Stochastic differential equations and numerical methods to approximate their solution

G14DIS = MATH4001

Mathematics 4th Year Dissertation

2021/22

School of Mathematical Sciences

University of Nottingham

Jake Denton

Supervisor: Professor Michael Tretyakov

Assessment type: Review

I have read and understood the School and University guidelines on plagiarism. I confirm that this work is my own, apart from the acknowledged references.

Abstract

In applications such as finance and molecular dynamics, many factors influencing the underlying process cannot be measured or modelled easily. As a result, use of models based on the theory of stochastic analysis, wherein the tiny effects of many factors are replaced by a random term, have been used in these settings effectively. This report aims to firstly introduce the foundations of stochastic analysis, including the probability theory surrounding it and its main object, Brownian motion, the involvement of which leads to a distinct class of differential equations - stochastic differential equations. The second main aim of this report is to derive and test numerical schemes for stochastic differential equations. These methods are split across two chapters, the first of which introduces the Euler and Milstein schemes, which can be used to simulate entire trajectories of a process. The second deals with weak methods, which can be used to estimate the moments of a process. Investigation of these schemes using the case studies of geometric Brownian motion and European option pricing with the Black-Scholes equation (from finance) allows us to confirm theoretical results on accuracy. The Euler scheme has mean-square order of accuracy of $\frac{1}{2}$ whilst the Milstein scheme has mean-square order of accuracy of 1. On the other hand, both of these schemes possess first weak order of accuracy, alongside a simpler weak Euler scheme. Finally, by using a number of different estimates obtained through the same method and manipulating results on their global error, we find weak order schemes of orders 2 and 3 through the Talay-Tubaro extrapolation method.

Contents

1 Introduction

Ordinary and partial differential equations allow us to see how a variable of interest evolves with respect to other independent variables, and have allowed mathematicians to model physical processes for hundreds of years, from the effect of gravity on enormous bodies to the evolution of wave functions on the quantum scale.

Some real-life processes depend on so many variables that it is not feasible or practical to consider the effect of every single one in turn. For example, in finance, stock prices depend on a huge number of factors, influenced not only by supply and demand but also broader economic trends. In this situation, how does one model the evolution of the stock price? It turns out that we can model this process effectively by taking a system which we already know something about (such as the general trend of the stock and how volatile it is) and introducing a random term to capture the tiny influences of other variables. This idea is the basis for an alternative family of differential equations, namely the stochastic differential equations.

This field, named Stochastic Analysis, is relatively new. The theory and the main object of this analysis, Brownian motion, was first studied in the PhD thesis of Louis Bachelier [2] in 1900. Other contributions and developments came from Robert Brown, Albert Einstein, Norbert Wiener and Jean Perrin (see [15, 51]). Later, in 1965, economist Paul Samuelson [56] introduced geometric Brownian motion, one model describing the movement of stock prices. This was used by Black, Scholes and Merton in 1973 [7] to derive a formula for the pricing of options which is still in use today. This subject provided the fundamentals for modern financial mathematics, and revolutionised the pricing of derivatives. It will come as no surprise that as a result, there is continued demand from financial institutions for research into the subject. The first aim of this report is to review the main concepts of stochastic analysis in detail, setting up the stage to introduce numerical methods.

One issue shared across every class of differential equation is how to approximate the solution numerically when we cannot find it exactly. For ordinary and partial differential equations, there are multi-stage and multi-step methods, finite difference and volume meth-

ods, just to give a few examples. Finding numerical schemes for their stochastic counterparts and testing their accuracy is the other main aim of this report.

To meet the first aim, this report begins with a review of the probability theory behind stochastic analysis [14]. Following this, Brownian motion is motivated, constructed and its properties are discussed in detail. Next, stochastic integrals are defined along with the basics of Ito calculus and stochastic differential equations are introduced [19]. This material is covered in chapters 2-4.

As for the second aim, the discussion of numerical approximation [66] to the solution of stochastic differential equations begins in chapter 5. The Euler scheme is generalised and implemented to approximate geometric Brownian motion, and then an extension, the Milstein scheme, is constructed and the same analysis is repeated for comparison. These methods can be used to simulate whole trajectories of a process - approximation in the mean-square sense. There is also a second type of approximation, the weak approximation, in which the moments of a stochastic process can be estimated. This is the topic of chapter 6. We discuss the schemes from chapter 5 as well as the weak Euler scheme, and a method for constructing more accurate approximations which is analogous to Richardson extrapolation [64].

The report is concluded with a reflection of the material covered and what further work might follow. A link to the full list of MATLAB codes used to generate the figures and tables in this report can be found in the appendix.

2 Review of Probability Theory

In order to develop the relevant background knowledge for defining and understanding stochastic differential equations (SDEs) and numerical methods associated with them, a discussion of probability is necessary. There are hundreds of books on probability and measure theory, here just a sample of their content is provided. Sources [19, 66, 67] provide short summaries of the probability theory covered here. For further reading on these topics please see sources [14] and [70], which provide more detail than is required here.

2.1 Probability spaces and σ-algebras

First, let us consider the following question - what setting do we refer to when we speak of 'randomness'? We approach this by introducing the following definitions.

Definition 2.1 (Elementary Events and Sample Space)**.** When an experiment is conducted, any outcome is an elementary event or sample point, denoted ω. The collection of all the elementary events is called the sample space, denoted Ω [19].

Definition 2.2 (Events)**.** An event A is a smaller collection of elementary events (a subset of the sample space Ω), for which each elementary event ω in the space Ω is in turn either in A ($\omega \in A$) or not in A ($\omega \notin A$).

Example 2.1. (Drawing cards from a deck) Suppose a card is drawn at random from a full deck. An example of an elementary event is that the card picked is the 10 of Hearts. An example of an event is that the card picked has a red suit. The sample space here is the collection of all the individual cards in the deck.

Definition 2.3. (σ-algebras) For a non-empty sample space Ω, let U be a collection of subsets of Ω. U is a σ-algebra if it has the following properties:

1. Ω and the empty set $\emptyset$ are in U. (U possesses empty and 'full' elements)

2. For a subset A$\in$U, the complement of A, $A^c = \Omega - A$ (meaning the elements left over when you remove A from Ω) is also in U. (U is closed under complement)

3. If $A_1, A_2, \ldots \in$U, then the infinite union $\cup_{k=1}^{\infty} A_k$ and infinite intersection $\cap_{k=1}^{\infty} A_k$ are also in U (U is closed under union and intersection).

Example 2.2. The trivial algebra U=$\{\Omega, \emptyset\}$ is a σ-algebra.

Definition 2.4. (Probability Measures) A function P that assigns a number in the interval [0,1] to each event in U is a probability measure provided:

1. P($\emptyset$)=0 and P(Ω)=1
2. If $A_1, A_2, \ldots \in$U, then $P(\cup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)$, which becomes an inequality if the sets $A_1, A_2, \ldots$ are disjoint meaning they share no elements.

Further remarks:

1. P(A) is called the probability of event A.
2. If A,B $\in$U and A$\subseteq$B (A is a subset of B), then P(A)$\leq$P(B).
3. A sample space coupled with a σ-algebra of it's subsets is a measurable space, (Ω,U).
4. The ordered triple (Ω, U,P) is called a probability space.
5. If U and G are σ-algebras of the sample space Ω and $G \subseteq U$ then G is called a sub-σ-algebra of U.

Consider now a system $\mathcal{A}$ of all the intervals on the real line $\mathbb{R}$ (plus the empty set). This can be written as unions of disjoint intervals (a,b] i.e. A $\in \mathcal{A}$ if A=$\cup_{i=1}^{n}(a_i, b_i]$. Then the smallest (meaning number of elements) σ-algebra that contains $\mathcal{A}$ is the Borel σ-algebra (in this case for the real line). Note that all open and closed subsets including those with infinite limits and points can be written as infinite unions or intersections of intervals of the form (a,b] and so also belong to the Borel-σ algebra. This is denoted as B($\mathbb{R}$). Subsets of B($\mathbb{R}$) are called Borel sets. This can be extended to higher dimensions by considering open subsets of $\mathbb{R}^n$ and proving that all other possible intervals can be constructed from these. Moving on, the above material can be used to define a random variable.

2.2 Random variables

Now that we have the backdrop of probability spaces to work with, we can understand the following set of definitions.

Definition 2.5. (Random Variable) An n-dimensional random variable X is a mapping from the Ω to $\mathbb{R}^n$ that is measurable with respect to the σ-algebra U. In other words, the set of elementary events ω with $X(\omega) \in B$ are in U (or equivalently, $X^{-1}(B) \in U$, where $X^{-1}(B) = \{\omega : X(\omega) \in B\}$) for every Borel set $B \in B(\mathbb{R}^n)$.

The requirement that X is U-measurable is due to the fact that the probability measure P is defined on the measurable space (Ω,U) and so since P assigns a probability to $\{X(\omega) \in B\}$, the condition in the definition is necessary [66].

Lemma 2.1. Let X be a random variable as described above. Then U(X)=$\{X^{-1}(B)|B \in B(\mathbb{R}^n)\}$ is a σ-algebra generated by X. This is a sub-σ-algebra of U, the smallest which measures X. U(X) effectively contains all the relevant information about X.

If a random variable Y is a function of X, which is U-measurable and so U(X)-measurable, and this function Φ is a reasonable function, then Y is also U(X)- measurable [19]. Conversely, if it is known that Y is U(X)-measurable then there exists a function Φ such that Y=$\Phi(X)$. To illustrate the meaning of reasonable here with respect to this function Φ, lets consider how to compute probabilities relating to Y in terms of a real-valued X:

$$\begin{aligned} P(Y \in B) &= P(\{\omega \in \Omega : Y(\omega) \in B\}) \\ &= P(\{\omega \in \Omega : \Phi(X(\omega) \in B\}) \\ &= P(\{\omega \in \Omega : X(\omega) \in \Phi^{-1}(B)\}) \\ &= P(X \in \Phi^{-1}(B)) \end{aligned}$$

where $\Phi^{-1}(B) = \{r \in \mathbb{R} : \Phi(r) \in B\}$. Reasonable therefore implies that this set can be found from the function Φ.

The probability measure P_x on the measurable space ($\mathbb{R}^n$,B($\mathbb{R}^n$)) defined as $P_x(B) = P(\omega : X(\omega) \in B)$ is the probability distribution of X. The distribution function F: $\Omega \to [0, 1]$ is

given by P($\omega : X(\omega) \leq x$) for $x \in \mathbb{R}^n$.

There are two distinct classes of random variables: discrete and continuous. Discrete variables take values from a countable set (finite or countably infinite e.g. $\mathbb{N}$ or $\mathbb{Z}$) whilst continuous variables have continuous distribution functions F. In this report, we focus almost entirely (with the exception of random walks, which we come to in chapter 6) on continuous random variables.

A random variable is absolutely continuous when there is a non-negative, integrable function ρ (density function) such that $F(x) = \int_{-\infty}^{x} \rho(x)dx$ (simplified here to one dimension, but this can easily be extended by using multiple integrals).

Example 2.3. (Absolutely continuous random variable) An example of an absolutely continuous random variable is the time between two successive lightning strikes hitting the Earth. Here the time can be modelled using an exponential distribution with density function:

$$\rho(x) = \lambda e^{-\lambda x} \text{ for } x > 0 \text{ and } \rho(x) = 0 \text{ otherwise.}$$

It is possible for a random variable to be continuous but not absolutely continuous (one example is a random variable possessing a Cantor distribution function [60]).

Definition 2.6. (Independence of Random Variables) The random variables $\eta_1, \eta_2, \ldots \eta_n$ are independent if either of the following properties hold:

1. Their joint distribution function F can be written as the product of the individual distributions:

   $F_{\eta_1,...,\eta_n}(x_1, ..., x_n) = \prod_{k=1}^{n} P(\eta_k \leq x_k)$

2. Joint density is the product of the individual densities:

   $\rho_{\eta_1,...,\eta_n}(x_1, ..., x_n) = \prod_{k=1}^{n} \rho_{\eta_k}(x_k)$

2.3 Expectation

One measure that can give some information about the behaviour of a random variable is the expectation. This subsection introduces the notion of expectation, and later we will

discuss conditional expectation. This part is also essential to later define convergence, a tool that tells us how quickly any method discussed approaches the object we wish to approximate.

Definition 2.7. (Expectations)

1. (Discrete) The expectation of a discrete random variable with probability distribution P on a sample space with M events ω can be calculated using the following sum: $E(X(\omega)) = \sum_{i=1}^{M} X(\omega_i)P(\omega_i)$

2. (Absolutely continuous) The above can be translated into the definition for an absolutely continuous random variable by replacing the sum with an integral and using the density function ρ: $E(X(\omega)) = \int_{-\infty}^{\infty} x\rho(x)dx$

Unfortunately, standard Riemann integration is not sufficient to understand expectation. This is because random variables are functions of ω, elementary events in the sample space. Thankfully, the concept of Lebesgue integration can be used instead.

Suppose that $X(\omega) \in [0, \infty)$ for all ω in the sample space. Create a partition of the positive semi-line: $0 = y_0 < y_1 < ...$ with δ the maximum distance between any two consecutive y_i. For each interval $[y_i, y_{i+1}]$, define the event $A_i = (\omega \in \Omega : X(\omega) \in [y_i, y_{i+1}])$. Then the lower Lebesgue sum is given by $F_n = \sum_{i=1}^{n} y_i P(A_i)$, and as $\delta \to 0$ this sum approaches the Lebesgue integral $\int_\Omega X(\omega)dP(\omega)$. Note that despite the sum looking very similar to discrete variable expectation, the integral is not quite what was presented in definition 2.7, which involved Riemann integration. What's the connection between the two? Starting from definition 2.7, the following theorem can be used when we have a function on a closed, bounded interval.

Theorem 2.2. If f is bounded on $[a, b]$ and is Riemann-integrable, then f is Lebesgue integrable and (denoting Riemann integration by (R)) [3]:

$$(R)\int_a^b f(x)dx = \int_{[a,b]} f(x)dx$$

It is clear from this that if an absolutely continuous X takes values in a closed subset of $\mathbb{R}$, the two definitions are equivalent. When the Riemann integral is improper (i.e. with

limits approaching infinity), or if f is discontinuous, this theorem is not enough to provide the connection. We don't go further here, but the intrigued reader can find out more from countless sources including [70, 61].

An illustration of the Lebesgue sum is given below. To motivate how this sum is formed, consider the first shaded box. The line $y = X(\omega)$ is between y_3 and y_4, so the product $y_3 * P(A_3)$ is added to the sum.

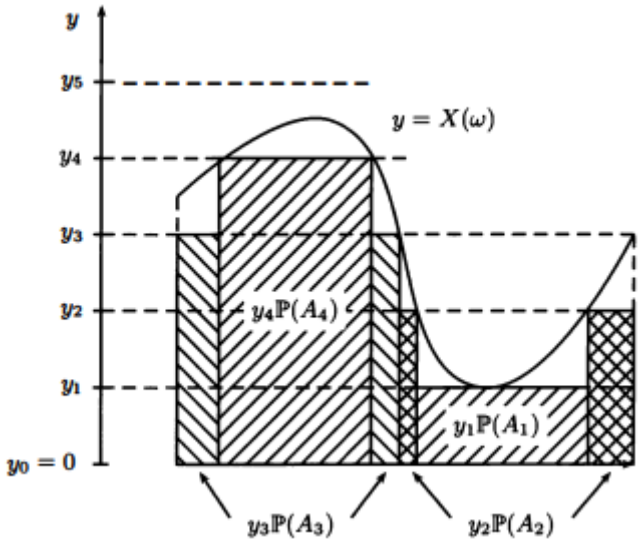

Figure 1: Calculating the Lebesgue sum [Source: [61] Chapter 1]

Of course, the assumption that the random variable is positive for all events ω can be lifted by splitting positive and negative values: $X^+(\omega) = \max(X(\omega), 0)$ and $X^-(\omega) = \max(-X(\omega), 0)$. The random variable X can then be written as $X(\omega) = X^+(\omega) - X^-(\omega)$ and integrated [67].

Definition 2.8. (Integrability) The expectation $E(X(\omega))$ is integrable if $E(|X(\omega)|) < \infty$.

Taking random variables X and Y along with real constants a and b plus a real convex function φ (convex meaning the line joining any two points on the function's graph is not below the graph between the two points), we can list some of the properties of expectation:

1. (Constant) $E(a) = a$.
2. (Linearity) $E(aX + bY) = aE(X) + bE(Y)$.
3. (Comparison) If $X \leq Y$ then $E(X) \leq E(Y)$.

4. (Jensen's inequality) $\varphi(E(X)) \leq E(\varphi(X))$

It is worth mentioning, in extension of the first property, that if the argument of the expectation is not random then taking the expectation returns the deterministic function.

Another important measure of a random variable is variance. This is the expectation of the squared distance between a random variable and its mean, and is one measure of the spread of a random variable.

Definition 2.9. (Variance) The variance of a random variable X, $Var(X)$, is given by:

$$Var(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

Example 2.4. (Normal Distribution) One extremely popular distribution used to model a number of real-life variables is the normal distribution. It takes the form:

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Here, the mean is μ and the variance is σ^2. Clearly, this distribution function only depends on two moments (specifically the first moment and second central moment), which determine the centre point and how spread out the density function is around this point. This property makes this distribution extremely useful.

Random variables defined on the measurable space ($\mathbb{R}^n$,B($\mathbb{R}^n$)) are called Borel functions and are Borel-measurable [60]. For example, the normal variable with probability distribution given in example 2.4 is a Borel-measurable function (in one dimension).

Proposition 2.3. For general Borel-measurable functions f on $\mathbb{R}$ with $E(|f(X)|) = \int_{\mathbb{R}} |f(x)| dP_X(x) < \infty$, the expectation of $f(X)$ is $E(f(X)) = \int_{\mathbb{R}} f(x) dP_X(x)$. If $X(\omega)$ has density function ρ, then $E(f(X)) = \int_{-\infty}^{\infty} f(x)\rho(x)dx$.

The above proposition is sometimes referred to as the law of the unconscious statistician (or LOTUS) in literature. The proof of the result in the continuous case is a simple application of the chain rule then a substitution of variables, and can be found in [55]. This result allows us to calculate the expectation for quite general f, as we see in the next part.

2.4 Characteristic Functions

One easy method to identify different types of random variables is by using an integral transform known as the characteristic function.

Definition 2.10. Let **X** be an $\mathbb{R}^n$-valued random variable. Then

$$\phi_{\mathbf{X}}(s) = E(e^{is\mathbf{X}})$$

is the characteristic function of **X**. Here i is the imaginary number.

Example 2.5. (Standard normal variable) Suppose we have X with a $\mathcal{N}(0,1)$ distribution (see example 2.4, setting $\mu = 0$ and $\sigma^2 = 1$). Then

$$\begin{aligned}\phi_X(s) &= \int_{-\infty}^{\infty} e^{isx}\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{isx-\frac{x^2}{2}}dx \\ &= e^{-\frac{s^2}{2}}\end{aligned}$$

where the integral in line 2 was computed using the central formula for Gaussian integration:

$$\int_{-\infty}^{\infty} e^{-ax^2+bx}dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}}.$$

This is valid for complex $a, a \neq 0$ and b. The real part of a must also be non-negative. The proof is simple for real values, involving completing the square and substitutions. For complex values, some knowledge of contour integration is required, see [63].

Lemma 2.4. 1. If X_1,...,X_m are independent random variables then for each $s \in \mathbb{R}^n$,

$$\phi_{X_1+\ldots+X_m}(s) = \prod_{i=1}^{m}\phi_{X_i}(s).$$

2. If X is a real-valued random variable then

$$\phi_X^{(k)}(0) = i^k E(X^k) \text{ for k=0,1,....}$$

Here, the left term is the k^{th}-derivative of the characteristic function evaluated at $s = 0$.

3. If X and Y are random variables with equal characteristic functions

$$\phi_X(s) = \phi_Y(s) \text{ for all } s \in \mathbb{R}^n,$$

then they share the same distribution. In other words, the characteristic function uniquely determines the distribution of a random variable.

A proof of these properties is given in [8]. An example using the first and third property will appear in the next chapter, when the uniqueness in distribution property is applied to Brownian motion.

2.5 Conditional Expectation

Now that the notion of expectation has been introduced for a random variable, we should consider how to define conditional expectation. To motivate what is meant by this, suppose a random experiment has been conducted and the only quantity measured is the value of a random variable $Y(\omega)$ but the value $X(\omega)$ is of primary interest (case 1). What is our best estimate for this value? Similarly, suppose instead that the only information we have from the experiment is that the outcome belongs to an event A which is an element of the sub-σ-algebra $\mathcal{G}$ (case 2). What now can we conclude about $X(\omega)$?

Definition 2.11. (Case 1) Let (Ω, U, P) be a probability space and Y be a random variable. Then the conditional expectation of X given Y, denoted $E(X|Y)$ is any $U(Y)$-measurable random variable such that:

$$\int_A E(X|Y)dP = \int_A XdP \text{ for all } A \in U(Y).$$

Looking at the right side of this equation, it isn't the random variable Y that is important here, but rather the σ-algebra generated by Y, U(Y). This observation motivates the following definition.

Definition 2.12. (Case 2) Take a probability space as before and suppose $\mathcal{G}$ is a σ-algebra. If X is integrable then $E(X|\mathcal{G})$ is any random variable such that:

1. $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable.
2. $\int_A E(X|\mathcal{G})dP = \int_A XdP$ for all $A \in \mathcal{G}$.

Before this chapter is concluded with a reminder of two limit theorems, it is important to mention the meaning of the term 'almost surely'. We say an event happens almost surely (abbreviated a.s) if it has probability 1 of taking place.

Example 2.6. (Infinite dice roll) Consider rolling a six-sided dice repeatedly and observing the scores. If this rolling continued indefinitely (number of rolls n $\to \infty$) then it's clear that eventually a five will be rolled. Then it can be said that the event 'the sequence of dice rolls contains at least one five' happens almost surely. On the other hand, if the rolling stopped after 100000 scores were recorded, it is not certain that at least one five will be rolled. The probability that no fives are seen is $\frac{5}{6}^{100000}$ which is not zero. In these circumstances, the event no longer happens almost surely, even though chance and experience suggests a five will be recorded.

Proposition 2.5. The conditional expectation $E(X|\mathcal{G})$ exists and is unique.

We prove uniqueness below. For a proof of existence see [14].

Proof. Let Y and Y' be two random variables satisfying the two conditions listed in definition 2.11. Then:

$$\int_A YdP = \int_A XdP = \int_A Y'dP$$

for all $A \in U$. Since the above is valid for all A, take $A = \{Y - Y' \geq \epsilon > 0\}$ where ϵ is small. Using the second condition from the definition:

$$0 = \int_A (X - X)dP = \int_A (Y - Y')dP \geq \epsilon \int_A dP = \epsilon P(A)$$

where in deriving the last inequality we used the definition of A. This shows that $P(A) = 0$ and since this holds for arbitrary ϵ, $Y' \geq Y$ a.s. The same argument can be repeated with

the roles of Y and Y' reversed, giving $Y \geq Y'$. Combining the two inequalities gives Y=Y' a.s. □

Another useful fact regarding conditional expectations is the following [67]:

Proposition 2.6. Let $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$ be integrable random variables defined on (Ω, U, P). Let $\mathcal{G}$ be a sub-σ-algebra of U, $X_1, X_2, ..., X_n$ be $\mathcal{G}$-measurable and $Y_1, Y_2, ..., Y_m$ be independent of $\mathcal{G}$. For a 'good' function (bounded and Borel-measurable) $f(x_1, ..., x_n, y_1, ..., y_m)$, define

$$g(x_1, ..., x_n) = E(f(x_1, ..., x_n, Y_1, ..., Y_m).$$

Then

$$E(f(X_1, ..., X_n, Y_1, ..., Y_m)|\mathcal{G}) = g(X_1, ..., X_n).$$

A proof of this proposition can be found in [60].

Some further properties of conditional expectations are given below (see [6] for proofs):

1. If X is $\mathcal{G}$-measurable then $E(X|\mathcal{G}) = X$ a.s.
2. $E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})$ a.s.
3. If X is $\mathcal{G}$-measurable and XY is integrable, then $E(XY|\mathcal{G}) = XE(Y|\mathcal{G})$ a.s.
4. If X is independent of Y, then $E(X|Y) = E(X)$ a.s.
5. If $X \leq Y$, then $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ a.s.

Definition 2.13. (Conditional density) If random variables ϵ and η have the joint density $\rho_{\epsilon,\eta}(x, y)$ (see definition 2.6), then the conditional density of ϵ given a value of $\eta = y$ is:

$$\rho_\epsilon(x|\eta = y) = \frac{\rho_{\epsilon,\eta}(x, y)}{\rho_\eta(y)}$$

Example 2.7. Suppose we have the following joint density function:

$$\rho_{\epsilon,\eta}(x,y) = \begin{cases} x + \frac{3}{2}y^2 & 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We can find the conditional density of ϵ given $\eta = y$ by first finding the marginal distribution of η then using the definition above. Integrating over x gives:

$$\rho_\eta(y) = \frac{1}{2} + \frac{3}{2}y^2$$

So the conditional density for these random variables is:

$$\rho_\epsilon(x|\eta = y) = \frac{\rho_{\epsilon,\eta}(x,y)}{\rho_\eta(y)} = \frac{x + \frac{3}{2}y^2}{\frac{1}{2} + \frac{3}{2}y^2}$$

In the final part of this chapter, we discuss two important theorems which allow us to later analyse the error associated with the use of estimators and Monte Carlo.

2.6 Limit theorems

The following are two results that describe asymptotic distributions of sums of random variables.

Theorem 2.7. (Strong law of large numbers (LLN)) Let ν_1,ν_2,... be a sequence of n independent random variables sharing the same distribution, which has mean μ. Introduce the arithmetic mean $S_n = \frac{1}{n}\sum_{i=1}^{n} \nu_i$. Then as n$\rightarrow \infty$, $S_n \rightarrow \mu$ with probability 1 i.e. the arithmetic mean converges almost surely.

The important observation to make here is that no assumption was made about which distribution the random variables come from.

Example 2.8. (Average time between calls) Suppose a call centre is interested in the average amount of time between calls received from 2pm to 3pm on Wednesdays, and he observes a large sample of these times. The assumption of independence between each

observation seems sensible here if the callers don't interact with each other, and there is no obvious reason for large changes in the number of callers between each Wednesday. Then the arithmetic mean of the times converges to the mean of the underlying distribution as the number of observations made approach infinity by the LLN.

Theorem 2.8. (Central Limit Theorem (CLT)) Let ν_1,ν_2,... be a sequence of n independent random variables sharing the same distribution, which has finite mean μ and variance σ^2. Denote their sum by $\theta_n = \nu_1 + ... + \nu_n$. Then $\widehat{\theta_n} = \frac{1}{\sqrt{n\sigma^2}} * (\theta_n - n\mu) \sim \mathcal{N}(0,1)$ as n $\to \infty$. This is a convergence in distribution, so $\widehat{\theta_n}$ has a standard normal distribution.

Example 2.9. (Continuing example 2.9...) Building on the previous example, the result in theorem 2.8 can be rearranged (using standard manipulation of normal variables) to obtain the approximate distribution for S_n, which is normal with mean μ and variance $\frac{\sigma^2}{n}$. Using this distribution, a confidence interval for the average time between calls could be constructed.

Proofs of these theorems can be found in [60]. These results conclude this chapter, which provided a review of probability theory. In the following chapter we construct Brownian motion, the main object of stochastic analysis.

3 Brownian Motion

The probability results from the previous chapter can now be used to develop the theory propagating from the idea that randomness can be added to a system. This chapter expands on this by motivating and constructing Brownian motion, an essential object within stochastic analysis.

3.1 Motivation and definition

Perhaps the most important and widely applicable field in mathematics is that of differential equations. How does a variable evolve according to other variables? This question is seemingly omnipresent, dictating the order of the universe itself - from the celestial to the sub-atomic.

Ordinary (ODEs) and partial differential equations (PDEs) represent a great portion of academic research and undergraduate study. Here, a change in the variable of interest is dependent on one or multiple arguments that can be measured accurately. However, if one has ever conducted say a chemical experiment, it is often observed that the results are not reproducible even if all possible efforts are made to keep the conditions the same. What is the cause of the discrepancy?

It is very common for the experimentally measured trajectories of a system to not behave exactly as they are predicted to. Consequently, we should think about how we might come up with a model that generates trajectories matching more closely to those we observe. It seems sensible to keep the variables included in the differential equation, since they are likely based on some prior knowledge of the experiment. Intuitively, we might modify this equation by adding a term that generates some kind of randomness. Doing this, and refining the idea by making this randomness possess particular properties leads to stochastic differential equations (SDEs).

