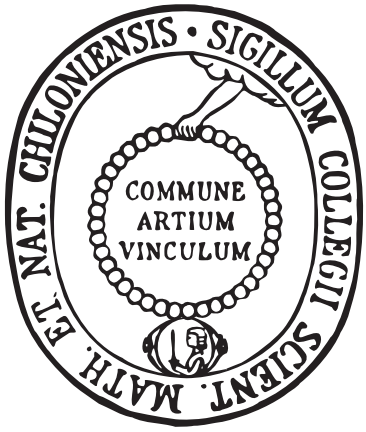
Computational Finance

Summer Term 2024



Mathias Vetter

This Version: April 23, 2024

Preface

These are the lecture notes for a Master-level course at the University of Kiel in the summer term 2024, and these notes will be updated and extended throughout the course.

Most material is taken from the lecture notes by Jan Kallsen (see http://www.math.uni-kiel.de/finmath/de/personen/kallsen/lec_notes). It, in principle, covers all the material needed for the course. This course is mainly based on [Sey06, Gla03, Rai00] and scattered literature. As an outlook we refer to the website of the *premia* project at INRIA, France, where many numerical algorithms for financial mathematics are implemented and well documented.

What are we talking about?

Mathematical finance tells us something about reasonable values of contingent claims, about hedging strategies and the like. But if the theory is to be applied in practice, we ultimately need numbers. Often these have to be obtained by numerical methods because explicit formulas are rare. In this course we discuss a number of approaches of quite different nature. Why do we not stick to the one method which proves superior in all circumstances? It simply does not exist. Some methods are faster, others easier to implement, even others more generally applicable.

In this course, we repeatedly consider the following basic tasks.

1. How can one compute the value of a European-style option? Mathematically, this boils down to calculating the expectation of the payoff under some appropriate probability measure.

2. How can one compute the value of an American option? This related question is more difficult because it involves solving an optimal stopping problem.

3. How can one compute the hedging strategy for such claims? In many models this means computing the so-called *Greeks*, i.e. derivatives of the option price relative to certain state variables.

4. How can one calibrate a parametric model to observed option prices? This means choosing the model parameters such that option prices in the model are as close as possible to quoted market prices.

This is only a selection of problems that must be solved in practice. Other questions involving numerical methods include parameter estimation in statistics or the computation of risks, probabilities, value at risk, etc. These are more of statistical nature and are not considered in this course. Neither do we discuss issues and models that require a deeper knowledge of financial mathematics as e.g. interest rate models, credit derivatives, commodity and energy markets, etc.

What is a good numerical method? Let us just mention a few aspects.

1. How precise is the method? For an approximation this basically means: is it precise enough for practical purposes? Many methods can be made arbitrarily precise by increasing computing power. This immediately leads to the next aspect.

2. How fast or expensive is the algorithm or, more precisely, how many arithmetic operations and how much memory does is take to reach a certain precision? The relative performance of two methods may depend on the precision that is actually required. Often, the absolute error ε of a method depends on the number N of operations approximately as

$$\varepsilon(N) \approx CN^{-p}$$

   with some constants $C > 0$ and $p > 0$ which depend on the method. For large N a method with higher rate p will ultimately be faster. However, for limited computing power and hence N, it may be more efficient to use a method with lower rate p if the multiplicative constant C is much smaller.

3. One should also keep in mind how much effort it takes to implement the method. If the result is needed only once, it may be more reasonable to spend one day to write a computer programme which needs another day to produce numbers rather than working one month on a very efficient implementation that takes only a second to compute the result.

4. Finally, it is of interest how broadly a method is applicable. The stochastic models used in Mathematical Finance are subject to change, depending on their ability to fit real data. Moreover, more and more complex derivatives are developed. Therefore, one might prefer methods that can easily be generalised to a large variety of models and payoffs.

In this course we focus on methods rather than models. We consider the above basic tasks repeatedly in the simple Black-Scholes model and in the slightly more complex Heston model, where closed-form solutions are no longer available. We do not consider other important model

classes as e.g. models with jumps, more involved continuous models, or discrete-time models beyond the binomial tree. We also stick to rather simple payoffs.

Concrete numerical examples play an important role in this course. We choose the programming language *Python* for this purpose because it is freely available, convenient, sufficiently fast for our purposes, and is used a lot in many fields. Therefore, it serves as a useful environment for research and development.

Contents

Recap of basic notions

This course assumes knowledge of financial mathematics as in the course *Mathematical Finance*, cf. e.g. [Vet21]. Here we briefly recall some basic notions that are used in the sequel. We mostly work in a market with two primary assets, a riskless *bond* written as $B = (B(t))_{t\geq 0}$ and a *risky asset* (stock or currency) $S = (S(t))_{t\geq 0}$. Typically, we assume fixed interest rate $r \geq 0$ in the sense that

$$B(t) = \exp(rt) \tag{0.1}$$

and a stock of geometric Brownian motion type

$$S(t) = S(0)\exp(\mu t + \sigma W(t)) \tag{0.2}$$

with parameters $\mu \in \mathbb{R}, \sigma > 0$ and standard Brownian motion W.

A *derivative* or *option* is an asset whose payoff depends on the underlying. The payoff X at time T may be a function $f(S(T))$ of the underlying at time T as e.g. $X = (S(T) - K)^+$ for a simple European call or it could be a more complex function of the whole past of S. In complete markets such options can be replicated perfectly. This means that there exists a self-financing portfolio $\varphi = (\varphi_0, \varphi_1)$ whose value at time T equals $V_\varphi(T) = X$. Absence of arbitrage implies that $V_\varphi(t)$ is the only reasonable option price of the option at time t. It can be computed as the conditional expectation

$$V_\varphi(t) = B(t)E_Q\left(\frac{X}{B(T)}\middle|\mathscr{F}_t\right) \tag{0.3}$$

of the discounted payoff under the unique equivalent martingale measure Q, i.e. the unique probability measure $Q \sim P$ such that S/B is a Q-martingale. The most prominent example of a complete market is the Black-Scholes model. There it holds that the asset price process S is also a geometric Brownian motion under Q of the form

$$S(t) = S(0)\exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W^Q(t)\right),$$

where W^Q is a Brownian motion under Q; see Mathematical Finance and Chapter 3 for additional details.

If the market is incomplete, the equivalent martingale measure Q is no longer unique. Absence of arbitrage then implies that option prices are still of the form (0.3) for at least one such Q. However, a replicating portfolio φ typically ceases to exist.

American options are specified by an entire exercise process $X = (X(t))_{t\in[0,T]}$ as e.g. $X(t) = (K - S(t))^+$ for an American put. In the complete case, the only reasonable price is the B-fold of the Snell envelope of X/B relative to Q. The Q-Snell envelope is the smallest Q-supermartingale dominating X/B. Again, Q denotes the unique equivalent martingale measure from above. One can write this fair price also as

$$V(t) = B(t) \sup_{\tau} E_Q \left(\frac{X(\tau)}{B(\tau)} \middle| \mathscr{F}_t \right), \tag{0.4}$$

where the supremum extends over all stopping times stopping between t and T. If $\tau^\star$ denotes the optimal stopping time, the holder should exercise the option at $\tau^\star$ in order to avoid unnecessary losses. One such stopping time is the first time t such that $V(t) = X(t)$, i.e. the market price of the option equals the exercise price. The stopped process $V^{\tau^\star}$ is a Q-martingale and it is the value process of a self-financing hedging strategy φ for the American option. In the incomplete case, such a hedge does not exist. But we can still say that absence of arbitrage implies that the option's value is of the form (0.4) for at least one equivalent martingale measure Q.

Martingale modelling means looking for the dynamics of the stock under such a martingale measure, which turns all discounted prices into martingales or, for American options, Snell envelopes. One typically starts a priori with a parametric family of conceivable models for the Q-dynamics of the stock. The unknown parameters are chosen such that the moments (0.3) resp. (0.4) match the observed option prices as closely as possible. This procedure is called *calibration*.

Chapter 1

Binomial trees

In this chapter, we summarize the main findings for the CRR binomial tree model and its connection to the Black-Scholes model. The goal is to have all information at hand for the implementation of tree models as approximations for continuous time models.

1.1 The Cox-Ross-Rubinstein model

We encountered the binomial or *Cox-Ross-Rubinstein* (CRR) model in the previous course as a simple discrete-time model for the stock. Here, we consider it as a numerical method for computing option prices and hedges in the Black-Scholes model that we briefly discussed at the end of that course. The point is that the stock price evolution as well as option prices in a properly chosen sequence of CRR models converges to the corresponding Black-Scholes notions if the number of time step tends to infinity. Put differently, option prices and hedges from a CRR model with a large number of time steps can be chosen as an approximation for Black-Scholes prices and hedges.

We consider equidistant times $0 = t_0, t_1, \ldots, t_M = T$ with $t_i = i\Delta t$ and hence $\Delta t = T/M$. The bank account or bond moves according to

$$B(t_i) = \exp(rt_i).$$

The stock goes up by a factor u resp. down by a factor d in each period, i.e.

$$S(t_i) = \begin{cases} S(t_{i-1})u & \text{with probability } p \\ S(t_{i-1})d & \text{with probability } 1-p, \end{cases}$$

where $d < \exp(r\Delta t) < u$. More precisely, p and $1-p$ denote the conditional probabilities of going up resp. down given the past.

For option pricing real-world transition probabilities do not matter. Instead we need to consider martingale probabilities, i.e. probabilities such that

$$E_Q(S(t_i)/B(t_i)|\mathscr{F}_{t_{i-1}}) = S(t_{i-1})/B(t_{i-1}).$$

If we denote the Q-transition probabilities by q resp. $1-q$, the left-hand side equals

$$q\frac{S(t_{i-1})u}{B(t_{i-1})\exp(r\Delta t)} + (1-q)\frac{S(t_{i-1})d}{B(t_{i-1})\exp(r\Delta t)},$$

which equals the right-hand side if and only if

$$\frac{qu+(1-q)d}{\exp(r\Delta t)} = 1$$

or

$$q = \frac{\exp(r\Delta t)-d}{u-d}. \tag{1.1}$$

This transition probability is the same for all branchings. Uniqueness of this probability implies that the model is complete.

We mean to use the CRR model as an approximation to the Black-Scholes model which we briefly sketched in the introductory chapter. The basic processes B, S in (0.1,0.2) involve parameters $r, S(0), \mu, \sigma$. Moreover, we consider a finite time horizon T. The parameter μ changes under the transition to the unique martingale measure as we will see in Chapter 3. In fact, the stock price (0.2) can be rewritten as

$$S(t) = S(0)\exp\left(\left(r-\frac{\sigma^2}{2}\right)t + \sigma W^Q(t)\right),$$

where W^Q denotes a standard Brownian motion relative to Q. In the binomial model we must fix parameters $r, S(0), T, M, u, d, q$. In order for the bond, the initial value of the stock, and the time horizon to coincide we naturally choose $r, S(0), T$ as in the Black-Scholes model that is to be approximated. The number of periods M depends on the precision that is needed. Large M means better approximation but higher costs of computation. In order for the models to behave similarly, we choose the remaining parameters u, d, q such that the first two conditional moments in the Black-Scholes model and its CRR approximation coincide. For the conditional expectation we have

$$E_Q(S(t_i)|\mathscr{F}_{t_{i-1}}) = \frac{S(t_{i-1})B(t_i)}{B(t_{i-1})}$$

in both the Black-Scholes and the CRR model because S/B is a martingale under any equivalent martingale measure Q. For the second moment in the Black-Scholes model we obtain

$$\begin{aligned}
&E_Q(S(t_i)^2|\mathscr{F}_{t_{i-1}}) \\
&\quad = S(t_{i-1})^2\exp((2r-\sigma^2)\Delta t)E_Q\left(\exp\left(2\sigma(W^Q(t_i)-W^Q(t_{i-1}))\right)\middle|\mathscr{F}_{t_{i-1}}\right) \\
&\quad = S(t_{i-1})^2\exp((2r-\sigma^2)\Delta t)\exp(\tfrac{1}{2}(2\sigma)^2\Delta t) \\
&\quad = S(t_{i-1})^2\exp((2r+\sigma^2)\Delta t),
\end{aligned}$$

where the second equality follows from the following two facts: $W^Q(t) - W^Q(s)$ is normally distributed with mean 0 and variance $t-s$, moreover $E(\exp(X)) = \exp(\sigma^2/2)$ for any Gaussian random variable with mean 0 and variance σ^2. In the CRR model we have

$$\begin{aligned} E_Q(S(t_i)^2|\mathscr{F}_{t_{i-1}}) &= qS(t_{i-1})^2u^2 + (1-q)S(t_{i-1})^2d^2 \\ &= S(t_{i-1})^2(qu^2 + (1-q)d^2) \end{aligned}$$

