Pricing a Multi-Asset American Option in a Parallel Environment by a Finite Element Method Approach

Deniz Kaya
Acknowledgments

I would like to thank my supervisor Professor Johan Tysk at the Department of Mathematics at Uppsala University for providing his support and guidance during this project. I am deeply grateful for his patience and for providing the UPPMAX account making this project possible. I would also like to thank the support team at UPPMAX, especially Tore Sundqvist and Lennart Karlsson.

The support that I have received from the PETSc support team has been invaluable during the application development, especially the help of Barry Smith and Matthew Knepley on PETSc functionality and implementational issues.

I would also like to acknowledge and thank Professor Per Lötstedt at the Division in Scientific Computing at Uppsala University for valuable discussion and guidance on implementing the finite element method. Also a special thank to John Burkardt at the Department of Scientific Computing at Florida State University for providing a C version of the program TRIANGULATION_TRIANGLE_NEIGHBORS.

Finally, I would like to thank my parents for their support in every way possible throughout this project.
Abstract

There is the need for applying numerical methods to problems that cannot be solved analytically and as the spatial dimension of the problem is increased the need for computational recourses increase exponentially, a phenomenon known as the “curse of dimensionality”. In the Black-Scholes-Merton framework the American option pricing problem has no closed form solution and a numerical procedure has to be employed for solving a PDE. The multi-asset American option introduces challenging computational problems, since for every added asset the dimension of the PDE is increased by one. One way to deal with the curse of dimensionality is threw parallelism. Here the finite element method-of-lines is used for pricing a multi-asset American option dependent on up to four assets in a parallel environment. The problem is also solved with the PSOR method giving an accurate benchmark used for comparison. In finance the put option is one of the most fundamental derivatives since it is basically asset-value insurance and a lot of research is done in the field of quantitative finance on accurate and fast pricing techniques for the multi-dimensional case.

“What most experimenters take for granted before they begin their experiments is infinitely more interesting than any results to which their experiments lead.”
Norbert Wiener

“As soon as an Analytical Engine exists, it will necessarily guide the future course of the science. Whenever any result is sought by its aid, the question will then arise – by what course of calculation can these results be arrived at by the machine in the shortest time?”
Charles Babbage
Sammanfattning


Pricing a Multi-Asset American Option
Contents

1 Introduction 1

2 Theory 2
   2.1 Stochastic Calculus 2
      2.1.1 Stochastic processes and stochastic integrals 3
      2.1.2 The Itô formula 7
   2.2 The Feynman-Kač representation theorem 9
   2.3 The Black-Scholes equation 12
      2.3.1 Risk Neutral Valuation 19
      2.3.2 The Black-Scholes formula 20
   2.4 European and American options 21
      2.4.1 European and American vanilla options 21
      2.4.2 Early exercise on a non–dividend–paying stock 22
   2.5 The American option pricing problem 25
      2.5.1 American put options as a free boundary problem 25
      2.5.2 Optimal stopping problem for American put options 29
      2.5.3 Variational Inequality Formulation and Localization to a bounded domain 33
      2.5.4 Penalty approximation 40

3 Numerical methods 41
   3.1 Monte Carlo simulation 42
   3.2 The Finite Difference Methods 43
   3.3 Iterative methods 44
      3.3.1 Projected Successive Over Relaxation 45
   3.4 The Finite Element Method 48
      3.4.1 Matrix representation 49
      3.4.2 Trial functions 50
      3.4.3 A simple 1D example 51

4 Problem setting 55
   4.1 Multi-Asset American options in the Black-Scholes model 55
      4.1.1 Multi-Asset American option on the geometric average 57

5 Implementation and numerical results 59

6 Conclusion and discussion 70

A The variation of a Brownian Motion 72

B Martingales 75

C The Banach fixed point theorem 76

D Transforming the Black-Scholes equation to the Heat equation 79

E A simple 2D example 80
1 Introduction

By definition an option is a contract between the writer (the issuer of the contract) and the holder (the buyer) of the contract. The contract is called an option precisely because it gives the holder the option, but not the obligation, to buy or sell a underlying asset at a pre-specified fixed price within a specified time period. The underlying asset could be anything from foreign currencies to stocks, oranges, timber or any other type of commodities. A call (put) option allows its holder to buy (sell) the underlying asset at the strike price (exercise price). If the holder has an European option it can only be exercised at the expiration (maturity) date. An American option on the other hand can be exercised at any time before it expires. The problem with option pricing is a fundamental financial problem, as trading with derivative securities in the financial market is a core activity for many flourishing companies in the investment business.

The famous Black-Scholes formula gives an explicit pricing formula for European call and put options on stocks which do not pay dividends, see [23]. This explicit pricing formulations were derived in 1973 by the late Fischer Sheffey Black and by Myron Samuel Scholes in their celebrated paper [9]. Robert Cox Merton published the same year independently a paper that contained the Black-Scholes formula with additional extensions. In 1997 Robert C. Merton and Myron S. Scholes received the Nobel prize in economics for the derivation of the Black-Scholes formula and related work, see [38]. Fischer S. Black had passed away in 1995 and the Nobel prize is not given posthumously unless the the winner of the prize has past away after the announcement. In the announcement of the Nobel prize in economics in 1997 the Nobel prize committee prominently mentioned Black’s contribution in the developed theory. As stated in [12] the action of early exercise regarding American options result in a free boundary value problem when the Black-Scholes model is used and finding an explicit closed-form solution is generally not possible. If the underlying assets follow a diffusion process then the value function of the derivative security, in our case an option, satisfies a second order parabolic partial differential equation (PDE) for European-style exercise or a second order parabolic partial differential variational inequality (PDVI) for American-style exercise, see [26].

When exact solutions are too difficult to be practically used, numerical methods can be used to solve the problem at hand. In financial applications the numerical method of choice have been binomial and trinomial trees (lattices). In cases of low-dimensional diffusion problems (one or two state variables) tree methods have the advantage in the ease of implementation over other numerical methods. In the case of multi-asset diffusion problems with more than two assets tree methods have limitations that need to be considered. Trinomial trees can be interpreted as first-order fully explicit finite-difference schemes for the pricing PDE. They are only first-order accurate in time and have the stability restrictions of the form $\Delta t \leq C(\Delta x)^2$, where $\Delta t$ is the time step and $\Delta x$ is the spatial step size. This indicates that large number of time steps may be required to be able to get a convergence within reasonable accuracies as stated in [26].

Other PDE-based methods like finite difference and finite element methods have been applied to the American option problem. This is done by first converting the optimal stopping problem to a PDVI and then applying a numerical method for differential variational inequalities. The PDVI is spatially discretized via a finite difference or finite element scheme and the resulting linear complementary problem (LCP) is solved by by a numerical algorithm such as the Projected Successive Over-Relaxation (PSOR). A method that is still first-order accurate in time which makes it unsuitable in a general problem setting for dimensions higher than two, see
The finite element method is a technique used in the field of structural mechanics and is increasingly used in other areas such as fluid dynamics. It is a numerical method and is excepted as a standard design tool and provides a simple framework within which the solution of a rich diversity of problems may be attempted, see [48]. A general high order-method applicable to multidimensional American-style options lies within the framework of the finite element method-of-lines. The variational inequality is converted to non-linear PDE by the method of penalization where a penalty term is introduced in the PDE to approximate the action of early exercise, see [26]. The finite element method-of-lines, as stated in [6], is a numerical procedure that approximates the solution of time-dependent PDEs by first reducing the problem through finite element discretization of the space variables to that of solving a system of ordinary differential equations (ODEs). There is efficient and reliable software containing the computational procedure needed for solving the system of ODEs. The method used here for solving the pricing problem for a multidimensional American-style option is thus based on approximating the PDVI by a non-linear penalized PDE with a penalty term with continuous Jacobian. The use of continuous Jacobian in contrast to the standard penalty terms with discontinuous Jacobians used in the literature improves the computational performance of the adaptive temporal integrator. The resulting PDE is converted to variational (weak) form which is discretized by the Galerkin finite element method to obtain a system of ODEs. Finally the resulting system of ODEs is integrated in time with an adaptive variable order and variable step size solver SUN-DIALS as done in [26].

In general a numerical approach leads to a set of equations that need large number of mathematical operations to solve and the computational time and the required memory usually increase exponentially as the number of operations increase. The cost associated with the need of higher performance on a sequential processor increases dramatically and one way to circumvent this problem is the use of parallelism, see [36]. The aim here is to develop a parallel application that solves the multidimensional American put option pricing problem by the finite element method-of-lines. A parallel machine consists of several sequential processors that are connected to one another in a specific way depending on the parallel architecture chosen. There is a variety of parallel architectures depending on the network used to interconnect the sequential processors and there is a variety of parallel algorithms. Problems are parallelizable to different degrees depending on the cost of communication. Thus the performance of a parallel machine depends mainly upon communication costs, see [36]. According to [28] some problems may be more time-consuming when solved in a parallel environment than solved on a sequential processor. In addition a problem may have different parallel formulations which result in a varying benefits.

2 Theory

2.1 Stochastic Calculus

Most of the general theory outlined in this section and its following subsections follow the theory outlined in [7] and in [15]. To be able to study and model the prices of assets, and thus options which are financial derivatives of those assets on financial markets, there is need for a stochastic price-model in continuous time. According to [49] asset prices move randomly because of the Efficient Market Hypothesis. All the different forms of this hypothesis say two
things: the price today depends on its past history and holds no further information, markets respond immediately to any new information about an asset price. The use of diffusion processes and stochastic differential equations gives the most complete theory. In dealing with formulas that depend on random variables taking the expected value seems natural and will thus be a powerful tool, in the analysis of the pricing models discussed. A stochastic process $X$ is a diffusion process if its local dynamics can be approximated by the stochastic differential equation given below:

$$X(t + \Delta t) - X(t) = \mu(t, X(t)) \Delta t + \sigma(t, X(t))Z(t), \quad (2.1)$$

where $Z(t)$ is a normally distributed disturbance term, independent of what has occurred up to time $t$ and $\mu$ and $\sigma$ are given deterministic functions. In the stochastic differential equation above the expression in front of $dt$ term is called the (local) drift of the process, whereas the expression in front of the $Z(t)$ term is called the diffusion. Thus the stochastic differential equation or the process above is driven by two separate terms, the drift which is deterministic and a Gaussian disturbance amplified by the magnitude of the diffusion. The drift represents the average rate of growth of the asset price and the diffusion is the volatility and measures the standard deviation of the returns for the asset, i.e. the random change in the asset price in response to external effects, such as unexpected news.

### 2.1.1 Stochastic processes and stochastic integrals

As a first step in trying to model the $Z(t)$ term in (2.1), an intuitive assumption is that the standard deviation regarding a stock price is $\propto \sqrt{\text{trading days}}$. The construction of such a stochastic processes is possible as follows. Choose $N$ and divide $t \in [0, T]$ into $0 = t_0 < t_1 < t_2 < \cdots < t_{N-1} < t_N = T$ sub-intervals, where $t_k = \frac{k}{N}T$. Let $x(t_0) = x(0) = 0$ and given $x(t_{k-1})$ define $x(t_k) = x(t_{k-1}) + \sqrt{\frac{T}{N}}Y_k$ where

$$Y_k = \begin{cases} 
1, & \text{with probability } \frac{1}{2}, \\
-1, & \text{with probability } \frac{1}{2}, 
\end{cases}$$

and the variance of the stochastic process is:

$$\text{Var}[X(T)] = E[X(T)]^2 - (E[X(T)])^2$$

$$= E[X(T)]^2 = E\left(\sum_{k=0}^{N-1} \sqrt{\frac{T}{N}}Y_k\right)^2$$

$$= \sum_{k=1}^{N} E\left(\sqrt{\frac{T}{N}}Y_k\right)^2 = \frac{T}{N} \sum_{k=1}^{N} E(Y_k^2) = T,$$

---

*Pricing a Multi-Asset American Option*
where \( E[\cdot] \) denotes the expected value and the standard deviation is thus \( \sigma = \sqrt{T} \). In the limit, i.e. as \( N \to \infty \) or as \( \Delta t \to 0 \), something very irregular is produced. This limit is called a Brownian Motion (BM). A stochastic process \( W \) is a BM (or, alternatively, Wiener process) if the conditions below are satisfied, see [7]:

i) \( W(0)=0 \) with probability 1.

ii) For any \( s < t \) the quantity \( W(t) - W(s) \) is a normal random variable with mean zero and variance \( (t-s) \), i.e. the random variable \( W(t) - W(s) \) has the Gaussian distribution \( N[0, \sqrt{t-s}] \).

iii) The process \( W \) has independent increments, i.e. for any \( 0 \leq t_1 < t_2 \leq t_3 < t_4 \), the random variables \( W(t_2) - W(t_1) \) and \( W(t_4) - W(t_3) \) are uncorrelated.

iv) The path taken by \( W \) can be chosen to be continuous.

It should be noted that \( W(t) \) is nowhere differentiable with respect to time. This can roughly be verified, as shown in [33], by noting that for \( s < t \),

\[
E \left[ \frac{W(t) - W(s)}{t-s} \right]^2 = \frac{1}{(t-s)^2} E[W(t) - W(s)]^2 \\
= \frac{1}{(t-s)^2} \left( \text{Var}[W(t) - W(s)] + \left( E[W(t) - W(s)] \right)^2 \right) \\
= \frac{t-s}{(t-s)^2} \to \infty,
\]

as \( s \to t \). In using a BM it is possible to rewrite (2.1) in the form

\[
X(t + \Delta t) - X(t) = \mu(t, X(t))\Delta t + \sigma(t, X(t))\Delta W(t),
\]

(2.2)

where \( \Delta W(t):= W(t + \Delta t) - W(t) \). In trying to solve (2.2) above division by \( \Delta t \) and letting \( \Delta t \to 0 \) results in

\[
\begin{cases}
\dot{X}(t) = \mu(t, X(t)) + \sigma(t, X(t))\nu(t), \\
X(0) = x_0,
\end{cases}
\]

where \( \nu(t) \) is the time derivative, i.e. \( \nu(t) = \frac{dW}{dt} \) of the BM and an initial condition has been added. But \( \nu \) is not an ordinary and well defined process and by recalling that a BM is nowhere differentiable, ordinary calculus cannot be applied here resulting in a ODE for each \( \nu \)-trajectory that can be solved in a straightforward fashion. One way to circumvent this problem is to let \( \Delta t \to 0 \) in (2.2) without first dividing by \( \Delta t \) obtaining the following expression

\[
\begin{cases}
dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \\
X(0) = x_0.
\end{cases}
\]

The stochastic process defined in the expression above can be written in integral form

\[Pricing a Multi-Asset American Option\]
\[ X(t) = x_0 + \int_0^t \mu(s, X(s)) \, ds + \int_0^t \sigma(s, X(s)) \, dW(s). \]

In the integral equation above the first integral can be viewed as an ordinary Riemann integral, whereas it may be tempting to view the \( dW(t) \) integral as an Riemann-Stieltjes integral. According to [27] the Riemann-Stieltjes integral is defined as

**Definition 2.1.** Let \( x \in C[a, b] \) (a function space that is continuous on the interval \( [a, b] \) is denoted by \( C[a, b] \)) and let \( w \in BV[a, b] \), where \( BV[a, b] \) denotes the normed space of all functions of bounded variation on \( [a, b] \). Define the partition of the interval as \( \Delta_n := \{a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b\} \) and define the mesh, i.e. the length of the largest interval in the partition, as \( \|\Delta_n\| := \max_{1 \leq i \leq n} (t_i - t_{i-1}) \). Then for every partition \( \Delta_n \) on \( [a, b] \) consider the sum

\[ s(\Delta_n) = \sum_{i=1}^n x(t_i)[w(t_i) - w(t_{i-1})]. \]  

There exists a number \( J \) with the property that for every \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that

\[ |J - s(\Delta_n)| < \varepsilon. \]

\( J \) is called the Riemann-Stieltjes integral of \( x \) over \( [a, b] \) w.r.t. \( w \) and is denoted by

\[ \int_a^b x(t) \, dw(t). \]

The integral (2.4) can be obtained as the limit of the sums (2.3) for a sequence \( \Delta_n \) of partitions of \( [a, b] \) satisfying \( \|\Delta_n\| \to 0 \) as \( n \to 0 \). But since \( W \) in \( dW(t) \) has unbounded variation this approach is not possible, see **Corollary A.3** in Section A for the derivation of the variation of a BM. The restriction on \( W(t) \) having bounded variation is to be certain that the integral discussed doesn’t blow up. Another approach is possible in using **Theorem A.1**, which states that the quadratic variation of a BM is finite, in giving a global\( (L^2) \)-definition of integrals of the form \( \int_0^t g(s) \, dW(s) \). The function \( g(t) \) is a stochastic process that belongs to the class \( L^2[a, b] \) and is determined by the information generated by \( W \), i.e there are integrability conditions imposed on \( g \) and \( g \) is adapted to the filtration \( \forall t: \{\mathcal{F}_t^W\}_{t \geq 0} \). Following [7] it is adequate here to define the integrability conditions imposed on \( g \) and what is meant by imposing that a function is adapted to the information generated by a stochastic process.

**Definition 2.2.**

(i) The process \( g \) belongs to class \( L^2[a, b] \) if the following conditions are satisfied.

- \( \int_a^b E[g^2] \, ds < \infty \).
• The process $g$ is adapted to the filtration $\{\mathcal{F}_t^W\}_{t \geq 0}$, i.e. $g(t)$ is determined by $\{W(s) : 0 \leq s \leq t\}$.

(ii) The process $g$ belongs to the class $L^2$ if $g \in L^2[0, t]$ for all $t > 0$.

**Definition 2.3.** Let $X$ be a stochastic process then

(i) An event $A$ is $\mathcal{F}_t^X$-measurable if it is possible to determine whether $A$ happened or not on observations of $X$ up to time $t$.

(ii) If a random variable $Z$ can be completely determined by observations of $\{X(s) : 0 \leq s \leq t\}$ then $Z \in \mathcal{F}_t^X$.

(iii) A stochastic process $Y$ such that $Y(t) \in \mathcal{F}_t^X$ for all $t$ is adapted to the filtration $\{\mathcal{F}_t^X\}_{t \geq 0}$.

If the function $g$ is simple, i.e. $g(s) = g(t_k)$ for $s \in [t_k, t_{k+1})$ for some partition $a = t_0 < t_1 < t_2 < \cdots < t_n = b$, and **Definition 2.2** is fulfilled the stochastic integral discussed can be defined by

$$\int_a^b g(s) \, dW(s) = \sum_{k=0}^{n-1} g(t_k)(W(t_{k+1}) - W(t_k)).$$

For a general $g \in L^2$, $g$ can be approximated with a simple function $g_k$ and the integral can be defined as $\int_a^b g(s) \, dW(s) = \lim_{k \to \infty} \int_a^b g_k(s) \, dW(s)$.

The most important properties of the stochastic integral are given in the proposition below.

**Proposition 2.1.** Assume that **Definition 2.2** is fulfilled, i.e. $g \in L^2$, then

(i) $E\left[\int_a^b g(s) \, dW(s)\right] = 0.$

(ii) $E\left[\left(\int_a^b g(s) \, dW(s)\right)^2\right] = \int_a^b E[g(s)^2] \, ds.$

(iii) $\int_a^b g(s) \, dW(s)$ is $\mathcal{F}_t^W$-measurable.

**Proof.** Assume $g$ is simple (the general case follows by approximations).

(i) $E\left[\int_a^b g(s) \, dW(s)\right] = E\left[\sum_{k=0}^{n-1} g(t_k)(W(t_{k+1}) - W(t_k))\right] = \sum_{k=0}^{n-1} E\left[g(t_k)(W(t_{k+1}) - W(t_k))\right]$

$= [g(t_k) \text{ and } (W(t_{k+1}) - W(t_k)) \text{ are independent}] = \sum_{k=0}^{n-1} E[g(t_k)] E[W(t_{k+1}) - W(t_k)] = 0.$

(ii) $E\left[\left(\int_a^b g(s) \, dW(s)\right)^2\right] = E\left[\sum_{k=0}^{n-1} [g(t_k)]^2 [W(t_{k+1}) - W(t_k)]^2\right]$

$= \sum_{k=0}^{n-1} E[g^2(t_k)] E[(W(t_{k+1}) - W(t_k))^2]$

$= \sum_{k=0}^{n-1} E[g^2(t_k)] Var[\Delta W]$

$= \sum_{k=0}^{n-1} E[g^2(t_k)] (E[\Delta W]^2)$

$= \sum_{k=0}^{n-1} E[g^2(t_k)] \Delta t = [\text{in the limit}] = \lim_{n \to \infty} \int_a^b E[g(s)]^2 \, ds.$
By definition d

other terms. These notations will be used interchangeably in the following text.

\[ \partial \]

The term

2.1

Remark

2.1.2 The Itô formula

As stated before an alternative way of writing the stochastic process of the form

\[ X(t) = x_0 + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s), \]

where \( x_0 \) is a constant, \( W \) is a BM and \( \mu \) and \( \sigma \) are \( \mathcal{F}^W \)-adapted processes, is with stochastic differentials:

\[
\begin{cases}
  dX(t) = \mu(t)dt + \sigma(t)dW(t), \\
  X(0) = x_0.
\end{cases}
\]

Assume that the function \( f(t, y) \in C^{1,2}, \) i.e. \( f \) is continuous and once differentiable w.r.t. \( t \) and twice differentiable w.r.t. \( y. \) Let \( Z(t) = f(t, X(t)) \) and in trying to find an expression for \( dZ, \) by the use of Taylor’s theorem, gives

\[
dZ = df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial y}dX + \frac{1}{2} \frac{\partial^2 f}{\partial y^2}(dX)^2 + \frac{\partial^2 f}{\partial t \partial y}dtdX + \frac{1}{2} \frac{\partial^2 f}{\partial t^2}(dW)^2. \tag{2.5}
\]

Remark 2.1. For readability reasons a lot of variables are suppressed in the expression above. The term \( \frac{\partial f}{\partial t}, \) for example, is shorthand notation for \( \frac{\partial f}{\partial t}(t, X(t)) \) and correspondingly for the other terms. These notations will be used interchangeably in the following text.

By definition \( dX(t) = \mu(t)dt + \sigma(t)dW(t) \) and thus

\[
(dX)^2 = (\mu dt + \sigma dW)^2 = \mu^2 dt^2 + 2\mu \sigma dtdW + \sigma^2 (dW)^2.
\]

It can be shown that the \( (dt)^2 \) - and the \( dtdW \)-terms are negligible compared to terms only containing \( dt \) when all the terms are plugged in (2.5). Before plugging in all expressions in (2.5) there is need for finding the value of \( (dW)^2. \) Following [7] this can be done by choosing \( n \) equally large subintervals and defining \( t_k = \frac{k}{n}t \) in the interval \( 0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = t. \) Now let

\[
S_n = \sum_{k=1}^{n} (W(t_{k+1}) - W(t_k))^2 = \sum_{k=1}^{n} (\Delta W(t_k))^2,
\]

then in the limit as \( n \to \infty \)

\[
E[S_n] = \sum_{k=1}^{n} E[(\Delta W^2(t_k))] = \sum_{k=1}^{n} [\text{Var}[\Delta W(t_k)] + (E[\Delta W(t_k)])^2] = \sum_{k=1}^{n} (t_k - t_{k-1}) = \sum_{k=1}^{n} \left( \frac{k}{n} - \frac{k-1}{n} \right) = \sum_{k=1}^{n} \left( \frac{1}{n} \right) = t. \tag{2.6}
\]
Taking into consideration that a BM has independent increments,

\[
Var[S_n] = \sum_{k=1}^{n} Var[(\Delta W(t_k))^2] = \sum_{k=1}^{n} [E[\Delta W^4(t_k)] - (E[\Delta W^2(t_k)])^2]
\]

\[
= \sum_{k=1}^{n} (3(t_k - t_{k-1})^2 - (t_k - t_{k-1})^2) = 2 \sum_{k=1}^{n} (t_k - t_{k-1})^2
\]

\[
= \sum_{k=1}^{n} \left( k - \frac{k-1}{n} \right)^2 = \frac{2}{n} \to 0 \quad \text{as } n \to \infty,
\]

where the fact that the fourth central moment of a normally distributed random variable with mean zero and variance \( (t_k - t_{k-1})^2 \) is \( 3(t_k - t_{k-1})^2 \) has been used.

**Remark 2.2.** The result derived in equation (2.6) is also given in A by **Theorem A.1**, where \( ||\Delta_n|| \to 0 \) is equivalent to the limit \( n \to \infty \).

Thus as \( n \to \infty \), \( E[S_n] = t \) and \( Var[S_n] \to 0 \), i.e. \( S_n \) tends to the deterministic limit \( t \), which motivates the following

\[
\int_0^t (dW)^2 = \int_0^t dt \Leftrightarrow (dW)^2 = dt.
\]

Following [7] and plugging in all the expressions for \( dX \), \( (dX)^2 \) and \( (dW)^2 \) in the Taylor expansion (2.5) gives the main result in the theory of stochastic calculus—the Itô formula.

**Theorem 2.1. (Itô’s formula)** Assume that the stochastic process \( X \) has a stochastic differential of the form

\[
dX(t) = \mu(t)dt + \sigma(t)dW(t),
\]

where \( W \) is a BM and \( \mu \) and \( \sigma \) are \( \mathcal{F}^W \)-adapted processes. Let the function \( f(t, y) \in C^{1,2} \) and define \( Z(t) = f(t, X(t)) \). The stochastic differential for \( Z \) is then given by

\[
dZ = df(t, X(t)) = \left\{ \frac{\partial f}{\partial t} + \mu \frac{\partial f}{\partial y} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial y^2} \right\} dt + \sigma \frac{\partial f}{\partial y} dW(t),
\]

which follows from the second order Taylor expansion

\[
df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial y} dX + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} (dX)^2
\]

where the following multiplication rules have been used:

\[
\left\{ \begin{array}{c}
(dt)^2 = 0, \\
(dt \cdot dW = 0, \\
(dW)^2 = dt.
\end{array} \right.
\]

**Remark 2.3.** All the reasoning in the derivation of **Theorem 2.1** has been purely motivational. There is an outline of the full proof in [8] and the full proof is given in [5].

The multidimensional Itô formula, following [7], for independent BMs is given by
Theorem 2.2. (The Multidimensional Itô formula) Assume that the \( n \)-dimensional stochastic process \( X \) has a stochastic differential of the form

\[
dX(t) = \mu(t)dt + \sigma(t)dW(t)
\]

i.e.

\[
X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \quad W = \begin{bmatrix} W_1 \\ \vdots \\ W_d \end{bmatrix} \quad \text{and} \quad \sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nd} \end{bmatrix}.
\]

Define the process \( Z = f(t, X(t)) \) where \( f : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R} \) is a \( C^{1,2} \) mapping. (The notation \( \mathbb{R}_+ := [0, \infty) \). The stochastic differential for \( Z \) is then given by

\[
dZ = df = \left\{ \frac{\partial f}{\partial t} + \sum_{i=1}^n \mu_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n C_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right\} dt + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \sigma_i dW_i,
\]

where \( \sigma_i \) is the \( i \)-th row vector of the matrix \( \sigma \), i.e. \( \sigma_i = [\sigma_{1i} \cdots \sigma_{ni}] \), and the matrix \( C = \sigma \sigma^* \) where \( ^* \) denotes the transpose. The differential is also given by the formula

\[
df = \frac{\partial f}{\partial t} dt + \sum_{i=1}^n \frac{\partial f}{\partial x_i} dX_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} (dX_i)(dX_j),
\]

where the following multiplication rules have been used:

\[
\begin{align*}
(d\tau)^2 &= 0, \\
d\tau \cdot dW &= 0, \\
(dW_i)^2 &= dt, \quad i = 1, \ldots, d, \\
dW_i \cdot dW_j &= 0, \quad i \neq j.
\end{align*}
\]

Remark 2.4. The multidimensional Itô formula for correlated Brownian Motions is derived and given in Section [8].

2.2 The Feynman-Kač representation theorem

Following [15] and [7] consider the terminal value problem: Find \( F(t, x) \) such that

\[
\begin{align*}
\frac{\partial F}{\partial t} (t, x) + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 F}{\partial x^2} (t, x) + \mu(t, x) \frac{\partial F}{\partial x} (t, x) &= 0, \\
F(T, x) &= \Phi(x),
\end{align*}
\]

Pricing a Multi-Asset American Option
where \( \sigma, \mu \) and \( \Phi \) are given scalar functions and \( t \in [0, T] \). Let \( F \) be the solution to (2.7) and define \( Z(t) = F(t, X(t)) \). Let the stochastic process \( X \) on the time interval \( s \in [t, T] \) be defined as

\[
\begin{aligned}
&\begin{cases}
  dX_s = \mu(s, X_s)ds + \sigma(s, X_s)dW_s, \\
  X_t = x.
\end{cases}
\end{aligned}
\]  

(2.8)

**Remark 2.5.** For readability reasons the term \( X_s \) in (2.8) is a shorthand notation for \( X(s) \) and correspondingly for the other terms. These notations will be used interchangeably in the following text.