Written formally (as in [19]), start with a well-posed ordinary differential equation with

fixed initial point $x_0 \in \mathbb{R}^n$:

$$\dot{\mathbf{x}}(t) = \mathbf{b}(\mathbf{x}(t)) \text{ for t} > 0,$$
$$\mathbf{x}(0) = x_0.$$

where here **b** is a vector field mapping $\mathbb{R}^n$ to itself. Disturbing the system involves adding a term to the right-hand side of the ODE:

$$\dot{\mathbf{X}}(t) = \mathbf{b}(\mathbf{X}(t)) + \mathbf{B}(\mathbf{X}(t))\xi(t),$$
$$\mathbf{X}(0) = x_0.$$

where **B** maps $\mathbb{R}^n$ to a matrix space with elements of dimension nxm. We'll call the added randomness $\xi(\cdot)$ white noise.

Where did this concept originate? In 1827, Scottish botanist Robert Brown observed that pollen particles floating on top of water had a jittery and unpredictable motion. He compared this observation with that of dust, which gave less exciting results, before coming to the conclusion that the reason for the behaviour is that pollen is "alive" whereas dust is not [18]. Other scientists verified Brown's experiment whenever small particles are suspended in fluid medium. It took 73 years for the first theory of Brownian motion to come about, with Louis Bachelier's PhD thesis "The theory of speculation" [2]. Einstein finally explained the motion using a probabilistic model 5 years later, by considering ink within a tube of water. He found that if the kinetic energy was just right, there was random movement created by collisions between molecules[15], the same movement Brown had described.

Einstein's experiment was later verified in 1926 by the french physicist Jean Perrin [51], who believed that the sample path formed by the movement was non-differentiable. It had recently been shown by Karl Weierstrass that continuous functions exist with this property.

The development of the theory of Brownian motion led a revolution in many other applications, perhaps most notably the subject of mathematical finance. Problems in finance such as stock pricing are difficult due to the multiplicity of factors involved. Having a good

approximation for future stock prices is key to make decisions that minimise losses. We will focus on the financial application of option pricing later in this report.

In the formal representation written on the previous page, the connection between white noise and Brownian motion is made explicit (in 1D) by considering $X(t)$ as a function of t and Brownian motion $W(t)$, $X(t) = f(t, W(t))$, and using the chain rule:

$$\begin{aligned} X(t) &= f(t, W(t)) \\ \text{(Differentiate)} \Rightarrow \frac{dX(t)}{dt} &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial W}\frac{dW(t)}{dt} \\ \text{(Compare)} \Rightarrow \dot{X}(t) &= b(X(t)) + B(X(t))\xi(t) \end{aligned}$$

The first equation from the previous page has been included at the bottom here. Comparing, it can easily be seen that the two equations are equal if $b(X(t)) = \frac{\partial f}{\partial t}$, $B(X(t)) = \frac{\partial f}{\partial W}$ and $\xi(t) = \frac{dW(t)}{dt}$. The last equality provides us with the desired connection - white noise is the time derivative of Brownian motion. Despite this seeming to make the connection clear, the sample path $t \to W(t, \omega)$ is not actually differentiable for positive t, therefore the white noise object $\xi(\cdot)$ does not really exist.

Definition 3.1. (Random function/process) A random function $\zeta_t(\omega)$ is a family of random variables depending on the parameter t, a member of a parameter set $\mathbb{T}$. In the case that $\mathbb{T}$ is a subset of the real line $\mathbb{R}$ and when t can be interpreted as time, we refer to $\zeta_t(\omega)$ as a random or stochastic process.

Fixing the event ω of a random process gives us a trajectory or realisation. If instead t is fixed, $\zeta_t(\omega)$ is a random variable.

Example 3.1. (Random walk) Suppose a fair two-sided coin is flipped a number of times. Assign the outcome of heads on a single flip to value 1, and tails to value -1. For each flip, the outcome can then be modelled by a Bernoulli random variable $\varsigma_i, i = 1, 2, ...$, taking values in $\{1, -1\}$ with equal probability $\frac{1}{2}$. The sum of these variables up to n flips of the coin is a process, $\eta_n = \sum_{i=1}^{n} \varsigma_i$.

Definition 3.2. (Brownian Motion) A real-valued stochastic process W(·) is called Brownian Motion (BM) or a Wiener Process if the following properties hold:

1. $W(0) = 0$ a.s. (Starts from zero)

2. $W(t) - W(s)$ has a $\mathcal{N}$(0,t-s) distribution for all $t \geq s \geq 0$. (Normality)

3. For all ordered partitions of time $0 < t_1 < ... < t_n$, the random variables $W(t_1)), W(t_2) - W(t_1), ..., W(t_n) - W(t_{n-1})$ are independent of each other. (Independent increments)

Processes with stationary, independent increments are known as L*é*vy processes. The second property implies that $W(t)$ has a $\mathcal{N}(0, t)$ distribution by setting $s = 0$ and using the first property.

The main question to consider at this point is how this process might be constructed, which we tackle in the next part.

3.2 Construction

We begin constructing one-dimensional Brownian motion by first starting with the simpler problem of constructing Brownian motion on the domain $0 \leq t \leq 1$. Once we have accomplished this, we can extend the construction for all times $t \geq 0$. Note that this section is mainly based off the derivation in [19].

Lemma 3.1. Suppose W(·) is one-dimensional Brownian motion. Then:

$$E(W(t)W(s)) = min\{s, t\} \text{ for t} \geq 0, \text{s} \geq 0.$$

Proof. Assume that $t \geq s \geq 0$. Then

$$\begin{aligned} E(W(t)W(s)) &= E((W(t) - W(s) + W(s))W(s)) \\ &= E((W(t) - W(s))W(s)) + E(W(s)^2) \\ &= E(W(t) - W(s))E(W(s)) + E(W(s)^2) \\ &= 0 + s = s = min\{s, t\} \end{aligned}$$

First, we used the fact that $t \geq s$ to rewrite $W(t)$ as the sum of the increment $W(t) - W(s)$ and $W(s)$. We then expanded the expectation. Following this, we used the independence of

increments property of Brownian motion to split the first expectation into a product. Finally, we used the fact that $W(s)$ is $\mathcal{N}(0, s)$ to identify that the first term is zero and the second term is equal to s, the minimum of t and s. We can repeat the same steps with the alternative condition $s \geq t \geq 0$. □

This lemma can be used to prove the following heuristic formula (remember that it was stated in the previous part that white noise does not really exist) for white noise $\xi(\cdot)$. Recall that the Dirac-delta function is the unit mass centred at 0. For more details on this object and the properties it possesses see [13].

Lemma 3.2. Suppose $\xi(\cdot)$ is white noise. Then

$$E(\xi(t)\xi(s)) = \delta_0(s-t) \text{ for } t \geq 0, s \geq 0$$

where δ_0 is the Dirac-delta function.

Proof. In the previous part it was noted that $\dot{W}(\cdot) = \xi(\cdot)$. Recall that a time derivative can be expressed as a limit in the following way:

$$\dot{W}(t) = \lim_{h \to 0} \left(\frac{W(t+h) - W(t)}{h} \right)$$

With this representation in mind, we prove the result as follows:

$$\begin{aligned}
\Psi_h(s) &= E\left(\frac{W(t+h) - W(t)}{h} \frac{W(s+h) - W(s)}{h} \right) \\
&= \frac{1}{h^2} E((W(t+h) - W(t))(W(s+h) - W(s))) \\
&= \frac{1}{h^2} [E(W(t+h)W(s+h)) - E(W(t+h)W(s)) - E(W(t)W(s+h)) \\
&+ E(W(t)W(s))] \\
&= \frac{1}{h^2} [min(t+h, s+h) - min(t+h, s) - min(t, s+h) + min(t, s)]
\end{aligned}$$

What does this function look like? Suppose $t-h < s < t$, then $\Psi_h(s) = \frac{1}{h} + \frac{s-t}{h^2}$, which is a line starting from 0 when $s = t-h$ going to $\frac{1}{h}$ when $s = t$. Similarly, if $t < s < t+h$, $\Psi_h(s) = \frac{1}{h} + \frac{t-s}{h^2}$, another line starting from $(t, \frac{1}{h})$ and ending at 0. For s outside of these intervals, $\Psi_h = 0$.

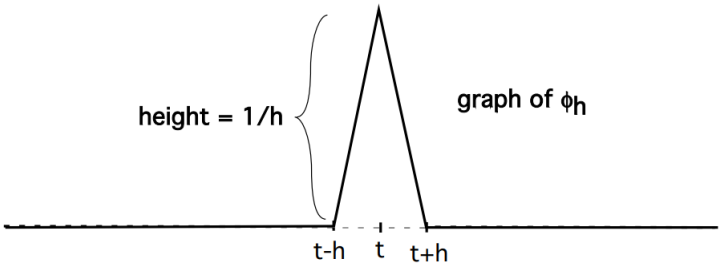

Figure 2: Graph of white noise autocorrelation, source [19]

The graph of this function (see figure below) is hence an isosceles triangle with height $\frac{1}{h}$ over the domain $t - h < s < t + h$ and zero elsewhere.

As this triangle is squeezed under the limit $h \to 0$, the area underneath it is fixed at value 1. It is clear that $\Psi_h \to \delta_0(s - t)$. Going back to the definition of Ψ_h, it is clear that it also approaches the quantity $E(\xi(t)\xi(s))$ using the limit-based definition of the derivative. □

The quantities calculated above can be generalised for other processes. When we have a real-valued stochastic process X(·) satisfying $E(X^2(t)) < \infty$, $r(t, s) = E(X(t)X(s))$ is called the autocorrelation function. When $r(t, s)$ is a function depending only on distance $t - s$ with the expectation $E(X(t))$ independent of t, the process $X(\cdot)$ is called stationary in the wide-sense.

The method we now use to construct Brownian motion on $0 \leq t \leq 1$ is laid out as follows (this procedure is similar to "wavelet analysis" [52]):

1. Obtain a formal expansion of white noise $\xi(\cdot)$ based on a chosen orthonormal basis of $L^2(0, 1)$, the space of square-integrable functions on (0,1). Recall that square-integrable simply means that the integral of the square of the function over the domain (0,1) is equal to 1.

2. Using the connection between white noise and Brownian motion, integrate the expansion in time.

3. Show that this series converges.

4. Finally, check the properties of Brownian motion hold.

Firstly, to obtain a formal expansion, suppose we have a complete, orthonormal basis for $L^2(0,1)$, $\{\psi_n\}_{n=0}^{\infty}$. By complete, we mean that any function in $L^2(0,1)$ can be written as an infinite sum of the basis functions. Orthonormality implies that:

$$\int_0^1 \psi_n(s)\psi_m(s)ds = \delta_{mn}$$

which is an inner product $< \psi_n , \psi_m >$ on $L^2(0,1)$, and δ_{mn} is the Kronecker-delta (equal to 1 when $m = n$ otherwise equal to 0). Since the basis is complete, $\xi(t)$ can be written as a sum of these basis functions:

$$\xi(t) = \sum_{n=0}^{\infty} A_n\psi_n(t),$$

where the coefficients A_n can be thought of as a 'weight' representing the amount that ψ_n contributes to $\xi(t)$. This makes it clear that:

$$A_n =< \psi_n, \xi >= \int_0^1 \psi_n(s)\xi(s)ds$$

What can we deduce about the coefficients A_n? The following calculation is revealing:

$$\begin{aligned} E(A_nA_m) &= E\left(\int_0^1 \int_0^1 \psi_n(s)\psi_m(t)\xi(s)\xi(t)dsdt \right) \\ &= \int_0^1 \int_0^1 \psi_n(s)\psi_m(t)E(\xi(s)\xi(t))dsdt \\ &= \int_0^1 \int_0^1 \delta_0(s-t)\psi_n(s)\psi_m dsdt \\ &= \int_0^1 \psi_n(s)\psi_m(s)ds = \delta_{nm} \end{aligned}$$

The definition of A_n is used before the expectation is taken under the integral (noting that the basis functions are deterministic). Then, the autocorrelation function of white noise is substituted and using the properties of the Dirac-delta function we arrive at the final line. The Kronecker-delta arises by realising the final line is the orthonormality condition.

Since $\xi(\cdot)$ is Gaussian, by the affine property of multivariate normal distributions [40] (sums of Gaussian variables are also Gaussian) we should expect the coefficients A_n to be Gaussian as well. In order to get an intuition for why this is, we can try expressing the

integral as a Riemann sum. It can be proven that infinite sums of independent Gaussian random variables converge to another Gaussian under certain conditions, using characteristic functions and L*é*vy's convergence theorem [23].

This implies that the coefficients are independent since when $n \neq m$ the covariance is zero, which is true only since the coefficients are Gaussian. To see this, note that if the covariance between two Gaussian variables is zero, then the covariance matrix of the Gaussian joint distribution between A_n and A_m is diagonal, and so there are no cross terms between the two variables. Consequently, this joint distribution can be split into a product of two terms, each of which depends exclusively on either A_n or A_m. Then by definition 2.6 in the previous chapter, A_n and A_m are independent. A more rigorous proof for a bivariate distribution using characteristic functions is given in [5]. Since we have independence, we can split the covariance into a product $E(A_nA_m) = E(A_n)E(A_m) = 0$ when $n \neq m$, so $E(A_n) = 0$ for all n. $E(A_n^2) = 1$ is also clear, setting $n = m$ in the above. All of this informs us that the coefficients A_n are independent standard normal variables.

Brownian motion should therefore be given by:

$$W(t) = \int_0^t \xi(s)ds = \sum_{n=0}^{\infty} A_n \int_0^t \psi_n(s)ds$$

Note that we have not actually selected an orthonormal basis yet, so this holds true for any basis on $L^2(0,1)$. We will now select one basis allowing us to easily calculate the integral in this formula.

Definition 3.3. (Haar functions) The family of functions $\{h_k(\cdot)\}_{k=0}^{\infty}$ are called Haar functions [16, 26] if they are defined on $0 \leq t \leq 1$ with:

$$h_0(t) = 1 \text{ for } 0 \leq t \leq 1,$$

$$h_1(t) = \begin{cases} 1, & 0 \leq t \leq \frac{1}{2}, \\ -1 & \frac{1}{2} < t \leq 1. \end{cases}$$

If $2^n \leq k < 2^{n+1}$, $n = 1, 2, ...$, set

$$h_k(t) = \begin{cases} 2^{n/2}, & \frac{k-2^n}{2^n} \leq t < \frac{k-2^n+\frac{1}{2}}{2^n}, \\ -2^{n/2}, & \frac{k-2^n+\frac{1}{2}}{2^n} < t \leq \frac{k-2^n+1}{2^n}, \\ 0 & \text{otherwise.} \end{cases}$$

So what do these functions look like as n increases? As n increases, the interval over which $h_k(t)$ is non-zero decreases (width of non-zero region is 2^{-n}) but the height/depth of the non-zero regions increases. There is symmetry around the point $t = \frac{k-2^n+\frac{1}{2}}{2^n}$. Summarising, the graph of h_k for $k \geq 0$ consists of three constant phases, and is symmetric.

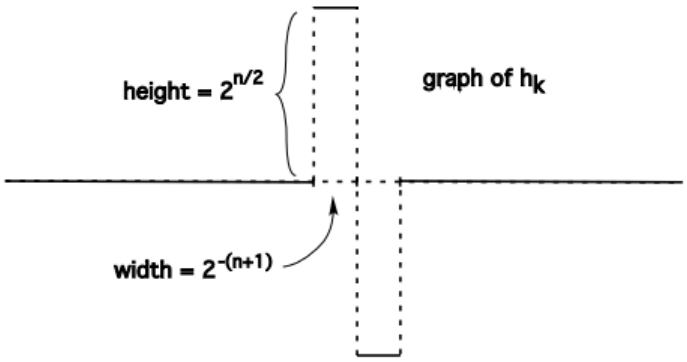

Figure 3: Graph of the Haar Function (Source:[19])

We should check that these Haar functions form a basis in order for us to apply them to the formulae derived earlier.

Lemma 3.3. The family of Haar functions forms a complete, orthonormal basis of the function space $L^2(0, 1)$.

Proof. We begin with the proof of orthonormality.

$$\int_0^1 h_k^2(t)dt = \int_{\frac{k-2^n}{2^n}}^{\frac{k-2^n+1}{2^n}} 2^n dt = 1$$

The last equality is due to the observation that the interval where $h_k(t)$ is non-zero has width

2^{-n}, and so the functions are normalised for all k. Now, for orthogonality, take $k < l$. There are two possibilities, the first being that the supports of the two Haar functions are disjoint, in which case we get zero (as in the case of $k = 2$, $l = 9$). The second is that $h_k(t)$ is constant on the support of $h_l(t)$. In this case, the constant value can be taken outside the integral, and we can note that the integral of any Haar function over $(0, 1)$ is zero. This is shown via the equation below.

$$\int_0^1 h_k(t)h_l(t)dt \propto \int_0^1 h_l(t)dt = 0$$

where the proportionality is due to $h_k(t)$ being constant (either $2^{\frac{n}{2}}$ or $2^{\frac{-n}{2}}$, where n is such that $2^n < k < 2^{n+1}$).

Secondly, we need to prove completeness. Take an arbitrary element f from the function space $L^2(0, 1)$ and assume that $\int_0^1 f(t)h_k(t)dt = 0$ for all $k = 0, 1, \ldots$. If we can prove that this condition implies that f is zero almost everywhere then any f can be represented as a sum of Haar functions, and completeness follows.

Starting from $k = 0$, we have $\int_0^1 f(t)dt = 0$. Moving on, $k = 1$ gives $\int_0^{\frac{1}{2}} f(t)dt - \int_{\frac{1}{2}}^1 f(t)dt = 0$ which, combining with the equation observed for $k = 0$, gives us that both these terms are zero. Continuing in this way gives $\int_{\frac{k}{2^{n+1}}}^{\frac{k+1}{2^{n+1}}} f(t)dt = 0$ for all $0 \leq k \leq 2^{n+1}$. This means that the integral of f across any sub-interval of $(0, 1)$ is zero. Using the fundamental theorem of calculus, we have $f(r) = \frac{d}{dr}\int_0^r f(t)dt = 0$ for all $r \leq 1$. For a more in-depth proof of completeness, see [27]. □

Before definition 3.3, it was noted that $W(\cdot)$ should be given by a series with standard Gaussian coefficients combined with integrals of basis functions. Now that we have a set of basis functions, we need to consider their associated integrals in order to obtain this series.

Definition 3.4. (Schauder functions) For $k = 1, 2, \ldots$, $s_k(t) = \int_0^t h_k(t)dt$ is the k-th Schauder function [58].

Using figure 3, it is clear that the integral grows from $t = \frac{k-2^n}{2^n}$ to $t = \frac{k-2^n+\frac{1}{2}}{2^n}$. This is where the Haar function takes positive value $2^{n/2}$. The width of this interval is $2^{-(n+1)}$ so the maximum value occurs at height $2^{-\frac{n+2}{2}}$. After this point the Haar function takes on a

negative value of the same magnitude for an interval of equal length, so the value of the integral decreases at the same rate until it reaches value 0 again at $t = \frac{k-2^n+1}{2^n}$. This analysis shows that the integral as a function of t is an equilateral triangle in the non-zero region with height given above, and is equal to zero otherwise.

So, the goal from here is to define $W(t) = \sum_{k=0}^{\infty} A_k s_k(t)$ where the coefficients A_k are independent, identically distributed standard normal random variables defined on some probability space. If this series converges and satisfies the properties of Brownian motion, then this goal has been achieved. But what does it mean for the series to converge?

Definition 3.5. (Uniform convergence of functions) We say a sequence of functions converges uniformly to a limiting function f on set E if given any arbitrarily small ϵ, a number N can be chosen such that the series truncated after N terms differs from f by no more than ϵ at every point t in E.

Considering $W(\cdot)$ as f and E as the set $(0, 1)$ here, this definition can be applied so that we know when the series converges.

Lemma 3.4. Let $\{a_k\}_{k=0}^{\infty}$ be a sequence of real numbers such that

$$|a_k| = O(k^{\delta}) \text{ as } k \to \infty$$

for some $0 \leq \delta < \frac{1}{2}$. Then the series

$$\sum_{k=0}^{\infty} a_k s_k(t)$$

converges uniformly for $0 \leq t \leq 1$.

Proof. Fix $\epsilon > 0$. We note from the above discussion of $h_k(\cdot)$ and $s_k(\cdot)$ that for $2^n \leq k < 2^{n+1}$, the Schauder functions have disjoint supports (e.g. $k = 8, 9, ..., 15$). Set:

$$b_n = \max_{2^n \leq k < 2^{n+1}} |a_k| \leq C(2^{n+1})^{\delta}$$

The inequality is a consequence of the assumption on the absolute value of a_k and the fact that in the interval considered for k in the definition of b_n, k is less than 2^{n+1}. Then for

$0 \leq t \leq 1$:

$$\begin{aligned}\sum_{k=2^m}^{\infty} |a_k||s_k(t)| &\leq \sum_{n=m}^{\infty} b_n \max_{2^n \leq k < 2^{n+1}} |s_k(t)| \\ &\leq C \sum_{n=m}^{\infty} (2^{n+1})^{\delta} 2^{-\frac{n}{2}-1} \\ &= C \sum_{n=m}^{\infty} 2^{n(\delta-\frac{1}{2})+\delta-1} < \epsilon \text{ for large m}\end{aligned}$$

In the first line we used the fact that b_n is the maximum of the absolute values of a_k and imposed a maximum on the absolute value of the Schauder function. We discussed on the previous page that the graph of this function obtains a maximum value at $2^{-\frac{n}{2}-1}$, so substituting this in gives the second line. Rearranging this gives the final equality. Considering the fact that $0 \leq \delta < \frac{1}{2}$, the exponent of 2 is negative and becomes more negative with increasing n. Consequently, a large enough m can be chosen so that the series is a sum of extremely small values which converge to a value less than ϵ. □

Comparing the series in the lemma with the series for Brownian motion, it is clear that if the normal coefficients A_k satisfy the condition in the lemma then we have uniform convergence over the domain. Indeed, it can be shown that the condition is satisfied.

Lemma 3.5. Suppose that $\{A_k\}_{k=1}^{\infty}$ are independent and share the $\mathcal{N}(0,1)$ distribution. Then for almost every event ω,

$$|A_k(\omega)| = O(\sqrt{\log k}) \text{ as } k \to \infty$$

See [19] for the proof of this, which uses the standard normal distribution and the Borel-Cantelli lemma (details of which can be found in [16]). This lemma tells us that the condition in lemma 3.4 holds. This follows as $k \to \infty$ since $\log k$ grows slower than k^{δ}, where δ is any positive number. The final lemma we need to prove to show that the series is Brownian motion is given below:

Lemma 3.6. $\sum_{k=0}^{\infty} s_k(s)s_k(t) = min(t,s)$ where $0 \leq s,t \leq 1$

Proof. Define $\phi_s(\tau) = \begin{cases} 1 \ 0 \leq \tau \leq s \\ 0 \ s \leq \tau \leq 1 \end{cases}$. If $s \leq t$, we can use the completeness of the Haar basis

to obtain:

$$s = \int_0^1 \phi_t(\tau)\phi_s(\tau)d\tau = \sum_{k=0}^{\infty} a_k b_k = \sum_{k=0}^{\infty} s_k(s)s_k(t)$$

as $a_k = \int_0^1 \phi_t(\tau)h_k(\tau)d\tau = \int_0^t h_k(\tau)d\tau = s_k(t)$ and $b_k = s_k(s)$ in the same way. All that has been done here is a substitution of the definition of ϕ given above which restricts the domain of integration. □

The lemmas of this section can be used to prove that the series derived is indeed a Brownian motion.

Theorem 3.7. Let $\{A_k\}_{k=0}^{\infty}$ be sequence of independent, identically distributed standard normal variables on the same probability space. Then the sum

$$W(t,\omega) = \sum_{k=0}^{\infty} A_k(\omega)s_k(t)$$

on the interval $0 \leq t \leq 1$ converges uniformly in t for almost every ω. On top of this, $W(\cdot)$ is a Brownian motion for $0 \leq t \leq 1$.

Proof. Uniform convergence is achieved by applying lemmas 3.4 and 3.5, which also implies that the sample paths of W(t,ω) are continuous for almost every ω. Moving on, we prove that this is a Brownian motion.

1. $W(0) = 0$ follows from definition 3.4 and the fact that integrating over a point returns zero.

2. We need to show that $W(t) - W(s)$ has a $\mathcal{N}(0, t-s)$ distribution, where $0 \leq s \leq t \leq 1$.

$$\begin{aligned}
E(e^{is(W(t)-W(s)}) &= E(e^{is\sum_{k=0}^{\infty} A_k(s_k(t)-s_k(s))}) \\
&= \prod_{k=0}^{\infty} E(e^{isA_k(s_k(t)-s_k(s))}) \\
&= \prod_{k=0}^{\infty} e^{-\frac{s^2}{2}(s_k(t)-s_k(s))^2} \\
&= e^{-\frac{s^2}{2}\sum_{k=0}^{\infty}(s_k(t)-s_k(s))^2} \\
&= e^{-\frac{s^2}{2}(t-s)}.
\end{aligned}$$

In step 1, we substituted the series before using the independence of the A_k to form a product in line 2. Then, the characteristic function of A_k is substituted as derived in chapter 2. The product can then be rewritten as a sum using the properties of the exponential function. The final step involved expanding the quadratic argument of the sum and using lemma 3.6. This is clearly the characteristic function of a normal variable with variance t-s and mean 0 as required, so by the uniqueness of characteristic functions we have proved the second property.

3. The last property to prove is that disjoint increments are independent. We claim that for $m = 1, 2, ...$, and for all $0 = t_0 < t_1 < ... < t_m \leq 1$:

$$E(e^{i\sum_{j=1}^{m} s_j(W(t_j)-W(t_{j-1}))}) = \prod_{j=1}^{m} e^{-\frac{s_j^2}{2}(t_j-t_{j-1})}$$

If this claim is true, then we note that each term in the product is the characteristic function of a normal random variable. Then by uniqueness of characteristic functions the joint distribution function has the following form:

$$F_{W(t_1),...,W(t_m)-W(t_{m-1})}(x_1, ...x_m) = F_{W(t_1)}(x_1)...F_{W(t_m)-W(t_{m-1})}(x_m)$$

This proves that the increments are independent using definition 2.6 from the last chapter. We'll prove the claim for $m = 2$, for higher values the same process can be

used.

$$
\begin{aligned}
E(e^{i(s_1 W(t_1+s_2(W(t_2)-W(t_1))}) &= E(e^{i((s_1-s_2)W(t_1)+s_2 W(t_2))}] \\
&= E(e^{i(s_1-s_2)\sum_{k=0}^{\infty} A_k s_k(t_1)+is_2\sum_{k=0}^{\infty} A_k s_k(t_2)}) \\
&= \prod_{k=0}^{\infty} e^{iA_k((s_1-s_2)s_k(t_1)+s_2 s_k(t_2))} \\
&= \prod_{k=0}^{\infty} e^{-\frac{1}{2}((s_1-s_2)s_k(t_1)+s_2 s_k(t_2))^2} \\
&= e^{-\frac{1}{2}\sum_{k=0}^{\infty}(s_1-s_2)^2 s_k^2(t_1)+2(s_1-s_2)s_2 s_k(t_1)s_k(t_2)+s_2^2 s_k^2(t_2)} \\
&= e^{-\frac{1}{2}((s_1-s_2)^2 t_1+2(s_1-s_2)s_2 t_1+s_2^2 t_2)} \\
&= e^{-\frac{1}{2}(s_1^2 t_1+s_2^2(t_2-t_1))},
\end{aligned}
$$

which is the same form as the claim made above. In the second line the series expansion is substituted for the Brownian motions. The product in line three comes about from the independence of the A_k. From the third line, we note that the terms in the bracket are not random. Then, using the characteristic function of A_k, a standard normal, the deterministic terms can be inserted as the argument. To get the penultimate line, lemma 3.6 is used. Lastly, we expand the quadratic within the exponential and rearrange.

□

In this section so far, we have constructed Brownian motion in the domain $0 \leq t \leq 1$ as an expansion of normal coefficients (random) and Schauder functions. (deterministic). How can this be extended to the domain $t \geq 0$?

Theorem 3.8. Let (Ω,U,P) be a probability space on which countably many standard normal ($\mathcal{N}(0,1)$) independent random variables $\{A_n\}_{n=0}^{\infty}$ are defined. Then a one-dimensional Brownian motion exists which is defined for $\omega \in \Omega$ and $t \geq 0$.

