Finance. I do still add new examples and make minor changes, so please check you have the latest version.

For those students who are only taking the Autumn Term of this course (also known as **Continuous Time Stochastic Processes**), Sections 2 and 3 are the key sections, although I shall also include some material from Section 4 (specifically Delta Hedging for the Binomial Model, to motive its continuous time analogue).

Some students will also be taking my Matlab course, but my Matlab notes are available to all:

`http://econ109.econ.bbk.ac.uk/brad/Methods/matlab_intro_notes.pdf`

My friend and colleague Raymond Brummelhuis provided very lucid notes for an earlier version of this course:

`http://econ109.econ.bbk.ac.uk/brad/Methods/old_methods_notes_RB.pdf`

Raymond's notes are still highly useful, but please do be remember that these notes define the current syllabus.

Past exams can be downloaded from

`http://econ109.econ.bbk.ac.uk/brad/FinEngExams/`

Many students will find my Numerical Analysis notes helpful too:

`http://econ109.econ.bbk.ac.uk/brad/Methods/nabook.pdf`

I wrote these notes for an undergraduate course in Numerical Analysis when lecturing at Imperial College, London, from 1995–2001. However, they have often been found useful by MSc students who need to improve their general understanding of theoretical Numerical Analysis. The first section of the notes is on matrix algebra and contain many examples and exercises, together with solutions.

Finally, there is lots of interesting material, including extensive notes for several related courses (e.g. Analysis) available on my office Linux server, so please do explore:

`http://econ109.econ.bbk.ac.uk/brad/`

Despite my providing you with lots of online notes, in my experience, students will benefit enormously from the old-fashioned method of taking notes as I lecture.

1.1. Reading List.

There are many books on mathematical finance, but very few good ones. My strongest recommendations are for the books by Higham and Kwok. However, the following books are all useful, but these notes are mostly either self-contained or refer to my own online notes.

(i) M. Baxter and A. Rennie, *Financial Calculus*, Cambridge University Press. This gives a fairly informal description of the mathematics of pricing, concentrating on martingales. It's not a source of information for efficient numerical methods.

(ii) A. Etheridge, *A Course in Financial Calculus*, Cambridge University Press. This does not focus on the algorithmic side but is very lucid, although it is probably better read once you are familiar with the contents of the first term.
(iii) D. Higham, *An Introduction to Financial Option Valuation*, Cambridge University Press. This book provides many excellent Matlab examples, although its mathematical level is undergraduate.
(iv) J. Hull, *Options, Futures and Other Derivatives*, 6th edition. [Earlier editions are probably equally suitable for much of the course.] Fairly clear, with lots of background information on finance. The mathematical treatment is lower than the level of much of our course (and this is *not* a mathematically rigorous book), but it's still the market leader in many ways.
(v) D. Kennedy, *Stochastic Financial Models*, Chapman and Hall. This is an excellent mathematical treatment, but probably best left until after completing the first term of Methods.
(vi) J. Michael Steele, *Stochastic Calculus and Financial Applications*, Springer. This is an excellent book, but is one to read near the end of this term, once you are more comfortable with fundamentals.
(vii) P. Wilmott, S. Howison and J. Dewynne, *The Mathematics of Financial Derivatives*, Cambridge University Press. This book is very useful for its information on partial differential equations. If your first degree was in engineering, mathematics or physics, then you probably spent many happy hours learning about the diffusion equation. This book is very much mathematical finance from the perspective of a traditional applied mathematician. It places much less emphasis on probability theory than most books on finance.
(viii) P. Wilmott, *Paul Wilmott introduces Quantitative Finance*, 2nd edition, John Wiley. More chatty than his previous book. The author's ego grew enormously between the appearance of these texts, but there's some good material here.
(ix) Y.-K. Kwok, *Mathematical Models of Financial Derivatives*, Springer. Rather dry, but very detailed treatment of finite difference methods. If you need a single book for general reference work, then this is probably it.

There are lots of books suitable for mathematical revision. The **Schaum series** publishes many good inexpensive textbooks providing worked examples. The inexpensive paperback *Calculus*, by K. G. Binmore (Cambridge University Press) will also be useful to students wanting an introduction to, say, multiple integrals, as will *Mathematical Methods for Science Students*, by Geoff Stephenson. At a slightly higher level, *All you wanted to know about Mathematics but were afraid to ask*, by L. Lyons (Cambridge University Press, 2 vols), is useful and informal.

The ubiquitous *Numerical Recipes in C++*, by S. Teukolsky et al, is extremely useful. Its coverage of numerical methods is generally reliable and it's available online at `www.nr.com`. A good hard book on partial differential equations is that of A. Iserles (Cambridge University Press).

At the time of writing, finance is going through a turbulent period, in which politicians profess their longstanding doubts that the subject was well-founded – surprisingly, many omitted to voice such doubts earlier! It is good to know that we have been here before. The following books are included for general cultural interest.

(i) M. Balen, *A Very English Deceit: The Secret History of the South Sea Bubble and the First Great Financial Scandal.*
(ii) C. Eagleton and J. William (eds), *Money: A History.*
(iii) C. P. Kindleberger, R. Aliber and R. Solow, *Manias, Panics, and Crashes: A History of Financial Crises*, Wiley. This is still a classic.
(iv) J. Lanchester, *How to Speak Money*, Faber. This is an **excellent** introduction to finance and economics for all readers. Lanchester is a journalist and author, as well as being a gifted expositor.
(v) N. N. Taleb, *The Black Swan.* In my view, this is greatly over-rated, but you should still read it.

No text is perfect: please report any slips to `b.baxter@bbk.ac.uk`.

2. The Geometric Brownian Motion Universe

We shall begin with a brisk introduction to the main topics, filling in the details later. The real economy is vastly complex, so mathematical finance begins with vast oversimplification.

Let r be the *risk-free* interest rate, which we shall assume constant. This is really the interest paid by the state when borrowing money via selling bonds, and it is **nominally** risk-free in any state that issues its own currency, although the real value of that currency can greatly decrease. We assume that everyone in our mathematical economy can borrow and lend at this rate, so that such debts (or investments, if lent) satisfy $B_t = B_0 \exp(rt)$. In reality, banks and companies borrow and lend at a higher rate $r + \delta$, where δ increases with the perceived risk of the lender, but this complication is ignored here.

Notation: In most (but not all) areas of mathematics, a function B depending on time t would be denoted $B(t)$, but mathematical finance often uses the alternative notation B_t which is very common in probability theory, statistics and economics. I shall be consistent in using $S(t)$ to denote the share price in Section 2, but we shall move to S_t in Section 3.

We shall assume that every **risky** asset (such as a share) is described by a random process called *geometric Brownian motion* (GBM):

$$S(t) = S(0)e^{\beta t + \sigma W_t}, \qquad t > 0, \tag{2.1}$$

where W_t denotes Brownian motion, $\beta \in \mathbb{R}$ and σ is a non-negative parameter called the *volatility* of the asset. You can think of Brownian motion as an important generalization of random walk, but we shall postpone its detailed definition and properties until Section 3. Fortunately all we need for now is the fundamental property that W_T is a normal (or Gaussian) random variable with mean zero and variance T, that is,

$$W_t \sim N(0, t), \qquad \text{for all } t > 0. \tag{2.2}$$

As we shall see later, option pricing requires us to use $\beta = r - \sigma^2/2$, that is,

$$S(t) = S(0)e^{(r-\sigma^2/2)t + \sigma W_t}, \qquad t > 0, \tag{2.3}$$

and this is usually called *risk neutral* GBM. The reason for the disappearance of the parameter β when pricing options is rather deep and extremely important, but will be explained later. All you need to know at present is that (2.1) is the mathematical model for share prices in the real world, but the risk neutral variant (2.3) is used when pricing options, i.e. contracts whose value depends on the asset price.

Thus, when pricing options, to generate sample prices $S(T)$ at some future time T given the initial price $S(0)$, we use

$$S(T) = S(0)\exp\Big((r - \sigma^2/2)T + \sigma\sqrt{T} Z_T\Big), \qquad \text{where } Z_T \sim N(0,1), \tag{2.4}$$

because $W_T \sim N(0, T)$, so that we can write $W_T = T^{1/2} Z_T$.

Example 2.1. *Generating sample prices at a fixed time T using* (2.4) *is particularly easy in Matlab and Octave:*

```
S = S0*exp((r-sigma^2/2)*T + sigma*sqrt(T)*randn(m,1)
```

will construct a column vector of m sample prices once you've defined $S0$, r, σ and T. To calculate the sample average price, we type `sum(S)/m`.

To analytically calculate $\mathbb{E}S(T)$ we need the following simple, yet crucial, lemma.

Lemma 2.1. *If $W \sim N(0,1)$, then $\mathbb{E}\exp(\lambda W) = \exp(\lambda^2/2)$.*

Proof. We have

$$\mathbb{E}e^{\lambda W} = \int_{-\infty}^{\infty} e^{\lambda t}(2\pi)^{-1/2}e^{-t^2/2}\,dt = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(t^2-2\lambda t)}\,dt.$$

The trick now is to *complete the square* in the exponent, that is,

$$t^2 - 2\lambda t = (t-\lambda)^2 - \lambda^2.$$

Thus

$$\mathbb{E}e^{\lambda W} = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}([t-\lambda]^2 - \lambda^2)\right)\,dt = e^{\lambda^2/2}.$$

This is also described in detail in Example 6.3. □

Lemma 2.2. *For every $\sigma \ge 0$, we have the expected growth*

$$\mathbb{E}S(t) = S(0)e^{rt}, \qquad t \ge 0. \tag{2.5}$$

Proof. This is an easy consequence of Lemma 2.1. □

The option pricing risk-neutral geometric Brownian motion universe might therefore seem rather strange, because *every* asset has the same expected growth e^{rt} as the risk-free interest rate. Thus our universe of all possible assets in a risk-neutral world is specified by one parameter only: the volatility σ. Please do remember that this is *not* the asset price in the market, but simply a mathematical device required for pricing options based on the asset.

A *financial derivative* is any function $f(S,t)$. We shall concentrate on the following particular class of derivatives.

Definition 2.1. *A* European option *is any function $f \equiv f(S,t)$ that satisfies the conditional expectation equation*

$$f(S(t),t) = e^{-rh}\mathbb{E}\Big(f(S(t+h),t+h)|S(t)\Big), \qquad \textit{for any } h > 0. \tag{2.6}$$

We shall often simply write this as

$$f(S(t),t) = e^{-rh}\mathbb{E}f(S(t+h),t+h)$$

but you should take care to remember that this is an expected future value given *the asset's current value $S(t)$. We see that (2.6) describes a contract $f(S,t)$ whose current value is the* discounted value *of its expected future value in the risk-neutral GBM universe.*

We can learn a great deal by studying the mathematical consequences of (2.6) and (2.3).

Example 2.2. *A* plain vanilla European put option *is simply an insurance contract that allows us to sell one unit of the asset, for exercise price K, at time T in the future. If the asset's price $S(T)$ is less than K at this expiry time, then the option is worth $K - S(T)$, otherwise it's worthless. Such contracts protect us if we're worried that the asset's price might drop. The pricing problem here to calculate the value of the contract at time zero given its value at expiry, namely*

$$f_P(S(T),T) = (K - S(T))_+\,, \tag{2.7}$$

where $(z)_+ := max\{z, 0\}$.

Typically, we know the value of the option $f(S(T), T)$ for all values of the asset $S(T)$ at some future time T. Our problem is to compute its value at some earlier time, because we're buying or selling this option.

Example 2.3. *A* plain vanilla European call option *gives us the right to buy one unit of the asset at the exercise price K at time T. If the asset's price $S(T)$ exceeds K at this expiry time, then the option is worth $S(T) - K$, otherwise it's worthless, implying the expiry value*

$$f_C(S(T), T) = (S(T) - K)_+\,, \tag{2.8}$$

using the same notation as Example 2.2. Such contracts protect us if we're worried that the asset's price might rise.

How do we compute $f(S(0), 0)$? The difficult part is computing the expected future value $\mathbb{E}f(S(T), T)$. This can be done analytically for a tiny number of options, including the European Put and Call (see Theorem 2.5), but usually we must resort to a numerical calculation. This leads us to our first algorithm: *Monte Carlo simulation*. Here we choose a large integer N and generate N pseudo-random numbers $Z_1, Z_2, \ldots, Z_N$ that have the normalized Gaussian distribution; in Matlab, we simply write `Z = randn(N,1)`. Using (2.3), these generate the future asset prices

$$S_k = S(0)\exp\Big((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}Z_k\Big), \qquad k = 1, \ldots, N. \tag{2.9}$$

We then approximate the future expected value by an average, that is, we take

$$f(S(0), 0) \approx \frac{e^{-rT}}{N}\sum_{k=1}^{N} f(S_k, T). \tag{2.10}$$

Monte Carlo simulation has the great advantage that it is extremely simple to program. Its disadvantage is that the error is usually a multiple of $1/\sqrt{N}$, so that very large N is needed for high accuracy (each decimal place of accuracy requires about a hundred times more work). We note that (2.10) will compute the value of *any* European option that is completely defined by a known final value $f(S(T), T)$.

We shall now use Monte Carlo to approximately evaluate the European Call and Put contracts. In fact, Put-Call parity, described below in Theorem 2.3, implies that we only need a program to calculate one of these, because they are related by the simple formula

$$f_C(S(0), 0) - f_P(S(0), 0) = S(0) - Ke^{-rT}. \tag{2.11}$$

Here's the Matlab program for the European Put.

```
%
% These are the parameters chosen in Example 11.6 of
% OPTIONS, FUTURES AND OTHER DERIVATIVES,
% by John C. Hull (Prentice Hall, 4th edn, 2000)
%
%% initial stock price
S0 = 42;
% unit of time = year
% 250 working days per year
```

```
% continuous compounding risk-free rate
r = 0.1;
% exercise price
K = 40;
% time to expiration in years
T = 0.5;
% volatility of 20 per cent annually
sigma = 0.2;
% generate asset prices at expiry
Z = randn(N,1);
ST = S0*exp( (r-(sigma^2)/2)*T + sigma*sqrt(T)*Z );
% calculate put contract values at expiry
fput = max(K - ST,0.0);
% average put values at expiry and discount to present
mc_put = exp(-r*T)*sum(fput)/N
% calculate analytic value of put contract
wK = (log(K/S0) - (r - (sigma^2)/2)*T)/(sigma*sqrt(T));
a_put = K*exp(-r*T)*Phi(wK) - S0*Phi(wK - sigma*sqrt(T))
```

The function `Phi` denotes the cumulative distribution function for the normalized Gaussian distribution, that is,

$$\Phi(x) = \mathbb{P}(Z \le x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-s^2/2}\, ds, \qquad \text{for } x \in \mathbb{R}, \tag{2.12}$$

where $Z \sim N(0,1)$.

Unfortunately, Matlab only provides the very similar *error function*, defined by

$$\operatorname{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y \exp(-s^2)\, ds, \qquad y \in \mathbb{R}.$$

It's not hard to prove that

$$\Phi(t) = \frac{1}{2}\Big(1 + \operatorname{erf}(t/\sqrt{2})\Big), \qquad t \in \mathbb{R}.$$

We can add this to Matlab using the following function.

```
function Y = Phi(t)
Y = 0.5*(1.0 + erf(t/sqrt(2)));
```

We have only revealed the tip of a massive iceberg in this brief introduction. Firstly, the Black–Scholes model, where asset prices evolve according to (2.3), is rather poor: reality is far messier. Further, there are many types of option which are *path-dependent*: the value of the option at expiry depends not only on the final price $S(T)$, but on its previous values $\{S(t) : 0 \le t \le T\}$. In particular, there are *American options*, where the contract can be exercised at any time before its expiry. All of these points will be addressed in our course, but you should find that Hull's book provides excellent background reading (although his mathematical treatment is often sketchy). Higham provides a clear Matlab-based exposition.

Although the future expected value usually requires numerical computation, there are some simple cases that are analytically tractable. These are particularly important because they often arise in examinations!

2.1. **European Puts and Calls.** It's not too hard to calculate the values of these options analytically. Further, the next theorem gives an important relation between the prices of call and put options.

Theorem 2.3 (Put-Call parity)**.** *European Put and Call options, each with exercise price K and expiry time T, satisfy*

$$f_C(S,t) - f_P(S,t) = S - Ke^{-r\tau}, \qquad \text{for } S \in \mathbb{R} \text{ and } 0 \le t \le T, \tag{2.13}$$

where $\tau = T - t$, the time-to-expiry.

Proof. The trick is the observation that

$$y = y_+ - (-y)_+,$$

for any $y \in \mathbb{R}$. Thus

$$\begin{aligned} S(T) - K &= \left(S(T) - K\right)_+ - \left(K - S(T)\right)_+ \\ &= f_C(S(T), T) - f_P(S(T), T), \end{aligned}$$

which implies

$$e^{-r\tau}\mathbb{E}\left(S(T) - K | S(t) = S\right) = f_C(S,t) - f_P(S,t).$$

Now

$$\begin{aligned} \mathbb{E}\left(S(T)|S(t) = S\right) &= (2\pi)^{-1/2}\int_{-\infty}^{\infty} Se^{(r-\sigma^2/2)\tau+\sigma\sqrt{\tau}w}e^{-w^2/2}\,dw \\ &= Se^{(r-\sigma^2/2)\tau}(2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(w^2-2\sigma\sqrt{\tau}w\right)}\,dw \\ &= Se^{r\tau}, \end{aligned}$$

and some simple algebraic manipulation completes the proof. □

This is a useful check on the Monte Carlo approximations of the options' values. To derive their analytic values, we shall need the cumulative distribution function

$$\Phi(y) = (2\pi)^{-1/2}\int_{-\infty}^{y} e^{-z^2/2}\,dz, \qquad y \in \mathbb{R}, \tag{2.14}$$

for the Gaussian probability density, that is, $\mathbb{P}(Z \le y) = \Phi(y)$ and $\mathbb{P}(a \le Z \le b) = \Phi(b) - \Phi(a)$, for any normalized Gaussian random variable Z. Further, we have the following relation which will be of use in subsequent formulae.

Lemma 2.4. *We have $1 - \Phi(a) = \Phi(-a)$, for any $a \in \mathbb{R}$.*

Proof. Observe that

$$\begin{aligned} 1 - \Phi(a) &= \int_a^{\infty} (2\pi)^{-1/2}e^{-s^2/2}\,ds \\ &= \int_{-\infty}^{-a} (2\pi)^{-1/2}e^{-u^2/2}\,du \\ &= \Phi(-a), \end{aligned}$$

where we have made the substitution $u = -s$. □

Theorem 2.5. *A European Put option satisfies*

$$f_P(S,t) = Ke^{-r\tau}\Phi(w(K)) - S\Phi(w(K) - \sigma\sqrt{\tau}), \quad \text{for } S \in \mathbb{R}, \tag{2.15}$$

where $\tau = T - t$, i.e. the time-to-expiry, and $w(K)$ is defined by the equation

$$K = Se^{(r-\sigma^2/2)\tau + \sigma\sqrt{\tau}w(K)},$$

that is

$$w(K) = \frac{\log(K/S) - (r - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}. \tag{2.16}$$

Proof. We have

$$\mathbb{E}\left(f_P(S(T),T)|S(t) = S\right) = (2\pi)^{-1/2}\int_{-\infty}^{\infty}\left(K - Se^{(r-\sigma^2/2)\tau+\sigma\sqrt{\tau}w}\right)_+ e^{-w^2/2}\,dw.$$

Now the function

$$w \mapsto K - S\exp((r - \sigma^2/2)\tau + \sigma\sqrt{\tau}w)$$

is strictly decreasing, so that

$$K - Se^{(r-\sigma^2/2)\tau+\sigma\sqrt{\tau}w} \geq 0$$

if and only if $w \leq w(K)$, where $w(K)$ is given by (2.16). Hence

$$\begin{aligned}
\mathbb{E}\left(f_P(S(T),T)|S(t) = S\right) &= (2\pi)^{-1/2}\int_{-\infty}^{w(K)}\left(K - Se^{(r-\sigma^2/2)\tau+\sigma\sqrt{\tau}w}\right)e^{-w^2/2}\,dw \\
&= K\Phi(w(K)) - Se^{(r-\sigma^2/2)\tau}(2\pi)^{-1/2}\int_{-\infty}^{w(K)} e^{-\frac{1}{2}\left(w^2-2\sigma\sqrt{\tau}w\right)}\,dw \\
&= K\Phi(w(K)) - Se^{r\tau}\Phi(w(K) - \sigma\sqrt{\tau}).
\end{aligned}$$

Thus

$$\begin{aligned}
f_P(S,t) &= e^{-r\tau}\mathbb{E}\left(f_P(S(T),T)|S(t) = S\right) \\
&= Ke^{-r\tau}\Phi(w(K)) - S\Phi(w(K) - \sigma\sqrt{\tau}).
\end{aligned}$$

□

There is an almost standard notation for Theorem 2.5, which is contained in the following corollary.

Corollary 2.6. *A European Put option satisfies*

$$f_P(S,t) = Ke^{-r\tau}\Phi(-d_-) - S\Phi(-d_+), \quad \text{for } S \in \mathbb{R}, \tag{2.17}$$

where $\tau = T - t$, i.e. the time-to-expiry, and

$$d_\pm = \frac{\log(S/K) + (r \pm \sigma^2/2)\tau}{\sigma\sqrt{\tau}}. \tag{2.18}$$

Proof. This is simply rewriting Theorem 2.5 in terms of (2.18). □

We can now calculate the price of a European call using Corollary 2.6 and the Put-Call parity Theorem 2.3.