Applying Itô’s formula (**Theorem 2.1**) on the process \( F(s, X_s) \), recall that by definition \( Z(s) = F(s, X_s) \), and using equation (2.7) gives

\[
\begin{aligned}
dZ(s) &= \frac{\partial F}{\partial s}ds + \mu \frac{\partial F}{\partial x}dx + \frac{1}{2} \sigma^2(s, X_s) \frac{\partial^2 F}{\partial x^2}ds + \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s)dW_s. \\
&= \frac{\partial F}{\partial s}ds + \mu(s, X_s) \frac{\partial F}{\partial x}(s, X_s) + \frac{1}{2} \sigma^2(s, X_s) \frac{\partial^2 F}{\partial x^2}(s, X_s) \frac{\partial F}{\partial x}(s, X_s)dW_s.
\end{aligned}
\]

Thus

\[
\begin{aligned}
Z(T) &= \int_t^T \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s)dW_s, \\
E[\Phi(X_T)] &= E[Z(t)] + E \left[ \int_t^T \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s)dW_s \right].
\end{aligned}
\]

As seen from the computations above \( F(t, x) = E_{t,x}[\Phi(X_T)] \), where the indexing on the expectation operator emphasizes that the expected value is to be taken given the initial value \( X_t = x \). In order to be able to use **Proposition 2.1** that will guarantee that the expected value of the stochastic integral equals zero there is need for the integrability assumption \( \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) \in L^2 \). The boundary value problem (2.7) is a so-called parabolic problem and will have infinitely many solutions of which only one is “nice” while the rest are rather “wild” solutions. The proposition below will only give the “nice” solution.

**Proposition 2.2.** (**Feynman-Kač representation theorem**) If \( F(t, x) \) satisfies

\[
\begin{aligned}
&\begin{cases}
  \frac{\partial F}{\partial t}(t, x) + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 F}{\partial x^2}(t, x) + \mu(t, x) \frac{\partial F}{\partial x}(t, x) = 0, \\
  F(T, x) = \Phi(x),
\end{cases}
\end{aligned}
\]

where \( \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) \in L^2 \), then \( F \) has the representation

\[
F(t, x) = E_{t,x}[\Phi(X_T)],
\]

where the stochastic process \( X \) satisfies the stochastic differential equation (SDE)

\[
\begin{aligned}
&\begin{cases}
  dX_s = \mu(s, X_s)ds + \sigma(s, X_s)dW_s, \\
  X_t = x.
\end{cases}
\end{aligned}
\]
Considering the closely related multidimensional boundary value problem

\[
\begin{aligned}
\frac{\partial F}{\partial t}(t, x) + \frac{1}{2} \sum_{i,j=1}^{n} C_{ij}(t, x) \frac{\partial^2 F}{\partial x_i \partial x_j}(t, x) + \sum_{i=1}^{n} \mu_i(t, x) \frac{\partial F}{\partial x_i}(t, x) - rF(t, x) &= 0, \\
F(T, x) &= \Phi(x),
\end{aligned}
\]  

(2.9)

where \( r \in \mathbb{R} \) is given constant and the matrix \( C(t, x) = (\sigma \sigma^*)(t, x) \). Using the technique of integrating factors w.r.t. the first variable which here is \( t \) when solving first order linear differential equations, for details see [1], the equation (2.9) is multiplied by the factor \( e^{r(s-t)} \) and in considering the process \( Z(s) = e^{-r(s-t)}F(s, X_s) \) according to [15] and [7] the following result is obtained.

**Proposition 2.3.** Assume that \( F(t, x) \) (where \( x \in \mathbb{R}^n \)) satisfies

\[
\begin{aligned}
\frac{\partial F}{\partial t}(t, x) + \frac{1}{2} \sum_{i,j=1}^{n} C_{ij}(t, x) \frac{\partial^2 F}{\partial x_i \partial x_j}(t, x) + \sum_{i=1}^{n} \mu_i(t, x) \frac{\partial F}{\partial x_i}(t, x) - rF(t, x) &= 0, \\
F(T, x) &= \Phi(x),
\end{aligned}
\]  

(2.10)

where \( r \in \mathbb{R} \) is given constant and \( C(t, x) = (\sigma \sigma^*)(t, x) \). Then \( F \) has the representation

\[ F(t, x) = e^{-r(T-t)}E_{t,x}[\Phi(X(T))], \]

where the stochastic process \( X \) satisfies the SDE

\[
\begin{aligned}
dx_i &= \mu_i(s, X_s)ds + \sigma_i(s, X_s)dW_s, \\
x_i &= x,
\end{aligned}
\]  

(2.11)

for \( s > t \) and where \( x \) is deterministic. Furthermore it is assumed that \( \sum_{i=1}^{n} \sigma_i(s, X_s) \frac{\partial F}{\partial x_i}(s, X_s) \in L^2 \).

**Proof.** Defining \( Z(s) = e^{-r(s-t)}F(s, X_s) \), applying the multidimensional Itô formula (Theorem 2.2) and using equation (2.10) gives

\[
\begin{aligned}
dZ_s &= \left(e^{-r(s-t)} \frac{\partial F}{\partial s}(s, X_s) + e^{-r(s-t)} A F(s, X_s) - re^{-r(s-t)} F(s, X_s) \right)ds \\
&\quad + e^{-r(s-t)} \sum_{i=1}^{n} \sigma_i(s, X_s) \frac{\partial F}{\partial x_i}(s, X_s)dW_s,
\end{aligned}
\]

\[ \text{Pricing a Multi-Asset American Option} \]
where $\mathcal{A}$ is the infinitesimal generator for any function $h(x) \in C^2(\mathbb{R}^n)$ and is, following **Definition** 5.4 in [7], given by

$$
\mathcal{A} h(t, x) = \sum_{i=1}^{n} \mu_i(t, x) \frac{\partial h}{\partial x_i}(x) + \frac{1}{2} \sum_{i,j=1}^{n} C_{ij}(t, x) \frac{\partial^2 h}{\partial x_i \partial x_j}(x).
$$

where as given earlier $C(t, x) = (\sigma \sigma^*)(t, x)$. Hence

$$
e^{-r(T-t)} = Z(T) = Z(t) + \int_{t}^{T} \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s
$$

$$
e^{-r(T-t)} F(T, X_T) = F(t, X_t) + \int_{t}^{T} \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s
$$

$$
e^{-r(T-t)} \Phi(X_T) = F(t, x) + \int_{t}^{T} \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s,
$$

where the stochastic process $X$ satisfies (2.11) and $\sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) \in L^2$. Taking expectations and using **Proposition** 2.1 gives

$$
e^{-r(T-t)} E_{t,x}[\Phi(X_T)] = E_{t,x}\left[F(t, x) + \int_{t}^{T} \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s\right]
$$

$$
= E_{t,x}[F(t, x)] + E_{t,x}\left[\int_{t}^{T} \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s\right] = F(t, x).
$$

### 2.3 The Black-Scholes equation

Most of the general theory outlined in this section and its following subsection follow the theory outlined in [7] and in [15]. Following [9], [7] and [49] there are some assumptions and definitions needed in the derivation of the celebrated Black-Scholes equation for valuing an option in terms of the price of the stock. The assumptions below give “ideal conditions” in the financial market for the stock and the option.

- **The short-term interest rate** is a known constant.
- **The underlying asset price**, or the stock price, follows a random walk in continuous time where the variance rate is constant. Thus the distribution of possible stock prices at the end of any finite time interval is lognormal, i.e. the stock price follows a lognormal random walk.
- **The underlying asset to the option does not pay any dividends or other distributions during the life of the option.**
- **There are no transaction costs** regarding the option or the stock.

---

**Pricing a Multi-Asset American Option**
• Short selling is allowed with no penalties and the assets are divisible, i.e. it is possible to borrow any fraction of the price of the security to buy it or to hold it at the short-term interest.

• There are no arbitrage possibilities (see Definition 2.4 below) on the financial market where the trading of the option and the stock takes place.

And the definitions needed are:

**Definition 2.4.** An arbitrage in a financial market is a self-financing portfolio \( \{ h(t) : 0 \leq t \leq T \} \) such that

\[
V^h(0) = 0, \\
P(V^h(T) \geq 0) = 1, \\
P(V^h(T) > 0) > 0,
\]

where \( V^h \) denotes the value process of the portfolio, i.e. \( V^h(t) = \sum_{i=1}^{N} h_i(t)S_i(t) \). By definition, see [7], \( \{ S(t) : t \geq 0 \} \) is an \( N \)-dimensional price process, i.e. \( S(t) = (S_1(t), \ldots, S_N(t)) \) and \( h_i(t) = \# \) of shares of stock \( i \) held at time \( t \). The portfolio held during time \( t \) is then \( h(t) = (h_1(t), \ldots, h_N(t)) \). Thus an arbitrage is a risk-free profit and the market is arbitrage-free if there are no arbitrage possibilities.

**Definition 2.5.** Consider a financial market with vector price process \( S \). A contingent claim with maturity \( T \) (\( T \)-claim) is any random variable \( X \in \mathcal{F}_T^3 \) and is simple if \( X = \Phi(S(T)) \), where \( \Phi(S(T)) \) is called the contract function, or alternatively the payoff function.

The aim in this section is to find the arbitrage-free price \( \Pi(t; X) \) of a simple \( T \)-claim, i.e. a contingent claim where \( X = \Phi(S(T)) \), assuming that the price process for the derivative asset is of the form \( \Pi(t; X) = F(t, S(t)) \) where \( F \) is some smooth function. The aim is thus to derive a PDE for \( F \).

**Definition 2.6.** The Black-Scholes model for the financial market consists of two assets, one risk-free and one risky, with dynamics given by

\[
dB(t) = rB(t)dt, \\
dS(t) = \alpha(t, S(t))S(t)dt + \sigma(t, S(t))S(t)d\bar{W}(t),
\]

where \( r, \alpha \) and \( \sigma \) are deterministic constants. The risk-free asset can be thought as representing a bank account whereas the risky asset represents a stock. The notation \( \bar{W} \) instead of \( W \) will be discussed when considering risk neutral valuation, see the following section. One more observation worth mentioning is that the process governing the stock is called a Geometric Brownian Motion (GBM). This type of process is one of the fundamental building blocks for the modelling of asset prices and should be given extra attention.

**Proposition 2.4.** (Geometric Brownian Motion) A process defined by the SDE

\[
\begin{aligned}
\begin{cases}
    dX_t = \alpha X_t dt + \sigma X_t dW_t, \\
    X_0 = x_0,
\end{cases}
\end{aligned}
\tag{2.12}
\]

where \( \alpha \) and \( \sigma \) are constants is called a Geometric Brownian Motion. By Proposition 5.2 in [7]
the solution to the equation above equation is

\[ X_t = x_0 e^{(a - \frac{1}{2} \sigma^2) t + \sigma W_t}, \]

with the expected value

\[ E[X_t] = x_0 e^{at}. \]

Figure 1 illustrates simulated trajectories of a 1000-timestep Geometric Brownian Motion for different values of \( \sigma \) while holding all other variables fixed. The simulations were done in MATLAB and even the expected value of the processes, where \( x_0 = 1 \), is plotted for comparison, i.e. the smooth line is is the plot of the function \( E[X_t] = 1 \cdot e^{at} \). A observable feature is that for small values of \( \sigma \) the trajectory of the process will follow the expected value curve fairly close, whereas higher values of \( \sigma \) give rise to larger random deviations from the expected value curve.

\[ \alpha = 1 \text{ and } \sigma = 0.2. \]

Figure 1: Geometric Brownian Motions with \( x_0 = 1 \) and different values for \( \sigma \).

Proof. Since equation (2.12) can be written more formally as

\[ \dot{X}_t = \left( \alpha + \sigma W_t \right) X_t, \]
the GBM can be viewed as a linear ordinary differential equation (ODE) with the stochastic coefficient driven by white noise. The deterministic linear equation, i.e. when $\sigma = 0$ in the ODE above, has a solution that is an exponential function of time and motivates the introduction of the process $Z_t = \ln X_t$. With the assumptions that $X$ is a solution and that $X > 0$ and applying Itô's formula (Theorem 2.1) on the process $Z_t$ gives

$$dZ_t = \frac{1}{X_t} dX_t + \frac{1}{2} \left( -\frac{1}{X_t^2} \right) (dX_t)^2 = \alpha dt + \sigma dW_t - \frac{1}{2} \sigma^2 dt$$

Thus

$$Z_t = \ln x_0 + \int_0^t \left( \alpha - \frac{1}{2} \sigma^2 \right) ds + \int_0^t \sigma dW_s = \ln x_0 + \left( \alpha - \frac{1}{2} \sigma^2 \right) t + \sigma W_t,$$

which means that $X$ is given by

$$X_t = x_0 e^{(\alpha - \frac{1}{2} \sigma^2) t + \sigma W_t}.$$

This completes the proof of the first part in the proposition above. Taking expectation and using the fact that $\alpha$ and $\sigma$ are constant gives the following equation

$$E[X_t] = x_0 e^{(\alpha - \frac{1}{2} \sigma^2) t} E[e^{\sigma W_t}],$$

where $E[e^{\sigma W_t}]$ needs to be computed. In defining the process $Y_t = e^{\sigma W_t}$ and $Y_t = f(t, X_t)$ where $X = W$ and $f$ is given by $f(t, x) = e^{\alpha x}$ the Itô formula gives the SDE

$$dY_t = d f = \sigma e^{\sigma W_t} dW_t + \frac{1}{2} \sigma^2 e^{\sigma W_t} dt = \frac{1}{2} \sigma^2 Y_t dt + \sigma Y_t dW_t,$$

with the initial condition $Y_0 = y_0 = 1$. Rewriting the above expression in integral form, taking expectations and using Proposition 2.1 leads to

$$E[Y_t] = E \left[ 1 + \frac{1}{2} \sigma^2 \int_0^t Y_s dW_s + \sigma \int_0^t Y_s dW_s \right]$$

$$= 1 + \frac{1}{2} \sigma^2 \int_0^t E[Y_s] ds + E \left[ \sigma \int_0^t Y_s dW_s \right]$$

$$\Leftrightarrow m_t = 1 + \frac{1}{2} \sigma^2 \int_0^t m_s ds.$$

Taking the derivative w.r.t. $t$ in the expression above the following ODE is obtained where the solution is straightforward.

$$\begin{cases}
  \dot{m}_t = \frac{1}{2} \sigma^2 m_t \\
  m_0 = 1,
\end{cases} \Rightarrow m_t = e^{\frac{1}{2} \sigma^2 t}.$$
Thus \( E[e^{\sigma W_t}] = E[Y_t] = m_t = e^{\frac{1}{2}\sigma^2 t} \) and the second part of the proposition above can now be proven.

\[
E[X_t] = x_0 e^{(\alpha - \frac{1}{2}\sigma^2) t} E[e^{\sigma W_t}] = x_0 e^{(\alpha - \frac{1}{2}\sigma^2) t} e^{\frac{1}{2}\sigma^2 t} = x_0 e^{\alpha t}.
\]

The idea is to construct a portfolio, which is self-financing and risk-free, based on the stock and the option price where \( \Pi(t) \) is such that the market \( (B(t), S(t), \Pi(t)) \) is arbitrage free. A portfolio with the value \( V \) that is self-financing means that the dynamics of the value process are given by \( dV = \sum_{i=1}^{N} h_i dS_i \), see Definition 2.4 above. Thus the relative weight of asset \( S_i \) at time \( t \) is \( \frac{h_i(t) S_i(t)}{V(t)} \) and \( \sum_{i=1}^{N} u_i = 1 \), where \( u_i \) can be negative if short selling is allowed. (Recall that in the Black-Scholes model there are two assets, the stock and the bank account.) The dynamics of the value process for the portfolio can be rewritten as

\[
dV = V \left( \sum_{i=1}^{N} \frac{h_i}{V} dS_i \right) = V \left( \sum_{i=1}^{N} \frac{h_i S_i}{V} \frac{dS_i}{S_i} \right) = V \sum_{i=1}^{N} u_i \frac{dS_i}{S_i}
\]

\[\Rightarrow \frac{dV}{V} = \sum_{i=1}^{N} u_i \frac{dS_i}{S_i}\]

If a portfolio has a value process \( dV \) without containing no driving Brownian Motion the portfolio is a \textit{locally risk-less portfolio} and the rate of return, i.e. the drift, for that portfolio must equal the short rate of interest \( r \). That is for preventing any arbitrage possibilities. Thus on a arbitrage-free market there can only be one short rate of interest. Applying Itô’s formula (Theorem 2.1) to the option price process \( \Pi(t) \), dividing both sides by \( F \) and multiplying both sides by \( \Pi \) (recall that \( \Pi(t, X) = F(t, S(t)) \)), gives

\[
d\Pi = \left( \frac{\frac{\partial F(t, S(t))}{\partial (t,S(t))}}{F} \right) dt + \frac{F_S}{F} dS + \frac{1}{2} \frac{F_{SS}}{F} (dS)^2 = \left( \frac{F_t + \alpha S F_S + \frac{1}{2} \sigma^2 S^2 F_{SS}}{F} \right) \Pi dt + \frac{\sigma S F_S}{F} \Pi d\bar{W}
\]

\[\Rightarrow d\Pi = \alpha \Pi dt + \sigma \Pi d\bar{W}, \quad (2.13)\]

where the dynamics for the stock process \( S \) is given in Definition 2.6. In the expression \( (dS)^2 = \alpha^2 S^2 dt^2 + \alpha \sigma S^2 dtd\bar{W} + \sigma^2 S^2 d\bar{W}^2 \) the fact that the \( (dt)^2 \)- and the \( dtd\bar{W} \)-terms are negligible compared to terms only containing \( dS \) and that \( (d\bar{W})^2 = dt \), see Section 2.1.2, has been used.

\textbf{Remark 2.6.} The notation \( F_t \) in the equation above denotes partial derivative w.r.t. \( t \) and correspondingly for the other terms.

\textit{Pricing a Multi-Asset American Option}
The Black-Scholes equation

From the discussion above let \((u_S, u_\Pi)\) be a relative portfolio consisting of the stock \(S\) and the option \(\Pi\), i.e. \(u_S + u_\Pi = 1\) where \(u_S\) is the relative weight in stocks and \(u_\Pi\) is the relative weight in the option. The value process of the portfolio is then given by

\[
dV = V\left(\frac{dS}{S} u_S + \frac{d\Pi}{\Pi} u_\Pi\right)
= V\left(\alpha_S dt + \sigma_S d\bar{W}\right) + u_S\left(\alpha_\Pi dt + \sigma_\Pi d\bar{W}\right)
= V\left((u_S \alpha + u_\Pi \alpha_\Pi) dt + \frac{(u_S \sigma + u_\Pi \sigma_\Pi) d\bar{W}}{=0}\right),
\]

where the fact that a risk-less portfolio doesn’t contain any \(dW\)-term has been used. So \((u_S, u_\Pi)\) are chosen so that

\[
\begin{align*}
    u_S + u_\Pi &= 1 \\
    u_S \sigma + u_\Pi \sigma_\Pi &= 0
\end{align*}
\]

\(\Rightarrow\)

\[
\begin{align*}
    u_S &= \frac{\sigma_\Pi}{\sigma_\Pi - \sigma} \\
    u_\Pi &= \frac{\sigma \sigma_\Pi}{\sigma_\Pi - \sigma}
\end{align*}
\]

(2.14)

With this choice of weights and from the discussion above regarding arbitrage

\[
dV = \left(u_S \alpha + u_\Pi \alpha_\Pi\right) V dt + 0 \cdot \bar{W}.
\]

If \((*) > r\) in the equation above, where it is assumed for simplicity and by Definition 2.6 that \((*)\) and \(r\) are constants, the owner of the portfolio can borrow from the bank at rate \(r\) and invest in the portfolio which has an higher rate of return without any risk (recall that the portfolio is a risk-less portfolio). At a later time the money borrowed can be returned to the bank and a profit has been made. But this gives an arbitrage possibility, i.e. the possibility to make an risk-less profit. By the assumptions made in the beginning of this section arbitrage is not allowed in the model. Similarly if \((*) < r\) the owner of the portfolio can sell a fraction of or the whole portfolio and invest this amount in the bank which gives a higher rate of return. The owner can later re-buy the amount sold with the money from the bank and still have a profit, which is another case of arbitrage. The case with non-constant and nondeterministic \((*)\) and \(r\) are handled in the same way. Thus to avoid arbitrage possibilities it must hold that \((u_S \alpha + u_\Pi \alpha_\Pi) = r\). So by equation (2.14) and the fact that \((*) = r\),

\[
\frac{\sigma_\Pi \alpha}{\sigma_\Pi - \sigma} - \frac{\sigma \alpha_\Pi}{\sigma_\Pi - \sigma} = r
\]

\[
\Leftrightarrow \sigma_\Pi \alpha - \sigma \alpha_\Pi = r \sigma_\Pi - r \sigma.
\]

In using the expressions for \(\alpha_\Pi\) and \(\sigma_\Pi\) from equation (2.13) in the equation above gives

\[
\alpha \frac{\sigma F_S}{F} - \sigma \frac{F_t + \alpha S F_S + \frac{1}{2} \sigma^2 S^2 F_{SS}}{F} = r \frac{\sigma F_S}{F} - r \sigma,
\]

and multiplying this expression by \(F\), dividing by \(\sigma\) and rearranging results in

\[
Pricing a Multi-Asset American Option
Theorem 2.3. (Black-Scholes equation) In the market
\[ \begin{align*}
    dB(t) &= rB(t)dt, \\
    dS(t) &= \alpha(t,S(t))S(t)dt + \sigma(t,S(t))d\bar{W}(t),
\end{align*} \] (2.15)

the arbitrage-free price \( \Pi(t;X) \) of a \( T \)-claim \( X = \Phi(S(T)) \) at \( t < T \) is \( \Pi(t) = F(t,S(t)) \) where \( F(t,s) \) satisfies

\[ \begin{cases} 
    F_t + rsF_s + \frac{1}{2}\sigma^2(t,s)s^2F_{ss}(t,s) - rF(t,s) = 0, \\
    F(T,s) = \Phi(s).
\end{cases} \]

Remark 2.7. The local mean rate of return does not appear in the pricing equation.

The Black-Scholes equation in higher dimensions is given by:

Theorem 2.4. (The Multidimensional Black-Scholes equation) In the market model
\[ \begin{align*}
    dB(t) &= rB(t)dt, \\
    dS(t) &= \alpha_i(t,S(t))S_i(t)dt + S_i(t)\sum_{j=1}^{n} \sigma_{ij}(t,S(t))d\bar{W}_j(t), \quad i = 1, \ldots, n
\end{align*} \]

where \( \bar{W}_1, \bar{W}_2, \ldots, \bar{W}_n \) are \( n \) independent Brownian Motions and where the \( n \times n \) volatility matrix \( \sigma = (\sigma_{ij})_{i,j=1}^{n} \) is non-singular. The arbitrage-free price of a simple \( T \)-claim \( X = \Phi(S_1(T),S_2(T), \ldots, S_n(T)) \) at \( t < T \) is \( \Pi(t;X) = F(t,S_1(t),S_2(t), \ldots, S_n(t)) \) where \( F(t,s_1,s_2, \ldots, s_n) \) satisfies

Pricing a Multi-Asset American Option
The Black-Scholes equation

\[
F_t + r \sum_{i=1}^{n} s_i F_i + \frac{1}{2} \sum_{i,j=1}^{n} s_i s_j F_{ij} C_{ij} - rF = 0,
\]

where \( F_i = \frac{\partial F}{\partial s_i} (t, s), \) \( F_{ij} = \frac{\partial^2 F}{\partial s_i \partial s_j} (t, s) \) and \( C = (\sigma \sigma^*) (t, s). \)

2.3.1 Risk Neutral Valuation

According to Feynman-Kač the solution to the Black-Scholes equation is

\[
F(t, s) = e^{-r(T-t)} E_Q^t_s [\Phi(S(T))],
\]

where the stochastic process \( S \) satisfies the SDE

\[
\begin{cases}
\frac{dS(u)}{S(u)} = rS(u) du + \sigma(u, S(u)) S(u) dW(u), \\
S(t) = s,
\end{cases}
\]

(2.16)

where \( W \) is a Brownian Motion. The price process \( dX(t) = \alpha(t, X(t)) dt + \sigma(t, X(t)) d\bar{W}(t) \) is of the same form as in equation (2.16) besides the fact that the process in equation (2.16) has the short rate \( r \) as local mean rate of return whereas the process \( X \) has the local mean rate of return \( \alpha \). There is also a notational difference between the Brownian Motions in the processes discussed. It is acceptable to call the \( X \)-process for \( S \) as long there is no confusion between the real process \( S \) defined in equation (2.15) and the new \( S \) process. This can be done by defining the probability measure which governs the real model (as in equation (2.15)) by the letter \( P \) and defining another probability measure \( Q \) under which the \( S \) process has a different probability distribution. Thus the \( P \)-dynamics of the \( S \)-process are that of equation (2.15), whereas the \( Q \)-dynamics of \( S \) are defined as

\[
\frac{dS(u)}{S(u)} = rS(u) du + \sigma(u, S(u)) S(u) dW(u),
\]

(2.17)

where \( W \) is a \( Q \)-Wiener process. (Recall that a Wiener process is an alternative name for a Brownian Motion.) So by the discussion above \( E \) denotes expectations taken under the \( P \)-measure whereas \( E_Q \) denotes expectations taken under the \( Q \)-measure. Furthermore \( \bar{W} \) denotes a \( P \)-Wiener process whereas \( W \) denotes \( Q \)-Wiener process. The following theorem concludes what has been discussed so far.

**Theorem 2.5.** The arbitrage free price of a \( T \)-claim \( \mathcal{X} = \Phi(S(T)) \) is \( \Pi(t; \mathcal{X}) = F(t, s) \) where

\[
F(t, s) = e^{-r(T-t)} E_Q^t_s [\Phi(S(T))],
\]

where the \( Q \)-dynamics are those of equation (2.17) and where the subscripts \( t, s \) on the expectation operator, as stated earlier, emphasizes that the expected value is to be taken given the initial condition \( S(t) = s. \)

The multidimensional model of the theorem above is

Pricing a Multi-Asset American Option
Theorem 2.6. Given the price process $S$ with $Q$-dynamics given by
\[ dS_i(t) = rS_i(t)dt + \sigma_iS_i(t)dW(t), \quad i = 1, \ldots, n, \]
the pricing function $F(t, s)$ has the following representation
\[ F(t, s) = e^{-r(T-t)}E^Q_{t,s}[\Phi(S(T))]. \]

2.3.2 The Black-Scholes formula

Given the Black-Scholes model for the financial market, see Definition 2.6,
\[ dB(t) = rB(t)dt, \quad dS(t) = \alpha(t, S(t))S(t)dt + \sigma(t, S(t))S(t)d\bar{W}(t), \]
where $r$, $\alpha$ and $\sigma$ are deterministic constants the arbitrage free price, see previous section, of a simple claim $\Phi(S(T))$ is given by
\[ F(t, s) = e^{-r(T-t)}E^Q_{t,s}[\Phi(S(T))]. \]
where the $Q$-dynamics of $S$ are given by
\[
\begin{cases}
  dS(u) = rS(u)du + \sigma(u, S(u))S(u)dW(u), \\
  S(t) = s.
\end{cases}
\]
Taking into account another assumption besides those presented in the beginning of Section 2.3, the assumption that
- The option is European, i.e. it can only be exercised at the expiration (maturity) date,
results in, following [7] and [15], the famous Black-Scholes formula given below. The payoff (contract) function associated with a European vanilla call option (standard call option) is $\Phi(S(T)) = (S(t) - K)^+$, where $(\cdot)^+ := \max(\cdot, 0)$.