The idea behind the proof of this theorem is as follows. Using the normal random variables countably many independent Brownian motions $W(\cdot)$ on the domain $0 \leq t \leq 1$ can

be constructed using the earlier results. Indexing these Brownian motions by $W^n(\cdot)$, a new object $\hat{W}(\cdot)$ can be assembled using the inductive formula:

$$\hat{W}(t) = \hat{W}(n-1) + W^n(t-(n-1)) \text{ for } n-1 \leq t \leq n$$

The argument of $W^n(\cdot)$ here stays within $[0,1]$ as required. For a more in-depth discussion of the existence of Brownian motion on the extended domain, see [14].

3.3 Further properties

In this section, we present and find the quadratic variation [11] of Brownian motion and compare it to the quadratic variation of functions with bounded derivatives. This is worthwhile firstly to become more familiar with how to apply the different properties given in definition 3.2 and perform calculations involving Brownian motions. Secondly, we can identify an unusual property of the motion not shared with smooth functions. At the end of this part, we also consider differentiability and higher dimensional motions.

Before we go on to define the quadratic variation, we should first define what it means to take a partition of an interval.

Definition 3.6. (Partition) For an interval $[a,b]$ with $b \geq a$, a partition P is a finite collection of points in $[a,b]$:

$$P = \{a = t_0 < t_1 < ... < t_m = b\}.$$

The mesh size of P is given by $|P| = \max_{0 \leq k \leq m-1} |t_{k+1} - t_k|$.

This will be particularly important later when we consider stochastic integration, which involves splitting up the domain into many parts and taking a limit i.e. letting the mesh size go to zero.

Definition 3.7. (Quadratic variation) The quadratic variation of a function f on the interval $[a,b]$ over some partition $\Pi = \{a = t_0, t_1, ..., t_n = b\}$ is defined by:

$$Q_\Pi(f) = \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|^2$$

The number of points n in the partition can be increased arbitrarily.

$$Q(f) = \lim_{n\to\infty} Q_\Pi(f)$$

Suppose we take a smooth function f, what will the quadratic variation be?

Proposition 3.9. A differentiable function whose derivative is bounded on $[a, b]$ has zero quadratic variation.

A proof of this result and others relating to the topic of variation, which often use the Mean Value Theorem, can be found in [11]. Note then that functions such as sine and cosine have bounded derivatives on any interval and consequently have zero quadratic variation. What can we find out about the quadratic variation of Brownian motion?

Example 3.2. Partition the interval $[t_0, T]$ into N parts $t_0, t_1, ..., t_N = T$ and let $h = \max_k(t_{k+1} - t_k)$ where $k = 0, 1, ..., N-1$. Then the quadratic variation of Brownian motion is given by:

$$\theta_h = \sum_{k=0}^{N-1} (W(t_{k+1}) - W(t_k))^2$$

How does this quantity behave as $h \to 0$? To answer this, consider the following limit:

$$\lim_{h\to 0} E(\theta_h - (T - t_0))^2$$

Limits of this form are particularly useful as they tell us how a random object converges in the mean-square sense, which we revisit later. This particular limit approaches zero as

$h \to 0$, which we now show. First, we calculate the expectation of θ_h:

$$\begin{aligned} E(\theta_h) &= \sum_{k=0}^{N-1} E((W(t_{k+1}) - W(t_k))^2) \\ &= \sum_{k=0}^{N-1} (E(W(t_{k+1})^2) - 2E(W(t_k W(t_{k+1})) + E(W(t_k)^2)) \\ &= \sum_{k=0}^{N-1} (t_{k+1} - 2t_k + t_k) \\ &= \sum_{k=0}^{N-1} (t_{k+1} - t_k) = T - t_0. \end{aligned}$$

Here, the definition of θ_h is substituted and the quadratic expanded. Then lemma 3.1 and the variance of Brownian motion is used to evaluate the expectations. Finally, the sum is tidied up and we note that many terms cancel out, leaving the quantity $T - t_0$. Comparing this to the limit above, we see that the limit is of the variance of the quadratic variation. Recall that for a random variable X, $Var(X) = E(X^2) - E(X)^2$. We walk through the steps in the calculation of the variance below.

$$\begin{aligned} E(\theta_h^2) &= E(\sum_{k=0}^{N-1} (W(t_{k+1}) - W(t_k))^4) \\ &= E(\sum_{k=0}^{N-1} (W(t_{k+1})^4 - 4W(t_{k+1})^3 W(t_k) + 6W(t_{k+1})^2 W(t_k)^2 \\ &\quad - 4W(t_{k+1}) W(t_k)^3 + W(t_k)^4). \end{aligned}$$

All that is done here is a substitution followed by an expansion using the binomial theorem. The second, third and fourth terms within the expansion have dependence whilst there are also higher central moments in the first and fifth terms than we've encountered before. To deal with the dependent terms here, rewrite $W(t_{k+1})$ as $W(t_{k+1}) - W(t_k) + W(t_k)$. This allows us to use the property of independent increments which enables the expectation to be split (recall that if X and Y are independent random variables, then $E(XY) = E(X)E(Y)$). Doing this creates some third-order central moment terms, which we can evaluate along with the fourth order terms using the fact that since $W(t)$ is $\mathcal{N}(0, t)$-distributed, it has a characteristic function of the form $\phi(s) = e^{-ts^2/2}$. Using property 2 of lemma 2.4, these moments can be

found easily. It turns out that $E(W(t)^3) = 0$ and $E(W(t)^4) = 3t^2$. Substituting all of this and going through the cumbersome calculations eventually leads to the simple formula given below.

$$E(\theta_h^2) = 3\sum_{k=0}^{N-1}(t_{k+1} - t_k)^2$$

Then the variance is calculated by subtracting $E(\theta_h)^2$, where it's easier to use the form of this expectation as a sum. Finally, we arrive at the following:

$$\begin{aligned}\lim_{h\to 0} E(\theta_h - (T - t_0))^2 &= \lim_{h\to 0}\sum_{k=0}^{N-1} 2(t_{k+1} - t_k)^2 \\ &\leq \sum_{k=0}^{N-1} 2h^2 \to 0 \text{ as } h \to 0.\end{aligned}$$

To get the inequality we used the fact that $h = \max_k(t_{k+1} - t_k)$. What we have found is unusual as for smooth functions with bounded derivatives, the quadratic variation approaches zero as the distance between points goes to zero. The same is not true for Brownian motion - due to its roughness the quadratic variation approaches a constant value that may be quite large if the interval of interest is wide. An even more alarming difference can be seen with first-order variations:

$$V(f) = \lim_{n\to\infty}\sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|$$

We have the following result for smooth functions (again, see [11] for proof):

Proposition 3.10. A differentiable function with a bounded derivative on the interval $[a, b]$ has bounded first-order variation.

On the other hand, for Brownian motion we have:

Proposition 3.11. The first variation of Brownian motion diverges almost surely as the distance between points in a partition on $[a, b]$ go to zero.

$$V(W) = \lim_{n\to\infty}\sum_{i=0}^{n-1} |W(t_{i+1}) - W(t_i)| \to \infty$$

Equivalently, almost all Brownian paths have unbounded variation, on any time interval [68].

Proof. Suppose towards a contradiction that the first variation is bounded on the interval. Then we have:

$$\sum_{k=0}^{N-1} |W(t_{k+1}) - W(t_k)|^2 \leq \max_{k=0,1,...} |W(t_{k+1}) - W(t_k)| \sum_{k=0}^{N-1} |W(t_{k+1}) - W(t_k)|$$
$$\leq \max_{k=0,1,...} |W(t_{k+1}) - W(t_k)| V(W)$$

To get the first inequality we pull out the maximum of the absolute differences along the interval from the sum. The sum left over we recognise as the definition of the first variation, which gives the second inequality.

Since $W(\cdot)$ is continuous almost surely on $[a, b]$, it is necessarily uniformly continuous on $[a, b]$. Therefore, as the number of points in the partition goes to infinity the maximum on the right-hand side goes to zero. If $V(W)$ is bounded, this implies that the right-hand side of the inequality goes to zero under the limit. However, the term on the left is the quadratic variation, which we know from earlier converges under the limit to a non-zero constant. We therefore have a contradiction and can conclude that the first variation is unbounded on the interval $[a, b]$ as the number of points in the partition approaches infinity [66]. □

This seems an extremely odd result - we have observed through this example that Brownian motion is an object with unbounded first-order variation but bounded quadratic variation.

Moving on, it has been mentioned previously that the trajectories of Brownian motion are continuous but not differentiable (almost everywhere). One other argument providing some intuition as to why that is is the following (from [67]). Using results from normal distributions, note that:

$$\frac{W(t+h) - W(t)}{h} \sim \mathcal{N}(0, \frac{1}{h})$$

Recall that this is an object that approaches the derivative of $W(\cdot)$ in the limit $h \to 0$, but in this limit the variance diverges - there is no limit in any probabilistic sense. For a rigorous proof of continuity (using Hölder continuity) and non-differentiability, see [19].

It is straight forward to project the definition of one-dimensional Brownian motion onto

$\mathbb{R}^n$. By the arguments in section 3.2, a probability space can be built and on this space n independent one-dimensional Brownian motions $W^k(\cdot)$ constructed.

Definition 3.8. An $\mathbb{R}^n$-valued stochastic process $\mathbf{W}(\cdot) = (W^1(\cdot), ..., W^n(\cdot))$ is an n-dimensional Brownian motion if:

1. For each $k = 1, ..., n$, $W^k(\cdot)$ is a 1D Brownian motion.

2. The σ-algebras $U(W^k(t)|t \geq 0)$ are independent for $k = 1, ..., n$.

The σ-algebras U in the above are those generated by the random processes themselves (see lemma 2.1).

Lemma 3.12. If $\mathbf{W}(\cdot)$ is an n-dimensional Brownian motion, then:

1. $E(W^k(t)W^l(s)) = \min(t, s)\delta_{kl}$ $(k, l = 1, ..., n)$

2. $E((W^k(t) - W^k(s))(W^l(t) - W^l(s))) = (t - s)\delta_{kl}$ $(k, l = 1, ..., n; t \geq s \geq 0)$

The proof of this lemma is very clear, since when $k \neq l$ there is independence and when $k = l$ we can use the results from section 3.2.

To discuss the distributional properties of n-dimensional Brownian motion, we need to define the multivariate normal distribution.

Definition 3.9. (Multivariate normal) $\mathbf{x}$ is distributed as a d-dimensional normal random variable with mean vector μ and covariance matrix Σ, denoted $\mathbf{x} \sim N_d(\mu, \Sigma)$ if it has the following density function with respect to the measure $d\mathbf{x}$:

$$f(\mathbf{x}) = \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{x} - \mu)^T\Sigma^{-1}(\mathbf{x} - \mu))$$

The mean vector has elements equal to the mean of each Gaussian component, whilst the covariance matrix is symmetric and contains elements equal to the covariance between components. If the normal components are all independent then the covariance matrix is diagonal, with elements on the diagonal corresponding to the variances of each component. In the below theorem we denote the nxn identity matrix by I_n.

Theorem 3.13. If $\mathbf{W}(\cdot)$ is n-dimensional Brownian motion, then $\mathbf{W}(t) \sim \mathcal{N}_n(0, tI_n)$ for each time $t > 0$. Therefore,

$$P(\mathbf{W}(t) \in A) = \frac{1}{(2\pi t)^{\frac{n}{2}}} \int_A e^{-\frac{|\mathbf{x}|^2}{2t}} d\mathbf{x} \text{ for each Borel subset } A \subseteq \mathbb{R}^n.$$

This covariance matrix arises since the different Brownian components are independent and identically distributed with variance t, a fact that makes the proof of this theorem simple (see [59]).

Having a mathematical definition of Brownian motion is useful, but what does it look like and what intuition might this give us about where we might find it in real life? To add a visual interpretation and connect this theory to numerical computation, the next section discusses how Brownian motion can be simulated.

3.4 Simulation in 1D

This part aims to present Brownian motion and its properties visually. As stated at the end of the last part, the second main aim of this report is to numerically approximate the solutions of stochastic differential equations. Of course, since solutions to SDEs depend on Brownian motion, it is obvious that we will not be able to perform any approximation if we cannot simulate Brownian motion itself.

So first of all, what should we expect a realisation of this process on say $[0, 1]$ to look like? Well, the properties of the process suggest that the process starts from zero and the difference between the value of the function at points t and s (with $t \geq s$) are normally distributed with zero mean and variance $t-s$. This means that the trajectory should be made up of many tiny jumps which differ in magnitude and direction. Also, since the property is true for any $t \geq s$, even when the difference between the two points is arbitrarily small these jumps still occur. Furthermore, since any subset of the domain is uncountable (contains too many elements to count), any such interval $[s, t]$ can be split into arbitrarily many sub-intervals which all have these normally distributed jumps. Considering this informs us that trajectories of Brownian motion are fractal. A fractal is a curve, each part of which has the

same statistical character as the whole [21].

Consequently, the graph of the process is going to be very rough. In fact, it was noted previously that Brownian motion is continuous but not differentiable. How do we make sense of this? Continuity implies that as you zoom in on a small portion of the trajectory, there are no big jumps. Note that a jump between two close points on the domain is proportional to the distance (see the variance) between them, so there is no argument from this to dismiss continuity. On the other hand, from a visual perspective, a differentiable trajectory is one where as you zoom in closer and closer, the sample path moves more and more on a straight line (tangent). Since Brownian motion is fractal, differentiability will clearly not hold.

In computing, we are restricted by limits on storage. As a result, plotting requires discretisation (partitioning of the domain) into a finite number of points, which in the context of Brownian motion does not allow us to see a 'true' realisation of the process (since we would need all the points on an uncountable set). However, using a large number of points, we can get some feel for what the process looks like. The following algorithm describes the steps involved in a 1D simulation. The appendix contains a link and description of how to run the code that implements the algorithm in MATLAB.

Algorithm 1 Algorithm for simulation of 1D Brownian Motion

1: Discretise $[0, T]$ into N uniformly spaced points
2: $h = \frac{T}{N}$
3: $B(0) = 0$
4: **for** $i = 0$ to $N - 1$ **do**
5: Sample once from $\mathcal{N}(0, h)$ distribution, set value equal to $Increment$
6: $B((i + 1) * h) = Increment + B(i * h)$
7: **end for**

The output of the code that makes use of the algorithm above is given below.

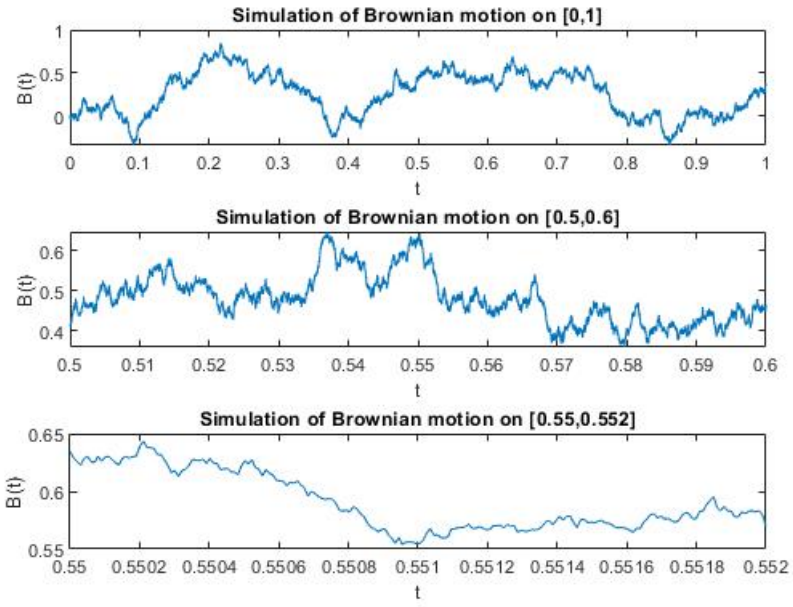

Figure 4: Simulation of 1D Brownian motion over [0,1]

The domain $t \in [0, 1]$ has been discretised into 100000 points. The second and third plots in the output above are magnifications of the first, targeting the sub-intervals $t \in [0.5, 0.6]$ and $t \in [0.55, 0.552]$ respectively. As expected, upon zooming in the character of the trajectory is similar to that on the larger domain and is very rough. The 'fingerprint' character of these plots allows Brownian motion to be identified as an argument in many real-life applications. Below is a figure showing the evolution of Facebook stock prices from the start of 2018 to 27 October 2021 from the Yahoo Finance website [73].

Figure 5: Price of Facebook shares (Source:[73])

The share price trajectory exhibits some similarities with the trajectories of 1D Brownian motion given in figure 5 in the sense that it is very rough and appears non-differentiable. Indeed, the log-returns (the logarithm of the price at time $t+1$ divided by the price at time t) of the share price can be modelled by a stochastic differential equation, which depends on Brownian motion and time.

In finance, objects tend to depend on time. Due to this, it is useful to have a record of the past behaviour of an object, as this might help forecast future movements. This motivates the need to define filtration, which we discuss in the next part.

3.5 Filtrations

Assume in this part that we have defined a probability space $(\Omega, \mathcal{F}, P)$. Filtrations are introduced in many of the references given in this report, including [66, 67, 60].

Definition 3.10. (Filtration) A filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is a set of σ-algebras $\mathcal{F}_t$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\mathcal{F}_t \subseteq \mathcal{F}$ for all t.

One special type of filtration is the natural filtration. The natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ encodes all information up to time t. At any time t, if we can say that an event A has occurred or not then $A \in \mathcal{F}_t$.

Example 3.3. Suppose we are going to throw a six-sided die twice. Let X_1 and X_2 be the outcomes of the first and second toss respectively, taking values in $\{1,2,3,4,5,6\}$. The sample space is the set of all ordered pairs, $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. As such, elementary events ω can be written as $\omega = (\omega_1, \omega_2)$, where $X_1(\omega) = \omega_1$ and $X_2(\omega) = \omega_2$.

There are three times of interest for this experiment: $t = 0$ (before any dice have been thrown), $t = 1$ (after the first die is thrown) and $t = 2$ (after the final die is thrown). For each of these times, the σ-algebra $\mathcal{F}_t$ encodes the information known at time t.

Firstly, at $t = 0$ no dice have been thrown, the only events that we can say for certain will or will not happen are $\varnothing$ and Ω. $\varnothing$ is the empty set - we know that rolling the dice will give us some outcome, so this won't happen. By the same argument the sample space Ω will happen. Therefore, we have that $\mathcal{F}_0 = \{\varnothing, \Omega\}$.

At $t = 1$, we have had one roll of the die, and having recorded the outcome we have some information that allows us to say for certain whether certain events have occurred. For example, if we take the event A as the collection of possible outcomes for which the first roll takes value one, then since we know after the first roll whether we got value one or not we can say for certain whether A has occurred. Then $\mathcal{F}_1$ is the collection of all events which we can say have or have not happened after roll one, and clearly also contains those elements that belong to $\mathcal{F}_0$.

Finally, at $t = 2$, we know the outcome of both rolls and so know exactly the elementary event ω that has occurred. As a result, if we found every subset of the sample space Ω, we could go through each of them and say for certain whether or not they happened. It is clear then that $\mathcal{F}_2$ is the collection of all the subsets of Ω, containing both $\mathcal{F}_1$ and $\mathcal{F}_0$.

We say that a stochastic process X(t) is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if $X(t)$ is measurable with respect to $\mathcal{F}_t$ for all $t \geq 0$.

Definition 3.11. (Filtered probability space) The quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ is called a filtered probability space.

Many stochastic models used in finance and other applications involve predicting values of a process in the future which depend on the present value alone. This memorylessness is an important notion, which we introduce in the following part.

3.6 Markov Processes

The fact that so many models in reality possess this useful property simplifies working with these processes. Essentially, the Markov property says that the future behaviour of a process $\zeta(t)$ depends only on its value at time t and so all times $s < t$ can be ignored [67]. At the end of this section we will prove that Brownian motion satisfies the Markov property. For more detail on this topic see [14], and for further reading see [36].

Definition 3.12. Let $\zeta(t)$ be a stochastic process defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. Then $\zeta(t)$ satisfies the Markov property if, for any bounded Borel-measurable

function f and for any times $s \leq t$, the following relationship holds:

$$E(f(\zeta(t))|\mathcal{F}_s) = E(f(\zeta(t))|\zeta(s)).$$

On the left side of the equality, we have an expectation given $\mathcal{F}_s$, which contains all the information available from time 0 to s. The right side tells us that we only need the value of the process at time s in order to find the expectation, so the extra information for all times before s can be ignored.

Equivalently, it follows that when we have a Markovian process $\zeta(t)$, for every bounded Borel-measurable function $f(x)$ there is another function $g(x)$ (depending on s and f) such that:

$$E(f(\zeta(t))|\mathcal{F}_s) = g(\zeta(s))$$

Proposition 3.14. Brownian motion, defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, is Markovian.

Proof. We need to prove that $W(t)$ satisfies the Markov property. From the definition, we take an arbitrary bounded Borel-measurable function $f(x)$ and form a second function associated with f, $h(x, y) = f(x + y)$. Then we have $f(x) = h(x - y, y)$, and so for any times $s \leq t$ we can write:

$$E(f(W(t))|\mathcal{F}_s) = E(h(W(t) - W(s), W(s))|\mathcal{F}_s)$$

Now, observe that $W(s)$ is $\mathcal{F}_s$-measurable and $W(t) - W(s)$ is independent of $\mathcal{F}_s$. Then, using proposition 2.6, the right side in the above is equal to $g(W(s))$ for some function $g(x)$. Comparing this with the equivalent definition of Markov property given above proves that $W(t)$ is Markovian. This proof is also given in [67]. □

If $\zeta(t)$ is a Markov process then we have for events $A \in \mathbb{R}$:

$$\begin{aligned} P(\zeta(t) \in A|\mathcal{F}_s) &= E(I_{\zeta(t)\in A}(\omega)|\mathcal{F}_s) \\ &= E(I_{\zeta(t)\in A}(\omega)|\zeta(s)) \\ &= P(\zeta(t) \in A|\zeta(s)) \\ &= P(s, \zeta(s); t, A) \text{ a.s.} \end{aligned}$$

Working through the calculation, note that $I_{\zeta(t)\in A}(\omega)$ is the indicator function, which takes value 1 in the event that $\zeta(t) \in A$ and takes value 0 otherwise. The probability of taking value 1 is the term on the left side, and the value 0 has no effect on the expectation so the first equality holds. The second equality arises using definition 3.12, which can easily be evaluated. We call the function in the final line, $P(s, x; t, y)$, the Markov transition function [67].

Now that we have constructed Brownian motion and discussed the properties it possesses, we can move on to developing the theory of stochastic integration and differential equations.

4 Stochastic Integrals and Ito Calculus

In section 3.1 the notion of stochastic differential equations was introduced and motivated. We had:

$$\dot{\mathbf{X}}(t) = \mathbf{b}(\mathbf{X}(t)) + \mathbf{B}(\mathbf{X}(t))\xi(t),$$
$$\mathbf{X}(0) = x_0.$$

We also had the heuristic equation $\xi(t) = \frac{dW(t)}{dt}$. If we substitute this into the equation above and multiply both sides by the differential dt, we obtain:

$$d\mathbf{X}(t) = \mathbf{b}(\mathbf{X}(t), t)dt + \mathbf{B}(\mathbf{X}(t), t)dW(t).$$

Recall that $\xi(t)$ does not exist, and consequently neither does $dW(t)$. It is therefore helpful to interpret the above in integral form:

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{b}(\mathbf{X}(s), s)ds + \int_0^t \mathbf{B}(\mathbf{X}(s), s)dW(s).$$

It is clear that this equation satisfies the earlier SDE. Differentiating both sides with respect to t returns the above using the fundamental theorem of calculus [38].

The first term in the integral equation will not cause any trouble since the initial condition is given. The second looks like it could be more complex, however we know that Brownian motion is continuous so this is a standard Riemann integral. Finally, the third term is an integration with respect to Brownian motion. How to compute this term is not at all obvious. Clearly, if we are interested in a solution to integral equations of this type, we must first define the integral:

$$\int_0^T \mathbf{G}dW$$

where here **G** is some function of stochastic processes and potentially time.

4.1 Paley-Wiener-Zygmund Integral

We approach the problem posed in the previous part by considering a simpler problem.

Definition 4.1. (Paley-Wiener-Zygmund Integral) Suppose we have a continuous, differentiable, deterministic function g on the domain [0,1] which takes values in $\mathbb{R}$ and satisfies the condition $g(0) = g(1) = 0$. Then the Paley-Wiener-Zygmund integral [50] is defined as:

$$\int_0^1 g dW = -\int_0^1 g' W dt.$$

We can see from the right term that the integral is actually a random variable. The following lemma and proof investigates the properties of the integral.

Lemma 4.1. The Paley-Wiener-Zygmund integral satisfies the following:

1. $E(\int_0^1 g dW) = 0.$
2. $E((\int_0^1 g dW)^2) = \int_0^1 g^2 dt.$

Proof.

$$E(\int_0^1 g dW) = -\int_0^1 g' E(W(t)) dt = 0$$

Here we substitute definition 4.1 then using the fact that g is deterministic we can move the expectation under the integral sign before using the property $E(W(t)) = 0$ from the previous chapter.

$$\begin{aligned} E\Big(\Big(\int_0^1 g dW\Big)^2\Big) &= E\Big(\Big(\int_0^1 g(t) dW(t)\Big)\Big(\int_0^1 g(s) dW(s)\Big)\Big) \\ &= \int_0^1 \int_0^1 g'(t) g'(s) E(W(t) W(s)) ds dt \\ &= \int_0^1 g'(t)\Big(\int_0^t s g'(s) ds + \int_t^1 t g'(s) ds\Big) dt \\ &= \int_0^1 g'(t)\Big(-\int_0^t g ds\Big) dt = \int_0^1 g^2 dt \end{aligned}$$

In the first line the square is split into two identical integrals. Then definition 4.1 is applied. Recall that the covariance of two Brownian motion at times s, t is the minimum of s and t.

This is used to obtain the third equality. From here we need to use integration by parts on the first term in the bracket, setting $u = s$ and $dv = g'(s)$. The outcome of this is $tg(t) - \int_0^t g(s)ds$, using the fact that $g(0) = g(1) = 0$. The second term in the bracket evaluates to $-tg(t)$, leaving the fourth line which again requires integration by parts, this time with u equal to the integral in the bracket and dv equal to $g'(t)$. This proof comes from [19]. □

Of course, the properties that the function g must satisfy for this definition are very restrictive, and so we need to devise a definition for a wider class of integrands. Nevertheless, the definition we come up with later in this chapter agrees with the properties above when g is a function of this type.

4.2 Ito Integral

In standard Riemann integration, an integral of function $f(x)$ is computed by splitting the domain into N sub-intervals $[x_i, x_{i+1}], i = 0, 1, ..., N-1$, then approximating the area under the curve by summing the areas of rectangles with height $f(t_i)$, where t_i is in the sub-interval $[x_i, x_{i+1}]$, before taking the limit $N \to \infty$. The sum discussed here is the Riemann sum [69]. Therefore, it seems a reasonable approach for our problem to construct a Riemann sum approximation in a similar way (but now with random f) and then take the mean-square limit. This idea was introduced by K.Ito [29].

Definition 4.2. (Ito stochastic integral) Let $h(t)$ be a continuous stochastic process adapted to the filtration $\mathcal{F}_t$ such that $\int_0^T E(h(t))^2 dt < \infty$. Then the Ito stochastic integral is defined as the following mean-square limit:

$$\int_0^T h(t)dW(t) = l.i.m_{h\to 0} \sum_{i=0}^{n-1} h(t_i)(W(t_{i+1}) - W(t_i))$$

Here, $0 = t_0 < t_1 < ... < t_n = T$ is a partition of $[0, T]$, and l.i.m is the limit in the mean-square sense, i.e.

$$E\left(\int_0^T h(t)dW(t) - \sum_{i=0}^{N-1} h(t_i)(W(t_{i+1}) - W(t_i)) \right)^2 \to 0 \text{ as } \delta \to 0$$

where δ is the mesh size of the partition.

The Ito integral exists for any continuous, bounded process $h(t)$ [72]. Note that in the sum $h(t)$ is evaluated at the point t_i, on the left of each sub-interval. This is essential as if $h(t)$ is evaluated at a different point within the sub-interval, we get alternative stochastic integrals which may approach another limit entirely. This is a consequence of the unbounded variation of Brownian motion (see section 3.3). This is something to be wary of, unless $h(t)$ has a continuous first derivative. In this case, the definition of the stochastic integral doesn't depend on the point in each sub-interval $[t_i, t_{i+1}]$ at which $h(t)$ is evaluated, this is explained below.

Proposition 4.2. If $h(t)$ has a continuous first derivative then:

$$l.i.m_{h\to 0}\sum_{i=0}^{N-1} h(t_i)(W(t_{i+1}) - W(t_i)) = l.i.m_{h\to 0}\sum_{i=0}^{N-1} h(t_{i+1})(W(t_{i+1}) - W(t_i))$$

where l.i.m is the limit in the mean-square sense.