Corollary 2.7. *A European Call option satisfies*

$$f_C(S,t) = S\Phi(d_+) - Ke^{-r\tau}\Phi(d_-), \quad \text{for } S \in \mathbb{R}, \tag{2.19}$$

where $\tau = T - t$*, i.e. the time-to-expiry, and* $d_\pm$ *is given by* (2.18)*.*

Proof. Theorem 2.3 implies that

$$\begin{aligned} f_C(S,t) &= f_P(S,t) + S - Ke^{-r\tau} \\ &= Ke^{-r\tau}\Phi(-d_-) - S\Phi(-d_+) + S - Ke^{-r\tau} \\ &= S\left(1 - \Phi(-d_+)\right) - Ke^{-r\tau}\left(1 - \Phi(-d_-)\right) \\ &= S\Phi(d_+) - Ke^{-r\tau}\Phi(d_-), \end{aligned}$$

using Lemma 2.4. □

Exercise 2.1. *Modify the proof of Theorem 2.5 to derive the analytic price of a European Call option. Check that your price agrees Corollary 2.7.*

2.2. **Digital Options.** A digital option is simply an option that only takes the values 0 and 1, that is, it is the indicator function for some event. Recall that, for any indicator function I_A, we have

$$\mathbb{E} I_A = \mathbb{P}(A).$$

Our first example is the digital call option with exercise price K and expiry time T is defined by

$$f_{DC}(S(T),T) = \begin{cases} 1 & \text{if } S(T) \ge K, \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

Theorem 2.8. *The digital call option* f_{DC} *satisfies*

$$f_{DC}(S,t) = e^{-r\tau}\Phi(d_-), \tag{2.21}$$

where $\tau = T - t$ *and* d_- *is defined by* (2.18)*.*

Proof. Its price at any earlier time $t \in [0,t)$ is therefore given by

$$\begin{aligned} f_{DC}(S,t) &= e^{-r\tau}\mathbb{E}\left(f_{DC}(S(T),T)|S(t) = S\right) \\ &= e^{-r\tau}\mathbb{E} f_{DC}(Se^{(r-\sigma^2/2)\tau + \sigma\tau^{1/2}Z}, \end{aligned} \tag{2.22}$$

(2.23)

where $Z \sim N(0,1)$.

Now

$$Se^{(r-\sigma^2/2)\tau+\sigma\tau^{1/2}Z} \ge K$$

if and only if

$$\log S + (r - \sigma^2/2)\tau + \sigma\tau^{1/2}Z \ge \log K,$$

because the logarithm is an increasing function. Rearranging this inequality, we find

$$Z \ge -\frac{\log(S/K) - (r - \sigma^2/2)\tau}{\sigma\tau^{1/2}} = -d_-.$$

Thus

$$\begin{aligned} f_{DC}(S,t) &= e^{-r\tau}\mathbb{P}\left(Z \ge -d_-\right) \\ &= e^{-r\tau}\left(1 - \Phi(-d_-)\right) \\ &= e^{-r\tau}\Phi(d_-), \end{aligned}$$

by Lemma 2.4. □

It is now simple to define and price the digital put option f_{DP}, which is defined by

$$f_{DP}(S(T), T) = \begin{cases} 1 & \text{if } S(T) < K, \\ 0 & \text{otherwise.} \end{cases} \tag{2.24}$$

A pair of digital put and call options with the same exercise price K and expiry time T satisfy a digital put-call parity relation, specifically

$$f_{DC}(S(T), T) + f_{DP}(S(T), T) \equiv 1,$$

at expiry, which implies

$$f_{DC}(S, t) + f_{DP}(S, t) \equiv e^{-r\tau}, \quad \text{for } S \in \mathbb{R}, \tag{2.25}$$

where $\tau = T - t$.

Theorem 2.9. *The digital put option f_{DP} satisfies*

$$f_{DP}(S, t) = e^{-r\tau}\Phi(-d_-), \tag{2.26}$$

where $\tau = T - t$ and d_- is defined by (2.18).

Proof. We use (2.25) and Lemma 2.4:

$$f_{DP}(S, t) = e^{-r\tau} - f_{DC}(S, t) = e^{-r\tau}\left(1 - \Phi(d_-)\right) = e^{-r\tau}\Phi(-d_-).$$

□

Another way to express digital calls and puts is as follows. Observe that

$$f_{DC}(S(T), T) = (S(T) - K)_+^0 \quad \text{and} \quad f_{DP}(S(T), T) = (K - S(T))_+^0. \tag{2.27}$$

Thus we have shown that

$$\begin{aligned}
\mathbb{E}\left((S(T) - K)_+ | S(t) = S\right) &= Se^{r\tau}\Phi(d_+) - K\Phi(d_-), \\
\mathbb{E}\left((S(T) - K)_+^0 | S(t) = S\right) &= \Phi(d_-), \\
\mathbb{E}\left((K - S(T))_+ | S(t) = S\right) &= K\Phi(-d_-) - Se^{r\tau}\Phi(-d_+), \\
\mathbb{E}\left((K - S(T))_+^0 | S(t) = S\right) &= \Phi(-d_-),
\end{aligned}$$

2.3. **European Puts and Calls.** The first and second partial derivatives of option prices are important both numerically and financially. A fairly standard nomenclature has evolved: for any option f, its partial derivative $\partial f/\partial S \equiv \partial_S f$ is called the "Delta" of the option. It is not our aim to provide an exhaustive list of the Greek (and non-Greek) letters used to denote these partial derivatives, but it is important to see how some of them are calculated.

Theorem 2.10. *If we let f_C denote the call option whose value is given by Theorem 2.7, then*

$$\partial_S f_C(S, t) = \Phi(d_+), \tag{2.28}$$

where d_+ is defined by (2.18).

This calculation is made easier by several preliminary results.

Lemma 2.11. *We have*

$$\partial_S d_\pm = \frac{1}{S\sigma\tau^{1/2}}. \tag{2.29}$$

Proof. Now

$$d_\pm = \frac{\log S - \log K + (r \pm \sigma^2/2)\tau}{\sigma\tau^{1/2}},$$

so that

$$\partial_S d_\pm = \frac{\partial_S \log S}{\sigma\tau^{1/2}} = \frac{1}{S\sigma\tau^{1/2}}.$$

□

Lemma 2.12. *The function* Φ *defined by* (2.12) *has derivative*

$$\Phi'(x) = (2\pi)^{-1/2} e^{-x^2/2}, \qquad \textit{for } x \in \mathbb{R}. \tag{2.30}$$

Proof. This is the fundamental theorem of calculus applied to the definition of (2.12). □

Lemma 2.13. *We have*

$$\frac{1}{2}d_+^2 - \frac{1}{2}d_-^2 = \log(S/K) + r\tau, \tag{2.31}$$

where $d_\pm$ *is defined by* (2.18)*.*

Proof. Let use write $d_\pm = a \pm b$, where

$$a = \frac{\log(S/K) + r\tau}{\sigma\tau^{1/2}}$$

and

$$b = \frac{\sigma^2\tau/2}{\sigma\tau^{1/2}}.$$

Hence

$$d_+^2 - d_-^2 = (a+b)^2 - (a-b)^2 = 4ab,$$

and so

$$\frac{1}{2}\left(d_+^2 - d_-^2\right) = 2ab = \log(S/K) + r\tau.$$

□

Lemma 2.14. *We have*

$$Se^{-d_+^2/2} - Ke^{-r\tau}e^{-d_-^2/2} = 0. \tag{2.32}$$

Proof. Using Lemma 2.13, we see that

$$\log K - r\tau - \frac{1}{2}d_-^2 = \log S - \frac{1}{2}d_+^2,$$

and taking the exponential we find

$$Ke^{-r\tau}e^{-d_-^2/2} = Se^{-d_+^2/2},$$

as required. □

We can now assemble these partial results to prove Theorem 2.10.

Proof. **Theorem 2.10:** Partially differentiating f_C with respect to S, we obtain

$$\partial_S f_C = \Phi(d_+) + L,$$

where

$$L = \frac{Se^{-d_+^2/2} - Ke^{-r\tau}e^{-d_-^2/2}}{S\sigma\tau^{1/2}(2\pi)^{1/2}},$$

by Lemma 2.11 and Lemma 2.12. Hence $\partial_S f_C = \Phi(d_+)$, by Lemma 2.14. □

Theorem 2.15. *If f_P denotes the put option with exercise price K and expiry time T, then*

$$\partial_S f_P(S,t) = -\Phi(-d_+), \tag{2.33}$$

where d_+ is defined by (2.18).

Proof. If we partially differentiate the put-call parity relation (2.13) with respect to S, then we obtain

$$\partial f_C - \partial f_P = 1.$$

Hence, using Lemma 2.4,

$$\partial f_P = \partial f_C - 1 = \Phi(d_+) - 1 = -\left(1 - \Phi(d_+)\right) = -\Phi(-d_+).$$

□

The *Vega* for any option V is simply $\partial_\sigma V$. It is also easily calculated for European plain vanilla options.

Theorem 2.16. *If f_C denotes the call option with exercise price K and expiry time T, then*

$$\partial_\sigma f_C(S,t) = S\tau^{1/2}(2\pi)^{-1/2}e^{-d_+^2/2}, \tag{2.34}$$

where d_+ is defined by (2.18).

Proof. Partially differentiating (2.18) with respect to σ, we obtain

$$\partial_\sigma d_\pm = -\left(\frac{\log(S/K) + r\tau}{\tau^{1/2}\sigma^2}\right) \pm \frac{1}{2}\tau^{1/2}. \tag{2.35}$$

Hence

$$\begin{aligned}
\partial_\sigma f_C &= S\Phi'(d_+)\partial_\sigma d_+ - Ke^{-r\tau}\Phi'(d_-)\partial_\sigma d_- \\
&= (2\pi)^{-1/2}\left(Se^{-d_+^2/2}\partial_\sigma d_+ - Ke^{-r\tau}e^{-d_-^2/2}\partial_\sigma d_-\right) \\
&= G_1 + G_2,
\end{aligned}$$

where

$$G_1 = (2\pi)^{-1/2}\left(Se^{-d_+^2/2} - Ke^{-r\tau}e^{-d_-^2/2}\right)\left(\frac{-\log(S/K) + r\tau}{\sigma^2\tau^{1/2}}\right)$$

and

$$G_2 = \frac{1}{2}\tau^{1/2}(2\pi)^{-1/2}\left(Se^{-d_+^2/2} + Ke^{-r\tau}e^{-d_-^2/2}\right).$$

Applying Lemma 2.14, we see that $G_1 = 0$ and

$$G_2 = S\tau^{1/2}(2\pi)^{-1/2}e^{-d_+^2/2}.$$

□

If we partially differentiate the put-call parity relation (2.13) with respect to σ, then we find

$$\partial_\sigma f_C = \partial_\sigma f_P.$$

The *Theta* of an option V is simply its partial derivative with respect to time, i.e. $\partial_t V$. For the plain vanilla calls, it's useful to notice that

$$\partial_t V = -\partial_\tau V.$$

Theorem 2.17. *If f_C denotes the call option with exercise price K and expiry time T, then*

$$\partial_t f_C(S,t) = -\frac{\sigma S e^{-d_+^2/2}}{2(2\pi\tau)^{1/2}} - rKe^{-r\tau}\Phi(d_-). \tag{2.36}$$

where $d_\pm$ is defined by (2.18).

Proof. We first note that $\partial_t = -\partial_\tau$, where $\tau = T - t$ is the time to expiry, as usual. Partially differentiating f_C with respect to τ, we obtain

$$\begin{aligned}\partial_\tau f_C(S,t) &= S\Phi'(d_+)\partial_\tau d_+ + rKe^{-r\tau}\Phi(d_+) - Ke^{-r\tau}\Phi'(d_-)\partial_\tau d_- \\ &= A_1 + A_2 + A_3.\end{aligned}$$

Now

$$\begin{aligned}A_1 + A_3 &= S\Phi'(d_+)\partial\tau d_+ - Ke^{-r\tau}\Phi'(d_-)\partial_\tau d_- \\ &= S\Phi'(d_+)\left(\partial_\tau d_+ - \partial_\tau d_-\right),\end{aligned}$$

using Lemma 2.14. It is not difficult to check that

$$\partial_\tau d_+ - \partial_\tau d_- = \sigma\tau^{-1/2}.$$

Hence

$$A_1 + A_3 = \frac{1}{2}\sigma\tau^{-1/2}\Phi'(d_+) = \frac{S\sigma}{2(2\pi\tau)^{1/2}}e^{-d_+^2/2}.$$

Adding this expression to A_2, we obtain $\partial_\tau f_C$. □

3. Brownian Motion

3.1. **Simple Random Walk.** Let $X_1, X_2, \ldots$ be a sequence of independent random variables all of which satisfy

$$\mathbb{P}\left(X_i = \pm 1\right) = 1/2 \tag{3.1}$$

and define

$$S_n = X_1 + X_2 + \cdots + X_n. \tag{3.2}$$

We can represent this graphically by plotting the points $\{(n, S_n) : n = 1, 2, \ldots\}$, and one way to imagine this is as a *random walk*, in which the walker takes identical steps forwards or backwards, each with probability $1/2$. This model is called *simple random walk* and, whilst easy to define, is a useful laboratory in which to improve probabilistic intuition.

Another way to imagine S_n is to consider a game in which a fair coin is tossed repeatedly. If I win the toss, then I win £1; losing the toss implies a loss of £1. Thus S_n is my fortune at time n.

Firstly note that

$$\mathbb{E}S_n = \mathbb{E}X_1 + \cdots + \mathbb{E}X_n = 0.$$

Further, $\mathbb{E}X_i^2 = 1$, for all i, so that $\operatorname{var} X_i = 1$. Hence

$$\operatorname{var} S_n = \operatorname{var} X_1 + \operatorname{var} X_2 + \cdots + \operatorname{var} X_n = n,$$

since $X_1, \ldots, X_n$ are independent random variables.

3.2. **Discrete Brownian Motion.** We begin with a slightly more complicated random walk this time. We choose a timestep $h > 0$ and let $Z_1, Z_2, \ldots$ be independent $N(0, h)$ Gaussian random variables. We then define a curve ${B^{(h)}}_t$ by defining ${B^{(h)}}_0 = 0$ and

$$B^{(h)}(kh) = Z_1 + Z_2 + \cdots + Z_k, \tag{3.3}$$

for positive integer k. We then join the dots to obtain a piecewise linear function. More precisely, we define

$${B^{(h)}}_t = {B^{(h)}}_{kh} + (t - kh)\left(\frac{{B^{(h)}}_{(k+1)h} - {B^{(h)}}_{kh}}{h}\right), \qquad \text{for } t \in (kh, (k+1)h).$$

The resultant random walk is called discrete Brownian motion.

Proposition 3.1. *If $0 \le a \le b \le c$ and $a, b, c \in h\mathbb{Z}$, then the discrete Brownian motion increments ${B^{(h)}}_c - {B^{(h)}}_b$ and ${B^{(h)}}_b - {B^{(h)}}_a$ are independent random variables. Further, ${B^{(h)}}_c - {B^{(h)}}_b \sim N(0, c-b)$ and ${B^{(h)}}_b - {B^{(h)}}_a \sim N(0, b-a)$.*

Proof. Exercise. □

3.3. **Basic Properties of Brownian Motion.** It's not obvious that discrete Brownian motion has a limit, in some sense, when we allow the timestep h to converge to zero. However, it can be shown that this is indeed the case (and will see the salient features of the *Lévy–Cieselski* construction of this limit later). For the moment, we shall state the defining properties of Brownian motion.

Definition 3.1. *There exists a stochastic process W_t, called Brownian motion, which satisfies the following conditions:*

(i) $W_0 = 0$*;*
(ii) *If $0 \le a \le b \le c$, then the Brownian increments $W_c - W_b$ and $W_b - W_a$ are independent random variables. Further, $W_c - W_b \sim N(0, c-b)$ and $W_b - W_a \sim N(0, b-a)$;*
(iii) *W_t is continuous almost surely.*

Proposition 3.2. *$W_t \sim N(0, t)$ for all $t > 0$.*

Proof. Just set $a = 0$ and $b = t$ in (ii) of Definition 3.1. □

The increments of Brownian motion are independent Gaussian random variables, but the actual values W_a and W_b are **not** independent random variables, as we shall now see.

Proposition 3.3. *If $a, b \in [0, \infty)$, then $\mathbb{E}\left(W_a W_b\right) = \min\{a, b\}$.*

Proof. We assume $0 < a < b$, the remaining cases being easily checked. Then

$$\begin{aligned}\mathbb{E}\left(W_a W_b\right) &= \mathbb{E}\left(W_a\left[W_b - W_a\right] + W_a^2\right)\\ &= \mathbb{E}\left(W_a\left[W_b - W_a\right]\right) + \mathbb{E}\left(W_a^2\right)\\ &= 0 + a\\ &= a.\end{aligned}$$

□

Brownian motion is continuous almost surely but it is easy to see that it cannot be differentiable. The key observation is that

$$\frac{W_{t+h} - W_t}{h} \sim N(0, \frac{1}{h}). \tag{3.4}$$

In other words, instead of converging to some limiting value, the variance of the random variable $(W_{t+h} - W_t)/h$ tends to infinity, as $h \to 0$.

3.4. **Martingales.** A martingale is a mathematical version of a fair game, as we shall first illustrate for simple random walk.

Proposition 3.4. *We have*

$$\mathbb{E}\left(S_{n+k} | S_n\right) = S_n.$$

Proof. The key observation is that

$$S_{n+k} = S_n + X_{n+1} + X_{n+2} + \cdots + X_{n+k}$$

and $X_{n+1}, \ldots, X_{n+k}$ are all independent of $S_n = X_1 + \cdots + X_n$. Thus

$$\mathbb{E}\left(S_{n+k} | S_n\right) = S_n + \mathbb{E}X_{n+1} + \mathbb{E}X_{n+2} + \cdots + \mathbb{E}X_{n+k} = S_n.$$

□

To see why this encodes the concept of a fair game, let us consider a biased coin with the property that

$$\mathbb{E}\left(S_{n+10} | S_n\right) = 1.1 S_n.$$

Hence

$$\mathbb{E}\left(S_{n+10\ell} | S_n\right) = 1.1^{\ell} S_n.$$

In other words, the expected fortune $S_{n+10\ell}$ grows exponentially with ℓ. For example, if we ensure that $S_4 = 3$, by fixing the first four coin tosses in some fashion, then our expected fortune will grow by 10% every 10 tosses thereafter.

3.5. **Brownian Motion and Martingales.**

Proposition 3.5. *Brownian motion is a martingale, that is,* $\mathbb{E}\left(W_{t+h} | W_t\right) = W_t$, *for any* $h > 0$.

Proof.

$$\begin{aligned}\mathbb{E}\left(W_{t+h} | W_t\right) &= \mathbb{E}\left(\left[W_{t+h} - W_t\right] + W_t | W_t\right)\\ &= \mathbb{E}\left(\left[W_{t+h} - W_t\right] | W_t\right) + W_t\\ &= \mathbb{E}\left(\left[W_{t+h} - W_t\right]\right) + W_t\\ &= 0 + W_t\\ &= W_t.\end{aligned}$$

□

We can sometimes use a similar argument to prove that functionals of Brownian motion are martingales.

Proposition 3.6. *The stochastic process* $X_t = W_t^2 - t$ *is a martingale, that is,* $\mathbb{E}\left(X_{t+h}|X_t\right) = X_t$*, for any* $h > 0$.

Proof.

$$\begin{aligned}
\mathbb{E}\left(X_{t+h}|X_t\right) &= \mathbb{E}\left([W_{t+h} - W_t + W_t]^2 - [t+h]|W_t\right) \\
&= \mathbb{E}\left([W_{t+h} - W_t]^2 + 2W_t\left[W_{t+h} - W_t\right] + W_t^2 - t - h|W_t\right) \\
&= \mathbb{E}[W_{t+h} - W_t]^2 + W_t^2 - t - h \\
&= h + W_t^2 - t - h \\
&= X_t.
\end{aligned}$$

□

The following example will be crucial.

Proposition 3.7. *Geometric Brownian motion*

$$Y_t = e^{\alpha + \beta t + \sigma W_t} \tag{3.5}$$

is a martingale, that is, $\mathbb{E}\left(Y_{t+h}|Y_t\right) = Y_t$*, for any* $h > 0$*, if and only if* $\beta = -\sigma^2/2$.

Proof.

$$\begin{aligned}
\mathbb{E}\left(Y_{t+h}|Y_t\right) &= \mathbb{E}\left(e^{\alpha + \beta(t+h) + \sigma W_{t+h}}|Y_t\right) \\
&= \mathbb{E}\left(Y_t e^{\beta h + \sigma(W_{t+h} - W_t)}|Y_t\right) \\
&= Y_t \mathbb{E} e^{\beta h + \sigma(W_{t+h} - W_t)} \\
&= Y_t e^{\left(\beta + \sigma^2/2\right)h}.
\end{aligned}$$

□

In this course, the mathematical model chosen for option pricing is risk-neutral geometric Brownian motion: we choose a geometric Brownian motion S_t with the property that $Y_t = e^{-rt} S_t$ is a martingale, where Y_t is given by (3.5). Thus we have

$$Y_t = e^{\alpha + (\beta - r)t + \sigma W_t}$$

and Proposition 3.7 implies that $\beta - r = -\sigma^2/2$, i.e.

$$S_t = e^{\alpha + (r - \sigma^2/2)t + \sigma W_t} = S_0 e^{(r - \sigma^2/2)t + \sigma W_t}.$$

3.6. **The Black–Scholes Equation.** We can also use (2.6) to derive the famous *Black–Scholes partial differential equation*, which is satisfied by any European option. The key is to choose a *small* positive h in (2.6) and expand. We shall need

Taylor's theorem for functions of two variables, which states that

$$
\begin{aligned}
G(x+\delta x, y+\delta y) &= G(x,y) + \left(\frac{\partial G}{\partial x}\delta x + \frac{\partial G}{\partial y}\delta y\right) \\
&\quad + \frac{1}{2}\left(\frac{\partial^2 G}{\partial x^2}(\delta x)^2 + 2\frac{\partial^2 G}{\partial x \partial y}(\delta x)(\delta y) + \frac{\partial^2 G}{\partial y^2}(\delta y)^2\right) + \cdots .
\end{aligned}
$$

(3.6)

Further, it simplifies matters to use "log-space": we introduce $u(t) := \log S(t)$, where $\log \equiv \log_e$ in these notes (*not* logarithms to base 10). In log-space, (2.3) becomes

$$u(t+h) = u(t) + (r - \sigma^2/2)h + \sigma \delta W_t, \tag{3.7}$$

where

$$\delta W_t = W_{t+h} - W_t \sim N(0,h). \tag{3.8}$$

We also introduce

$$g(u(t),t) := f(\exp(u(t),t)), \tag{3.9}$$

so that (2.6) takes the form

$$g(u(t),t) = e^{-rh}\mathbb{E}g(u(t+h),t+h). \tag{3.10}$$

Now Taylor expansion yields the (initially daunting)

$$
\begin{aligned}
g(u(t+h),t+h) &= g(u(t) + (r-\sigma^2/2)h + \sigma\delta W_t, t+h) \\
&= g(u(t),t) + \frac{\partial g}{\partial u}\left((r-\sigma^2/2)h + \sigma\delta W_t\right) + \\
&\quad \frac{1}{2}\frac{\partial^2 g}{\partial u^2}\sigma^2(\delta W_t)^2 + h\frac{\partial g}{\partial t} + \cdots ,
\end{aligned} \tag{3.11}
$$

ignoring all terms of higher order than h. Further, since $\delta W_t \sim N(0,h)$, i.e. $\mathbb{E}\delta W_t = 0$ and $\mathbb{E}[(\delta W_t)^2] = h$, we obtain

$$\mathbb{E}g(u(t+h),t+h) = g(u(t),t) + h\left(\frac{\partial g}{\partial u}(r-\sigma^2/2) + \frac{1}{2}\frac{\partial^2 g}{\partial u^2}\sigma^2 + \frac{\partial g}{\partial t}\right) + \cdots . \tag{3.12}$$

Recalling that

$$e^{-rh} = 1 - rh + \frac{1}{2}(rh)^2 + \cdots ,$$

we find

$$
\begin{aligned}
g &= (1 - rh + \cdots)\Big(g + h\left[\frac{\partial g}{\partial u}(r-\sigma^2/2) + \frac{1}{2}\frac{\partial^2 g}{\partial u^2}\sigma^2 + \frac{\partial g}{\partial t}\right] + \cdots\Big) \\
&= g + h\Big(-rg + \frac{\partial g}{\partial u}(r-\sigma^2/2) + \frac{1}{2}\frac{\partial^2 g}{\partial u^2}\sigma^2 + \frac{\partial g}{\partial t}\Big) + \cdots .
\end{aligned} \tag{3.13}
$$

For this to be true for all $h > 0$, we must have

$$-rg + \frac{\partial g}{\partial u}(r-\sigma^2/2) + \frac{1}{2}\frac{\partial^2 g}{\partial u^2}\sigma^2 + \frac{\partial g}{\partial t} = 0, \tag{3.14}$$

and this is the celebrated Black–Scholes partial differential equation (PDE). Thus, instead of computing an expected future value, we can calculate the solution of the Black–Scholes PDE (3.14). The great advantage gained is that there are highly efficient numerical methods for solving PDEs numerically. The disadvantages are complexity of code and learning the mathematics needed to exploit these methods effectively.