Proposition 2.5. (The Black-Scholes formula) The price of a European call option with strike price $K$, time of maturity $T$ and payoff function $\mathcal{X} = \Phi(S(t)) = (S(t) - K)^+$ is $\Pi(t; \mathcal{X}) = F(t, S(t))$, where
\[ F(t, s) = sN[d_1(t, s)] - e^{-r(T-t)}KN[d_2(t, s)]. \]
Here $N$ is the cumulative distribution function of $N[0, 1]$ and
\[
\begin{align*}
d_1(t, s) &= \frac{\ln \frac{s}{K} + \left( r + \frac{1}{2}\sigma^2 \right)(T-t)}{\sigma \sqrt{T-t}}, \\
d_2(t, s) &= \frac{\ln \frac{s}{K} + \left( r - \frac{1}{2}\sigma^2 \right)(T-t)}{\sigma \sqrt{T-t}} = d_1(t, s) - \sigma \sqrt{T-t}.
\end{align*}
\]
Proof. By Proposition 2.4 the solution to the equation
\[ \begin{cases} dS(u) = rS(u)du + \sigma(u, S(u))S(u)dW(u), \\ S(t) = s, \end{cases} \]
is given by
\[ S(u) = se^{(r-\frac{1}{2}\sigma^2)(u-t)+\sigma(W(u)-W(t))}. \]
With the last equation in mind and using risk neutral evaluation gives
\[
F(t, s) = e^{-r(T-t)}E_{\mathbb{Q}^T}^Q[(S(T)-K)^+]
= e^{-r(T-t)}\int_{-\infty}^{\infty} \left( se^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}y} - K \right)^+ \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy
= e^{-r(T-t)}\int_{-d_2}^{\infty} \left( se^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}y} - K \right)^+ \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy
= \int_{-d_2}^{\infty} se^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}y} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy - Ke^{-r(T-t)}\int_{-d_2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy
= sN[d_1(t, s)] - e^{-r(T-t)}KN[d_2(t, s)].
\]

2.4 European and American options

2.4.1 European and American vanilla options

There are different types of options traded in the financial market such as stock options, foreign currency options, index options and futures options. Here the primary concern is with stock options. According to [33] there are four features associated with the details of an option—the description of the asset, whether a call or a put, the exercise price and the expiration date (including whether American or European in style, a distinction that has nothing to do with geographical location). A final and separate feature is the price of the option itself, i.e. the premium of the option. The notation style vanilla option is used when referring to a standard option of the specified style. In the beginning of Section 1 some of the features such as: European- or American-style call/put options were defined and can be revisited. It is worth pointing out that there are two sides to any option: the side that grants or issues the option is said to write the option and the side that obtains the option is said to buy the option. The only risk that the purchasing party faces is the price, i.e. the original premium, of the option whereas the writing party may face a large loss. The latter party must buy or sell the underlying asset at the specified terms if the option is exercised. In the case of an exercised call option the underlying asset must be bought in the market if not already owned by the writer in order to deliver it at the specified strike price. The current market price might be much higher than the specified strike price. In the case of an exercised put option the underlying asset must be bought by the writer at the specified strike price, a price that might be much higher than the current market price.
2.4.2 Early exercise on a non–dividend–paying stock

Following the discussion in [22] about American options some properties regarding the possibility of early exercise are worth taking into consideration. It is never optimal to exercise an American call option on a non–dividend–paying stock before the expiration date. The following example illustrates the intuition of the previous statement. Consider an American call option on a non–dividend–paying stock with its strike price $K$ when the underlying stock is worth more than $K$ today and the time to expiration is one month. If the option is deep in the money, i.e. if the strike price is significantly below the market price of the underlying asset, the owner of the option may be tempted to exercise the option. If the owner of the option is determined to hold on to the stock, that is obtained by exercising the option, for more than one month the best strategy is to keep the option and exercise it at the end of the month. In taking this course of action the strike price $K$ is paid out one month later which gives the opportunity to earn interest on the amount $K$ for that month and no income is sacrificed since the stock does not pay any dividends. Another argument for not exercising early is the fact that there is some chance for the stock price to fall below the strike price in which case the option will not be exercised. In the case where the owner of the option thinks that the stock is overpriced and wants to sell the stock in the market by exercising the option, it is better to sell the option. The option will be bought by an investor that wants to hold on to the stock and such investors must exist or the current stock price would not be what it is today.

A more formal argument uses the following equation $c \geq \max(S_0 - Ke^{-rT}, 0)$ giving a lower bound (besides the value zero in which case the option expires worthless) for the European call option value on a non–dividend–paying stock:

$$c \geq S_0 - Ke^{-rT}$$

where $c$ is the value of an European call option to buy one share, $S_0$ is the current stock price, $T$ is time to expiration of the option and $r$ is the continuously compounded risk-free rate of interest for an investment maturing in time $T$. It is assumed that $r > 0$ due to the fact that an risk-free investment would otherwise not have advantages over cash.

Proof. Consider two portfolios $A$ and $B$ consisting of:

Portfolio A: One European call option plus an amount of cash equal to $Ke^{-rT}$.

Portfolio B: One share.

Considering portfolio $A$ and assuming that the cash is invested at the risk-free interest $r$, the cash will grow to $K$ in time $T$. If the stock price at $T$, denoted by $S_T$, is greater than $K$ the option is exercised and the portfolio is worth the amount $S_T$. If on the other hand $S_T < K$, the call option expires worthless and the portfolio is worth the amount $K$. Thus at time $T$ portfolio $A$ is worth $\max(S_T, K)$. In considering portfolio $B$’s value it is $S_T$ at time $T$. Thus it can be deduced that portfolio $A$’s value is at least equal to or greater than the value of portfolio $B$ at expiration. Since the assumption that there are no arbitrage possibilities on the financial market where trading takes place must also hold today,

$$c + Ke^{-rT} \geq S_0$$

$$\iff c \geq S_0 - Ke^{-rT}.$$
The owner of an American call option, compared to the corresponding European call option, has the right to exercise at any time during the lifetime of the option and thus,

\[ C \geq c \]
\[ \iff C \geq S_0 - Ke^{-rT}, \]

where \( C \) denotes the value of the American call option. The assumption that \( r > 0 \) results in \( C > S_0 - K \) and if it was optimal to exercise early \( C \) would be equal to \( S_0 - K \), thus it can be deduced that it can never be optimal to exercise an American call option on a non–dividend–paying–stock. In Figure 2 it can be seen that the call option price is always above its intrinsic value (the payoff function) of \( \max(S_0 - K, 0) \), where the intrinsic value function is represented by the dotted line. The arrows indicate that as \( r \) or \( T \) or \( \sigma \) increases the call price moves upwards, i.e further away from the intrinsic value, for details see [22].

It can on the other hand be optimal to exercise an American put option on a non–dividend–paying stock before the expiration date. A put option held together with the stock provides insurance against the stock price falling below a certain level. At any given time during its lifetime the option should be exercised if it is sufficiently deep in the money. The following example illustrates the intuition of the last statement. Consider an American put option on a non–dividend–paying stock with strike price \( K \) and where the stock price is almost zero. The owner of the put can make the profit \( K \) by exercising immediately and by waiting the profit might become lesser than \( K \) since stock prices are never negative. Thus a major difference between an American call option of the type discussed, i.e on a non–dividend–paying stock, and an American put of the same type is that it can be optimal to exercise the put during its lifetime to realize the strike price immediately. In general as \( S_0 \) decreases, \( r \) increases while \( \sigma \) decreases and the early exercise of a put option becomes more attractive.

![Figure 2: Variation of price of an American or European call option on a non–dividend–paying stock with stock price \( S_0 \).](image)

A more formal argument uses the following equation \( p \geq \max(Ke^{-rT} - S_0, 0) \), which this gives a upper bound (besides the value zero in which case the option expires worthless) for the
European put option value on a non-dividend-paying stock:

\[ p \geq Ke^{-rT} - S_0, \]

where \( p \) is the value of an European put option to buy one share, \( S_0 \) is the current stock price, \( T \) is time to expiration of the option and \( r \) is the continuously compounded risk-free rate of interest for an investment maturing in time \( T \). It is assumed that \( r > 0 \) due to the fact that an risk-free investment would otherwise not have advantages over cash. A similar proof to the previous one is given below.

**Proof.** Consider two portfolios \( A \) and \( B \) consisting of:

*Portfolio A:* One European put option plus one share.

*Portfolio B:* An amount of cash equal to \( Ke^{-rT} \).

The option in portfolio \( A \) is exercised at maturity if \( S_T < K \) and the value of this portfolio becomes \( K \). On the other hand if \( S_T > K \) the option expires worthless and the value of \( A \) becomes \( S_T \). Thus the value of \( A \) is \( \max(S_T, K) \) at expiration, i.e. at time \( T \). Considering portfolio \( B \) and assuming that the cash is invested at the risk-free interest \( r \), the cash invested will grow to \( K \) in time \( T \). Thus it can be deduced that portfolio \( A \)'s value is at least equal to or greater than the value of portfolio \( B \) at expiration. Since the assumption that there are no arbitrage possibilities on the financial market where trading takes place must also hold today,

\[ p + S_0 \geq Ke^{-rT} \]

\[ \Leftrightarrow p \geq Ke^{-rT} - S_0. \]

Figure 3: Variation of price of an American put option on a non–dividend–paying stock with stock price \( S_0 \).

*Pricing a Multi-Asset American Option*
The owner of an American put option, compared to the corresponding European put option, has the right to exercise at any time during the lifetime of the option and due to the fact that immediate exercise is always possible a stronger condition than \( P \geq p \geq Ke^{-rT} - S_0 \), where \( P \) is the price of an American put, holds

\[
P \geq K - S_0.
\]

In Figure 3 it is shown how the price of an American put option on a non–dividend–paying stock varies with \( S_0 \). Given \( r > 0 \) it is always optimal to exercise the option when the stock price is sufficiently low. This happens when the value of the option is its intrinsic value \( K - S_0 \) and this is why the curve representing the value of the put merges with the put’s intrinsic value at point \( A \) for a sufficiently small \( S_0 \). The arrows indicate how the line relating the put price to the stock price moves when \( r \) decreases, or \( T \) increases, or \( \sigma \) increases.

When dividends are expected it is sometimes optimal to exercise an American call option immediately prior to an ex–dividend date. Thus the statement that an American call option will not be exercised early no longer holds if the underlying stock pays dividends, for more details see [22].

### 2.5 The American option pricing problem

From the discussion in the previous section the pricing problem treated here will mainly focus on the valuation of the American put option. It should further be stated that for simplicity only the one-dimensional case is considered in this section, i.e. the option has only one stock as underlying asset. Following [49] and [52] on the discussion regarding the valuation of American options it is stated that the valuation of American options is more complicated than the valuation of European options due to the property of early exercise. There must be some values of \( S \) for which it is optimal for the holder of the option to exercise it otherwise the option would have the European value and the Black-Scholes equation would hold for all \( S \). At each time besides determining the option value, for each value of \( S \), it must be determined if the option should be exercised resulting in a free boundary problem. At each time \( t \) there is particular value of \( S \) that marks the boundary between two regions: to one side the option should be held and to the other side the option should be exercised. This dividing price is called the optimal exercise price and will be denoted by \( b(t) \). There is just one free boundary \( S = b(t) \) for the American put option. For values of \( S \) where \( S < b(t) \) the option should be exercised and for values of \( S \) where \( b(t) < S \) the option can be held. It should however be noted that in the more general case considering American options there may be several exercise prices, separating ranges where the options should be exercised from ranges where the option should be held, but this issue is not discussed here. The value of \( b(t) \) is not known a priori as a function of time nor is it known where to apply this boundary condition and \( b(t) \) is thus called a free boundary for the associated Black-Scholes equation.

#### 2.5.1 American put options as a free boundary problem

Considering the American put option with value \( V(t,x) \) and exercise boundary \( x = b(t) \) there are some possibilities for the delta of the option, i.e \( \Delta = \frac{\partial V}{\partial x} \), that need be considered. (It should
be noted that in the theory developed earlier the arbitrage free price of an option is denoted by $F(t, s)$ and it is only for convenience that the asset price $S$ is replaced by $x$ and similarly for the notation of the value function.) Assuming that $b(t) < K$ the slope of the payoff function $(K - x)^+$, where as earlier $(\cdot)^+ := \max(\cdot, 0)$, at the contact point with $b(t)$ is $-1$. This gives three possibilities for the delta of the option at $x = b(t)$ in assuming that $\partial V / \partial x$ does exist and is continuous, i.e. that $x \mapsto V(x, t)$ is $C^1$ at $x = b(t)$ and it can be shown that it does, see Lemma 2.2 or [41] for proof of existence and continuity. These three possible boundary conditions are:

\begin{itemize}
  \item $\partial V / \partial x < -1$;
  \item $\partial V / \partial x > -1$;
  \item $\partial V / \partial x = -1$.
\end{itemize}

In the first case $\partial V / \partial x < -1$ which indicates that as $x$ increases from $b(t)$ the value function of the option $V(x, t)$ drops below the payoff function $(K - x)^+$ since its slope is more negative than the slope of the payoff function. By arbitrage arguments, see previous Section 2.4.2, this is not possible and the case $\partial V / \partial x < -1$ can be discarded.

The second case $\partial V / \partial x > -1$ is discussed throw the strategy adopted by the holder of the option. The exercise strategy, as discussed in [52], focuses on the decision the holder must make on how far $x$ should fall before exercising the option in order to maximize the profit of the exercise. It is further argued that the correct boundary condition is given in the third case, i.e. $\partial V / \partial x = -1$ at $x = b(t)$, where the benefit is maximized and arbitrage is avoided. This boundary condition is called smooth fit.

In recalling that the Black-Scholes equation follows from an arbitrage argument this argument is only partially valid for American options. The arbitrage argument used for the European option no longer leads to a unique value for the return on the portfolio but to an inequality. It can only be said that the return of the portfolio cannot exceed the return from a bank deposit which gives the following PDE for an American put option:

$$\frac{\partial V}{\partial t}(t, x) + \frac{1}{2}\sigma^2(t, x)x^2\frac{\partial^2 V}{\partial x^2}(t, x) + r x \frac{\partial V}{\partial x}(t, x) - r V(t, x) \leq 0.$$ 

From the discussion above the American put option problem is written as a free boundary problem as follows. For each time $t$ the $x$ axis is divided in two distinct regions of which the first, for values $0 \leq x < b(t)$, is where early exercise is optimal and

$$V(t, x) = K - x, \quad \frac{\partial V}{\partial t}(t, x) + \frac{1}{2}\sigma^2(t, x)x^2\frac{\partial^2 V}{\partial x^2}(t, x) + r x \frac{\partial V}{\partial x}(t, x) - r V(t, x) < 0.$$ 

The second region, for values $b(t) < x < \infty$, early exercise is not optimal and

$$V(t, x) > K - x, \quad \frac{\partial V}{\partial t}(t, x) + \frac{1}{2}\sigma^2(t, x)x^2\frac{\partial^2 V}{\partial x^2}(t, x) + r x \frac{\partial V}{\partial x}(t, x) - r V(t, x) = 0.$$ 

The boundary conditions at $x = b(t)$ are that $V(t, x)$ and its delta are continuous:

$$V(t, b(t)) = (K - b(t))^+, \quad \frac{\partial V}{\partial x}(t, b(t)) = -1.$$ 

---

**Pricing a Multi-Asset American Option**
Remark 2.8. It should be noted that the smooth fit condition is not implied by the fact that \( V(t, b(t)) = K - b(t) \) since it is not known \textit{a priori} where \( b(t) \) is. The extra condition needed for determining its location is that the gradient of \( V \) should be continuous which follows from arbitrage arguments.

The smooth fit condition above was derived intuitively and for a mathematical derivation of the proof, following [37], there is need for some definitions and lemmas.

\begin{definition}
Excessive functions. A non-negative \( \mathcal{F} \) measurable function \( f \) is \( \lambda \)-excessive provided \( P^t_\lambda f \leq f \) for all \( t \geq 0 \) and \( P^t_\lambda \to f \) as \( t \to 0 \) pointwise. (\( \mathcal{F} \) denotes a \( \sigma \)-algebra, for details see [17].) The notation \( P \) is a measure with the property
\[
P_t f(x) = \int P_t(x, dy) f(y) = E^x f(X_t),
\]
and more generally for any \( \lambda > 0 \), \( P^t_\lambda f(x) \) is defined as
\[
P^t_\lambda f(x) = e^{-\lambda t} P_t f(x) = E^x [e^{-\lambda t} f(X_t)],
\]
where the superscript on the expectation operator denotes that \( X_0 = x \).
\end{definition}

\begin{lemma}
The American put value function \( V \) is \( r \)-excessive to the space-time stock price process and therefore on \([0, T] \times \mathbb{R}_+\),
\[
\mathbb{L}[e^{-rt} V(t, x)] \leq 0,
\]
in the sense of Schwartz distributions and where \( \mathbb{L} \) is the Kolmogorov backward operator of the Geometric Brownian Motion in equation (2.12) given by
\[
\mathbb{L} = r x \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial t}.
\]
\end{lemma}

\textbf{Proof.} In showing that \( V(t, x) \) is \( r \)-excessive, i.e. that Definition 2.7 is fulfilled for \( V(t, x) \), equation (2.18) is implied since any excessive function is the increasing limit of a sequence of infinitely differentiable excessive functions. This can be done by first showing that, for every \( t \in [0, T] \),
\[
V(0, x) \geq E^Q_{0, t} [e^{-rt} V(t, X_t)].
\]
Choose \( \varepsilon > 0 \) and a stopping time \( \tau_\varepsilon \), see next section on \textit{stopping time}, from the set
\[
\{ \tau \in \mathcal{T}_{t,T} | E^Q [e^{-r(t-\tau)} (K - X_\tau)^+ | X_t] \geq V(t, X_t) - \varepsilon \}.
\]

The set chosen is necessarily nonempty for \( t \in [0, T) \). Then, letting \( E^Q_{0,x} \) denote expectation conditional on \( X_0 = x \),
\[
E^Q_{0,x} [e^{-r\tau_\varepsilon} (K - X_{\tau_\varepsilon})^+] = E^Q_{0,x} [e^{-r\tau_\varepsilon} E^Q [e^{-r(t-\tau)} (K - X_\tau)^+ | X_t]]
\geq E^Q_{0,x} [e^{-rt} V(t, X_t)] - \varepsilon e^{-rt}.
\]

\textit{Pricing a Multi-Asset American Option}
But, for any stopping time $\tau$, 

$$V(0, x) \geq E^Q_{0,\lambda}\left[e^{-r\tau}(K - X_\tau)^+\right].$$

so, 

$$V(0, x) \geq E^Q_{0,\lambda}\left[e^{-r\tau}V(t, X_\tau)\right] - \varepsilon e^{-rt}.$$ 

Since $\varepsilon$ is arbitrary and letting $\varepsilon \downarrow 0$ gives the desired result. Thus the first condition in Definition 2.7, i.e. that $P^t f \leq f$ for all $t \geq 0$, is shown to hold. The second condition, i.e. that $P^t f \to f$ as $t \to 0$ pointwise, is obvious and it can be concluded that the value function $V(t, x)$ is $r$-excessive.

Following [37] the lemma given below can now be proven

**Lemma 2.2.** *(Smooth fit)*. $V_x$ is continuous almost everywhere across the stopping boundary $b$. That is, for almost every $t \in [0, T]$, 

$$\lim_{x \downarrow b(t)} V_x(t, x) = -1.$$ 

**Proof.** Since the value function of the American put option is $r$-excessive and by Lemma 2.1 

$$\mathbb{L}\left[e^{-rt}V(t, x)\right] \leq 0, \quad (t, x) \in [0, T] \times \mathbb{R}_+.$$ 

By introducing the variable $\xi := \ln(x)$ and letting $\widetilde{V}(t, \xi) := V(t, \xi(x))$ the above equation can be written as 

$$\frac{1}{2}\sigma^2 \widetilde{V}_{\xi\xi} \leq -\left(r - \frac{1}{2}\sigma^2\right)\widetilde{V}_\xi - \widetilde{V} + r \widetilde{V}.$$ 

Figure 4: Integration across the optimal stopping boundary.
Choosing $\varepsilon > 0$ and integrating the above expression on a region $\Sigma$ of width $2\varepsilon$ over the log stopping boundary $\xi^\ast$, defined by $\xi^\ast := \ln(b)$, from $t_1$ to $t_2$ as in Figure 4 gives
\[
\int_{t_1}^{t_2} \frac{1}{2} \sigma^2 \left[ \widetilde{V}_\xi(t, \xi^\ast + \varepsilon) - \widetilde{V}_\xi(t, \xi^\ast - \varepsilon) \right] dt \\
\leq - \int_{t_1}^{t_2} \left( r - \frac{1}{2} \sigma^2 \right) \left[ \widetilde{V}(t, \xi^\ast + \varepsilon) - \widetilde{V}(t, \xi^\ast - \varepsilon) \right] dt - \int_\Sigma [\tilde{V}_t + r \tilde{V}] d\xi dt.
\]
Defining the horizontal strips of $\Sigma$ as $\Sigma_\xi$ which start at time $t_1 - (\xi)$ and end at time $t_2 + (\xi)$ the last equation becomes
\[
\int_{t_1}^{t_2} \frac{1}{2} \sigma^2 \left[ \widetilde{V}_\xi(t, \xi^\ast + \varepsilon) - \widetilde{V}_\xi(t, \xi^\ast - \varepsilon) \right] dt \\
\leq - \int_{t_1}^{t_2} \left( r - \frac{1}{2} \sigma^2 \right) \left[ \widetilde{V}(t, \xi^\ast + \varepsilon) - \widetilde{V}(t, \xi^\ast - \varepsilon) \right] dt \\
- \int_\Sigma \left[ \tilde{V}(t^+, \xi) - \tilde{V}(t^-, \xi) \right] d\xi + \int_\Sigma r \tilde{V} d\xi dt.
\]
Since $\varepsilon$ is arbitrary and as $\varepsilon \downarrow 0$, by dominated convergence and since $\widetilde{V}_\xi = -e^\xi$ on the stopping set $\bar{D}$, see next section for definition of $\bar{D}$,
\[
\int_{t_1}^{t_2} \lim_{\xi \downarrow \xi^\ast} \widetilde{V}_\xi + e^\xi \right] dt \leq 0.
\]
Since the spatial derivative of $V$ is bounded from below, i.e $V_x(t, x) \in [-1, 0]$ for $x \in \bar{D}$ (see Lemma 4.1 in [37]),
\[
\widetilde{V}_\xi \geq -e^\xi,
\]
Plugging in $\xi = \ln(x)$, the fact that $\frac{\partial \xi}{\partial x} = \frac{1}{x}$ and using the chain rule on the expression above yields
\[
\frac{\partial V}{\partial x} = \frac{\partial \widetilde{V}}{\partial \xi} \cdot \frac{\partial \xi}{\partial x} \geq -e^\xi \cdot \frac{\partial \xi}{\partial x} \\
\iff \frac{\partial V}{\partial x} \geq -e^{\ln x} \cdot \frac{1}{x} = -1.
\]
By the same arbitrage arguments given in the beginning of this section it can be concluded that the smooth fit condition must hold, i.e that equality must hold for the delta of the value function $V$ across the optimal stopping boundary $b(t)$.

### 2.5.2 Optimal stopping problem for American put options

Following [41] the problem of valuing the American put option with finite horizon, i.e. $t \in [0, T]$, is equivalent to that of solving the optimal stopping problem:
\[
V(t, x) = \sup_{\tau \in \mathcal{T}_x} E^{Q}_{t,x} \left[ e^{-r(\tau - t)} \psi(\tau, X_\tau) \right].
\]

---

*Pricing a Multi-Asset American Option*
where $\psi(\tau, X_\tau) = (K - X_\tau)^+$ is the payoff function at $\tau$ and where $\tau$ is a stopping time of the Geometric Brownian Motion $X$ solving the SDE

$$\begin{align*}
\begin{cases}
  dX_t = rX_t dt + \sigma X_t dW_t \\ 
  X_0 = x.
\end{cases}
\end{align*}$$

It can be concluded that a stopping time exists for problem (2.19), for details see [41], where the continuation set $C$ and the stopping set $\bar{D}$ equals

$$C = \{(t, x) \in [0, T) \times (0, \infty) : V(t, x) > \psi(x)\},$$

$$\bar{D} = \{(t, x) \in [0, T) \times (0, \infty) : V(t, x) = \psi(x)\},$$

which means that the stopping time $\tau_0$ defined by

$$\tau_0 = \inf\{0 \leq s \leq T : X_s \in \bar{D}\}$$

is optimal in (2.19).

The optimal stopping problem discussed is equivalent to the problem discussed in the previous section, where it was deduced that the American put option pricing problem can be written as a free boundary problem. In following [41] the free boundary problem can be rewritten in a more suitable form:

$$V_t + \mathbb{L}_X V = rV \quad \text{in } C,$$

$$V(t, x) = (K - x)^+ \quad \text{for } x = b(t),$$

$$V_s(t, x) = -1 \quad \text{for } x = b(t) \quad \text{(smooth fit)},$$

$$V(t, x) > (K - x)^+ \quad \text{in } C,$$

$$V(t, x) = (K - x)^+ \quad \text{in } D,$$

where now the continuation set $C$ equals

$$C = \{(t, x) \in [0, T) \times (0, \infty) : x > b(t)\}$$

and the stopping set $\bar{D}$ is the closure of the set

$$D = \{(t, x) \in [0, T) \times (0, \infty) : x < b(t)\},$$

joined with the remaining points $(T, x)$ for $x \geq b(T)$. The infinitesimal generator $\mathbb{L}_X$ for the Geometric Brownian Motion discussed is given by

$$\mathbb{L}_X = rx \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2}.$$

Given the free boundary problem above, a Wiener space $(\Omega, \mathcal{F}, P, \{\mathcal{F}_s\}_{s \geq 0}, W)$, denoting the solution to the SDE (2.12) starting at time $t$ from $x \in \mathbb{R}$ by $X^{t,x}$ and following [39] it is now possible to give the following theorem

**Theorem 2.7.** Let $V$ be a strong (unique) solution to the free boundary problem above such that the growth assumption

$$|V(t, x)| \leq Ce^{\lambda |x|^2}, \quad (t, x) \in (0, T) \times (0, \infty)$$

(2.20)
for some constants $C$, $\lambda$ with $\lambda$ sufficiently small so that the standard maximal estimate $E\left[e^{\lambda \sup_{t \in [0,T]} |X_t|}\right] < \infty$ holds. (The superscripts on the process $X$ emphasize the initial condition $X_t = x$.) Then

$$V(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} E\left[\psi(\tau, X^t_\tau) e^{-\int_\tau^T r(s, X^t_s) ds} - \int_\tau^T f(s, X^t_s) e^{-\int_\tau^s r(p, X^t_p) dp} ds\right],$$

(2.21)

where

$$\mathcal{T}_{t,T} = \{\tau \in \mathcal{T} | \tau \in [t, T] \text{ almost surely}\}$$

and $\mathcal{T}$ is the set of all stopping times with respect to the filtration $(\mathcal{F}_s)$. In particular, such a solution is unique.

In the theorem above it is assumed that $r$ and $f$ are non-constant functions typically representing the locally risk free interest rate and some transaction costs, respectively. Almost surely in the definition of the set $\mathcal{T}_{t,T}$ means with the probability of one. Thus the theorem describes a more general case than what is needed here. In recalling the Black-Scholes model, see Definition 2.6, the interest rate $r$ is assumed to be a deterministic constant. Furthermore by the fourth assumption made in the beginning of Section 2.3 it was assumed that there are no transaction costs regarding the option or the stock. Taking these assumptions under consideration and setting $f(s, X^t_s) = 0$ and $r(s, X^t_s) = r$, where $r > 0$ is a constant, in equation (2.21) gives

$$V(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} E\left[e^{-r(\tau-t)} \psi(\tau, X^t_\tau)\right].$$

Remark 2.9. It should be noted that the above equation is identical to equation (2.19) given in the beginning of this section besides notational differences. In recalling the notation for taking expectations under the $Q$-measure (see Section 2.3.1, given the SDE (2.12)),

$$\sup_{\tau \in \mathcal{T}_{t,T}} E\left[e^{-r(\tau-t)} \psi(\tau, X^t_\tau)\right] \Leftrightarrow \sup_{\tau \in \mathcal{T}_{t,T}} E^Q_{t,t}\left[e^{-r(\tau-t)} \psi(\tau, X^t_\tau)\right].$$

Before beginning the proof for the theorem above there is need for the lemma given below, following [39]

**Lemma 2.3.** The American put value is $r$-harmonic to the space-time stock price process on the continuation region and therefore on the set $C$:

$$\mathcal{L}\left[e^{-r \tau} V(t, x)\right] = 0,$$

where, as earlier, $\mathcal{L}$ is the Kolmogorov backward operator.

**Proof.** Following [39] it is for simplicity considered the case $r = 0$. The proof exploits a truncation and regularization technique to exploit the weak interior regularity properties of $V$ and it also uses Itô’s formula, where the formula can be applied directly due to the fact that a strong solution need not have the required regularity.