Proof. First, we combine the sums by subtracting the left side from the right:

$$\sum_{i=0}^{N-1}(h(t_{i+1}) - h(t_i)(W(t_{i+1}) - W(t_i)).$$

Using the mean value theorem [38], we can rewrite the first bracket as:

$$h(t_{i+1}) - h(t_i) = h'(t_i^*)(t_{i+1} - t_i)$$

where we used the fact that $h(t)$ has a continuous first derivative. t_i^* here is some number in the sub-interval $[t_i, t_{i+1}]$ for each $i = 0, 1, ..., N-1$. Then, replacing this in the sum, we obtain:

$$l.i.m_{h\to 0}\sum_{i=0}^{N-1}(h(t_{i+1}) - h(t_i))(W(t_{i+1}) - W(t_i)) = l.i.m_{h\to 0}\sum_{i=0}^{N-1} h'(t_i^*)(t_{i+1} - t_i)(W(t_{i+1}) - W(t_i))$$

Indeed, we can see that on the right side we have the change in t multiplied with the change in Brownian motion. As the change in t goes to zero, both terms in the product approach zero. Note that evaluating $h(t)$ at any point in each $[t_i, t_{i+1}]$ would give the same result using

the mean-value theorem, showing that the claim we made above is valid. □

In fact, when $h(t)$ is differentiable we can express the Ito integral in terms of a Riemann integral and random term:

$$\int_0^T h(t)dW(t) = h(T)W(T) - \int_0^T W(t)h'(t)dt.$$

This can be proved using the definition of the stochastic integral as a sum and the summation by parts formula [12], and it is a direct consequence of the Ito formula which we discuss later in this chapter.

So why did we pick to evaluate $h(t_i)$? Notice in the Ito formulation above that using this point means that $h(t_i)$ (which may depend on Brownian motion) is independent of the increment it is multiplied with, and so we are guaranteed to obtain zero when we take the expectation [62]. This justifies the first property given below. A proof of these properties can be found in [19, 33].

Properties of Ito integrals: Let a, b be real numbers and $g(t)$, $h(t)$ be continuous stochastic processes. Then

1. $\int_0^T ag(t) + bh(t)dW(t) = a\int_0^T g(t)dW(t) + b\int_0^T h(t)dW(t)$
2. $E\Big(\int_0^T h(t)dW(t)\Big) = 0$
3. $E\Big(\Big(\int_0^T h(t)dW(t)\Big)^2\Big) = E\Big(\int_0^T h(t)^2dt\Big)$
4. $E\Big(\int_0^T h(t)dW(t)\int_0^T g(t)dW(t)\Big) = E\Big(\int_0^T h(t)g(t)dt\Big)$

The following example illustrates how the sum given in definition 4.2 can be used to evaluate the Ito stochastic integral.

Example 4.1. We calculate $\int_0^T W(t)dW(t)$. Using definition 4.2, the integral is the mean-square limit of the following sum:

$$\sum_{i=0}^{n-1} W(t_i)(W(t_{i+1}) - W(t_i))$$

It can easily be checked that

$$W(t_i)(W(t_{i+1}) - W(t_i)) = \frac{1}{2}(W(t_{i+1})^2 - W(t_i^2)) - \frac{1}{2}(W(t_{i+1}) - W(t_i))^2$$

The second quantity on the right is the quadratic variation which we evaluated earlier. Therefore, we obtain:

$$\sum_{i=0}^{n-1} W(t_i)(W(t_{i+1}) - W(t_i)) = \frac{1}{2}\sum_{i=0}^{n-1}(W(t_{i+1})^2 - W(t_i^2)) - \frac{1}{2}\sum_{i=0}^{n-1}(W(t_{i+1}) - W(t_i))^2$$
$$\rightarrow \frac{W(T)^2}{2} - \frac{T}{2} \text{ almost surely in the limit } n \rightarrow \infty.$$

We can do a simple sanity check here by taking the expectation. Using the properties of Brownian motion from the previous chapter, $E(W(T)^2) = T$ and so the expectation of the integral is indeed zero.

Based on the definition of Ito integral that has been given in this part, we now have everything we need to rigorously define stochastic differential equations.

4.3 Stochastic Differential Equations

We are ready to introduce the definition of stochastic differential equations, plus discuss a few examples and how they might model objects in real-life applications. First we consider the definition with one-dimensional Brownian motion, then extend to multiple dimensions.

Definition 4.3. (Stochastic differential equation in 1D) Assume that $b(t, x)$and $B(t, x)$ have good analytical properties [19]. Then the solution of the stochastic differential equation

$$dX = b(t, X)dt + B(t, X)dW(t), X(t_0) = x,$$

is a stochastic process $X(t)$ that is measurable with respect to the filtration $\mathcal{F}_t$ and satisfies the integral equation

$$X(t) = x + \int_{t_0}^{t} b(s, X_{t_0,x}(s))ds + \int_{t_0}^{t} B(s, X_{t_0,x}(s))dW(s).$$

The notation $X_{t_0,x}(\cdot)$ simply captures the fact that the initial state of $X(t)$ is $X(t_0) = x$. Besides this, the integral equation consists of a Riemann integral and a stochastic Ito integral. An important note to make is that the two forms are equivalent, but the integral form is more rigorous (since SDEs have dW, which we know doesn't exist).

We now discuss the analytical properties that $b(t, x)$ and $B(t, x)$ might satisfy. First, it is worth mentioning that both of these functions share the domain $[t_0, t] \times \mathbb{R}$. Both of these functions should be measurable with respect to the filtration $\mathcal{F}_t$. One strong constraint that can be placed on them, which guarantees that a solution exists, is the global Lipschitz condition. This states that there should exist constants C and D such that:

$$|b(x,t)| + |B(x,t)| \leq C(1 + |x|);$$
$$|b(x,t) - b(y,t)| + |B(x,t) - B(y,t)| \leq D|x - y|,$$

for all t and all x, y in $\mathbb{R}$. A proof of the existence and uniqueness of the solutions under these conditions is given in [30]. This is just one set of possible constraints that ensure a solution exists and is unique, and a pretty strong set of conditions at that. Other less restrictive properties are sufficient in some cases for a solution to exist, which is a large area of research (see [42] for further reading).

When we have multiple Brownian motions, there is an analogous definition of stochastic differential equations, given below.

Definition 4.4. (Stochastic differential equation in n dimensions) Assume that $b(t, X)$ and $B_r(t, X)$ have good analytical properties, and that $X(t)$ is progressively measurable with respect to $\mathcal{F}_t$. Then $X(t)$ is a solution of the SDE

$$dX = b(t, X)dt + \sum_{r=1}^{q} B_r(t, X)dW_r(t), X(t_0) = x,$$

if it can be written in integral form as:

$$X(t) = x + \int_{t_0}^{t} b(s, X_{t_0,x}(s))ds + \sum_{r=1}^{q} \int_{t_0}^{t} B_r(s, X_{t_0,x}(s))dW_r(s).$$

Here, q is the number of independent Brownian motions involved. The global Lipschitz condition, which we might impose on the coefficients to guarantee a solution, in the multi-dimensional case is:

$$|b(x,t) - b(y,t)| + \sum_{r=1}^{q} |B_r(x,t) - B_r(y,t)| \leq K|x-y|$$

where K is a constant, x, y are vectors, and $|\cdot|$ is the Euclidean norm.

We now introduce a few examples of SDEs. The examples we choose originate from finance. For more discussion on their background, please see [66].

Example 4.2. (Geometric Brownian motion) In finance, modelling stock prices is essential in order to price options and make investment decisions. One such model is **Geometric Brownian motion**, or GBM for short. It suggests that the stock price S satisfies the following SDE:

$$dS = \mu S\, dt + \sigma S\, dW$$

There are two coefficients here: μ is the drift, which indicates by how much we expect the stock price to grow over time (normally measured in years). Note that if we ignored the Brownian motion term in this equation we would have a simple deterministic ODE with solution $S(t) = S_0 e^{\mu t}$. This is identical to what we would expect of a bank account with continuously compounded interest rate μ.

The volatility σ dictates how much the stock price can bounce around and vary. If the value of the coefficient is high then investment is more risky. In GBM, the volatility is set equal to a constant, which is not realistic. One other disadvantage of GBM is that there are no occasional large sudden jumps or drops in stock price which we know can occur in real life. One case study worth mentioning is the drop in Viacom CBS stock prices by over a half over just one week, observed in March 2021. The drop was largely due to the actions of Archegos capital management, a family office which had enormous investments (billions) in the stock (see [57]).

A simple extension of GBM is the constant elasticity of variance (CEV) model, which is

of the form:

$$dS = \mu S\,dt + \sigma S^{\gamma} dW,$$

This adds more flexibility, as setting $\gamma > 1$ increases volatility for larger values for stock prices which is appropriate in some markets whilst setting $\gamma < 1$ is appropriate in markets where volatility increases as prices fall.

These models could be improved by instead allowing volatility to be considered as a random variable as well. σ can be replaced by a function v that obeys it's own SDE. This motivates **stochastic volatility models**, which have systems of SDEs [4]:

$$dS = \mu S\,dt + \sqrt{v} S\,dW_1$$
$$dv = \alpha_{v,t} dt + \beta_{v.t} dW_2$$

Here, W_1 and W_2 are correlated Brownian motions, and $\alpha_{v,t}$ and $\beta_{v,t}$ are functions of the volatility v and time. One popular form of the stochastic differential equation for v is given by the Heston model [28]:

$$dv = \theta(\eta - v)dt + \varepsilon\sqrt{v} dW_2,$$

where η is the mean long-term volatility, θ is the rate at which v reverts to its long-term mean and ε is the volatility of v. W_1 and W_2 have constant correlation ρ.

The benefit of introducing GBM and using it to test our numerical schemes later in the report is that it has an explicit solution with which we can compare approximations. We start to discuss methods for solving stochastic differential equations next.

4.4 Ito formula

If we take a perturbation of a sufficiently smooth deterministic function of x, say $g(x + \Delta x)$, where Δx is small, then Taylor's theorem (see [1] for details) allows us to express g as a polynomial incorporating the derivatives of g as coefficients of powers of Δx. Specifically,

including terms up to quadratic power, we may write $g(x + \Delta x)$ as:

$$g(x + \Delta x) = g(x) + \frac{dg}{dx}\Delta x + \frac{1}{2}\frac{d^2g}{dx^2}(\Delta x)^2 + ...$$

Suppose that x is a function of t, so that $\Delta x = x(t + \Delta t) - x(t)$. Then $g(x + \Delta x) = g(x(t + \Delta t))$, substituting this into the above and rearranging gives:

$$g(x(t + \Delta t)) - g(x(t)) = \frac{dg}{dx}\Delta x + \frac{1}{2}\frac{d^2g}{dx^2}(\Delta x)^2 + ...$$

Letting Δt go to zero and ignoring terms of $(\Delta x)^2$ and higher gives:

$$dg = \frac{dg}{dx}dx \Rightarrow \frac{dg}{dt} = \frac{dg}{dx}\frac{dx}{dt}.$$

Of course, this is the chain rule. What about if $x(t)$ was a stochastic process? Specifically, suppose that $x(t)$ is Brownian motion in one dimension. Then, we have from the Taylor expansion:

$$g(W(t + \Delta t)) - g(W(t)) = \frac{dg(W)}{dx}\Delta W + \frac{1}{2}\frac{d^2g(W)}{dx^2}(\Delta W)^2 + ...$$

In the limit $\Delta t \to 0$, we can no longer ignore the quadratic term since the mean-square limit of the quadratic variation is non-zero. In fact, it was shown that the expectation of the quadratic variation is Δt and the variance approaches zero under the mean-square limit. Therefore, we can replace $(\Delta W)^2$ with Δt in the limit, which gives (ignoring terms of order higher than Δt):

$$dg(W) = \frac{dg(W)}{dx}\Delta W + \frac{1}{2}\frac{d^2g(W)}{dx^2}dt.$$

This is Ito's formula in one variable.

Example 4.3. Suppose $g(x) = x^2$, and x is Brownian motion. Then $\frac{dg}{dx} = 2x$ and $\frac{d^2g}{dx^2} = 2$, using Ito's formula we have:

$$dg = d(W^2) = 2WdW + dt.$$

This is the stochastic differential equation of W^2. We can check this using the integral form:

$$\begin{aligned} g(W(t)) &= g(W(t_0)) + \int_{t_0}^{t} ds + \int_{t_0}^{t} 2W(s)dW(s) \\ &= W(t_0)^2 + t - t_0 + W(t)^2 - t - W(t_0)^2 + t_0 \\ &= W(t)^2, \end{aligned}$$

where we used example 4.1 to evaluate the stochastic integral (we split it into the difference of two integrals as the limits do not match with those in the example).

Ito's formula can be generalised to the case when $X(t)$ is a stochastic process depending on both time and Brownian motion in a similar way. Suppose $S(t)$ is the solution of the following SDE:

$$dS = a(S)dt + b(S)dW.$$

Then using the Taylor expansion and our knowledge of the quadratic variation:

$$dg = \left(a(S)\frac{dg}{dx} + \frac{b(S)^2}{2}\frac{d^2g}{dx^2}\right)dt + b(S)\frac{dg}{dx}dW.$$

Note that if we set $a = 0$ and $b = 1$, then we retrieve the Ito formula that we had before.

Example 4.4. (Solving GBM) Take the stock price S, and assume that it satisfies the GBM stochastic differential equation,

$$dS = \mu S\,dt + \sigma S\,dW \text{ with } S(0) = S_0.$$

Let g(x)=ln(x), the natural logarithm of x. Then $\frac{dg}{dx} = \frac{1}{x}$ and $\frac{d^2g}{dx^2} = \frac{-1}{x^2}$. Using the generalised Ito formula, we have that:

$$dg = (\mu - \frac{\sigma^2}{2})dt + \sigma dW.$$

Putting this into integral form and solving, plus substituting back the formula for g, we obtain:

$$ln(S(t)) = ln(S_0) + (\mu - \frac{\sigma^2}{2})t + \sigma W(t)$$
$$\Rightarrow S(t) = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t)}.$$

We have found the explicit solution.

We can also extend Ito's formula to smooth functions g that depend on both time and stochastic processes. Again, assume $S(t)$ depends on time and Brownian motion, which can be described by an SDE as before. Then from the Taylor expansion we obtain:

$$dg = \left(\frac{\partial g}{\partial t} + a\frac{\partial g}{\partial x} + \frac{b^2}{2}\frac{\partial^2 g}{\partial x^2}\right)dt + b\frac{\partial g}{\partial x}dW.$$

This is the same as we had before with a new term accounting for the time dependence of g. This is Ito's formula in two variables.

Finally, in the case that we have multiple dimensions such that the stochastic process $X(t)$ has an SDE of the form given in definition 4.4, the analogous Ito formula involves replacing previous partial derivatives with respect to X with the gradient and hessian operators [30]. We can rewrite the sum on the right side of definition 4.4 as a matrix-vector product $\mathbf{G}_t d\mathbf{W}_t$, giving:

$$\begin{aligned} dg(t, \mathbf{X}_t) &= \frac{\partial g}{\partial t}dt + (\nabla_X g)^T d\mathbf{X}_t + \frac{1}{2}(d\mathbf{X}_t)^T(\mathbf{H}_X g)d\mathbf{X}_t \\ &= \left\{\frac{\partial g}{\partial t} + (\nabla_X g)^T \mathbf{b}(t, \mathbf{X}) + \frac{1}{2}Tr(\mathbf{G}_t^T(\mathbf{H}_X g)\mathbf{G}_t)\right\}dt + (\nabla_X g)^T \mathbf{G}_t d\mathbf{W}_t. \end{aligned}$$

Recall that the gradient of a scalar function f is a vector with elements equal to the partial derivative of f with respect to each component $x_1, x_2, ...x_n$. The Hessian (denoted $\mathbf{H}_X$ here) is the matrix of second derivatives and cross derivatives with respect to the components of f above. The trace of a matrix is the sum of values on the diagonal and the transpose swaps the dimension of a matrix or vector around. Since the Brownian motions in definition 4.4 are assumed to be independent, we get no cross terms in the final formula, only the diagonal terms from the Hessian are present (since the trace is calculated). This is due to

the fact that in the mean-square limit the quadratic variation goes to the differential dt.

This part has provided justification of Ito's formula, rigorous proofs can be found in [30, 33]. We go through two more examples of how Ito's formula can be used to solve SDEs and other tricks that are similar to those from ODEs below.

Example 4.5. (Using Ito formula in two variables) Suppose we have a stochastic process $f(t, W(t))$ which satisfies the following SDE:

$$df(t) = Wdt + tdW$$
$$f(0, W(0)) = 0.$$

How do we solve for the process $f(t)$? Let us assume that $f(t) = g(t, W(t))$ and apply Ito's formula, replacing W(t) with x:

$$dg = \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\right)dt + \frac{\partial g}{\partial x}dW.$$

We now equate the coefficients of the differentials of the SDE for f and Ito's formula. The right term looks the most simple to solve, so we start with that.

$$\frac{\partial g}{\partial x} = t \Rightarrow g(t, x) = tx + A(t),$$

where $A(t)$ is some unknown function of t. Computing the other derivatives gives:

$$\frac{\partial g}{\partial t} = x + A'(t),$$
$$\frac{\partial^2 g}{\partial x^2} = 0.$$

Comparing with the coefficient of dt in the SDE gives:

$$x + A'(t) = x$$
$$\Rightarrow A(t) = a$$
$$\Rightarrow g(t, x) = tx + a.$$

Reintroducing $W(t)$ and replacing g with f gives the general solution $f(t) = tW(t) + a$. Lastly, we find that $a = 0$ by using the initial condition, giving the solution $f(t) = tW(t)$. A more applied example is given below, which shows another simple method in which the solution of an SDE can be found.

Example 4.6. (Langevin's equation in one dimension) The modelling of particles using Brownian motion can be improved by taking frictional forces into account. Including the effects of these forces leads to an SDE of Langevin's equation [34]:

$$dV = -\alpha V dt + \sigma dW$$
$$V(0) = V_0.$$

Consider deriving an SDE for $Ve^{\alpha t}$, using Ito's formula in two variables (here t and V), we have $a = -\alpha V$ and $b = \sigma$:

$$d(Ve^{\alpha t}) = ((\alpha e^{\alpha t})V - \alpha V(e^{\alpha t}))dt + \sigma e^{\alpha t} dW = \sigma e^{\alpha t} dW.$$

The cancellation of the bracket forms a new SDE which is very simple to solve. Integrating both sides gives:

$$V(t)e^{\alpha t} - V(0) = \sigma \int_0^t e^{\alpha s} dW(s)$$

Finally, we can rearrange for V(t) to give:

$$V(t) = V(0)e^{-\alpha t} + \sigma \int_0^t e^{\alpha(s-t)} dW(s).$$

Note that this equation is also used in financial mathematics to model instantaneous rates (see [66] for further reading), solutions of which are Ornstein-Uhlenbeck processes.

Other very useful formulae from standard calculus include the product rule and integration by parts. What are the equivalent formulae in Ito calculus? The following is based on the notes in [19].

Lemma 4.3. (Ito product rule) Suppose that u and v satisfy the following stochastic differ-

ential equations:

$$du = F_1 dt + G_1 dW$$

$$dv = F_2 dt + G_2 dW,$$

where F_i and G_i $(i = 1, 2)$ have good analytical properties (such as the global Lipschitz condition discussed earlier). Then:

$$d(uv) = udv + vdu + dudv = udv + vdu + G_1 G_2 dt.$$

The final term on the right-hand side is the Ito correction term. The integral form of the product rule can be rearranged to give the Ito integration by parts formula:

$$\int_s^r udv = u(r)v(r) - u(s)v(s) - \int_s^r vdu - \int_s^r G_1 G_2 dt.$$

An interesting remark is that if either G_1 or G_2 equal zero (which is the case if u or v depend solely on time t), then we get back the Paley-Wiener-Zygmund definition from section 4.1. A proof of the Ito product rule is given below.

Proof. The first thing to note is that $uv = \frac{1}{2}(u+v)^2 - \frac{1}{2}u^2 - \frac{1}{2}v^2$. We can now find the differential of each of these terms to find $d(uv)$. By Taylor expansion, $d(u^2) = 2udu + (du)^2$ and $d(v^2)$ is of the same form where v by symmetry. Now, let $g = u + v$, so that $dg = du + dv$. In the same way, $d(g^2) = 2gdg + (dg)^2$. Substituting $dg = du + dv$ back in and expanding $(dg)^2$ gives:

$$d(g^2) = d((u+v)^2) = 2(u+v)(du+dv) + (du)^2 + 2dudv + (dv)^2.$$

We now have all the different aspects we need to compute $d(uv)$.

$$\begin{aligned} d(uv) &= \frac{1}{2}d((u+v)^2) - \frac{1}{2}d(u^2) - \frac{1}{2}d(v^2) \\ &= (u+v)(du+dv) + \frac{1}{2}(du)^2 + dudv + \frac{1}{2}(dv)^2 - udu - \frac{1}{2}(du)^2 - vdv - \frac{1}{2}(dv)^2 \\ &= (u+v)(du+dv) + dudv - udu - vdv \\ &= udv + vdu + dudv, \end{aligned}$$

□

In the next part, we describe how Ito's formula can be applied to help us solve parabolic partial differential equations and obtain probabilistic representations.

4.5 Probabilistic representations of solutions

In this part, we consider how Ito's formula can be used to find the solution of parabolic PDEs. We will begin with a reminder of what it means for a PDE to be parabolic before giving an example from finance of how to find the probabilistic solution.

Definition 4.5. (Parabolic PDE) Take a real-valued function of two independent real variables τ and x, $u(\tau, x)$. A second-order, linear partial differential equation for u takes the form:

$$A\frac{\partial^2 u}{\partial x^2} + 2B\frac{\partial^2 u}{\partial x \partial \tau} + C\frac{\partial^2 u}{\partial \tau^2} + D\frac{\partial u}{\partial x} + E\frac{\partial u}{\partial \tau} + Fu = G$$

where here A, B, C, D, E, F, G are coefficients which may depend on τ and x. The PDE is *parabolic* type if $B^2 - AC = 0$ for all τ and x.

For further reading on the topic of partial differential equations including more background on the definition above, see [20]. We move on to an example which combines many of the techniques seen so far in this chapter. The main reference for this example is [66].

Example 4.7. In this example, we denote time by τ which is considered over the interval $t \leq \tau \leq T$. Firstly, we derive the Black-Scholes equation for a European option. An option is a type of contract that gives the holder the right, but not the obligation, to buy from or sell

to the writer a stock at some future date.This relies on the assumption that the underlying stock price $S(\tau)$ can be modelled by geometric Brownian motion (introduced earlier), so $S(\tau)$ satisfies the SDE:

$$dS = rS\,d\tau + \sigma S\,dW.$$

Related to an option is the payoff function f which describes the potential profit of holding the option. There are two main types of options, namely the call (holder wants to buy) and put (holder wants to sell) options. In the case of European options, the holder may exercise their right to buy or sell a stock for a strike price K only at the maturity time T. The payoff of owning a call option at this final time is $f(S(T)) = \max(S(T) - K, 0)$. The idea behind this function is as follows: if the stock price at time T exceeds the strike price K, the holder of the option can buy the stock for price K and instantly sell at time T for a profit of S(T)-K, otherwise they won't exercise the option as it's cheaper for them to buy the stock at price S(T) i.e. the option is worthless.

Pricing an option correctly from the writer's perspective is finding the amount which allows them to hedge the option. Hedging is a strategy that the writer can follow so that regardless of the stock price at maturity T, they can meet their obligation to the option holder. The writer's portfolio under this strategy involves two assets: an amount of stock (denoted ϕ) and an amount of money held in the bank (denoted ψ). Summarising, the writer needs to find $V(\tau, S(\tau))$ for $t \leq \tau \leq T$ such that $V(T, S(T)) = f(S(T))$, with $S(t) = S_t$ assumed known. We derive the Black-Scholes equation below.

Applying the Ito formula to $V(\tau, S(\tau))$ we get:

$$dV = (\frac{\partial v}{\partial \tau} + rS\frac{\partial v}{\partial x} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 v}{\partial x^2})d\tau + \sigma S\frac{\partial v}{\partial x}dW.$$

Let $(\phi(\tau), \psi(\tau))$ be the writers portfolio. The value of the portfolio at time τ is

$$V(\tau, S(\tau)) = \phi(\tau)S(\tau) + \psi(\tau),$$

and a small time $\Delta\tau$ later the value of the portfolio becomes (assuming no change to amount

of stock held, and using the fact that the bank account accrues interest $\frac{B(\tau+\Delta\tau)}{B(\tau)}$ over this period):

$$V(\tau+\Delta\tau, S(\tau+\Delta\tau)) = \phi(\tau)S(\tau+\Delta\tau) + \psi(\tau)\frac{B(\tau+\Delta\tau)}{B(\tau)}.$$

Finding the difference between these two times and taking the limit that $\Delta\tau \to 0$:

$$V(\tau+\Delta\tau, S(\tau+\Delta\tau)) - V(\tau, S(\tau)) = \phi(\tau)(S(\tau+\Delta\tau) - S(\tau)) + \psi(\tau)\frac{B(\tau+\Delta\tau) - B(\tau)}{B(\tau)}$$

$dV = \phi dS + \psi\frac{dB}{B}$ under the limit $\Delta\tau \to 0$.

Now, the bank account can be modelled by $B(\tau) = B(0)e^{r\tau}$ so that we have $\frac{dB}{d\tau} = rB$ and $dB = rBd\tau$. Substituting this along with the SDE of the stock price $S(\tau)$, we obtain:

$$dV = (\psi r + \phi rS)d\tau + \phi\sigma S\, dW.$$

Comparing the coefficients of this SDE with the one we got from applying the Ito formula, we have:

$$\begin{aligned}
&\sigma S\frac{\partial V}{\partial x} = \phi\sigma S \\
&\Rightarrow \phi = \frac{\partial V}{\partial x} \\
&\frac{\partial v}{\partial \tau} + rS\frac{\partial v}{\partial x} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 v}{\partial x^2} = \psi r + \phi rS \\
&\Rightarrow \frac{\partial v}{\partial \tau} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 v}{\partial x^2} = \phi r.
\end{aligned}$$

Considering the last equation, note the following:

$$\begin{aligned}
&V = \phi S + \psi \\
&\Rightarrow \phi r = rV - \phi S r \\
&\Rightarrow \phi r = rV - rS\frac{\partial V}{\partial x}.
\end{aligned}$$

In the last line, we substituted ϕ for the partial derivative which we found earlier. Finally,

substituting this into the right side of the above and collecting the terms gives:

$$\frac{\partial V}{\partial \tau} + rS\frac{\partial V}{\partial x} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial x^2} - rV = 0.$$

This is the Black-Scholes equation, which we aimed to derive. Checking that the PDE is parabolic is simple using the definition, B and C are equal to zero so $B^2 - AC = 0$. The rest of this example solves this equation to give a probabilistic representation. Consider the product of V with another function $u(\tau) = e^{-r(\tau - t)}$, which is such that

$$\begin{aligned}\frac{du}{d\tau} &= -re^{-r(\tau-t)} = -ru \\ \Rightarrow du &= -rud\tau.\end{aligned}$$

Using Ito's formula and substituting x for S, we have that if g is any smooth function of τ and $x(\tau)$ then it satisfies:

$$dg(\tau, S(\tau)) = (\frac{\partial g}{\partial \tau} + rS\frac{\partial g}{\partial x} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 g}{\partial x^2})d\tau + \sigma S\frac{\partial g}{\partial x}dW.$$

This holds for any function g, so it also holds for $V(\tau, S(\tau))$ which we can replace in the above. Note that the equation in front of the $d\tau$ term is close to the PDE which we aim to solve, it is just missing the term $-rV$. Applying the product rule gives:

$$\begin{aligned}d(uV) &= udV + Vdu + dudV \\ &= u((\frac{\partial V}{\partial \tau} + rS\frac{\partial V}{\partial x} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial x^2})d\tau + u\sigma S\frac{\partial V}{\partial x}dW - ruVd\tau \\ &= u(\frac{\partial V}{\partial \tau} + rS\frac{\partial V}{\partial x} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial x^2} - rV)d\tau + u\sigma S\frac{\partial V}{\partial x}dW.\end{aligned}$$

We can identify the term in the first bracket as the parabolic PDE which is equal to zero.