If we want the conditional second moments to coincide in both models, we must require

$$\exp((2r+\sigma^2)\Delta t) = qu^2 + (1-q)d^2. \tag{1.2}$$

We obtain two conditions (1.1, 1.2) for three variables u, d, q. As a third condition we impose

$$ud = 1, \tag{1.3}$$

which implies that the median of the possible final stock prices $S(0)d^M, S(0)d^{M-1}u, \ldots, S(0)d^{M/2}u^{M/2}, \ldots S(0)du^{M-1}, S(0)u^M$ equals the initial stock price $S(0)$. Equations (1.1–1.3) for u, d, q lead to

$$\begin{aligned} \alpha^2\exp(\sigma^2\Delta t) &= \frac{\alpha - u^{-1}}{u-u^{-1}}u^2 + \frac{u-\alpha}{u-u^{-1}}u^{-2} \\ &= \frac{\alpha u^2 - u + u^{-1} - \alpha u^{-2}}{u-u^{-1}} \\ &= \alpha u - 1 + \alpha u^{-1} \end{aligned}$$

with $\alpha = \exp(r\Delta t)$. Put differently,

$$u^2 - (\alpha^{-1} + \alpha\exp(\sigma^2\Delta t))u + 1 = 0$$

or

$$u = \beta \pm \sqrt{\beta^2 - 1}$$

for

$$\beta = \frac{\alpha^{-1} + \alpha\exp(\sigma^2\Delta t)}{2} = \frac{\exp(-r\Delta t) + \exp((r+\sigma^2)\Delta t)}{2}$$

and hence

$$d = \frac{1}{u} = \frac{1}{\beta \pm \sqrt{\beta^2-1}} = \frac{\beta \mp \sqrt{\beta^2-1}}{\beta^2 - \beta^2 + 1} = \beta \mp \sqrt{\beta^2 - 1}.$$

Since we require $d < u$, we end up with

$$u = \beta + \sqrt{\beta^2 - 1} \tag{1.4}$$

$$d = u^{-1} = \beta - \sqrt{\beta^2 - 1} \tag{1.5}$$

$$\beta = \frac{1}{2}(\exp(-r\Delta t) + \exp((r+\sigma^2)\Delta t)) \tag{1.6}$$

$$q = \frac{\exp(r\Delta t) - d}{u-d}, \tag{1.7}$$

which indeed solves (1.1–1.3).

For $M \to \infty$ this sequence of CRR models converges in law in some proper sense to the corresponding Black-Scholes model. This statement can be made precise and shown based on the central limit theorem.

1.2 European options

Let us consider a European option with payoff $g(S(T))$ for some function g as e.g. $g(x) = (x-K)^+$ (call) or $g(x) = (K-x)^+$ (put). We denote the fair price of the option at time t_i by $V(S(t_i), t_i)$. The Q-martingale property for discounted option prices reads as

$$E_Q(V(S(t_i), t_i)/B(t_i)|\mathscr{F}_{t_{i-1}}) = V(S(t_{i-1}), t_{i-1})/B(t_{i-1}) \tag{1.8}$$

or

$$q\frac{V(S(t_{i-1})u, t_i)}{B(t_{i-1})\exp(r\Delta t)} + (1-q)\frac{V(S(t_{i-1})d, t_i)}{B(t_{i-1})\exp(r\Delta t)} = \frac{V(S(t_{i-1}), t_{i-1})}{B(t_{i-1})},$$

which is

$$V(S(t_{i-1}), t_{i-1}) = \exp(-r\Delta t)\left(qV(S(t_{i-1})u, t_i) + (1-q)V(S(t_{i-1})d, t_i)\right). \tag{1.9}$$

This allows to compute option prices starting from

$$V(S(T), T) = g(S(T))$$

and moving backwards in time.

Since the model is complete, the payoff $g(S(T))$ can be replicated perfectly by a self-financing portfolio $\varphi = (\varphi_0, \varphi_1)$. Its value equals

$$\varphi_0(t_i)B(t_i) + \varphi_1(t_i)S(t_i) = V(S(t_i), t_i). \tag{1.10}$$

The same equation for t_{i-1} and self-financeability

$$\varphi_0(t_{i-1})B(t_{i-1}) + \varphi_1(t_{i-1})S(t_{i-1}) = \varphi_0(t_i)B(t_{i-1}) + \varphi_1(t_i)S(t_{i-1})$$

lead to

$$V(S(t_{i-1}), t_{i-1}) = \varphi_0(t_i)B(t_{i-1}) + \varphi_1(t_i)S(t_{i-1}). \tag{1.11}$$

(1.10) yields

$$\begin{aligned} \varphi_0(t_i)B(t_i) + \varphi_1(t_i)S(t_{i-1})u &= V(S(t_{i-1})u, t_i) \\ \varphi_0(t_i)B(t_i) + \varphi_1(t_i)S(t_{i-1})d &= V(S(t_{i-1})d, t_i) \end{aligned}$$

and hence

$$\varphi_1(t_i) = \frac{V(S(t_{i-1})u, t_i) - V(S(t_{i-1})d, t_i)}{S(t_{i-1})(u-d)} \tag{1.12}$$

by taking differences. From (1.11) we get

$$\varphi_0(t_i) = \frac{V(S(t_{i-1}), t_{i-1}) - \varphi_1(t_i)S(t_{i-1})}{B(t_{i-1})}. \tag{1.13}$$

Observe that the option price and the hedging strategy depend only on the current value of the stock and the time. We use these objects as approximations for the corresponding quantities in the Black-Scholes model, which also depend only on $S(t)$ and t. As noted above, one can show that convergence holds for $M \to \infty$.

1.3 Non-recombining trees and exotic options

In principle one can repeat the above arguments for options that depend on the whole stock price path $S(t_0), S(t_1), \ldots, S(T)$. In this case the option price at time t_i depends on $S(t_0), S(t_1), \ldots, S(t_i)$ as well. However, this makes the algorithm much more expensive because the structure of the recombining tree with only $(M+1)(M+2)/2$ instead of 2^M nodes cannot be exploited anymore.

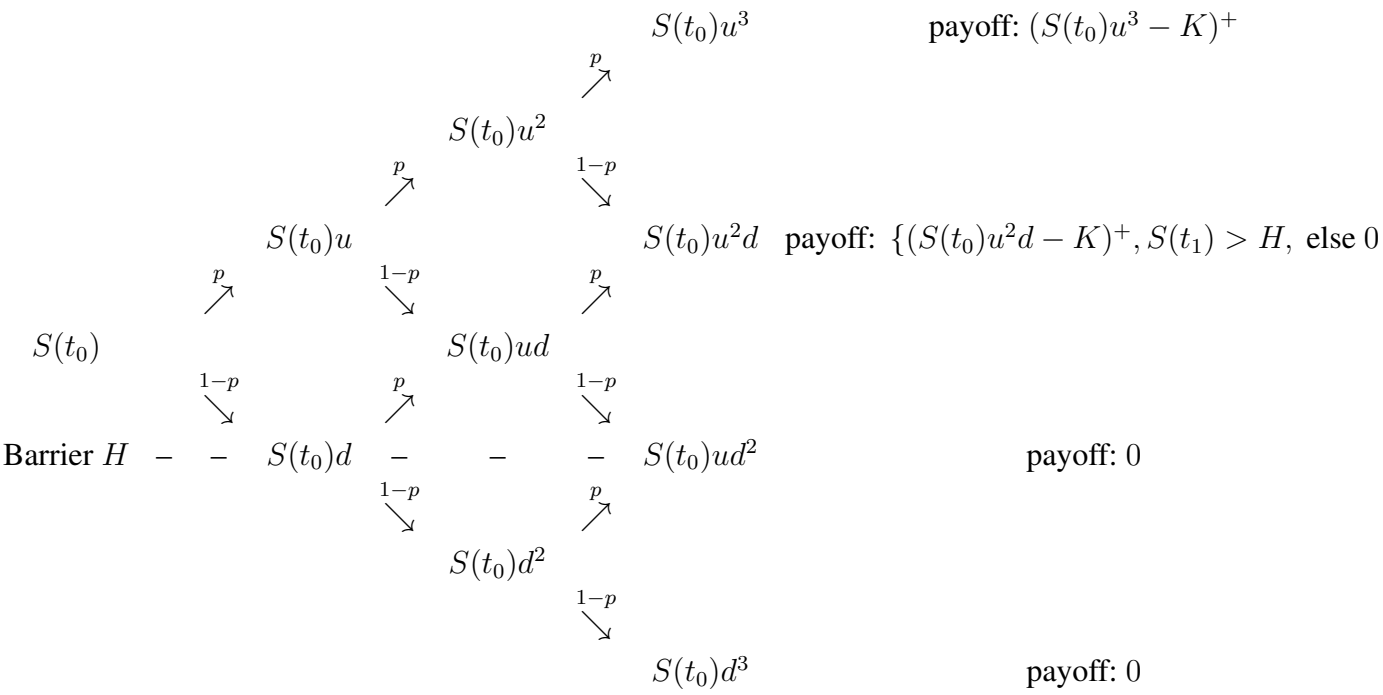
Nevertheless, knock-out barrier options can be treated efficiently. Their payoff vanishes if the stock price exceeds a certain barrier at least once before expiration. Although their fair value at time t_i depends on the whole past, their value *given that the option has not been knocked out yet* does not. It can be calculated using a recursion along the same lines as above.

More precisely, consider the *Knock-out barrier option*. This is an option which loses its value when a certain barrier is reached, for example the *down-and-out call*

$$g(S(t_0), S(t_1), \ldots, S(T)) = (S(T) - K)^+ 1_{\left\{\min_{j=0,1\ldots,M} S(t_j) > H\right\}}$$

$$1_{\left\{\min_{j=1,\ldots,M} S(t_j) > H\right\}} = \begin{cases} 1 & \text{all prices are above barrier } H \\ 0 & \text{barrier is hit} \end{cases}$$

To represent it, it first seems that we need a *non-recombining tree*



How can we compute the price V_0 of this Barrier option without using the non-recombining tree? We define $\tilde{V}_i$ as the price of a Barrier option O_i which is *written at time i* with the same T, K and H, starting at $S(t_i)$. Then we set

(i) $V_i = \tilde{V}_i$ if $S(t_0) > H, \ldots, S(t_i) > H$, in particular $V_0 = \tilde{V}_0$ if $S(t_0) > H$.

(ii) $V_i = \tilde{V}_i = 0$ if $S(t_i) \leq H$.

(iii) $\tilde{V}_i = \frac{1}{\exp(r\Delta t)} \left[q\tilde{V}_{i+1}(S(t_i)u) + (1-q)\tilde{V}_{i+1}(S(t_i)d) \right]$, if $S(t_i) > H$, where now $\tilde{V}_{i+1}$ is the price of O_{i+1} at time t_{i+1}.

Note that the recursion in (iii) does not lead to the same option prices as for the regular call because certain $\tilde{V}_{i+1}$ are now zero by construction.

Approximating continuous time models by binomial tree models introduces a further error as the actual barrier is often different from the barrier used for the tree model, leading to slow convergence.

Numerical accuracy often becomes better when using a trinomial model. Such a model is no longer complete so the risk neutral Q has to be chosen by calibration.