Exercise 3.1. *Use the substitution $S = \exp(u)$ to transform (3.14) into the non-linear form of the Black–Scholes PDE.*

3.7. **Itô Calculus.** Equation (3.12) is really quite surprising, because the second derivative contributes to the $O(h)$ term. This observation is at the root of the Itô rules. We begin by considering the *quadratic variation* $I_n[a,b]$ of Brownian motion on the interval $[a,b]$. Specifically, we choose a positive integer n and let $nh = b - a$. We then define

$$I_n[a,b] = \sum_{k=1}^{n} \left(W_{a+kh} - W_{a+(k-1)h}\right)^2. \tag{3.15}$$

We shall prove that $\mathbb{E}I_n[a,b] = b - a$, for every positive integer n, but that $\operatorname{var} I_n[a,b] \to 0$, as $n \to \infty$. We shall need the following simple property of Gaussian random variables.

Lemma 3.8. *Let $Z \sim N(0,1)$. Then $\mathbb{E}Z^4 = 3$.*

Proof. Integrating by parts, we obtain

$$\begin{aligned}
\mathbb{E}Z^4 &= \int_{-\infty}^{\infty} s^4 (2\pi)^{-1/2} e^{-s^2/2}\, ds \\
&= (2\pi)^{-1/2} \int_{-\infty}^{\infty} s^3 \frac{d}{ds}\left(-e^{-s^2/2}\right) ds \\
&= (2\pi)^{-1/2} \left\{ \left[-s^3 e^{-s^2/2}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 3s^2 \left(-e^{-s^2/2}\right) ds \right\} \\
&= 3.
\end{aligned}$$

□

Exercise 3.2. *Calculate $\mathbb{E}Z^6$ when $Z \sim N(0,1)$. More generally, calculate $\mathbb{E}Z^{2m}$ for any positive integer m.*

Proposition 3.9. *We have $\mathbb{E}I_n[a,b] = b - a$ and $\operatorname{var} I_n[a,b] = 2(b-a)^2/n$.*

Proof. Firstly,

$$\mathbb{E}I_n[a,b] = \sum_{k=1}^{n} \mathbb{E}\left(W_{a+kh} - W_{a+(k-1)h}\right)^2 = \sum_{k=1}^{n} h = nh = b - a.$$

Further, the Brownian increments $W_{a+kh} - W_{a+(k-1)h}$ are independent $N(0,h)$ random variables. We shall define independent $N(0,1)$ random variables $Z_1, Z_2, \ldots, Z_n$ by

$$W_{a+kh} - W_{a+(k-1)h} = \sqrt{h} Z_k, \qquad 1 \le k \le n.$$

Hence

$$\begin{aligned}
\operatorname{var} I_n[a,b] &= \sum_{k=1}^{n} \operatorname{var}\left(\sqrt{h} Z_k\right)^2 \\
&= \sum_{k=1}^{n} \operatorname{var}\left[h Z_k^2\right] \\
&= \sum_{k=1}^{n} h^2 \operatorname{var}\left[Z_k^2\right] \\
&= \sum_{k=1}^{n} h^2 \left(\mathbb{E} Z_k^4 - \left[\mathbb{E} Z_k^2\right]^2\right) \\
&= \sum_{k=1}^{n} 2h^2 \\
&= 2nh^2 \\
&= 2(b-a)^2/n.
\end{aligned}$$

□

With this in mind, we *define*

$$\int_a^b (dW_t)^2 = b - a$$

and observe that we have shown that

$$\int_a^b (dW_t)^2 = \int_a^b dt,$$

for any $0 \le a < b$. Thus we have really shown the Itô rule

$$dW_t^2 = dt.$$

Using a very similar technique, we can also prove that

$$dt dW_t = 0.$$

We first define

$$J_n[a,b] = \sum_{k=1}^{n} h\left(W_{a+kh} - W_{a+(k-1)h}\right),$$

where $nh = b - a$, as before.

Proposition 3.10. *We have* $\mathbb{E} J_n[a,b] = 0$ *and* $\operatorname{var} J_n[a,b] = (b-a)^3/n^2$.

Proof. Firstly,

$$\mathbb{E} J_n[a,b] = \sum_{k=1}^{n} \mathbb{E} h\left(W_{a+kh} - W_{a+(k-1)h}\right) = 0.$$

The variance satisfies

$$\begin{aligned}
\operatorname{var} J_n[a,b] &= \sum_{k=1}^{n} \operatorname{var} h\left(W_{a+kh} - W_{a+(k-1)h}\right) \\
&= \sum_{k=1}^{n} h^2 \operatorname{var}\left(W_{a+kh} - W_{a+(k-1)h}\right) \\
&= \sum_{k=1}^{n} h^3 \\
&= nh^3 \\
&= (b-a)^3/n^2.
\end{aligned}$$

□

With this in mind, we *define*

$$\int_a^b dt dW_t = 0, \qquad \text{for any } 0 \le a < b,$$

and observe that we have shown that

$$dt dW_t = 0.$$

Exercise 3.3. *Setting $nh = b - a$, define*

$$K_n[a,b] = \sum_{k=1}^{n} h^2.$$

Prove that $K_n[a,b] = (b-a)^2/n \to 0$, as $n \to \infty$. Thus

$$\int_a^b (dt)^2 = 0,$$

for any $0 \le a < b$. Hence we have $(dt)^2 = 0$.

Proposition 3.11 (Itô Rules). *We have $dW_t^2 = dt$ and $dW_t dt = dt^2 = 0$.*

Proof. See Propositions 3.9, 3.10 and Exercise 3.3 □

The techniques used in Propositions 3.9 and 3.10 are crucial examples of the basics of stochastic integration. We can generalize this technique to compute other useful stochastic integrals, as we shall now see. However, computing these stochastic integrals directly from limits of stochastic sums is cumbersome compared to direct use of the Itô rules: compare the proof of Proposition 3.12 to the simplicity of Example 3.3.

Proposition 3.12. *We have*

$$\int_0^t W_s dW_s = \frac{1}{2}\left(W_t^2 - t\right).$$

Proof. We have already seen that, when $h = t/n$,

$$\sum_{k=1}^{n} \left(W_{kh} - W_{(k-1)h}\right)^2 \to t, \tag{3.16}$$

as $n \to \infty$. Further, we shall use the telescoping sum

$$\sum_{k=1}^{n} \left(W_{kh}^2 - W_{(k-1)h}^2 \right) = W_{nh}^2 - W_0^2 = W_t^2. \tag{3.17}$$

Subtracting (3.16) from (3.17), we obtain

(3.18)

$$\sum_{k=1}^{n} \left[\left(W_{kh}^2 - W_{(k-1)h}^2 \right) - \left(W_{kh} - W_{(k-1)h} \right)^2 \right] = 2 \sum_{k=1}^{n} W_{(k-1)h} \left(W_{kh} - W_{(k-1)h} \right).$$

Now the LHS converges to $W_t^2 - t$, whilst the RHS converges to

$$2 \int_0^t W_s dW_s,$$

whence (3.12). □

Example 3.1. *Here we shall derive a useful formula for*

$$\int_0^t f(s) dW_s, \tag{3.19}$$

where f is continuously differentiable. The corresponding discrete stochastic sum is

$$S_n = \sum_{k=1}^{n} f(kh) \left(W_{kh} - W_{(k-1)h} \right) \tag{3.20}$$

where $nh = t$, as usual. The key trick is to introduce another telescoping sum:

$$\sum_{k=1}^{n} \left(f(kh) W_{kh} - f((k-1)h) W_{(k-1)h} \right) = f(t) W_t. \tag{3.21}$$

Subtracting (3.21) *from* (3.20) *we find*

$$\begin{aligned} S_n - f(t) W_t &= - \sum_{k=1}^{n} \left(f(kh) - f((k-1)h) \right) W_{(k-1)h} \\ &= - \sum_{k=1}^{n} \left(h f'(kh) + O(h^2) \right) W_{(k-1)h} \\ &\to - \int_0^t f'(s) W_s ds, \end{aligned} \tag{3.22}$$

as $n \to \infty$. Hence

$$\int_0^t f(s) dW_s = f(t) W_t - \int_0^t f'(s) W_s \, ds. \tag{3.23}$$

Exercise 3.4. *Modify the technique of Example 3.1 to prove that*

$$\mathbb{E} \left[\left(\int_0^t h(s) dW_s \right)^2 \right] = \int_0^t h(s)^2 \, ds. \tag{3.24}$$

This is the Itô isometry property*.*

We now come to Itô's lemma itself.

Lemma 3.13 (Itô's Lemma for univariate functions)**.** *If f is any infinitely differentiable univariate function and $X_t = f(W_t)$, then*

$$dX_t = f'(W_t)dW_t + \frac{1}{2}f^{(2)}(W_t)dt. \tag{3.25}$$

Proof. We have

$$\begin{aligned} X_{t+dt} &= f(W_{t+dt}) \\ &= f(W_t + dW_t) \\ &= f(W_t) + f'(W_t)dW_t + \frac{1}{2}f^{(2)}(W_t)dW_t^2 \\ &= X_t + f'(W_t)dW_t + \frac{1}{2}f^{(2)}(W_t)dt. \end{aligned}$$

Subtracting X_t from both sides gives (3.25). □

Example 3.2. *Let $X_t = e^{cW_t}$, where $c \in \mathbb{C}$. Then, setting $f(x) = \exp(cx)$ in Lemma 3.13, we obtain*

$$dX_t = X_t \left(cdW_t + \frac{1}{2}c^2 dt \right).$$

Example 3.3. *Let $X_t = W_t^2$. Then, setting $f(x) = x^2$ in Lemma 3.13, we obtain*

$$dX_t = 2W_t dW_t + dt.$$

If we integrate this from 0 to T, say, then we obtain

$$X_T - X_0 = 2\int_0^T W_t dW_t + \int_0^T dt,$$

or

$$W_T^2 = 2\int_0^T W_t dW_t + T,$$

that is

$$\int_0^T W_t dW_t = \frac{1}{2}\left(W_T^2 - T\right).$$

This is an excellent example of the Itô rules greatly simplifying direct calculation with stochastic sums, because it is much easier than the direct proof of Proposition 3.12.

Example 3.4. *Let $X_t = W_t^n$, where n can be any positive integer. Then, setting $f(x) = x^n$ in Lemma 3.13, we obtain*

$$dX_t = nW_t^{n-1}dW_t + \frac{1}{2}n(n-1)W_t^{n-2}dt.$$

We can also integrate Itô's Lemma, as follows.

Example 3.5. *Integrating* (3.25) *from a to b, we obtain*

$$\int_a^b dX_t = \int_a^b f'(W_t)dW_t + \frac{1}{2}\int_a^b f^{(2)}(W_t)dt,$$

i.e.

$$X_b - X_a = \int_a^b f'(W_t)dW_t + \frac{1}{2}\int_a^b f^{(2)}(W_t)dt.$$

Lemma 3.13 is not quite sufficient to deal with geometric Brownian motion, hence the following bivariate variant.

Lemma 3.14 (Itô's Lemma for bivariate functions)**.** *If $g(x_1, t)$, for $x_1, t \in \mathbb{R}$, is any infinitely differentiable function and $Y_t = g(W_t, t)$, then*

$$dY_t = \frac{\partial g}{\partial x_1}(W_t, t)dW_t + \left(\frac{1}{2}\frac{\partial^2 g}{\partial x_1^2}(W_t, t) + \frac{\partial g}{\partial t}(W_t, t)\right) dt. \tag{3.26}$$

Proof. We have

$$\begin{aligned}
Y_{t+dt} &= g(W_{t+dt}, t + dt) \\
&= g(W_t + dW_t, t + dt) \\
&= g(W_t, t) + \frac{\partial g}{\partial x_1}(W_t, t)dW_t + \frac{1}{2}\frac{\partial^2 g}{\partial x_1^2}(W_t, t)dW_t^2 + \frac{\partial g}{\partial t}(W_t, t)dt \\
&= g(W_t, t) + \frac{\partial g}{\partial x_1}(W_t, t)dW_t + \left(\frac{1}{2}\frac{\partial^2 g}{\partial x_1^2}(W_t, t) + \frac{\partial g}{\partial t}(W_t, t)\right) dt
\end{aligned}$$

Subtracting Y_t from both sides gives (3.26). □

Example 3.6. *Let $X_t = e^{\alpha+\beta t+\sigma W_t}$. Then, setting $f(x_1, t) = \exp(\alpha + \beta t + \sigma x_1)$ in Lemma 3.13, we obtain*

$$dX_t = X_t \left(\sigma dW_t + \left(\frac{1}{2}\sigma^2 + \beta\right) dt\right).$$

Example 3.7. *Let $X_t = e^{\alpha+(r-\sigma^2/2)t+\sigma W_t}$. Then, setting $\beta = r - \sigma^2/2$ in Example 3.6, we find*

$$dX_t = X_t (\sigma dW_t + r dt).$$

Exercise 3.5. *Let $X_t = W_t^2 - t$. Find dX_t.*

3.8. **Itô rules and SDEs.** Suppose now that the asset price S_t is given by the SDE

$$dS_t = S_t (\mu dt + \sigma dW_t), \tag{3.27}$$

that is, S_t is a geometric Brownian motion. Then the Itô rules imply that

$$(dS_t)^2 = \sigma^2 S_t^2 dt. \tag{3.28}$$

Hence, if we define $X_t = f(S_t)$, then

$$dX_t = f'(S_t)dS_t + \frac{1}{2}f^{(2)}(S_t)(dS_t)^2 = \sigma f'(S_t)S_t dW_t + dt \left(\mu f'(S_t)S_t + \frac{1}{2}\sigma^2 S_t^2 f^{(2)}(S_t)\right). \tag{3.29}$$

We illustrate this with the particularly important example of solving the SDE for geometric Brownian motion.

Example 3.8. *If $f(x) = \log x$, then $f'(x) = 1/x$, $f^{(2)}(x) = -1/x^2$ and (3.29) becomes*

$$dX_t = \sigma \frac{1}{S_t} S_t dW_t + dt \left(\mu \frac{1}{S_t} S_t + \frac{1}{2}\sigma^2 S_t^2 \left(\frac{-1}{S_t^2}\right)\right) = \sigma dW_t + dt \left(\mu - \sigma^2/2\right).$$

Integrating from t_0 to t_1, say, we obtain

$$X_{t_1} - X_{t_0} = \sigma\left(W_{t_1} - W_{t_0}\right) + \left(\mu - \sigma^2/2\right)\left(t_1 - t_0\right),$$

or

$$\log\frac{S_{t_1}}{S_{t_0}} = \sigma\left(W_{t_1} - W_{t_0}\right) + \left(\mu - \sigma^2/2\right)\left(t_1 - t_0\right).$$

Taking the exponential of both sides, we obtain

$$S_{t_1} = S_{t_0} e^{\left(r - \sigma^2/2\right)(t_1 - t_0) + \sigma\left(W_{t_1} - W_{t_0}\right)}.$$

3.9. **Multivariate Geometric Brownian Motion.** So far we have considered one asset only. In practice, we need to construct a multivariate GBM model that allows us to incorporate dependencies between assets via a covariance matrix. To do this, we first take a vector Brownian motion $\mathbf{W}_t \in \mathbb{R}^n$: its components are independent Brownian motions. Its covariance matrix C_t at time t is simply a multiple of the identity matrix:

$$C_t = \mathbb{E}\mathbf{W}_t\mathbf{W}_t^T = tI.$$

Now take any real, invertible, symmetric $n \times n$ matrix A and define

$$\mathbf{Z}_t = A\mathbf{W}_t.$$

The covariance matrix D_t for this new stochastic process is given by

$$D_t = \mathbb{E}\mathbf{Z}_t\mathbf{Z}_t^T = \mathbb{E}A\mathbf{W}_t\mathbf{W}_t^T A = A\left(\mathbb{E}\mathbf{W}_t\mathbf{W}_t^T\right)A = tA^2,$$

and A^2 is a symmetric positive definite matrix.

Exercise 3.6. *Prove that A^2 is symmetric positive definite if A is real, symmetric and invertible.*

In practice, we calculate the covariance matrix M from historical data, hence must construct a symmetric A satisfying $A^2 = M$. Now a covariance matrix is precisely a symmetric positive definite matrix, so that the following linear algebra is vital. We shall use $\|\mathbf{x}\|$ to denote the Euclidean norm of the vector $\mathbf{x} \in \mathbb{R}^n$, that is

$$\|\mathbf{x}\| = \left(\sum_{k=1}^n x_k^2\right)^{1/2}, \qquad \mathbf{x} \in \mathbb{R}^n. \tag{3.30}$$

Further, great algorithmic and theoretical importance attaches to those $n\times n$ matrices which preserve the Euclidean norm. More formally, an $n \times n$ matrix Q is called *orthogonal* if $\|Q\mathbf{x}\| = \|\mathbf{x}\|$, for all $\mathbf{x} \in \mathbb{R}^n$. It turns out that Q is an orthogonal matrix if and only if $Q^TQ = I$, which is equivalent to stating that its columns are orthonormal vectors. See Section 7 for further details.

Theorem 3.15. *Let $M \in \mathbb{R}^{n\times n}$ be symmetric. Then it can be written as $M = QDQ^T$, where Q is an orthogonal matrix and D is a diagonal matrix. The elements of D are the eigenvalues of M, while the columns of Q are the eigenvectors. Further, if M is positive definite, then its eigenvalues are all positive.*

Proof. Any good linear algebra textbook should include a proof of this fact; a proof is given in my numerical linear algebra notes. □

Given the *spectral decomposition* $M = QDQ^T$, with $D = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$, we define

$$D^{1/2} = \text{diag}\,(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2})$$

when M is positive definite. We can now define the *matrix square-root* $M^{1/2}$ by

$$M^{1/2} = QD^{1/2}Q^T. \tag{3.31}$$

Exercise 3.7. *Prove that* $(M^{1/2})^2 = M$ *directly from* (3.31).

Given the matrix square-root $M^{1/2}$ for a chosen symmetric. positive definite matrix M, we now define the assets

$$S_k(t) = e^{(r - M_{kk}/2)t + \left(M^{1/2}\mathbf{W}_t\right)_k}, \qquad k = 1, 2, \ldots, n, \tag{3.32}$$

where $\left(M^{1/2}\mathbf{W}_t\right)_k$ denotes the kth element of the vector $M^{1/2}\mathbf{W}_t$. We now need to check that our assets remain risk-neutral.

Proposition 3.16. *Let the assets' stochastic processes be defined by* (3.32). *Then*

$$\mathbb{E}S_k(t) = e^{rt},$$

for all $k \in \{1, 2, \ldots, n\}$.

Proof. The key calculation is

$$\begin{aligned}
\mathbb{E}e^{\left(M^{1/2}\mathbf{W}_t\right)_k} &= \mathbb{E}e^{\sum_{\ell=1}^n (M^{1/2})_{k\ell} W_t(\ell)} \\
&= \mathbb{E}\prod_{\ell=1}^n e^{(M^{1/2})_{k\ell} W_t(\ell)} \\
&= \prod_{\ell=1}^n \mathbb{E}e^{(M^{1/2})_{k\ell} W_t(\ell)} \\
&= \prod_{\ell=1}^n e^{(M^{1/2})^2_{k\ell} t/2} \\
&= e^{(t/2)\sum_{\ell=1}^n (M^{1/2})^2_{k\ell}} \\
&= e^{(t/2)M_{kk}},
\end{aligned}$$

using the independence of the components of $\mathbf{W}_t$. □

Exercise 3.8. *Compute* $\mathbb{E}[S_k(t)^2]$.

Exercise 3.9. *What's the covariance matrix for the assets* $S_1(t), \ldots, S_n(t)$*?*

In practice, it is usually easier to describe the covariance structure of multivariate Brownian motion via the Itô rules, which take the simple form

$$d\mathbf{W}_t d\mathbf{W}_t^T = M\ dt, \tag{3.33}$$

where $M \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix and $\mathbf{W}_t$ is an n-dimensional Brownian motion.