Then:

$$
\begin{aligned}
&d(uV) = u\sigma S\frac{\partial V}{\partial x}dW \\
&\Rightarrow \int_t^T d(uV) = \int_t^T u\sigma S\frac{\partial V}{\partial x}dW \\
&\Rightarrow u(T)V(T,S(T)) - u(t)V(t,S(t)) = \int_t^T u\sigma S\frac{\partial V}{\partial x}dW \\
&\Rightarrow E(e^{-r(T-t)}V(T,S(T))) - V(t,S(t)) = 0 \\
&\Rightarrow V(t,S(t)) = e^{-r(T-t)}E(f(S(T))).
\end{aligned}
$$

From the top equation, we integrate then solve the left side. Thirdly, we take expectations (the expectation of an Ito integral is zero), where we recall that t and $S(t)$ are known and finally rearrange and substitute $f(S(T))$ [48].

We can use the techniques from these examples to approach more general problems.

Example 4.8. (Feynman-Kac formula) The parabolic differential operator is given by:

$$L = \frac{\partial}{\partial\tau} + a(\tau,x)\frac{\partial}{\partial x} + \frac{b^2(\tau,x)}{2}\frac{\partial^2}{\partial x^2}.$$

The Cauchy problem is:

$$
\begin{aligned}
&Lu = 0,\ t \in [t_0, T] \\
&u(T,x) = \phi(x),\ x \in \mathbb{R},
\end{aligned}
$$

along with an associated SDE for X given by

$$
\begin{aligned}
&dX = a(s,X)ds + b(s,X)dW,\ s \geq t \\
&X(t) = x.
\end{aligned}
$$

Assume that a sufficiently smooth unique solution $u(s, X_{t,x}(s))$ exists (this is shown in [22] when a and b are subject to certain conditions). The notation $X_{t,x}(s)$ is introduced to indicate

a solution to the associated SDE with the initial condition. Applying the Ito formula gives:

$$
\begin{aligned}
&du(s, X_{t,x}(s)) = Luds + b(s, X_{t,x}(s))\frac{\partial u}{\partial x}dW(s) \\
&\Rightarrow u(T, X_{t,x}(T)) - u(t,x) = \int_t^T b(s, X_{t,x}(s))\frac{\partial u}{\partial x}dW(s) \\
&\Rightarrow u(t,x) = E(u(T, X_{t,x}(T))) = E(\phi(X_{t,x}(T))).
\end{aligned}
$$

Here, we integrated the stochastic differential equation, recognised that the term Lu vanishes and solved the left side. We then took expectations which sends the right side to zero, and rearranged to get the final formula.

An even more general form of the problem above is:

$$
\begin{aligned}
&Lu + c(t,x)u + g(t,x) = 0, \ t \in [t_0, T] \\
&u(T,x) = \phi(x), \ x \in \mathbb{R}.
\end{aligned}
$$

It turns out that the solution has the probabilistic representation:

$$
\begin{aligned}
u(t,x) =& E\Big(\phi(X_{t,x}(T)) \exp \int_t^T c(s, X_{t,x}(s))ds \\
&+ \int_t^T g(s, X_{t,x}(s)) \exp \int_t^s c(s', X_{t,x}(s'))ds'ds\Big).
\end{aligned}
$$

This representation is commonly called the Feynman-Kac formula. For further reading on this, including the multidimensional form of the formula, see [72, 33].

4.6 Summary of chapter 4

This chapter has introduced the notions of stochastic integrals and stochastic differential equations as well as methods that can be used to solve them explicitly. Next, ideas of how we might approximate the solutions to stochastic differential equations using numerical methods are discussed. The derivation and implementation of these numerical schemes is the second main aim of this report.

5 Stochastic Numerics: Strong Taylor Approximations

One large portion of undergraduate scientific computation modules is the study and analysis of numerical methods to solve ordinary and partial differential equations. As the power of computing has grown, so have the number of methods that we can use to approximate these problems. Much is known about the stability and accuracy of these schemes, so it makes sense to ask the question: what numerical methods might be used to solve stochastic differential equations, and what are their properties?

5.1 Generalising the Euler method

One incredibly popular and simple numerical method for ODEs is the explicit (or forward) Euler method [31]. The idea behind this is as follows, suppose we have an ordinary differential equation of the form:

$$\frac{dx}{dt} = b(t, x), \text{ where } t \in [t_0, T],$$
$$x(t_0) = x_0,$$

and we wish to obtain approximations for x for values of t in the interval $[t_0, T]$. We form a partition of this interval into $N + 1$ points: $t_i = t_0 + ih, i = 0, 1, ..., N$, where $h = \frac{T-t_0}{N}$ is called the step-size. If the step-size h is small, then approximating x at each of the points t_i and interpolating should give a reasonable estimate for the function x over the domain. This is the general setup for any numerical method for this problem, the variation comes through how these approximations are calculated (see [17] for more details on construction and analysis of different approximations).

An alternative way to express the above problem, which resembles what we saw in the previous chapter, can be obtained by integrating:

$$x(t) = x_0 + \int_{t_0}^{t} b(s, x(s))ds.$$

Consider estimating x at the point $t_1 = t_0 + h$. If the step-size h is small, a sensible idea

might be to assume that $b(s,x)$ is almost constant over the interval, $b(s,x(s)) \approx b(t_0,x_0)$. Solving the integral is then easily done:

$$x(t_1) \approx x_0 + hb(t_0,x_0).$$

For the next time point, we can do a similar thing. The integral now has an upper limit of $t_2 = t_0 + 2h$, but we can split it into the sum of two integrals, the first having the form above and the second being the integral over the interval $[t_1,t_2]$. The only difference between the approximations of the successive points is this second integral, the integrand of which we can assume constant on each sub-interval. Continuing this for N steps gives approximations (denoted by Y_i at time t_i) over the interval:

$$Y_{i+1} = Y_i + hb(t_i,Y_i),\ i = 0,...,N-1.$$

This method is explicit because the right side of the equation depends only on the current time t_i and not any future time. Analysis of the error in these approximations is relatively simple, and can be studied by truncating the Taylor expansion of $x(t)$ around t_0. The local error (error that comes from two successive approximations) is proportional to h^2 whilst the error accumulated over the whole interval, or global error, is proportional to h. As a result, we say that the forward Euler method is first-order accurate in the step-size h. Simple, more accurate extensions such as the Adams-Bashforth and Runge-Kutta methods exist and truncate the Taylor expansion later but require more evaluations of $b(t,x)$ which can be computationally expensive [10].

This all seems simple enough, but what happens when we incorporate Brownian motion? Extending the problem above, we have the following SDE:

$$dx = b(t,x)dt + \sum_{r=1}^{q} \sigma_r(t,x)dW_r(t),$$

$$x(t_0) = x_0 \text{ where } t \in [t_0,T].$$

Note there are many independent Brownian motions here. Again, we set the problem up in

the same way by partitioning the interval into $N+1$ points and we wish to approximate x at each of these points. The integral form of the above is:

$$x(t) = x_0 + \int_{t_0}^{t} b(s, x(s))ds + \sum_{r=1}^{q} \int_{t_0}^{t} \sigma_r(s, x(s))dW_r(s).$$

Well, we can treat the integral over s as before, but it is not obvious how we should treat the stochastic integrals. From the previous chapter, we know that the stochastic integral can be written as a mean-square limit of a sum depending on the function σ_r and the increment of Brownian motion. A reasonable idea if h is very small is to assume that this function is constant over the interval. We can split the integrals in the same way as before to get approximations for every time point. This method can be summarised in the following scheme (again, using Y_i to denote approximations):

$$\begin{aligned} Y_{i+1} &= Y_i + hb(t_i, Y_i) + \sum_{r=1}^{q} \sigma_r(t_i, Y_i)(W_r(t_{i+1}) - W_r(t_i)) \\ &= Y_i + hb(t_i, Y_i) + \sqrt{h} \sum_{r=1}^{q} \sigma_r(t_i, Y_i)Z_{r,i}, \end{aligned}$$

where for the second line we used the properties of Gaussian variables, $W_r(t_{i+1}) - W_r(t_i) = \sqrt{h}Z_{r,i}$ where $Z_{r,i}$ has a standard normal distribution. This gives us a clear idea of how we can implement this method and others on a computer.

In the next part, we discuss how the explicit Euler scheme can be implemented by applying it to a finance-based problem of tracking stock price.

5.2 Implementing the Explicit Euler scheme

Suppose we wish to study the evolution of a stocks price over a period of 8 years. One way to do this is by assuming the stock price follows geometric Brownian motion (which was first mentioned in example 4.2), which requires constants μ and σ representing the drift and volatility respectively. We showed earlier that the solution of GBM is given by $S(t) = S(0)e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t)}$.

In this part, we simulate trajectories of the stock price under GBM and compare them to

approximations produced via the explicit Euler scheme. Following this, a general function that implements the numerical scheme will be discussed and tested on a second SDE which also has a known solution.

First of all, how does one simulate a trajectory of the stock price using the GBM solution? Since this depends on Brownian motion, we need to partition the interval of interest. We have for the first time point that:

$$\begin{aligned} S(t+h) &= S(0)e^{(\mu-\frac{\sigma^2}{2})(t+h)+\sigma W(t+h)} \\ &= S(0)e^{(\mu-\frac{\sigma^2}{2})t+\sigma W(t)}e^{(\mu-\frac{\sigma^2}{2})h+\sigma(W(t+h)-W(t))} \\ &= S(t)e^{(\mu-\frac{\sigma^2}{2})h+\sigma(W(t+h)-W(t))} \end{aligned}$$

Now, from the properties of Brownian motion, the terms in the product on the right side are independent and the increment in the argument of the exponential term is normally distributed with mean zero and variance h. Therefore, we can replace the increment by a product of a standard normal random variable with the square root of h. This makes it clear how to use the previous time point to obtain the next time point.

We move on to approximating GBM using the explicit Euler scheme. Recall that the SDE that the stock price satisfies under GBM is:

$$dS = \mu S\,dt + \sigma S\,dW.$$

Comparing this to the formulation of the method given in the previous section, the explicit Euler scheme for GBM can be written as:

$$\begin{aligned} Y_0 &= S(0) \\ Y_{i+1} &= Y_i + h\mu Y_i + \sqrt{h}\sigma Y_i Z_i. \end{aligned}$$

Here, we replaced the increment again so that it is in a form more easy to implement on a computer.

To analyse the accuracy of the approximation, we need to use the same sample path for

both the true solution and the approximation. This allows us to get a sense of the closeness of the approximation and also to see how this changes as the number of time-steps used varies. Forcing the sample paths to match adds some extra difficulty in coding, as the vector of observations of the standard normal random variables sometimes needs to be passed between multiple functions. To motivate what is meant by this, consider the case that the coefficients of the SDE depend on Brownian motion. Then the sample path must also be passed to the function that evaluates these coefficients, and appropriate portions of this must be summed in order to evaluate Brownian motion at any point in the partition.

The function *explicitEulerGBM.m* (which can be found through the link in the appendix) takes inputs of the initial value of the stock, the annual drift, volatility and interval of interest as well as two values for the number of time steps - one for the true solution and another for the approximation. It also takes the optional input of a vector zValues, which represents the sample path of the true solution (which in this case we know) from a previous call to the function. The outputs include the approximation, true solution and the sample path used to produce each of them.

To use this function, it needs to be called from a separate driver script (also available via the link). In the first instance, a large number of time steps (large to obtain a realistic trajectory) should be passed for the true solution, ideally a number that has many divisors so that many different numbers of time steps can be passed through further calls, for example 2048. This number should stay constant for every following call in order to check that the approximation can be formed using the sample path generated. Any smaller number of time steps can be passed to form the first approximation. In any given call to the function, an error will be displayed if the number of time steps for the approximation is less than that of the true solution. Furthermore, in order to evaluate the increment for the time-step of the approximation exactly, the number of time steps of the approximation needs to be a factor of that for the true solution. This is easy to understand - any increment of the approximation needs to be able to be written as a sum of increments of the true solution otherwise we would need further evaluations of the random variables in the sample path to match them up, but these are not available to us. Therefore, if these inputs do not satisfy this, a warning

is displayed stating that the number of time steps passed for the approximation is not a factor of that passed for the true solution. The number of time steps in the approximation is then altered within the function to the nearest value that makes it a factor of the number of time steps in the solution so that the same sample path may be used. A message is displayed to the console window notifying the user of the new number of time steps used in the approximation.

In the first call from the driver, the input zValues can be omitted. The code for the function is written so that if this argument is not passed, the sample path is generated. In all the calls thereafter, if the user wishes to use the same sample path they can collect it as an output from this first call and pass it as the zValues argument. In this way the sample paths are matched and approximations using different numbers of time steps can be obtained for comparison. The figure below shows a number of explicit Euler approximations along with the true solution where (μ,σ)=(0.08,0.08) and $S_0 = 160$, over the interval [0,8], which can be interpreted as years.

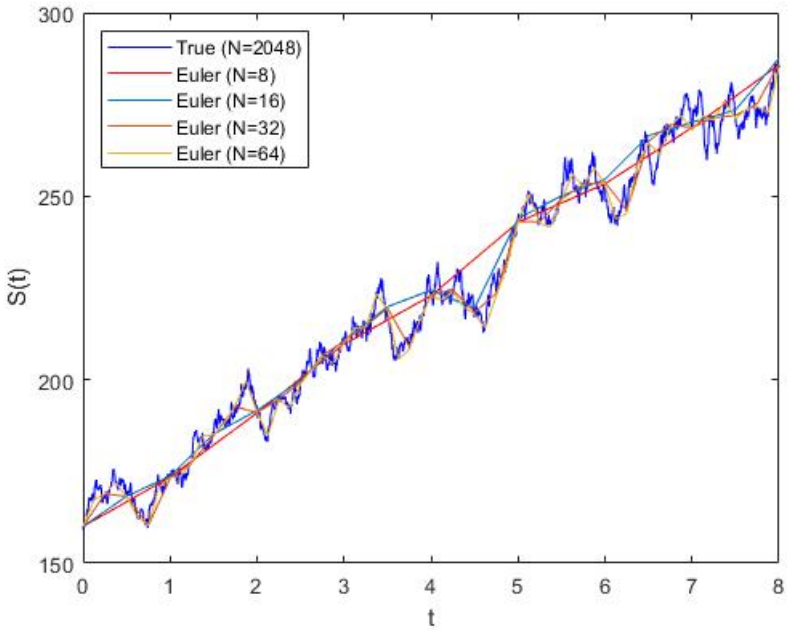

Figure 6: Plot of true solution of GBM with approximations for various numbers of time-steps N.

It can clearly be seen that the approximation improves as the number of time-steps increases. For 8 time-steps, the approximation is quite poor, and does not manage to

capture the up and down behaviour especially between times 4 and 5. On the other hand, the yellow line with 64 time-steps captures the general shape of the true trajectory well. Below is a magnification of the above figure that allows this to be seen much more clearly.

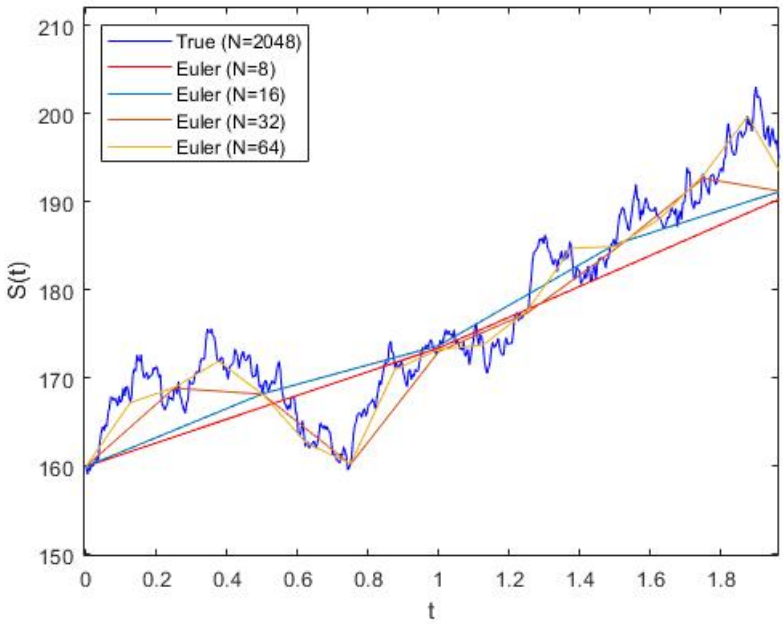

Figure 7: Zoomed-in image of figure 7.

In this part, we discussed how to carry out the explicit Euler scheme for a specific problem depending on one Brownian motion, where the coefficients depend on the process itself. This hopefully motivates the reader on how the scheme may be implemented in general, but it is worth discussing how the scheme is applied for other problems, particularly in the more complex case that the coefficients of the SDE depend on Brownian motion. This particular difficulty is covered in the next part, which designs a general function for approximating the solution of SDEs using the explicit Euler method.

5.3 Generalising the implementation of the Explicit Euler Method

Now that we have managed to implement the explicit Euler scheme in the simple case of GBM, we can extend its application further by creating a function that may be used for more general problems including those with more than one Brownian motion. Recall the general

SDE that we wish to approximate is given by:

$$dX = b(t, X)dt + \sum_{r=1}^{q} \sigma_r(t, X)dW_r(t),$$

$$X(t_0) = X_0 \text{ where } t \in [t_0, T].$$

The explicit Euler scheme for this SDE is:

$$\begin{aligned} Y_{i+1} &= Y_i + hb(t_i, Y_i) + \sum_{r=1}^{q} \sigma_r(t_i, Y_i)(W_r(t_{i+1}) - W_r(t_i)) \\ &= Y_i + hb(t_i, Y_i) + \sum_{r=1}^{q} \sigma_r(t_i, Y_i)\sqrt{h}Z_{ri}, \end{aligned}$$

where Z_{ri} is a standard normal variable. In order to find the approximation, it is necessary to write both a function that allows one to evaluate $b(t_i, Y_i)$ and $\sigma_r(t_i, Y_i)$ for all i as well as a function that uses these evaluations to successively produce the approximation along the interval.

The first function mentioned above may depend on other parameters (for example, we had μ and σ in GBM), so it is worth including a vector of parameters as an input along with t and x. On top of this, if the coefficients depend on Brownian motion then the sample path and step-size needs to be passed on which adds an extra layer of difficulty. Ideally, the output of such a function should be a single vector, the first element of which is $b(t, x)$ followed by q elements containing $\sigma_r(t, x)$ for $r = 1, ..., q$. Access to this vector at each time point in the partition allows the approximations to be calculated. See the link in the appendix for an example of a function (namely *evaluateFunctions.m*) that can achieve this in the case that GBM is the SDE of interest, it is mentioned in the comments of the code how this can be adapted for any SDE which has coefficients that do not depend on Brownian motion.

The second function that implements the explicit Euler scheme should take minimum inputs of: the function that evaluates $b(t, x)$ and $\sigma_r(t_i, x_i)$, the vector of parameters for the aforementioned function (this makes it easier to investigate the effect of the parameter values, and can contain the step-size if required for the calculation of Brownian motion along the interval), the initial state of the system, the number of time-steps, and values for

the start and endpoints of the interval. Besides these inputs, a sample path can be passed along with the factor - this is how many times the approximate number of time-steps goes into that used for the sample path. The output should of course be the vector of explicit Euler approximations as well as the sample path used if requested by the user. An algorithm that allows one to generate a sample path and obtain the approximation is given below.

Algorithm 2 Algorithm for explicit Euler approximation of SDE

Discretise $[t_0, T]$ into N uniformly spaced points
Set $h = \frac{T-t_0}{N}$ and $t = t_0$
Set the first element in the approximation equal to x_0
for $i = 1$ to N **do**
    Evaluate $b(t_{i-1}, x_{i-1})$ and $\sigma_r(t_{i-1}, x_{i-1})$ for $r = 1, ..., q$.
    Set $x_i = x_{i-1} + h * b(t_{i-1}, x_{i-1})$
    for $r = 1$ to q **do**
        Sample once from the $\mathcal{N}(0, 1)$ distribution, set this equal to z
        Add onto the approximation by evaluating $x_i = x_i + \sqrt{h} * \sigma_r(t_{i-1}, x_{i-1}) * z$
    end for
    Update the time, $t = t + h$
end for

For an implementation of this algorithm in MATLAB, see *explicitEulerMethod.m* through the link in the appendix. Note that if the inputs zValues and factor are not provided by the user, the function generates zValues and sets the factor equal to 1. This is useful for when the function is first called from a driver and no sample path has been generated.

These functions allowing the approximation of general SDEs were tested using first the geometric Brownian motion problem which we had previously and secondly with $f(t) = tW(t)$, which satisfies the following SDE:

$$df(t) = W(t)dt + tdW(t).$$

Note that having a coefficient depending on Brownian motion makes matching all the sample paths much more difficult. The step-size and sample path both need to be passed to the function that evaluates the coefficients. For the true solution $f(t)$, the following relation can be used, which is easily seen by applying Ito's formula to the left side:

$$(t + h)W(t + h) = (t + h)W(t) + (t + h)(W(t + h) - W(t)).$$

The first term on the right-hand side can be evaluated using the approximation from the previous time point whilst the second has an increment which is independent and normally distributed with mean zero and variance h. The following figure shows the true solution of this problem in the interval $[0, 8]$ along with explicit Euler approximations of varying step-size.

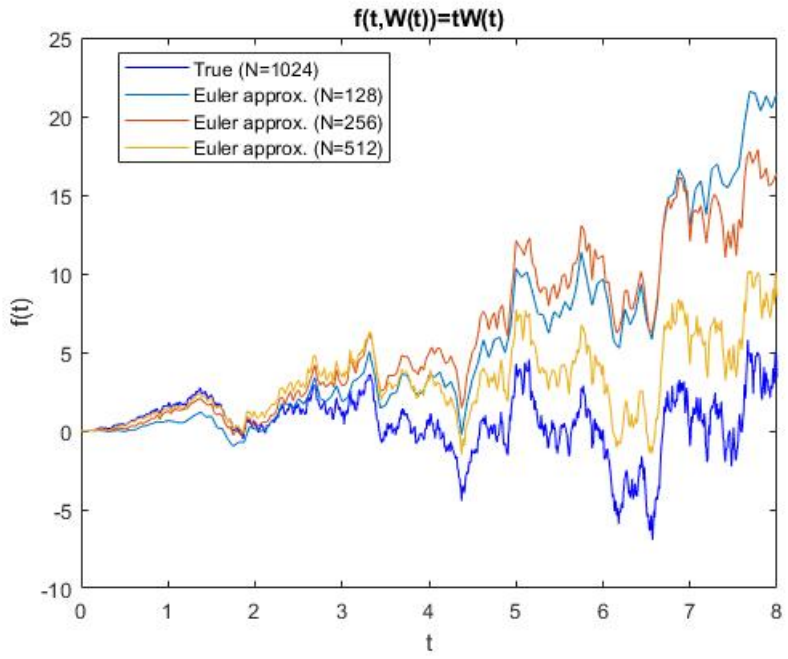

Figure 8: Plot of true and approximate solutions to f(t,W)=tW(t).

The error all the way along the interval is observable even with the largest number of time-steps $N = 512$. Despite this, the approximation gets noticeably closer to the true solution as the number of time-steps increases. The reason for the discrepancy may be due to the fact that the coefficient C bounding the absolute error depends on the sample path ω. This can be large. Indeed, this seems to be the case as when we consider another trajectory such as the one below we get:

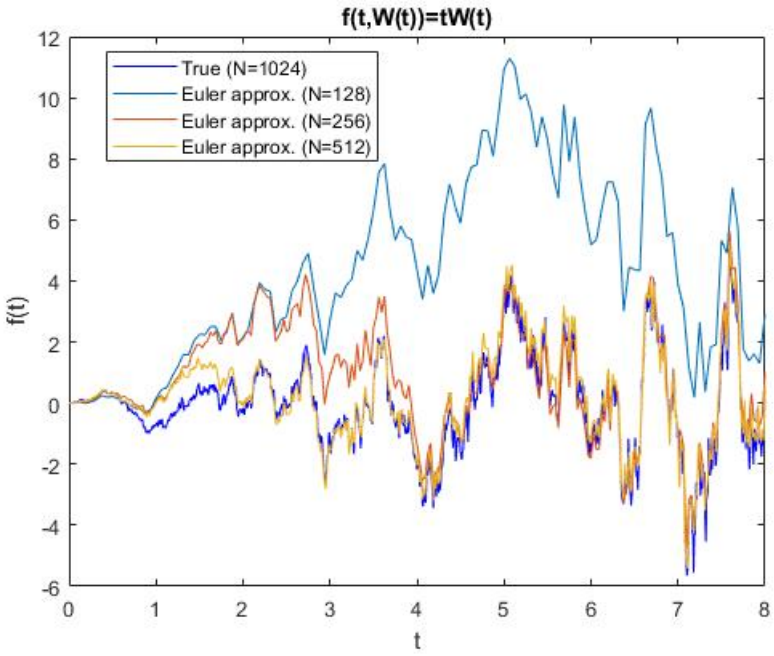

Figure 9: Plot of true and approximate solutions to f(t,W)=tW(t) for another trajectory.

In this case, the approximation with 512 time-steps follows closely the true solution all along the interval whilst the approximation with 128 time-steps is much further away. This variability is why it is important to take caution when interpreting these approximations and also to consider more than one time-step.

One thing that we need to consider is how accurate the method is. Since we are dealing with random objects, the approximate solutions will be a family of trajectories and this adds a degree of complexity, how do we decide how accurate these trajectories are to the true solution, which is also random? The next part introduces the notion of mean-square approximation, which allows us to answer questions like this.

5.4 Mean-square approximation

In this section, we aim to create a framework which allows us to analyse the accuracy of numerical schemes for SDEs.

Definition 5.1. Consider the general multidimensional SDE given in the previous chapter.

If for a numerical scheme we have:

$$(E(x(t_k) - Y_k)^2)^{\frac{1}{2}} \leq Ch^p,$$

then we say that the mean-square order of accuracy of the method is p.

For example, with the explicit Euler method given in the previous section, subject to conditions on the functions $b(t, x)$ and $\sigma_r(t, x)$ G.Marujama proved mean-square convergence in [41]. Then, I.Gichman and A.Skorkhod [24] showed that this method had mean-square order of accuracy of $p = \frac{1}{2}$.

In order to introduce the fundamental theorem on mean-square order of convergence, we need to describe the conditions that the functions $b(t, x)$ and $\sigma_r(t, x)$ should satisfy as well as introduce the so-called one-step approximation. What will the theorem tell us? Recall that for a deterministic scheme, a local error of order p+1 leads to a corresponding global error of order p, which can be easily seen by summing the N local errors and making the observation that $N = \frac{T-t_0}{h}$. A similar result will be presented in the theorem, which is proven in [43].

Let $(X(t), \mathcal{F}_t)$, $t_0 \leq t \leq T$ be a solution to the general multidimensional SDE with $E(|X(t_0)|^2) < \infty$. The one-step approximation $\bar{Y}_{t,x}(t+h)$, $t_0 \leq t < t+h \leq T$ depending on x, t, h and the set of independent Brownian increments $\{w_1(\theta) - w_1(t), ..., w_q(\theta) - w_q(t), t \leq \theta \leq t+h\}$ is defined as:

$$\bar{Y}_{t,x}(t + h) = x + A(t, x, h; w_i(\theta) - w_i(t), i = 1, ..., q, t \leq \theta \leq t + h).$$

This can be used to construct approximations at further points in the partition, which we denote $\bar{Y}_k$, that are $\mathcal{F}_{t_k}$-measurable for all $k = 0, 1, ..., N$. Keeping things general, let $h_{k+1} = t_{k+1} - t_k$:

$$\bar{Y}_0 = x(t_0),$$

$$\bar{Y}_{k+1} = \bar{Y}_k + A(t_k, \bar{Y}_k, h_{k+1}; w_i(\theta) - w_i(t_k), i = 1, ..., q, t_k \leq \theta \leq t_{k+1}),$$

where $k = 0, 1, ..., N-1$. Linking back to the Euler method for SDEs, when the partition is uniform (h constant over the interval) and $\theta = t_{k+1}$ for all k, setting the function $A(t_k, \bar{Y}_k, h; w_i(\theta) - w_i(t), i = 1, ..., q)$ equal to $hb(t_k, \bar{Y}_k) + \sum_{r=1}^{q} \sigma_r(t_k, \bar{Y}_k)(w_r(t_{k+1}) - w_r(t_k))$ gives back the scheme that we defined. Now that we have this, we can move on to discussing the conditions underlying the fundamental theorem.

For simplicity, assume the partition of the interval is into $N+1$ uniform points so the step-size h is constant and equal to $h = \frac{T-t_0}{N}$. Assume that $b(t, x)$ and $\sigma_r(t, x)$ are defined and continuous on the interval $t \in [t_0, T]$, x is real, and they satisfy the global Lipschitz condition (mentioned in section 4.3). That is, for all $t \in [t_0, T]$, x, y real, the following inequality is satisfied:

$$|b(t, x) - b(t, y)| + \sum_{r=1}^{q} |\sigma_r(t, x) - \sigma_r(t, y)| \leq K|x - y|.$$

Here, $|\cdot|$ is the Euclidean norm and K is a constant.