For $x \in B_R$, where $B_R = \{x \in \mathbb{R} | |x| < R\}$ and $R > 0$, the first exit time of $X^x$ from $B_R$ is denoted by $\tau_R$. For every $(t, x) \in [0, T] \times B_R$ and $\tau \in \mathcal{T}$ such that $t \leq \tau \leq \tau_R$ almost surely the following equation holds

$$V(t, x) = E\left[V(\tau, X^t_\tau) - \int_t^\tau \mathcal{L} V(s, X^t_s) \, ds\right],$$

(2.22)
where $\mathbb{L}$, as earlier, is the Kolmogorov backward operator of the Geometric Brownian Motion in equation (2.12), i.e. $\mathbb{L} = r x \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial t} - r$. As a first step the equation above is proven. For a fixed $\epsilon$ positive and suitably small this is done by considering the function $V^{\epsilon,R}$ on $\mathbb{R}$ with compact support such that $V^{\epsilon,R} = V$ on $[t, T - \epsilon[\times B_R]$. The regularizing sequence obtained by convolution of $V^{\epsilon,R}$ with the usual mollifiers is denoted by $(V^{\epsilon,R,n})_{n \in \mathbb{N}}$. Then, for any $p \geq 1$, $V^{\epsilon,R,n} \in \mathcal{S}^p(\mathbb{R})$, where $\mathcal{S}^p$ is a Sobolev space with properties given in [39], and

$$\lim_{n \to \infty} \|L V^{\epsilon,R,n} - L V^{\epsilon,R}\|_{L^p([t, T - \epsilon[\times B_R])} = 0. \quad (2.23)$$

Applying the Itô formula to the smooth function $V^{\epsilon,R,n}$ gives

$$V^{\epsilon,R,n}(\tau, X^{\epsilon,x}_\tau) = V^{\epsilon,R,n}(t, x) + \int_t^\tau \mathbb{L} V^{\epsilon,R,n}(s, X^{\epsilon,x}_x) \, ds + \int_t^\tau \sigma(s, X^{\epsilon,x}_x) \frac{\partial V}{\partial x}(s, X^{\epsilon,x}_x) \, dW_s \quad (2.24)$$

for $\tau \in \mathcal{T}$ such that $t \leq \tau \leq T \wedge (T - \epsilon)$ almost surely, where $a \wedge b := \min(a, b)$. Since $\sigma \frac{\partial V}{\partial x}$ is a bounded function on $[t, T - \epsilon[\times B_R$. Proposition 2.1 gives

$$E \left[ \int_t^\tau \sigma(s, X^{\epsilon,x}_x) \frac{\partial V}{\partial x}(s, X^{\epsilon,x}_x) \, dW_s \right] = 0.$$  

Moreover,

$$\lim_{n \to \infty} V^{\epsilon,R,n}(t, x) = V^{\epsilon,R}(t, x)$$

and by dominated convergence,

$$\lim_{n \to \infty} E[V^{\epsilon,R,n}(\tau, X^{\epsilon,x}_\tau)] = E[V^{\epsilon,R}(\tau, X^{\epsilon,x}_\tau)].$$

According to [39] the transition density of $X^{\epsilon,x}$ satisfies

$$\Gamma(t, x; \cdot, \cdot) \in L^\tilde{q}(\mathbb{R}) \quad (2.26)$$

for some $\tilde{q} > 1$. Using this fact and Hölder’s inequality, together with equation (2.23), it can be shown that the convergence of the $\mathcal{D}$-integral is, for complete proof and details on the convergence of this integral see [39],

$$\lim_{n \to \infty} E \left[ \int_t^\tau \mathbb{L} V^{\epsilon,R,n}(s, X^{\epsilon,x}_x) \, ds \right] = E \left[ \int_t^\tau \mathbb{L} V^{\epsilon,R}(s, X^{\epsilon,x}_x) \, ds \right].$$

Since $V^{\epsilon,R} = V$ on $[t, T - \epsilon[\times B_R$ and $\epsilon$ is arbitrary, equation (2.22) is proven. Next, since by Lemma 2.1 $\mathbb{L} V \leq 0$ almost everywhere and the law of $X^{\epsilon,x}$ is absolutely continuous with respect to the Lebesgue measure, the following holds

$$E \left[ \int_t^\tau \mathbb{L} V(s, X^{\epsilon,x}_x) \, ds \right] \leq 0$$

for any $\tau \in \mathcal{T}_{t,T}$ and from (2.22) it be concluded that

$$V(t, x) \geq E[V(\tau \wedge R, X^{\epsilon,x}_{\tau \wedge T})] \quad (2.25)$$

**Pricing a Multi-Asset American Option**
for any $\tau \in T_{t,T}$. Next as passing to the limit $R \to +\infty$, $\lim_{R \to +\infty} \tau \wedge \tau_R = \tau$ pointwise and by the growth assumption (2.20)

$$|V(\tau \wedge \tau_R, X^{t,x}_{\tau \wedge \tau_R})| \leq Ce^{-\lambda \sup_{\tau \in T_{t,T}} |X^{t,x}_{\tau}|^2}.$$ 

By the standard maximal estimate given in Theorem 2.7 the right-hand side of the estimate above is integrable and thus Lebesgue’s theorem can be applied. From equation (2.25) and as $R \to +\infty$ it can be deduced that

$$V(t, x) \geq E[V(\tau, X^{t,x}_{\tau})] \geq E[\psi(\tau, X^{t,x}_{\tau})].$$

This shows that

$$V(t, x) \geq \sup_{\tau \in T_{t,T}} E[\psi(\tau, X^{t,x}_{\tau})].$$

The proof is concluded by putting

$$\tau_0 = \inf\{s \in [t, T] | V(s, X^{t,x}_{s}) = \psi(s, X^{t,x}_{s})\}.$$ 

In using Lemma 2.3 which gives that $\mathbb{L}V = 0$ almost everywhere, where $V > \psi$, the following holds

$$E \left[ \int_{\tau_0 \wedge \tau_R} \mathbb{L}V(s, X^{t,x}_{s}) \, ds \right] = 0$$

and from equation (2.22) it can be deduced that

$$V(t, x) = E[V(\tau_0 \wedge \tau_R, X^{t,x}_{\tau_0 \wedge \tau_R})].$$

Repeating the previous argument to pass to the limit $R$, gives

$$V(t, x) = E[V(\tau_0, X^{t,x}_{\tau_0})] = E[\psi(\tau_0, X^{t,x}_{\tau_0})].$$

Remark 2.10. The proof of Theorem 2.7 above is easily extended to the $\mathbb{N}$-dimensional case by defining $B_R = \{x \in \mathbb{R} | |x| < R\}$ as $B_R = \{x \in \mathbb{R}^N | |x| < R\}$ and the Sobolev space $\mathcal{S}^p(\mathbb{R})$ as $\mathcal{S}^p(\mathbb{R}^{N+1})$. Furthermore in the equations given in the proof the operator $\mathbb{L}$ in the multidimensional case for a American multi-asset vanilla option is

$$\mathbb{L} = \frac{1}{2} \sum_{i,j=1}^{\mathbb{N}} x_i x_j C_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{\mathbb{N}} x_i \frac{\partial}{\partial x_i} + \frac{\partial}{\partial t} - r,$$

where as earlier $C = (\sigma \sigma^*) (t, x)$. An obvious extension, mentioned for completeness, is that the computational domain in the theorem discussed, i.e. in equation (2.20), should be $(t, x) \in (0, T) \times \mathbb{R}^{\mathbb{N}}$ in the $\mathbb{N}$-dimensional case.

2.5.3 Variational Inequality Formulation and Localization to a bounded domain

In following [37] it is possible to derive a variational characterization of the stopping problem discussed in the previous section. The variational inequality approach has the advantage over the free boundary problem formulation in the lack of need to introduce the stopping boundary $b(t)$. This gives an valuable advantage when considering the stopping of multidimensional processes. Before stating the one-dimensional variational inequality related to the computation of
the value function $V(t, x)$ in equation (2.19) there is need for introducing the weighted Sobolev space $H^{m, \mu}$ in Definition 2.9 below. But before defining $H^{m, \mu}$ the following definition, given in [55], is adequate

**Definition 2.8. (Sobolev spaces of integer order).** Considering functions $\nu : \Omega \to \mathbb{R}$ and where $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$ is a multi-index of modulus $|k| = \sum_{i=1}^d k_i$ then the notation

$$D^k f := \frac{\partial^{|k|} f}{\partial x_{k_1} \cdots \partial x_{k_d}}.$$ 

Let $L^p(\Omega)$, where $\Omega \subset \mathbb{R}^d (d = 1, \ldots, n)$, be the class of all measurable functions $\nu$ defined on $\Omega$ with bounded norm

$$\|\nu\|_{L^p(\Omega)} := \left\{ \begin{array}{ll} \left( \int_{\Omega} |\nu|^p \right)^{1/p} & \text{if } 1 \leq p \leq \infty \\ \text{ess sup}_{x \in \Omega} |\nu(x)| & \text{if } p = \infty \end{array} \right. \tag{2.26}$$

Sobolev spaces of integer order are normed spaces of functions with finite weak derivatives in the $L^p$-norm, i.e. for any nonnegative integer $m \in \mathbb{N}$ and $1 \leq p \leq \infty$, the space $W^{m,p}(\Omega)$ is defined to be the subset of $L^p(\Omega)$ such that $\nu$ and its weak derivatives up to order $m$ have finite $L^p(\Omega)$-norm. Thus the Sobolev space admits the natural norm given below and with the use of equation (2.26)

$$\|\nu\|_{W^{m,p}(\Omega)} := \left( \sum_{|\xi| \leq m} \|D^\xi \nu\|_{L^p(\Omega)}^p \right)^{1/p} = \left( \sum_{|\xi| \leq m} \int_{\Omega} |D^\xi \nu|^p \, dx \right)^{1/p}.$$ 

The $W^{m,2}(\Omega)$ space constitutes a Hilbert space and is thus is denoted by $\mathcal{H}^m(\Omega)$ and $W^{0,2}(\Omega)$ is identified with $L^2(\Omega)$.

**Definition 2.9. (Weighted Sobolev spaces $W^{m,2,\mu}$).** The weighted Sobolev space $H^{m,\mu}$, see Definition 2.8, is the set of measurable, real-valued functions $f$ on $\mathbb{R}$ whose distributional derivatives of all orders $\leq m$ belong to $L^2(\mathbb{R}, e^{-\mu|x|} \, dx)$, where $m$ denotes a nonnegative integer and $0 < \mu < \infty$. This space is given the norm

$$\|f\|_{H^{m,\mu}} := \left( \sum_{|\xi| \leq m} \int_{\mathbb{R}} |D^\xi f(x)|^2 \cdot e^{-\mu|x|} \, dx \right)^{1/2}.$$ 

The space $L^2([0, T] ; \mathcal{H}^{m,\mu})$ consists of the set of measurable functions $g : [0, T] \to \mathcal{H}^{m,\mu}$ such that $\int_{[0, T]} \|g(t)\|^2 \, dt < \infty$.

The theorem given below states the variational inequality related to the computation of the one-dimensional model.

**Theorem 2.8. (One dimensional (strong) variational inequality).** Suppose a continuous function $(t,x) \to f(t, x)$, defined on $[0, T] \times \mathbb{R}_+$, such that $f(t, e^x) \in L^2([0,T]; \mathcal{H}^{2,\mu})$ and $f(t, e^x) \in L^2([0,T]; \mathcal{H}^{0,\mu})$, satisfies the following system on $[0, T] \times \mathbb{R}_+$:

$$\frac{\partial f}{\partial t} + Af \leq 0, \quad f \geq \psi,$$

$$\left\{ \frac{\partial f}{\partial t} + Af \right\} \cdot (f - \psi) = 0, \text{ a.e. in } [0, T) \times \mathbb{R}_+.$$
where, in recalling that the payoff function $\psi$ for a American put option is $(K - x)^+$,

$$f(T, x) = \psi(T, x),$$

and where the notation

$$A := \frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2} + rx \frac{\partial}{\partial x} - r.$$

Then $f$ is unique and is the American put option value function $V$.

The above theorem was given for completeness and since the main interest here lies in the computation of a multidimensional model it is natural to give the multidimensional variational characterization of the optimal stopping problem. In following [34], $m$ independent exogeneous sources of uncertainty are considered represented by the $m$-dimensional Brownian Motion $\{B_t\}_{t \geq 0}$, defined on the Wiener space $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t \geq 0}, B)$. Without any loss of generality and for convenience it is supposed that the stochastic process $X_t \in \mathbb{R}^n$ satisfies the SDE

$$dX_t = \mu(X_t, s) ds + \sigma(X_t, s) dB_t,$$

for $s > t$ and where $x$ is deterministic. For completeness there is need for the additional assumptions given below.

(A1) The drift vector $\mu : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^n$ is $C^1$ and has bounded derivatives.

(A2) The volatility matrix $\sigma : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^{n \times m}$ is $C^{2,1}$, bounded and has bounded derivatives.

(A3) The diffusion matrix is the $(n \times n)$-matrix $a(x, t) := \sigma(x, t)\sigma(x, t)^*$ where the coercivity (positive definite) condition holds, i.e. there exists a constant $\eta > 0$ such that

$$\frac{\xi^* a \xi}{\|\xi\|^2} \geq \eta, \quad \forall \xi \in \mathbb{R}^n - \{0\} \text{ and each } (x, t) \in \mathbb{R}^n \times (0, \infty).$$

(A4) The payoff function $\psi \in L^p([0, T]; W^{2,p')}$ is given the restrictions that it is nonnegative and bounded, where $p > \frac{q}{2} + 1$ and $W^{2,p'}$ denotes the weighted Sobolev space given in Definition 2.10 below.

For there to exist a unique $t$-strong solution to equation (2.27) the functions $\mu$ and $\sigma$ need to satisfy the Lipschitz condition, see Proposition 5.1 in [7],

$$\|\mu(x, t) - \mu(t, y)\| + \|\sigma(x, t) - \sigma(y, t)\| \leq K\|x - y\|,$$

for every $0 \leq t < \infty$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$ and for some constant $K > 0$. 

---

**Pricing a Multi-Asset American Option**
Remark 2.11. For all $x \in \mathbb{R}$, (2.27) is taken under the $P$-measure, where, in recalling that the payoff function $\psi$ for an American put option is $(K - x)^+$, 

$$V(T, x) = \psi(T, x),$$

(2.28c)

Thus Theorem 2.9 describes a general variational characterization of the stopping problem. Here the interest lies in pricing American put options in the Black-Scholes-Merton model under the $Q$-measure. In recalling the Black-Scholes model for the financial market, see Definition 2.6, the $Q$-dynamics of the multidimensional price process associated with the risky assets given

\[ V(x, t) = \sup_{\tau \in T, \mathbb{R}} E_x \left[ \psi(X_{x,t}^\tau, \tau) e^{-\int_0^\tau a_0(X_{x,t}^\tau, s) ds} \right]. \]
in **Theorem 2.6** and the multidimensional Black-Scholes equation stated in **Theorem 2.4**, it can be concluded that here \( a_{ij} := \sigma_j(x,t)\sigma_i(x,t)^* \), \( a_0 = r \) and \( \mu_j := r \). These definitions can only be made after first doing a log normal transformation of the stochastic process in equation (2.27) to the dimensionless variables say \( Y_i = \ln \left( \frac{X_i}{\mathcal{F}} \right) \) where \( K \) is the strike of the option.

It is possible to simplify the variational inequality stated in **Theorem 2.9** with the introduction of the time-value function \( u(x,t) = V(x,t) - \psi(x,t) \). The function \( u \) is the excess option premium over the payoff and is always nonnegative for American options and the variational inequality (2.28) can be reformulated as

\[
\frac{\partial u}{\partial t} - \mathcal{A}u - a_0u - f \geq 0, \quad u \geq 0, \quad \text{(2.29a)}
\]

\[
\begin{aligned}
\left\{ \frac{\partial u}{\partial t} - \mathcal{A}u - a_0u - f \right\} \cdot u = 0, \\
\text{(2.29b)}
\end{aligned}
\]

with the inhomogeneous term \( f = - \frac{\partial \psi}{\partial t} + \mathcal{A}\psi + a_0\psi \) and the initial homogeneous condition

\[ u(x,0) = 0, \quad \text{(2.29c)} \]

for \( t \in (0,T) \) and all \( x \in \mathbb{R}^n \).

The state space is unbounded in **Theorem 2.9**, i.e. \( \Omega = \mathbb{R}^n \), and according to [34] in order to be able to solve the problem numerically there is need for localizing the computational domain to a bounded computational domain. The solution \( u \) may be realized as the limit of a the pointwise convergent sequence \( u_k \rightarrow u \), see **Definition 2.11** below, obtained through approximating (2.28) on increasing bounded domains that exhaust the state space \( \Omega \).

**Definition 2.11.** Given functions \( f : X \rightarrow Y \), where the set \( X \) is the domain of \( f \) and the set \( Y \) is the codomain of \( f \), a sequence \( \{ f_n \}_{n=1}^{\infty} \) of functions \( X \rightarrow Y \) is said to be **pointwise convergent** to another function \( f \) if \( \lim_{n \rightarrow \infty} f_n(x) = f(x) \forall x \in X \). This is usually denoted by \( f_n \rightarrow f \).

Now let \( \{ \Omega_k \}_{k=1}^{\infty} \) denote a sequence of increasing bounded open domains such that \( \bigcup_{k=1}^{\infty} \Omega_k = \mathbb{R}^n \). For the localized problem, by **Proposition 4.1** in [23], it then follows that

\[
\max_{t \in [0,T]} ||u_k(x,t) - u(x,t)||_{L^\infty(G)} \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty, \quad \text{(2.30)}
\]

for any compact set \( \overline{G} \subset \Omega \) such that \( \overline{G} \subset \Omega_k \) for all \( k \) and for essentially arbitrary choice of artificial boundary conditions. Here \( \overline{G} \subset \Omega_k \) is referred to as the **approximation domain** and \( \Omega_k \subset \mathbb{R}^n \) as the **computation domain**. Intuitively equation (2.30) says that the asymptotic behavior, i.e \( u_k |_{\partial \Omega_k} \), cannot affect the solution in any fixed bounded region within a finite interval of time. More precisely, the solution near the distant boundary \( \partial \Omega_k \) does not affect the solution on any fixed bounded region \( \overline{G} \subset \Omega \) in the limit \( k \rightarrow \infty \). For this reason any well-posed problem on \( \Omega_k \) is suitable as an approximation of the original problem on the unbounded state space \( \Omega \), provided that \( k \) is taken sufficiently large and the validity of the approximation is considered on the compact region \( \overline{G} \) only. It is this result that justifies the use of essentially arbitrary boundary conditions associated with (2.29). From a theoretical standpoint any choice of artificial boundary conditions will work as long as the problem is well-posed on the bounded domain whereas from a practical standpoint of numerical computation the artificial boundary conditions should
approximate the solution on the boundary as closely as possible. If the latter is the case the bounded computational domain can be chosen to be of relatively moderate size reducing the number of numerical computations needed and thus gaining computational efficiency. Here the approximation domain is fixed to $G = (\underline{x}_1, \overline{x}_1) \times \cdots \times (\underline{x}_n, \overline{x}_n)$ in considering a sequence of computational domains $\Omega_k = (\underline{x}_1^k, \overline{x}_1^k) \times \cdots \times (\underline{x}_n^k, \overline{x}_n^k)$ where, due to the requirement $G \subset \Omega_k$, $-\infty < \underline{x}_i^k < \overline{x}_i < \overline{x}_i^k < \infty$ for $i = 1, \ldots, n$ and $k = 1, 2, \ldots$. A boundary condition that is a good approximation for the American option problem is the Dirichlet boundary condition $u_k|_{\partial \Omega_k} = 0$. Furthermore the transformation to the excess premium over the payoff, i.e. $u = V - \psi$, suits the American option problem well since the initial and the boundary conditions for $u$ are vanishing. The concepts of the approximation domain and the exhausting sequence of computational domains are illustrated in Figure 5.

Figure 5: Localization procedure: $\overline{G}$ is the approximation domain, where the interest in the solution lies, and $\Omega_1, \Omega_2, \ldots$ is an increasing sequence of computational domains covering the state space, i.e. a sequence constituting the exhausting sequence.

The weak variational formulation on bounded domains is given in Theorem 2.11 bellow, where the following theorem will be useful in the derivation of the formulation.

**Theorem 2.10.** (Green’s Formula).

$$\int_{\Omega} v \Delta u \, d\Omega = - \int_{\Omega} \nabla v \cdot \nabla u \, d\Omega + \int_{\Gamma} v \partial_n u \, d\Gamma,$$

where $\partial_n$ denotes the normal derivative with respect to the outward unit normal vector $\hat{n}$ to $\Gamma$, $\Delta$ the Laplace operator, $\nabla$ the gradient and $\Gamma$ the boundary of the integration domain $\Omega$. It should be noted that the boundary integral in the above equation can be written as

$$\int_{\Gamma} v \partial_n u \, d\Gamma = \int_{\Gamma} v \hat{n} \cdot \nabla u \, d\Gamma.$$
Multiplying equation (2.29a) with a test function \( \nu \) such that \( \nu \geq 0 \) it then follows that for any given \( t \)

\[
\left\{ \frac{\partial u}{\partial t} - Au - a_0 u - f \right\} \cdot \nu \geq 0.
\]

The weak form of (2.29) on bounded domains then follows by integrating the above expression over \( \Omega_k \) and using Green’s formula with the Dirichlet boundary condition in mind, i.e. that \( u_k |_{\partial \Omega} = 0 \). This type of boundary conditions are known as essential boundary conditions and are imposed explicitly on all candidate solutions. In following [34] it is for simplicity assumed that \( a_{ij} \in \mathbb{R} \) and in using the notations \( dx := d\Omega_k \) and \( \Gamma := \partial \Omega_k \) the following equation is obtained

\[
\int_{\Omega_k} 0 \cdot \nu \, dx \leq \int_{\Omega_k} \frac{\partial u_k}{\partial t} \nu \, dx + \frac{1}{2} \sum_{ij} \int_{\Omega_k} \frac{\partial^2 u_k}{\partial x_i \partial x_j} \nu \, dx \\
- \sum_j \int_{\Omega_k} a_{ij} \frac{\partial u_k}{\partial x_j} \nu \, dx - \int_{\Omega_k} a_0 u_k \nu \, dx - \int_{\Omega_k} f \nu \, dx \\
= \int_{\Omega_k} \frac{\partial u_k}{\partial t} \nu \, dx - \frac{1}{2} \sum_{ij} \int_{\Omega_k} a_{ij} \frac{\partial u_k}{\partial x_j} \frac{\partial \nu}{\partial x_i} \, dx + \frac{1}{2} \sum_{ij} \int_{\Gamma} a_{ij} \frac{\partial u_k}{\partial x_j} \frac{\partial \nu}{\partial x_i} \, d\Gamma \\
- \sum_j \int_{\Omega_k} a_{ij} \frac{\partial u_k}{\partial x_j} \nu \, dx - \int_{\Omega_k} a_0 u_k \nu \, dx - \int_{\Omega_k} f \nu \, dx \\
= \int_{\Omega_k} \frac{\partial u_k}{\partial t} \nu \, dx - \frac{1}{2} \sum_{ij} \int_{\Omega_k} a_{ij} \frac{\partial u_k}{\partial x_j} \frac{\partial \nu}{\partial x_i} \, dx \\
- \sum_j \int_{\Omega_k} a_{ij} \frac{\partial u_k}{\partial x_j} \nu \, dx - \int_{\Omega_k} a_0 u_k \nu \, dx - \int_{\Omega_k} f \nu \, dx.
\]

Before giving the main result of this section there is need for the following definition.

**Definition 2.12. (Bilinear forms).** Given two Hilbert spaces \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \), a mapping \( a : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{R} \) is known as a bilinear form if is linear in both arguments, i.e. in assuming that the vectors \( x, y, z \in \mathcal{H}_1 \), the vectors \( u, v, w \in \mathcal{H}_2 \) and that \( \alpha, \beta \) are scalars the following relations hold \( \forall x, y, z \in \mathcal{H}_1, \forall u, v, w \in \mathcal{H}_2, \forall \alpha, \beta \in (\mathbb{R} \text{ or } \mathbb{C}) \):

- \( a(x + y, w) = a(x, w) + a(y, w) \),
- \( a(\alpha x, w) = \alpha a(x, w) \),
- \( a(z, u + v) = a(z, u) + a(z, v) \),
- \( a(z, \beta u) = \beta a(z, u) \),
- \( a \) is a bounded \( \Rightarrow \) continuous bilinear form if the norm of \( a \) is bounded, i.e. if

\[
||a|| = \sup_{x \in \mathcal{H}_1, v \in \mathcal{H}_2} \frac{|a(x, w)|}{||x|| ||w||} \text{ is bounded,}
\]
• \( a \) is positive definite (alternatively, coercive or bounded and continuous) if
\[
\inf_{x \in \mathcal{H} \setminus \{0\}} \frac{a(x, x)}{\|x\|^2} > 0.
\]

In defining the bilinear form for each \( t \in (0, T) \), \( u, \nu \in \mathcal{H} \), where it is assumed that \( a_{ij} \in \mathbb{R} \),
\[
a_k(t; u, \nu) := \frac{1}{2} \sum_{ij} a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial \nu}{\partial x_i} dx + \sum_j \int_{\Omega_k} a_j \frac{\partial u}{\partial x_j} \nu dx + \int_{\Omega_k} a_0 u \nu dx,
\]
and denoting the inner product in \( L^2(\Omega_k) \) by \((\cdot, \cdot)\), i.e. \((u, \nu) = \int_{\Omega_k} u(x) \nu(x) \, dx\), it is now possible to give the main result of this section in the following theorem.

**Theorem 2.11.** (Weak variational formulation on bounded domains). In considering that \( \text{supp}(\psi) \subseteq \Omega_k \), where \( \text{supp} \) is shorthand for support, in which case \( u_k|_\Gamma = 0 \) for all \( k \) sufficiently large. The weak variational problem is then find \( u_k \in \mathcal{K}_t := \{ \nu \in \mathcal{H}^{2,1}(\Omega_k \times (0, T)) \mid \nu \geq 0 \text{ a.e. in } \Omega_k \times (0, T); \nu = 0 \text{ a.e. on } \Gamma \times (0, T) \} \), such that, for all \( \nu \in \mathcal{K}_t \),
\[
\left( \frac{\partial u_k}{\partial t}, \nu \right) - a_k(t; u_k, \nu) \geq (f, \nu) \quad \text{a.e. in } t,
\]
where
\[
u_k(x, T) = \psi(x, T).
\]

With the assumptions made in Remark 2.11, i.e. that \( a_{ij} := \sigma_i(x, t) \sigma_j(x, t)^* \), \( a_0 = r \) and \( \mu_j := r \), the theorem above describes the problem setting in the Black-Scholes model.

### 2.5.4 Penalty approximation

In order to solve the variational inequality formulation given in equation (2.29) numerically there is the need for transforming the constrained problem into an unconstrained one. In following [26] this can be done by constructing a penalty approximation by adding a penalty term in problem (2.29). This penalty term penalizes solutions if they do not satisfy the given constraint. The problem (2.29) can be reconstructed to to yield the penalty approximation
\[
\frac{\partial u_e}{\partial t} - Au_e - a_0 u_e - \frac{1}{\varepsilon} (u_e)^{-} - f = 0
\]
with the homogeneous initial condition
\[
u_e(x, 0) = 0,
\]
where \( \varepsilon > 0 \) is some small fixed parameter and the notation \( (z)^- := \max(-z, 0) \). The action of the early exercise is approximated by the non-linear penalty term \( \frac{1}{\varepsilon} (u_e)^- \). The solution \( u_e \) of the
penalized non-linear PDE above converges to the solution of the variational inequality (2.29) as \( \varepsilon \to 0 \). In particular the following relations hold for the penalty approximation.

\[
\max \|u_{\varepsilon}(\cdot, t) - u(\cdot, t)\|_{L^\infty(\Omega)} \leq C \varepsilon
\]

for the penalization error estimate where \( C > 0 \) does not depend on \( \varepsilon \), and

\[
u_{\varepsilon} \geq -C \varepsilon.
\]

These two relations have the following intuitive formulation. The solution \( u \) of the variational inequality is constrained to stay non-negative and corresponds to the value function \( V \) taking on values \( V \geq \psi \), i.e. values that are not less than the payoff that can be obtained through an early exercise of the option. The solution \( u_{\varepsilon} \) of the penalized PDE on the other hand can take on values that are below zero. While \( u_{\varepsilon} \geq 0 \) the penalty term vanishes, whereas when \( u_{\varepsilon} < 0 \) the value of the penalty term \( \frac{1}{\varepsilon}(u_{\varepsilon})^- \) is positive and is increasing rapidly as the solution falls below zero and thus forcing the solution back above zero. The coefficient \( \frac{1}{\varepsilon} \) in the penalty term becomes larger as the value of \( \varepsilon \) becomes smaller giving the penalty term a better approximation to the action of the early exercise constraint. If the penalized solution \( u_{\varepsilon} \) falls below zero it is bounded by \( u_{\varepsilon} \geq -C \varepsilon \), where \( C > 0 \) is independent of \( \varepsilon \).