Proposition 3.17. *If* $X_t = f(\mathbf{W}_t)$, *then*

$$dX_t = \nabla f(\mathbf{W}_t)^T d\mathbf{W}_t + \frac{1}{2} d\mathbf{W}_t^T D^2 f(\mathbf{W}_t) d\mathbf{W}_t \tag{3.34}$$

or

$$dX_t = \sum_{j=1}^n \frac{\partial f}{\partial x_j} dW_{j,t} + \frac{dt}{2}\sum_{j=1}^n\sum_{k=1}^n \frac{\partial^2 f}{\partial x_j \partial x_k} M_{jk}. \tag{3.35}$$

Proof. This is left as an exercise. □

Example 3.9. *If $n = 2$ and*

$$f(x_1, x_2) = e^{a_1 x_1 + a_2 x_2},$$

and the correlated Brownian motions $W_{1,t}$ and $W_{2,t}$ satisfy

$$dW_{1,t} dW_{2,t} = \rho dt,$$

for some constant correlation coefficient $\rho \in [-1, 1]$, then $X_t = f(W_{1,t}, W_{2,t})$ satisfies

$$dX_t = \left(a_1 dW_{1,t} + a_2 dW_{2,t} + \frac{1}{2} dt \left(a_1^2 + 2\rho a_1 a_2 + a_2^2 \right) \right) X_t.$$

Example 3.10. *If $n = 3$ and*

$$f(x_1, x_2, x_3) = e^{a_1 x_1 + a_2 x_2 + a_3 x_3},$$

and the correlated Brownian motions $W_{1,t}, W_{2,t}, W_{3,t}$ satisfy

$$dW_{1,t} dW_{2,t} = M_{12} dt, \quad dW_{2,t} dW_{3,t} = M_{23} dt, \quad dW_{3,t} dW_{1,t} = M_{31} dt,$$

where $M \in \mathbb{R}^{3\times 3}$ is a symmetric positive definite matrix which also satisfies

$$M_{11} = M_{22} = M_{33} = 1,$$

then $X_t = f(W_{1,t}, W_{2,t}, W_{3,t})$ satisfies

$$dX_t = \left(a_1 dW_{1,t} + a_2 dW_{2,t} + a_3 dW_{3,t} + \frac{1}{2} dt \left(a_1^2 + a_2^2 + a_3^2 + 2a_2 a_3 M_{23} + 2a_3 a_1 M_{31} + 2a_1 a_2 M_{12} \right) \right) X_t.$$

Example 3.11. *If*

$$f(\mathbf{x}) = e^{\mathbf{a}^T \mathbf{x}}, \qquad \mathbf{x} \in \mathbb{R}^n,$$

then

$$\nabla f(\mathbf{x}) = \mathbf{a} f(\mathbf{x})$$

and

$$D^2 f(\mathbf{x}) = \mathbf{a}\mathbf{a}^T f(\mathbf{x}).$$

Let $\mathbf{W}_t$ be any n-dimensional Brownian motion satisfying

$$d\mathbf{W}_t d\mathbf{W}_t^T = M dt,$$

where $M \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. Then $X_t = f(\mathbf{W}_t)$ satisfies

$$dX_t = \left(\mathbf{a}^T d\mathbf{W}_t + \frac{1}{2} d\mathbf{W}_t^T M d\mathbf{W}_t \right) f(\mathbf{x}),$$

or, in coordinate form,

$$dX_t = \left(\sum_{j=1}^{n} a_j dW_{j,t} + \frac{1}{2} dt \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k M_{jk} \right) f(\mathbf{x}).$$

Proposition 3.18. *If $Y_t = g(\mathbf{W}_t, t)$, then*

$$dY_t = \nabla g(\mathbf{W}_t)^T d\mathbf{W}_t + \frac{\partial g}{\partial t} dt + \frac{1}{2} d\mathbf{W}_t^T D^2 g(\mathbf{W}_t) d\mathbf{W}_t \tag{3.36}$$

or

$$dY_t = \sum_{j=1}^{n} \frac{\partial g}{\partial x_j} dW_{j,t} + \frac{\partial g}{\partial t} dt + \frac{dt}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_k} M_{jk}. \tag{3.37}$$

Proof. Exercise. □

Example 3.12. *If $n = 2$ and*

$$g(x_1, x_2) = e^{a_1 x_1 + a_2 x_2 + bt},$$

and the correlated Brownian motions $W_{1,t}$ and $W_{2,t}$ satisfy

$$dW_{1,t} dW_{2,t} = \rho dt,$$

for some constant correlation coefficient $\rho \in [-1, 1]$, then $X_t = f(W_{1,t}, W_{2,t})$ satisfies

$$dX_t = \left(a_1 dW_{1,t} + a_2 dW_{2,t} + b dt + \frac{1}{2} dt \left(a_1^2 + 2\rho a_1 a_2 + a_2^2 \right) \right) X_t.$$

Example 3.13. *If*

$$g(\mathbf{x}, t) = e^{\mathbf{a}^T \mathbf{x} + bt}, \qquad \mathbf{x} \in \mathbb{R}^n,$$

then

$$\nabla g(\mathbf{x}, t) = \mathbf{a} g(\mathbf{x}, t), \quad \frac{\partial g}{\partial t} = b g(\mathbf{x}, t),$$

and

$$D^2 g(\mathbf{x}, t) = \mathbf{a}\mathbf{a}^T g(\mathbf{x}, t).$$

Let W_t be any n-dimensional Brownian motion satisfying

$$dW_t dW_t^T = M dt,$$

where $M \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. Then $Y_t = g(\mathbf{W}_t, t)$ satisfies

$$dY_t = \left(\mathbf{a}^T d\mathbf{W}_t + b dt + \frac{1}{2} d\mathbf{W}_t^T M d\mathbf{W}_t \right) Y_t,$$

or, in coordinate form

$$dY_t = \left(\sum_{j=1}^{n} a_j dW_{j,t} + b dt + \frac{1}{2} dt \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k M_{jk} \right) Y_t.$$

3.10. **Asian Options.** European Put and Call options provide a useful laboratory in which to understand and test methods. However, the main aim of Monte Carlo is to calculate option prices for which there is no convenient analytic formula. We shall illustrate this with *Asian options*. Specifically, we shall consider the option

$$f_A(S,T) = \left(S(T) - \frac{1}{T}\int_0^T S(\tau)\,d\tau\right)_+ . \tag{3.38}$$

This is a *path dependent* option: its value depends on the history of the asset price, not simply its final value.

Why would anyone trade Asian options? Consider a bank's corporate client trading in, say, Britain and the States. The client's business is exposed to exchange rate volatility: the pound's value in dollars varies over time. Therefore the client may well decide to *hedge* by buying an option to trade dollars for pounds at a set rate at time T. This can be an expensive contract for the writer of the option, because currency values can "blip". An alternative contract is to make the exchange rate at time T a *time-average*, as in (3.38). Any contract containing time-averages of asset prices is usually called an Asian option, and there are many variants of these. For example, the option dual to (3.38) (in the sense that a call option is dual to a put option) is given by

$$g_A(S,T) = \left(\frac{1}{T}\int_0^T S(\tau)\,d\tau - S(T)\right)_+ . \tag{3.39}$$

Pricing (3.38) via Monte Carlo is fairly simple. We choose a positive integer M and subdivide the time interval $[0,T]$ into M equal subintervals. We evolve the asset price using the equation

$$S\left(\frac{(k+1)T}{M}\right) = S\left(\frac{kT}{M}\right)e^{(r-\sigma^2/2)\frac{T}{M}+\sigma\sqrt{\frac{T}{M}}Z_k}, \qquad k = 0,1,\ldots,M-1, \tag{3.40}$$

where $Z_0, Z_1, \ldots, Z_{M-1}$ are independent $N(0,1)$ independent pseudorandom numbers. We can use the simple approximation

$$T^{-1}\int_0^T S(\tau)\,d\tau \approx M^{-1}\sum_{k=0}^{M-1} S\left(\frac{kT}{M}\right).$$

Exercise 3.10. *Write a Matlab program to price the discrete Asian option defined by*

$$f_M(S,T) = \left(S(T) - M^{-1}\sum_{k=0}^{M-1} S(kT/M)\right)_+ . \tag{3.41}$$

We can also study the average

$$A(T) = T^{-1}\int_0^T S(t)\,dt \tag{3.42}$$

directly, and this is the subject of a recent paper of Raymond and myself. For example,

$$\mathbb{E}A(T) = T^{-1}\int_0^T \mathbb{E}S(t)\,dt. \tag{3.43}$$

Exercise 3.11. *Prove that*

$$\mathbb{E}A(T) = S(0)\left(\frac{e^{rT}-1}{rT}\right). \tag{3.44}$$

Exercise 3.12. *In a similar vein, find expressions for* $\mathbb{E}S(a)S(b)$ *and* $\mathbb{E}\left(A(T)^2\right)$.

3.11. **The Ornstein–Uhlenbeck Process.** This interesting SDE displays mean-reversion and requires a slightly more advanced technique. We consider the SDE

$$dX_t = -\alpha X_t dt + \sigma dW_t, \qquad t \geq 0, \tag{3.45}$$

where $\alpha > 0$ and $\sigma \geq 0$ are constants and $X_0 = x_0$.

It's very useful to consider the special case $\sigma = 0$ first, in which case the SDE (3.45) becomes the ODE

$$\frac{dX_t}{dt} + \alpha X_t = 0. \tag{3.46}$$

There is a standard method for solving (3.46) using an *integrating factor*. Specifically, if we multiply (3.46) by $\exp(\alpha t)$, then we obtain

$$\frac{d}{dt}\left(X_t e^{\alpha t}\right) = 0, \tag{3.47}$$

so that $X_t \exp(\alpha t)$ is constant. Hence, recalling the initial condition $X_0 = x_0$, we must have

$$X_t = x_0 e^{-\alpha t}. \tag{3.48}$$

Thus the solution decays exponentially to zero, at a rate determined by the positive constant α, for any initial value x_0.

Fortunately the integrating factor method also applies to the $\sigma > 0$ case, with a little more work. Multiplying (3.45) by $\exp \alpha t$, we obtain

$$e^{\alpha t}\left(dX_t + \alpha X_t dt\right) = \sigma e^{\alpha t} dW_t,$$

or

$$d\left(X_t e^{\alpha t}\right) = \sigma e^{\alpha t} dW_t, \tag{3.49}$$

using the infinitesimal increments variant on the product rule for differentiation. Integrating (3.49) from 0 to s, we find

$$X_s e^{\alpha s} - x_0 = \int_0^s d\left(X_t e^{\alpha t}\right) = \sigma \int_0^s e^{\alpha t} dW_t, \tag{3.50}$$

or

$$X_s = x_0 e^{-\alpha s} + e^{-\alpha s} \int_0^s e^{\alpha t} dW_t. \tag{3.51}$$

We can say more using the following important property of stochastic integrals.

Proposition 3.19. *Let $f : [0, \infty) \to \mathbb{R}$ be any infinitely differentiable function and define the stochastic process*

$$F_s = \int_0^s f(t) dW_t. \tag{3.52}$$

Then $\mathbb{E} F_s = 0$ and

$$\operatorname{var} F_s = \int_0^s f(t)^2 dt. \tag{3.53}$$

Proof. The key point is that (3.52) is the limit of the stochastic sum

$$S_n = \sum_{k=1}^{n} f(kh)\left(W_{kh} - W_{(k-1)h}\right),$$

where $h > 0$ and $nh = s$. Now the increments $W_{kh} - W_{(k-1)h}$ are independent and satisfy

$$W_{kh} - W_{(k-1)h} \sim N(0, h),$$

by the axioms of Brownian motion, so

$$\mathbb{E} S_n = 0,$$

for all n. By independence of the terms in the sum, we see that

$$\begin{aligned}
\operatorname{var} S_n &= \sum_{k=1}^{n} \operatorname{var}\left(f(kh)\left(W_{kh} - W_{(k-1)h}\right)\right) \\
&= \sum_{k=1}^{n} f(kh)^2 \operatorname{var}\left(W_{kh} - W_{(k-1)h}\right) \\
&= h \sum_{k=1}^{n} f(kh)^2 \\
&\to \int_0^s f(t)^2\, dt,
\end{aligned}$$

as $n \to \infty$. $\square$

Applying Proposition 3.19 to the Ornstein–Uhlenbeck process solution (3.51), we obtain $\mathbb{E} X_s = x_0 \exp(-\alpha s)$ and

$$\begin{aligned}
\operatorname{var} X_s &= \operatorname{var} \sigma e^{-\alpha s} \int_0^s e^{\alpha t} dW_t \\
&= \sigma^2 e^{-2\alpha s} \int_0^s e^{2\alpha t} dt \\
&= \sigma^2 e^{-2\alpha s} \left(\frac{e^{2\alpha s} - 1}{2\alpha}\right) \\
&= \sigma^2 \left(\frac{1 - e^{-2\alpha s}}{2\alpha}\right).
\end{aligned}$$

3.12. **Feynman–Kac.** The derivation of the Black–Scholes PDE earlier is really the first example of a much more general link between expectations of functions of Brownian motion and PDEs.

If we consider the stochastic process $X_t = u(W_t, t)$, then Itô's Lemma 3.14 states that

$$du(W_t, t) = u_x dW_t + \left(u_t + \frac{1}{2}u_{xx}\right) dt. \tag{3.54}$$

Thus, if u satisfies the PDE

$$u_t + \frac{1}{2}u_{xx} = 0, \tag{3.55}$$

then the dt component vanishes in (3.54), implying

$$du(W_t, t) = u_x(W_t, t)dW_t. \tag{3.56}$$

If we now choose any times $0 \le t_0 < t_1$, then integrating (3.56) yields

$$u(W_{t_1}, t_1) - u(W_{t_0}, t_0) = \int_{t_0}^{t_1} du(W_t, t) = \int_{t_0}^{t_1} u_x(W_t, t)dW_t. \tag{3.57}$$

Taking the expectation of (3.57), conditioned on $W_{t_0} = X$, say, we obtain

$$\mathbb{E}\left(u(W_{t_1}, t_1)|W_{t_0} = X\right) - u(X, t_0) = \mathbb{E}\left(\int_{t_0}^{t_1} u_x(W_t, t)\, dW_t | W_{t_0} = X\right) = 0, \tag{3.58}$$

by the independent increments property of Brownian motion. In other words, the solution to the PDE (3.55) is given by

$$u(X, t_0) = \mathbb{E}\left(u(W_{t_1}, t_1)|W_{t_0} = X\right). \tag{3.59}$$

Now

$$W_{t_1} = W_{t_1} - W_{t_0} + W_{t_0} = W_{t_1} - W_{t_0} + X,$$

so we can rewrite (3.59) as

$$u(X, t_0) = \mathbb{E}u(X + W_{t_1} - W_{t_0}, t_1), \tag{3.60}$$

or

$$u(X, t_0) = \mathbb{E}u(X + \sqrt{t_1 - t_0}Z, t_1), \tag{3.61}$$

where $Z \sim N(0, 1)$. Now that we have derived (3.61), it's clearer to replace t_0 by t, t_1 by T and X by x, respectively, to obtain

$$u(x, t) = \mathbb{E}u(x + \sqrt{\tau}Z, T), \tag{3.62}$$

where $\tau = T - t$ is the time to expiry. The key point here is that we have a solution to the PDE (3.55) (which is called the time-reversed diffusion equation) expressed as an expectation. If we express this expectation in terms of the $N(0, 1)$ PDF, then we obtain

$$u(x, t) = \int_{-\infty}^{\infty} u(x + \sqrt{\tau}s, T)(2\pi)^{-1/2}e^{-s^2/2}\, ds. \tag{3.63}$$

Example 3.14. *Suppose the expiry value of u is given by*

$$u(y, T) = \begin{cases} 1 & \text{if } a \le y \le b, \\ 0 & \text{otherwise.} \end{cases} \tag{3.64}$$

Then (3.63) *implies that the solution of* (3.55) *is given by*

$$
\begin{aligned}
u(x,t) &= \int_{-\infty}^{\infty} u(x+\sqrt{\tau}s, T)(2\pi)^{-1/2}e^{-s^2/2}\,ds \\
&= \int_{\frac{a-x}{\sqrt{\tau}}}^{\frac{b-x}{\sqrt{\tau}}} (2\pi)^{-1/2}e^{-s^2/2}\,ds \\
&= \Phi\left(\frac{b-x}{\sqrt{\tau}}\right) - \Phi\left(\frac{a-x}{\sqrt{\tau}}\right),
\end{aligned}
$$

because

$$a \le x + \sqrt{\tau}s \le b$$

if and only if

$$\frac{a-x}{\sqrt{\tau}} \le x \le \frac{b-x}{\sqrt{\tau}}.$$

Exercise 3.13. *Show that, in terms of the time to expiry* $\tau = T - t$*, the PDE* (3.55) *becomes the* **diffusion equation**

$$u_\tau = \frac{1}{2}u_{xx},$$

with solution

$$u(x,\tau) = \mathbb{E}u(x + \sqrt{\tau}Z, 0).$$

3.13. **Feynman–Kac and Black–Scholes I.** Suppose we consider the stochastic process S_t defined by the SDE

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t, \tag{3.65}$$

which is, of course, geometric Brownian motion. We can solve this SDE using the technique of Example 3.8.

Example 3.15. *The stochastic process* $\log S_t$ *satisfies the SDE*

$$d\log S_t = \left(r - \sigma^2/2\right)dt + \sigma dW_t, \tag{3.66}$$

whence, integrating from t_0 *to* t_1*, we obtain*

$$\log\frac{S_{t_1}}{S_{t_0}} = \left(r - \sigma^2/2\right)(t_1 - t_0) + \sigma\left(W_{t_1} - W_{t_0}\right). \tag{3.67}$$

Taking the exponential, we find

$$S_{t_1} = S_{t_0}e^{\left(r-\sigma^2/2\right)(t_1-t_0)+\sigma\left(W_{t_1}-W_{t_0}\right)}. \tag{3.68}$$

Applying Itô's Lemma to the stochastic process $V(S_t,t)$, where S_t is defined by the SDE (3.65), we obtain

$$dV(S_t,t) = \left(V_t + \frac{1}{2}\sigma^2 S_t^2 V_{SS} + rS_tV_S\right)dt + \sigma S_t V_S dW_t. \tag{3.69}$$

Hence, if $V(S,t)$ satisfies the PDE

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S = 0, \tag{3.70}$$

together with the boundary condition

$$V(S,T) = F(S), \tag{3.71}$$

for some known function $F(S)$, then integrating (3.69) from $t = t_0$ to $t = T$, we find

$$V(S_T,T) - V(S_{t_0},t_0) = \int_{t_0}^{T} \sigma S_t V_S dW_t, \tag{3.72}$$

or, using the boundary condition (3.71),

$$F(S_T) - V(S_{t_0},t_0) = \int_{t_0}^{T} \sigma S_t V_S dW_t, \tag{3.73}$$

If we now take the expectation of (3.72), conditioned on $S_{t_0} = S$, then

$$\mathbb{E}\left(F(S_T)|S_{t_0} = S\right) - V(S,t_0) = 0, \tag{3.74}$$

the RHS vanishing because of the independent increments property of Brownian motion. Now, setting $t_0 = t$ and $t_1 = T$ in (3.68), we can rewrite (3.74) as

$$V(S,t) = \mathbb{E}F(Se^{(r-\sigma^2/2)\tau+\sigma\tau^{1/2}Z}), \tag{3.75}$$

or

$$V(S,t) = \mathbb{E}V(Se^{(r-\sigma^2/2)\tau+\sigma\tau^{1/2}Z},T), \tag{3.76}$$

where

$$\tau = T - t \tag{3.77}$$

and $Z \sim N(0,1)$.

3.14. **Feynman–Kac and Black–Scholes II.** To obtain Black–Scholes from Feynman–Kac, we substitute

$$V(S,t) = e^{-rt}U(S,t) \tag{3.78}$$

in (3.70). Now

$$V_S = e^{-rt}U_S, \quad V_{SS} = e^{-rt}U_{SS} \quad \text{and} \quad V_t = e^{-rt}\left(U_t - rU\right),$$

so (3.70) becomes the Black–Scholes equation

$$U_t + \frac{1}{2}\sigma^2 S^2 U_{SS} + rSU_S - rU = 0. \tag{3.79}$$

Hence (3.76) becomes

$$e^{-rt}U(S,t) = e^{-rT}\mathbb{E}U(Se^{(r-\sigma^2/2)\tau+\sigma\tau^{1/2}Z},T), \tag{3.80}$$

or

$$U(S,t) = e^{-r\tau}\mathbb{E}U(Se^{(r-\sigma^2/2)\tau+\sigma\tau^{1/2}Z},T). \tag{3.81}$$

4. The Binomial Model Universe

The geometric Brownian Motion universe is an infinite one and, for practitioners, has the added disadvantage of the mathematical difficulty of Brownian motion. It is also possible to construct finite models with similar properties. This was first demonstrated by Cox, Ross and Rubinstein in the late 1970s.

Our model will be entirely specified by two parameters, $\alpha > 0$ and $p \in [0, 1]$. We choose $S_0 > 0$ and define

$$S_k = S_{k-1} \exp(\alpha X_k), \qquad k > 0, \tag{4.1}$$

where the independent random variables $X_1, X_2, \ldots$ satisfy

$$\mathbb{P}\left(X_k = 1\right) = p, \qquad \mathbb{P}\left(X_k = -1\right) = 1 - p =: q. \tag{4.2}$$

Thus

$$S_m = S_0 e^{\alpha(X_1 + X_2 + \cdots + X_m)}, \qquad m > 0. \tag{4.3}$$

It is usual to display this random process graphically.

$$\begin{tikzcd}
e^{4\alpha} & & & & \\
e^{3\alpha} \arrow[u, "p"] \arrow[r, "q"] & e^{2\alpha} & & & \\
e^{2\alpha} \arrow[u, "p"] \arrow[r, "q"] & e^{\alpha} \arrow[u, "p"] \arrow[r, "q"] & 1 & & \\
e^{\alpha} \arrow[u, "p"] \arrow[r, "q"] & 1 \arrow[u, "p"] \arrow[r, "q"] & e^{-\alpha} \arrow[u, "p"] \arrow[r, "q"] & e^{-2\alpha} & \\
1 \arrow[u, "p"] \arrow[r, "q"] & e^{-\alpha} \arrow[u, "p"] \arrow[r, "q"] & e^{-2\alpha} \arrow[u, "p"] \arrow[r, "q"] & e^{-3\alpha} \arrow[u, "p"] \arrow[r, "q"] & e^{-4\alpha}
\end{tikzcd}$$

At this stage, we haven't specified p and α. However, we can easily price a European option given these parameters. If S_k denotes our Binomial Model asset price at time kh, for some positive time interval h, then the Binomial Model European option requirement is given by

$$\begin{aligned}
f(S_{k-1}, (k-1)h) &= e^{-rh}\mathbb{E} f(S_{k-1} e^{\alpha X_k}, kh) \\
&= e^{-rh}\left(pf(S_{k-1}e^{\alpha}, kh) + (1-p)f(S_{k-1}e^{-\alpha}, kh)\right). \qquad (4.4)
\end{aligned}$$

Thus, given the $m+1$ possible asset prices at expiry time mh, and their corresponding option prices, we use (4.4) to calculate the m possible values of the option at time $(m-1)h$. Recurring this calculation provides the value of the option at time 0. Let's illustrate this with an example.