Theorem 5.1. (Fundamental theorem on mean-square order of convergence) For arbitrary $t_0 \leq t \leq T - h$ and real x, suppose the following inequalities hold:

$$|E(X_{t,x}(t + h) - \bar{Y}_{t,x}(t + h))| \leq K(1 + |x|^2)^{\frac{1}{2}} h^{p_1}$$
$$[E|X_{t,x}(t + h) - \bar{Y}_{t,x}(t + h)|^2]^{\frac{1}{2}} \leq K(1 + |x|^2)^{\frac{1}{2}} h^{p_2}.$$

Also, let:

$$p_2 \geq \frac{1}{2}$$
$$p_1 \geq p_2 + \frac{1}{2}.$$

Then for any N, $k = 0, 1, ..., N$ the following inequality holds:

$$[E|X_{t,x}(t_k) - \bar{Y}_{t,x}(t_k)|^2]^{\frac{1}{2}} \leq K(1 + E|X_0|^2)^{\frac{1}{2}} h^{p_2 - \frac{1}{2}}$$

In other words, if the one-step approximation $\bar{Y}_{t,x}(t + h)$ has p_1 order of accuracy in the expectation of the deviation with the true solution and p_2 order of accuracy in the mean-

square deviation, subject to the inequalities above, then the method has $p = p_2 - \frac{1}{2}$ mean-square order of accuracy. The proof of this theorem is given in many texts including [45] (also see [43, 44]).

Similar to the mean-square order of accuracy is the strong order of accuracy [31]. If for some method,

$$E(|X(t_k) - Y_k|) \leq Kh^p,$$

where K is a positive constant independent of h and k, we say the strong order of accuracy of the method is p. Note that a method with mean-square order of accuracy p necessarily has strong order p. This is not immediately obvious, but can be cleared up in the following way. Suppose we have a p-th order accurate method in the mean-square sense, so that $(E(X(t_k) - Y_k)^2)^{\frac{1}{2}} \leq Ch^p$. Recall that for a random variable X, $Var(X) = E(X^2) - E(X)^2 \geq 0$ so $E(X^2) \geq E(X)^2$. Now, let $X = |X(t_k) - Y_k|$. Substituting this into the inequality and taking the square root gives:

$$E(|X(t_k) - Y_k|) \leq (E(X(t_k) - Y_k)^2)^{\frac{1}{2}} \leq Ch^p.$$

Later in this chapter, we will confirm the strong order of accuracy for the Euler method computationally.

Now knowing how the explicit Euler method can be implemented, we should use this framework to discuss how accurate its approximations are. In the next two sections, we consider criteria based on the inequalities in theorem 5.1.

5.5 Analysing the Euler scheme: The absolute error criterion

For general SDEs, the true solution is often not known explicitly, so approximations allow us to find out about how they behave. In the case of geometric Brownian motion, we do know the explicit solution, and so we can measure how well the approximation performs using the absolute error criterion [31]:

$$\varepsilon = E(|X(T) - Y(T)|)$$

This quantity tells us how close we can expect the approximation to be to the true solution at the end of the interval, after all the local errors have accumulated. Recall from section 5.2 the notion of strong order of accuracy. The expectation considered is the same as the absolute error criterion when at the final time. We noted that a mean-square order of accuracy equal to p necessarily has strong order p. For this part, we will continue the example of geometric Brownian motion and check that the strong order of accuracy of the explicit Euler method is indeed equal to $\frac{1}{2}$ through computer experiment.

In order to do this, we need to find a way to estimate the expectation. One way we can do this is by simulating many independent sample paths (say N) of the Ito process and its associated approximation, and finding the absolute difference between each of them at time T before averaging. Denoting the appropriate values of the k-th simulated trajectory at time T by $X_{T,k}$ and $Y_{T,k}$, we have:

$$\varepsilon \approx \frac{1}{N}\sum_{k=1}^{N}|X_{T,k} - Y_{T,k}|$$

The expectation of each term in this sum is the absolute error criterion, so the expectation of the whole of the right side is the absolute error criterion. In other words, this estimator is unbiased. As for the variance, using $Var(aX) = a^2 Var(X)$, we have that:

$$\begin{aligned} Var(\frac{1}{N}\sum_{k=1}^{N}|X_{T,k} - Y_{T,k}|) &= \frac{1}{N^2}\sum_{k=1}^{N}Var(|X_{T,k} - Y_{T,k}|) \\ &= \frac{Var(|X_T - Y_T|)}{N} \to 0 \text{ as } N \to \infty. \end{aligned}$$

Since N is large, we know from the central limit theorem that the estimator is asymptotically Gaussian and converges in distribution to the absolute error criterion as the number of simulated trajectories goes to infinity. The knowledge of the asymptotic distribution of the estimator means that we can construct confidence intervals for the absolute error criterion.

We will return to confidence intervals later in this part, but first we will look at some estimates for the absolute error criterion when increasing numbers of time-steps are con-

sidered. Using the GBM problem posed in section 5.3 and 100 simulations, we evaluate the estimator with step-sizes ranging from $h = 0.125$ to $h = \frac{8}{2048}$, where h is successively halved. The results are given in the following table:

h	Absolute error	Ratio
0.125	1.125	1
0.0625	0.87925	1.2796
0.03125	0.58617	1.4923
0.015625	0.42463	1.3875
0.0078125	0.29532	1.4379
0.0039063	0.21159	1.3957

The table shows that as the step-size h decreases, the absolute error also decreases monotonically. The ratio in the third column is calculated by dividing the absolute error from the previous step-size with the absolute error at the current step-size. When the step-size gets very small, this ratio settles around the value $\sqrt{2}$. This implies that when the step-size is halved the absolute error criterion shrinks by a factor of $\sqrt{2}$, which is what we should expect given the strong order of accuracy of the method is $\frac{1}{2}$.

For the larger step-sizes in the table, the ratios are quite variable. The most likely cause of this is due to the bound (proportional to $\sqrt{h}$) being an asymptotic result, so it is only revealed for small enough h. This is further reason to consider confidence intervals for the estimators, as for other problems we may be restricted in terms of the number of simulations we can conduct due to large computational or time cost.

To construct these confidence intervals, we arrange the simulations into M batches with N simulations in each. The purpose of this is so that we can use the Student's t-distribution to construct the confidence intervals. Note that if $X_1, ..., X_n$ are n independent Gaussian variables ($n > 3$) with known mean μ and unknown variance, then if we define the sample mean $\overline{\mu}$ in the standard way (i.e. the arithmetic mean) and the sample variance using

$$\overline{\sigma}^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{\mu})^2,$$

the following quantity satisfies a Student's t-distribution with $n-1$ degrees of freedom, zero mean and variance $\frac{(n-1)}{(n-3)}$:

$$T_n = \frac{\overline{\mu} - \mu}{\sqrt{\frac{\overline{\sigma^2}}{n}}}.$$

For further reading on the Student's t-distribution, see [37]. For this problem, there is no evidence that each term in the estimator is Gaussian, but we can form approximate Gaussian variables by arranging the simulations into batches and using the Central Limit theorem before considering the quantity asymptotically. Therefore, denoting by $Y_{T,k,j}$ the value of the approximation at the final time on the k-th trajectory of the j-th batch and $X_{T,k,j}$ the associated Ito process, we form a set of average errors.

$$\overline{\varepsilon}_j = \frac{1}{N}\sum_{k=1}^{N} |X_{T,k,j} - Y_{T,k,j}| \text{ for j=1,...,M.}$$

Each average in the set is independent and identically distributed (approximately Gaussian for large N). Taking the mean of the batch averages $\overline{\varepsilon}$, an unbiased estimate for the variance is given by:

$$\overline{\sigma}_{\varepsilon}^2 = \frac{1}{M-1}\sum_{j=1}^{M}(\overline{\varepsilon}_j - \overline{\varepsilon})^2.$$

Finally, using the fact that we have multiple independent approximate Gaussian variables, we take the critical value for the confidence interval from the Student's t -distribution with $M-1$ degrees of freedom. A $100(1-\alpha)\%$ confidence interval takes the form:

$$\left(\overline{\varepsilon} - \Delta\overline{\varepsilon}, \overline{\varepsilon} + \Delta\overline{\varepsilon}\right) = \left(\overline{\varepsilon} - t_{1-\alpha/2,M-1}\sqrt{\frac{\overline{\sigma}_{\varepsilon}^2}{M}}, \overline{\varepsilon} + t_{1-\alpha/2,M-1}\sqrt{\frac{\overline{\sigma}_{\varepsilon}^2}{M}}\right)$$

As expected, the width of the confidence interval decreases as the number of batches M increases. This is confirmed by the following, which considers three different numbers of batches for fixed step-size. The size of each batch is fixed at 100.

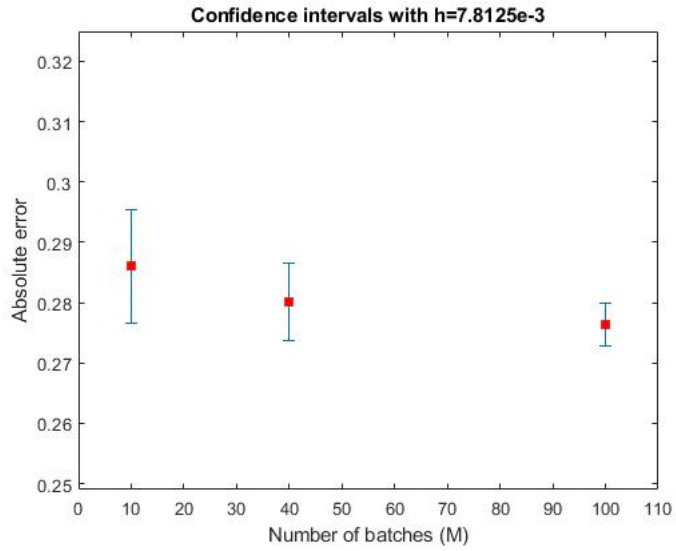

Figure 10: Confidence intervals for varying numbers of batches.

Next, fixing the number of batches and the size of each batch at 100, and varying the step-size, we have:

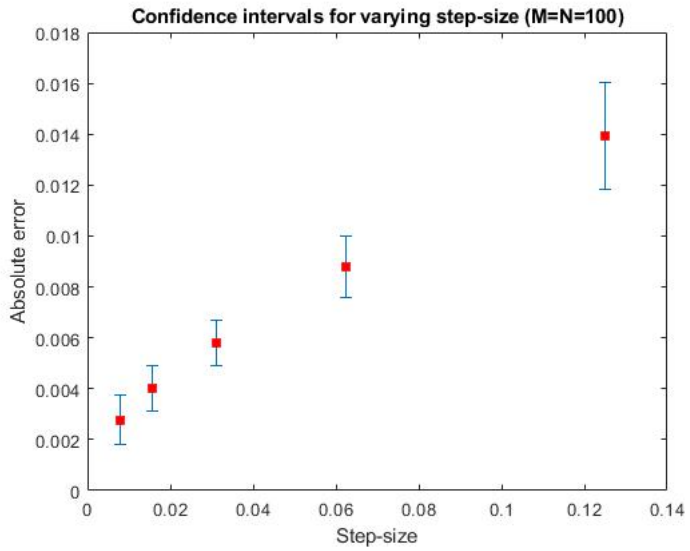

Figure 11: Confidence intervals for varying step-size.

The absolute error decreases with decreasing step-size, and so does the width of the confidence interval. This is due to the standard error being proportional to the variance - if more time-steps are used then the absolute difference between the trajectories at the end of the interval is smaller which translates into a smaller variance. To check that the error decreases at a rate that matches the strong order of accuracy, a log-log plot using the mean of the batch averages above is given below:

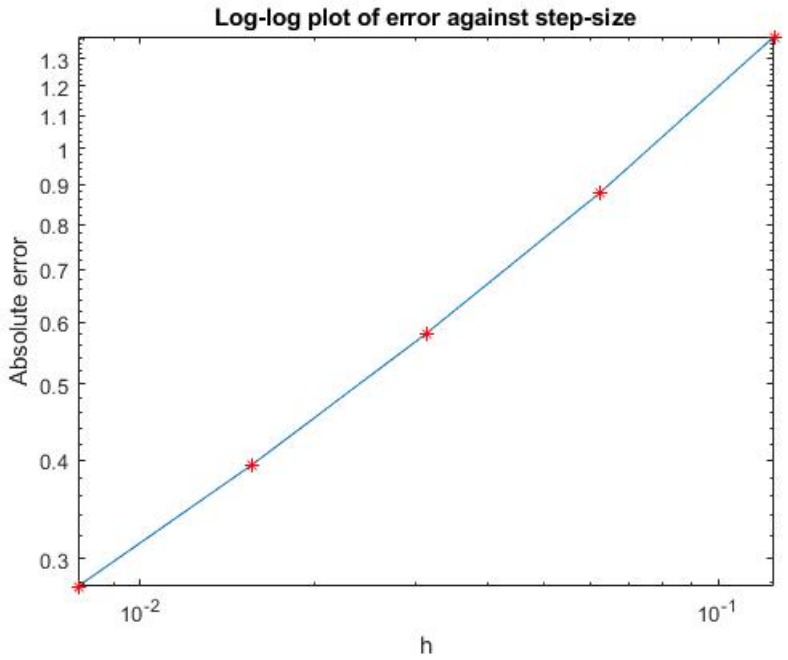

Figure 12: Log-log plot displaying rate of convergence.

The slope of the line between the two smallest step-sizes is equal to 0.5147, which is close to the theoretical value of 0.5 confirming the strong order of accuracy of the explicit Euler method.

If we required an approximation very close (in the mean-square sense) to the solution of an SDE, we would need to use a tiny step-size h due to the fact that we only have a strong order of accuracy of $\frac{1}{2}$. This would require a great number of time-steps which in turn increases the number of computations involved and consequently the time it takes for the approximation to be found. As a result, it is worth considering other methods with a stronger order of accuracy. The next section introduces a way to find new numerical schemes by considering a series of iterated integrals analogous to a Taylor expansion.

5.6 Ito-Taylor expansion

As mentioned previously, one incredibly useful tool that can be used to obtain numerical schemes for ODEs is the Taylor expansion [65]. The expansion being in terms of polynomials means any method derived is simple to implement, and after truncation the error term is known which gives us an idea of how much uncertainty is present in approximations.

Obtaining a more accurate method is as easy as including more terms in the expansion. It would be incredibly useful if we had an analogous formula for SDEs from which we could obtain new methods. It turns out that such an expansion does exist, and it can be derived in a similar way as this part will show. A very useful and clear reference for this section is [35].

Consider the following SDE with one Brownian motion (for simplicity):

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t), \text{ for } t \in [t_0, T].$$

For a smooth function f depending on X and t, the Ito formula gives:

$$df(X(t)) = \left(\frac{\partial f(X(t))}{\partial t} + a\frac{\partial f(X(t))}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 f(X(t))}{\partial x^2}\right)dt + b\frac{\partial f(X(t))}{\partial x}dW(t).$$

We can simplify this by introducing some operators, define

$$\begin{aligned}
\mathcal{L}^0 &= \frac{\partial}{\partial t} + a\frac{\partial}{\partial x} + \frac{1}{2}b^2\frac{\partial^2}{\partial x^2}, \\
\mathcal{L}^1 &= b\frac{\partial}{\partial x}, \\
\Rightarrow df(X(t)) &= \mathcal{L}^0 f(X(t))dt + \mathcal{L}^1 f(X(t))dW(t).
\end{aligned}$$

Incorporating the initial condition and switching to integral form gives:

$$f(X(t)) = f(X(t_0)) + \int_{t_0}^{t} \mathcal{L}^0 f(X(s))ds + \int_{t_0}^{t} \mathcal{L}^1 f(X(s))dW(s).$$

Setting $f(X(t)) = X(t)$ gives:

$$X(t) = X(t_0) + \int_{t_0}^{t} a(s, X(s))ds + \int_{t_0}^{t} b(s, X(s))dW(s).$$

If we instead set $f(X(t)) = a(t, X(t))$ and then $f(X(t)) = b(t, X(t))$, we get:

$$a(t, X(t)) = a(t_0, X(t_0)) + \int_{t_0}^{t} \mathcal{L}^0 a(s, X(s))ds + \int_{t_0}^{t} \mathcal{L}^1 a(s, X(s))dW(s),$$
$$b(t, X(t)) = b(t_0, X(t_0)) + \int_{t_0}^{t} \mathcal{L}^0 b(s, X(s))ds + \int_{t_0}^{t} \mathcal{L}^1 b(s, X(s))dW(s).$$

Notice that we now have expressions to substitute for the integrands of the integral form of $X(t)$ above. This leads us to:

$$\begin{aligned} X(t) = X(t_0) &+ \int_{t_0}^{t} \left(a(t_0, X(t_0)) + \int_{t_0}^{s_1} \mathcal{L}^0 a(s_2, X(s_2))ds_2 + \int_{t_0}^{s_1} \mathcal{L}^1 a(s_2, X(s_2))dW(s_2) \right) ds_1 \\ &+ \int_{t_0}^{t} \left(b(t_0, X(t_0)) + \int_{t_0}^{s_1} \mathcal{L}^0 b(s_2, X(s_2))ds_2 + \int_{t_0}^{s_1} \mathcal{L}^1 b(s_2, X(s_2))dW(s_2) \right) dW(s_1). \end{aligned}$$

Now, we can use the definitions of the operators to find the arguments of the integrals in the above (using the simplified notation $\frac{\partial c}{\partial x} = c'$, $\frac{\partial^2 c}{\partial x^2} = c''$ and $\frac{\partial c}{\partial t} = \dot{c}$):

$$\begin{aligned} \mathcal{L}^0 a &= \dot{a} + aa' + \frac{1}{2}b^2 a'', \\ \mathcal{L}^0 b &= \dot{b} + ab' + \frac{1}{2}b^2 b'', \\ \mathcal{L}^1 a &= ba', \\ \mathcal{L}^1 b &= bb'. \end{aligned}$$

Separating the double integral terms from the rest gives:

$$\begin{aligned} X(t) &= X(t_0) + a(t_0, X(t_0)) \int_{t_0}^{t} ds_1 + b(t_0, X(t_0)) \int_{t_0}^{t} dW(s_1) + R \\ &= X(t_0) + a(t_0, X(t_0))(t - t_0) + b(t_0, X(t_0))(W(t) - W(t_0)) + R. \end{aligned}$$

This should be recognisable to the reader as the explicit Euler method we derived at the start of this chapter plus a remainder. The remainder term R containing all the double

integrals takes the form:

$$R = \int_{t_0}^{t}\int_{t_0}^{s_1} \mathcal{L}^0 a(s_2, X(s_2))ds_2 ds_1 + \int_{t_0}^{t}\int_{t_0}^{s_1} \mathcal{L}^1 a(s_2, X(s_2))dW(s_2)ds_1 + \int_{t_0}^{t}\int_{t_0}^{s_1} \mathcal{L}^0 b(s_2, X(s_2))dW(s_1)ds_2 + \int_{t_0}^{t}\int_{t_0}^{s_1} \mathcal{L}^1 b(s_2, X(s_2))dW(s_2)dW(s_1).$$

The double integral which stands out is the final term, which is with respect to two Brownian motions. This is the lowest order term in Δt within R (using what we know of the quadratic variation being non-vanishing and equal to the width of the interval in time). Isolating this term and substituting the integral form of $b(X(t))$ derived above gives:

$$\int_{t_0}^{t}\int_{t_0}^{s_1} \mathcal{L}^1 b(s_2, X(s_2))dW(s_2)dW(s_1) = \int_{t_0}^{t}\int_{t_0}^{s_1} \Big(\mathcal{L}^1 b(t_0, X(t_0)) + \int_{t_0}^{s_2} \mathcal{L}^0\mathcal{L}^1 b(s_3, X(s_3))ds_3 + \int_{t_0}^{t} \mathcal{L}^1\mathcal{L}^1 b(s_3, X(s_3))dW(s_3)\Big)dW(s_2)dW(s_1).$$

Considering the first term on the right side (which is a constant) and using our knowledge of how the operator acts on b, $\mathcal{L}^1 b(t_0, X(t_0)) = b(t_0, X(t_0)b'(t_0, X(t_0)$, gives:

$$\int_{t_0}^{t}\int_{t_0}^{s_1} \mathcal{L}^1 b(s_2, X(s_2))dW(s_2)dW(s_1) = b(t_0, X(t_0)b'(t_0, X(t_0) \int_{t_0}^{t}\int_{t_0}^{s_1} dW(s_2)dW(s_1).$$

We can easily evaluate the double integral on the right in the following way:

$$\begin{aligned}\int_{t_0}^{t}\int_{t_0}^{s_1} dW(s_2)dW(s_1) &= \int_{t_0}^{t}(W(s_1) - W(t_0))dW(s_1)\\ &= \int_{t_0}^{t} W(s_1)dW(s_1) - \int_{t_0}^{t} W(t_0)dW(s_1)\\ &= \frac{1}{2}(W^2(t) - W^2(t_0) - (t - t_0)) - W(t_0)(W(t) - W(t_0))\\ &= \frac{1}{2}(W(t) - W(t_0))^2 - \frac{1}{2}(t - t_0).\end{aligned}$$

Firstly, we integrated with respect to $W(s_2)$. The first integral on line 2 was calculated earlier whilst the second is again just a constant. Rearranging gives the final value of the double

integral. Incorporating this in the expansion of $X(t)$ gives:

$$\begin{aligned} X(t) &= X(t_0) + a(t_0, X(t_0)) \int_{t_0}^{t} ds_1 + b(t_0, X(t_0)) \int_{t_0}^{t} dW(s_1) + R \\ &= X(t_0) + a(t_0, X(t_0))(t - t_0) + b(t_0, X(t_0))(W(t) - W(t_0)) \\ &+ \frac{1}{2} b(t_0, X(t_0) b'(t_0, X(t_0)) \Big((W(t) - W(t_0))^2 - (t - t_0) \Big) + \hat{R}. \end{aligned}$$

The new remainder term $\hat{R}$ is as before, but with the double Ito integral replaced with the triple integrals from the substitution. It is clear that we could continue to substitute the integral forms of the coefficients into parts of the remainder indefinitely and obtain further terms in the expansion, thereby obtaining more and more numerical schemes which should give us increasingly accurate approximations. Pausing at this point in the expansion, we can develop the Milstein scheme [46], first derived in 1974.

5.7 Milstein scheme

The expansion given at the end of the last section provides an intuition for how to develop a new method (for problems with one Brownian motion), building on top of the Euler scheme. Partitioning the interval of interest as before, an estimate for the solution at each time point is given by:

$$\begin{aligned} Y(t_{k+1}) =& Y(t_k) + a(t_k, Y(t_k))(t_{k+1} - t_k) + b(t_k, Y(t_k)(W(t_{k+1}) - W(t_k)) \\ &+ \frac{1}{2} b(t_k, Y(t_k) b'(t_k, Y(t_k)) \Big((W(t_{k+1}) - W(t_k))^2 - (t_{k+1} - t_k) \Big) \end{aligned}$$

The first line in the above is the Euler method. The term on the second line is new. Noticeably, it depends on the derivative of the coefficient $b(\cdot, X(\cdot))$ with respect to x. Evaluating this can be quite cumbersome for some problems, which is why it can be replaced using a finite difference term to get a basic Runge-Kutta method (see [31] for more details, and [55] for further reading). It can be proved that the Milstein method has first order mean-square order of accuracy. We will now confirm that this is the case by implementing the scheme on the geometric Brownian motion problem, which will allow us to directly compare

the approximations with those of the Euler method.

Recall that the coefficients of the GBM stochastic differential equation are given by $a(t, X(t)) = \mu X$ and $b(t, X(t)) = \sigma X$. The derivative of the second coefficient with respect to x is easily evaluated to obtain the constant σ. Therefore, the scheme (specifically for this problem) is given by:

$$\begin{aligned} Y(t_{k+1}) =& Y(t_k) + \mu Y(t_k)(t_{k+1} - t_k) + \sigma Y(t_k)(W(t_{k+1}) - W(t_k)) \\ &+ \frac{1}{2}\sigma^2 Y(t_k)\Big((W(t_{k+1}) - W(t_k))^2 - (t_{k+1} - t_k)\Big) \end{aligned}$$

To obtain these approximations, the code for the Euler method was adapted by adding this new term, and then the same calls were made from the driver script to produce the figure below. The GBM problem has the same parameters as before, with $(\mu, \sigma) = (0.08, 0.08)$ and initial value $S(0) = 160$ considered over the interval $t \in [0, 8]$.

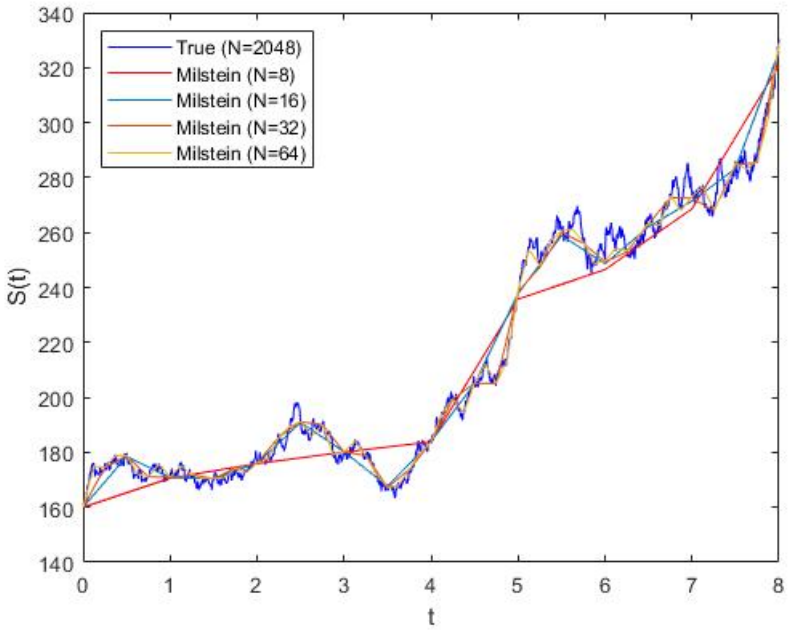

Figure 13: Milstein approximations for GBM problem.

The fact that this method has a theoretical mean-square order of accuracy of 1 implies that on average, the approximations should be closer to the true solution than the Euler method for a fixed step-size. Of course, this depends largely on the sample path so observing single trajectories will not make it clear in general which approximation is more accurate.

For that, we need to discuss the absolute error criterion introduced in section 5.4.

Once again, we use the unbiased estimator of the arithmetic mean of the absolute differences between trajectories of the approximation and true solution given by:

$$\varepsilon \approx \frac{1}{N} \sum_{k=1}^{N} |X_{T,k} - Y_{T,k}|.$$

We showed in section 5.5 that the variance of this estimator goes to zero in the limit that N approaches infinity. For each time-step in the table below, an estimate for the absolute error is calculated using 500 simulations of the geometric Brownian motion problem with parameters defined above. The ratio of the absolute error of successive step-sizes is also given.

h	Absolute error	Ratio
0.125	0.970	1
0.0625	0.541	1.794
0.03125	0.260	2.081
0.015625	0.126	2.066
0.0078125	0.061	2.055

As before, the decrease in the error with respect to the step-size is monotonic. Comparing with the table in section 5.5, the absolute error of the Milstein scheme is always below that of the Euler scheme. Furthermore, an absolute error below 0.1 is achieved for the smallest time-step here whereas the Euler method never achieves this. As for the ratio (the factor by which the absolute error decreases between each step-size), it very quickly settles around 2. In other words, when h is halved, the absolute error tends to approximately half as well, supporting the theory that the mean-square order of accuracy of this method is equal to 1.

We could also repeat an analysis of the uncertainty in these estimates using confidence intervals as we did before. The plot corresponding to that of figure 11 with the Milstein scheme is given below, where 100 batches of 100 simulations are used.

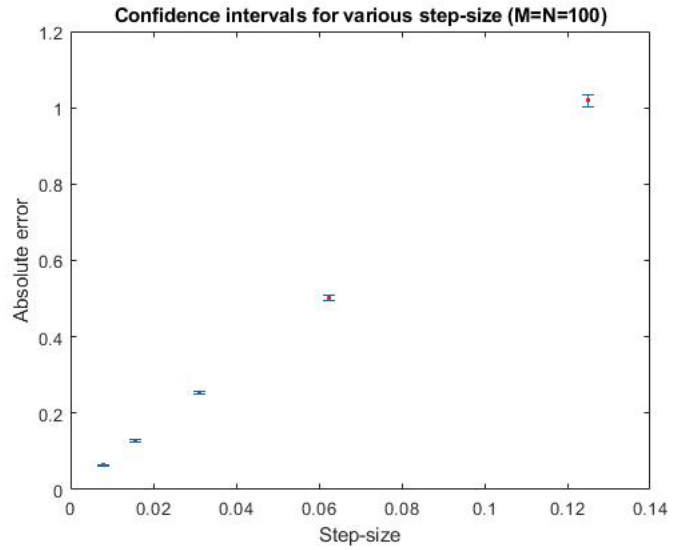

Figure 14: Milstein absolute errors (with confidence intervals) for GBM problem.