The curse of dimension

We look at a model with more than one security ($k > 1$), so we have price processes $S^1, \dots, S^k$. So at each time i we have a vector of stock prices.

$$\begin{bmatrix} S^1 \\ \vdots \\ S^k \end{bmatrix} \text{ which moves to } \begin{bmatrix} S^1 \cdot \alpha^{(1)} \\ \vdots \\ S^k \cdot \alpha^{(k)} \end{bmatrix}$$

with $\alpha^{(1)}, \dots, \alpha^{(k)} \in \{u, d\}$. Look at the tree for $k = 2$:
At $t = 1$ we have the possible states

$$\begin{bmatrix} u \\ u \end{bmatrix}, \begin{bmatrix} d \\ u \end{bmatrix}, \begin{bmatrix} u \\ d \end{bmatrix}, \begin{bmatrix} d \\ d \end{bmatrix} : \#4 = 2^2$$

at $t = 2$ we have for the recombining tree:

$$\begin{bmatrix} u^2 \\ u^2 \end{bmatrix}, \begin{bmatrix} u^2 \\ ud \end{bmatrix}, \begin{bmatrix} u^2 \\ d^2 \end{bmatrix}, \begin{bmatrix} ud \\ u^2 \end{bmatrix}, \begin{bmatrix} ud \\ ud \end{bmatrix}, \begin{bmatrix} ud \\ d^2 \end{bmatrix}, \begin{bmatrix} d^2 \\ u^2 \end{bmatrix}, \begin{bmatrix} d^2 \\ ud \end{bmatrix}, \begin{bmatrix} d^2 \\ d^2 \end{bmatrix} : \#9 = 3^2$$

So at time t_M we have for each stock $M + 1$ possible states, so for $S^1_{t_M}, \dots, S^k_{t_M}$ we have $(M + 1)^k$ possible states $[M = 99,\ k = 6,\ \#10^{12}]$
This *curse of dimension* in higher dimensions has led to a variety of new methods in recursive option pricing, in particular for American options.

1.4 American options

At first glance, it seems that for American options also non-recombining trees are needed. However, from the introductory course, we know that the above recursive algorithm for European options needs only little modification for an American option with exercise price $g(S(t_i))$ at time t_i. General theory from the previous course tells us that its fair value at time t_{i-1} satisfies

$$V(S(t_{i-1}), t_{i-1}) = \max\left\{g(S(t_{i-1})), B(t_{i-1})E_Q(V(S(t_i), t_i)/B(t_i)|\mathscr{F}_{t_{i-1}})\right\}$$

rather than (1.8). Repeating the above calculation this leads to

$$V(S(t_{i-1}), t_{i-1}) = \max\left\{g(S(t_{i-1})), \exp(-r\Delta t)\left(qV(S(t_{i-1})u, t_i) + (1-q)V(S(t_{i-1})d, t_i)\right)\right\}$$

instead of (1.9). General theory tell us that $\tau_f = \inf\{t_i : V(S(t_i), t_i) = g(S(t_i))\}$ and

$$\tau_s = \inf\left\{t_i : i = M \text{ or } g(S(t_i)) > \exp(-r\Delta t)\left(qV(S(t_i)u, t_{i+1}) + (1-q)V(S(t_i)d, t_{i+1})\right)\right\}$$

are the first and the last reasonable times to exercise the option. Before that, they can be hedged perfectly using the strategy φ satisfying (1.12, 1.13).

Recall that the prices of European and American call coincide for nonnegative interest rate r because the second term in the maximum always dominates the first.

1.5 Algorithm and discussion

In order to write down an algorithm in pseudo code, we need some notation. $S_{ji} = S(0)u^j d^{i-j}$ denotes the stock price at time t_i after j upward and hence $i-j$ downward movements. Likewise, V_{ji} denotes the option price at time t_i after j upward movements of the stock, which means

$$V_{ji} = \exp(-r\Delta t)\left(qV_{j+1,i+1} + (1-q)V_{j,i+1}\right). \tag{1.14}$$

in the European case and

$$V_{ji} = \begin{cases} \max\left\{(S_{ji}-K)^+, \exp(-r\Delta t)\left(qV_{j+1,i+1} + (1-q)V_{j,i+1}\right)\right\} & \text{for a call,} \\ \max\left\{(K-S_{ji})^+, \exp(-r\Delta t)\left(qV_{j+1,i+1} + (1-q)V_{j,i+1}\right)\right\} & \text{for a put} \end{cases} \tag{1.15}$$

for American options, both with initial condition

$$V_{jM} = \begin{cases} (S_{jM}-K)^+ & \text{for a call,} \\ (K-S_{jM})^+ & \text{for a put.} \end{cases} \tag{1.16}$$

We can now formulate the algorithm for computing European and American calls and puts in the Black-Scholes model based on the binomial approximation:

Input $r, \sigma, S(0), T, K$, call/put, European/American, M

Compute $\Delta t := T/M$

Set u, d, q according to (1.4–1.7).

Set $S_{00} := S(0)$.

$S_{ji} = S_{00}u^j d^{i-j}$ for $i = 1, \ldots, M$ and $j = 0, \ldots, i$.

V_{jM} according to (1.16).

V_{ji} according to (1.14) resp. (1.15) for $i = M-1, \ldots, 0$ and $j = 0, \ldots, i$.

Output V_{00}

The output V_{00} is the desired approximation for the initial option price in the Black-Scholes model.

We end this section with a short discussion. The binomial approximation provides a simple algorithm for both European and American options, which is easy to understand and converges to the right value for $M \to \infty$. Moreover, the output can be interpreted as the correct answer in a different financial market model. On the other hand, other methods based e.g. on numerical solution of the Black-Scholes partial differential equation (PDE) turn out to be more efficient. Moreover, the binomial approximation delivers the option price only for a specific value $S(0)$, not simultaneously for all inital stock prices. Finally, a generalisation to more general models is not immediately obvious. But other tree-based approximations have been suggested in the literature to improve and generalise the above binomial approach.

Chapter 2

Monte Carlo simulation

2.1 Numerical integration using Monte Carlo

We will later discuss numerical methods relying on solving PDE's or using integral transforms as a tool for option pricing. These methods may face severe difficulties when applied to very complex models or for complicated, in particular path-dependent payoff structures. In these cases Monte Carlo simulation may help although it does not seem very efficient on first glance. Recall that prices of European options with payoff $f(S(T))$ are of the form

$$V(0) = \exp(-rT)E_Q(f(S(T)))$$

if the numeraire price process equals $B(t) = \exp(rt)$.

In order to compute such expectations, we start by considering the general question how to approximate expectations $V = E(f(X))$ numerically, where X denotes a real- or vector-valued random variable and f some real-valued function. We assume that we know how to simulate X on the computer. More precisely, the computer is supposed to produce a realization of N independent, identically distributed random variables $X_1, \ldots, X_N$, which have the same distribution as X. Then we use the *average* or *empirical mean*

$$\hat{V}_N = \frac{1}{N}\sum_{n=1}^{N} f(X_n) \tag{2.1}$$

as an approximation for V. Why does this make sense?

Unbiasedness Note that $\hat{V}_N$ is random because it is computed from random numbers. But at least its expectation equals

$$\begin{aligned} E(\hat{V}_N) &= \frac{1}{N}\sum_{n=1}^{N} E(f(X_n)) \\ &= \frac{1}{N}\sum_{n=1}^{N} E(f(X)) \\ &= E(f(X)) = V, \end{aligned}$$

i.e. $\hat{V}_N$ is an *unbiased estimator* for the desired expectation $V = E(f(X))$. This is nice but does not necessarily mean that $\hat{V}_N$ is close to V because the variance may be very large.

Consistency But according to the *law of large numbers* we have that $\hat{V}_N$ converges for $N \to \infty$ to $V = E(f(X))$. In that sense $\hat{V}_N$ is a *consistent estimator* for V. Hence we can get an arbitrarily precise approximation by simulating sufficiently often.

Rate of convergence On the other hand, due to limited time we have to stop simulating at some point. So what is the precision of the approximation for fixed N? As we have seen, the error is random and zero on average. Its variance equals

$$\begin{aligned} \mathrm{Var}(\hat{V}_N) &= \frac{1}{N^2}\sum_{n=1}^{N} \mathrm{Var}(f(X_n)) \\ &= \frac{1}{N^2}\sum_{n=1}^{N} \mathrm{Var}(f(X)) \\ &= \frac{1}{N}\mathrm{Var}(f(X)), \end{aligned}$$

which implies that the standard deviation is

$$\sigma(\hat{V}_N) = \sqrt{\mathrm{Var}(\hat{V}_N)} = \frac{1}{\sqrt{N}}\sigma(f(X)). \tag{2.2}$$

In this sense the rate of convergence is $1/\sqrt{N}$. In order to halve the error, we need on average four times as many simuations. This rate of convergence seems rather slow. On the other hand, in particular for high-dimensional random vectors it is hard to beat this seemingly low rate of convergence.

Asymptotic normality and confidence intervals For large N we can in fact tell more than just the variance of the error. By the *central limit theorem* the standardised error

$$\frac{\hat{V}_N - V}{\sigma(\hat{V}_N)} = \frac{\sqrt{N}(\hat{V}_N - V)}{\sigma(f(X))}$$

converges in law to some standard Gaussian random variable Z. This implies that

$$P(|\hat{V}_N - V| > \varepsilon) \approx P\left(|Z| > \frac{\varepsilon\sqrt{N}}{\sigma(f(X))}\right) = 2\Phi\left(-\frac{\sqrt{N}\varepsilon}{\sigma(f(X))}\right) \tag{2.3}$$

for arbitrary $\varepsilon > 0$ and large N, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-z^2/2)dz$ as usual denotes the cumulative distribution function of the standard normal law. Suppose that we want the probability of an error greater than ε to be smaller than e.g. 0.01. In principle, we can now compute how large N must be chosen to reach this precision. The inequality

$$2\Phi\left(-\frac{\sqrt{N}\varepsilon}{\sigma(f(X))}\right) \leq 0.01$$

can be rewritten as

$$N > \left(-\Phi^{-1}(0.005)\right)^2 \frac{\mathrm{Var}(f(X))}{\varepsilon^2} \approx 6.635 \frac{\mathrm{Var}(f(X))}{\varepsilon^2}. \tag{2.4}$$

However, it is unlikely that we can compute the right hand side of (2.4) explicitly. Indeed, if even the expectation of $f(X)$ is unknown, we probably do not know its variance either.

The way out is to approximate it as well by an average of random variables. From statistics we know that the *sample variance*

$$\begin{aligned} \hat{\sigma}_N^2(f(X)) &= \frac{1}{N-1} \sum_{n=1}^{N} \left(f(X_n) - \hat{V}_N\right)^2 \\ &= \frac{N}{N-1} \left(\frac{1}{N} \sum_{n=1}^{N} \left(f(X_n)\right)^2 - \hat{V}_N^2\right) \end{aligned} \tag{2.5}$$

is an unbiased and consistent estimator for the variance

$$\sigma^2(f(X)) = \mathrm{Var}(f(X)).$$

Using Slutsky's theorem it is not hard to show that

$$\frac{\sqrt{N}(\hat{V}_N - V)}{\sqrt{\hat{\sigma}_N^2(f(X))}} \tag{2.6}$$

converges to a standard Gaussian random variable as well. This means that we can just replace the unknown variance in (2.3) resp. (2.4) by its approximation $\hat{\sigma}_N^2(f(X))$ and still get a reasonable criterion on how large N should be chosen.

From a related perspective, the asymptotic normality of (2.6) can be used to construct asymptotic confidence bounds. Since 95% of all observations lie within 1.96 standard deviations of the mean, we may say that–provided that N is not too small–

$$V \in \left[\hat{V}_N - 1.96\sqrt{\frac{\hat{\sigma}_N^2(f(X))}{N}}, \hat{V}_N + 1.96\sqrt{\frac{\hat{\sigma}_N^2(f(X))}{N}}\right]$$

holds approximately with probability 0.95.

Mean squared error in the biased case Sometimes it is not obvious how to simulate X but we know how to simulate Y with similar law or, more precisely, with

$$E(f(Y)) \approx E(f(X)).$$

In this case we take

$$\widetilde{V}_n = \frac{1}{N}\sum_{n=1}^{N} f(Y_n)$$

as an approximation for V, where $Y_1, \ldots, Y_N$ denote independent random variables with the same law as Y. Since $\widetilde{V}_n$ may no longer be an unbiased estimator for V, the *mean squared error* now equals

$$\begin{aligned} MSE(\widetilde{V}_n) &= E((\widetilde{V}_n - V)^2) \\ &= (E(\widetilde{V}_n - V))^2 + \mathrm{Var}(\widetilde{V}_n - V) \\ &= (E(\widetilde{V}_n) - V)^2 + \mathrm{Var}(\widetilde{V}_n) \\ &= \Big(E(f(Y)) - V\Big)^2 + \frac{1}{N}\mathrm{Var}(f(Y)). \end{aligned}$$

The first term represents the squared *bias*, the second the *variance* of the approximation. Often there is a tradeoff between bias and variance. E.g. one could simulate random variables Y with smaller absolute bias $|E(f(Y)) - V|$, which however, require more simulation time so that N must be reduced.