Example 4.1. *Suppose $e^{\alpha} = 2$, $p = 1/2$ and $D = e^{-rh}$. Let $m = 4$ and let's use the Binomial Model to calculate all earlier values of the call option whose expiry value is*

$$f(S(mh), mh) = (S(mh) - 1)_+ \,.$$

Using (4.4), *we obtain the following diagram for the asset prices.*

$$\begin{tikzcd}
16 \\
8 \arrow[u, "p"] \arrow[r, "q"] & 4 \\
4 \arrow[u, "p"] \arrow[r, "q"] & 2 \arrow[u, "p"] \arrow[r, "q"] & 1 \\
2 \arrow[u, "p"] \arrow[r, "q"] & 1 \arrow[u, "p"] \arrow[r, "q"] & 1/2 \arrow[u, "p"] \arrow[r, "q"] & 1/4 \\
1 \arrow[u, "p"] \arrow[r, "q"] & 1/2 \arrow[u, "p"] \arrow[r, "q"] & 1/4 \arrow[u, "p"] \arrow[r, "q"] & 1/8 \arrow[u, "p"] \arrow[r, "q"] & 1/16
\end{tikzcd}$$

The corresponding diagram for the option prices is as follows.

$$\begin{tikzcd}
15 \arrow[d] \\
9D \arrow[d] & 3 \arrow[l] \arrow[d] \\
21D^2/4 \arrow[d] & 3D/2 \arrow[l] \arrow[d] & 0 \arrow[l] \arrow[d] \\
3D^3 \arrow[d] & 3D^2/4 \arrow[l] \arrow[d] & 0 \arrow[l] \arrow[d] & 0 \arrow[l] \arrow[d] \\
27D^4/16 & 3D^3/8 \arrow[l] & 0 \arrow[l] & 0 \arrow[l] & 0 \arrow[l]
\end{tikzcd}$$

How do we choose the constants α and p? One way is to use them to mimic geometric Brownian motion. Thus we choose a positive number h and use

$$S(kh) = S((k-1)h)e^{\left(r-\sigma^2/2\right)h+\sigma\sqrt{h}Z}, \qquad k > 0, \tag{4.5}$$

where, as usual $Z \sim N(0,1)$.

Lemma 4.1. *In the Geometric Brownian Motion Universe, we have*

$$\mathbb{E}S(kh)|S((k-1)h) = S((k-1)h)e^{rh} \tag{4.6}$$

and

$$\mathbb{E}S(kh)^2|S((k-1)h) = S((k-1)h)^2e^{(2r+\sigma^2)h}. \tag{4.7}$$

Proof. These are easy exercises if you have digested Lemma 2.1 and Lemma 2.2: everything rests on using the fact that $\mathbb{E}\exp(cZ) = c^2/2$ when $Z \sim N(0,1)$, for any real (or complex) number c. □

There are analogous quantities in the Binomial Model.

Lemma 4.2. *In the Binomial Model Universe, we have*

$$\mathbb{E}S_k|S_{k-1} = \left(pe^{\alpha} + (1-p)e^{-\alpha}\right)S_{k-1} \tag{4.8}$$

and

$$\mathbb{E}S_k^2|S_{k-1} = \left(pe^{2\alpha} + (1-p)e^{-2\alpha}\right)S_{k-1}^2 \tag{4.9}$$

Proof. You should find these to be very easy given the definitions (4.1), (4.2) and (4.3); revise elementary probability theory if this is not so! □

One way to choose p and α is to require that the right hand sides of (4.6), (4.8) and (4.7), (4.9) agree, that is,

$$e^{rh} = pe^{\alpha} + (1-p)e^{-\alpha}, \tag{4.10}$$

$$e^{(2r+\sigma^2)h} = pe^{2\alpha} + (1-p)e^{-2\alpha}. \tag{4.11}$$

This ensures that our Binomial Model preserves risk neutrality.

Rearranging (4.10) and (4.11), we find

$$p = \frac{e^{rh} - e^{-\alpha}}{e^{\alpha} - e^{-\alpha}} = \frac{e^{(2r+\sigma^2)h} - e^{-2\alpha}}{e^{2\alpha} - e^{-2\alpha}}. \tag{4.12}$$

Further, the elementary algebraic identity

$$e^{2\alpha} - e^{-2\alpha} = \left(e^{\alpha} + e^{-\alpha}\right)\left(e^{\alpha} - e^{-\alpha}\right)$$

transforms (4.12) into

$$e^{rh} - e^{-\alpha} = \frac{e^{(2r+\sigma^2)h} - e^{-2\alpha}}{e^{\alpha} + e^{-\alpha}}. \tag{4.13}$$

Exercise 4.1. *Show that* (4.12) *implies the equation*

$$e^{\alpha} + e^{-\alpha} = e^{(r+\sigma^2)h} + e^{-rh}, \tag{4.14}$$

How do we solve (4.14)? The following analysis is a standard part of the theory of hyperbolic trigonometric functions[1], but no background knowledge will be assumed. If we write

$$y = \frac{1}{2}\left(e^{(r+\sigma^2)h} + e^{-rh}\right), \tag{4.15}$$

then (4.14) becomes

$$e^{\alpha} + e^{-\alpha} = 2y, \tag{4.16}$$

that is

$$(e^{\alpha})^2 - 2y(e^{\alpha}) + 1 = 0. \tag{4.17}$$

This quadratic in e^{α} has solutions

$$e^{\alpha} = y \pm \sqrt{y^2 - 1} \tag{4.18}$$

and, since (4.15) implies $y \geq 1$, we see that each of these possible solutions is positive. Thus the possible values for α are

$$\alpha_1 = \log_e\left(y + \sqrt{y^2 - 1}\right) \tag{4.19}$$

and

$$\alpha_2 = \log_e\left(y - \sqrt{y^2 - 1}\right). \tag{4.20}$$

Now

$$\begin{aligned} \alpha_1 + \alpha_2 &= \log_e\left[\left(y + \sqrt{y^2-1}\right)\right] + \log_e\left[\left(y - \sqrt{y^2-1}\right)\right] \\ &= \log_e\left[\left(y + \sqrt{y^2-1}\right)\left(y - \sqrt{y^2-1}\right)\right] \\ &= \log_e\left[y^2 - (y^2 - 1)\right] \\ &= \log_e 1 \\ &= 0. \end{aligned} \tag{4.21}$$

Since $y + \sqrt{y^2 - 1} \geq 1$, for $y \geq 1$, we deduce that $\alpha_1 \geq 0$ and $\alpha_2 = -\alpha_1$. Since we have chosen $\alpha > 0$, we conclude

$$\alpha = \log_e\left[y + \sqrt{y^2 - 1}\right], \tag{4.22}$$

where y is given by (4.15).

Now (4.22) tells us the value of α required, but the expression is somewhat complicated. However, if we return to (4.14), that is,

$$e^{\alpha} + e^{-\alpha} = e^{(r+\sigma^2)h} + e^{-rh},$$

and consider small h, then α is also small and Taylor expansion yields

$$2 + \alpha^2 + \cdots = 1 + (r + \sigma^2)h + \cdots + 1 - rh + \cdots,$$

that is,

$$\alpha^2 + \cdots = \sigma^2 h + \cdots. \tag{4.23}$$

[1]Specifically, this is the formula for the inverse hyperbolic cosine.

Cox and Ross had the excellent idea of ignoring the messy higher order terms, since the model is only an approximation in any case. Thus the Cox–Ross Binomial Model chooses

$$\alpha = \sigma h^{1/2}. \tag{4.24}$$

The corresponding equation for the probability p becomes

$$p = \frac{e^{rh} - e^{-\sigma h^{1/2}}}{e^{\sigma h^{1/2}} - e^{-\sigma h^{1/2}}} \tag{4.25}$$

It's useful, but tedious, to Taylor expand the RHS of (4.25). We obtain

$$\begin{aligned}
p &= \frac{1 + rh + \cdots - \left(1 - \sigma h^{1/2} + \frac{1}{2}\sigma^2 h + \cdots\right)}{2\left(\sigma h^{1/2} + \sigma^3 h^{3/2}/6 + \cdots\right)} \\
&= \frac{\sigma h^{1/2} + (r - \sigma^2/2)h + \cdots}{2\sigma h^{1/2}\left(1 + \sigma^2 h/6 + \cdots\right)} \\
&= \frac{1}{2}\left[\frac{1 + \sigma^{-1}h^{1/2}(r - \sigma^2/2) + \cdots}{1 + \sigma^2 h/6 + \cdots}\right] \\
&= \frac{1}{2}\left[\left(1 + \sigma^{-1}h^{1/2}(r - \sigma^2/2) + \cdots\right)\left(1 - \sigma^2 h/6 + \cdots\right)\right] \\
&= \frac{1}{2}\left[1 + \sigma^{-1}h^{1/2}(r - \sigma^2/2) + \cdots\right], \qquad (4.26)
\end{aligned}$$

to highest order, so that

$$1 - p = \frac{1}{2}\left[1 - \sigma^{-1}h^{1/2}(r - \sigma^2/2) + \cdots\right], \tag{4.27}$$

It's tempting to omit the higher order terms, but we would then lose risk neutrality in our Binomial Model.

Is the Binomial Model consistent with the Geometric Brownian Motion universe as $h \to 0$? We shall now show that the definition of a sufficiently smooth European option in the Binomial Model still implies the Black–Scholes PDE in the limit as $h \to 0$.

Proposition 4.3. *Let $f : [0,\infty)\times[0,\infty) \to \mathbb{R}$ be an infinitely differentiable function satisfying*

$$f(S, t-h) = e^{-rh}\left(pf(Se^{\sigma h^{1/2}}, t) + (1-p)f(Se^{-\sigma h^{1/2}}, t))\right), \tag{4.28}$$

for all $h > 0$, where p is given by (4.25). Then f satisfies the Black–Scholes PDE.

Proof. As usual, it is much more convenient to use log-space. Thus we define $u = \log S$ and

$$g(u, t) = f(S, t).$$

Hence (4.28) becomes

$$g(u, t-h) = e^{-rh}\left(pg(u + \sigma h^{1/2}, t) + (1-p)g(u - \sigma h^{1/2}, t)\right), \tag{4.29}$$

Using (4.26) and (4.27) and omitting terms whose order exceeds h for clarity, we obtain

$$
\begin{aligned}
g - hg_t + \cdots &= e^{-rh}\Big(\frac{1}{2}\Big[1 + h^{1/2}\sigma^{-1}(r - \sigma^2/2)\Big]\Big[g + \sigma h^{1/2} g_u + \frac{1}{2}\sigma^2 h g_{uu}\Big] \\
&\qquad + \frac{1}{2}\Big[1 - h^{1/2}\sigma^{-1}(r - \sigma^2/2)\Big]\Big[g - \sigma h^{1/2} g_u + \frac{1}{2}\sigma^2 h g_{uu}\Big]\Big) \\
&= e^{-rh}\Big(g + h\Big[(r - \sigma^2/2) g_u + \frac{1}{2}\sigma^2 g_{uu}\Big]\Big) \\
&= \Big(1 - rh + O(h^2)\Big)\Big(g + h\Big[(r - \sigma^2/2) g_u + \frac{1}{2}\sigma^2 g_{uu}\Big]\Big) \\
&= g + h\Big[-rg + (r - \sigma^2/2) g_u + \frac{1}{2}\sigma^2 g_{uu}\Big].
\end{aligned}
\tag{4.30}
$$

Equating the $O(h)$ terms on both sides of equation (4.30) yields the Black–Scholes equation

$$-g_t = -rg + (r - \sigma^2/2) g_u + \frac{1}{2}\sigma^2 g_{uu}.$$

□

4.1. **The Binomial Model and Delta Hedging.** We begin with (4.1), as before, but this time do *not* impose risk neutrality to determine the parameters p and α. Instead, we use a delta hedging argument.

At time $t_{n-1} = (n-1)h$, we construct a new portfolio

$$\Pi_{n-1} = f(S_{n-1}, t_{n-1}) - \Delta_{n-1} S_{n-1}. \tag{4.31}$$

and we choose Δ_{n-1} so that the evolution of Π_{n-1} is deterministic. Now, at time $t_n = nh$, the portfolio Π_{n-1} has the new value

$$\Pi_n = f(S_{n-1} e^{\alpha X_n}, t_n) - \Delta_{n-1} S_{n-1} e^{\alpha X_n}. \tag{4.32}$$

Thus Π_n is deterministic if the two possible values of (4.32) are equal, that is,

$$f(S_{n-1} e^{\alpha}, t_n) - \Delta_{n-1} S_{n-1} e^{\alpha} = f(S_{n-1} e^{-\alpha}, t_n) - \Delta_{n-1} S_{n-1} e^{-\alpha}. \tag{4.33}$$

It is useful to introduce the notation

$$f_\pm = f(S_{n-1} e^{\pm\alpha}, t_n). \tag{4.34}$$

Then (4.33) and (4.34) imply that

$$\Delta_{n-1} S_{n-1} = \frac{f_+ - f_-}{e^{\alpha} - e^{-\alpha}}. \tag{4.35}$$

Thus the resulting portfolio values are given by

$$\Pi_{n-1} = f(S_{n-1}, t_{n-1}) - \frac{f_+ - f_-}{e^{\alpha} - e^{-\alpha}} \tag{4.36}$$

and

$$
\begin{aligned}
\Pi_n &= f(S_{n-1} e^{\alpha}, t_n) - \frac{f_+ - f_-}{e^{\alpha} - e^{-\alpha}} e^{\alpha} \\
&= \frac{f_- e^{\alpha} - f_+ e^{-\alpha}}{e^{\alpha} - e^{-\alpha}}.
\end{aligned}
\tag{4.37}
$$

Now that the portfolio's evolution from Π_{n-1} to Π_n is deterministic, we must have $\Pi_n = \exp(rh)\Pi_{n-1}$, i.e.

$$\frac{f_- e^\alpha - f_+ e^{-\alpha}}{e^\alpha - e^{-\alpha}} = e^{rh}\left(f(S_{n-1}, t_{n-1}) - \frac{f_+ - f_-}{e^\alpha - e^{-\alpha}}\right). \tag{4.38}$$

The key point here is that $f(S_{n-1}, t_{n-1})$ is a linear combination of f_+ and f_-. Specifically, if we introduce

$$P = \frac{e^{rh} - e^{-\alpha}}{e^\alpha - e^{-\alpha}}, \tag{4.39}$$

then (4.38) becomes

$$f(S_{n-1}, t^{n-1}) = e^{-rh}\left(Pf_+ + (1-P)\, f_-\right). \tag{4.40}$$

Note that original model probability p does not occur in this formula: instead, it is as if we had begun with the alternative binomial model

$$S_n = S_{n-1} e^{\alpha Y_n}, \tag{4.41}$$

where the independent Bernoulli random variables $Y_1, Y_2, \ldots, Y_n$ satisfy $\mathbb{P}(Y_k = 1) = P$ and $\mathbb{P}(Y_k = -1) = 1 - P$, where P is given by (4.39). Indeed, we have

$$\mathbb{E} S_n | S_{n-1} = S_{n-1}\mathbb{E} e^{\alpha Y_n} = S_{n-1} e^{rh}. \tag{4.42}$$

Exercise 4.2. *Prove that* $\mathbb{E} S_n | S_{n-1} = S_{n-1} e^{rh}$.

4.2. **Δ-Hedging for GBM.** We begin with the real world asset price

$$S_t = e^{\alpha + \beta t + \sigma W_t}, \tag{4.43}$$

where we do **not** assume there is any connection between the parameters α, β and σ: this is **not** risk-neutral GBM. It is a simple exercise in Itô calculus (see Example 3.6) to prove that

$$dS_t = S_t\left(\sigma dW_t + \left(\beta + \sigma^2/2\right) dt\right) \tag{4.44}$$

and

$$\left(dS_t\right)^2 = \sigma^2 S_t^2 dt. \tag{4.45}$$

By analogy with delta hedging in the Binomial Model (4.31), let us assume that $S_t = S$ and define the portfolio

$$\Pi_t = f(S_t, t) - \Delta S_t, \tag{4.46}$$

where Δ is a constant. Then

$$\begin{aligned}
\Pi_{t+dt} &= f(S_{t+dt}, t + dt) - \Delta S_{t+dt} \\
&= f(S + dS_t, t + dt) - \Delta S - \Delta dS_t \\
&= f(S,t) + dS_t f_S + \frac{1}{2} dS_t^2 f_{SS} + dt f_t - \Delta S - \Delta dS_t \\
&= \Pi_t + dS_t\left(f_S - \Delta\right) + \frac{1}{2} dS_t^2 f_{SS} + dt f_t \\
&= \Pi_t + dS_t\left(f_S - \Delta\right) + \left(\frac{1}{2}\sigma^2 S^2 f_{SS} + f_t\right) dt. \qquad (4.47)
\end{aligned}$$

In other words, we have the infinitesimal increment

$$\Pi_{t+dt} - \Pi_t = d\Pi_t = dS_t\left(f_S - \Delta\right) + \left(\frac{1}{2}\sigma^2 S^2 f_{SS} + f_t\right) dt. \tag{4.48}$$

Thus we eliminate the stochastic dS_t component by setting

$$\Delta = f_S \tag{4.49}$$

and (4.47) then becomes

$$d\Pi_t = \left(f_t + \frac{1}{2}\sigma^2 S^2 f_{SS}\right) dt, \tag{4.50}$$

or

$$\frac{d\Pi_t}{dt} = f_t + \frac{1}{2}\sigma^2 S^2 f_{SS}. \tag{4.51}$$

Now there is no stochastic component in (4.50), so we must also have

$$\frac{d\Pi_t}{dt} = r\Pi_t = r\left(f - f_S S\right), \tag{4.52}$$

because all deterministic assets must grow at the risk-free rate. Equating (4.51) and (4.52) yields

$$f_t + \frac{1}{2}\sigma^2 S^2 f_{SS} = r\left(f - f_S S\right), \tag{4.53}$$

or

$$f_t - rf + rf_S S + \frac{1}{2}\sigma^2 S^2 f_{SS} = 0, \tag{4.54}$$

which is the Black–Scholes PDE.

It is often useful to restate the Black–Scholes PDE in terms of the logarithm of the asset price, i.e. via $S = e^x$. Thus

$$\partial_S \frac{dS}{dx} = \partial_x,$$

or

$$S\partial_S = \partial_x. \tag{4.55}$$

Hence

$$\begin{aligned} \partial_{xx} &= S\partial_S\left(S\partial_S\right) \\ &= S\left(\partial_S + S\partial_{SS}\right) \\ &= S\partial_S + S^2\partial_{SS}, \qquad (4.56) \end{aligned}$$

or, using (4.55),

$$S^2\partial_{SS} = \partial_{xx} - \partial_x. \tag{4.57}$$

Therefore substituting (4.55) and (4.57) in (4.54) yields

$$\begin{aligned} 0 &= f_t - rf + rf_x + \frac{1}{2}\sigma^2\left(f_{xx} - f_x\right) \\ &= f_t - rf + \left(r - \sigma^2/2\right) f_x + \frac{1}{2}\sigma^2 f_{xx}. \qquad (4.58) \end{aligned}$$

(4.59)

5. The Partial Differential Equation Approach

One important way to price options is to solve the Black–Scholes partial differential equation (PDE), or some variant of Black–Scholes. Hence we study the fundamentals of the numerical analysis of PDEs.

5.1. **The Diffusion Equation.** The diffusion equation arises in many physical and stochastic situations. In the hope that the baroque will serve as a mnemonic, we shall model the diffusion of poison along a line. Let $u(x,t)$ be the density of poison at location x and time t and consider the stochastic model

$$u(x,t) = \mathbb{E}u(x+\sigma\sqrt{h}Z, t-h), \qquad x \in \mathbb{R}, t \geq 0, \tag{5.1}$$

where σ is a positive constant and $Z \sim N(0,1)$. The idea here is that the poison molecules perform a random walk along the line, just as share prices do in time. If we assume that u has sufficiently many derivatives, then we obtain

$$\begin{aligned} &u(x,t) \\ &= \mathbb{E}\; u(x,t) + \sigma\sqrt{h}\frac{\partial u}{\partial x}Z + \frac{1}{2}\sigma^2 hZ^2\frac{\partial^2 u}{\partial x^2} + O(h^{3/2}) - h\frac{\partial u}{\partial t} + O(h^2) \\ &= u(x,t) + h\Big(\frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t}\Big) + O(h^{3/2}). \end{aligned}$$

In other words, dividing by h, we obtain

$$\Big(\frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t}\Big) + O(h^{1/2}) = 0.$$

Letting $h \to 0$, we have derived the *diffusion equation*

$$\frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}. \tag{5.2}$$

This important partial differential equation is often called the *heat equation*.

Exercise 5.1. *The d-dimensional form of our stochastic model for diffusion is given by*

$$u(\mathbf{x},t) = \mathbb{E}u(\mathbf{x}+\sigma\sqrt{h}\mathbf{Z}, t-h), \qquad \mathbf{x} \in \mathbb{R}^d, t \geq 0.$$

Here $\mathbf{Z} \in \mathbb{R}^d$ is a normalized Gaussian random vector: its component are independent $N(0,1)$ random variables and its probability density function is

$$p(\mathbf{z}) = (2\pi)^{-d/2}\exp(-\|\mathbf{z}\|^2/2), \qquad \mathbf{z} \in \mathbb{R}^d.$$

Assuming u is sufficiently differentiable, prove that u satisfies the d-dimensional diffusion equation

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2}\sum_{k=1}^{d}\frac{\partial^2 u}{\partial x_k^2}.$$

Variations on the diffusion equation occur in many fields including, of course, mathematical finance. For example, the neutron density[2] $N(\mathbf{x},t)$ in Uranium 235 or Plutonium approximately obeys the partial differential equation

$$\frac{\partial N}{\partial t} = \alpha N + \beta\sum_{k=1}^{d}\frac{\partial^2 N}{\partial x_k^2}.$$

[2] In mathematical finance, we choose our model to avoid exponential growth, but this is not always the aim in nuclear physics.