This plot reveals that the drop in absolute error is much more drastic. Finally, a log-plot of the absolute error against step-size is presented below.

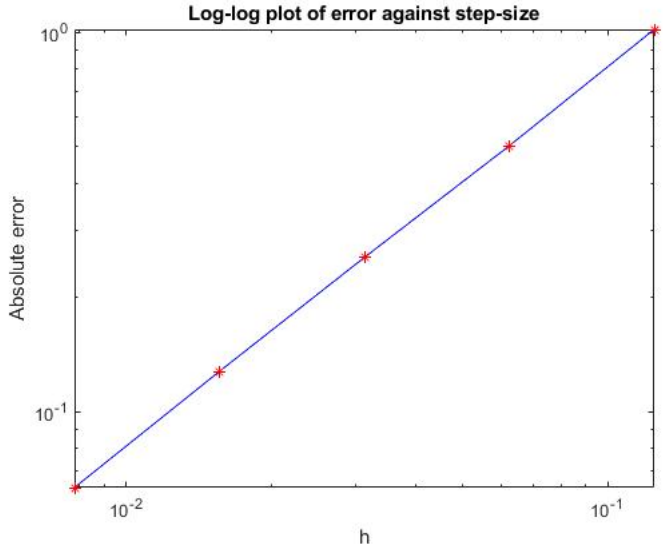

Figure 15: Log-log plot displaying rate of convergence of Milstein method.

The slope this time is about 1, confirming the theoretical order of accuracy of the Milstein method. For one-dimensional problems, this is a very useful method since it has a higher order of accuracy in comparison to the Euler method. However, note that when we derived the Ito-Taylor expansion in the previous part, it was specific to the case that one Brownian motion is involved. What happens when more than one Brownian motion is involved? Very quickly, we get terms in which one Brownian motion is integrated with respect to another

(see [67] for a more general Milstein method). Therefore, we face the additional challenge of having to consider how to simulate these integrals. It is for this reason that in practice, the Euler scheme is very popular, as it easily extends into multiple dimensions despite requiring a smaller step-size for accuracy.

5.8 Summary of chapter 5

In this chapter, we introduced an extension of the Euler scheme applicable to stochastic differential equations with multiple dimensions. We considered how we can implement the scheme, tested this implementation and the closeness of approximations to the true solution by conducting computational experiments using geometric Brownian motion as a case study. The mean-square order of accuracy for the Euler method was confirmed to be $p = \frac{1}{2}$.

When just a single Brownian motion is involved, it is simple to use the Ito formula to obtain representations of the coefficients of an SDE. Substituting these back into the integral form of the process repeatedly gives the Ito-Taylor expansion, which upon truncation motivates the definition of the higher order Milstein scheme, which has first mean-square order of accuracy as we confirmed.

Comparing the two schemes, we see that despite the fact that the Milstein scheme has a higher mean-square order of accuracy, it is not feasible to implement in higher dimensions. Besides this disadvantage, it also suffers in the sense that the derivative of the second coefficient needs to be computed, which can be computationally expensive. To get around this, finite difference methods can replace the derivative to form a Runge-Kutta method. In contrast, the Euler scheme is simple to apply in the higher dimensional case and is less computationally expensive.

When we are not interested in the qualitative behaviour of a model (for example, when we are interested in the moments of a model), we can derive much simpler models to implement. For these purposes, we can consider approximation in the weak sense.

6 Stochastic Numerics: Weak methods

The methods introduced in the last chapter can be used to observe trajectories near to (in the mean-square sense) the process of interest. This is useful when one is interested in stock prices, as it gives an idea of the behaviour over an arbitrary interval. In contrast, if we were considering a problem in molecular dynamics, where billions of particles are involved, individual trajectories are not important - it would be far more helpful to know say the average position of the particles after a certain length of time and the variance around that average. These quantities are moments of the underlying process, so if we had methods to approximate these, that would be extremely useful.

Methods to do this exist (see [43, 31]) - and the problem of matching moments is fundamentally simpler than matching trajectories. These methods do not require the modelling of complicated random variables which made the implementation of previous methods we've seen difficult. This chapter introduces some of the theory plus examples of schemes that approximate in the weak-sense.

6.1 Approximation in the weak-sense

Consider the multi-dimensional SDE given (as before) by

$$dX = b(t, X)dt + \sum_{r=1}^{q} \sigma_r(t, X)dW_r(t),$$

and assume that the coefficients are smooth and satisfy the global Lipschitz condition with respect to x (introduced in section 4.3).

Definition 6.1. If an approximation Y is such that

$$|E(f(Y(T)) - E(f(X(T))| \leq Kh^p,$$

for f a class of functions with polynomial growth at infinity, then we say the weak order of accuracy of the approximation Y is p. Here, the constant K depends on $b(t, X)$, $\sigma_r(t, X)$, the function f and T [67].

It turns out that despite the higher order of accuracy in the mean-square sense of the Milstein scheme, the weak order of accuracy of both the Euler and Milstein schemes is 1 [43]. This is interesting, perhaps you would expect an increase in the order of accuracy in the mean-square sense to imply an increase in the weak order of accuracy, but this is not the case. In fact, there is a method even simpler than Euler's which also has first weak order of accuracy, in which the Brownian increments are replaced by another set of random variables which possess similar moments. What kind of random variable might be appropriate? We look at a particular one in the next part.

6.2 Symmetric random walk

Consider the following set of random variables J_i, with distribution:

$$P(J_i = 1) = \frac{1}{2},$$
$$P(J_i = -1) = \frac{1}{2},$$

and the sum of these variables given by $S(n) = \sum_{i=1}^{n} J_i$. Assuming these J_i are independent, the process $S(n)$ for $n = 1, 2, ...$ is a symmetric random walk (see [71] for further reading, [66] is the main reference for this part). Using the definition of discrete expectation from chapter 2, we have

$$E(J_i) = 1 * \frac{1}{2} + (-1) * \frac{1}{2} = 0,$$
$$E(J_i^2) = 1^2 * \frac{1}{2} + (-1)^2 * \frac{1}{2} = 1,$$
$$Var(J_i) = E(J_i^2) - (E(J_i))^2 = 1,$$

so that

$$E(S(n)) = \sum_{i=1}^{n} E(J_i) = 0$$
$$Var(S(n)) = \sum_{i=1}^{n} Var(J_i) = n.$$

Since we assume that the J_i are independent, the increments given by

$$\Delta_n S = S(n) - S(n-1) = J_n,$$

are also independent for $n = 1, 2, ...$ It is easy to see from this that the quadratic variation is equal to n using the fact that for all i, $J_i^2 = 1$.

Now, divide the interval of interest $[0, T]$ into N equal parts so that $\Delta t = \frac{T}{N}$ and $t_n = n\Delta t$. Let η be a scaled random walk - a process such that $\eta(0) = 0$ and $\eta(t_n) = \sqrt{\Delta t}\sum_{i=1}^{n} J_i$, $n = 1, 2, ..., N$ with the J_is as above. The expectation is still zero, whilst the increments have been scaled so that $\Delta_n \eta = \eta(t_n) - \eta(t_{n-1}) = \sqrt{\Delta t} J_n$. This leads to a quadratic variation of $n\Delta t = t_n$ and a variance (squaring the scaling factor and summing) of t_n. Finally, if there are two time points $t_l < t_k$, then:

$$\begin{aligned} Var(\eta(t_k) - \eta(t_l)) &= Var(\sqrt{\Delta t}\sum_{i=1}^{k} J_i - \sqrt{\Delta t}\sum_{i=1}^{l} J_i) \\ &= Var(\sqrt{\Delta t}\sum_{i=l+1}^{k} J_i) \\ &= \Delta t Var(\sum_{i=l+1}^{k} J_i) \\ &= (k-l)\Delta t = t_k - t_l. \end{aligned}$$

In this way, the variance is proportional to time. This should remind the attentive reader of Brownian motion. However, in this section we have considered a process which is clearly discrete, whilst Brownian motion is a continuous process. Luckily, we can interpolate linearly to make the above continuous. Consider straight lines connecting $\eta(t_{n-1})$ and $\eta(t_n)$.

The gradient of this line is given by:

$$m = \frac{\eta(t_n) - \eta(t_{n-1})}{t_n - t_{n-1}} = \frac{J_n}{\sqrt{\Delta t}},$$

whilst the intercept can be found by setting $t = t_n$ and rearranging $\eta(t_n) = mt_n + c$,

$$c = \eta(t_n) - \frac{J_n}{\sqrt{\Delta t}} t_n.$$

Substituting the above gives the following straight line:

$$\eta(t) = \eta(t_n) - \frac{t_n - t}{\sqrt{\Delta t}} J_n.$$

Now, what is very interesting is what happens in the limit that this $\Delta t \to 0$. Fixing t, there must be a k such that $t_{k-1} < t < t_k$. Then the straight line between these two points can be rewritten as:

$$\begin{aligned} \eta(t) &= \sqrt{\Delta t} \sum_{i=1}^{k} J_i - \frac{t_k - t}{\sqrt{\Delta t}} J_k \\ &= \frac{\sqrt{t}}{\sqrt{t}} \frac{\sqrt{t_k}}{\sqrt{k}} \sum_{i=1}^{k} J_i - \frac{t_k - t}{\sqrt{\Delta t}} J_k, \end{aligned}$$

using the fact that $t_k = k\Delta t$. Then, in the limit that $\Delta t \to 0$, we have $\frac{t_k}{t} \to 1$ (since the interval t sits in becomes ever narrower). Furthermore, note that $t_k - t < \Delta t$ since by definition, $t_{k-1} = t_k - \Delta t$. This gives us that:

$$0 \leq \frac{t_k - t}{\sqrt{\Delta t}} \leq \frac{\Delta t}{\sqrt{\Delta t}} = \sqrt{\Delta t} \to 0,$$

so $\eta(t) \to \frac{\sqrt{t}}{\sqrt{k}} \sum_{i=1}^{k} J_i$. Using the Central Limit theorem, $\sum_{i=1}^{k} J_i \to \mathcal{N}(0, k)$ as $k \to \infty$ (which occurs in the limit $\Delta t \to 0$). Combining all of these results, we have $\eta(t) \to \mathcal{N}(0, t)$ as $\Delta t \to 0$. Furthermore, it can be shown that the increment $\eta(t) - \eta(s) \to \mathcal{N}(0, t - s)$ in the limit (see [66]). This reveals that the process under the limit has the same properties as Brownian motion, and motivates why we might use a random walk as a substitute.

This is illustrated in the following figure. It is clear that as more points are considered, the behaviour of the trajectory gets closer to that of Brownian motion as expected given the results in this section.

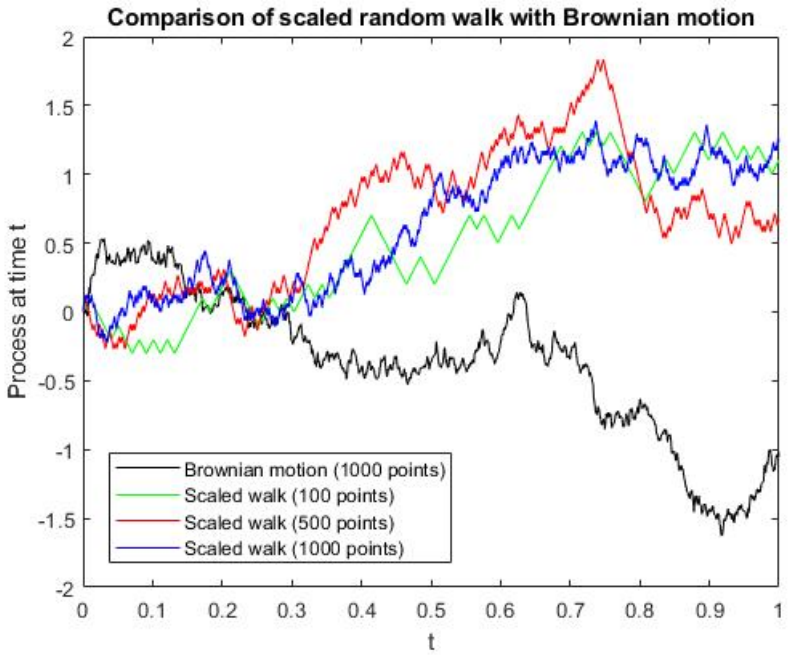

Figure 16: Trajectories of scaled random walks with Brownian motion.

6.3 The Weak Euler scheme

Replacing the Brownian increment in the Euler scheme with a random walk, the weak Euler scheme for the multidimensional SDE takes the form:

$$X_{k+1} = X_k + b(t_k, X_k)h + \sqrt{h}\sum_{r=1}^{q} \sigma_{rk} J_{rk},$$

where $J_{rk},\ r = 1, ..., q,\ k = 0, ..., N-1$ are independent and identically distributed with $P(J_{rk} = \pm 1) = \frac{1}{2}$. As mentioned in section 6.1, this scheme has first weak order of accuracy. So far we have simply stated that the three methods discussed in this report have first weak order of accuracy. We should confirm this through an example. In the following section, we consider the continuous Black-Scholes formula for the pricing of European options, derived and solved in section 4.5. This should motivate to the reader as to how methods such as those discussed in this report are used in practice.

6.4 A numerical example: Continuous Black-Scholes formula

The continuous Black-Scholes formula [7], which we derived and solved in example 4.7, takes the following form:

$$\frac{\partial V}{\partial \tau} + \frac{\sigma^2 S^2}{2} + rS\frac{\partial V}{\partial x} - rV = 0 \text{ where } t \leq \tau \leq T.$$

Here, V is the European option price, S is the stock price, σ is the stock volatility, r is the interest rate of a bank account and T is the maturity time of the option. In example 4.7, we found that the probabilistic representation of the solution subject to a terminal condition is:

$$V(t, S(t)) = e^{-r(T-t)}E(f(S(T))) \Rightarrow V(0, S(0)) = e^{-rT}E(f(S(T))).$$

We did not discuss how to use this solution to calculate the option price. Given that f is a payoff function, this is not entirely obvious. First of all, consider the argument of the payoff function. Since Black-Scholes assumes that the underlying stock price follows GBM, we know what $S(T)$ is as we solved GBM in example 4.4:

$$S(T) = S(0)e^{(r-\frac{\sigma^2}{2})T+\sigma W(T)} = S(0)e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}Z},$$

where for the second equality we replaced $W(T)$ by $\sqrt{T}Z$ (Z being a standard normal random variable) using the properties of Brownian motion. Suppose we are interested in the price of a call option, which has payoff function $f(S) = max\{S-K, 0\} = (S-K)_+$ where K is the strike price. Substituting this into the solution at time 0 gives:

$$\begin{aligned} V(0, S(0)) &= e^{-rT}E((S(0)e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}Z} - K)_+ \\ &= E((S(0)e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}Z} - Ke^{-rT})_+) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(S(0)e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}z} - Ke^{-rT})_+ e^{-\frac{z^2}{2}}dz. \end{aligned}$$

The multiplier e^{-rT} is brought inside the expectation, then we recognise that the only random term inside the expectation is the standard normal random variable Z so we use the definition of continuous expectation to rewrite as an integral. From here, we need to identify for which values of z the argument of the integral is non-zero, which should restrict the bounds over which we are integrating.

$$\begin{aligned}
&S(0)e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}z} > Ke^{-rT} \\
&e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}z} > \frac{K}{S(0)}e^{-rT} \\
&-\frac{\sigma^2 T}{2} + \sigma\sqrt{T}z > log(\frac{K}{S(0)}) - rT \\
&z > \frac{log(\frac{K}{S(0)}) - rT + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}} = -d_2.
\end{aligned}$$

From the second line, the logarithm is taken then rearranged. The constant $d_2 = \frac{log(\frac{S(0)}{K})+rT-\frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}$ is defined as a short-hand. Introducing this into the integral gives:

$$\begin{aligned}
V(0,S(0)) &= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty}((S(0)e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}z} - Ke^{-rT})e^{-\frac{z^2}{2}}dz \\
&= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty}S(0)e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}z-\frac{1}{2}z^2}dz - \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty}Ke^{-rT-\frac{1}{2}z^2}dz.
\end{aligned}$$

The first integral can be simplified significantly by recognising that $-\frac{\sigma^2 T}{2} + \sigma\sqrt{T}z - \frac{1}{2}z^2 = -\frac{1}{2}(z-\sigma\sqrt{T})^2$. Using the substitution $y = z - \sigma\sqrt{T}$ and introducing a second short-hand $d_1 = d_2 - \sigma\sqrt{T}$, we get:

$$\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty}S(0)e^{-\frac{\sigma^2 T}{2}+\sigma\sqrt{T}z-\frac{1}{2}z^2}dz = \frac{1}{\sqrt{2\pi}}\int_{-d_1}^{\infty}S(0)e^{-\frac{y^2}{2}}dy.$$

The term on the right can now be recognised as equal to $S(0)(1-\Phi(-d_1))$ where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. By the symmetry of the normal distribution, $1-\Phi(-d_1) = \Phi(d_1)$ so the first integral has value $S(0)\Phi(d_1)$. As for the second integral, pull Ke^{-rT} outside and recognise the cumulative distribution function in the same way, so the second integral is equal to $Ke^{-rT}\Phi(d_2)$. Combining all of this, we have

that the price of a European call option at time 0 (see [66] for more details) is:

$$V(0, S(0)) = S(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where d_1 and d_2 were defined earlier. Now that we know the exact solution and a closed form way of calculating it, we can compare it with approximations using the three schemes we have considered. To do this, we need to define the problem specifically.

Consider a European call option with strike price $K = 100$ and maturity time $T = 3$. Assume that the stock has volatility $\sigma = 0.05$ and the writer has access to a bank account with a continuously compounding interest rate of $r = 0.5$. Finally, assume the initial price of the stock is $S(0) = 85$. Then, using the explicit formula given above, the exact price of this option can be calculated.

$$\begin{aligned}
d_2 &= \frac{log(\frac{85}{100}) - 0.5 * 3 + 0.5 * 0.05^2 * 3}{0.05 * \sqrt{3}} \approx 15.401 \\
d_1 &= d_2 + 0.05 * \sqrt{3} \approx 15.487 \\
V(0, S(0)) &= 85 * \Phi(15.487) - 100 * e^{-0.5*3}\Phi(15.401) = 62.687.
\end{aligned}$$

If we were not aware of how to calculate this option price, how might we approximate it using the numerical schemes introduced over the last two chapters? Stepping back a moment, the probabilistic representation of the solution is given by:

$$V(0, S(0)) = e^{-rT}E(f(S(T)).$$

If we defined an estimator for the term within the expectation, then using the Central Limit theorem, we could sample many times before finding the mean of the estimates. Well, we know that the payoff function for a call option is given by $f(S) = (S - K)_+$ and that the underlying stock price S is modelled by geometric Brownian motion, so if we approximate the stock price at the final time using one of our numerical schemes then apply the payoff

function, we will obtain an estimate. Mathematically, this can be written as:

$$V(0,S(0)) = e^{-rT}E(f(S(T)) \approx e^{-rT}E(f(\overline{S}(T)) \approx e^{-rT}\left(\frac{1}{M}\sum_{i=1}^{M} f(\overline{S}(T))\right).$$

Approximating in this way introduces two forms of uncertainty. First, we approximate the stock price process via one of our numerical schemes, this is the numerical integration error. Then, forming the arithmetic mean of these values introduces the statistical, or Monte Carlo error (see [54] for a comparison between this method and the finite-difference scheme). To see the effect of choosing any one approximation method over another (i.e. analyse the numerical integration error), we need to try and make the Monte Carlo error much smaller. As a result, for the table below a million samples of the estimator are taken for each method. The error refers to the absolute difference between the approximation and the true value:

h	Monte-Carlo (99.7% CI)	Error (Euler)	Error (Milstein)	Error (Weak Euler)
0.25	± 0.018023	7.0467	7.0584	7.0537
0.2	± 0.018739	5.7807	5.7772	5.7732
0.1	± 0.020334	3.0422	3.0242	3.0331
0.05	± 0.021202	1.5563	1.5541	1.549
0.025	± 0.021641	0.79121	0.78223	0.78429

The Monte-Carlo column gives us the amount by which the error in the final column varies around the quantity $e^{-rT}E(f(\overline{S}(T))$ where $\overline{S}(T)$ is estimated using the weak-Euler method. It is calculated using the confidence interval given on page 48 of [67], which also links the two forms of error given earlier to the bias and variance of the estimator $e^{-rT}\left(\frac{1}{M}\sum_{i=1}^{M} f(\overline{S}(T))\right)$. Evidently, this counts for very little of the overall error with this many samples (the reader is invited to use *BlackScholes.m* to check that this is also true for the Euler and Milstein methods). Hence, the difference should be almost entirely due to numerical integration. This column is also given in tables we consider later, and can be interpreted in the same way.

It is clear that all three methods perform similarly in terms of accuracy. However, in terms of time cost, the weak Euler scheme can be implemented significantly quicker. The

results in the above table were obtained after just 19.163 seconds. On the other hand, the Euler scheme took 3629.740 seconds (just over an hour) whilst the Milstein scheme took 3572.41 seconds (just under an hour). Note that for the Milstein scheme, the differentiation of the second coefficient was done by hand before being written into code, and for harder problems it is likely that it would take longer than the Euler scheme. This is an enormous difference in computation time, and for this reason the weak Euler scheme is the most suitable for this problem.

Also worth noting is how the absolute error decays with decreasing step-size. For smaller step-sizes, the ratio between the errors of successive estimates is near to 2. This is in line with the theory which tells us that these methods all possess first weak order of accuracy, and is confirmed through the figure below. Note that all three lines coincide, and comparing magnitudes of the errors with the Monte-Carlo results from the table, the confidence interval around each point is negligible.

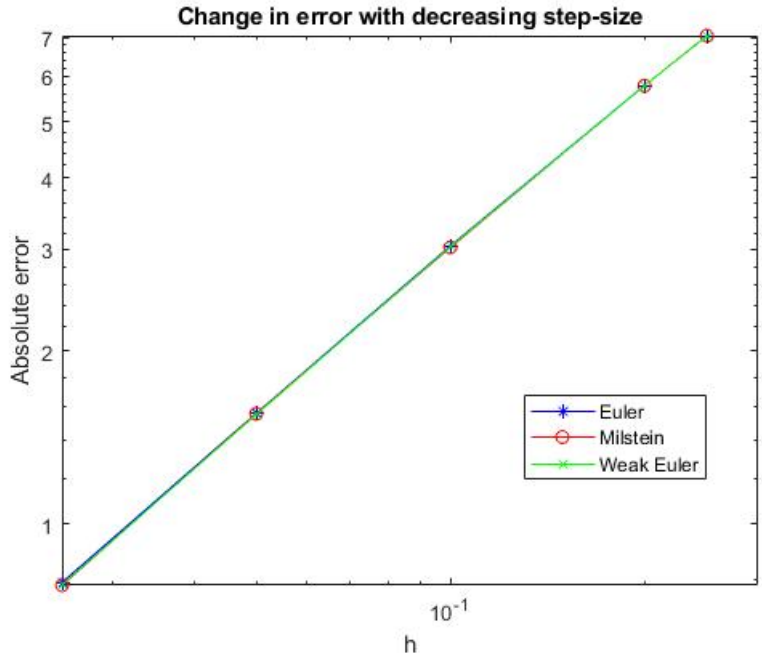

Figure 17: Plot of errors with decreasing step-size, confirming weak order of accuracy.

This example has confirmed the result that the three methods covered so far all have first weak order of accuracy. It would be very useful if there was a general convergence theorem, similar to the theorem on mean-square convergence in section 5.4, allowing one to find the weak order of accuracy of a scheme by checking if a scheme satisfies some

assumptions. It turns out there is such a theorem, discussed in the next part.

6.5 The general convergence theorem

Let us define the general approximation through a weak scheme in order to introduce the convergence theorem. The main reference for this section is [67].

For the multidimensional SDE given at the start of section 6.1, consider $Y_{t,x}(t+h) = x + A(t,x,h;\epsilon)$. In this notation, $Y_{t,x}(t+h)$ is the approximation at time $t+h$ which satisfies $Y(t) = x$, and since we are considering the multidimensional case the approximation Y as well as x are vectors, and the function A is a vector function with dimension matching the problem in 6.1. On top of this, ϵ is a random vector which has moments of a sufficiently high order. Then, partitioning the interval $[t_0, T]$ into N equal parts $t_0 < t_1 = t_0 + h < ... < t_N = T = t_0 + Nh$ and using the above approximation allows us to construct a sequence to approximate the process X:

$$Y_0 = X(t_0)$$
$$Y_{k+1} = Y_k + A(t, Y_k, h; \epsilon_k), k = 0, 1, ..., N-1,$$

where ϵ_k is independent of all the previous approximations $Y_0, Y_1, ..., Y_{k-1}$ as well as the previous vectors $\epsilon_0, \epsilon_1, ..., \epsilon_{k-1}$. Now, let the increment vectors be denoted by $\Delta = X_{t,x}(t+h) - x$ (true process) and $\bar{\Delta} = Y_{t,x}(t+h) - x$ (approximation). Finally, let $X(t) = X_{t_0,X_0}(t)$ be the solution of the general SDE (with initial condition $X(t_0) = X_0$) and $Y_{t_0,X_0}(t_k) = Y_k$ be the approximation satisfying the aforementioned initial condition. Being more specific on the class of functions f relevant within the context of weak approximation (see definition 6.1), we introduce the following definition [47]:

Definition 6.2. We say that a function $f(x)$ belongs to the class **F**, written $f \in \mathbf{F}$, if we can find constants $K > 0$, $\kappa > 0$ such that for all $x \in \mathbb{R}^d$, the following inequality holds:

$$|f(x)| \leq K(1 + |x|^{\kappa})$$

If a function $f(s,x)$ depends not only on $x \in \mathbb{R}^d$ but also a parameter $s \in S$, then we say that $f(s,x)$ belongs to **F** (with respect to x) if the inequality above holds for all values of s.

We are now in a position to introduce the general convergence theorem, proved in [43].

Theorem 6.1. Assume that:

- The coefficients in the SDE of interest satisfy the Lipschitz condition, and along with their partial derivatives in respect to x up to order $2p+2$ belong to the class **F**.
- The approximation given in this section is such that:

$$\left| E\Big(\prod_{j=1}^{s} \Delta^{i_j} - \prod_{j=1}^{s} \overline{\Delta}^{i_j} \Big) \right| \leq K(x) h^{p+1} \ \ s = 1, ..., 2p+1, K(x) \in \mathbf{F}$$

$$E\Big(\prod_{j=1}^{2p+2} \left| \overline{\Delta}^{i_j} \right| \Big) \leq K(x) h^{p+1}, \ \ K(x) \in \mathbf{F}.$$

- the function $f(x)$ along with its partial derivatives up to and including order $2p+2$ belong to **F**.
- for a sufficiently large m, the expectations $E|\overline{X}_k|^{2m}$ exist and are uniformly bounded with respect to N and $k = 0, 1, ..., N$.

Then, for all N and all $k = 0, 1, ..., N$ the following inequality holds:

$$|E(f(X_{t_0,X_0}(t_k)) - E(f(Y_{t_0,X_0}(t_k))| \leq K h^p,$$

Stepping back a moment, this theorem tells us that subject to some (quite restrictive) conditions, a method is guaranteed to have an order of accuracy p in the sense of weak approximation. This is useful, but the assumptions are strong. In many applications these conditions do not hold, for example in financial engineering where payoff functions are used (which are sometimes not differentiable at particular points). Study into methods of approximation when these restrictions are loosened is an active area of research, see [42] for further reading.

Theorem 6.1 may allow us to check the order of accuracy of a method, but it does not give us much indication of how to actually construct new methods of higher orders. It is possible to construct new, more accurate methods from estimates we obtain from lower order methods using a technique analogous to Richardson extrapolation (see [9] for more details). In the final part of this chapter, we introduce this technique and use it to form second and third weak order methods.