Numerical integration As a side remark we note that Monte Carlo simulations can be used to compute integrals $\int_a^b f(x)dx$ because $\int_a^b f(x)dx = (b-a)E(f(X))$ for a random variable X with uniform distribution on the interval $[a, b]$. This is particularly useful in higher dimensions. We have

$$\int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_d)dx_1 \ldots dx_d = E(f(X)),$$

where $X = (X_1, \ldots X_d)$ denotes a random vector with uniform law on $[0, 1]^d$, i.e. $X_1, \ldots X_d$ are independent an identically distributed with uniform law on $[0, 1]$.

2.2 Random number generation

We turn now to the question how to simulate random variables with a given law. *Random number generators* on computers usually generate independent random variables with uniform law on $[0, 1]$, i.e. with cumulative distribution function (cdf)

$$F(x) = P(X \le x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 \le x < 1, \\ 1 & \text{if } x \ge 1. \end{cases}$$

Linear congruential generator Clearly, algorithms on computers are deterministic by nature so they cannot produce pure randomness. So what they rather do is produce a sequence $U_1, U_2, \ldots$ which for practical purposes behaves like a sequence of independent uniformly distributed random variables, so which would e.g. be accepted by statistical tests.

We study a basic procedure (with modern modifications), the *linear congruential generator*: Generators of pseudo-random numbers are usually based on mathematical concepts of number theory. A typical algorithm is the following one: Set

$$s_n = (As_{n-1} + C) \mod M,$$

where A, C and M are integers and for $x = kM + y$, $y \in 0, 1, \ldots, M-1$, we have $y = x \mod M$. Then $u_n = \frac{s_n}{M}$ plays the role of the n-th generated uniform distribution.

Note that there are at most M distinct s_n, and hence u_n, generated. Also, as soon as for some l

$$s_{l+d} = s_l,$$

then $s_{l+d+1} = s_{l+1}$, so there will be cycles of length d, the smallest such d being called the period. The period should be as large as possible, preferably M. One needs number theory to choose good A, C and M. One popular choice was $M = 2^{31} - 1$, $A = 7^5$, $C = 0$ but nowadays one mixes different congruential generators to obtain much larger periods.

Depending on the software package, random variables with other laws are also available or must be built from uniform ones. We mention a few methods how this can be done.

Inversion Suppose that we want to simulate a random variable with arbitrary cumulative distribution function F. We denote the inverse function of F as F^{-1}, which can be defined even in the case that F is not invertible, namely as *pseudo inverse*

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

If U is a random variable whose law is uniform on $[0, 1]$, then $X = F^{-1}(U)$ is indeed a random variable with cdf F because

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

If $U_1, \ldots, U_n$ denote independent random variables with uniform law on $[0, 1]$, then $X_i := F^{-1}(U_i)$, $i = 1, \ldots, n$, are n independent random variables with cdf F.

As an example consider

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \exp(-\lambda x) & \text{if } x \geq 0 \end{cases}$$

with some parameter λ, which means that we want to simulate random variables whose law is *exponential* with parameter λ. The pseudo inverse of F is

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda}.$$

Therefore $X = -\log(1-U)/\lambda$ is exponentially distributed with parameter λ if U has uniform law on $[0,1]$. One can in fact also choose $X = -\log(U)/\lambda$ because $1-U$ has the same law as U.

Acceptance/rejection method Unfortunately, the pseudo inverse is not easily available for many cumulative distribution functions. An alternative is provided by the *acceptance/rejection* method. Suppose that we want to simulate a random variable X with probability density function f. We may not know right away how to do this but are able to simulate random variables Y whose probability density function g satisfies $f(x) \leq Cg(x)$ for some constant C which is not too large.

In order to simulate X, we first simulate independent random variables Y, U, where Y has density g and U is uniformly distributed on $[0,1]$. We accept Y as a simulation of X if $U \leq f(Y)/(Cg(Y))$. Otherwise we repeat the first step.

In order to convince ourselves that X has indeed the right density, observe that

$$\begin{aligned}
P(X \leq t) &= P(Y \leq t | U \leq f(Y)/(Cg(Y))) \\
&= \frac{P(Y \leq t, U \leq f(Y)/(Cg(Y)))}{P(U \leq f(Y)/(Cg(Y)))} \\
&= \frac{\int_{-\infty}^{t} \int_0^{f(y)/(Cg(y))} du g(y) dy}{\int_{-\infty}^{\infty} \int_0^{f(\widetilde{y})/(Cg(\widetilde{y}))} du g(\widetilde{y}) d\widetilde{y}} \\
&= \frac{\int_{-\infty}^{t} f(y)/(Cg(y)) g(y) dy}{\int_{-\infty}^{\infty} f(\widetilde{y})/(Cg(\widetilde{y})) g(\widetilde{y}) d\widetilde{y}} \\
&= \frac{\int_{-\infty}^{t} f(y) dy}{\int_{-\infty}^{\infty} f(\widetilde{y}) d\widetilde{y}} \\
&= \int_{-\infty}^{t} f(y) dy
\end{aligned}$$

for any $t \in \mathbb{R}$.

As an example we want to simulate a standard normal random variable X. Note that $X = |X|\mathrm{sgn}(X)$. It is not hard to show that $\mathrm{sgn}(X) = \pm 1$, each with probability $1/2$, the modulus $|X|$ has density

$$f(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \geq 0,$$

and $\mathrm{sgn}(X), |X|$ are independent. In order to simulate $|X|$ we consider

$$g(x) = \exp(-x), \quad x \geq 0,$$

which is the density of an exponential random variable with parameter 1. The ratio $h(x) = f(x)/g(x) = \exp(x - x^2/2)\sqrt{2/\pi}$ reaches is maximum at $x = 1$, which implies that $f(x) \leq Cg(x)$ for $C = h(1) = \sqrt{2e/\pi} \approx 1.32$. Note that

$$\frac{f(y)}{Cg(y)} = \exp\left(\frac{-(y-1)^2}{2}\right).$$

Hence we obtain the following algorithm for simulating a standard normal random variable X:

1. Simulate an exponential random variable Y with parameter 1, i.e. simulate a uniform random variable U_1 on $[0, 1]$ and set $Y = -\log U_1$.
2. Simulate another uniform random variable U_2 on $[0, 1]$.
3. Set $|X| = Y$ if $U_2 \leq f(Y)/(Cg(Y))$, otherwise repeat the first two steps.
4. Generate another uniform random variable U_3 on $[0, 1]$.
5. Set $X = |X|$ if $U_3 \geq 0.5$ and $X = -|X|$ otherwise.

Transformations of random vectors The change of variables formula can be restated in probabilistic terms as follows:

Theorem 2.1 *Suppose that $U = (U_1, \ldots, U_n)$ is a random vector with positive density f on $A \subset \mathbb{R}^n$. Moreover, let $h : A \to \mathbb{R}^n$ be a function with continuously differentiable inverse. Then the random vector $X = h(U)$ has a density g given by*

$$g(x) = f(h^{-1}(x)) \cdot \left|\det(Dh^{-1}(x))\right|$$

on $h(A)$.

As an example consider a pair U_1, U_2 of independent random variables with uniform distribution on $[0, 1]$. Now set

$$\begin{aligned} Z_1 &= \sqrt{-2\log(U_1)}\cos(2\pi U_2), \\ Z_2 &= \sqrt{-2\log(U_1)}\sin(2\pi U_2). \end{aligned}$$

Then Z_1, Z_2 is a pair of independent standard normal random variables, which can be verified by using the above theorem. This approach to simulating Gaussian random variables is called *Box-Muller* method. In spite of its simplicity, other methods run faster on computers because they avoid trigonometric functions.

Similar transformations as above exist also for a number of other distributions. Now suppose that we need to simulate a Gaussian random vector $X = (X_1, \ldots, X_d)$ with mean vector $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. We start by simulating d independent standard normal random variables $Z_1, \ldots, Z_d$. This implies that the random vector $Z = (Z_1, \ldots, Z_d)$ is Gaussian with mean vector 0 and covariance matrix 1_d. Determine a matrix $A \in \mathbb{R}^{d\times d}$ with $AA^\top = \Sigma$, e.g. the *Cholesky decomposition* is a lower triangular matrix. Now set $X = AZ + \mu$, X is Gaussian because it is an affine transformation of a Gaussian random vector. Its expectation equals $E(X) = AE(Z) + \mu$. Similarly, we obtain its covariance matrix $\mathrm{Cov}(X) = \mathrm{Cov}(AZ) = A\mathrm{Cov}(Z)A^T = AA^\top = \Sigma$. Consequently, X has the desired law.

2.3 Variance reduction

Recall that the rate of convergence of the Monte Carlo method is relatively low. A way out is to reduce the numerator $\sigma(f(X))$ in the expression (2.2) of the standard deviation $\hat{V}_N$. We briefly discuss three of many approaches here. For more background on variance reduction cf. [Gla03].

Antithetic variables Suppose that the law of X is symmetric in the sense that X and $-X$ have the same distribution. Now replace the standard Monte Carlo estimate (2.1) by

$$\hat{V}_N^{AV} = \frac{1}{2}(\hat{V}_N + \hat{V}_N^-) = \frac{1}{2N}\sum_{n=1}^{N}\Big(f(X_n) + f(-X_n)\Big)$$

with

$$\hat{V}_N^- = \frac{1}{N}\sum_{n=1}^{N} f(-X_n).$$

$\hat{V}_N^{AV}$ is an unbiased estimate for V as well because

$$E(\hat{V}_N^{AV}) = \frac{1}{2}(E(\hat{V}_N) + E(\hat{V}_N^-)) = V$$

by symmetry of the law of X. Its variance equals

$$\begin{aligned}
\mathrm{Var}(\hat{V}_N^{AV}) &= \frac{1}{4}\Big(\mathrm{Var}(\hat{V}_N) + \mathrm{Var}(\hat{V}_N^-) + 2\mathrm{Cov}(\hat{V}_N, \hat{V}_N^-)\Big)\\
&= \frac{1}{2}\mathrm{Var}(\hat{V}_N) + \frac{1}{2}\mathrm{Cov}(\hat{V}_N, \hat{V}_N^-)\\
&\leq \frac{1}{2}\mathrm{Var}(\hat{V}_N) + \frac{1}{2}\sqrt{\mathrm{Var}(\hat{V}_N)\mathrm{Var}(\hat{V}_N^-)}\\
&= \mathrm{Var}(\hat{V}_N),
\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz inequality. Therefore the methods seems at least not to increase the variance. However, if we take into account that it may take up to

twice the time to compute $\hat{V}_N^{AV}$ compared to $\hat{V}_N$, we should require $\mathrm{Var}(\hat{V}_N^{AV}) \leq \frac{1}{2}\mathrm{Var}(\hat{V}_N)$ or, equivalently, $\mathrm{Cov}(\hat{V}_N, \hat{V}_N^-) \leq 0$. Since

$$\begin{aligned}\mathrm{Cov}(\hat{V}_N, \hat{V}_N^-) &= \frac{1}{N^2}\sum_{m,n=1}^{N}\mathrm{Cov}(f(X_m), f(-X_n)) \\ &= \frac{1}{N^2}\sum_{m=1}^{N}\mathrm{Cov}(f(X_m), f(-X_m)) \\ &= \frac{1}{N}\mathrm{Cov}(f(X), f(-X)),\end{aligned}$$

this holds if and only if $f(X)$ and $f(-X)$ are negatively correlated. This holds e.g. for in- or decreasing functions.

In the Black-Scholes model the payoff of the call can be written as an increasing function of a standard and hence symmetric Gaussian random variable. Therefore antithetic variables may help to reduce the variance if Monte Carlo simulation is used to compute call resp. put prices in the Black-Scholes model.