In fact, the Black–Scholes PDE is really the diffusion equation in disguise, as we shall now show. In log-space, we consider any solution $f(\tilde{S}, t)$ of the Black–Scholes equation, that is,

$$-rf + (r - \sigma^2/2)\frac{\partial f}{\partial \tilde{S}} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial \tilde{S}^2} + \frac{\partial f}{\partial t} = 0. \tag{5.3}$$

The inspired trick, a product of four centuries of mathematical play with differential equations, is to substitute

$$f(\tilde{S}, t) = u(\tilde{S}, t)e^{\alpha\tilde{S} + \beta t} \tag{5.4}$$

and to find the PDE satisfied by u. Now

$$\frac{\partial f}{\partial \tilde{S}} = \left(u\alpha + \frac{\partial u}{\partial \tilde{S}}\right) e^{\alpha\tilde{S} + \beta t}$$

and

$$\frac{\partial^2 f}{\partial \tilde{S}^2} = \left(u\alpha^2 + 2\alpha\frac{\partial u}{\partial \tilde{S}} + \frac{\partial^2 u}{\partial \tilde{S}^2}\right) e^{\alpha\tilde{S} + \beta t}.$$

Substituting in the Black–Scholes equation results in

$$-ru + \left(r - \sigma^2/2\right)\left(\alpha u - \frac{\partial u}{\partial \tilde{S}}\right) + \frac{\sigma^2}{2}\left(\alpha^2 u + 2\alpha\frac{\partial u}{\partial \tilde{S}} + \frac{\partial^2 u}{\partial \tilde{S}^2}\right) + \beta u + \frac{\partial u}{\partial t} = 0,$$

or

$$0 = \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial t} + \frac{\partial u}{\partial \tilde{S}}\left((r - \sigma^2/2) + \alpha\sigma^2\right) + u\left(-r + \alpha(r - \sigma^2/2) + \alpha^2\sigma^2/2 + \beta\right).$$

We can choose α and β to be any real numbers we please. In particular, if we set $\alpha = -\sigma^{-2}(r - \sigma^2/2)$, then the $\partial u/\partial \tilde{S}$ term vanishes. We can then solve for β to kill the u term.

Exercise 5.2. *Find the value of β that annihilates the u term.*

The practical consequence of this clever trick is that every problem involving the Black–Scholes PDE can be transformed into an equivalent problem for the diffusion equation. Therefore we now study methods for solving the diffusion equation.

There is an analytic solution for the diffusion equation that is sometimes useful. If we set $h = t$ in (5.1), then we obtain

$$u(x, t) = \mathbb{E}u(x + \sigma\sqrt{t}Z, 0), \tag{5.5}$$

that is,

$$u(x, t) = \int_{-\infty}^{\infty} u(x + \sigma\sqrt{t}z, 0)(2\pi)^{-1/2}\exp(-z^2/2)\,dz \tag{5.6}$$

$$= \int_{-\infty}^{\infty} u(x - w, 0)G(w, t)\,dw, \tag{5.7}$$

using the substitution $w = -\sigma\sqrt{t}z$, where

$$G(w, t) = (2\pi\sigma^2 t)^{-1/2}\exp\left(-\frac{w^2}{2\sigma^2 t}\right), \qquad w \in \mathbb{R}.$$

This is called the *Green's function* for the diffusion equation. Of course, we must now evaluate the integral. As for European options, analytic solutions exist for some simple cases, but numerical integration must be used in general.

5.2. **Finite Difference Methods for the Diffusion Equation.** The simplest finite difference method is called *explicit Euler* and it's a **BAD** method. Fortunately the insight gained from understanding why it's bad enables us to construct good methods. There is another excellent reason for you to be taught bad methods and why they're bad: stupidity is a renewable resource. In other words, simple bad methods are often rediscovered.

We begin with some *finite difference approximations* to the time derivative

$$\frac{\partial u}{\partial t} \approx \frac{u(x,t+k)-u(x,t)}{k}$$

and the space derivative, using the *second central difference*

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u(x+h,t)-2u(x,t)+u(x-h,t)}{h^2}.$$

Exercise 5.3. *Show that*

$$\frac{g(x+h)-2g(x)+g(x-h)}{h^2} = g^{(2)}(x) + \frac{h^2}{12}g^{(4)}(x) + O(h^4)$$

and find the next two terms in the expansion.

Our model problem for this section will be the *zero boundary value problem*:

$$\begin{aligned} &\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 \le x \le 1, \quad t \ge 0, \\ &u(x,0) = f(x), \quad 0 \le x \le 1, \qquad u(0,t) = u(1,t) = 0, \quad t \ge 0. \end{aligned} \tag{5.8}$$

We now choose a positive integer M and positive numbers T and k. We then set $h = 1/M$, $N = T/k$ and generate a discrete approximation

$$\{U_m^n : 0 \le m \le M, 0 \le n \le N, \}$$

to the values of the solution u at the points of the rectangular grid

$$\{(mh, nk) : 0 \le m \le M, 0 \le n \le N\}$$

using the recurrence

$$U_m^{n+1} = U_m^n + \mu\left(U_{m+1}^n - 2U_m^n + U_{m-1}^n\right), \qquad n \ge 0, 1 \le m \le M-1, \tag{5.9}$$

where

$$\mu = \frac{k}{h^2} \tag{5.10}$$

and the *boundary values* for u imply the relations

$$U_0^n = U_M^n = 0 \quad \text{and} \quad U_m^0 = u(mh,0), \quad 0 \le m \le M. \tag{5.11}$$

This is called *explicit Euler*.

In matrix terms[3], we have

$$\mathbf{U}^n = T\mathbf{U}^{n-1}, \qquad n \ge 1, \tag{5.12}$$

[3]How do we find T? Equation ((5.12)) implies

$$U_m^n = (T\mathbf{U}^{n-1})_m = \sum_{\ell=1}^{M-1} T_{m\ell} U_\ell^{n-1} = \mu U_{m-1}^n + (1-2\mu)U_m^n + \mu U_{m+1}^n.$$

where

$$\mathbf{U}^n = \begin{pmatrix} U_1^n \\ \vdots \\ U_{M-1}^n \end{pmatrix} \in \mathbb{R}^{M-1} \tag{5.13}$$

and $T \in \mathbb{R}^{(M-1)\times(M-1)}$ is the *tridiagonal symmetric Toeplitz* (TST) matrix defined by

$$T = \begin{pmatrix} 1-2\mu & \mu & & & \\ \mu & 1-2\mu & \mu & & \\ & & \ddots & \ddots & \ddots \\ & & \mu & 1-2\mu & \mu \\ & & & \mu & 1-2\mu \end{pmatrix} \tag{5.14}$$

Hence

$$\mathbf{U}^n = T^n \mathbf{U}^0. \tag{5.15}$$

Unfortunately, explicit Euler is an *unstable method* unless $\mu \le 1/2$. In other words, the numbers $\{U_m^n : 1 \le m \le M-1\}$ grow exponentially as $n \to \infty$. Here's an example using Matlab.

Example 5.1. *The following Matlab fragment generates the explicit Euler approximations.*

```
% Choose our parameters
mu = 0.7;
M=100; N=20;
% Pick (Gaussian) random initial values
uold = randn(M-1,1);
% construct the tridiagonal symmetric Toeplitz matrix T
T = (1-2*mu)*diag(ones(M-1,1)) + mu*( diag(ones(M-2,1),1) + diag(ones(M-2,1),-1) );
% iterate and plot
plot(uold)
hold on
for k=1:N
  unew = T*uold;
  plot(unew)
  uold = unew;
end
```

If we run the above code for $M = 6$ *and*

$$\mathbf{U}^0 = \begin{pmatrix} -0.034942 \\ 0.065171 \\ -0.964159 \\ 0.406006 \\ -1.450787 \end{pmatrix},$$

then we obtain

$$\mathbf{U}^{20} = \begin{pmatrix} -4972.4 \\ 8614.6 \\ -9950.7 \\ 8620.5 \\ -4978.3 \end{pmatrix}.$$

Further, $\|\mathbf{U}^{40}\| = 2.4 \times 10^8$. *The exponential instability is obvious. Experiment with different values of* μ, M *and* N.

The restriction $\mu \le 1/2$, that is $k \le h^2/2$, might not seem particularly harmful at first. However, it means that small h values require *tiny* k values, and tiny timesteps imply lots of work: an inefficient method. Now let's derive this stability requirement. We begin by studying a more general problem based on (5.15). Specifically, let $A \in \mathbb{R}^{n\times n}$ be any symmetric matrix[4]. Its *spectral radius* $\rho(A)$ is simply its largest eigenvalue in modulus, that is,

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\}. \tag{5.16}$$

Theorem 5.1. *Let* $A \in \mathbb{R}^{n\times n}$ *be any symmetric matrix and define the recurrence* $\mathbf{x}_k = A\mathbf{x}_{k-1}$, *for* $k \ge 1$, *where* $\mathbf{x}_0 \in \mathbb{R}^n$ *can be any initial vector.*

(i) *If* $\rho(A) < 1$, *then* $\lim_{k\to\infty} \|\mathbf{x}_k\| = 0$, *for any initial vector* $\mathbf{x}_0 \in \mathbb{R}^n$.
(ii) *If* $\rho(A) \le 1$, *then the norms of the iterates* $\|\mathbf{x}_1\|, \|\mathbf{x}_2\|, \ldots$ *remain bounded.*
(iii) *If* $\rho(A) > 1$, *then we can choose* $\mathbf{x}_0 \in \mathbb{R}^n$ *such that* $\lim_{k\to\infty} \|\mathbf{x}_k\| = \infty$.

Proof. We use the spectral decomposition introduced in Theorem 3.15, so that $A = QDQ^T$, where $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix and $D \in \mathbb{R}^{n\times n}$ is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A. Then

$$\mathbf{x}_k = A\mathbf{x}_{k-1} = A^2\mathbf{x}_{k-2} = \cdots = A^k\mathbf{x}_0$$

and

$$\begin{aligned} A^k &= \left(QDQ^T\right)\left(QDQ^T\right)\cdots\left(QDQ^T\right) \\ &= QD^kQ^T. \end{aligned}$$

Hence

$$\mathbf{x}_k = QD^kQ^T\mathbf{x}_0$$

and, introducing $\mathbf{z}_k := Q^T\mathbf{x}_k$, we obtain

$$\mathbf{z}_k = D^k\mathbf{z}_0,$$

and it is important to observe that $\|\mathbf{z}_k\| = \|Q^T\mathbf{x}_k\| = \|\mathbf{x}_k\|$, because Q^T is an orthogonal matrix. Since D is a diagonal matrix, this matrix equation is simply n linear recurrences, namely

$$z_k(\ell) = \lambda_\ell^k z_0(\ell), \qquad \ell = 1, 2, \ldots, n.$$

The following consequences are easily checked.

(i) If $\rho(A) < 1$, then each of these scalar sequences tends to zero, as $k \to \infty$, which implies that $\|\mathbf{x}_k\| \to 0$.
(ii) If $\rho(A) = 1$, then each of these scalar sequences is bounded, which implies that the sequence $\|\mathbf{x}_k\|$ remains bounded.

[4]All of this theory can be generalized to nonsymmetric matrices using the Jordan canonical form, but this advanced topic is not needed in this course.

(iii) If $\rho(A) > 1$, then there is at least one eigenvalue, λ_i say, for which $|\lambda_i| > 1$. Hence, if $z_0(i) \neq 0$, then the sequence $|z_k(i)| = |\lambda_i^k z_0(i)| \to \infty$, as $k \to \infty$.

□

Definition 5.1. *Let $A \in \mathbb{R}^{n\times n}$ be any symmetric matrix. We say that A is* spectrally stable *if its spectral radius satisfies $\rho(A) \leq 1$.*

Example 5.2. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix whose distinct eigenvalues are $0.1, 0.1, \ldots, 0.1, 10$. If $\mathbf{x}_0 \in \mathbb{R}^n$ contains no component of the eigenvector corresponding to the eigenvalue 10, i.e. $z_0(n) = 0$, then, in* exact arithmetic*, we shall still obtain $\|\mathbf{x}_k\| \to 0$, as $k \to \infty$. However, a computer uses finite precision arithmetic, which implies that, even if $z_0(n) = 0$, it is highly likely that $z_0(m) \neq 0$, for some $m > 0$, since the matrix-vector product is not computed exactly. This (initially small) nonzero component will grow exponentially.*

Theorem 5.1 is only useful when we can deduce the magnitude of the spectral radius. Fortunately, this is possible for an important class of matrices.

Definition 5.2. *We say that a matrix $T(a, b) \in \mathbb{R}^{m\times m}$ is* tridiagonal, symmetric and Toeplitz *(TST) if it has the form*

$$T(a,b) = \begin{pmatrix} a & b & & & \\ b & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & b & a & b \\ & & & b & a \end{pmatrix}, \qquad a, b \in \mathbb{R}. \tag{5.17}$$

TST matrices arise naturally in many applications. Fortunately they're one of the few nontrivial classes of matrices for which the eigenvalues and eigenvectors can be analytically determined rather easily. In fact, every TST matrix has the same eigenvectors, because

$$T(a, b) = aI + 2bT_0, \tag{5.18}$$

where $T_0 = T(0, 1/2)$ (this is *not* a recursive definition, simply an observation given (5.18)). Hence, if $T_0\mathbf{v} = \lambda\mathbf{v}$, then $T(a, b)\mathbf{v} = (a + 2b\lambda)\mathbf{v}$. Thus we only need to study T_0.

In fact, every eigenvalue of T_0 lies in the interval $[-1, 1]$. The proof is interesting because it's our only example of using a different *norm*. For any vector $\mathbf{w} \in \mathbb{R}^m$, we define its *infinity norm* to be

$$\|\mathbf{w}\|_\infty = \max\{|w_1|, |w_2|, \ldots, |w_m|\}.$$

Exercise 5.4. *Show that*

$$\|T_0\mathbf{z}\|_\infty \leq \|\mathbf{z}\|_\infty,$$

for any vector $\mathbf{z} \in \mathbb{R}^m$.

We shall state our result formally for ease of reference.

Lemma 5.2. *Every eigenvalue λ of T_0 satisfies $|\lambda| \leq 1$.*

Proof. If $T_0\mathbf{v} = \lambda\mathbf{v}$, then

$$|\lambda|\|\mathbf{v}\|_\infty = \|\lambda\mathbf{v}\|_\infty = \|T_0\mathbf{v}\|_\infty \leq \|\mathbf{v}\|_\infty.$$

Hence $|\lambda| \leq 1$. □

Proposition 5.3. *The eigenvalues of $T_0 \in \mathbb{R}^{m\times m}$ are given by*

$$\lambda_j = \cos\left(\frac{j\pi}{m+1}\right), \qquad j = 1, \ldots, m, \tag{5.19}$$

and the corresponding eigenvector $\mathbf{v}^{(j)}$ has components

$$\mathbf{v}_k^{(j)} = \sin\left(\frac{\pi jk}{m+1}\right), \qquad j, k = 1, \ldots, m. \tag{5.20}$$

Proof. Suppose $\mathbf{v}$ is an eigenvector for T_0, so that

$$v_{j+1} + v_{j-1} = 2\lambda v_j, \qquad 2 \le j \le m-1,$$

and

$$v_2 = 2\lambda v_1, \quad v_{m-1} = 2\lambda v_m.$$

Thus the elements of the vector $\mathbf{v}$ are m values of the recurrence relation defined by

$$v_{j+1} + v_{j-1} = 2\lambda v_j, \qquad j \in \mathbb{Z},$$

where $v_0 = v_{m+1} = 0$. Here's a rather slick trick: we know that $|\lambda| \le 1$, and a general theoretical result states that the eigenvalues of a real symmetric matrix are real, so we can write $\lambda = \cos\theta$, for some $\theta \in \mathbb{R}$. The associated equation for this recurrence is therefore the quadratic

$$t^2 - 2t\cos\theta + 1 = 0$$

which we can factorize as

$$\left(t - e^{i\theta}\right)\left(t - e^{-i\theta}\right) = 0.$$

Thus the general solution is

$$v_j = re^{ij\theta} + se^{-ij\theta}, \qquad j \in \mathbb{Z},$$

where r and s can be any complex numbers. But $v_0 = 0$ implies $s = -r$, so we obtain

$$v_j = \sin j\theta, \qquad j \in \mathbb{Z},$$

on using the fact that every multiple of a sequence satisfying the recurrence is another sequence satisfying the recurrence. The only other condition remaining to be satisfied is $v_{m+1} = 0$, so that

$$\sin\left((m+1)\theta\right) = 0,$$

which implies $(m+1)\theta$ is some integer multiple of π. □

Exercise 5.5. *Prove that the eigenvectors given in Proposition 5.3 are orthogonal by direct calculation.*

The spectral radius of the matrix T driving explicit Euler, defined by (5.15), is now an easy consequence of our more general analysis of TST matrices.

Corollary 5.4. *Let $T \in \mathbb{R}^{(M-1)\times(M-1)}$ be the matrix driving explicit Euler, defined by (5.15). Then $\rho(T) \le 1$ if and only if $\mu \le 1/2$. Hence explicit Euler is spectrally stable if and only if $\mu \le 1/2$*

Proof. We need only observe that $T = T(1-2\mu, \mu)$, so that Proposition 5.3 implies that its eigenvalues are

$$\begin{aligned} \lambda_k &= 1 - 2\mu + 2\mu \cos(\frac{\pi k}{M}) \\ &= 1 - 4\mu \sin^2\left(\frac{\pi k}{2M}\right), \qquad k = 1, 2, \ldots, M-1. \end{aligned}$$

Thus $\rho(T) \leq 1$ if and only if $\mu \leq 1/2$, for otherwise $|1 - 4\mu| > 1$. □

We can also use TST matrices to understand *implicit Euler*: here we use

$$U_m^{n+1} = U_m^n + \mu \left(U_{m+1}^{n+1} - 2U_m^{n+1} + U_{m-1}^{n+1}\right), \qquad 1 \leq m \leq M-1. \tag{5.21}$$

In matrix form, this becomes

$$T(1+2\mu, -\mu)\mathbf{U}^{n+1} = \mathbf{U}^n, \tag{5.22}$$

using the notation of (5.17). Before using Proposition 5.3 to derive its eigenvalues, we need a simple lemma.

Lemma 5.5. *Let $A \in \mathbb{R}^{n\times n}$ be any symmetric matrix, having spectral decomposition $A = QDQ^T$. Then $A^{-1} = QD^{-1}Q^T$.*

Proof. This is a very easy exercise. □

Proposition 5.6. *Implicit Euler is spectrally stable for all $\mu \geq 0$.*

Proof. By Proposition 5.3, the eigenvalues of $T(1 + 2\mu, -\mu)$ are given by

$$\begin{aligned} \lambda_k &= 1 + 2\mu - 2\mu \cos(\frac{\pi k}{M}) \\ &= 1 + 4\mu \sin^2(\frac{\pi k}{2M}). \end{aligned}$$

Thus every eigenvalue of $T(1 + 2\mu, -\mu)$ exceeds 1, which implies (by Lemma 5.5) that every eigenvalue of its inverse lies in the interval $(0, 1)$. Thus implicit Euler is spectrally stable for all $\mu \geq 0$. □

We have yet to prove that the answers produced by these methods converge to the true solution as $h \to 0$. We illustrate the general method using explicit Euler, for $\mu \leq 1/2$, applied to the diffusion equation on $[0, 1]$ with zero boundary (5.8). If we define the error

$$E_m^n := u(mh, nk) - U_m^n. \tag{5.23}$$

then

$$E_m^{n+1} - E_m^n - \mu \left(E_{m+1}^n - 2E_m^n + E_{m-1}^n\right) = L(x, t), \tag{5.24}$$

where the *Local Truncation Error (LTE)* $L(x, t)$ is defined by

$$L(x, t) = u(x, t + k) - u(x, t) - \mu \left(u(x + h, t) - 2u(x, t) + u(x - h, t)\right), \tag{5.25}$$

recalling that, by definition,

$$0 = U_m^{n+1} - U_m^n - \mu \left(U_{m+1}^n - 2U_m^n + U_{m-1}^n\right), \qquad 1 \leq m \leq M-1. \tag{5.26}$$

Thus we form the LTE by replacing U_m^n by $u(x,t)$ in (5.26)[5]. Taylor expanding and recalling that $k = \mu h^2$, we obtain

$$\begin{aligned} L(x,t) &= ku_t(x,t) + O(k^2) - \mu\left(h^2 u_{xx}(x,t) + O(h^4)\right) \\ &= \mu h^2 u_{xx}(x,t) - \mu h^2 u_{xx}(x,t) + O(h^4) \\ &= O(h^4), \end{aligned} \tag{5.27}$$

using the fact that $u_t = u_{xx}$. Now choose a time interval $[0,T]$. Since $L(x,t)$ is a continuous function, (5.27) implies the inequality

$$|L(x,t)| \le Ch^4, \qquad \text{for } 0 \le x \le 1 \text{ and } 0 \le t \le T, \tag{5.28}$$

where the constant C depends on T. Further, rearranging (5.24) yields

$$E_m^{n+1} = E_m^n + \mu\left(E_{m+1}^n - 2E_m^n + E_{m-1}^n\right) + L(x,t), \tag{5.29}$$

and applying inequality (5.28), we obtain

$$|E_m^{n+1}| \le (1-2\mu)|E_m^n| + \mu|E_{m+1}^n| + \mu|E_{m-1}^n| + Ch^4, \tag{5.30}$$

because $1 - 2\mu \ge 0$ for $\mu \le 1/2$. If we let η_n denote the maximum modulus error at time nk, i.e.

$$\eta_n = \max\{|E_1^n|, |E_2^n|, \ldots, |E_{M-1}^n|\} \tag{5.31}$$

then (5.31) implies

$$|E_m^{n+1}| \le (1-2\mu)\eta_n + 2\mu\eta_n + Ch^4 = \eta_n + Ch^4, \tag{5.32}$$

whence

$$\eta_{n+1} \le (1-2\mu)\eta_n + 2\mu\eta_n + Ch^4 = \eta_n + Ch^4. \tag{5.33}$$

Therefore, recurring (5.33)

$$\eta_n \le \eta_{n-1} + Ch^4 \le \eta_{n-2} + 2Ch^4 \le \cdots \le \eta_0 + nCh^4 = Cnh^4, \tag{5.34}$$

since $E_m^0 \equiv 0$. Now

$$n \le N := \frac{T}{k} = \frac{T}{\mu h^2}, \tag{5.35}$$

so that (5.34) and (5.35) jointly provide the upper bound

$$|U_m^n - u(mh, nk)| \le \left(\frac{CT}{\mu}\right) h^2, \tag{5.36}$$

for $1 \le m \le M-1$ and $0 \le n \le N$. Hence we have shown that the explicit Euler approximation has uniform $O(h^2)$ convergence for $0 \le t \le T$. The key here is the *order*, not the constant in the bound: halving h reduces the error uniformly by 4.