6.6 Talay-Tubaro extrapolation method

One major advantage of the Euler/weak Euler schemes is how simple they are to implement computationally. The main issue with these schemes is that they are only first-order accurate. As a result, it would be incredibly useful if we could construct an improved estimate, based on those of these schemes, which is even more accurate. Using an expansion of the global error, we are able to do this. An incredibly useful reference for this part is [32], and an original paper on the topic is [64]. In this part, we will stick with approximations of the weak Euler scheme, due to how much quicker it is to obtain approximations. The main theorem that we need is as follows:

Theorem 6.2. The global error of a weak approximation of order p has the following expansion, where $C_0, ..., C_n$ are independent of h and n is an integer, $n \geq 0$:

$$R = C_0 h^p + ... + C_n h^{p+n} + O(h^{p+n+1}).$$

A corollary of this theorem is that the weak Euler scheme, with weak order of accuracy $p = 1$, has global error given by:

$$R = C_0 h + ... + C_n h^{1+n} + O(h^{2+n}).$$

Now, suppose we are interested in simulating $u(t_0, x) = E(f(X_{t_0,x}(T)))$, the Feynman-Kac formula introduced in example 4.8. Then the global error R is equivalent to

$$E(f(X_{t_0,x}(T))) - E(f(Y^h_{t_0,x}(T))) = u - \bar{u}^h,$$

where Y represents the approximation and we abbreviate on the right, with $\bar{u}^h$ being the estimate formed through use of the step-size h. Again, we can form this estimate by using the numerical scheme to estimate many values of $X_{t_0,x}(T)$ before applying f and averaging. Form two estimates with distinct time steps $h_1 = h$ and $h_2 = \alpha h$ where $\alpha > 0$, $\alpha \neq 1$ denoted:

$$\bar{u}^{h_1} = E(f(Y^{h_1}_{t_0,x}(T))),$$

$$\bar{u}^{h_2} = E(f(Y^{h_2}_{t_0,x}(T))).$$

Using theorem 6.2, we can expand u in terms of these estimates. Cutting the expansion off at first order gives:

$$u = \bar{u}^{h_1} + Ch_1 + O(h^2), \tag{6.1}$$

$$u = \bar{u}^{h_2} + Ch_2 + O(h^2). \tag{6.2}$$

The idea now is to solve for the constant C in terms of the two estimates and their associated step-sizes, then substitute into one of these equations to obtain an improved estimate with second weak-order accuracy. Subtracting (6.2) from (6.1) and rearranging gives:

$$\begin{aligned} 0 &= \bar{u}^{h_1} - \bar{u}^{h_2} + C(h_1 - h_2) + O(h^2) \\ C(h_2 - h_1) &= \bar{u}^{h_1} - \bar{u}^{h_2} + O(h^2) \\ \Rightarrow C &= \frac{\bar{u}^{h_1} - \bar{u}^{h_2}}{h_2 - h_1} + O(h). \end{aligned}$$

The $O(h)$ in the final line comes from the fact that $h_2 - h_1$ is proportional to h. Inserting this into (6.1), we have the improved second-order approximation:

$$\bar{u}_{imp} = \bar{u}^{h_1} + \frac{\bar{u}^{h_1} - \bar{u}^{h_2}}{h_2 - h_1} h_1 = \bar{u}^{h_1} \frac{h_2}{h_2 - h_1} - \bar{u}^{h_2} \frac{h_1}{h_2 - h_1}.$$

It should be clear to the reader that:

$$u = \bar{u}_{imp} + O(h^2).$$

To test this method against the previous methods, we apply it to the same problem as before (see section 6.4). We set $\alpha = \frac{1}{2}$, and form the expectation using $M = 2e7$ (2 billion) samples. We use this many samples so that the error we see is overwhelmingly due to numerical integration rather than the Monte Carlo error. The results obtained for this method are summarised below:

h	Monte-Carlo (99.7% CI)	Error (Talay-Tubaro order 2)	Ratio (Talay-Tubaro order 2)
1	± 0.0071751	4.3131	1
0.5	± 0.0087334	1.4572	2.96
0.25	± 0.097715	0.43218	3.37
0.125	± 0.010379	0.11979	3.61
0.0625	± 0.010709	0.02958	4.05

Note how much more accurate the improved estimates are in comparison with the previous three methods. For example, the improved estimate with $h = 1$ is more accurate than the estimate of the previous three methods with $h = 0.2$. Furthermore, the ratio between successive estimates gets larger as the step-size decreases, reaching the theoretical ratio of 4 at the final step-size. However, caution should be taken with regards to this observation as the Monte-Carlo error is the same magnitude as the combined error for this step-size. Taking even more samples can reduce the statistical error.

We can go even further and construct a third-order method based on estimates from the weak Euler scheme. To this end, consider three estimates formed (for simplicity in later algebra) from the step-sizes $h_1 = h, h_2 = \frac{h}{2}$ and $h_3 = \frac{h}{4}$. Then, as before, using theorem 6.2 we have the following expansions:

$$u = \bar{u}^h + C_0 h + C_1 h^2 + O(h^3), \tag{A1}$$

$$u = \bar{u}^{\frac{h}{2}} + C_0 \frac{h}{2} + C_1 \frac{h^2}{4} + O(h^3), \tag{A2}$$

$$u = \bar{u}^{\frac{h}{4}} + C_0 \frac{h}{4} + C_1 \frac{h^2}{16} + O(h^3). \tag{A3}$$

Subtracting $(A2)$ from $(A1)$, then $(A3)$ from $(A1)$ and finally $(A3)$ from $(A2)$, we obtain a

second system of equations:

$$0 = \bar{u}^h - \bar{u}^{\frac{h}{2}} + C_0\frac{h}{2} + C_1\frac{3h^2}{4} + O(h^3), \tag{B1}$$

$$0 = \bar{u}^h - \bar{u}^{\frac{h}{4}} + C_0\frac{3h}{4} + C_1\frac{15h^2}{16} + O(h^3), \tag{B2}$$

$$0 = \bar{u}^{\frac{h}{2}} - \bar{u}^{\frac{h}{4}} + C_0\frac{h}{4} + C_1\frac{3h^2}{16} + O(h^3). \tag{B3}$$

Now, we can cancel out the C_0 terms by forming the new equations below. The first is formed from computing $3(B2) - 3(B1)$ whilst the second is formed through $(B1) + (B3)$:

$$0 = 3(\bar{u}^{\frac{h}{2}} - \bar{u}^{\frac{h}{4}}) + C_0\frac{3h}{4} + C_1\frac{9h^2}{16} + O(h^3), \tag{C1}$$

$$0 = \bar{u}^h - \bar{u}^{\frac{h}{4}} + C_0\frac{3h}{4} + C_1\frac{15h^2}{16} + O(h^3). \tag{C2}$$

Taking the difference between $(C2)$ and $(C1)$ gives:

$$\begin{aligned} 0 &= 3(\bar{u}^{\frac{h}{2}} - \bar{u}^{\frac{h}{4}}) - \bar{u}^h + \bar{u}^{\frac{h}{4}} - C_1\frac{3h^2}{8} + O(h^3) \\ \Rightarrow C_1 &= \frac{8}{3h^2}(3\bar{u}^{\frac{h}{2}} - 2\bar{u}^{\frac{h}{4}} - \bar{u}^h) + O(h). \end{aligned}$$

Reinserting this constant into $4(B2) - 4(B1)$ gives:

$$\begin{aligned} 0 &= 4(\bar{u}^{\frac{h}{2}} - \bar{u}^{\frac{h}{4}}) + C_0 h + C_1\frac{3h^2}{4} + O(h^3) \\ -C_0 h &= 4(\bar{u}^{\frac{h}{2}} - \bar{u}^{\frac{h}{4}}) + \frac{3}{4}(\frac{8}{3h^2}(3\bar{u}^{\frac{h}{2}} - 2\bar{u}^{\frac{h}{4}} - \bar{u}^h) + O(h^3) \\ \Rightarrow C_0 &= \frac{1}{h}(2\bar{u}^h + 8\bar{u}^{\frac{h}{4}} - 10\bar{u}^{\frac{h}{2}}) + O(h^2). \end{aligned}$$

Finally, combining the above, the third-order weak method is:

$$\begin{aligned} \hat{u}_{imp} =& \bar{u}^h + C_0 h + C_1 h^2 + O(h^3) \\ =& \frac{1}{3}(\bar{u}^h - 6\bar{u}^{\frac{h}{2}} + 8\bar{u}^{\frac{h}{4}}) + O(h^3), \end{aligned}$$

where for the second line we simply substituted the constants C_0 and C_1 and collected

coefficients.

We test this method numerically in the same way as we did the second-order method, but now we expect that when the step-size is halved the error should go down by a factor of around 8. The table below presents the results for 3 values of h and 5 billion samples:

h	Monte-Carlo (99.7% CI)	Error (Talay-Tubaro order 3)	Ratio (Talay-Tubaro order 3)
1.5	± 0.0071895	1.2306	1
0.75	± 0.008535	0.24929	4.936
0.375	± 0.0093987	0.029212	8.534

Using this method, we can use a larger step-size even compared to the second-order method and obtain a more accurate result. The ratio between the bottom two estimates exceeds 8, but again we must be wary of the fact that the statistical error is almost the same magnitude as the combined error for the final step-size.

6.7 Summary of chapter 6

In this chapter, we have introduced and motivated the concept of the weak approximation, which estimate the moments of a process instead of trajectories.

We found a method far simpler to compute than the Euler and Milstein methods from chapter 5, namely the weak Euler scheme in which Brownian motion is replaced with a random walk process. All three of these schemes possess the same weak order of accuracy, $p = 1$, however we found that when we wish to estimate the price of a European option, the weak Euler scheme took much less time compared to the other two methods. On top of this, we discussed the general convergence theorem, which allows us to check the order of accuracy of a given scheme.

We finally discussed an extrapolation method which uses estimates calculated from the first-order schemes to construct new estimates with higher orders of accuracy. Second and third order methods were constructed before estimates of the weak Euler scheme were used to confirm their order of accuracy through the same Black-Scholes problem.

7 Conclusion

In conclusion, we have found that equipped with some probability theory and basic definitions, many methods exist to approximate solutions of stochastic differential equations. We have seen 2 methods which can be used for mean-square approximation of trajectories, and 3 methods for weak approximation. These were derived through Ito-Taylor expansion (in the case of Euler and Milstein schemes), substitution of processes with similar properties (weak Euler scheme), and the Talay-Tubaro extrapolation method. We found that the implementation of these schemes can be useful for the pricing of European options through our test problem of the Black-Scholes formula. Below, we link the material covered back to the two main aims of this study, and then discuss some options for further development of this content.

7.1 Reflection

The first aim of this report was to dive deep into the theory surrounding Stochastic Analysis. We approached this aim by first setting the scene - introducing the components of the ordered triple, random variables and their expectations (including conditional) plus the characteristic functions which can be used to find their moments. Chapter 2 was closed with a discussion of two limit theorems, one being the Central Limit theorem. Later, we used this to construct confidence intervals for the absolute error criterion as well as Monte Carlo sampling for option pricing through the Feynman-Kac formula.

We continued on to the main object of Stochastic Analysis, Brownian motion. We constructed the object on the space of square-integrable functions over the domain $0 \leq t \leq 1$ using a formal expansion of white noise in the Haar basis, and then integrated with respect to time to obtain a series representing Brownian motion. Finally, we checked that the series converged and that the properties of Brownian motion held. Following this, we found that Brownian motion has bounded quadratic variation equal to the width of the interval of interest, and unbounded first variation. We discussed how one might simulate the process, as well as showing that Brownian motion is Markovian.

Equipped with this knowledge, we moved on to defining Ito's integral, one way of defining integration with respect to Brownian motion. This is the mean-square limit of a sum, a similar idea to Riemann integration from regular calculus. Building off the integral, we were able to define stochastic differential equations. We had a short discussion of the properties an SDEs coefficients might satisfy which guarantee that a solution exists, before considering a specific example in geometric Brownian motion, the model which we tested all the methods on later in the report. Furthermore, we motivated Ito's formula, which gave us some intuition on how to solve some simple SDEs. Using this, we derived the solution to GBM. The last topic we covered in order to achieve the first aim was the probabilistic representation of solutions, which connects the theory of Stochastic Analysis to the solutions of partial differential equations. We derived the Black-Scholes equation then found its solution in the probabilistic sense, another important result which we used as a motivating example in the final chapter of the report. This material covered all the aspects of the theory necessary to begin our study of numerical methods for SDEs, and so we met the first aim.

The second objective was to find numerical schemes for SDEs and test their accuracy. We began to work towards this objective by considering the simplest scheme, that is the Euler scheme. We saw that we could derive this method through very similar ideas to the derivation for ODEs. We saw some approximate trajectories for GBM for varying step-sizes alongside the true solution. It was clear that with a smaller step-size, the approximation we observed was closer to the true solution, but to see just how much closer we had to introduce the notion of mean-square approximation. The Euler method should theoretically have a mean-square order of accuracy of $p = \frac{1}{2}$, and we confirmed this result numerically using the absolute error criterion. Using GBM as a test problem and the arithmetic mean as an unbiased estimator, we saw that the ratio at which the absolute error decreased as the step-size was successively halved approached $\sqrt{2}$, as expected. We were also able to construct confidence intervals for the absolute error criterion, which gave us even more evidence confirming the theoretical result. In order to improve on the mean-square order of accuracy, we performed an Ito-Taylor expansion. Truncating beyond the Euler scheme, we found the Milstein scheme. This method possesses first order accuracy in the mean-square

sense, which we again confirmed using the GBM test problem.

In chapter 6, we considered a second type of approximation, where instead of whole trajectories we considered estimating the moments of a process. The weak order of accuracy was introduced, and we stated that the weak order of accuracy of the Milstein and Euler schemes were both 1. The scaled random walk was motivated as a substitute for more complicated Brownian motion, and through replacement in the Euler scheme we obtained the far simpler weak Euler scheme, which also possesses first weak order of accuracy. We used the Black-Scholes equation as a test problem for the three methods, and they all performed very similarly however the computational cost in using the weak Euler scheme was signifcantly less.

Finally, after introducing the general convergence theorem, we attempted to construct higher order weak methods using an extrapolation technique. Using an expansion of the global error, we were able to construct an improved estimate that has second weak order of accuracy from two estimates with different step-sizes. Furthermore, using a third estimate with a distinct step-size, we were able to construct another estimate with third weak order of accuracy.

The above material met our second aim - we covered three unique schemes, two of which can be used for mean-square approximations and a third for which the moments of a process can be considered. On top of this, we constructed methods of second and third weak order of accuracy using the Talay-Tubaro extrapolation method. All of these methods had their accuracy confirmed using the GBM and Black-Scholes test problems.

7.2 Suggestions for further work

The material covered in this report could be extended in any number of directions. Firstly, in terms of the theory of stochastic analysis, as mentioned in chapter 4 there are a number of ways to define a stochastic integral. One way popular in literature is the Stratonovich integral (see [48]), which we did not discuss. Besides this, the topic of how to compute multiple stochastic integrals was not delved into deeply for the purposes of this report. This would be essential to derive higher-order methods from the Ito-Taylor expansion. Furthermore, with

respect to the methods considered, we did not discuss their stability. This is an essential part of analysis (see [31] for more details on multiple integrals and stability).

Branching off now, the methods we considered were all explicit, meaning the process at a future time was always calculated using the information we had from the present. We could well consider implicit methods, which estimate the process at a future time using information at both present and future times (again, [31] is an excellent book to start with). One other very simple extension on the concept of stochastic differential equations is stochastic partial differential equations, where a random term is added to a partial differential equation (see [49]). This can lead to problems with regularity. Lastly, we briefly mentioned Monte Carlo error in chapter 6. It is possible to reduce this error using variance reduction techniques (see [25]).

References

[1] T. Apostol. *Calculus*. Wiley, 1967.

[2] L. Bachelier. *Théorie de la Spéculation*. Annales Scientifique de l'Ecole Normale Supérieure 3^e série 17, 1900.

[3] R. Bartle. *The Elements of Integration and Lebesgue Measure*. Wiley, 1995.

[4] L. Bergomi. *Stochastic Volatility Modeling*. Routledge, 2015.

[5] D.Bertsekas and J.Tsitsiklis. *Introduction to Probability (First edition)*. Athena Scientific, 2002.

[6] P. Billingsley. *Probability and Measure, 3rd edition* Wiley, 1995.

[7] F. Black and M. Scholes. *The Pricing of Options and Corporate Liabilities*. The Journal of Political Economy, Vol. 81, No. 3 (May - Jun., 1973), pp. 637-654.

[8] L. Breiman. *Probability*. Addison–Wesley, 1968.

[9] C.Brezinski and M. Zaglia. *Extrapolation Methods. Theory and Practice*. North-Holland, 1991.

[10] J. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2008.

[11] A. Carlstein. *Stochastic Calculus: Understanding Brownian Motion and Quadratic Variation*. Link: *http://math.uchicago.edu/ may/REU2019/REU Papers/Carlstein.pdf*. Last accessed: 22/11/2021.

[12] W.Chu. *Abel's lemma on summation by parts and basic hypergeometric series*. Advances in Applied Mathematics. 39 (4): 490–514, 2007.

[13] P.Dirac. *The Principles of Quantum Mechanics*. Oxford University Press, 1947.

[14] R.Durrett *Probability: Theory and Examples*. Cambridge University Press, 1990.

[15] A. Einstein. *Investigations on the Theory of the Brownian Movement*. Dover, 1956.

[16] *Encyclopedia of Mathematics: Haar System*. EMS Press, 2001. (Available online: last accessed 07/11)

[17] J. Epperson. *An Introduction to Numerical Methods and Analysis*. Wiley, 2013.

[18] A. Ermogenous. *Brownian Motion and its applications in the stock market*, from https://academic.udayton.edu. Last accessed: 31/10/21.

[19] L.C. Evans. *An Introduction To Stochastic Differential Equations*. American Mathematical Society, 2013.

[20] L.C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.

[21] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley-Blackwell, 1990.

[22] A. Friedman. *Partial Differential Equations of Parabolic Type*. Dover, 1964.

[23] B. Fristedt and L.Gray. *A Modern Approach to Probability Theory*. Springer, 1997.

[24] I.I. Gichman and A.V. Skorkhod. *Stochastic Differential Equations*. Naukova Dumka, Kiev, 1968; Springer, 1972.

[25] C. Graham and D. Talay. *Stochastic Simulation and Monte Carlo Methods*. Springer, 2013.

[26] A. Haar. *Zur Theorie der orthogonalen Funktionensysteme*. Math. Ann. 69, 331-371, 1910.

[27] *The Haar functions and the Brownian motion*. Link: *http://math.stanford.edu/ ryzhik/STANFORD/STANF227-10/notes-bm.pdf*. Last accessed: 10/11/21.

[28] S. Heston. *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*. The Review of Financial Studies 1993 Volume 6, number 2, pp. 327-343.

[29] K. Ito. *Differential equations determining a Markov process*. Zenkoku Shijo Danwakai 1077 (1942), 1352-1400.

[30] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[31] P.E.Kloeden and E.Platen. *Numerical solution of stochastic differential equations*. Springer, 1992.

[32] P. Kloeden, E. Platen and N. Hofmann. *Extrapolation Methods for the Weak Approximation of Ito Diffusions*. SIAM Journal on Numerical Analysis , October 1995, Vol. 32, No. 5 (Oct., 1995), pp. 1519-1534.

[33] N.V. Krylov. *Introduction to the Theory of Random Processes*. AMS, 2002.

[34] P. Langevin. *Sur la théorie du mouvement brownien*. C. R. Acad. Sci. Paris. 146: 530–533, 1908.

[35] *Lecture notes on Stochastic Taylor expansion*. New York University. Link: *https://www.math.nyu.edu/∼cai/Courses/Derivatives/compfin_lecture_5.pdf* Last accessed: 07/03/2022.

[36] J. Leiner. *Brownian Motion and the Strong Markov Property*. Link: *https://math.uchicago.edu/ may/REU2012/REUPapers/Leiner.pdf*. Last accessed:29/11/2021.

[37] R.Li and S.Nadarajah. *A review of Student's t distribution and its generalizations.* Link: *https://www.research.manchester.ac.uk/portal/files/78743532/studentt.pdf*. Last accessed: 16/02/2022.

[38] L.Loomis and S. Sternberg. *Advanced Calculus*. WSPC, 1968.

[39] *Lp spaces*. Link: *https://www.math.ucdavis.edu/∼hunter/measure_theory/ measure_notes_ch7.pdf*. Last accessed: 13/12/21.

[40] K. Mardia et al. *Multivariate Analysis*. London Academic Press, 1979.

[41] G. Marujama. *Continuous Markov processes and stochastic equations*. Rend. Mat. Circ. Palermo, Ser.2, 4 (1955), 48-90.

[42] G.N. Milstein and M.V. Tretyakov. *Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients*. SIAM J. Numer. Anal. 43 (2005), 1139-1154.

[43] G.N.Milstein and M.V.Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.

[44] G.N.Milstein. *Numerical integration of stochastic differential equations*. Ural State University, Sverdlovsk, 1988.

[45] G.N.Milstein. *A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations*. Theor. Prob. Appl. 32 (1987), 809-811.

[46] G.N.Milstein. *Approximate integration of stochastic differential equations.* Teoriya Veroyatnostei i ee Primeneniya (in Russian). 19 (3): 583–588 (1974).

[47] G.N.Milstein. *A method with second order accuracy for the integration of stochastic differential equations.* Theor. Prob. Appl. 23 (1978), 414-419.

[48] B. Øksendal. *Stochastic Differential Equations : An Introduction with Applications*. Springer, 1998.

[49] B. Øksendal, H. Holden, J. Ubøe and T. Zhang. *Stochastic Partial Differential Equations. A Modeling, White Noise Functional Approach*. Springer, 2010.

[50] R. Paley, N. Wiener, and A. Zygmund. *Notes on random functions*. Math. Z. 37 (1959), 647–668.

[51] J. Perrin. *Brownian Motion and Molecular Reality*. Dover, 1910.

[52] M. Pinsky. *Introduction to Fourier Analysis and Wavelets*. Brooks/Cole, 2002.

[53] A.J.Roberts. *Modifying the Improved Euler scheme to integrate stochastic differential equations*. Link: *https://arxiv.org/pdf/1210.0933.pdf*. Last accessed: 13/03/22.

[54] M.Rosa. *Numerical Methods for Stochastic Differential Equations with Applications to Finance*. University of Lisbon, 2016.

[55] A. Rumbos. *Probability lecture notes*. Link:*https://pages.pomona.edu/ ajr04747/ Spring2008/Math151/Math151NotesSpring08.pdf*. Last accessed 27/10/21.

[56] P. Samuelson. *Proof That Properly Anticipated Prices Fluctuate Randomly*. Industrial Management Review, 6, 41–49, 1965.

[57] E.Schatzker, S.Natarajan and K.Burton. *Bill Hwang Had $20 Billion, Then Lost It All in Two Days*. Bloomberg, 08/04/21.

[58] J. Schauder. *Eine Eigenschaft des Haarschen Orthogonalsystem*. Math. Z 28, 1928.

[59] K. Schorning. *Brownian Motion* (lecture notes). Link: *https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ws13/Stochastik_II/Skript_4 .pdf*. Last accessed:22/11/2021.

[60] A.N. Shiryaev. Probability. Springer, 1996.

[61] S.E. Shreve. *Stochastic Calculus for Finance II: continuous-time models*. Springer, 2004.

[62] L.Smith. *Ito and Stratonovich; a guide for the perplexed*. Link: *https://www.robots.ox.ac.uk/∼lsgs/posts/2018-09-30-ito-strat.html*. Last accessed: 05/12/21.

[63] E.Stein and R. Shakarchi. *Complex analysis*. Princeton University Press, 2007.

[64] D. Talay and L. Tubaro. *Expansion of the global error for numerical schemes solving stochastic differential equations*. Stoch.Anal.Appl. 8 (1990), 483-509.

[65] G. Thomas and R. Finney. *Calculus and analytical geometry*. Addison Wesley, 1996.

[66] M.V.Tretyakov *Introductory Course On Financial Mathematics*. Imperial College Press, 2013.

[67] M.V.Tretyakov *Mini-course on numerical methods for SDEs*. Umea University, 2016.

[68] *Variation of Brownian motion*. Link: *http://stat.math.uregina.ca/~kozdron/Teaching/Regina/862Winter06/Handouts/quad_var_cor.pdf*. Last accessed:01/12/2021.

[69] J.Virtanen. *Real Analysis II Notes*. Link: *http://www.personal.reading.ac.uk/~vn904206/ra2-2015.pdf*. Last accessed:02/12/21.

[70] S.Wang *https://www.whitman.edu/Documents/Academics/Mathematics/2017/Wang.pdf*. Last accessed: 27/10/21.

[71] G.Weiss and R.Rubin. *Random Walks: Theory and Selected Applications*. Wiley, 1982.

[72] A.D. Wentzell. *A Course in the Theory of Stochastic Processes*. McGraw-Hill, New York, 1981.

[73] Facebook stock price chart. Yahoo Finance Website. Link: *https://finance.yahoo.com › quote › FB*. Last accessed: 14/11/2021.

8 Appendix

The link to the github project containing the code that was written to produce all the results in this report is:

https://github.com/pmyjd11/StochasticNumericsDissertation.git

The scripts and functions are listed alphabetically. The information on which files are relevant to which part/figure/table in the report can be found in the README.md file. Simply click this in the list of files or scroll down from the main page to read it. It not only states which functions/drivers are relevant to which section, but also mentions which code sections need to be run to produce specific figures. This information will also be repeated here in the appendix.

Within each script/function, there are comments either at the top or following the function definition which describe briefly what the code does. The code itself is commented regularly for the readers understanding.

The below gives a list of sections where code has been used to produce a figure or table. For each section, the files and any dependencies are stated, and specific code sections are referenced within drivers if necessary.

- Section 3.4 (Brownian Motion - Simulation in 1D): The script BrownianSimulation.m is used to produce figure 4 (page 41). This code is self-contained and does not make calls to any other functions, thus can be ran as is by the reader.

- Section 5.2 (Stochastic Numerics: Strong Taylor Approximations - Implementing the Explicit Euler scheme): The function explicitEulerGBM is called from code section 1 within GBMDriver.m to produce figure 6/7 (page 72/73). Use the 'Run Section' feature to run this code.

- Section 5.3 (Stochastic Numerics: Strong Taylor Approximations - Generalising the implementation of the Explicit Euler Method): The function explicitEulerMethod.m, which calls the function evaluateFunctions.m, is called from code sections 4 and 5 (evaluateFunctions.m is called from explicitEulerMethod.m through GBMDriver.m in

sections 4) within GBMDriver.m to produce figure 8/9 (page 76/77). Note that in sections 4 and 5, there is a function test2 used as an input of explicitEulerMethod.m. To run these sections, the reader will need to copy and paste each section in turn to the bottom of the script, otherwise an error will be displayed (functions must be written at the bottom of script files). Use the 'Run Section' feature to run this code.

- Section 5.5 (Stochastic Numerics: Strong Taylor Approximations - Analysing the Euler scheme: The absolute error criterion): The function explicitEulerGBM.m is called from code section 2 within GBMDriver.m to produce the table on page 82. The function explicitEulerMethod.m, which calls the function evaluateFunctions.m, is called from code sections 6 and 7 (evaluateFunctions.m is called from explicitEulerMethod.m through GBMDriver.m in section 6) within GBMDriver.m to produce figure 10/11 (page 84). Use the 'Run Section' feature to run this code.

- Section 5.7 (Stochastic Numerics: Strong Taylor Approximations - Milstein scheme): The function milsteinGBM.m is called from code section 8 within GBMDriver.m to produce figure 13 on page 90, and it is called from code section 11 to produce the table on page 91. Figure 14 (page 92) is produced through running code section 12 of GBMDriver.m. Use the 'Run Section' feature to run this code.

- Section 6.2 (Stochastic Numerics: Weak methods - Symmetric random walk): The script RandomWalk.m is used to produce figure 16 (page 98). This code is self-contained and does not make calls to any other functions, thus can be ran as is by the reader.

- Section 6.4 (Stochastic Numerics: Weak methods - A numerical example: Continuous Black-Scholes formula): The functions explicitEulerGBM.m, milsteinGBM.m and randomWalkGBM.m are called in code sections 2,3 and 4 of the driver BlackScholes.m to produce the table on page 102, the columns of which are used to produce figure 17 (page 103). Note that the first code section of BlackScholes.m must be run before any others, as this defines the variables of the problem. Use the 'Run Section' feature to run this code.

- Section 6.6 (Stochastic Numerics: Weak methods - Talay-Tubaro extrapolation method): The functions TalayTubaroOrderTwo.m and TalayTubaroOrderThree.m (which both call the function randomWalkGBM.m) are called in code sections 5 and 6 of the driver BlackScholes.m to produce the tables on pages 108 and 110. Note that again, the first code section of BlackScholes.m must be run before any others. Use the 'Run Section' feature to run this code.