Control variates The control variate approach is applicable if we know some $Y \approx f(X)$ such that $E(Y)$ can be computed explicitly. Since

$$E(f(X)) = E(Y) + E(f(X) - Y),$$

it remains to simulate $E(f(X)-Y)$, which can be done hopefully with smaller variance. To this end, simulate independent pairs (X_n, Y_n), $n = 1, \ldots, N$, each having the same law as (X, Y). Now replace $\hat{V}_N$ from (2.1) by

$$\begin{aligned}\hat{V}_N^{CV} &= \hat{V}_N - \frac{1}{N}\sum_{n=1}^{N} Y_n + E(Y) \\ &= \frac{1}{N}\sum_{n=1}^{N}(f(X_n) - Y_n) + E(Y).\end{aligned}$$

The estimate $\hat{V}_N^{CV}$ is an unbiased estimate for V because

$$E(\hat{V}_N^{CV}) = E(\hat{V}_N) - \frac{1}{N}\sum_{n=1}^{N} E(Y_n) + E(Y) = V.$$

For the variance we obtain

$$\begin{aligned}\mathrm{Var}(\hat{V}_N^{CV}) &= \mathrm{Var}(\hat{V}_N) + \mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^{N} Y_n\right) - 2\mathrm{Cov}\left(\hat{V}_N, \frac{1}{N}\sum_{n=1}^{N} Y_n\right) \\ &= \mathrm{Var}(\hat{V}_N) + \frac{1}{N^2}\sum_{n=1}^{N}\mathrm{Var}(Y_n) - \frac{2}{N^2}\sum_{m,n=1}^{N}\mathrm{Cov}(f(X_m), Y_n)) \\ &= \mathrm{Var}(\hat{V}_N) + \frac{1}{N}(\mathrm{Var}(Y) - 2\mathrm{Cov}(f(X), Y)). \qquad (2.7)\end{aligned}$$

The variance is reduced compared to $\hat{V}_N$ if $\mathrm{Cov}(f(X), Y) \geq \mathrm{Var}(Y)/2$. One should note, however, that more operations are needed for a single simulation. In any case, $f(X)$ and Y should ideally be highly correlated for the method to make sense.

One can improve the method further by considering βY instead of Y for some appropriately chosen constant β. In view of (2.7) we need to minimize

$$\begin{aligned} &\mathrm{Var}(\beta Y) - 2\mathrm{Cov}(f(X), \beta Y) \\ &= \beta^2 \mathrm{Var}(Y) - 2\beta \mathrm{Cov}(f(X), Y) \\ &= \mathrm{Var}(Y) \left(\beta - \frac{\mathrm{Cov}(f(X), Y)}{\mathrm{Var}(Y)} \right)^2 - \frac{\mathrm{Cov}(f(X), Y)^2}{\mathrm{Var}(Y)} \end{aligned}$$

as a function of β. The optimal choice is obviously

$$\beta = \frac{\mathrm{Cov}(f(X), Y)}{\mathrm{Var}(Y)}, \tag{2.8}$$

in which case the variance is reduced by $\mathrm{Cov}(f(X), Y)^2/N\mathrm{Var}(Y)$. Note, however, that it may not be obvious how to compute the optimal β in concrete cases because $\mathrm{Cov}(f(X), Y)$ is typically not known explicitly. A way out is to replace the numerator and possibly also the denominator of (2.8) by a first Monte-Carlo simulation.

A classical application of the control variate approach is the computation of the price of *Asian options* e.g. with payoff $(\sum_{i=1}^{M} S(Ti/M)/M - K)^+$, which corresponds to a call on the arithmetic mean on the stock price path. In the Black-Scholes model, the fair price of the related payoff $Y := ((\prod_{i=1}^{M} S(Ti/M))^{1/M} - K)^+$ can be evaluated explicitly. Indeed, since $\log S(Ti/M), i = 1, \ldots, M$, are jointly normally distributed, the payoff is of the form $Y = (\exp(Z) - K)^+$ for some Gaussian random variable Z which leads to a Black-Scholes type formula for $E(Y)$; see (2.10) below. Since the geometric mean provides a natural proxy of the arithmetic mean, the control variate approach is likely to be useful here. Note that we need to simulate the whole paths $S(Ti/M)$, $i = 0, \ldots, M$, in this application, which will be considered in Section 5.2 below.

Importance sampling Sometimes $f(X)$ attains relatively large values with only small probability. The results of the Monte Carlo estimate depends crucially on how often these unlikely events appear in the simulation. The idea now is to change the distribution of X in order to make the relevant events more likely.

Suppose that the law of X has density g. Let h be another probability density such that $g(x) > 0$ implies $h(x) > 0$. Finally, denote by Y a random variable with probability density function h. The importance sampling estimate is based on independent random variables $Y_1, \ldots, Y_N$ with

the same law as Y. Specifically, it is defined as

$$\hat{V}_N^{IS} = \frac{1}{N}\sum_{n=1}^{N} f(Y_n)\frac{g(Y_n)}{h(Y_n)}. \tag{2.9}$$

It is unbiased for V because

$$\begin{aligned} E(\hat{V}_N^{IS}) &= \frac{1}{N}\sum_{n=1}^{N} E\left(f(Y_n)\frac{g(Y_n)}{h(Y_n)}\right) \\ &= \int f(y)\frac{g(y)}{h(y)}h(y)dy \\ &= \int f(y)g(y)dy \\ &= E(f(X)) = V. \end{aligned}$$

Its variance equals

$$\begin{aligned} \mathrm{Var}(\hat{V}_N^{IS}) &= \frac{1}{N^2}\sum_{n=1}^{N} \mathrm{Var}\left(f(Y_n)\frac{g(Y_n)}{h(Y_n)}\right) \\ &= \frac{1}{N}\left(\int \left(f(y)\frac{g(y)}{h(y)}\right)^2 h(y)dy - V^2\right) \\ &= \frac{1}{N}\left(\int (f(y))^2\frac{g(y)}{h(y)}g(y)dy - V^2\right) \\ &= \frac{1}{N}\left(E\Big((f(X))^2\frac{g(X)}{h(X)}\Big) - V^2\right), \end{aligned}$$

which should be compared to

$$\mathrm{Var}(\hat{V}_N) = \frac{1}{N}\left(E\Big((f(X))^2\Big) - V^2\right).$$

In general, this variance can be both smaller or larger. Ideally, we have $h(x) \approx cf(x)g(x)$ with some norming constant c because this implies $\mathrm{Var}\left(f(Y)\frac{g(Y)}{h(Y)}\right) \approx \mathrm{Var}(1/c) = 0$ and hence $\mathrm{Var}(V_{IS}) \approx 0$. However, even the calculation of the constant

$$\frac{1}{c} = \int f(x)g(x)dx = E(f(X)) = V$$

is as hard as the original problem! Nevertheless, it seems reasonable to look for a density h which is approximately proportional to fg resp. attains large values where this is the case for fg. Let us finally remark that an unbiased estimate of

$$\mathrm{Var}(\hat{V}_N^{IS}) = \frac{1}{N}\mathrm{Var}\left(f(Y)\frac{g(Y)}{h(Y)}\right)$$

is easily obtained from (2.5), which here reads as

$$\widehat{\mathrm{Var}}(\hat{V}_N^{IS}) = \frac{1}{N-1}\left(\frac{1}{N}\sum_{n=1}^{N}\left(f(Y_n)\frac{g(Y_n)}{h(Y_n)}\right)^2 - (\hat{V}_N^{IS})^2\right).$$

As an example consider a call $(S(T)-K)^+$ in the Black-Scholes model which is far out of the money, i.e. K is very large compared to $S(0)$. We already know and will furthermore detail in (3.20) or (3.21) below that its fair price equals

$$V = \exp(-rT)E\left(\left(S(0)\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}X\right)-K\right)^+\right) \tag{2.10}$$

where X denotes some standard normal random variable, i.e. $V = E(f(X))$ for

$$f(x) = \exp(-rT)\left(S(0)\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}x\right)-K\right)^+.$$

In a Monte Carlo simulation

$$\hat{V}_N = \frac{1}{N}\sum_{n=1}^{N}\exp(-rT)\left(S(0)\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}X_n\right)-K\right)^+$$

according to (2.1), only the few outcomes $X_n > d$ contribute to the empirical mean, where

$$d = \frac{\log\frac{K}{S(0)} - (r-\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}.$$

Instead of the density $g(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$ of standard normal random variable X_n let us consider Gaussian random variables Y_n with density

$$h(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2}\right),$$

where the mean is chosen as $\mu = d$. Indeed, fg has its largest values close to d because f vanishes below d and $g(x)$ decays very quickly for increasing x. Since h is concentrated around μ, the above reasoning suggest to choose $\mu \approx d$. The Monte Carlo estimate (2.9) now reads as

$$\hat{V}_N^{IS} = \frac{1}{N}\sum_{n=1}^{N}\exp\left(-rT-Y_n\mu+\frac{\mu^2}{2}\right)\left(S(0)\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}Y_n\right)-K\right)^+,$$

where the Y_n are independent Gaussian random variables with mean μ and variance 1.

Chapter 3

Continuous market models and Itō calculus

Most models in financial mathematics are formulated in continuous time. One cannot work with them without a certain background in the theory of stochastic processes. This is provided here on an informal level and with a view towards finance. The mathematical details are provided in the course on stochastic integration and continuous time finance.

3.1 General theory of stochastic processes

Most notions in the general theory of stochastic processes have a counterpart in discrete time, which we already encountered in the previous course on Mathematical Finance. We provide them here only as a transition to the next section, where the concepts are reformulated for the more specific setup that we really need, namely the case of Itō processes. For more precise definitions, motivation, and examples we refer to the previous course.

Mathematical framework for random phenomena The starting point is a *probability space* $(\Omega, \mathscr{F}, P)$ consisting of the set Ω of possible outcomes, a *σ-field* $\mathscr{F}$ on Ω and a *probability measure* P on $\mathscr{F}$. Recall that a σ-field is a collection of subsets of Ω called *events*. The probability measure P assigns probabilities to any of these events. For deep mathematical reasons it is typically not possible to assign probabilities to all subsets of Ω, which is why the σ-field $\mathscr{F}$ is introduced. Using the σ-field $\mathscr{F}$ instead of the full set of all subsets can be seen as a technical issue (and for the QF-students you can always think of each subset of Ω to be an event). However, the following σ-algebras have an important interpretation: We model the dynamic structure on the probability space using a *filtration* $\mathbf{F} = (\mathscr{F}_t)_{t\geq 0}$. As in discrete time, this

denotes an increasing family of σ-fields. The σ-field $\mathscr{F}_t$ contains all events that are no longer random or undecided at time t. The whole structure $(\Omega, \mathscr{F}, \mathbf{F}, P)$ is called *filtered probability space*. In practice, typically neither Ω nor $\mathscr{F}$ nor $\mathbf{F}$ nor P is explicitly specified. Only some of its properties are needed.

Stochastic processes The real objects of interest are stochastic processes $X = (X(t))_{t\geq 0}$, i.e. collections of random variables $X(t)$, indexed by time t. This random variable $X(t)$ stands for the value of some quantity at time t, e.g. a stock price, the number of shares in a portfolio, etc. It could be vector-valued as $X(t) = (X_1(t), \ldots, X_d(t))$, denoting e.g. the value at time t of d stocks in the market. From a different viewpoint, X can be interpreted as a random function of time. In this course we consider mostly *continuous processes*, i.e. the random path $(X(\omega, t))_{t\geq 0}$ depends continuously on t. Advanced models in particular for asset prices do also allow for jumps. Nevertheless, the paths are generally still assumed to be *càdlàg*, i.e. right continuous with limits from the left (*continu à droite avec des limites à gauche*). This implies that $X(t)$ denotes the value of the process after a jump has happened, whereas the left-hand limit $X(t-) := \lim_{s\uparrow t} X(s)$ denotes the value before the jump. The jump itself is denoted by $\Delta X(t) := X(t) - X(t-)$.

Functions considered in calculus are typically of *finite variation*, i.e. a difference of two increasing functions. Surprisingly, this is not the case for stochastic processes, which tend to be very shaky or rapidly oscillating. In mathematical terms, they are of *infinite variation*. This can be interpreted such that the path of the process has infinite length on finite intervals because it has so many kinks.