Exercise 5.6. *Refine the expansion of the LTE in* (5.27) *to obtain*

$$L(x,t) = ku_t(x,t) + \frac{1}{2}k^2 u_{tt}(x,t) + O(k^3) - \mu\left(h^2 u_{xx}(x,t) + \frac{1}{12}h^4 u_{xxxx}(x,t) + O(h^6)\right).$$

Hence prove that

$$L(x,t) = \frac{1}{2}\mu h^4 u_{tt}(x,t)\,(\mu - 1/6) + O(h^6).$$

[5] You will see the same idea in the next section, where this will be called the associated functional equation.

Hence show that, if $\mu = 1/6$ in explicit Euler, we obtain the higher-order uniform error

$$|U_m^n - u(mh, nk)| \le Dh^4,$$

for $1 \le m \le M-1$ and $0 \le n \le T/k$, where D depends on T.

Implicit Euler owes its name to the fact that we must solve linear equations to obtain the approximations at time $(n+1)h$ from those at time nh. This linear system is tridiagonal, so Gaussian elimination only requires $O(n)$ time to complete, rather than the $O(n^3)$ time for a general $n \times n$ matrix. In fact, there is a classic method that provides a higher order than implicit Euler together with excellent stability: Crank–Nicolson is the implicit method defined by

$$\begin{aligned} U_m^{n+1} - \frac{\mu}{2}\left(U_{m+1}^{n+1} - 2U_m^{n+1} + U_{m-1}^{n+1}\right) \\ = U_m^n + \frac{\mu}{2}\left(U_{m+1}^n - 2U_m^n + U_{m-1}^n\right). \end{aligned} \tag{5.37}$$

In matrix form, we obtain

$$T(1+\mu, -\mu/2)\mathbf{U}^{n+1} = T(1-\mu, \mu/2)\mathbf{U}^n, \tag{5.38}$$

or

$$\mathbf{U}^{n+1} = T(1+\mu, -\mu/2)^{-1}T(1-\mu, \mu/2)\mathbf{U}^n. \tag{5.39}$$

Now every TST matrix has the same eigenvectors. Thus the eigenvalues of the product of TST matrices in (5.39) are given by

$$\lambda_k = \frac{1-\mu+\mu\cos(\frac{\pi k}{M})}{1+\mu-\mu\cos\frac{\pi k}{M}} = \frac{1-2\mu\sin^2(\frac{\pi k}{2M})}{1+2\mu\sin^2(\frac{\pi k}{2M})}. \tag{5.40}$$

Hence $|\lambda_k| \in (0,1)$ for all $\mu \ge 0$.

Exercise 5.7. *Calculate the LTE of Crank-Nicolson when $h = k$.*

5.3. **The Fourier Transform and the von Neumann Stability Test.** Given any univariate function $f : \mathbb{R} \to \mathbb{R}$ for which the integral

$$\int_{-\infty}^{\infty} |f(x)|\,dx \tag{5.41}$$

is finite, we define its *Fourier transform* by the relation

$$\widehat{f}(z) = \int_{-\infty}^{\infty} f(x)\exp(-ixz)\,dx, \qquad z \in \mathbb{R}. \tag{5.42}$$

The Fourier transform is used in this course to understand stability properties, solve some partial differential equations and calculate the local truncation errors for finite difference methods. It can also be used to derive analytic values of certain options, as well as providing several key numerical methods.

Proposition 5.7. (i) *Let*

$$T_a f(x) = f(x+a), \qquad x \in \mathbb{R}. \tag{5.43}$$

We say that $T_a f$ is the translate *of f by a. Then*

$$\widehat{T_a f}(z) = \exp(iaz)\widehat{f}(z), \qquad z \in \mathbb{R}. \tag{5.44}$$

(ii) *The Fourier transform of the derivative is given by*

$$\widehat{f'}(z) = iz\widehat{f}(z), \qquad z \in \mathbb{R}. \tag{5.45}$$

Proof. (i)

$$\widehat{T_a f}(z) = \int_{-\infty}^{\infty} T_a f(x) e^{-ixz}\, dx$$
$$= \int_{-\infty}^{\infty} f(x+a) e^{-ixz}\, dx$$
$$= \int_{-\infty}^{\infty} f(y) e^{-i(y-a)z}\, dy$$
$$= e^{iaz}\widehat{f}(z).$$

(ii) Integrating by parts and using the fact that $\lim_{x\to\pm\infty} f(x) = 0$, which is a consequence of (5.41), we obtain

$$\widehat{f'}(z) = \int_{-\infty}^{\infty} f'(x) e^{-ixz}\, dx$$
$$= \left[f(x)e^{-ixz}\right]_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} f(x)\left(-ize^{-ixz}\right)\, dx$$
$$= iz\widehat{f}(z).$$

□

Exercise 5.8. *We have* $\widehat{f^{(2)}}(z) = (iz)^2\widehat{f}(z) = -z^2\widehat{f}(z)$. *Find* $\widehat{f^{(k)}}(z)$

Many students will have seen some use of the Fourier transform to solve differential equations. It is also vitally important to finite difference operators.

Example 5.3. *Let's analyse the second order central difference operator using the Fourier transform. Thus we take*

$$g(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.$$

and observe that

$$\widehat{g}(z) = h^{-2}\left(e^{ihz} - 2 + e^{-ihz}\right)\widehat{f}(z) = 2h^{-2}\left(\cos(hz) - 1\right)\widehat{f}(z).$$

Now[6]

$$\cos(hz) = 1 - \frac{h^2z^2}{2} + \frac{h^4z^4}{4!} - \cdots,$$

so that

$$\begin{aligned}
\widehat{g}(z) &= 2h^{-2}\left(-\frac{h^2z^2}{2} + \frac{h^4z^4}{4!} - \cdots\right)\widehat{f}(z) \\
&= -z^2\widehat{f}(z) + \frac{h^2z^4}{12}\widehat{f}(z) + \cdots \\
&= \widehat{f^{(2)}}(z) + h^2\frac{\widehat{f^{(4)}}(z)}{12} + \cdots.
\end{aligned}$$

Taking the inverse transform, we have computed the Taylor expansion of g:

$$g(x) = f^{(2)}(x) + (1/12)h^2 f^{(4)}(x) + \cdots.$$

[6]Commit this Taylor expansion to memory if you don't already know it!

Of course, there's no need to use the Fourier transform to analyse the second order central difference operator, but we have to learn to walk before we can run!

We shall also need the Fourier transform for functions of more than one variable. For any bivariate function $f(x_1, x_2)$ that tends to zero sufficiently rapidly at infinity, we define

$$\widehat{f}(z_1, z_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) e^{-i(x_1 z_1 + x_2 z_2)} \, dx_1 dx_2, \qquad z_1, z_2 \in \mathbb{R}. \tag{5.46}$$

In fact, it's more convenient to write this using a slightly different notation:

$$\widehat{f}(\mathbf{z}) = \int_{\mathbb{R}^2} f(\mathbf{x}) \exp(-i\mathbf{x}^T \mathbf{z}) \, d\mathbf{x}, \qquad \mathbf{z} \in \mathbb{R}^2. \tag{5.47}$$

This is still a double integral, although only one integration sign is used. Similarly, for a function $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, we define

$$\widehat{f}(\mathbf{z}) = \int_{\mathbb{R}^n} f(\mathbf{x}) \exp(-i\mathbf{x}^T \mathbf{z}) \, d\mathbf{x}, \qquad \mathbf{z} \in \mathbb{R}^n. \tag{5.48}$$

Here

$$\mathbf{x}^T \mathbf{z} = \sum_{k=1}^{n} x_k z_k, \qquad \mathbf{x}, \mathbf{z} \in \mathbb{R}^n.$$

The multivariate version of Proposition 5.7 is as follows.

Proposition 5.8. (i) *Let*

$$T_{\mathbf{a}} f(\mathbf{x}) = f(\mathbf{x} + \mathbf{a}), \qquad \mathbf{x} \in \mathbb{R}^n. \tag{5.49}$$

We say that $T_{\mathbf{a}} f$ is the translate *of f by $\mathbf{a}$. Then*

$$\widehat{T_{\mathbf{a}} f}(\mathbf{z}) = \exp(i\mathbf{a}^T \mathbf{z}) \widehat{f}(\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n. \tag{5.50}$$

Further, if $\alpha_1, \ldots, \alpha_n$ are non-negative integers and $|\alpha| = \alpha_1 + \cdots + \alpha_n$, then

(ii)

$$\frac{\widehat{\partial^{|\alpha|} f}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}(z) = (iz_1)^{\alpha_1} (iz_2)^{\alpha_2} \cdots (iz_n)^{\alpha_n} \widehat{f}(\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n. \tag{5.51}$$

Proof. The proof is not formally examinable, but very similar to the multivariate result. □

5.4. **Stability and the Fourier Transform.** We can also use Fourier analysis to avoid eigenanalysis when studying stability. We shall begin abstractly, but soon apply the analysis to explicit and implicit Euler for the diffusion equation.

Suppose we have two sequences $\{u_k\}_{k \in \mathbb{Z}}$ and $\{v_k\}_{k \in \mathbb{Z}}$ related by

$$\sum_{k \in \mathbb{Z}} b_k v_k = \sum_{k \in \mathbb{Z}} a_k u_k. \tag{5.52}$$

In most applications, $u_k \approx u(kh)$, for some underlying function, so we study the *associated functional equation*

$$\sum_{k \in \mathbb{Z}} b_k v(x + kh) = \sum_{k \in \mathbb{Z}} a_k u(x + kh). \tag{5.53}$$

The advantage of widening our investigation is that we can use the Fourier transform to study (5.53). Specifically, we have

$$\widehat{v}(z) \sum_{k\in\mathbb{Z}} b_k e^{ikhz} = \widehat{u}(z) \sum_{k\in\mathbb{Z}} a_k e^{ikhz}, \tag{5.54}$$

or

$$\widehat{v}(z) = \left(\frac{a(hz)}{b(hz)}\right) \widehat{u}(z) =: R(hz)\widehat{u}(z), \tag{5.55}$$

where

$$a(w) = \sum_{k\in\mathbb{Z}} a_k e^{ikw} \quad \text{and} \quad b(w) = \sum_{k\in\mathbb{Z}} b_k e^{ikw}. \tag{5.56}$$

Example 5.4. *For explicit Euler, we have*

$$v_k = \mu u_{k+1} + (1-2\mu)u_k + \mu u_{k-1}, \qquad k \in \mathbb{Z},$$

so that the associated functional equation is

$$v(x) = \mu u(x+h) + (1-2\mu)u(x) + \mu u(x-h), \qquad x \in \mathbb{R},$$

whose Fourier transform is given by

$$\begin{aligned} \widehat{v}(z) &= \left(\mu e^{ihz} + 1 - 2\mu + \mu e^{-ihz}\right) \widehat{u}(z) \\ &= \left(1 - 2\mu(1-\cos(hz))\right) \widehat{u}(z) \\ &= \left(1 - 4\mu\sin^2(hz/2)\right) \widehat{u}(z). \end{aligned} \tag{5.57}$$

Thus $\widehat{v}(z) = r(hz)\widehat{u}(z)$*, where* $r(w) = 1 - 4\mu\sin^2(w/2)$*.*

When we advance forwards n steps in time using explicit Euler, we obtain in Fourier transform space

$$\widehat{u_n}(z) = r(hz)\widehat{u_{n-1}}(z) = (r(hz))^2\,\widehat{u_{n-2}}(z) = \cdots = (r(hz))^n\,\widehat{u_0}(z). \tag{5.58}$$

Thus, if $|r(w)| < 1$, for all $w \in \mathbb{R}$,then $\lim_{n\to\infty} \widehat{u_n}(z) = 0$, for all $z \in \mathbb{R}$. However, if $|r(hz_0)| > 1$, then, by continuity, $|r(hz)| > 1$ for z sufficiently close to z_0. Further, since $r(hz)$ is periodic, with period π/h, we deduce that $|r(hz)| > 1$ on π/h-integer shifts of an interval centred at z_0. Hence $\lim_{n\to\infty} \widehat{u_n}(z) = \infty$. Further, there is an intimate connection between u and $\widehat{u}$ in the following sense.

Theorem 5.9 (Parseval's Theorem)**.** *If* $f : \mathbb{R} \to \mathbb{R}$ *is continuous, then*

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\widehat{f}(z)|^2\,dz. \tag{5.59}$$

Proof. Not examinable. □

Hence $\lim_{n\to\infty} \widehat{u_n}(z) = \infty$ implies

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |u_n(x)|^2\,dx = \infty.$$

this motivated the brilliant Hungarian mathematician John von Neumann to analyse the stability of finite difference operators via the Fourier transform.

Definition 5.3. *If* $|r(hz)| \le 1$*, for all* $z \in \mathbb{R}$*, then we say that the finite difference operator is* **von Neumann stable***, or Fourier stable.*

Theorem 5.10. *Explicit Euler is von Neumann stable if and only if* $\mu \le 1/2$*, while implicit Euler is von Neumann stable for all* $\mu > 0$*.*

Proof. For explicit Euler, we have already seen that

$$\widehat{u_n}(z) = r(hz)\widehat{u_{n-1}}(z),$$

where

$$r(w) = 1 - 4\mu \sin^2(w/2).$$

Thus $|r(w)| \le 1$, for all $w \in \mathbb{R}$, if and only if $|1-4\mu| \le 1$, i.e. $\mu \le 1/2$.

For implicit Euler, we have the associated functional equation

$$u_{n+1}(x) = u_n(x) + \mu\left(u_{n+1}(x+h) - 2u_{n+1}(x) + u_{n+1}(x-h)\right), \qquad x \in \mathbb{R}.$$

Hence

$$\left(-\mu e^{ihz} + (1+2\mu) - \mu e^{-ihz}\right)\widehat{u_{n+1}}(z) = \widehat{u_n}(z), \qquad z \in \mathbb{R},$$

or

$$\left(1 + 4\mu \sin^2(hz/2)\right)\widehat{u_{n+1}}(z) = \widehat{u_n}(z), \qquad z \in \mathbb{R}.$$

Therefore

$$\widehat{u_{n+1}(z)} = \frac{1}{1+4\mu\sin^2(hz/2)}\widehat{u_n}(z) =: r(hz)\widehat{u_n}(z), \qquad z \in \mathbb{R},$$

and $0 \le r(hz) \le 1$, for all $z \in \mathbb{R}$. □

Exercise 5.9. *Prove that Crank–Nicolson* (5.37) *is von Neumann stable for all* $\mu \ge 0$.

5.5. **Option Pricing via the Fourier transform.** The Fourier transform can also be used to calculate solutions of the Black–Scholes equation, and its variants, and this approach provides a powerful analytic and numerical technique.

We begin with the Black–Scholes equation in "log-space":

$$0 = -rg + (r - \sigma^2/2)g_x + (\sigma^2/2)g_{xx} + g_t, \tag{5.60}$$

where the asset price $S = e^x$ and subscripts denote partial derivatives. We now let $\widehat{g}(z,t)$ denote the Fourier transform of the option price $g(x,t)$ at time t, that is,

$$\widehat{g}(z,t) = \int_{-\infty}^{\infty} g(x,t)e^{-ixz}\,dx, \qquad z \in \mathbb{R}, \tag{5.61}$$

The Fourier transform of (5.60) is therefore given by

$$0 = -r\widehat{g} + iz(r-\sigma^2/2)\widehat{g} - \frac{1}{2}\sigma^2 z^2\widehat{g} + \widehat{g}_t. \tag{5.62}$$

In other words, we have, for each fixed $z \in \mathbb{R}$, the ordinary differential equation

$$\widehat{g}_t = -\Big(-r + iz(r-\sigma^2/2) - \frac{1}{2}\sigma^2 z^2\Big)\widehat{g},, \tag{5.63}$$

with solution

$$\widehat{g}(z,t) = \widehat{g}(z,t_0)e^{-\left(-r+iz(r-\sigma^2/2)-\frac{1}{2}\sigma^2 z^2\right)(t-t_0)}. \tag{5.64}$$

When pricing a European option, we know the option's expiry value $g(x,T)$ and wish to calculate its initial price $g(x,0)$. Substituting $t = T$ and $t_0 = 0$ in (5.64), we therefore obtain

$$\widehat{g}(z,0) = e^{\left(-r+iz(r-\sigma^2/2)-\frac{1}{2}\sigma^2 z^2\right)T}\widehat{g}(z,T). \tag{5.65}$$

In order to apply this, we shall need to know the Fourier transform of a Gaussian.

Proposition 5.11. *Let* $G(\mathbf{x}) = \exp(-\lambda\|\mathbf{x}\|^2)$*, for* $\mathbf{x} \in \mathbb{R}^d$*, where* λ *is a positive constant. Then its Fourier transform is the Gaussian*

$$\widehat{G}(\mathbf{z}) = (\pi/\lambda)^{d/2} \exp(-\|\mathbf{z}\|^2/(4\lambda)), \qquad \mathbf{z} \in \mathbb{R}^d. \tag{5.66}$$

Proof. It's usual to derive this result via contour integration, but here is a neat proof via Itô's lemma and Brownian motion. Let $c \in \mathbb{C}$ be any complex number and define the stochastic process $X_t = \exp(cW_t)$, for $t \geq 0$. Then a straightforward application of Itô's lemma implies the relation

$$dX_t = X_t \left(cdW_t + (c^2/2)dt\right).$$

Taking expectations and defining $m(t) = \mathbb{E}X_t$, we obtain the differential equation $m'(t) = (c^2/2)m(t)$, whence $m(t) = \exp(c^2t/2)$. In other words,

$$\mathbb{E}e^{cW_t} = e^{c^2t/2},$$

which implies, on recalling that $W_t \sim N(0,t)$ and setting $\alpha = ct^{1/2}$,

$$\mathbb{E}e^{\alpha Z} = e^{\alpha^2/2},$$

for any complex number $\alpha \in \mathbb{C}$.

□

Corollary 5.12. *The Fourier transform of the univariate Gaussian probability density function*

$$p(x) = (2\pi\sigma^2)^{-1/2} e^{-x^2/(2\sigma^2)}, \qquad x \in \mathbb{R},$$

is

$$\widehat{p}(z) = e^{-\sigma^2 z^2/2}, \qquad z \in \mathbb{R}.$$

Proof. We simply set $\lambda = 1/(2\sigma^2)$ in Proposition 5.11. □

Exercise 5.10. *Calculate the Fourier transform of the multivariate Gaussian probability density function*

$$p(\mathbf{x}) = (2\pi\sigma^2)^{-d/2} e^{-\|\mathbf{x}\|^2/(2\sigma^2)}, \mathbf{x} \in \mathbb{R}^d.$$

The cumulative distribution function (CDF) for the Gaussian probability density $N(0,\sigma^2)$ is given by

$$\Phi_{\sigma^2}(x) = \int_{-\infty}^{x} (2\pi\sigma^2)^{-1/2} e^{-y^2/(2\sigma^2)}\, dy, \qquad x \in \mathbb{R}. \tag{5.67}$$

Thus the fundamental theorem of calculus implies that

$$\Phi'_{\sigma^2}(x) = (2\pi\sigma^2)^{-1/2} e^{-x^2/(2\sigma^2)}.$$

Exercise 5.11. *Calculate the price of the option whose expiry price is given by*

$$f(S(T),T) = \begin{cases} 1 & \text{if } a \leq S(T) \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, this option is simply a bet that pays £1 if the final asset price lies in the interval $[a,b]$*.*

5.6. **Fourier Transform Conventions.** There are several essentially identical Fourier conventions in common use, but their minor differences are often confusing. The most general definition is

$$\widehat{f}(z) = A \int_{-\infty}^{\infty} f(x) e^{iCxz}\, dx, \tag{5.68}$$

where A and C are nonzero real constants. The Fourier Inversion Theorem then takes the form

$$f(x) = \frac{A^{-1}C}{2\pi} \int_{-\operatorname{sign}(C)\infty}^{\operatorname{sign}(C)\infty} \widehat{f}(z) e^{-iCxz}\, dx, \tag{5.69}$$

where

$$\operatorname{sign}(C) = \begin{cases} 1 & C > 0, \\ -1 & C < 0. \end{cases}$$

Example 5.5. *The following four cases are probably the most commonly encountered.*

(i) $C = -1$, $A = 1$:

$$\begin{cases} \widehat{f}(z) = \int_{-\infty}^{\infty} f(x) e^{-ixz}\, dx \\ f(x) = \frac{-1}{2\pi} \int_{+\infty}^{-\infty} \widehat{f}(z) e^{ixz}\, dz = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(z) e^{ixz}\, dz. \end{cases}$$

(ii) $C = 2\pi$, $A = 1$:

$$\begin{cases} \widehat{f}(z) = \int_{-\infty}^{\infty} f(x) e^{2\pi ixz}\, dx \\ f(x) = \int_{-\infty}^{\infty} \widehat{f}(z) e^{-2\pi ixz}\, dz. \end{cases}$$

(iii) $C = 1$, $A = 1/\sqrt{2\pi}$:

$$\begin{cases} \widehat{f}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{ixz}\, dx \\ f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widehat{f}(z) e^{-ixz}\, dz. \end{cases}$$

(iv) $C = 1$, $A = 1$:

$$\begin{cases} \widehat{f}(z) = \int_{-\infty}^{\infty} f(x) e^{ixz}\, dx \\ f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(z) e^{-ixz}\, dz. \end{cases}$$

It's not hard to show that

$$\widehat{T_a f}(z) = e^{-iaCz} \widehat{f}(z) \tag{5.70}$$

and

$$\widehat{f'}(z) = -iCz \widehat{f}(z), \tag{5.71}$$

where $T_a f(x) = f(x + a)$, for any $a \in \mathbb{C}$.

Example 5.6. *For the same four examples given earlier, we obtain the following shifting and differentiation formulae.*

(i) $C = -1$, $A = 1$:

$$\widehat{T_a f}(z) = e^{iaz} \widehat{f}(z), \qquad \widehat{f'}(z) = iz\widehat{f}(z).$$

(ii) $C = 2\pi$, $A = 1$:

$$\widehat{T_a f}(z) = e^{-2\pi iaz} \widehat{f}(z), \qquad \widehat{f'}(z) = -2\pi iz\widehat{f}(z).$$

(iii) $C = 1,\ A = 1/\sqrt{2\pi}$:

$$\widehat{T_a f}(z) = e^{-iaz}\widehat{f}(z), \qquad \widehat{f'}(z) = -iz\widehat{f}(z).$$

(iv) $C = 1,\ A = 1$:

$$\widehat{T_a f}(z) = e^{-iaz}\widehat{f}(z), \qquad \widehat{f'}(z) = -iz\widehat{f}(z).$$

Which, then, should we choose? It's entirely arbitrary but, once made, the choice is likely to be permanent, since changing convention greatly increases the chance of algebraic errors. I have chosen $C = -1$ and $A = 1$ in lectures, mainly because it's probably the most common choice in applied mathematics. It was also the convention chosen by my undergraduate lecturers at Cambridge, so the real reason is probably habit!