The fact that the penalty term \( \frac{1}{\varepsilon}(u_{\varepsilon})^- \) has discontinuous first derivative w.r.t. \( u_{\varepsilon} \) gives rise to computational challenges in the numerical solution. After suitable spatial discretization and implicit time discretization there is need for using Newton-type iterations to solve a non-linear system of algebraic equations resulting from the discretized PDE. The discontinuity in the Jacobian of this system, that is inherited from the discontinuity in the derivative of the penalty term w.r.t. \( u_{\varepsilon} \), forces the use of non-smooth Newton-type iterative schemes for non-linear systems with discontinuous Jacobians. To circumvent this problem more general penalty terms of the form \( \left(\frac{1}{\varepsilon}(u_{\varepsilon})^-\right)^p \) for some \( p \geq 1 \) can be used, which will be the case here. If the value \( p > 1 \) is taken the continuity of the derivative of the penalty term w.r.t. \( u_{\varepsilon} \) is restored and standard Newton iteration with continuous Jacobian can be used. The penalized non-linear PDE is thus reformulated as

\[
\frac{\partial u_{\varepsilon}}{\partial t} - Au_{\varepsilon} - a_0 u_{\varepsilon} - \left(\frac{1}{\varepsilon}(u_{\varepsilon})^-\right)^p - f = 0
\]

with the homogeneous initial condition

\[
u_{\varepsilon}(x, 0) = 0.
\]

3 Numerical methods

In this section, mainly following [33], [24], [49] and [48] numerical procedures for valuing financial derivatives are introduced when there are no exact formulas available. Also following [42] and [18] some iterative methods for solving linear system of equations are also presented in this section. The first of these procedures is only briefly discussed since its mention here is for orientational purposes only. Two popular and closely related numerical methods for option
Monte Carlo simulation

Monte Carlo simulation is used for financial derivatives where the payoff depends on the past history of the underlying asset or where the model is multidimensional, i.e. there are several underlying assets. One of the main advantages with the method is its easy implementation but the model is mostly suited for European style options, where there is no possibility for early exercise. The second methods discussed are the finite difference methods (FDM) where the methods, as suggested by their name, use finite-difference approximations to the partial derivatives in the partial differential equation associated with the financial derivative needed to be valued. One advantage with the methods is that they can handle American put options that impose boundary conditions other than terminal time conditions. Besides approximation errors, the methods use approximations to the actual partial differential equation and the solutions inherent instabilities and inconsistencies that are associated with the discretization of the continuous problem. A feature that is not characteristic of the partial differential equation itself.

The final method discussed is the finite element method (FEM) which uses an approximate solution known as a Galerkin approximation to solve variational formulations. It is this method that will be used here for solving the penalized problem discussed in Section 2.5.4 and some of the related features associated with this method are discussed in Section 3.4.

3.1 Monte Carlo simulation

Due its ease of implementation and its ease of handling high-dimensionality Monte Carlo simulation is a popular method used for pricing options. On the other hand the early exercise condition is not handled easily or impossible to implement. There is a brief intuitive reasoning in [10] behind the last statement. To able to exploit the early exercise opportunities optimally there is the need of going backwards in time to be able to compare the value of exercising or keeping the option, a decision that is based on the value of the option. But since the simulations are run forwards in time this presents a type of complexity that is not inherent in the PDE methods (FDM and FEM). In [32] a comparison is made between a space-time adaptive finite difference method a Monte Carlo method and a Quasi-Monte Carlo method when pricing a one, two and three dimensional European option. According to [10] it is only in recent years the belief that simulation should not be applied to American style options has diminished considerably. This mainly due to the introduction of new algorithms, using simulations that can handle early exercise conditions as proposed in [31], during the last decade.

The basic idea behind Monte Carlo simulations is to estimate the value of $F(t, s)$ in Theorem 2.6, i.e. to compute the value of the estimator $\hat{F}$ by

$$\hat{F} = e^{-r(T-t)}\text{average}[\Phi(S(T))].$$

This can be done by simulating the stochastic dynamics of the price process $S$ given in equation (2.17). Dividing the time interval $[0, T]$ into periods of length $\Delta t$ the simulation equation is defined as follows

$$S(t_k + \Delta t) = S(t_k) + rS(t_k)\Delta t + \sigma S(t_k)\epsilon(t_k),$$

where $\epsilon(t_k) \in N[0, \sqrt{\Delta t}]$ is chosen by a random number generator. A value of $\Phi(S(T))$ is calculated for each simulation path and the estimate $\hat{F}$ is then computed. One major disadvantage
of the method is that the error decays as $M^{-1/2}$, where $M$ is the number of simulation trials, i.e. to get one more digit of accuracy the number of trials have to be increased 100 times. As stated in [33] to get two decimal place accuracy tens of thousands simulation trials are often required. On the other hand the error decay is independent of the dimension of the problem, a major advantage compared to PDE methods when dealing with high dimensional problems.

### 3.2 The Finite Difference Methods

In the finite difference method the partial derivatives in the backward Black-Scholes equation, see Remark 3.1, are approximated by finite-difference approximations. Besides introducing the three different approximations forward difference, backward difference and central difference accuracy and stability issues are also addressed in [49]. Depending on the choice of approximation type three different finite difference schemes arise, where here the values of $u(\tau, S)$ are restricted on the regular finite difference mesh with $u_i^m = u(m\Delta\tau, i\Delta S)$, namely

#### The Explicit Finite-difference scheme:

$$\frac{u_i^m - u_i^{m-1}}{\Delta\tau} = \frac{1}{2}\sigma^2 S_i^2 \frac{u_{i+1}^{m-1} - 2u_i^{m-1} + u_{i-1}^{m-1}}{(\Delta S)^2} + rS_i \frac{u_{i+1}^{m-1} - u_{i-1}^{m-1}}{2\Delta S} - ru_i^{m-1}$$

#### The Implicit Finite-difference scheme:

$$\frac{u_i^m - u_i^{m-1}}{\Delta\tau} = \frac{1}{2}\sigma^2 S_i^2 \frac{u_{i+1}^m - 2u_i^m + u_{i-1}^m}{(\Delta S)^2} + rS_i \frac{u_{i+1}^m - u_{i-1}^m}{2\Delta S} - ru_i^m$$

#### The Crank-Nicolson Finite-difference scheme:

$$\frac{u_i^m - u_i^{m-1}}{\Delta\tau} = \frac{1}{4}\sigma^2 S_i^2 \frac{u_{i+1}^{m-1} - 2u_i^{m-1} + u_{i-1}^{m-1}}{(\Delta S)^2} + \frac{1}{2}rS_i \frac{u_{i+1}^{m-1} - u_{i-1}^{m-1}}{2\Delta S} - \frac{1}{2}rS_i(u_i^{m-1} + u_i^m)$$

The explicit and implicit schemes are first-order accurate in $\Delta\tau$ and second-order accurate in $\Delta S$, i.e. they are accurate to $O(\Delta\tau)$ and $O((\Delta S)^2)$, while the accuracy of the Crank-Nicolson scheme is $O((\Delta\tau)^2)$ and $O((\Delta S)^2)$, for details see [50]. Essentially the Crank-Nicolson scheme is the average of the implicit and explicit schemes and is used not only to get better rate of convergence but the Crank-Nicolson as the implicit scheme do not require the stability constraint that is associated in using a explicit scheme, see [43] or [51] where the latter reference contains a probabilistic interpretation of the stability condition.

**Remark 3.1.** When applying a numerical method to the Black-Scholes equation there is the need to make a change of variables regarding the time variable, transforming the terminal value.
problem into a initial value problem. Performing the substitution of \(t\) to \(\tau = T - t\), i.e. \(u(\tau, s) \coloneqq F(t, s)\), and using the chain rule \(\left(\frac{\partial F}{\partial t}\right) = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial s} = -\frac{\partial u}{\partial s}\) the terminal value problem in Theorem 2.3 becomes the initial value problem

\[
\begin{align*}
    u_\tau - r s u_s - \frac{1}{2} \sigma^2 s^2 u_{ss} + ru &= 0, \\
    u(0, s) &= \Phi(s).
\end{align*}
\]

3.3 Iterative methods

The methods discussed in this section are concerned with the solving of a system of linear equations, i.e. in solving a system of linear equations \(Ax = b\) for the unknown vector \(x\), where \(A \in \mathbb{R}^{n \times n}\), \(x \in \mathbb{R}^n\) and \(b \in \mathbb{R}^n\). After choosing a suitable regular matrix \(M \in \mathbb{R}^{n \times n}\) the system can be written as

\[
    Mx = Mx - Ax + b = (M - A)x + b
\]

\[
    \Rightarrow x = (I - M^{-1}A)x + M^{-1}b = Gx + c =: f(x).
\]

The solution to the system of linear equations is equivalent to the fixed point problem \(f(x) = x\). The Banach fixed point theorem, see Theorem C.5 in Section C, leads to the corresponding iteration equation

\[
    x^{(k+1)} = f(x) = Gx^{(k)} + c,
\]

where \(k\) denotes the number of the iteration. Subtracting the fixed point equation from the last one gives the error term \(e^{(k+1)} = x^{(k+1)} - x = G(x^{(k+1)} - x) = Ge^{(k)}\). It follows by Lemma 4.2 in [42] that if an explicit scheme is used for discretization then convergence is achieved if the spectral radius of \(G\) is less than one, i.e if \(\lim_{k \to \infty} e^{(k)} = 0 \iff \rho(G) < 1\) where \(\rho(G) \coloneqq \max \lambda_i^G\) and \(\lambda_1^G, \ldots, \lambda_n^G\) are the eigenvalues of \(G \in \mathbb{R}^{n \times n}\). The values of \(x^{(k+1)}\) are found by solving the system \(Mx^{(k+1)} = (M - A)x^{(k)} + b\) and the construction of the matrix \(M\), also referred to as the preconditioner matrix, should be simple and such that rapid convergence is achieved. Classical relaxation methods use additive splitting of the matrix \(A\) as in \(A = D - L - U\), where \(D\) is a diagonal, \(L\) a strict lower-triangular and \(U\) a strictly upper-triangular matrix. Below are given some widely used iterative methods. It should also be mentioned that a suitable convergence criteria for a iterative method is \(\|x^{(k+1)} - x^{(k)}\|_2 \leq \epsilon\), where here \(\epsilon \ll 1\) is a user defined tolerance parameter.

Jacobi’s Method:

The matrix \(M \coloneqq D\) gives the iteration

\[
    Dx^{(k+1)} = (L + U)x^{(k)} + b \implies x^{(k+1)} = D^{-1}(L + U)x^{(k)} + D^{-1}b.
\]

The component form of the solution in the Jacobi method is

\[
    x_i^{(k+1)} = -\frac{1}{a_{ii}} \sum_{j=1}^{n} a_{ij} x_j^{(k)} + \frac{b_i}{a_{ii}} \quad \text{for } i = 1, 2, 3, \ldots, n \text{ in any order}.
\]
Gauss-Seidel Method:
If instead the matrix is chosen as $M := D - L$ then
\[
(D - L)x^{(k+1)} = UX^{(k)} + b
\]
\[
\Leftrightarrow x^{(k+1)} = (D - L)^{-1}UX^{(k)} + (D - L)^{-1}b.
\]
The Gauss-Seidel method is essentially derived from the component form of the Jacobi’s method. The component form the Jacobi’s method can be rewritten as
\[
x_i^{(k+1)} = -\frac{1}{a_{ii}} \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \frac{1}{a_{ii}} \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + \frac{b_j}{a_{ii}} \quad \text{for } i = 1, 2, \ldots, n \text{ in any order.}
\]
Replacing the superscript of $x_j$ from $k$ to $k + 1$ in the first term of the right-hand side in the equation above allows for the usage of the most up-to-date values available, since the values of $x_j^{(k)}$ for $j = 1, 2, \ldots, i - 1$ are already computed and known if proceeding in the natural order. Thus the component form of the solution in the Gauss-Seidel method is
\[
x_i^{(k+1)} = -\frac{1}{a_{ii}} \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \frac{1}{a_{ii}} \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + \frac{b_j}{a_{ii}} \quad \text{for } i = 1, 2, \ldots, n \text{ in the natural order.}
\]
Successive Over Relaxation:
The SOR method is based on the Gauss-Seidel method where a relaxation parameter $0 < \omega < 2$ is introduced and chosen, see [44], so that better rate of convergence is achieved. In viewing the Gauss-Seidel method as adding the increment $x^{(k+1)} - x^{(k)}$ to $x^{(k)}$, i.e. $x^{(k+1)} = x^{(k)} + (x^{(k+1)} - x^{(k)})$, it is possible to speed up the convergence by multiplying the increment by $\omega$ in giving
\[
x^{(k+1)} = x^{(k)} + \omega(x^{(k+1)} - x^{(k)})
\]
\[
= x^{(k)} + \omega(D^{-1}Lx^{(k+1)} + D^{-1}UX^{(k)} + D^{-1}b - x^{(k)})
\]
\[
= (1 - \omega)x^{(k)} + \omega(D^{-1}Lx^{(k+1)} + D^{-1}UX^{(k)} + D^{-1}b).
\]
The component form, enabling the computation of the solution without computing matrix inverses, for $i = 1, 2, \ldots, n$ in the natural order is then
\[
x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \omega \left( -\frac{1}{a_{ii}} \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \frac{1}{a_{ii}} \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + \frac{b_j}{a_{ii}} \right).
\]

### 3.3.1 Projected Successive Over Relaxation

It is possible to transform the Black-Scholes equation to the dimensionless form of the heat equation by change of variables, see Section D, by the following transformation
\[
S = Ke^x, \quad t = T - \frac{\tau}{\sigma^2} \quad \text{where } \tau := \frac{2\sigma^2 \tau}{\sigma^2}, \quad q := \frac{2(r - \delta)}{\sigma^2},
\]
\[
V(t, S) = V(T - \frac{2\tau}{\sigma^2}, Ke^x) =: v(\tau, x),
\]
\[
v(\tau, x) := Ke^{-\frac{1}{2}(q_1-1)x - \frac{1}{2}(q_1-1)^2\tau}y(\tau, x).
\]
Thus the linear complementary problem (LCP) for a American put option where the underlying asset is paying the dividend $\delta \geq 0$ can be formulated as

$$\begin{align*}
\begin{cases}
\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} & = 0 \\
\frac{\partial y}{\partial \tau} & \geq 0, \\ y(0, \tau) & = g(0, \tau), \\ y(\infty, \tau) & = g(\infty, \tau), \\ y & \geq 0
\end{cases}
\end{align*}$$

(3.1)

where $q_i = \frac{2(q_i - \sigma^2)}{\sigma^2}$, $q = \frac{2\delta}{\sigma^2}$ and $(\cdot)^+ := \max(\cdot, 0)$. In practice the boundary conditions are formulated as $y(\infty, x_{\text{min}}) = g(\infty, x_{\text{min}})$ and $y(\infty, x_{\text{max}}) = g(\infty, x_{\text{max}})$. Solving for $y(\tau, x)$ the transformation is applied again to get the value of the option $V(t, S)$ in the original variables.

Now consider discretizing with finite differences where $u_i^m \approx y(x_m, t_i)$ and $g_i^m := g(x_m, t_i)$. It is possible to combine the explicit, implicit and Crank-Nicolson finite difference schemes into one equation. The weight parameter $\theta$ gives explicit scheme if $\theta = 0$, implicit scheme if $\theta = 1$ and the Crank-Nicolson scheme if $\theta = \frac{1}{2}$ in the equation below

$$\begin{align*}
\frac{u_i^m - u_i^{m-1}}{\Delta t} = \theta \frac{u_i^{m+1} - 2u_i^m + u_i^{m-1}}{(\Delta x)^2} + (1 - \theta) \frac{u_i^{m+1} - 2u_i^m + u_i^{m-1}}{(\Delta x)^2}.
\end{align*}$$

Using $\lambda := \frac{\Delta \tau}{(\Delta x)^2}$ the differential inequality $\left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2}\right) \geq 0$ is discretized as

$$\begin{align*}
u_i^m - \lambda \theta (u_i^{m+1} - 2u_i^m + u_i^{m-1}) - \lambda (1 - \theta) (u_i^{m+1} - 2u_i^m + u_i^{m-1}) & \geq 0 \\
\Leftrightarrow u_i^m - \lambda \theta (u_i^{m+1} - 2u_i^m + u_i^{m-1}) & \geq u_i^{m-1} + \lambda (1 - \theta) (u_i^{m+1} - 2u_i^m + u_i^{m-1}) \\
& = b_i^{m-1} \quad \text{for } i = 2, \ldots, n-2
\end{align*}$$

With the choice of $b_i^{m-1}$ as above and the vectors

$$\begin{align*}
b^{(m-1)} := \begin{bmatrix} b_1^{m-1} \\ \vdots \\ b_n^{m-1} \end{bmatrix}, \quad u^{(m-1)} := \begin{bmatrix} u_1^{m-1} \\ \vdots \\ u_n^{m-1} \end{bmatrix}, \quad g^{(m-1)} := \begin{bmatrix} g_1^{m-1} \\ \vdots \\ g_n^{m-1} \end{bmatrix}
\end{align*}$$

and

$$A := \begin{bmatrix} 1 + 2\lambda \theta & -\lambda \theta & \cdots & \cdots \\ -\lambda \theta & \ddots & \ddots & \cdots \\ \vdots & \ddots & \ddots & \ddots \\ \cdots & \cdots & \cdots & 0 \end{bmatrix},$$

where $A \in \mathbb{R}^{(n-1) \times (n-1)}$, it is possible to rewrite the discretized version as $Au^{(m)} \geq b^{(m-1)}$, $\forall m \geq 1$. In defining $b_i^{m-1}$ and $b_n^{m-1}$ the boundary conditions $u_0^{m-1} = g_0^{m-1}$, $u_n^{m-1} = g_n^{m-1}$, $m \geq 2$ have to be realized in $b_i^{m-1}$, i.e.

$$\begin{align*}
b_1^{m-1} & = u_1^{m-1} + \lambda (1 - \theta) (u_2^{m-1} - 2u_1^{m-1} + g_0^{m-1}) + \lambda \theta g_0^{m-1} \\
b_n^{m-1} & = u_n^{m-1} + \lambda (1 - \theta) (g_n^{m-1} - 2u_{n-1}^{m-1} + u_n^{m-1}) + \lambda \theta g_n^{m-1}.
\end{align*}$$

---

**Pricing a Multi-Asset American Option**
Also the initial conditions are $u^{(0)} = g^{(0)}$ and $u_i^0 = g_i^0$, $i = 1, 2, \ldots, n-1$. A solving procedure for the system 3.1 can now be summarized in pseudocode in following Algorithm 4.8 in [42],

Finite difference discretised LCP

for $m = 0$ to $m = m_{\text{max}} - 1$ do
  Compute $g := g^{(m+1)}$, $b := b^{(m)}$.
  Compute the solution $u$ to
  \[ Au - b \geq 0, \quad u \geq g, \quad (Au - b)^\top (u - g) = 0. \]
  Set $u^{m+1} = u$.
end for

As suggested by the pseudocode above at each time level $m$ a discretized LCP problem must be solved. Since $V(t,S)$ in not $C^2$-smooth across the free boundary $b(t)$, i.e. $\frac{\partial V}{\partial S}$ and $\frac{\partial^2 V}{\partial S^2}$ are not continuous across $b(t)$, see Section 2.5.1 on discussion of the free boundary, the order of accuracy in the method used is usually less than suggested by the method. One feature of iterative methods is the property of smoothing. The SOR method for the problem discussed can be formulated as, where $x := u - g$ and $y := Au - g$,

SOR formulation

Find $x$ and $y$ such that for $\hat{b} := b - Ag$
\[ Ax - y = \hat{b}, \quad x \geq 0, \quad y \geq 0, \quad x^\top y = 0. \]

The componentwise formulation of the iteration within the SOR method for solving $Ax = \hat{b}$ is, where again $k$ is the iteration number and $\omega$ the relaxation parameter,

\[ \vartheta_i^{(k+1)} = \hat{b}_i - a_{ii} x_i^{(k)} - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n-1} a_{ij} x_j^{(k)} \]
\[ x_i^{(k+1)} = x_i^{(k)} + \omega \frac{\vartheta_i^{(k+1)}}{a_{ii}} \]

The projected SOR (PSOR) is a modification of the SOR algorithm. The constraint $x_i^{(k+1)} \geq 0$ is enforced by modifying the second component above to

\[ x_i^{(k+1)} = \max \left\{ 0, x_i^{(k)} + \omega \frac{\vartheta_i^{(k+1)}}{a_{ii}} \right\}. \]

Transforming back to $u$ the PSOR method for solving the LCP for the American option can be formulated as

Pricing a Multi-Asset American Option
3.4 The Finite Element Method

The FEM is an alternative to the finite difference method when solving PDEs. Compared to finite difference methods, which are mostly suited for equidistant grids, the FEM framework enables an easy way to handle complex geometries, boundaries and operators. Having a grid with finer regions where the solution has steep gradients and coarser regions where the gradient is nearly flat is a flexibility feature that is natural within the unstructured mesh generated in finite element formulations. This type of flexibility is hard to obtain with finite difference methods. According to [42] this increase in degrees of freedom becomes more important as the dimension of the problem increases and for certain type of options that are path dependent, e.g. Exotic options, finite elements is the natural method of choice when pricing this type of derivatives. The best way to explain the principles of the finite element method is through an example but firstly there is the need for some theoretical discussion in following [48] and [27], where proofs of given theorems and lemmas can be found.

Definition 3.1. (Linear operators). Given two Hilbert spaces \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) a mapping \( A : \mathcal{H}_1 \rightarrow \mathcal{H}_2 \) is linear operator if \( \forall x, y \in \mathcal{H}_1 \)

- \( A(x, y) = A(x) + A(y) \), where \( A(x) \) is often written as \( Ax \).
- \( A(ax) = aAx \).
- \( A \) is a bounded \( \Rightarrow \) continuous linear operator if the norm of \( A \) is bounded, i.e. if

\[
\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \text{ is bounded.}
\]

Definition 3.2. (Linear functional). A linear functional \( F \) is a linear operator with domain, \( \mathcal{D}(F) \), in a vector space \( \mathcal{V} \) and range in the scalar field \( \mathcal{K} \) of \( \mathcal{V} \), i.e.

\[
F : \mathcal{D}(F) \rightarrow \mathcal{K},
\]

where \( \mathcal{K} = \mathbb{R} \) if \( \mathcal{V} \) is real and \( \mathcal{K} = \mathbb{C} \) if \( \mathcal{V} \) is complex. (Here it is assumed that \( \mathcal{V} \) is real.)

Theorem 3.1. (Reisz’s representation theorem). Given a Hilbert space \( \mathcal{V} \), \( F \) is a linear functional on \( \mathcal{V} \) iff there exists \( u \in \mathcal{V} \) such that \( \forall v \in \mathcal{V} \)

\[
F(v) = (u, v),
\]

where \( \langle \cdot, \cdot \rangle \) denotes the inner product.
Remark 3.2. It should be mentioned that in Calculus of Variations, where the “variables” (or the unknowns) are functions or derivatives of functions, \((\cdot,\cdot)\) is an inner product defined by integrals, i.e. \((v,w) = \int_\Omega v(x)w(x)\,dx\). Here it is assumed that \(\Omega \subseteq \mathbb{R}^n\). For example in the one dimensional Poisson problem: 
\[-u'' = f\] with zero Dirichlet boundary conditions, i.e. \(u|_{\partial\Omega} = 0\), 
\[a(u,w) = \int_\Omega u'w'\,dx\]. The theoretical framework using FEM for solving this problem will be given in Section 3.4.3.

Definition 3.3. (Variational problem). Given a bounded bilinear form \(a\), see Definition 2.12, and a bounded linear functional \(L\) both defined on a Hilbert space \(\mathcal{V}\) the variational problem is formulated as find \(u \in \mathcal{V}\) such that 
\[a(u,\nu) = L(\nu) \quad \forall \nu \in \mathcal{V}.\]

Theorem 3.2. (The Lax-Milgram theorem). Given a continuous and positive definite bilinear form \(a(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}\) and a continuous linear functional \(L : \mathcal{V} \to \mathbb{R}\) on the Hilbert space \(\mathcal{V}\), there is a unique and bounded (\(\|u\| \leq \kappa\|L\|\), where \(\kappa\) is some constant) solution \(u \in \mathcal{V}\) to 
\[a(u,\nu) = L(\nu) \quad \forall \nu \in \mathcal{V}.\]

The idea behind the Galerkin approximation is to find and approximate the solution to the variational problem and replace the infinite dimensional problem by a finite dimensional one. Let \(\mathcal{V}\) be a given Hilbert space with an \(n\)-dimensional subspace \(\mathcal{V}_h \subset \mathcal{V}\) then the variational problem 
\[\text{Find } u \in \mathcal{V} \text{ satisfying } \quad a(u,\nu) = L(\nu) \quad \forall \nu \in \mathcal{V}\]
is replaced by 
\[\text{Find } U_h \in \mathcal{V}_h \subset \mathcal{V} \text{ satisfying } \quad a(U_h,\nu) = L(\nu) \quad \forall \nu \in \mathcal{V}_h.\]

Lemma 3.1. (Cea’s lemma). Given a positive definite and symmetric bilinear form \(a(\cdot,\cdot)\), the unknown solution \(u \in \mathcal{V}\) and the Galerkin approximation \(U_h \in \mathcal{V}_h \subset \mathcal{V}\) 
\[\|u - U_h\|_E \leq \|u - \nu_h\|_E \quad \forall \nu_h \in \mathcal{V}_h,\]
where \(\|u\|_E = a(u,u)^{1/2}\), i.e. the Galerkin approximation is best in energy.

3.4.1 Matrix representation

In assuming that the functions \(\phi_i \in \mathcal{V}\) for \(i = 1,\ldots,n\) are linearly independent and thus form a basis for a \(n\)-dimensional subspace \(\mathcal{V}_h\), it is possible to represent the Galerkin equations \(a(U,\nu) = L(\nu) \forall \nu \in \mathcal{V}_h\) in matrix form. Since the \(\phi_i\) form a basis for \(\mathcal{V}_h\) it is possible to express any \(U \in \mathcal{V}_h\) as a linear combination of the basis functions \(\phi_i\), i.e.
\[U = \sum_{j=1}^{n} \alpha_j \phi_j,\]
where \(\alpha_j \ j = 1,\ldots,n\) are scalars. The Galerkin equations can be written as
\[a(U,\phi_i) = L(\phi_i) \quad i = 1,\ldots,n\]
and plugging in the expression for \( U \) gives

\[
a(U, \phi_i) = a(\sum_{j=1}^{n} \alpha_j \phi_j, \phi_i) = \sum_{j=1}^{n} \alpha_j a(\phi_j, \phi_i) = L(\phi_i) \quad i = 1, \ldots, n.
\]

In matrix form the expression above becomes

\[
A x = b, \quad a_{ji} = a(\phi_j, \phi_i), \quad x := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad b := \begin{bmatrix} L(\phi_1) \\ \vdots \\ L(\phi_n) \end{bmatrix},
\]

where the matrix \( A \), the Gram matrix, is \( n \times n \) and since the bilinear form is symmetric, i.e. \( a(u, \nu) = a(\nu, u) \), \( A \) is symmetric (\( A^T = A \)). Also since the bilinear form is positive definite the matrix \( A \) is positive definite, i.e. for every \( x \in \mathbb{R}^d - \{0\} \) the expression \( x^T A x > 0 \) holds.

### 3.4.2 Trial functions

When solving PDEs in the Galerkin solution \( U = \sum_{j=0}^{n} U_j \phi_j(x) \) the basis or trial functions \( \{\phi_i\}_{j=0}^{n} \) are known and the free parameters \( \{U_j\}_{j=0}^{n} \) are to be determined. The choice of suitable basis functions should be based on ease of implementation and accuracy. Also if the matrix \( A \) is sparse the linear system of equations can be solved efficiently even when the system is large. The simplest choice for the \( \phi_i \)'s in one dimension are the Hat functions, see Figure 6, being piecewise linear trial functions. The hat functions are defined as

\[
\phi_i(x) = \begin{cases} 
\frac{x - x_{i-1}}{x_i - x_{i-1}} & \text{for } x_{i-1} \leq x < x_i \\
\frac{x_{i+1} - x}{x_{i+1} - x_i} & \text{for } x_i \leq x < x_{i+1} \\
0 & \text{otherwise}
\end{cases}
\]

\[
\phi_i(x) = \begin{cases} 
1 & i = j \\
0 & \text{otherwise}
\end{cases}
\]

![Figure 6: The Hat function \( \phi_i(x) \).](image)

The boundary functions are half-hats, see Figure 7, and are defined as
\[ \phi_0(x) = \begin{cases} \frac{x_1 - x}{x_1 - x_0} & \text{for } x_0 \leq x \leq x_1 \\ 0 & \text{otherwise} \end{cases} \]

\[ \phi_n(x) = \begin{cases} \frac{x - x_{n-1}}{x_n - x_{n-1}} & \text{for } x_{n-1} \leq x \leq x_n \\ 0 & \text{otherwise} \end{cases} \]

(a) The half-hat function \( \phi_0(x) \).

(b) The half-hat function \( \phi_n(x) \).

Figure 7: The half-hat functions \( \phi_0(x) \) and \( \phi_n(x) \).