All processes are typically assumed to be *adapted* to the filtration, i.e. the value $X(t)$ is supposed to be observable at t at the latest. For stochastic integration in discrete time we also introduced processes which are known one period ahead. Such *predictable* processes which are observable an infinitesimal time before the respective time t, also play an important role in continuous time. However, we skip the somewhat technical definition here.

Any process X generates a filtration, namely the smallest filtration it is adapted to. This means that all randomness in the model originates in the process X. This assumption is sometimes needed for statements concerning in particular market completeness. (This is natural: If a claim is based on information not coming from X, we cannot expect to hedge it using X only.) A *stopping time* T is a random time that is observable, i.e. we know at any time t whether T has already happened or not. If X denotes a process and T a stopping time, the *stopped process* X^T equals X up to time T and stays frozen at the value $X(T)$ afterwards.

A *martingale* is an adapted process X satisfying $E(X(t)|\mathscr{F}_s) = X(s)$ for any $s \leq t$, i.e. we expect on average the present value for the future. A *supermartingale* is decreasing on average,

i.e. we have $E(X(t)|\mathscr{F}_s) \leq X(s)$ for any $s \leq t$, whereas a *submartingale* grows on average, i.e. $E(X(t)|\mathscr{F}_s) \geq X(s)$ for any $s \leq t$. Martingales, supermartingales, and submartingales X are *stable under stopping*, i.e. the stopped process X^T is again a martingale, supermartingale resp. submartingale for any stopping time T. Random variables Y naturally generate a martingale X via $X(t) = E(Y|\mathscr{F}_t)$. An important example of this type is connected to measure changes to some probability measure $Q \sim P$. If $\frac{dQ}{dP}$ denotes the density of Q, then the martingale $Z(t) = E(\frac{dQ}{dP}|\mathscr{F}_t)$ is called *density process* of Q relative to P.

A large class of adapted stochastic processes X allow for a decomposition of the form $X = X(0) + M + A$ with a martingale M and a predictable process of finite variation A. This *Doob-Meyer decomposition* generalizes the discrete time Doob decomposition. The process A stands for the trend or predictable growth whereas the martingale part M represents the unsystematic random deviation from that trend.

Stochastic integral The key concept from stochastic calculus needed in finance is the *stochastic integral* representing in particular financial gains. If the integrand H is piecewise constant, i.e. of the form

$$H(t) = \sum_{i=1}^{n} V_i 1_{(t_{i-1},t_i]}(t),$$

then the integral $H \bullet X = \int H(t)dX(t)$ is defined as

$$H \bullet X(t) := \int_0^t H(s)dX(s) := \sum_{i=1}^{n} V_i(X(t \wedge t_i) - X(t \wedge t_{i-1})). \tag{3.1}$$

If $H(t)$ stands for the number of shares in the portfolio at time t and $X(t)$ denotes the stock price at time t, then simple bookkeeping yields that $\int_0^t H(s)dX(s)$ stands for the financial gains of the portfolio from time 0 to t. If both $H = (H_1, \ldots, H_d)$ and $X = (X_1, \ldots, X_d)$ are vector-valued, then (3.1) needs to be replaced by

$$H \bullet X(t) := \int_0^t H(s)dX(s) := \sum_{j=1}^{d} \sum_{i=1}^{n} V_{i,j}(X_j(t \wedge t_i) - X_j(t \wedge t_{i-1}))$$

because the total financial gains of the portfolio are the sum of the gains in any of the d assets.

The stochastic integral is defined for more general integrands H than piecewise constant ones but its mathematical definition is quite involved. Roughly speaking, it may be viewed as a limit of expressions of the form (3.1) if the integrand H is approximated closer and closer by piecewise constant ones.

Later, a differential notation turns out to be useful. We write

$$dY(s) = H(s)dX(s)$$

as a shorthand for

$$\int_0^t 1dY(s) = \int_0^t H(s)dX(s),$$

which — by $\int_0^t 1dY(s) = Y(t) - Y(0)$ — in turn means nothing else than

$$Y(t) = Y(0) + \int_0^t H(s)dX(s).$$

The stochastic integral is linear in both H and X, moreover it is continuous in H. The stochastic integral of a martingale is typically a martingale if we forget about important mathematical subtleties as we generally do in this course. We have the rule

$$H \bullet (K \bullet X) = (HK) \bullet X,$$

which one applies without noticing in the differential notation:

$$H(t)(K(t)dX(t)) = (H(t)K(t))dX(t).$$

From discrete time, we also know the *integration by parts* rule

$$X(t)Y(t) = X(0)Y(0) + \int_0^t X(s-)dY(s) + \int_0^t Y(s-)dX(s) + [X,Y](t)$$

or

$$d(X(t)Y(t)) = X(t-)dY(t) + Y(t-)dX(t) + d[X,Y](t)$$

in differential notation. Here, $[X,Y](t)$ denotes the *covariation process*, which is the limit of expressions

$$\sum_{k=1}^{n} \left(X(\tfrac{k}{n}t) - X(\tfrac{k-1}{n}t)\right)\left(Y(\tfrac{k}{n}t) - (Y(\tfrac{k-1}{n}t)\right) \tag{3.2}$$

if we let n tend to infinity. In other words, $[X,Y]$ sums up products of increments of X and Y over infinitesimal time intervals. For $X = Y$ it is called quadratic variation. Sometimes, the possibly confusing notation

$$dX(t)dY(t) = d[X,Y](t)$$

is used in the literature, which is motivated by (3.2). We will see in the next section how it is computed in more specific situation. At this point we only note that it vanishes if X or Y is a continuous process of finite variation. Moreover, $[X,Y]$ is itself of finite variation.

The covariation is linear in both X and Y. Moreover, it commutes with stochastic integration in the sense that

$$\left[\int_0^{\cdot} H(s)dX(s), Y\right](t) = \int_0^t H(s)d[X,Y](s).$$

The most important formula in stochastic calculus is *Itō's formula*, which is

$$\begin{aligned} f(X(t)) &= f(X(0)) + \int_0^t f'(X(s-))dX(s) + \frac{1}{2}\int_0^t f''(X(s-))d[X^c,X^c](s) \\ &\quad + \sum_{s\le t}\left(f(X(s)) - f(X(s-)) - f'(X(s-))\Delta X(s)\right) \end{aligned}$$

for one-dimensional processes and smooth functions $f : \mathbb{R} \to \mathbb{R}$. Here, X^c stands for the continuous martingale part of X which we do not want to define exactly. It can be shown, however, that $[X^c, X^c]$ simplifies to $[X, X]$ for continuous processes. Consequently, Itō's formula reads as

$$f(X(t)) \;=\; f(X(0)) + \int_0^t f'(X(s))dX(s) + \frac{1}{2}\int_0^t f''(X(s))d[X,X](s)$$

or, in differential notation,

$$df(X(t)) = f'(X(t))dX(t) + \frac{1}{2}f''(X)d[X,X](t)$$

in the continuous case.

If $X = (X_1, \ldots, X_d)$ is d-dimensional and $f : \mathbb{R}^d \to \mathbb{R}$, the partial derivatives $\partial_i f$ and second partial derivatives $\partial_{ij} f$ come into play:

$$f(X(t)) = f(X(0)) + \sum_{i=1}^d \int_0^t \partial_i f(X(s))dX_i(s) + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \partial_{ij} f(X(s))d[X_i, X_j](s) \quad (3.3)$$

or

$$df(X(t)) = \sum_{i=1}^d \partial_i f(X(t))dX_i(t) + \frac{1}{2}\sum_{i,j=1}^d \partial_{ij} f(X(t))d[X_i, X_j](t).$$

in differential notation.

General SDEs Some stochastic processes are defined in terms of integral equations that involve the as yet unknown process on both sides. More specifically, suppose that a process X, a function $f : \mathbb{R} \to \mathbb{R}$ and an initial value z are given. We look for a process Z of the form

$$Z(t) = z + \int_0^t f(Z(s-))dX(s)$$

or

$$dZ(t) = f(Z(t-))dX(t), \quad Z(0) = z \quad (3.4)$$

in differential notation. As a particular example consider the *stochastic exponential* or *Doléans exponential* $Z = \mathscr{E}(X)$, which is the solution to (3.4) for $f(x) = x$ and $z = 1$, i.e. satisfying

$$dZ(t) = Z(t-)dX(t), \quad Z(0) = 1.$$

Intuitively, this means that the changes of Z are proportional to the present value of Z. We have encountered the discrete-time version of this process in the previous course.

Since Z appears on both sides, it is not a priori clear that a unique solution to (3.4) exists. However, it can be shown this is generally the case for sufficiently well-behaved functions f. In

particular, this holds for the stochastic exponential, which has an explicit representation of the form

$$\mathscr{E}(X)(t) = \exp\left(X(t) - X(0) - \frac{1}{2}[X,X](t)\right) \prod_{s\leq t} \left(e^{-\Delta X(s)}(1+\Delta X(s))\right).$$

For continuous processes X, this reduces to

$$\mathscr{E}(X)(t) = \exp\left(X(t) - X(0) - \frac{1}{2}[X,X](t)\right).$$

If the probability measure is changed from P to $Q \sim P$, martingales do not remain martingales any more. Denote the density process of Q relative to P by Z. Then a process X is a martingale relative to Q if and only if XZ is a martingale relative to P. This simple statement can be used to show the following version of *Girsanov's theorem*: if X denotes a P-martingale, then

$$X(t) - \int_0^t \frac{1}{Z(s-)} d[Z,X](s)$$

is a martingale under Q.

Lévy processes and Brownian motion *Lévy processes* X are defined as processes with *stationary* and *independent increments*. This means that the increment $X(t) - X(s)$ of the process for $s \leq t$ is independent of the information up to time s and its law depends only on the length $t - s$ of the interval, but not on s and t themselves. Intuitively, this means that Lévy processes stand for constant growth in a stochastic sense. In intervals of the same length they grow in the same way. They can also be interpreted as the continuous time counterpart of *random walks*, i.e. the sum of independent and identically distributed random variables. Arguably they can be viewed as the most basic and important class of stochastic processes.

Lévy processes may have jumps. If we restrict attention to processes with continuous paths, we end up with the much smaller class of *Brownian motions* X. Their law is specified by only two parameters, namely the expectation $\mu = E(X(1))$ and the variance $\sigma^2 = \mathrm{Var}(X(1))$. In fact, the random variable $X(t)$ is Gaussian with expected value $E(X(t)) = \mu t$ and variance $\mathrm{Var}(X(t)) = \sigma^2 t$. Moreover, the process can be written as

$$X(t) = \mu t + \sigma W(t), \tag{3.5}$$

where W denotes *standard Brownian motion* or a *Wiener process*, i.e. a Brownian motion with $\mu = 0$, $\sigma^2 = 1$. As far as its law is concerned, any continuous process of constant growth can hence be viewed as a linear combination of two building blocks: a deterministic linear trend t and an erratically moving Wiener process W. One may note that Brownian motion are of infinite variation and nowhere differentiable unless they are deterministic linear functions. The

Wiener process may be regarded as the most important stochastic process altogether. It has been introduced in 1900 by Louis Bachelier in his Ph.D. thesis on financial mathematics and by Albert Einstein in his 1905 paper on thermal molecular motion.

Brownian motion as the only continuous process with stationary and independent increments has a natural extension to the multivariate case. The law of such a continuous $\mathbb{R}^d$-valued Lévy process $X = (X_1, \dots, X_d)$ is specfied by the expectation vector $\mu = E(X(1)) = (E(X_1(1)), \dots, E(X_d(1)))$ and the covariance matrix $c = \Sigma\Sigma^\top = \mathrm{Cov}(X(1)) = (\mathrm{Cov}(X_i(1), X_j(1))_{i,j=1,\dots,d}$. For $\mu = 0$ and $c = 1_d$ we call X standard Brownian motion in $\mathbb{R}^d$. As for $d = 1$, both expectation and covariance matrix of the Gaussian random vector $X(t)$ are linear in t. Moreover, any d-dimensional Brownian motion is of the form

$$X(t) = \mu t + \Sigma W(t)$$

with standard Brownian motion W in $\mathbb{R}^d$.