6. Mathematical Background Material

I've collected here a miscellany of mathematical methods used (or reviewed) during the course.

6.1. **Probability Theory.** You may find my more extensive notes on Probability Theory useful:

http://econ109.econ.bbk.ac.uk/brad/Probability_Course/probnotes.pdf

A *random variable* X is said to have (continuous) *probability density function* $p(t)$ if

$$\mathbb{P}(a < X < b) = \int_a^b p(t)\,dt. \tag{6.1}$$

We shall assume that $p(t)$ is a continuous function (no jumps in value). In particular, we have

$$1 = \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} p(t)\,dt.$$

Further, because

$$0 \le \mathbb{P}(a < X < a + \delta a) = \int_a^{a+\delta a} p(t)\,dt \approx p(a)\delta a,$$

for small δa, we conclude that $p(t) \ge 0$, for all $t \ge 0$. In other words, a probability density function is simply a non-negative function $p(t)$ whose integral is one. Here are two fundamental examples.

Example 6.1. *The Gaussian probability density function, with mean μ and variance σ^2, is defined by*

$$p(t) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right). \tag{6.2}$$

We say that the Gaussian is normalized *if $\mu = 0$ and $\sigma = 1$.*

To prove that this is truly a probability density function, we require the important identity

$$\int_{-\infty}^{\infty} e^{-Cx^2}\,dx = \sqrt{\pi/C}, \tag{6.3}$$

which is valid for any $C > 0$. [In fact it's valid for any complex number C whose real part is positive.]

Example 6.2. *The Cauchy probability density function is defined by*

$$p(t) = \frac{1}{\pi(1+t^2)}. \tag{6.4}$$

This distribution might also be called the Mad Machine Gunner *distribution; imagine our killer sitting at the origin of the (x, y) plane. He*[7] *is firing (at a constant rate) at the infinite line $y = 1$, his angle θ (with the x-axis) of fire being uniformly distributed in the interval $(0, \pi)$. Then the bullets have the Cauchy density.*

[7] The sexism is quite accurate, since males produce vastly more violent psychopaths than females.

If you draw some graphs of these probability densities, you should find that, for small σ, the graph is concentrated around the value μ. For large σ, the graph is rather flat. There are two important definitions that capture this behaviour mathematically.

Definition 6.1. *The* mean, *or* expected value, *of a random variable X with p.d.f $p(t)$ is defined by*

$$\mathbb{E}X := \int_{-\infty}^{\infty} tp(t)\,dt. \tag{6.5}$$

It's very common to write μ instead $\mathbb{E}X$ when no ambiguity can arise. Its variance *Var X is given by*

$$Var\ X := \int_{-\infty}^{\infty} (t-\mu)^2\, p(t)\,dt. \tag{6.6}$$

Exercise 6.1. *Show that the Gaussian p.d.f. really does have mean μ and variance σ^2.*

Exercise 6.2. *What happens when we try to determine the mean and variance of the Cauchy probability density defined in Example 6.4?*

Exercise 6.3. *Prove that Var $X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.*

We shall frequently have to calculate the expected value of *functions* of random variables.

Theorem 6.1. *If*

$$\int_{-\infty}^{\infty} |f(t)|p(t)\,dt$$

is finite, then

$$\mathbb{E}\left(f(X)\right) = \int_{-\infty}^{\infty} f(t)p(t)\,dt. \tag{6.7}$$

Example 6.3. *Let X denote a normalized Gaussian random variable. We shall show that*

$$\mathbb{E}e^{\lambda X} = e^{\lambda^2/2}, \tag{6.8}$$

Indeed, applying (6.7), we have

$$\mathbb{E}e^{\lambda X} = \int_{-\infty}^{\infty} e^{\lambda t}(2\pi)^{-1/2}e^{-t^2/2}\,dt = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(t^2-2\lambda t)}\,dt.$$

The trick now is to complete the square *in the exponent, that is,*

$$t^2 - 2\lambda t = (t-\lambda)^2 - \lambda^2.$$

Thus

$$\mathbb{E}e^{\lambda X} = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}([t-\lambda]^2 - \lambda^2)\right)\,dt = e^{\lambda^2/2}.$$

Exercise 6.4. *Let W be any Gaussian random variable with mean zero. Prove that*

$$\mathbb{E}\left(e^W\right) = e^{\frac{1}{2}\mathbb{E}\left(W^2\right)}. \tag{6.9}$$

6.2. **Differential Equations.** A *differential equation*, or *ordinary differential equation* (ODE), is simply a functional relationship specifying first, or higher derivatives, of a function; the *order* of the equation is just the degree of its highest derivatives. For example,

$$y'(t) = 4t^3 + y(t)^2$$

is a univariate first-order differential equation, whilst or

$$\mathbf{y}'(t) = A\mathbf{y}(t),$$

where $\mathbf{y}(t) \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d\times d}$ is a first-order differential equation in d-variables. A tiny class of differential equations can be solved analytically, but numerical methods are required for the vast majority. The numerical analysis of differential equations has been one of the most active areas of research in computational mathematics since the 1960s and excellent free software exists. It is extremely unlikely that any individual can better this software without years of effort and scholarship, so you should use this software for any practical problem. You can find lots of information at `www.netlib.org` and `www.nr.org`. This section contains the minimum relevant theory required to make use of this software.

You should commit to memory one crucial first-order ODE:

Proposition 6.2. *The general solution to*

$$y'(t) = \lambda y(t), \qquad t \in \mathbb{R}, \tag{6.10}$$

where λ can be any complex number, is given by

$$y(t) = c\exp(\lambda t), \qquad t \in \mathbb{R}. \tag{6.11}$$

Here $c \in \mathbb{C}$ is a constant. Note that $c = y(0)$, so we can also write the equation as $y(t) = y(0)\exp(\lambda t)$.

Proof. If we multiply the equation $y' - \lambda y = 0$ by the *integrating factor* $\exp(-\lambda t)$, then we obtain

$$0 = \frac{d}{dt}\left(y(t)\exp(-\lambda t)\right),$$

that is

$$y(t)\exp(-\lambda t) = c,$$

for all $t \in \mathbb{R}$. □

In fact, there's a useful **slogan** for ODEs: try an exponential $\exp(\lambda t)$ or use reliable numerical software.

Example 6.4. *If we try $y(t) = \exp(\lambda t)$ as a trial solution in*

$$y'' + 2y' - 3y = 0,$$

then we obtain

$$0 = \exp(\lambda t)\left(\lambda^2 + 2\lambda - 3\right).$$

Since $\exp(\lambda t) \neq 0$*, for any t, we deduce the* associated equation

$$\lambda^2 + 2\lambda - 3 = 0.$$

The roots of this quadratic are 1 and -3, which is left as an easy exercise. Now this ODE is linear: any linear combination of solutions is still a solution. Thus we have a general family of solutions

$$\alpha\exp(t) + \beta\exp(-3t),$$

for any complex numbers α and β. We need two pieces of information to solve for these constants, such as $y(t_1)$ and $y(t_2)$, or, more usually, $y(t_1)$ and $y'(t_1)$. In fact this is the general solution of the equation.

In fact, we can always change an mth order equation is one variable into an equivalent first order equation in m variables, a technique that I shall call *vectorizing* (some books prefer the more pompous phrase "reduction of order"). Most ODE software packages are designed for first order systems, so vectorizing has both practical and theoretical importance.

For example, given

$$y''(t) = \sin(t) + (y'(t))^3 - 2\,(y(t))^2\,,$$

we introduce the *vector function*

$$\mathbf{z}(t) = \begin{pmatrix} y(t) \\ y'(t) \end{pmatrix},$$

Then

$$\mathbf{z}'(t) = \begin{pmatrix} y' \\ y'' \end{pmatrix} = \begin{pmatrix} y' \\ \\ \sin(t) + (y')^3 - 2\,(y)^2 \end{pmatrix}.$$

In other words, writing

$$\mathbf{z}(t) = \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} \equiv \begin{pmatrix} y(t) \\ y'(t) \end{pmatrix},$$

we have derived

$$\mathbf{z}' = \begin{pmatrix} z_2 \\ \sin(t) + z_2^3 - 2z_1^2 \end{pmatrix},$$

which we can write as

$$\mathbf{z}' = \mathbf{f}(\mathbf{z}, t).$$

Exercise 6.5. *You probably won't need to consider ODEs of order exceeding two very often in finance, but the same trick works. Given*

$$y^{(n)}(t) = \sum_{k=0}^{n-1} a_k(t) y^{(k)}(t),$$

we define the vector function $\mathbf{z}(t) \in \mathbb{R}^{n-1}$ by

$$z_k(t) = y^{(k)}(t), \qquad k = 0, 1, \ldots, n-1.$$

Then $\mathbf{z}'(t) = M\mathbf{z}(t)$. Find the matrix M.

6.3. **Recurrence Relations.** In its most general form, a *recurrence relation* is simply a sequence of vectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots$ for which some functional relation generates $\mathbf{v}^{(n)}$ given the earlier iterates $\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n-1)}$. At this level of generality, very little more can be said. However, the theory of linear recurrence relations is simple and very similar to the techniques of differential equations.

The first order linear recurrence relation is simply the sequence $\{a_n : n = 0, 1, \ldots\}$ of complex numbers defined by

$$a_n = c a_{n-1}.$$

Thus

$$a_n = c a_{n-1} = c^2 a_{n-2} = c^3 a_{n-3} = \cdots = c^n a_0$$

and the solution is complete.

The second order linear recurrence relation is slightly more demanding. Here

$$a_{n+1} + pa_n + qa_{n-1} = 0$$

and, inspired by the solution for the first order recurrence, we try $a_n = c^n$, for some $c \neq 0$. Then

$$0 = c^{n-1}\left(c^2 + pc + q\right),$$

or

$$0 = c^2 + pc + q.$$

If this has two distinct roots c_1 and c_2, then one possible solution to the second order recurrence is

$$u_n = p_1 c_1^n + p_2 c_2^n,$$

for constants p_1 and p_2. However, is this the full set of solutions? What happens if the quadratic has only one root?

Proposition 6.3. *Let $\{a_n : n \in \mathbb{Z}\}$ be the sequence of complex numbers satisfying the recurrence relation*

$$a_{n+1} + pa_n + qa_{n-1} = 0, \qquad n \in \mathbb{Z}.$$

If α_1 and α_2 are the roots of the associated quadratic

$$t^2 + pt + q = 0,$$

then the general solution is

$$a_n = c_1\alpha_1^n + c_2\alpha_2^n$$

when $\alpha_1 \neq \alpha_2$. If $\alpha_1 = \alpha_2$, then the general solution is

$$a_n = (v_1 n + v_2)\,\alpha_1^n.$$

Proof. The same vectorizing trick used to change second order differential equations in one variable into first order differential equations in two variables can also be used here. We define a new sequence $\{\mathbf{b}^{(n)} : n \in \mathbb{Z}\}$ by

$$\mathbf{b}^{(n)} = \begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix}.$$

Thus

$$\mathbf{b}^{(n)} = \begin{pmatrix} a_{n-1} \\ -pa_{n-1} - qa_{n-2} \end{pmatrix},$$

that is,

$$\mathbf{b}^{(n)} = A\mathbf{b}^{(n-1)}, \tag{6.12}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -q & -p \end{pmatrix}. \tag{6.13}$$

This first order recurrence has the simple solution

$$\mathbf{b}^{(n)} = A^n \mathbf{b}^{(0)}, \tag{6.14}$$

so our analytic solution reduces to calculation of the matrix power A^n. Now let us begin with the case when the eigenvalues λ_1 and λ_2 are distinct. Then the corresponding eigenvectors $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ are linearly independent. Hence we can write our initial vector $\mathbf{b}^{(0)}$ as a unique linear combination of these eigenvectors:

$$\mathbf{b}^{(0)} = b_1\mathbf{w}^{(1)} + b_2\mathbf{w}^{(2)}.$$

Thus
$$\mathbf{b}^{(n)} = b_1 A^n \mathbf{w}^{(1)} + b_2 A^n \mathbf{w}^{(2)} = b_1 \lambda_1^n \mathbf{w}^{(1)} + b_1 \lambda_2^n \mathbf{w}^{(2)}.$$
Looking at the second component of the vector, we obtain
$$a_n = c_1 \lambda_1^n + c_2 \lambda_2^n.$$
Now the eigenvalues of A are the roots of the quadratic equation
$$\det(A - \lambda I) = \det \begin{pmatrix} -\lambda & 1 \\ -q & -p-\lambda \end{pmatrix},$$
in other words the roots of the quadratic
$$\lambda^2 + p\lambda + q = 0.$$
Thus the associated equation is precisely the characteristic equation of the matrix A in the vectorized problem. Hence $a_n = c_1 \alpha_1^n + c_2 \alpha_2^n$.

We only need this case in the course, but I shall lead you through a careful analysis of the case of coincident roots. It's a good exercise for your matrix skills.

First note that the roots are coincident if and only if $p^2 = 4q$, in which case
$$A = \begin{pmatrix} 0 & 1 \\ -p^2/4 & -p \end{pmatrix},$$
and the eigenvalue is $-p/2$. In fact, subsequent algebra is simplified if we substitute $\alpha = -p/2$, obtaining
$$A = \begin{pmatrix} 0 & 1 \\ -\alpha^2 & 2\alpha \end{pmatrix}.$$
The remainder of the proof is left as the following exercise. □

Exercise 6.6. *Show that*
$$A = \alpha I + \mathbf{u}\mathbf{v}^T,$$
where
$$\mathbf{u} = \begin{pmatrix} 1 \\ \alpha \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} -\alpha \\ 1 \end{pmatrix}$$
and note that $\mathbf{v}^T\mathbf{u} = 0$. *Show also that*
$$A^2 = \alpha^2 I + 2\alpha \mathbf{u}\mathbf{v}^T, \qquad A^3 = \alpha^3 I + 3\alpha^2 \mathbf{u}\mathbf{v}^T,$$
and use proof by induction to demonstrate that
$$A^n = \alpha^n I + n\alpha^{n-1}\mathbf{u}\mathbf{v}^T.$$
Hence find the general solution for a_n.

6.4. **Mortgages – a once exotic instrument.** The objective of this section is to illustrate some of the above techniques for analysing difference and differential equations via mortgage pricing. You are presumably all too familiar with a repayment mortgage: we borrow a large sum M for a fairly large slice T of our lifespan, repaying capital and interest using N regular payments. The interest rate is assumed to be constant and it's a secured loan: our homes are forfeit on default. How do we calculate our repayments?

Let $h = T/N$ be the interval between payments, let $D_h : [0, T] \to \mathbb{R}$ be our debt as a function of time, and let $A(h)$ be our payment. We shall assume that our initial debt is $D_h(0) = 1$, because we can always multiply by the true initial cost M of our house after the calculation. Thus D must satisfy the equations

$$D_h(0) = 1, \quad D_h(T) = 0 \quad \text{and} \quad D_h(\ell h) = D_h((\ell-1)h)e^{rh} - A(h). \tag{6.15}$$

We see that $D_h(h) = e^{rh} - A(h)$, while

$$D_h(2h) = D_h(h)e^{rh} - A(h) = e^{2rh} - A(h)\left(1 + e^{rh}\right).$$

The pattern is now fairly obvious:

$$D_h(\ell h) = e^{\ell rh} - A(h)\sum_{k=0}^{\ell-1} e^{krh}, \tag{6.16}$$

and summing the geometric series[8]

$$D_h(\ell h) = e^{\ell rh} - A(h)\left(\frac{e^{\ell rh} - 1}{e^{rh} - 1}\right). \tag{6.17}$$

In order to achieve $D(T) = 0$, we choose

$$A(h) = \frac{e^{rh} - 1}{1 - e^{-rT}}. \tag{6.18}$$

Exercise 6.7. *What happens if $T \to \infty$?*

Exercise 6.8. *Prove that*

$$D_h(\ell h) = \frac{1 - e^{-r(T-\ell h)}}{1 - e^{-rT}}. \tag{6.19}$$

Thus, if $t = \ell h$ is constant (so we increase ℓ as we reduce h), then

$$D_h(t) = \frac{1 - e^{-r(T-t)}}{1 - e^{-rT}}. \tag{6.20}$$

Almost all mortgages are repaid by 300 monthly payments for 25 years. However, until recently, many mortgages calculated interest *yearly*, which means that we choose $h = 1$ in Exercise 6.7 and then divide $A(1)$ by 12 to obtain the monthly payment.

Exercise 6.9. *Calculate the monthly repayment $A(1)$ when $M = 10^5$, $T = 25$, $r = 0.05$ and $h = 1$. Now repeat the calculation using $h = 1/12$. Interpret your result.*

[8] Many students forget the simple formula. If $S = 1 + a + a^2 + \cdots + a^{m-2} + a^{m-1}$, then $aS = a + a^2 + \cdots + a^{m-1} + a^m$. Subtracting these expressions implies $(a-1)S = a^m - 1$, all other terms cancelling.

In principle, there's no reason why our repayment could not be continuous, with interest being recalculated on our constantly decreasing debt. For continuous repayment, our debt $D : [0, T] \to \mathbb{R}$ satisfies the relations

$$D(0) = 1, \quad D(T) = 0 \quad \text{and} \quad D(t + h) = D(t)e^{rh} - hA. \tag{6.21}$$

Exercise 6.10. *Prove that*

$$D'(t) - rD(t) = -A, \tag{6.22}$$

where, in particular, you should prove that (6.21) implies the differentiability of $D(t)$. Solve this differential equation using the integrating factor e^{-rt}. You should find the solution

$$D(t)e^{-rt} - 1 = A \int_0^t \left(-e^{-r\tau}\right) d\tau = A\left(\frac{e^{-rt} - 1}{r}\right). \tag{6.23}$$

Hence show that

$$A = \frac{r}{1 - e^{-rT}} \tag{6.24}$$

and

$$D(t) = \frac{1 - e^{-r(T-t)}}{1 - e^{-rT}}, \tag{6.25}$$

agreeing with (6.20)*, i.e.* $D_h(kh) = D(kh)$*, for all* k*. Prove that* $\lim_{r\to\infty} D(t) = 1$ *for* $0 < t < T$ *and interpret.*

Observe that

$$\frac{A(h)}{Ah} = \frac{e^{rh} - 1}{rh} \approx 1 + (rh/2), \tag{6.26}$$

so that continuous repayment is optimal for the borrower, but that the mortgage provider is making a substantial profit. Greater competition has made yearly recalculations much rarer, and interest is often paid daily, i.e. $h = 1/250$, which is rather close to continuous repayment.

Exercise 6.11. *Construct graphs of $D(t)$ for various values of r. Calculate the time $t_0(r)$ at which half of the debt has been paid.*

6.5. **Pricing Mortgages via lack of arbitrage.** There is a very slick arbitrage argument to deduce the continuous repayment mortgage debt formula (6.25). Specifically, the simple fact that $D(t)$ is a deterministic financial instrument implies, via arbitrage, that $D(t) = a + b\exp(rt)$, so we need only choose the constants a and b to satisfy $D(0) = 1$ and $D(T) = 1$, which imply $a + b = 1$ and $a + b\exp(rT) = 0$. Solving these provides $a = \exp(rT)/(\exp(rT) - 1)$ and $b = -1/(\exp(rT) - 1)$, and equivalence to (6.25) is easily checked.

7. Numerical Linear Algebra

I shall not include much explicitly here, because you have my longer lecture notes on numerical linear algebra:

`http://econ109.econ.bbk.ac.uk/brad/Methods/nabook.pdf`

Please do revise the first long chapter of those notes if need to brush up on matrix algebra.

You will also find my Matlab notes useful:

`http://econ109.econ.bbk.ac.uk/brad/Methods/matlab_intro_notes.pdf`

7.1. Orthogonal Matrices.

Modern numerical linear algebra began with the computer during the Second World War, its progress accelerating enormously as computers became faster and more convenient in the 1960s. One of the most vital conclusions of this research field is the enormous practical importance of matrices which leave Euclidean length invariant. More formally:

Definition 7.1. *We shall say that* $Q \in \mathbb{R}^{n\times n}$ *is* distance-preserving *if* $\|Q\mathbf{x}\| = \|\mathbf{x}\|$, *for all* $\mathbf{x} \in \mathbb{R}^n$.

The following simple result is very useful.

Lemma 7.1. *Let* $M \in \mathbb{R}^{n\times n}$ *be any symmetric matrix for which* $\mathbf{x}^T M\mathbf{x} = 0$, *for every* $\mathbf{x} \in \mathbb{R}^n$. *Then* M *is the zero matrix.*

Proof. Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ be the usual coordinate vectors. Then

$$M_{jk} = \mathbf{e}_j^T M\mathbf{e}_k = \frac{1}{2}\left(\mathbf{e}_j + \mathbf{e}_k\right)^T M\left(\mathbf{e}_j + \mathbf{e}_k\right) = 0, \qquad 1 \le j, k \le n.$$

□

Theorem 7.2. *The matrix* $Q \in \mathbb{R}^n$ *is distance-preserving if and only if* $Q^T Q = I$.

Proof. If $Q^T Q = I$, then

$$\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^T(Q\mathbf{x}) = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2,$$

and Q is distance-preserving. Conversely, if $\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$, for all $\mathbf{x} \in \mathbb{R}^n$, then

$$\mathbf{x}^T\left(Q^T Q - I\right)\mathbf{x} = 0, \qquad \mathbf{x} \in \mathbb{R}^n.$$

Since $Q^T Q - I$ is a symmetric matrix, Lemma 7.1 implies $Q^T Q - I = 0$, i.e. $Q^T Q = I$. □

The condition $Q^T Q = I$ simply states that the columns of Q are orthonormal vectors, that is, if the columns of Q are $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$, then $\|\mathbf{q}_1\| = \cdots = \|\mathbf{q}_n\| = 1$ and $\mathbf{q}_j^T\mathbf{q}_k = 0$ when $j \ne k$. For this reason, Q is also called an *orthogonal matrix*. We shall let $O(n)$ denote the set of all (real) $n \times n$ orthogonal matrices.

Exercise 7.1. *Let* $Q \in O(n)$. *Prove that* $Q^{-1} = Q^T$. *Further, prove that* $O(n)$ *is closed under matrix multiplication, that is,* $Q_1 Q_2 \in O(n)$ *when* $Q_1, Q_2 \in O(n)$. *(In other words,* $O(n)$ *forms a group under matrix multiplication. This observation is important, and* $O(n)$ *is often called the* Orthogonal Group.*)*

Department of Economics, Mathematics and Statistics, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England

Email address: `b.baxter@bbk.ac.uk`