The support \( \Omega_i \) of the function \( \phi_i \) is defined as the union of the subintervals on which \( \phi_i \) is non-zero, i.e. \( \Omega_i := I_i \cup I_{i+1} = [x_{i-1}, x_{i+1}] \). Thus on every subinterval \( I_i = [x_i, x_{i+1}] \) only \( \phi_i \) and \( \phi_{i+1} \) are non-zero which implies \( \phi_j \phi_i = 0 \) and \( \phi_j' \phi_i' = 0 \) for \( |j - i| > 1 \), a property that results in a sparse matrix \( A \).

3.4.3 A simple 1D example

In following [42] consider on \( \Omega \subset \mathbb{R} \) the problem with zero Dirichlet boundary conditions

\[
\begin{align*}
-u'' &= f(x) \quad \text{in } \Omega \\
u &= 0 \quad \text{on } \partial \Omega.
\end{align*}
\]

Pricing a Multi-Asset American Option
Let \( \nu \) be a test function. Multiplying by \( \nu \) and integrating by parts gives

\[
\int_\Omega f \nu \, dx = \int_\Omega -u'' \nu \, dx
\]

\[
= \int_\Omega u' \nu' \, dx - u' \nu|_{\partial \Omega}
\]

Using the boundary conditions a suitable space for test functions

\[
\mathcal{V} = \left\{ \nu : \int_\Omega (\nu^2 + \nu'^2) \, dx < \infty \text{ and } \nu = 0 \text{ on } \partial \Omega \right\}
\]

and the variational form is

\[
\text{Find } u \in \mathcal{V} \text{ such that } a(u, \nu) = L(\nu) \quad \forall \nu \in \mathcal{V}
\]

where

\[
L(\nu) = \int_\Omega f \nu \, dx \quad \text{and} \quad a(u, \nu) = \int_\Omega u' \nu' \, dx
\]

Plugging in the Galerkin approximation \( U = \sum_{i=0}^n U_i \phi_i(x) \) and \( \nu = \phi_j(x) \) into the variational formulation above yields

\[
\sum_{i=0}^n U_i \int_{\Omega} \phi'_i \phi'_j \, dx = \sum_{i=0}^n f_i \int_{\Omega} \phi_i \phi_j \, dx \quad j = 0, \ldots, n.
\]

In the equation above the function \( f \) is approximated by its continuous piecewise linear interpolant \( \pi_h f \), i.e. \( \pi_h f := \sum_{i=0}^n f_i \phi_i \), where \( f_i := f(x_i) \). The values of the interpolant \( \pi_h f \) and \( f \) on the subinterval \( I = [x_i, x_{i+1}] \) are the same at the nodes \( x_i \) and \( x_{i+1} \), e.g. \( \pi_h f(x_i) = f(x_i) \) and \( \pi_h f(x_{i+1}) = f(x_{i+1}) \). In following [29] where proofs can be found the interpolation error, i.e. \( f - \pi_h f \), is

\[
\|f - \pi_h f\|_{L^2(I)} \leq C h^2 \|f_{xx}\|_{L^2(I)}
\]

\[
\|(f - \pi_h f)_x\|_{L^2(I)} \leq C \|f_{xx}\|_{L^2(I)}
\]

where \( h = x_{i+1} - x_i \), \( C \) is a constant and the norm \( L^2(I) \) is defined as \( \|w\|_{L^2(I)} := \left( \int_I w^2 \, dx \right)^{1/2} \). Thus the error estimate for continuous piecewise linear interpolation is

\[
\|f - \pi_h f\|_{L^2(I)}^2 \leq C \sum_{i=1}^n h_i^4 \|f_{xx}\|_{L^2(I)}^2
\]

\[
\|(f - \pi_h f)_x\|_{L^2(I)}^2 \leq C \sum_{i=1}^n h_i^2 \|f_{xx}\|_{L^2(I)}^2
\]

In matrix form the expression above is \( \mathbf{A} \mathbf{U} = \mathbf{M} \mathbf{f} \), where \( \mathbf{A} \) is the stiffness matrix and \( \mathbf{M} \) the mass matrix. In order to assemble (construct) the \( \mathbf{A} \) and \( \mathbf{M} \) matrices the integrals over the region \( \Omega \) are evaluated as the sums of integrals over elements, i.e.

\[
\int_\Omega = \sum_e \int_e
\]
On each element $e = I_i = [x_{i-1}, x_i]$ only the basis functions $\phi_i^{(e)} = \frac{x - x_{i-1}}{x_i - x_{i-1}}$ and $\phi_i^{(e)} = \frac{x_i - x}{x_i - x_{i-1}}$ are non-zero, thus the element mass and element stiffness matrices are of size $(2 \times 2)$ and are defined as

$$
M^{(e)} := \begin{bmatrix}
\int_{x_{i-1}}^{x_i} (\phi_i^{(e)})^2 \, dx & \int_{x_{i-1}}^{x_i} \phi_i^{(e)} \phi_i^{(e)'} \, dx \\
\int_{x_{i-1}}^{x_i} \phi_i^{(e)'} \phi_i^{(e)} \, dx & \int_{x_{i-1}}^{x_i} (\phi_i^{(e)'})^2 \, dx
\end{bmatrix},
A^{(e)} := \begin{bmatrix}
\int_{x_{i-1}}^{x_i} (\phi_i^{(e)'})^2 \, dx & \int_{x_{i-1}}^{x_i} \phi_i^{(e)'} \phi_i^{(e)''} \, dx \\
\int_{x_{i-1}}^{x_i} \phi_i^{(e)''} \phi_i^{(e)} \, dx & \int_{x_{i-1}}^{x_i} (\phi_i^{(e)''})^2 \, dx
\end{bmatrix}.
$$

Now evaluating the integrals, with $h_{i-1} = x_i - x_{i-1}$, gives

$$
M^{(e)} := \frac{1}{(x_i - x_{i-1})^2} \begin{bmatrix}
\int_{x_{i-1}}^{x_i} (x_i - x)^2 \, dx & \int_{x_{i-1}}^{x_i} (x_i - x)(x - x_{i-1}) \, dx \\
\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x_i - x) \, dx & \int_{x_{i-1}}^{x_i} (x - x_{i-1})^2 \, dx
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\int_{x_{i-1}}^{x_i} (x_i - x)^2 \, dx &== \left[-\frac{1}{3} (x_i - x)^3\right]_{x_{i-1}}^{x_i} = \frac{1}{3} (x_i - x_{i-1})^3 = \frac{1}{3} h_{i-1}^3 \\
\int_{x_{i-1}}^{x_i} (x_i - x)(x - x_{i-1}) \, dx &== \int_{x_{i-1}}^{x_i} (x_i - x - x^2 + xx_{i-1}) \, dx \\
&= \left[\frac{1}{2} x_i x^2 - x_i x_{i-1} x - \frac{1}{3} x^3 + \frac{1}{2} x^2 x_{i-1}\right]_{x_{i-1}}^{x_i} = \frac{1}{6} x_i^3 - \frac{1}{2} x_i x_{i-1} + \frac{1}{2} x_i x_{i-1} - \frac{1}{6} x_{i-1}^3 \\
&= \frac{1}{6} (x_i - x_{i-1})^3 = \frac{1}{6} h_{i-1}^3
\end{bmatrix}
$$

and similarly

$$
\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x_i - x) \, dx = \frac{1}{6} h_{i-1}^3, \quad \int_{x_{i-1}}^{x_i} (x - x_{i-1})^2 \, dx = \frac{1}{3} h_{i-1}^3
$$

$$
= \frac{1}{h_{i-1}^2} \begin{bmatrix}
\frac{1}{3} h_{i-1}^3 & \frac{1}{6} h_{i-1}^3 \\
\frac{1}{6} h_{i-1}^3 & \frac{1}{3} h_{i-1}^3
\end{bmatrix} = \frac{1}{6} h_{i-1} \begin{bmatrix}
2 & 1 \\
1 & 2
\end{bmatrix},
$$

$$
A^{(e)} := \frac{1}{(x_i - x_{i-1})^2} \begin{bmatrix}
\int_{x_{i-1}}^{x_i} (-1)^2 \, dx & \int_{x_{i-1}}^{x_i} (-1) \, dx \\
\int_{x_{i-1}}^{x_i} 1 \, dx & \int_{x_{i-1}}^{x_i} (1)^2 \, dx
\end{bmatrix} = \frac{1}{h_{i-1}^2} \begin{bmatrix}
h_{i-1} & -h_{i-1} \\
-h_{i-1} & h_{i-1}
\end{bmatrix}
$$

$$
= \frac{1}{h_{i-1}} \begin{bmatrix}
1 & -1 \\
-1 & 1
\end{bmatrix}.
$$

If the grid is equidistant, i.e. $h = h_{i-1}$ for $i = 1, \ldots, n$, the element matrices become

$$
M^{(e)} := \frac{1}{6} h \begin{bmatrix}
2 & 1 \\
1 & 2
\end{bmatrix}, \quad A^{(e)} := \frac{1}{h} \begin{bmatrix}
1 & -1 \\
-1 & 1
\end{bmatrix}.
$$
The global stiffness matrix $A^{(n+1)\times(n+1)}$ is assembled elementwise, i.e. for each element $e = I_i$ for $i = 1, \ldots, n$

$$a_{ii} = a_{ii} + a_{22}^{(e)}$$
$$a_{i,i-1} = a_{i,i-1} + a_{21}^{(e)}$$
$$a_{i-1,i} = a_{i-1,i} + a_{12}^{(e)}$$
$$a_{i-1,i-1} = a_{i-1,i-1} + a_{11}^{(e)}.$$

The global mass matrix $M^{(n+1)\times(n+1)}$ is assembled in similar fashion. The structures of the global matrices are

$$M = \frac{1}{6h} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \\ \vdots & \ddots & \ddots \\ 1 & 4 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad A = \frac{1}{h} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$}

The matrix $A$ is singular implying that there is no unique solution to problem 3.2. Now using the boundary conditions $u(x_0) = u(x_n) = 0$ it is possible to decouple $U_0$ and $U_n$ from the system of equations, since these values are known to be 0, by removing the first row, the first column, the last row and the last column in matrix $A$. Thus the system of equations that needs to be solved can be written as, where now $A^{(n-1)\times(n-1)}$ and $M^{(n-1)\times(n-1)}$.

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ \vdots & \ddots & \ddots \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \frac{1}{6h} \begin{bmatrix} 1 & 4 & 1 \\ 1 & 4 & 1 \\ \vdots & \ddots & \ddots \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{n-2} \\ U_{n-1} \end{bmatrix} = \frac{1}{h} \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{n-1} \\ f_n \end{bmatrix}.$$}

It is straightforward to extend the concepts discussed to higher spatial dimensions than one. It should be noted that the tridiagonal structure of the mass and stiffness matrices is apparent only in one dimensional problems. In higher dimensions, where the matrices still exhibit a sparse structure, there is no systematic way to number the nodes in the mesh that give tridiagonal structures, since the matrices exhibit the adjacency structure of the nodes in the mesh. In Section E the general formulation of piecewise linear approximation for problem 3.2 is extended to two spatial dimensions.

**Pricing a Multi-Asset American Option**
4 Problem setting

4.1 Multi-Asset American options in the Black-Scholes model

In the beginning of Section 2.3 some assumptions were made about the market where the assets are traded. It is assumed that there are no transaction costs, that trading takes place continuously, that the term structure of interest rates is flat and non-stochastic and that the market is efficient in the sense that there are no arbitrage possibilities.

Following [26] the Black-Scholes-Merton model with \( n \) assets follow the risk-neutral price dynamics described by the \( n \)-dimensional stochastic differential equation:

\[
dS_i(t) = (r - \delta_i)S_i(t)dt + \sigma_i S_i(t)dW_i(t), \quad S_i(0) = S_i, \quad i = 1, \ldots, n. \tag{4.1}
\]

where \( r \geq 0 \) is the risk-free interest rate, \( \sigma_i > 0 \) are the asset volatilities, \( \delta_i \geq 0 \) are asset dividends yields and \( S_i > 0 \) are initial asset prices. The \( \{W_i(t) : t \geq 0\} \) are \( n \) correlated standard Brownian Motions with the correlation matrix \( \rho_{ij} \) where

\[
dW_i(t)dW_j(t) = \rho_{ij}dt.
\]

The payoff (contract) function associated with the considered American vanilla option (standard put or call) with expiration at date \( T > 0 \) is \( \psi(t, S_i(t), \ldots, S_n(t)) \), i.e. this is the payoff delivered to the holder of the option if the option is exercised at time \( t \in [0, T] \). The payoff function is a function of the prices of the \( n \) underlying assets at the time of exercise. There are many types of multi-asset options with different payoff functions. An example of the value of the payoff function at the time of exercise for call and put options on the arithmetic average of \( n \) assets is given below.

\[
\psi(t, S_i(t), \ldots, S_n(t)) = \begin{cases} 
\left( \frac{1}{n} \sum_{i=1}^{n} S_i(t) - K \right)^+ & \text{for a call,} \\
K - \left( \frac{1}{n} \sum_{i=1}^{n} S_i(t) \right)^+ & \text{for a put,}
\end{cases}
\]

where \((\cdot)^+ := \max(\cdot, 0)\) and \( K \) is the strike price. As discussed in Section D it is sometimes convenient to use the log transform to the dimensionless variables \( X_i(t) = \ln(\frac{S_i(t)}{K}) \) where the SDE for \( S_i(t) \) is given in equation 4.1. In a similar fashion as in the proof of Proposition 2.4 it can be deduced that the solution to the SDE for the new variable is

\[
X_i(t) = x_i + \left( r - \delta_i - \frac{1}{2} \sigma_i^2 \right) t + \sigma_i W_i(t) \quad \text{with} \quad x_i = \ln\left( \frac{S_i(0)}{K} \right). \tag{4.2}
\]

Following the theory developed in Sections 2.5, 3.4 and in [26] the weak variational formulation for the penalized non-linear PDE for the American option is

\[
\left( \frac{\partial u_{\epsilon}}{\partial t}, v \right) + a(u_{\epsilon}, v) + (\pi_{\epsilon}(u_{\epsilon}), v) = (f, v)
\]
with the non-linear penalty term
\[ \pi_c(u_e) = -\left( \frac{1}{c}(u_e)^p \right) \]

and the bilinear form
\[ a(u, v) = \frac{1}{2} \int_{\Omega} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \, dx - \int_{\Omega} \sum_{i=1}^{n} \left( r - \delta_i - \frac{1}{2} \sigma_i^2 \right) \frac{\partial u}{\partial x_i} v \, dx + \int_{\Omega} ru v \, dx . \]

Also in the inhomogeneous term \( f = G \psi - r \psi \) the infinitesimal generator for the underlying process (4.2) is given by
\[ Gw := \frac{1}{2} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} \frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} \left( r - \delta_i - \frac{1}{2} \sigma_i^2 \right) \frac{\partial w}{\partial x_i} . \]

Using the boundary conditions a suitable test space for admissible functions is
\[ \mathcal{V} = \mathcal{H}_0^1(\Omega) = \left\{ v : \int_{\Omega} \left( v^2 + |\nabla v|^2 \right) \, dx < \infty, \quad v = 0 \text{ on } \partial \Omega \right\} \]

and the variational problem is: Find \( u_e \in \mathcal{V} \) such that
\[
\begin{align*}
\frac{1}{2} \int_{\Omega} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} \frac{\partial \psi}{\partial x_i} \frac{\partial v}{\partial x_j} \, dx + \int_{\Omega} \sum_{i=1}^{n} \left( r - \delta_i - \frac{1}{2} \sigma_i^2 \right) \frac{\partial \psi}{\partial x_i} v \, dx - \int_{\Omega} r \psi v \, dx \\
= \int_{\Omega} \frac{\partial u_e}{\partial t} v \, dx + \frac{1}{2} \int_{\Omega} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} \frac{\partial u_e}{\partial x_i} \frac{\partial v}{\partial x_j} \, dx - \int_{\Omega} \sum_{i=1}^{n} \left( r - \delta_i - \frac{1}{2} \sigma_i^2 \right) \frac{\partial u_e}{\partial x_i} v \, dx + \int_{\Omega} ru_e v \, dx \\
+ \int_{\Omega} \pi_c(u_e) v \, dx \quad \forall v \in \mathcal{V}
\end{align*}
\]

Solving the variational problem by the FEM the Galerkin approximation \( U_e = \sum_{J=0}^{N} u_{e,J}(t) \phi_j(x) \) is now a linear combination of the basis functions with time dependent coefficients, where \( x \in \mathbb{R}^n \) and \( N + 1 \) is the number of nodes in the \( n \)-dimensional mesh. The basis functions used are the same as in Section E. The penalty term is thus now approximated by \( \pi_c(U_e)(t) = -\sum_{J=0}^{N} \left( \frac{1}{c}(u_{e,J}(t))^p \right) \phi_j(x) \). In the FEM framework, with \( \hat{U}_e(t) := \left( \hat{U}_{e,0}(t), \ldots, \hat{U}_{e,N}(t) \right) \) , \( \hat{U}_{e,J}(t) := \frac{\partial \hat{U}_{e,J}(t)}{\partial t} \), \( \pi_c(U_e)(t) := (\pi_c(U_{e,0}(t)), \ldots, \pi_c(U_{e,N}(t)))^{\top} \), the matrix representation leads to the ODE system
\[
M \ddot{U}_e(t) + AU_e(t) + M \pi_c(U_e)(t) - F(t) = 0, \quad t \in [0, T] \tag{4.3a}
\]

with the homogeneous initial condition
\[
U_e(0) = 0. \tag{4.3b}
\]
The matrices $M$ and $A$ are the mass and stiffness matrices, where $m_{i,j} := \langle \phi_j, \phi_i \rangle$, $a_{i,j} := a(\phi_j, \phi_i)$ respectively and where the inner product in $L^2(\Omega)$ is denoted by $(\cdot, \cdot)$, i.e. $(u, v) = \int_{\Omega} u(x)v(x) \, dx$. The load vector is $F(t) := (f_0(t), \ldots, f_N(t))^T$ with $f_j(t) := -a(\psi, \phi_j)$.

Spatial discretization of the variational formulation by the finite element method reduces the American option problem to the ODE system (4.3). This numerical procedure is the finite element method-of-lines. To be able to solve the semi-discr etized system, since continuous in time, there is the need to integrate numerically w.r.t to $t$ by some time stepping scheme. (A freely available software package for numerical time integration is discussed in Section 5.)

### 4.1.1 Multi-Asset American option on the geometric average

The American-style option considered here will be an put option on the geometric average of $n$ assets for reasons that will become apparent further down in this section. The geometric average processes is defined by

$$I_t := \left( \prod_{i=1}^{n} S_i(t) \right)^{\frac{1}{n}}, \quad t \geq 0,$$

with the put payoff function $\psi_t = (K - I_t)^+$ if the option is exercised at time $t \in [0, T]$, where the maturity date $T > 0$. The geometric average $\{I_t(t) : t \geq 0\}$ of $n$ geometric Brownian motion processes $\{S_i(t) : t \geq 0\}$ is itself a Brownian motion process with volatility $\sigma_i$ and risk-neutral drift $(r - \delta_i)$. Thus the problem of valuing options on the geometric average of $n$ assets can be reduced to the one of valuing options on the one-dimensional process $I$:

$$dI_t = (r - \delta_i)I_t dt + \sigma_i I_t dB_t, \quad I_0 = \left( \prod_{i=1}^{n} S_i(0) \right)^{\frac{1}{n}}.$$

**Remark 4.1.** For readability reasons the term $I_t$ is a shorthand notation for $I(t)$ and correspondingly for the other terms.

The derivations of the volatility $\sigma_i$ and the effective dividend yield $\delta_i$ in the one-dimensional stochastic differential equation follow from **Proposition 2.4**. Thus the solution to the GBM $dI_t(t) = (r - \delta_i)I_t(t) dt + \sigma_i I_t(t) dB_t(t)$ is given by $I_t = I_0 e^{\sum_{i=1}^{n} [(r-\delta_i - \frac{1}{2} \sigma_i^2) t + \sigma_i W_i(t)]}$, where in the case of a geometric average process the initial condition is given by $I_0 = (\prod_{i=1}^{n} S_i(0))^{\frac{1}{n}}$. Rewriting the solution in the form:

$$I_t = \left( \prod_{i=1}^{n} S_i(0) \right)^{\frac{1}{n}} e^{\sum_{i=1}^{n} [(r-\delta_i - \frac{1}{2} \sigma_i^2) t + \sigma_i W_i(t)]} = I_0 e^{(r-\delta_i - \frac{1}{2} \sigma_i^2) t + \sigma_i B_t},$$

where the volatility $\sigma_i$ and the effective dividend yield $\delta_i$ are given by:

$$\sigma_i^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} \rho_{ij} \sigma_i \sigma_j, \quad \delta_i = \frac{1}{n} \sum_{i=1}^{n} \left( \delta_i + \frac{1}{2} \sigma_i^2 \right) - \frac{1}{2} \sigma_i^2,$$
and where

\[ B_t := \frac{1}{n \sigma_I} \sum_{i=1}^{n} \sigma_i W_i(t) \]

is a one-dimensional continuous martingale, see Lemma B.1 in B, with quadratic variation \( \langle B_t \rangle = t \) and thus a standard BM, the problem of valuating on the geometric average of \( n \) assets is reduced to valuating options on the one-dimensional process \( I \). The one-dimensional problem can be used as a benchmark in the analysis of the general multidimensional method for the numerical solution of the general \( n \)-asset problem. This is why a American-style option on the geometric average of \( n \) assets is considered here.

**Remark 4.2.** With the log normal transform to dimensionless variables, see equation (4.2), and the assumption that the process is in state \( x \) at time \( t \) the payoff function \( \psi_t \) is reformulated as

\[ \psi_t = \left( K - Ke^{\frac{1}{n} \sum_{i=1}^{n} x_i} \right)^+. \]
5 Implementation and numerical results

(a) The free boundary $S_f(t)$ with $R = 0.32$, $h = 0.002$ and 2500 time steps.

(b) The free boundary $S_f(t)$ together with the option value and the option payoff function with $R = 0.32$, $h = 0.002$ and 250 time steps.

Figure 8: The free boundary $S_f(t)$, the option value and the payoff function for a one dimensional American put option on the geometric average.

The problem discussed in Section 4.1.1 is solved by the PSOR method described in 3.3.1. The dimension reduction to one dimension enables the accurate computation of a benchmark since it is possible to use a very fine spatial mesh size with large number of time steps and stringent error tolerance within the time stepping in the PSOR algorithm. The aim here is to develop a parallel application that runs on any number of processors and that can be used in a more general problem setting for the American option pricing within the FEM framework. Thus computations are made for the geometric average and the arithmetic average of $n = 1, 2, 3, 4$ underlying assets (stocks) in the American put option setting. As in [26] the values of the

Pricing a Multi-Asset American Option
parameters for the numerical results in the PSOR and FEM implementations are: $K = 100$, $T = 0.25$, $\sigma_i = 0.20$, $r = 0.03$, $\delta_i = 0$ and $\rho_{ij} = 0.50$ for $i \neq j$. The benchmark was computed sequentially, i.e. on one processor, with MATLAB.

![Plots of the outer and inner mesh, the payoff function and the option value function.](image)

Figure 9: Plots of the outer and inner mesh, the payoff function and the option value function in the case of right-angled triangulation at $T = 0.25$ with $R = 0.32$ and $h = 0.04$ in the $x_i$ variables (recall that $x_i = \ln \frac{S_i}{K}$).

Before setting up $M$, $A$ and $F$ in the ODE system (4.3) the computational domain needs to be partitioned into $n$-simplices ($n$ being the spatial dimension, e.g. a 2-simplex is a triangle and a 3-simplex is a tetrahedron), since here the FEM is applied with discretizing spatially by a simplicial mesh. This done by the freely available C written open source software package Qhull. It can compute the Delaunay triangulation of a set of points in $n$-dimensions, see [4]. Here it is sufficient to just mention that the simplicial tessellation is uniquely defined if...
the Delaunay algorithm is used to define the triangulation and that MATLAB uses Qhull in some of its computational geometry functions. Once the Delaunay triangulation is computed by Qhull, with the initial distribution of points constituting a regular grid, the structure of the mesh (e.g. nodal coordinates, element adjacency, indices of boundary nodes and element nodal composition) is written to and saved in text files. In the parallel implementation this initial step, the triangulation of the spatial domain, is the only step that is done sequentially.

In creating and computing the parallel structures of $M$, $A$ and $F$ the PETSc (Portable, Extensible Toolkit for Scientific Computations) package is used. It uses MPI (Message Passage Interface) for the parallel communications and is also freely available and written in C. Further details and
documentation on the PETSc package can be found in [3]. After the mesh is generated on the 
zero rank processor and after writing and storing the mesh, tessellation structure is distributed 
in a straightforward fashion, i.e. the zero rank processor reads in a chunk of vertex and element 
(simplex) data and sends it of to each other processor, where any leftovers are assigned to and 
stored in the sending processor. Given this resulting distribution a second partitioning of the 
elements (simplices) is made with the external package PARMETiS available in PETSc, see 
[25].

![Plots of the outer and inner mesh, the payoff function and the option value function](image)

Figure 11: Plots of the outer and inner mesh, the payoff function and the option value function 
in using the MATLAB function delaunayn (where Qhull is used for the Delaunay tessellation) 
at $T = 0.25$ with $R = 0.32$ and $h = 0.04$ in the $x_i$ variables (recall that $x_i = \ln \frac{S_i}{K}$).

This new partitioning by PARMETiS is done so that the number of cut edges between the ele-
ments is reduced. A second time data is distributed according to the second partitioning but this

---

Pricing a Multi-Asset American Option
time in parallel with the `vecScatter` context available in PETSc to give the final element distribution, see [2] for discussion on managing unstructured grids with PETSc. Now that every processors holds a part of the mesh the assembly of $M$, $A$ and $F$ is done in parallel elementwise with each processor computing the nodal values for the elements that constitute its part of the mesh.

After the assembly there is the need to integrate w.r.t time. A free and parallel software suit written in C that integrates ODE and differential algebraic equation (DAE) initial value problems is SUNDIALS (SUite of Nonlinear and DIfferential/ALgebraic equation Solvers), see [11].

![Figure 12: Plots of the outer and inner mesh, the payoff function and the option value function in using the MATLAB function delaunayn (where Qhull is used for the Delaunay tessellation) at $T = 0.25$ with $R = 0.30$ and $h = 0.05$ in the $x_i$ variables (recall that $x_i = \ln \frac{S_i}{K}$).](image)

The library within the SUNDIALS family of solvers used here is the IDA (Implicit Differential-
Algebraic solver) library. In following [21] and [26] IDA solves the initial value problem for a DAE system of the general form

$$F(t, y, \dot{y}) = 0, \quad y(t_0) = y_0, \quad \dot{y}(t_0) = \dot{y}_0,$$  \hfill (5.1)

where \( t \) is the independent variable, \( y \in \mathbb{R}^n \) a vector of dependent variables, \( \dot{y} = \frac{dy}{dt} \) and where \( F: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \) called the DAE residual is a non-linear function. The ODE system given in (4.3) is thus integrated with IDA by setting \( F(t, \mathbf{U}_e, \dot{\mathbf{U}}_e) = M\mathbf{U}_e(t) + A\dot{\mathbf{U}}_e(t) + M\tau_e(U_e)(t) - F(t) \). IDA uses a variable-order, variable coefficient BDF (Backward Differentiation Formula), in fixed leading coefficient form which makes the integrator useful for stiff systems. The order of the method ranges from 1 to 5. The BDF of order \( q \) is obtained by

$$\sum_{i=0}^{q} \alpha_{n,i} \mathbf{U}_{e,n-i} = h_n \dot{\mathbf{U}}_e, \quad \text{with } h_n = t_n - t_{n-1},$$

where \( \mathbf{U}_{e,n} \) and \( \dot{\mathbf{U}}_e \) are approximations to \( \mathbf{U}_e(t_n) \) and \( \dot{\mathbf{U}}_e(t_n) \) respectively. The coefficients \( \alpha_{n,i} \) are independent of \( \mathbf{U}_e(t) \) and are uniquely determined by the order \( q \) and the history of the time steps \( \Delta t_n, \Delta t_{n-1}, \ldots, \Delta t_{n-q} \), see [16] for details on the coefficients \( \alpha_{n,i} \) and on the approximation of \( \dot{\mathbf{U}}_e(t_n) \). With BDF the DAE system (5.1) is reduced to a non-linear algebraic equation that needs to be solved at each time step from the results obtained at earlier time steps:

$$G(t_n, \mathbf{U}_{e,n}, h_n^{-1} \sum_{i=0}^{q} \alpha_{n,i} \mathbf{U}_{e,n-i}) = 0,$$

where the vector \( \mathbf{U}_{e,n} \) is unknown. The non-linear algebraic system given in the equation above is solved by Newton iteration. The Newton correction is determined iteratively by solving the linear system

$$J(\mathbf{U}_{e,n(m+1)} - \mathbf{U}_{e,n(m)}) = -G(\mathbf{U}_{e,n(m)}),$$

where \( \mathbf{U}_{e,n(m)} \) is the Newton approximation to \( \mathbf{U}_{e,n} \) in the \( m \)-th iteration. The matrix \( J \), evaluated at the previous time \( \mathbf{U}_{e,n(m)} \), is some approximation to the system Jacobian of \( G \) given by

$$J = \frac{\partial G}{\partial \mathbf{U}_e} = \frac{\partial F}{\partial \dot{\mathbf{U}}_e} + \frac{\alpha_{0,n}}{\Delta t_n} \frac{\partial F}{\partial \dot{\mathbf{U}}_e}.$$

IDA provides several choices of solvers, organized in two families, for solving the linear system within the Newton correction. The linear solver used here is the SPGMR, a scaled preconditioned GMRES (Generalized Minimal Residual method) without restarts, and belongs to the spils family (scaled preconditioned iterative (Krylov) linear solvers). In using SPGMR within IDA the user has to provide the function that computes the residual \( F \) in the DAE for given values of the independent variable \( t \), state vector \( y \), and derivative \( \dot{y} \). In the parallel implementation here all the matrix and vector operations needed for computing \( F \) were done with PETSc. Since the boundary conditions are zero Dirichlet boundary conditions the solution is known to be zero on the boundary of the computational domain and the problem needs just to be solved.