The quadratic variation of Brownian motion (3.5) equals $[X, X](t) = \sigma^2 t$ and in particular

$$[W, W](t) = t$$

for a Wiener process W. More generally, the covariation process of Brownian motion $X = (X_1, \dots, X_d)$ in $\mathbb{R}^d$ with covariance matrix c is given by $[X_i, X_j](t) = c_{ij}t$ and in particular

$$[W_i, W_j](t) = \begin{cases} t & \text{if } i = j, \\ 0 & \text{otherwise} \end{cases}$$

for standard Brownian motion W in $\mathbb{R}^d$.

3.2 Itō processes

We do not need the theory of stochastic processes in full generality for our purposes. We focus instead on processes with continuous paths, i.e. without jumps. Although jumps play an important role in theory and practice, continuity may be considered as an acceptable first approximation which simplifies the mathematical treatment considerably. Among continuous models Itō processes play a predominant role. For an intuitive understanding let us recall the basic idea underlying ordinary calculus. Deterministic functions of constant growth can be written as

$$X(t) = \mu t$$

with slope or growth rate μ. Of course, not all real world phenomena exhibit constant growth. But the fundamental idea of analysis is the observation that they often do so on a local scale. In other words, a large class of functions are of the form

$$dX(t) = \mu(t)dt,$$

i.e. in a neighbourhood of t they resemble a linear function with growth rate $\mu(t)$. Of course, this local growth rate $\mu(t)$ is just the derivative of X in the language of calculus, which now depends on t.

Now, we want to mimic this idea in a random setting. As argued above, continuous processes of constant growth are of the form

$$X(t) = \mu t + \sigma W(t)$$

with standard Brownian motion W and constants μ, σ. Along the same lines as in deterministic setups, we want to consider processes that locally resemble such a Brownian motion with *drift rate* μ and *diffusion coefficient* σ. To this end, we have to make sense of an equation of the form

$$dX(t) = \mu(t)dt + \sigma(t)dW(t), \tag{3.6}$$

where the local drift rate $\mu(t)$ and the local diffusion coefficient $\sigma(t)$ now depend on time t and possibly also on randomness. (3.6) does not make sense in ordinary calculus but it can be interpreted as an integral equation

$$X(t) = X(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s) \tag{3.7}$$

in the sense of the previous section. Processes of this form are called *Itō processes*.

We will reformulate the results and rules of Section 3.1 for the special case of Itō processes. (3.7) is also the Doob-Meyer decomposition of X. The martingale part is the second integral, whereas the first signifies the trend. X is essentially a martingale if the drift rate μ vanishes. For positive μ the process X is a submartingale, for negative μ a supermartingale. In particular, we obtain

$$E\Big(\int_0^t \sigma(s)dW(s)\Big) = 0$$

because the expectation of a martingale is its initial value. Stopping an Itō process X at some stopping time T just means that μ and σ vanish after T, i.e.

$$dX^T(t) = \mu(t)1_{\{t\leq T\}}dt + \sigma(t)1_{\{t\leq T\}}dW(t).$$

Since the quadratic variation of standard Brownian motion W is $[W, W](t) = t$, the rules of the previous section yield

$$d[X, X](t) = \sigma^2(t)dt.$$

More generally, we have

$$d[X, \tilde{X}](t) = \sigma(t)\tilde{\sigma}(t)dt$$

for Itō processes X and $\tilde{X}$ with

$$d\tilde{X}(t) = \tilde{\mu}(t)dt + \tilde{\sigma}(t)dW(t).$$

Sometimes, one can find the shorthand

$$dtdt = dtdW(t) = dW(t)dt = 0, \quad dW(t)dW(t) = dt$$

to help working with Itō processes. Itō's formula reads as

$$df(X(t)) = \left(f'(X(t))\mu(t) + \frac{1}{2}f''(X(t))\sigma^2(t)\right)dt + f'(X(t))\sigma(t)dW(t)$$

or

$$\begin{aligned} df(t,X(t)) &= \left(\partial_1 f(t,X(t)) + \partial_2 f(t,X(t))\mu(t) + \frac{1}{2}\partial_{22} f(t,X(t))\sigma^2(t)\right)dt \quad (3.8)\\ &\quad + \partial_2 f(t,X(t))\sigma(t)dW(t) \quad (3.9)\end{aligned}$$

if the function f depends on both t and X. If we apply this to the function $f(x) = x^2$, we obtain

$$dX^2(t) = (2X(t)\mu(t) + \sigma^2(t))dt + 2X(t)\sigma(t)dW(t).$$

Since the expectation of the martingale part vanishes, we obtain for $\mu(t) = 0$ the following *Itō isometry*

$$\begin{aligned} \mathrm{Var}\left(\int_0^t \sigma(s)dW(s)\right) &= E\left(\left(\int_0^t \sigma(s)dW(s)\right)^2\right)\\ &= E\left(\int_0^t \sigma^2(s)ds\right).\end{aligned}$$

Similarly, one can show

$$\mathrm{Cov}\left(\int_0^t \sigma(s)dW(s), \int_0^t \tilde{\sigma}(s)dW(s)\right) = E\left(\int_0^t \sigma(s)\tilde{\sigma}(s)ds\right). \quad (3.10)$$

As a side remark, for deterministic μ, σ

$$X(t) = \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s)$$

is a Gaussian random variable with mean $\int_0^t \mu(s)ds$ and variance $\int_0^t \sigma^2(s)ds$.

The stochastic exponential of an Itō process readily follows from the general formula as

$$\mathscr{E}(X)(t) = \exp\left(\int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s)\right)ds + \int_0^t \sigma(s)dW(s)\right). \quad (3.11)$$

For Brownian motion we have the following version of Girsanov's theorem. If the density process Z of $Q \sim P$ reads as

$$dZ(t) = Z(t)\sigma(t)dW(t),$$

the Wiener process W can be written as

$$dW(t) = dW^Q(t) + \sigma(t)dt$$

with some Q-standard Brownian motion W^Q, i.e. it acquires a drift with rate $\sigma(t)$ under Q.

Finally, we want to mention the important *martingale representation theorem* for Brownian motion. It basically states that any martingale X has a representation as a stochastic integral of a Wiener process W, i.e. it is of the form

$$X(t) = X(0) + \int_0^t \sigma(s)dW(s)$$

for some process σ. However, this holds only if the filtration is generated by W, i.e. if all the randomness in the model can be traced back to the Brownian motion W under consideration. This result, which also extends to the d-dimensional case, plays an important role for market completeness in Mathematical Finance.

Stochastic differential equations (SDEs) For the description of a dynamical system over time, one usually uses differential equations. We start with two examples from history:

(i) *Malthusian growth* [Malthus, 1798]

$$X(t) = a_0 e^{at},\ t \geq 0$$

then

$$\frac{dX}{dt} = a_0 a e^{at} = aX(t).$$

$X(\cdot)$ is the unique solution of the differential equation

$$\frac{dX}{dt} = aX$$

with initial value $X(0) = a_0$. This is also written as

$$dX(t) = aX(t)dt,$$

or in integral form

$$X(t) = X(0) + \int_0^t aX(s)ds$$

Such X is said to fulfill *Malthusian law* of growth for populations, economics, etc. Here, the growth rate is proportional to size of population, etc. with *unlimited resources*.

(ii) *Logistic growth* [Verhulst, 1837]

$$\frac{dX}{dt} = aX - bX^2 = aX(1 - \frac{b}{a}X)$$

Logistic law of growth: In contrast to the previous example, the growth rate gets lower the closer X is to $\frac{a}{b}$; this reflects *limited resources*. *Solution* with initial value $X(0) = a_0$:

$$X(t) = \frac{a \cdot a_0}{ba_0 + (a - ba_0)e^{-at}}.$$

For Itō processes *stochastic differential equations* (SDE's) are typically stated in the form

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad X(0) = x \tag{3.12}$$

with deterministic functions μ, σ of the current state $X(t)$ of the solution process and possibly also of time t. If the functions μ, σ are regular enough, then this equation or *initial value problem* has a unique solution. As a side remark, note that (3.12) reduces to

$$dX(t) = \mu(t, X(t))dt, \quad X(0) = x$$

for $\sigma = 0$, which is another way of writing the ordinary differential equation

$$\frac{dX(t)}{dt} = \mu(t, X(t)), \quad X(0) = x.$$

In that sense stochastic differential equations of the form (3.12) truly generalise ordinary differential equations by adding a stochastic term.

The following linear SDE

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \quad X(0) = x$$

with constants μ, σ can be rephrased as

$$dX(t) = X(t)dY(t), \quad X(0) = x$$

with $Y(t) = \mu t + \sigma W(t)$. Using formula (3.11) for the stochastic exponential, we conclude

$$X(t) = x\mathscr{E}(Y)(t) = x \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right).$$

This process X is called *geometric Brownian motion*.

As a second example we consider the *Ornstein-Uhlenbeck process*, which solves the SDE

$$dX(t) = -\lambda X(t)dt + \sigma dW(t), \quad X(0) = x \tag{3.13}$$

with constants λ, σ. We will verify that X is of the form

$$X(t) = xe^{-\lambda t} + \int_0^t e^{-\lambda(t-s)}\sigma dW(s). \tag{3.14}$$

The initial condition $X(0) = x$ is obviously satisfied by X in (3.14). Set $U(t) = e^{-\lambda t}$ and $V(t) = \int_0^t e^{\lambda s}\sigma dW(s)$. We must show that our candidate $X(t) = U(t)(x + V(t))$ from (3.14) satisfies SDE (3.13). Firstly, note that

$$dU(t) = -\lambda e^{-\lambda t}dt = -\lambda U(t)dt$$

e.g. by Itō's formula and

$$d(x + V(t)) = e^{\lambda t}\sigma dW(t).$$

Integration by parts yields

$$\begin{aligned} dX(t) &= d(U(t)(x+V(t))) \\ &= (x+V(t))dU(t) + U(t)d(x+V(t)) + d[U, x+V](t) \\ &= -\lambda U(t)(x+V(t))dt + U(t)e^{\lambda t}\sigma dW(t) + 0 \\ &= -\lambda X(t)dt + \sigma dW(t), \end{aligned}$$

which is indeed the SDE under consideration.

3.3 Essential notions of Mathematical Finance

We do not study stochastic processes here for their own sake but in the context of finance. In this section we introduce resp. recall important notions from financial mathematics whose discrete-time counterparts we encountered in the previous course.

The starting point is the $d+1$-dimensional *price process* $S(t) = (S_0(t), \ldots, S_d(t))$ of $d+1$ traded securities in the market. We mostly work with two assets (i.e. $d = 1$) whose price processes $S_0(t) = e^{rt}$ and $S_1(t) = S_1(0)\exp(\mu t + \sigma W(t))$ are of the form (0.1, 0.2). As in the Recap-Section (p. 1) we will later denote them as $B(t), S(t)$ for ease of notation.

Trading is expressed in terms of a likewise $d+1$-dimensional *trading strategy* or *portfolio* $\varphi(t) = (\varphi_0(t), \ldots, \varphi_d(t))$. The *value of the portfolio* at time t is

$$V_\varphi(t) = \varphi(t)^\top S(t) = \sum_{i=0}^{d} \varphi_i(t) S_i(t).$$

The strategy is called *self-financing* if

$$dV_\varphi(t) = \varphi(t)^\top dS(t) \left(= \sum_{i=0}^{d} \varphi_i(t) dS_i(t) \right), \tag{3.15}$$

which as usually means that no funds are added or withdrawn after inception of the portfolio.

Bookkeeping is often simplified by using asset S_0 as a *numeraire* for discounting, i.e. all prices are expressed as multiples of S_0.

$$\hat{S}(t) := S(t)/S_0(t) = (1, S_1(t)/S_0(t), \ldots, S_d(t)/S_0(t))$$

is called *discounted price process* and $\hat{V}_\varphi(t) = V_\varphi(t)/S_0(t)$ *discounted value process* of portfolio φ. Observe that a strategy is self-financing if and only if

$$d\hat{V}_\varphi(t) = \varphi(t) d\hat{S}(t),$$

i.e. if (3.15) holds in discounted terms. As in discrete time we can identify self-financing strategies with their arbitrarily chosen initial capital $V(0)$ and the partial portfolio $\varphi_1(t), \dots, \varphi_d(t)$. The missing component $\varphi_0(t)$ is then uniquely determined by the self-financing condition.