---

**Pricing a Multi-Asset American Option**
for the inner nodes of the triangulation. Thus the problem is reduced and fewer computations are needed, but this problem reductions is only made prior to solving the problem with IDA. The assembly of the mass matrix, the stiffness matrix and the load vector is done over the whole computational domain including the boundary.

For the one dimensional American put option on the geometric average the free boundary (optimal exercise boundary) \( S_f(t) \) is presented in Figure 8(a) and in Figure 8(b) \( S_f(t) \) is plotted together with the option value and the option payoff function. The numerical data needed for these plots, i.e. \( U_e \) (the value of the excess option premium over the payoff), was produced from a run on 32 processors (cores).

When computing \( U_e \) for two underlying assets, i.e. a 2-D problem, the triangulation of the spatial domain was done in three different ways. The first type of triangulation consisting of right-angled triangles, given in Figure 9(a), is done in using the MATLAB computational engine. The inner mesh, where the problem is solved by IDA, together with the boundary elements is also visualized. The payoff function and the option value at expiration are given given in Figure 9(b) and Figure 9(c), respectively. In Figure 9(d) the option value is plotted together with the payoff. The second type uses Delaunay triangulation iteratively to produce a high quality mesh with the DISTMESH algorithm, see [40]. The algorithm is written in MATLAB and uses delaunayn, a MATLAB function utilizing Qhull for its Delaunay triangulation. Here DISTMESH was used with the MATLAB computational engine except for the Delaunay triangulation step within the algorithm. Instead Qhull was used for this step, giving the same meshing results but reducing the wall-time needed for the triangulation step by a factor of more than 2 for a 2-D and 3-D problem (4-D was never tested with delaunayn). This reduces wall-time significantly since DISTMESH is an iterative algorithm where the Delaunay triangulation is the most time consuming part. Analogously as for the right-angled mesh Figure 10 presents results obtained with DISTMESH. A very useful feature in DISTMESH its ability to construct constrained triangulation making it possible to triangulate complex regions, a feature that is not implemented in Qhull. The algorithm does perturbate the positions of the input points in order to increase the quality of the mesh, but it is possible to fix points of choice by the

Figure 13: Cross-section plots of the 3-D Delaunay triangulation with \( R = 0.32 \) and \( h = 0.04 \).
user. In Figure 10(a) the center point, i.e. $S_1 = S_2 = 100$, of the computational domain was fixed since the solution at that point will be compared with the solution obtained by the benchmark. The algorithm was terminated after 11 Delaunay retriangulations and if more triangulation were used the quality of the simplices containing the center point would have been higher than what is seen in Figure 10(a).

Figure 14: The free boundary $S_f(t)$ for a two dimensional American put option on the geometric average.

Lastly the mesh was generated with Qhull with the results given in Figure 11. Depending on the computational domain radius $R$ and the step size $h$ for the $x_i$ variable the triangulation with Qhull will give different connectivity patterns in the resulting mesh. This can be visualized in comparing Figure 11(a) with Figure 12(a). For $R = 0.32$ and $h = 0.002$, being the smallest

---

Pricing a Multi-Asset American Option
step size used in 2-D, the latter connectivity pattern is the resulting one. Also the connectivity patterns for a 3-D Delaunay triangulation are apparent in Figure 13. Since Figures 9, 10, 11, 12 and 13 are mainly aimed at visually emphasizing the different geometrical characteristics of

<table>
<thead>
<tr>
<th>Number of assets (n)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometric average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benchmark price</td>
<td>3.66778</td>
<td>3.18468</td>
<td>3.00447</td>
<td>2.90979</td>
</tr>
<tr>
<td>FEM price Qhull</td>
<td>3.66776</td>
<td>3.18515</td>
<td>3.05238</td>
<td>2.94941</td>
</tr>
<tr>
<td>Error</td>
<td>2 · 10⁻⁵</td>
<td>4.7 · 10⁻⁴</td>
<td>4.8 · 10⁻²</td>
<td>4.0 · 10⁻²</td>
</tr>
<tr>
<td>FEM price DISTMESH</td>
<td>–</td>
<td>3.18567</td>
<td>3.00801</td>
<td>2.92226</td>
</tr>
<tr>
<td>Error</td>
<td>–</td>
<td>9.9 · 10⁻⁴</td>
<td>3.5 · 10⁻³</td>
<td>1.2 · 10⁻²</td>
</tr>
<tr>
<td>FEM price Right-angled</td>
<td>–</td>
<td>3.18464</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Error</td>
<td>–</td>
<td>4.0 · 10⁻⁵</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>Arithmetic average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEM price Qhull</td>
<td>3.66780</td>
<td>3.14047</td>
<td>2.99199</td>
<td>2.88134</td>
</tr>
<tr>
<td>FEM price DISTMESH</td>
<td>–</td>
<td>3.14029</td>
<td>2.94777</td>
<td>2.85496</td>
</tr>
<tr>
<td>FEM price Right-angled</td>
<td>–</td>
<td>3.13960</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

### Mesh size (h) | 0.002 | 0.002 | 0.02 | 0.04 |
### Number of nodes in comp. domain | 321 | 103041 | 35937 | 83521 |

| **Qhull** |   |   |   |   |
| Mesh generation time (sec.) | 10.07 | 20.74 | 21.84 | 305.33 |
| Number of elements (n-simplices) | 320 | 204800 | 196608 | 1556307 |

| **DISTMESH** |   |   |   |   |
| Number of nodes in comp. domain | – | 118585 | 35937 | 83521 |
| Number of iterations | – | 11 | 51 | 51 |
| Mesh generation time (sec.) | – | 77.26 | 285.26 | 3303.13 |
| Number of elements (n-simplices) | – | 235793 | 197924 | 2096101 |

| **Right-angled triangulation** |   |   |   |   |
| Right-angled mesh gen. time (sec.) | – | 46.34 | – | – |
| Number of elements (n-simplices) | – | 204800 | – | – |

| **Computation time** |   |   |   |   |
| Sequential wall-time (sec.) | – | 2022.35 | 340.57 | 4280.65 |
| Shortest wall-time (sec.) | 0.71 | 6.41 | 5.43 | 32.42 |
| Number of cores | 1 | 100 | 40 | 60 |
| IDA tolerance | 10⁻⁶ | 10⁻⁶ | 10⁻⁶ | 10⁻⁶ |
| Penalty parameter ε | 10⁻⁵ | 10⁻⁴ | 10⁻⁴ | 10⁻⁴ |

Table 1: Numerical results for the three month American put option on the geometric and arithmetic averages. Numerical parameters: \( S_i = K = 100 \), \( T = 0.25 \), \( r = 0.03 \) (annual interest rate), \( \ln S_i / K = x_i \in \Omega := [-0.32, 0.32]^n \) and \( p = 2 \) for the penalty term \( \pi_\varepsilon \).

*For the geometric average with mesh generation time not included.

Pricing a Multi-Asset American Option
the mesh structures within the different triangulation techniques employed a large $h$ was used. The optimal stopping boundary for each 2-D triangulation at different times $t$ for the American option on the geometric average is plotted in Figure 14. In Figure 14(c) a zoomed in part of the free boundary is presented in able to visualize the un-smooth structure. In Figure 14(d) the same part is plotted but this time the data was firstly smoothed out with the MATLAB function `sgolayfilt` before producing the plot. This function performs the Savitzky-Golay filtering, a method that basically does local smoothing while maintaining the shape of the curve. (Here the framesize was set to 15 with a second order filter). Also the numerical data for the 2-D problem was produced on 32 cores. The implementational results obtained are summarized in Table 1 and the speedup $(S_p = \frac{T_1}{T_p})$ and the efficiency $(E_p = \frac{T_1}{pT_p})$ of the parallel implementation are given in Figure 15. Here $T_1$ is the sequential execution time and $T_p$ the parallel execution time on $p$ processors.

![Speedup and Efficiency](image)

**Figure 15:** Speedup and Efficiency for the mesh sizes given in Table 1.

*Pricing a Multi-Asset American Option*
It should be noted that timing for producing the mesh is not included in Figure 15, since for a given dimension the mesh and the files defining its structure are produced only once and reused. It is evident from the plots that adding more cores for solving a problem of fixed size would just result in more time being spent for solving the problem, this results from the increase of communication and now the communication costs become the bottleneck. All the development with implementation and the numerical experiments presented here was done on the Kalkyl cluster at UPPMAX (Uppsala Multidisciplinary Center for Advanced Scientific Computations). The Kalkyl cluster, see Figure 16, consists of 2784 CPU cores located in 348 computer nodes, where each node consists of two Quad-core Intel® Xeon 5520 Nehalem 2.26 Ghz, 8 MB cache processors. The nodes are interconnected with a 4:1 oversubscribed DDR Infiniband fabric and the memory is 9504 GB of total RAM and 113 TB of total disk, see [47] for more details on Kalkyl and other system resources available at UPPMAX.

Figure 16: The Kalkyl cluster and the Panasas parallel file storage system at UPPMAX.
6 Conclusion and discussion

Solving the general high-dimensional American put option pricing problem numerically introduces challenging computational problems, since for every added asset (dimension) the need for computational resources and memory grows exponentially, a phenomenon known as the curse of dimensionality. To develop an application that meets these needs and overcomes this complexity would almost decisively involve the utilization of parallelism. Here an application was developed that is almost fully parallel. Only the initial step, the simplicial spatial discretization of the computational domain, runs sequentially and thus restricting the number of dimensions for which the problem is solvable within reasonable timing limits. Also, as suggested by the numerical results in Table 1, depending on the simplicial meshing procedure employed the solution at the center point varies in accuracy, especially for high dimensional problems where choosing a large spatial step size in comparison to the computational domain radius is unavoidable for keeping mesh generation time down. On the other hand the mesh needs to be constructed only once for a certain stepsize and computational domain radius. As in [26] it is possible to perform a more extensive error analysis. The approximation error between the computed solution and the benchmark is computed in the maximum norm over the approximation domain $G$, see section 2.5.3, thus reducing the computations needed.

In can be seen from Figures 15(a) and 15(b) that the speedup is superlinear and that the efficiency is greater than hundred percent. Often these results are given by a more efficient resource utilization, e.g. caching effects. When using multiple processors the problem size per processor is reduced and more or even all the core data set can fit into the caches giving higher cache hit rate. The wall-time (mesh generation not included) for the 4D problem when run on 5 cores with the allocation of 72 GB of RAM was 287.20 seconds whereas when no RAM allocation was made the wall-time was 752.84 seconds. (The Kalkyl cluster consists of 316 nodes with 24 GB, 16 nodes with 48 GB and 16 nodes with 72 GB of RAM available). Also running the 4D problem sequentially was done on a 72 GB node, since if run otherwise the core starts to thrash memory, i.e. the processor starts using the swap memory. Thus in considering Figure 15 more justified results for the speedup and efficiency would have been obtained in the cases when say $p < 10$, if enough RAM to fit the problem was allocated prior to execution. Nevertheless utilizing parallelization seems to be a necessity when dealing with large numerical problems if the problems are to be solved within reasonable time.

Since the problems discussed here do not exhibit any complex geometry or boundaries a structured grid could be used instead of a simplicial mesh as described in [26]. One advantage of triangular discretization as mentioned earlier is the ease of implementation when pricing exotic options since irregular barriers such as time-dependent barriers are handled easily. A feature that is not associated with quadrilateral grids in general. As shown in the previous section and as mentioned in [30] the mesh generation within the finite element method is indeed a non-trivial task. There has been a lot of research involving the development of mesh free methods for solving PDEs where the problem is solved with the use of RBFs (radial basis functions). A mesh free parallel implementation could significantly speed up the computation and increase the dimension for which the American put option pricing problem is solved.

As suggestions on expanding the framework developed here, some of the assumption made within the Black-Scholes-Merton model approximate real life events crudely, e.g. constant volatilities. A more realistic approach would be to model the volatility by a stochastic pro-
cesses, as function of time and space. Also it is assumed that in the SDEs describing the dynamics of the asset prices the disturbance terms are given by Brownian Motions. A better approximation would be a processes that allow jumps, since asset prices in real life could have sudden jumps, e.g. see [35]. More complexity of this form can be handled easily within the parallel implementation developed here, since it is possible to utilize the computational power available on a cluster such as Kalkyl.
A  The variation of a Brownian Motion

Let \( f : [a, b] \rightarrow \mathbb{R} \) be a real-valued function defined on \( t \in [a, b] \) and define the partition of the interval as \( \Delta_n := \{a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b\} \). The mesh of the partition is the length of the largest interval in the partition and is thus defined by

\[
\|\Delta_n\| := \max_{1 \leq i < n}(t_i - t_{i-1}),
\]

and for every \( p > 0 \), let

\[
Q_p(f; a, b, \Delta_n) := \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p.
\]

There are some definitions that need be considered when investigating the limiting behavior of \( Q_p(W; a, b, \Delta_n) \) where \( W \) is a Brownian Motion (or, alternatively, Wiener process).

**Definition A.1.** The true \( p \)-th variation of \( f \) on \([a, b]\) is defined as

\[
V_p(f; a, b) := \sup_{\Delta_n} Q_p(f; a, b, \Delta_n)
\]

where the supremum is taken over all possible partitions of \([a, b]\).

The function \( f \) has finite true \( p \)-th variation on \([a, b]\) if \( V_p(f; a, b) < \infty \) and when \( p = 1 \), i.e. if \( V_1(f; a, b) < \infty \) the function \( f \) is said to be of bounded variation or its total variation on \([a, b]\) is said to be finite. It is shown in [46] that for a Brownian Motion \( V_p(f; a, b) < \infty \iff p > 2 \). If the restriction to only consider those sequences of partitions \( \{\Delta_n\} \) for which \( ||\Delta_n|| \to 0 \) is made, instead of considering the supremum over all possible partitions, the 2nd or alternatively the quadratic variation for a Brownian Motion is finite, see **Theorem A.1** below.

**Theorem A.1.** If \( \{\Delta_n, n = 1, 2, 3, \ldots\} \) is a sequence of partitions of \([a, b]\), then

\[
Q_2(W; a, b, \Delta_n) \to b - a \quad \text{in L}^2
\]

as \( ||\Delta_n|| \to 0 \).

**Proof.** Notice that \( \sum_{i=1}^{n}(t_i - t_{i-1}) = (b - a) \), and let

\[
Y_n = \sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}|^2 - (b - a) = \sum_{i=1}^{n} \left[ |W_{t_i} - W_{t_{i-1}}|^2 - (t_i - t_{i-1}) \right] = \sum_{i=1}^{n} X_i \tag{A.1}
\]

where \( X_i = |W_{t_i} - W_{t_{i-1}}|^2 - (t_i - t_{i-1}) \) and note that

\[
Y_i^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j = \sum_{i=1}^{n} X_i^2 + 2 \sum_{i<j} X_i X_j.
\]

---

**Pricing a Multi-Asset American Option**
Taking the expected value of $Y_n^2$ and taking into consideration that a Brownian Motion has independent increments gives for $i \neq j$

$$E[X_iX_j] = E \left\{ \left[ \sum_{i=1}^{n} (|W_{t_i} - W_{t_{i-1}}|^2 - (t_i - t_{i-1})) \right] \left[ \sum_{j=1}^{n} (|W_{t_j} - W_{t_{j-1}}|^2 - (t_j - t_{j-1})) \right] \right\}$$

$$= E \left\{ \sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}|^2 \sum_{j=1}^{n} |W_{t_j} - W_{t_{j-1}}|^2 - \sum_{i=1}^{n} (t_i - t_{i-1}) \sum_{j=1}^{n} |W_{t_i} - W_{t_{i-1}}|^2 \right\}$$

$$- \sum_{i=1}^{n} (t_i - t_{i-1}) \sum_{j=1}^{n} |W_{t_j} - W_{t_{j-1}}|^2 + \sum_{i=1}^{n} (t_i - t_{i-1}) \sum_{j=1}^{n} (t_j - t_{j-1})$$

$$= \sum_{i=1}^{n} E|W_{t_i} - W_{t_{i-1}}|^2 \sum_{j=1}^{n} E|W_{t_j} - W_{t_{j-1}}|^2 - \sum_{i=1}^{n} (t_i - t_{i-1}) \sum_{j=1}^{n} E|W_{t_i} - W_{t_{i-1}}|^2$$

$$- \sum_{i=1}^{n} (t_i - t_{i-1}) \sum_{j=1}^{n} E|W_{t_j} - W_{t_{j-1}}|^2 + \sum_{i=1}^{n} (t_i - t_{i-1}) \sum_{j=1}^{n} (t_j - t_{j-1})$$

$$= (b - a)^2 - (b - a)^2 - (b - a)^2 + (b - a)^2 = 0,$$

hence $E[Y_n^2] = \sum_{i=1}^{n} E(X_i^2)$ and taking the expected value of the last expression gives

$$E[X_i^2] = E \left\{ \left[ |W_{t_i} - W_{t_{i-1}}|^2 - (t_i - t_{i-1}) \right]^2 \right\}$$

$$= E \left\{ (W_{t_i} - W_{t_{i-1}})^4 - 2(t_i - t_{i-1}) (W_{t_i} - W_{t_{i-1}})^2 + (t_i - t_{i-1})^2 \right\}$$

$$= E(W_{t_i} - W_{t_{i-1}})^4 - 2E(t_i - t_{i-1})E(W_{t_i} - W_{t_{i-1}})^2 + E(t_i - t_{i-1})^2$$

$$= 3(t_i - t_{i-1})^2 - 2(t_i - t_{i-1})^2 + (t_i - t_{i-1})^2$$

$$= 2(t_i - t_{i-1})^2$$

where the fact that the fourth central moment of a normally distributed random variable with mean zero and variance $(t_i - t_{i-1})$ is $3(t_i - t_{i-1})^2$ has been used. Thus,

$$E[Y_n^2] = \sum_{i=1}^{n} E(X_i^2) = 2 \sum_{i=1}^{n} (t_i - t_{i-1})^2 \leq 2\|\Delta_n\|^2 \sum_{i=1}^{n} (t_i - t_{i-1}) = 2\|\Delta_n\|(b - a) \rightarrow 0$$

as $\|\Delta_n\| \rightarrow 0$. Thus as a result from the computations above it follows that $Y_n \rightarrow 0$ in $L^2$ and by recalling equation (A.1) it is easily seen that $Q_2(W; a, b, \Delta_n) \rightarrow b - a$ in $L^2$ as $\|\Delta_n\| \rightarrow 0$, which completes the proof.

By Theorem A.1 the quadratic variation of a Brownian Motion is defined as this limit, see definition below.
Definition A.2. The \textit{quadratic variation} of a Brownian Motion \( W \) on the interval \([a, b]\) is defined as

\[
Q_2(W; a, b) := \lim_{||\Delta_n|| \to 0} Q_2(W; a, b) \text{ in } L^2.
\]

Corollary A.2. If \( \{\Delta_n, n = 1, 2, 3, \ldots\} \) is a sequence of partitions of \([a, b]\) with

\[
\sum_{n=1}^{\infty} ||\Delta_n|| < \infty
\]

then

\[
Q_2(W; a, b) \to b - a.
\]

Proof. Given \( \varepsilon > 0 \) and recalling the Chebychev inequality, for proof and details see \cite{45}, which states that for any random variable \( X \)

\[
P(|X| \geq a) \leq \frac{1}{a^2} E[X^2],
\]

where \( a > 0 \) it follows that

\[
\sum_{n=1}^{\infty} P(|Y_n| \geq \varepsilon) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2} E[Y_n^2] \leq \frac{2(b - a)}{\varepsilon^2} \sum_{n=1}^{\infty} ||\Delta_n|| < \infty
\]

where the notation in the proof of the previous theorem has been used. The Borel-Cantelli lemma states that, for proof and details see \cite{45}: \textit{Given a collection of events} \( \{A_n; n \geq 1\} \) \textit{and letting} \( A \) \textit{be the event} \( \{A_n \text{ i.o.}\} \), \textit{where i.o. is shorthand for infinitely often, i.e. that infinitely many of the} \( A_n \) \textit{occur. If the sum of the probabilities of the} \( A_n \) \textit{is finite, i.e. if} \( \sum_{n=1}^{\infty} P(A_n) < \infty \), \textit{then the probability for} \( A \) \textit{is zero, i.e.} \( P(A) = 0 \). \textit{By the Borel-Cantelli lemma} it can be concluded that \( Y_n \to 0 \).

Corollary A.3. If \( \{\Delta_n, n = 1, 2, 3, \ldots\} \) is a sequence of partitions of \([a, b]\) then

\[
Q_1(W; a, b, \Delta_n) = \sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}| \to \infty
\]

as \( ||\Delta_n|| \to 0 \), i.e. almost all Brownian Motion paths are of unbounded variation on every time interval.

Proof. The proof follows from contradiction. If \( W \) is a function of bounded variation with \( V_1(W; a, b) \) denoting the total variation of \( W \) on \([a, b]\), see Definition A.1 earlier in this section, it follows that

\[
\sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}|^2 \leq \max_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}| \leq \max_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \cdot V_1(W; a, b).
\]

Since \( W \) is continuous on \([a, b]\) it is necessarily uniformly continuous on \([a, b]\), hence

\textit{Pricing a Multi-Asset American Option}
\[
\max_{1 \leq i \leq n} |W_i - W_{i-1}| \to 0 \quad \text{as} \quad \|\Delta_n\| \to 0.
\]

It then follows that
\[
\sum_{i=1}^{n} |W_i - W_{i-1}|^2 \to 0,
\]
but this is in contradiction to Corollary A.2, which completes the proof.

\section{B Martingales}

The theory given here follows the brief discussion regarding martingales in [7]. Considering a given filtration \( \{\mathcal{F}_t\}_{t \geq 0} \), i.e. \( \mathcal{F}_t \) can be viewed as the information generated by all observed events up time \( t \), let the expected value \( E[X|\mathcal{F}_t] \) of any stochastic variable \( X \) denote the expected value of \( X \), given the information available at time \( t \). The following proposition gives some rules of calculation.

\begin{proposition}
Assume that \( X \) and \( Y \) are stochastic variables and that \( Y \) is adapted to the filtration \( \{\mathcal{F}_t\}_{t \geq 0} \) (see Definition 2.3), then
\[
E[X \cdot Y|\mathcal{F}_t] = X \cdot E[Y|\mathcal{F}_t].
\]
In the above formula the expected value is taken given the information available at time \( t \) and since \( X \) is adapted to the filtration \( \{\mathcal{F}_t\}_{t \geq 0} \) its value is known at time \( t \) and can be treated as a constant.

Assume that \( X \) is a stochastic variable and that \( s < t \), then
\[
E[E[X|\mathcal{F}_t]|\mathcal{F}_s]] = E[X|\mathcal{F}_s].
\]
\end{proposition}

Given these computational rules the martingale concept can be defined.

\begin{definition}
If the following conditions hold, a stochastic process \( X \) is a \((\mathcal{F}_t)\)-martingale:
\begin{enumerate}
\item \( X \) is adapted to the filtration \( \{\mathcal{F}_t\}_{t \geq 0} \).
\item \( E[|X(t)|] < \infty, \quad \forall t. \)
\item For \( \forall s, t \) such that \( s \leq t \)
\[
E[X(t)|\mathcal{F}_s]] = X(s).
\]
\end{enumerate}
A super-martingale is a process satisfying the inequality
\[
E[X(t)|\mathcal{F}_s] \leq X(s), \quad \forall s, t \text{ with } s < t,
\]
\end{definition}
and a process satisfying
\[ E[X(t)|\mathcal{F}_s] \geq X(s), \quad \forall s, t \text{ with } s \leq t \]
is called a sub-martingale.

The most important condition in the above definition of a process as a \((\mathcal{F}_t)\)-martingale is the third one. It states that the expected future value of the stochastic process is equal to the value observed today, given the available information today, i.e. a martingale has no systematic drift.

**Corollary B.1.** The process \(X\) defined by
\[ X(t) = \int_0^t g(s)dW(s), \]
where \(g \in L^2\) and \(W\) is a Brownian Motion, is a \((\mathcal{F}_t)\)-martingale, for proof see [7].

**Lemma B.1.** Assuming enough integrability, i.e. \(g \in L^2\), a stochastic process \(X\) is a martingale iff the stochastic differential of the process discussed has no \(dt\)-term, i.e. the stochastic differential for \(X\) is given by
\[ dX(t) = g(t)\,dW(t). \]

## C The Banach fixed point theorem

Before giving the theorem and the proof, in following [53] and [54], there is need for some prerequisites regarding metric spaces and sequences in form of theorems and definitions.

**Definition C.1.** Given some set \(X\) and any \(x, y, z \in X\), a function \(d : X \times X \to \mathbb{R}\) is called a metric or a distance function on \(X\), if the following properties are satisfied

i) \(d(x, y) \geq 0\), where \(d(x, y) = 0\) iff \(x = y\).

ii) \(d(x, y) = d(y, x)\) \(\forall x, y \in X\).

iii) \(d(x, y) \leq d(x, z) + d(y, z)\) \(\forall x, y, z \in X\). (This condition is also known as the triangle inequality.)

The pair \((X, d)\) is called a metric space.

Two examples of commonly used distance functions \(d(x, y)\) on \(\mathbb{R}^n\) are the standard and the Euclidean metric, i.e. \(d(x, y) = \sum_{i=1}^n |x_i - y_i|\) and \(d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}\) respectively. (The metric space in both cases is denoted by \((\mathbb{R}^n, d)\).

**Definition C.2.** A sequence \(\{x_n\}_{n=1}^{\infty}\) converges to a limit \(x \in \mathbb{R}\) if for every \(\varepsilon > 0\), there is an integer \(N_\varepsilon\) such that for every \(n \geq N_\varepsilon\) implies \(d(x_n, x) < \varepsilon\). Usually a limit is denoted by \(\lim_{n \to \infty} x_n = x\) or \(x_n \to x\). (A sequence \(\{x_n\}_{n=1}^{\infty}\) diverges if it does not converge.)

**Theorem C.1.** A convergent sequence has at most one limit.

**Proof.** The proof follows from contradiction. Suppose that \(x\) and \(y\) are both limits of the sequence \(\{x_n\}_{n=1}^{\infty}\), i.e. that \(x_n \to x\) and \(x_n \to y\). If \(x \neq y\) then \(d(x, y) > 0\) and now choose \(0 < \varepsilon < \frac{d(x, y)}{2}\). By definition there is a \(N_x\) and \(N_y\) such that \(n \geq N_x\) implies \(d(x_n, x) < \varepsilon\) and

**Pricing a Multi-Asset American Option**
n \geq N_y$ implies $d(x_n, y) < \epsilon$. Therefore, for any $n > \max \{N_x, N_y\}$, in using the triangle inequality
\[ d(x, y) \leq d(x, x_n) + d(x_n, y) < \epsilon + \epsilon = 2\epsilon, \]
which contradicts the choice of $\epsilon < \frac{d(x, y)}{2}$. ■

**Definition C.3.** A sequence $\{x_n\}_{n=1}^{\infty}$ is nonincreasing if $x_n \leq x_{n+1}$ $\forall n$ and increasing if $x_n < x_{n+1}$. Similarly, a sequence $\{x_n\}_{n=1}^{\infty}$ is nondecreasing if $x_n \geq x_{n+1}$ $\forall n$ and decreasing if $x_n > x_{n+1}$. (A sequence which is neither nondecreasing or nonincreasing is called a monotonic sequence).

The sequence $\{x_n\}_{n=1}^{\infty}$ is said to be bounded (above or below) if $\{x_n : n = 1, 2, 3, \ldots\}$ is bounded (above or below). Given a set $X = x_n : n \in \mathbb{N}$ the number $M \in \mathbb{R}$ is the supremum of $X$ if $M$ is an upper bound of $X$ and $M \leq y$ for any upper bound of $X$. Also given a set $X = x_n : n \in \mathbb{N}$ the number $M \in \mathbb{R}$ is the infimum of $X$ if $M$ is a lower bound of $X$ and $M \geq x$ for any lower bound of $X$. The supremum, also called the least upper bound and the infimum, sometimes called the greatest lower bound are denoted by $\sup X$ (or $\sup_n x_n$) and $\inf X$ (or $\inf_n x_n$) respectively.

**Theorem C.2.** Suppose that a sequence $\{x_n\}_{n=1}^{\infty}$ is nondecreasing and bounded above. Then the least upper bound is the limit of $\{x_n\}_{n=1}^{\infty}$, i.e. $x_n \to x$ where $x = \sup_n x_n$.

**Proof.** Suppose that $\sup_n x_n = x$ and choose any $\epsilon > 0$. Then there exists a $n_\epsilon$ such that $x - \epsilon \leq x_{n_\epsilon}$. Since $\{x_n\}_{n=1}^{\infty}$ nondecreasing implies $x_n \leq x \forall n \geq n_\epsilon$. This implies that
\[ x - \epsilon \leq x_n \leq x \leq x + \epsilon \quad \forall n \geq n_\epsilon. \]
Thus $x_n \to x$. ■

**Theorem C.3.** Suppose that a sequence $\{x_n\}_{n=1}^{\infty}$ is nonincreasing and bounded below. Then the greatest lower bound is the limit of $\{x_n\}_{n=1}^{\infty}$, i.e. $x_n \to x$ where $x = \inf_n x_n$.

**Proof.** The proof is similar to the previous one. Suppose that $\inf_n x_n = x$ and choose any $\epsilon > 0$. Then there exists a $n_\epsilon$ such that $x_{n_\epsilon} \leq x + \epsilon$. Since $\{x_n\}_{n=1}^{\infty}$ nonincreasing implies $x_n \leq x_n \forall n \geq n_\epsilon$. This implies that
\[ x + \epsilon \geq x_n \geq x \geq x - \epsilon \quad \forall n \geq n_\epsilon. \]
Thus $x_n \to x$. ■

Even if it is not possible to specify the limit for a sequence the following definition gives a necessary and sufficient condition for the convergence of the sequence.

**Definition C.4.** A sequence $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence if $\forall \epsilon > 0$, there is an integer $N$ such that $\forall m, n \geq N$ implies $d(x_m, x_n) < \epsilon$.

**Theorem C.4.** If the sequence $\{x_n\}_{n=1}^{\infty}$ converges then it is a Cauchy sequence and if $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence then the sequence converges.

**Proof.** In proving the first part of the theorem above assume that $x_n \to x$ and choose any $\epsilon > 0$. The assumption $x_n \to x$ implies that there is an integer $N$ such that for any $n \geq N$ implies $d(x_n, x) < \frac{\epsilon}{2}$. Thus for any $m, n \geq N$
\[ d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \]

---

Pricing a Multi-Asset American Option
The second part part is proven in three steps in assuming that \( \{x_n\}_{n=1}^{\infty} \) is a Cauchy sequence. For each \( n \) let \( B_n = \{x_i : i \geq n\} \). Firstly there is the need to show that each \( B_n \) is bounded. Since the sequence is a Cauchy sequence and in choosing any \( \varepsilon > 0 \) there is a integer \( N \) such that \( d(x_m, x_{N+1}) < \varepsilon \) for \( m \geq N \). This implies that \( x_m < x_{N+1} + \varepsilon \) for \( m \geq N \) and therefore \( x_m < \max\{x_n, \ldots, x_{N+1}\} + \varepsilon \) for all \( x_m \in B_n \). In a similar fashion \( x_m > \min\{x_n, \ldots, x_{N+1}\} - \varepsilon \).

Now in letting \( b_n = \sup_n B_n \) it is possible to show that \( b_n \) converges. Since \( B_n+1 \subset B_n \) implies that \( b_{n+1} \leq b_n \) and also \( b_n \geq \inf B_1 \) for all \( n \). Thus since \( b_n \) is nonincreasing and bounded below it is possible to apply Theorem C.3 giving that \( b_n \to b \), where \( b = \inf_n b_n \) and \( n = 1, 2, \ldots \).

Lastly there is the need to show that \( x_n \to b \). In choosing any \( \varepsilon > 0 \) there is a \( N_x \) such that \( m, n > N_x \) implies \( d(x_m, x_n) < \frac{\varepsilon}{3} \). Since \( b_n \to b \) there is also a \( N > N_x \) such that \( d(b_N, b) < \frac{\varepsilon}{3} \). Also since \( b_N = \inf N B_N \) there is a \( m > N \) such that \( d(x_m, b_N) < \frac{\varepsilon}{3} \). Thus

\[
d(x_n, b) \leq d(x_n, x_m) + d(x_m, b_N) + d(b_N, b) < \varepsilon \frac{1}{3} + \varepsilon \frac{1}{3} + \varepsilon \frac{1}{3} = \varepsilon.
\]

Definition C.5. A metric space \((X, d)\) is said to be a complete metric space if every Cauchy sequence \( \{x_n\}_{n=1}^{\infty} \) from \( X \) converges to a point in \( X \), i.e. there is an \( x \in X \) such that \( x_n \to x \).

The space \( \mathbb{R}^n \) is complete for each of the following metrics: \( d_1(x, y) = \sum_{i=1}^n |x_i - y_i| \), \( d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \) and \( d_{\infty}(x, y) = \max\{|x_i - y_i| : i = 1, \ldots, n\} \).

Definition C.6. Suppose that \((X, d)\) is a metric space. A function \( f : X \to X \) is said to be a contraction mapping if there is a constant \( 0 < \beta < 1 \) such that

\[
d(f(x), f(y)) \leq \beta d(x, y) \quad \forall x, y \in X.
\]

Definition C.7. Let \( X \) be some interval of \( \mathbb{R} \). A function \( f : X \to \mathbb{R} \) is said to be continuous on \( X \) if for every \( x \in X \) and \( \varepsilon > 0 \) there is a \( \delta > 0 \), depending on both \( x \) and \( \varepsilon \), such that \( d(f(x), f(y)) < \varepsilon \) whenever \( d(x, y) < \delta \) in \( X \). If however it is possible to find a number \( \delta \) that depends only on \( \varepsilon \) such that \( d(f(x), f(y)) < \varepsilon \) and \( d(x, y) < \delta \) then the function \( f \) is said to be uniformly continuous on \( X \).

Given the last two definitions above it is now possible to check that a contraction mapping is uniformly continuous. In given \( 0 < \beta < 1 \) and \( \varepsilon > 0 \) and choosing \( \delta = \frac{\varepsilon}{\beta} \) implies that if \( d(x, y) < \delta \) then \( d(f(x), f(y)) \leq \beta d(x, y) < \beta \delta = \varepsilon \).

It is now possible possible to give the main theorem of this section.

Theorem C.5. (Banach fixed point theorem). Suppose \( f : X \to X \) is a contraction mapping on a complete metric space \( X \neq \emptyset \) then \( f \) has a unique fixed point, i.e. there is a \( x \in X \) such that \( f(x) = x \).

Proof. Choose an arbitrary point \( x_0 \in (X, d) \) and for any \( n > 0 \) define \( x_n \) recursively, i.e. by

\[
x_n = f(x_{n-1}) \quad n = 1, 2, 3, \ldots
\]

Since \( f \) is a contraction there is a \( \beta < 1 \) such that, in using induction

\[
d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) < \beta d(x_n, x_{n-1}) < \beta d(x_1, x_0)
\]

induction step

Pricing a Multi-Asset American Option
Now in considering any $\varepsilon > 0$ find a $N$ sufficiently large so that $\beta^N < \varepsilon \frac{1-\beta}{d(x_1, x_0)}$. Then for any \(n, m > N\) where $n > m$

\[
d(x_n, x_m) \leq d(x_n, x_{n-1}) + d(x_{n-1}, x_{n-2}) + \cdots + d(x_{m+1}, x_m) = \sum_{k=m}^{n-1} d(x_{k+1}, x_k)
\]

\[
< \beta^{n-1}d(x_1, x_0) + \beta^{n-2}d(x_1, x_0) + \cdots \beta^m d(x_1, x_0) = \sum_{k=m}^{n-1} \beta^k d(x_1, x_0)
\]

\[
= d(x_1, x_0)\beta^m \sum_{k=0}^{n-m-1} \beta^k < \beta^m d(x_1, x_0)\sum_{k=0}^{\infty} \beta^k = \beta^m \frac{d(x_1, x_0)}{1-\beta}
\]

\[
< \beta^N \frac{d(x_1, x_0)}{1-\beta} < \varepsilon \frac{1-\beta}{d(x_1, x_0)} \frac{d(x_1, x_0)}{1-\beta} = \varepsilon.
\]

This shows that the sequence \(\{x_n\}_{n=1}^\infty\) is a Cauchy sequence, thus there exists an \(x \in X\) such that \(x_n \to x\). Also since \(f\) is continuous implies \(f(x_n) = f(x)\) but since each \(f(x_n) = x_{n+1}\) and \(x_{n+1} \to x\) it follows that \(f(x) = x\).

In completing the proof it remains to show that \(x\) is unique. Suppose that \(f(x) = x\) and \(f(y) = y\) for \(x \neq y\), i.e. that \(f\) has another fixed point \(y\) that is different from \(x\). Then

\[
0 < d(x, y) = d(f(x), f(y)) < \beta d(x, y) < d(x, y),
\]

which is not possible since it is a contradiction. Therefore \(x = y\) and \(f\) has a unique fixed point in \(X\). □

## D Transforming the Black-Scholes equation to the Heat equation

In assuming that the underlying asset, in this case the stock, pays dividends it can be concluded by arbitrage arguments that the value of the asset price must decrease by the amount payed out. Suppose dividends are payed out with the continuously compounding yield $\delta \geq 0$. This flow can be modeled by a decrease of the asset prise $S$ in each infinitesimal time interval $dt$ by $\delta S$ $dt$. If it were not so, by purchasing the asset at time $t$ and selling it immediately after the dividend is payed out the risk-free profit $\delta S$ $dt$ could be made, a violation to the last assumption made in the beginning of Section 2.3. Thus the drift coefficient in equation (2.12) in Proposition 2.4 changes from being only $\alpha$ to being $\alpha - \delta$. Bringing this change under consideration and applying the theory developed in Section 2.3 it can be concluded that the corresponding Black-Scholes equation in Theorem 2.3 now becomes (where $V$ denotes the arbitrage free price and $\delta \geq 0$ the dividend)

\[
\begin{cases}
V_t + (r - \delta)sV_s + \frac{1}{2}\sigma^2 s^2 V_{ss} - rV = 0, \\
V(T, s) = \Phi(s).
\end{cases}
\]

Pricing a Multi-Asset American Option
When solving differential equations it is sometimes preferable to scale the equation in order to get dimensionless quantities and collect the dependent and independent variables as far as possible by variable transformation. In the Black-Scholes equation terms of the type $S^j \frac{\partial^j V}{\partial S^j}$, for $j = 0, 1, 2$ suggest introducing a new independent and scaled variable $x$ by the transformation $S = Ke^x$. There is also the need to transform the backward running time $\tau$ to a forward running time $\tau$ in order to make a terminal value problem into an initial value problem. Transforming to the independent variables $S = Ke^x$, $\tau = T - \frac{x}{\sigma^2}$ and the dependent variable $\nu(\tau, x) := V(t, S) = V(T - \frac{q}{\sigma^2}, Ke^x)$ and in using the chain rule gives

$$\frac{\partial V}{\partial t} = \frac{\partial \nu}{\partial \tau} \frac{\partial \tau}{\partial t} = -\frac{1}{2} \sigma^2 \frac{\partial \nu}{\partial \tau},$$

$$\frac{\partial V}{\partial S} = \frac{\partial \nu}{\partial x} \frac{\partial x}{\partial S} = \frac{\partial \nu}{\partial x},$$

$$\frac{\partial^2 V}{\partial S^2} = \frac{\partial}{\partial S} \left( \frac{1}{2} \frac{\partial \nu}{\partial \tau} \right) = \left( -\frac{1}{2} \frac{\partial \nu}{\partial x} + \frac{1}{2} \frac{\partial^2 \nu}{\partial x^2} \right) = \frac{1}{2} \left( \frac{\partial^2 \nu}{\partial x^2} - \frac{\partial \nu}{\partial x} \right).$$

Plugging the expressions above and $q_0 = \frac{2(\nu - \delta)}{\sigma^2}$, $q = \frac{2r}{\sigma^2}$ into the Black-Scholes equation transforms it to a equation with constant coefficients, i.e. to

$$\nu_\tau = \nu_{xx} + (q_0 - 1) \nu_x - (q_0 + q) \nu. \quad (D.1)$$

This equation needs further simplification and in considering the following constants $\gamma = \frac{1}{2}(q_0 - 1)$, $\beta = \frac{1}{2}(q_0 + 1)$ so that $\beta^2 = \gamma^2 + q_0$ the transformation becomes $\nu(\tau, x) = Ke^{-\gamma x - (\beta^2 + q_0)\gamma}(\tau, x)$. Now the partial derivatives become

$$\nu_\tau = Ke^{-\gamma x - (\beta^2 + q_0)\gamma}(-\beta y + y_\tau),$$

$$\nu_x = Ke^{-\gamma x - (\beta^2 + q_0)\gamma}(-\gamma y + y_x),$$

$$\nu_{xx} = Ke^{-\gamma x - (\beta^2 + q_0)\gamma}(\gamma^2 y - 2\gamma y_x + y_{xx}).$$

Using these derivatives in equation (D.1) gives

$$y_\tau = y_{xx} + (-2\gamma + q_0 - 1)y_x + \gamma(2\gamma - q_0 + 1)y.$$

Plugging in $\gamma = \frac{1}{2}(q_0 - 1)$ in the equation above gives the dimensionless form of the heat equation

$$\frac{\partial y}{\partial \tau} = \frac{\partial^2 y}{\partial x^2}.$$

### E  A simple 2D example

In two dimensions the region is partitioned into non-overlapping elements, triangles or quadrilaterals. Here the discussion is only concerned with a mesh consisting of triangles since when implementing the Galerkin method the spatial discretization is done by $n$-simplices ($n$ being the spatial dimension, e.g. a 2-simplex is a triangle and a 3-simplex is a tetrahedron). Also the elements are piecewise linear elements, i.e. the basis functions are piecewise linear functions, and for discussion on higher order elements on a triangular mesh see [19]. The trial functions...
\( \phi_i \) should be chosen so that each \( \phi_i \) is non-zero on only few elements and each \( \phi_i \) should be continuous over the whole domain \( \Omega \subset \mathbb{R}^2 \). Consider the following 2D Poisson problem on \( \Omega \subset \mathbb{R}^2 \), where \( \Omega \) is square partitioned into a mesh consisting of linear Lagrange triangles and \( \Delta \) is the Laplacian,

\[
\begin{cases}
-\Delta u = f & \text{in } \Omega, \\
u = 0 & \text{on } \partial \Omega
\end{cases}
\] (E.1)

Multiplying by a test function \( \varphi \), integrating, using Green’s formula (see Theorem 2.10), and using the boundary conditions problem E.1 becomes

\[
\iint_{\Omega} f \varphi \, dx \, dy = \iint_{\Omega} \left( \frac{\partial u}{\partial x} \frac{\partial \varphi}{\partial x} + \frac{\partial u}{\partial y} \frac{\partial \varphi}{\partial y} \right) \, dx \, dy - \iint_{\partial \Omega} \varphi \frac{\partial u}{\partial n} \, dx \, dy.
\]

Thus a suitable test space of admissible functions is

\[
\mathcal{V} = \mathcal{H}_0^1(\Omega) = \left\{ \varphi : \iint_{\Omega} \left( \varphi^2 + |\nabla \varphi|^2 \right) \, dx \, dy < \infty, \varphi = 0 \text{ on } \partial \Omega \right\}
\]

The variational form is then: find \( u \in \mathcal{H}_0^1(\Omega) \) such that \( a(u, \varphi) = L(\varphi) \forall \varphi \in \mathcal{H}_0^1(\Omega) \) where \( L(\varphi) = \iint_{\Omega} f \varphi \, dx \, dy \) and \( a(u, \varphi) = \iint_{\Omega} \nabla u \nabla \varphi \, dx \, dy \).

In the general case each element \( e \) has \( N \) nodes (here \( N = 3 \)) and each trial function \( \phi_i \) is identified with a unique node \( x_i \) on the mesh, i.e.

\[
\phi_i(x_j) = \begin{cases} 
1 & i = j \\
0 & i \neq j
\end{cases}
\]

The small region \( \Omega_i \) surrounding the node \( x_i \) is defined as the support of \( \phi_i \) and \( \phi_i(x) = 0 \) if \( x \notin \Omega_i \). It is possible to define \( \phi_i \) in terms of the local functions \( \phi_i^{(e)} \) on the element \( e \) as \( \phi_i = \sum_e \phi_i^{(e)} \). Thus the requirements on the local trial functions, when restricting \( \phi_i \) to a triangular element \( e \), are \( \phi_i^{(e)}(x, y) = 1 \) at node \( i \) and \( \phi_i^{(e)}(x, y) = 0 \) at the other \( N - 1 \) nodes, see Figure 17. Also \( \phi_i^{(e)}(x, y) = 0 \) outside the element \( e \). For the arbitrary triangle \( e \) with vertices \( (x_i, y_i), (x_j, y_j) \) and \( (x_k, y_k) \) in defining

\[
A = \begin{vmatrix} 
1 & x_i & y_i \\
1 & x_j & y_j \\
1 & x_k & y_k 
\end{vmatrix}
\]

gives the two dimensional piecewise linear basis functions on \( e \), where \( A \) is twice the area of \( e \),

\[
\phi_i^{(e)} = \frac{1}{A} \begin{vmatrix} 
1 & x & y \\
1 & x_j & y_j \\
1 & x_k & y_k 
\end{vmatrix}, \quad \phi_j^{(e)} = \frac{1}{A} \begin{vmatrix} 
1 & x_i & y_i \\
1 & x & y \\
1 & x_k & y_k 
\end{vmatrix}, \quad \phi_k^{(e)} = \frac{1}{A} \begin{vmatrix} 
1 & x_i & y_i \\
1 & x_j & y_j \\
1 & x & y 
\end{vmatrix}
\]
A simple 2D example

Figure 17: Hat functions on a triangular mesh.

A function $f$ can be approximated by the global piecewise linear interpolating function in terms of basis functions, i.e.

$$\pi_h f(x, y) = \sum_i f(x_i, y_i) \phi_i(x, y) \quad \text{if} \quad (x, y) \in \Omega$$

$$= f(x_i, y_i) \phi_i^{(x)} + f(x_j, y_j) \phi_j^{(x)} + f(x_k, y_k) \phi_k^{(x)} \quad \text{if} \quad (x, y) \in e$$

The interpolation error on the triangle $e$ given a continuous function $f$ on $e$, i.e. $f \in C(e)$, is $f - \pi_h f$. The values of the interpolant $\pi_h f$ and $f$ are the same at the nodes of the triangle, e.g. $\pi_h f(x_i, y_i) = f(x_i, y_i)$, $\pi_h f(x_j, y_j) = f(x_j, y_j)$ and $\pi_h f(x_k, y_k) = f(x_k, y_k)$. Thus the interpolant $\pi_h f$ is a plane having the same values as $f$ at the nodal points of the triangle $e$. Following [29] and defining the size of the the first and second order derivatives of $f$ as $Df := \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right)^{\frac{1}{2}}$.

**Pricing a Multi-Asset American Option**
and \( D^2 f := \left( \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 \right)^{\frac{1}{2}} \) the interpolation error estimate is

\[
\|f - \pi_h f\|_{L^2(\Omega)} \leq Ch^2 \|D^2 f\|_{L^2(\Omega)}
\]

\[
\|D(f - \pi_h f)\|_{L^2(\Omega)} \leq Ch^2 \|D^2 f\|_{L^2(\Omega)}
\]

where \( C \) is a constant, \( h_e \) the longest edge length of \( e \) and the norm \( L^2(\Omega) \) given a function \( w(x, y) \) is defined as \( \|w\|_{L^2(\Omega)} := \left( \iint_{\Omega} |w|^2 \, dx \, dy \right)^{\frac{1}{2}} \). Thus with the mesh \( \mathcal{E} = \{ e \} \) on a given domain \( \Omega \) the error estimate for continuous piecewise linear interpolation is

\[
\|f - \pi_h f\|_{L^2(\Omega)}^2 \leq C \sum_{e \in \mathcal{E}} h_e^2 \|D^2 f\|_{L^2(e)}^2
\]

\[
\|D(f - \pi_h f)\|_{L^2(\Omega)}^2 \leq C \sum_{e \in \mathcal{E}} h_e^2 \|D^2 f\|_{L^2(e)}^2
\]

Now in applying the Finite Element Method to solve problem E.1, where the Galerkin approximation is \( U(x, y) = \sum_{j=0}^{n} U_j \phi_j(x, y) \), the (weak) variational formulation is formulated as

\[
\sum_{j=0}^{n} U_j \iint_{\Omega} \left( \frac{\partial \phi_j}{\partial x} \frac{\partial \phi_i}{\partial x} + \frac{\partial \phi_j}{\partial y} \frac{\partial \phi_i}{\partial y} \right) \, dx \, dy = \sum_{j=0}^{n} f_j \iint_{\Omega} \phi_j \phi_i \, dx \, dy \quad i = 0, \ldots, n.
\]

The system of equations which can be written in matrix form as \( \mathbf{A} \mathbf{U} = \mathbf{M} \mathbf{f} \), where \( \mathbf{A} \) is the stiffness matrix and \( \mathbf{M} \) is the mass matrix. Since the region \( \Omega \) is partitioned into linear triangular elements, all integrals over the region \( \Omega \) are evaluated as the sums of integrals over linear triangular elements, i.e. as

\[
\iint_{\Omega} = \sum_{e} \iint_{e}
\]

and the formulation is rewritten as

\[
\sum_{e} \sum_{j=0}^{n} U_j \iint_{e} \left( \frac{\partial \phi_j^{(e)}}{\partial x} \frac{\partial \phi_i^{(e)}}{\partial x} + \frac{\partial \phi_j^{(e)}}{\partial y} \frac{\partial \phi_i^{(e)}}{\partial y} \right) \, dx \, dy = \sum_{e} \sum_{j=0}^{n} f_j \iint_{e} \phi_j^{(e)} \phi_i^{(e)} \, dx \, dy \quad i = 0, \ldots, n.
\]

**Remark E.1.** The local approximation for \( U \) on each triangular element with vertices \( i, j \) and \( k \) is denoted as \( U|_e = U_i \phi_i^{(e)} + U_j \phi_j^{(e)} + U_k \phi_k^{(e)} \) and the gradient components of \( U \) as \( \begin{align*}
\frac{\partial U|_e}{\partial x} &= U_i \frac{\partial \phi_i^{(e)}}{\partial x} + U_j \frac{\partial \phi_j^{(e)}}{\partial x} + U_k \frac{\partial \phi_k^{(e)}}{\partial x} \\
\frac{\partial U|_e}{\partial y} &= U_i \frac{\partial \phi_i^{(e)}}{\partial y} + U_j \frac{\partial \phi_j^{(e)}}{\partial y} + U_k \frac{\partial \phi_k^{(e)}}{\partial y}.
\end{align*} \)

The partial derivatives of \( \phi_i^{(e)}, \phi_j^{(e)} \) and \( \phi_k^{(e)} \) are constants and are evaluated as

\[
\begin{align*}
\frac{\partial \phi_i^{(e)}}{\partial x} &= \frac{y_j - y_k}{A}, & \frac{\partial \phi_j^{(e)}}{\partial x} &= \frac{y_k - y_i}{A}, & \frac{\partial \phi_k^{(e)}}{\partial x} &= \frac{y_i - y_j}{A}, \\
\frac{\partial \phi_i^{(e)}}{\partial y} &= \frac{x_k - x_j}{A}, & \frac{\partial \phi_j^{(e)}}{\partial y} &= \frac{x_j - x_k}{A}, & \frac{\partial \phi_k^{(e)}}{\partial y} &= \frac{x_k - x_i}{A}.
\end{align*}
\]
Thus the approximated partial derivatives of $U$ are

\[
\begin{align*}
\frac{\partial U}{\partial x}_e &= \frac{1}{A} \left( U_i(y_j - y_k) + U_j(y_k - y_i) + U_k(y_i - y_j) \right) \\
\frac{\partial U}{\partial y}_e &= \frac{1}{A} \left( U_i(x_k - x_j) + U_j(x_i - x_k) + U_k(x_j - x_i) \right)
\end{align*}
\]

When evaluating the integrals in the mass and stiffness matrices the reference element can be used, with the reference (or master) element being the unit triangle $E$ with vertices $(0,0)$, $(1,0)$ and $(0,1)$ at nodes 1, 2 and 3. Each triangular element $e$ with coordinates $(x, y)$ can be transformed onto the reference element $E$ with coordinates $(p, q)$, i.e. $e \mapsto E$, see [13] for mappings. The integrals are then evaluated together with the Jacobian of the transformation, e.g.

\[
\int_E \left( \frac{\partial \phi_i^e}{\partial x} \frac{\partial \phi_j^e}{\partial x} + \frac{\partial \phi_i^e}{\partial y} \frac{\partial \phi_j^e}{\partial y} \right) \, dp \, dq \]

where the Jacobian of the transformation is

\[
\frac{\partial (x,y)}{\partial (p,q)} = \begin{vmatrix}
\frac{\partial x}{\partial p} & \frac{\partial x}{\partial q} \\
\frac{\partial y}{\partial p} & \frac{\partial y}{\partial q}
\end{vmatrix}.
\]

Since the triangles are linear the transformation $e \mapsto E$ is not necessary. Using a reference element is advantageous when dealing with higher order elements. The element stiffness matrix $A(e)$ and the element mass matrix $M(e)$ are of size $(3 \times 3)$ ($N = 3$ for linear triangular elements). Evaluating the integrals in the element stiffness matrix, as the derivatives of the element basis functions are constants in recalling that the basis functions $\phi_i^e$, $\phi_j^e$ and $\phi_k^e$ are known linear functions, the components of $A(e)$ become

\[
a_{ij}^e = \left( \frac{\partial \phi_i^e}{\partial x} \frac{\partial \phi_j^e}{\partial x} + \frac{\partial \phi_i^e}{\partial y} \frac{\partial \phi_j^e}{\partial y} \right) \frac{A}{2} \quad i, j = 1, 2, 3.
\]

**Remark E.2.** In the expression above the dummy indices $i, j = 1, 2, 3$ now correspond to the three nodes in each triangular element $e$, whereas in the previous discussion the vertices were indexed by $i$, $j$ and $k$. But since the above expression now gives element indexing within a matrix it is more convenient to use 1, 2 and 3 for the indexing.

When evaluating the integrals in the element mass matrix the following formula, given in [14], for integrating products of hat functions can be used

\[
\int_e \phi_1^{m_1} \phi_2^{m_2} \cdots \phi_n^{m_n} \, de = \frac{m_1! m_2! \cdots m_n!}{(m_1 + m_2 + \cdots + m_n + n)!} n! |e|, \quad (E.2)
\]

where $n$ is the dimension and $|e|$ the volume of the $n$-simplex. Here

\[
\int_e \phi_1^{m_1} \phi_2^{m_2} \phi_3^{m_3} \, de = \frac{m_1! m_2! m_3!}{(m_1 + m_2 + m_3 + 2)!} 2! A.
\]

---

**Pricing a Multi-Asset American Option**
Thus the components in the element mass matrix are

\[ m^{(e)}_{ij} = \frac{A}{24}(1 + \delta_{ij}) \quad i, j = 1, 2, 3 \]

where \( \delta_{ij} \) is Kronecker’s delta, i.e. \( \delta_{ij} = 1 \) if \( i = j \) and 0 otherwise. When performing the assembly for the global stiffness and global mass matrices all integrals associated with an element are computed together, i.e.

\[ a_{ij} = \sum_e a^{(e)}_{ij}, \quad m_{ij} = \sum_e m^{(e)}_{ij} \]

An algorithm performing an assembly that is element oriented would have three nested loops, the \( e \)-loop, the \( i \)-loop and the \( j \)-loop. For further details on algorithms for the assembly and computation of the matrices and vectors within the finite element framework see [20].

Pricing a Multi-Asset American Option
References


[14] C. C. Cuvalier, A. Segal, and A. A. van Steenhoven. *Finite Element Methods and Navier-
REFERENCES


[47] UPPMAX Web page. http://www.uppmax.uu.se/, Apr. 2011. Uppsala Multidisciplinary Center for Advanced Scientific Computations (UPPMAX) is a member of the na-
tional metacenter Swedish National Infrastructure for Computing (SNIC) and is Uppsala University’s resource of high-performance computers, large-scale storage, and expertise in the field of high-performance computing (HPC).