The key concept of financial mathematics is *arbitrage*, which essentially refers to a self-financing strategy φ with initial value $V_\varphi(0) = 0$, and final value $V_\varphi(T) \geq 0$ with $P(V_\varphi(T) > 0) > 0$. In fact, this definition must be modified or refined slightly in order to obtain a reasonable theory. We do not discuss these very delicate and nontrivial aspects here. Under absence of arbitrage the *law of one price* holds, which states that two self-financing portfolios φ, ψ with the same final value $V_\varphi(T) = V_\psi(T)$ have the same value at any time $t \leq T$.

The *(first) fundamental theorem of asset pricing (FTAP)* links absence of arbitrage with martingales. Up to subtle technicalities, a market does not allow for arbitrage opportunities if and only if there exists an *equivalent martingale measure (EMM)*, i.e. a probability measure $Q \sim P$ such that the discounted price processes $\hat{S}_i$ are Q-martingales for $i = 0, \dots, d$. We generally assume that absence of arbitrage holds.

The notion of arbitrage opens the door towards statements on derivative prices. In its most basic form a *derivative*, *option* or *contingent claim* is expressed in terms of a random variable X, namely its payoff at time T. Often, it is of the form $X = f(S_1(T))$ with some funtion f as e.g. $f(x) = (x - K)^+$ for a European call. General theory tells us that the only reasonable option price at time $t \leq T$ is given by the conditional expectation

$$V(t) = S_0(t) E_Q(X/S_0(T)|\mathscr{F}_t), \tag{3.16}$$

where Q denotes an EMM as in the FTAP. Otherwise arbitrage opportunities would be introduced into the market by trading the option. If there several EMM's Q exist, (3.16) does not yet determine option prices uniquely. However, if several options are traded, they all have to be priced using the same EMM Q. The right-hand side of (3.16) does not depend on the chosen EMM Q if and only if the option can be *replicated* by dynamic trading. In other words, if there exists a self-financing *hedging* strategy φ with final value $V_\varphi(T) = X$. This holds in particular if there is only one EMM. More specifically, the *second fundamental theorem of asset pricing* states that there exists only one EMM if and only if *any* contingent claim can be replicated. In this case, the market is called *complete*.

In this course we consider market models which are driven by an n-dimensional Brownian motion. As a rule of thumb such markets are complete if and only if the number n of Brownian motions is less than the number of traded assets, i.e. if we have $n \leq d$. However, this general rule may be violated in concrete cases.

American options are defined in terms of their exercise price process $X(t)$, which specifies how much money the owner of the option receives if the option is exercised at time t. General

theory tells us that the only reasonable prices of an American option are of the form $V(t) = S_0(t)\hat{V}(t)$, where $\hat{V}$ denotes the *Q-Snell envelope* of $\hat{X}(t) := X(t)/S_0(t)$ under some EMM Q, i.e. the smallest Q-supermartingale satisfying $\hat{V} \geq \hat{X}$. Otherwise, arbitrage opportunities arise. As a supermartingale, the Snell envelope has negative drift rate relative to Q. However, whenever $\hat{V}(t-) > \hat{X}(t-)$ it even behaves as a martingale, i.e. its Q-drift rate vanishes on the set $\{\hat{V}(t-) > \hat{X}(t-)\}$. As noted in the course on Mathematical Finance, the Snell envelope is linked to an optimal stopping problem. We have

$$V(t) = S_0(t) \sup_{\tau} E_Q(X(\tau)/S_0(\tau)|\mathscr{F}_t),$$

where the supremum extends over all stopping times stopping at t or later. For the particular case of an American call with payoff process $X(t) = (S(t) - K)^+$, the fair option price can be computed more easily, provided that the numeriare S_0 is an increasing process. In this situation the optimal stopping time is to wait till expiry T, which means that the fair value of an European and American call with the same strike and maturity coincide. This holds for the put as well if the numeriare S_0 is constant, which, however, is typically not the case.

3.4 Black-Scholes model

We turn now to the Black-Scholes model

$$B(t) = e^{rt}, \tag{3.17}$$

$$S(t) = S(0)\exp(\mu t + \sigma W(t)), \tag{3.18}$$

which was already mentioned above. With the help of the martingale convergence theorem one can show that this market model is complete if one works with the filtration generated by W. This is in line with the above rule of thumb because the number of assets beyond the numeraire coincides with the number of driving Brownian motions.

In order to compute option prices we need the law of the stock price under the unique EMM Q. To this end, observe that the discounted price process

$$\hat{S}(t) = S(0)\exp((\mu - r)t + \sigma W(t))$$

solves the stochastic differential equation

$$d\hat{S}(t) = \hat{S}(t)\left(\left(\mu - r + \frac{\sigma^2}{2}\right)dt + \sigma dW(t)\right)$$

because of (3.11). From this Itō process decomposition one can observe that $\hat{S}$ is a martingale only if $\mu = r - \frac{\sigma^2}{2}$. But if we consider the measure $Q \sim P$ with density process

$$dZ(t) = -Z(t)\frac{\mu - r + \sigma^2/2}{\sigma}dW(t), \quad Z(0) = 1$$

Girsanov's theorem as in Section 3.2 yields that

$$dW(t) = dW^Q(t) - \frac{\mu - r + \sigma^2/2}{\sigma} dt$$

with some Q-Wiener process W^Q. This yields

$$d\hat{S}(t) = \hat{S}(t)\sigma dW^Q(t), \tag{3.19}$$

which implies that $\hat{S}$ is a Q-martingale because the drift part in the Itō process decomposition relative to Q vanishes. Hence we have found the unique EMM Q. According to (3.11) and using $S = \hat{S}B$ we obtain

$$S(t) = S(0) \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W^Q(t)\right).$$

Pricing by integration

We can now compute the fair price of an European option with payoff $X = f(S(T))$, which is given by

$$V(0) = B(0)E_Q(f(S(T))/B(T)). \tag{3.20}$$

We only need to observe that $W^Q(T)$ has the same law as $\sqrt{T}Y$ for a standard normal random variable Y. Moreover, since we have

$$E(g(Y)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) e^{-\frac{x^2}{2}} dx$$

for any function g, we conclude that

$$V(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f\left(S(0) \exp\left(\left(r - \sigma^2/2\right)T + \sigma\sqrt{T}x\right)\right) e^{-rT} e^{-\frac{x^2}{2}} dx. \tag{3.21}$$

Hence option prices in the Black-Scholes model can be computed by numerical integration. This extends to all models where the density of the law of $S(T)$ or $\log(S(T))$ under the pricing measure Q is known in closed form.

In order to compute the fair option price

$$V(t) = B(t)E_Q(f(S(T))/B(T)|\mathscr{F}_t) \tag{3.22}$$

at time t, we use the representation

$$S(T) = S(t) \exp\left(\left(r - \frac{\sigma^2}{2}\right)(T-t) + \sigma(W^Q(T) - W^Q(t))\right)$$

and observe that $W^Q(T) - W^Q(t)$ has the same law as $\sqrt{T-t}Y$ for a standard normal random variable Y. Moreover, $S(t)$ is known at time t and is hence to be treated as a constant in the computation of the conditional expectation (3.22). This leads to the generalisation

$$V(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f\left(S(t) \exp\left(\left(r - \sigma^2/2\right)(T-t) + \sigma\sqrt{T-t}x\right)\right) e^{-r(T-t)} e^{-\frac{x^2}{2}} dx.$$

of (3.21).

In some cases as e.g. for call and put options, $V(t)$ can be expressed more explicitly. For $f(x) = (x-K)^+$ this was done in the previous course. For this payoff we obtained the following Black-Scholes formula:

$$V(t) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \tag{3.23}$$

with

$$\begin{aligned} d_1 &= \frac{\log\frac{S(t)}{K} + r(T-t) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= \frac{\log\frac{S(t)}{K} + r(T-t) - \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}} \end{aligned}$$

and the cumulative distribution function

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$$

of the standard normal distribution.

Note that the payoffs of calls and puts with the same strike and maturity are related by

$$(K - S(T))^+ = Ke^{-rT}B(T) - S(T) + (S(T) - K)^+.$$

The law of one price yields that a put has always the same value as a portfolio of Ke^{-rT} bonds, -1 shares of stock, and $+1$ call. This relationship is called *put-call parity*. For the European put with payoff function $f(x) = (K-x)^+$ we therefore conclude

$$V(t) = Ke^{-r(T-t)}\Phi(-d_2) - S(t)\Phi(-d_1).$$

Partial differential equation (PDE) approach

Integration is not the only way to compute option prices. We will now consider an alternative approach relying on the solution of partial differential equations. To this end we focus on a European style contingent claim whose payoff $X = f(S(T))$ is a function of the stock at time T. Recall that the discounted fair option price price $\hat{V}(t) = V(t)/B(t)$ is a martingale relative to the unique EMM Q. It is not surprising and can in fact be shown that this price $\hat{V}(t)$ at time t is a deterministic function of the present stock price $S(t)$ and time t. In other words, it does not depend on the past stock prices which do not play a role for the expected payoff. We denote this function also as $\hat{v}$, i.e. we write the discounted option price as

$$\hat{V}(t) = \hat{v}(t, S(t))$$

with some function $\hat{v}$ on $[0,T]\times\mathbb{R}_+$. Similarly, we write

$$V(t) = v(t, S(t)),$$

which implies $v(t,x) = \hat{v}(t,x)e^{rt}$ because $V(t) = \hat{V}(t)B(t) = \hat{V}(t)e^{rt}$. Since

$$dS(t) = S(t)rdt + S(t)\sigma dW^Q(t),$$

Itō's formula yields

$$\begin{aligned} d\hat{V}(t) &= d\hat{v}(t,S(t)) \\ &= \left(\partial_1\hat{v}(t,S(t)) + \partial_2\hat{v}(t,S(t))rS(t) + \frac{1}{2}\partial_{22}\hat{v}(t,S(t))\sigma^2S^2(t)\right)dt \\ &\quad + \partial_2\hat{v}(t,S(t))\sigma S(t)dW^Q(t). \end{aligned} \tag{3.24}$$

Since $\hat{V}$ is a Q-martingale, it drift part has to vanish regardless of t and the present value of $S(t)$. This is only possible if $\hat{v}$ satisfies the *partial differential equation (PDE)*

$$\partial_1\hat{v}(t,x) + \partial_2\hat{v}(t,x)rx + \frac{1}{2}\partial_{22}\hat{v}(t,x)\sigma^2x^2 = 0 \tag{3.25}$$

or

$$\frac{\partial}{\partial t}\hat{v}(t,x) = -rx\frac{\partial}{\partial x}\hat{v}(t,x) - \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}\hat{v}(t,x)$$

in a possibly more familiar notation. Since $v(t,x) = \hat{v}(t,x)e^{rt}$, this leads to the *Black-Scholes PDE*

$$\frac{\partial}{\partial t}v(t,x) = rv(t,x) - rx\frac{\partial}{\partial x}v(t,x) - \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}v(t,x), \tag{3.26}$$

which is complemented by the final value

$$v(T,x) = f(x) \tag{3.27}$$

from $v(T,S(T)) = V(T) = f(S(T))$. In other words, European option prices can be computed by solving (3.26, 3.27). It is straightforward to verify that the function

$$\begin{aligned} v(t,x) &= x\Phi\left(\frac{\log\frac{x}{K} + r(T-t) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad - Ke^{-r(T-t)}\Phi\left(\frac{\log\frac{x}{K} + r(T-t) - \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

from (3.23) solves (3.26, 3.27) for $f(x) = (x-K)^+$. For other payoffs there may not be a closed-form solution. In Chapter 6 we will study how to solve the Black-Scholes PDE numerically.

We can now determine the replicating strategy as well. If $\varphi = (\varphi_0, \varphi_1)$ denotes a self-financing strategy, its discounted value satisfies

$$d\hat{V}_\varphi(t) = \varphi_1(t)d\hat{S}(t) = \varphi_1(t)\hat{S}(t)\sigma dW^Q(t). \tag{3.28}$$

If φ is the replicating strategy for the option, its value coincides with the option price. Comparing (3.24) and (3.28) yields

$$\varphi_1(t)\hat{S}(t)\sigma = \partial_2\hat{v}(t, S(t))\sigma S(t)$$

or

$$\begin{aligned}\varphi_1(t) &= \frac{S(t)}{\hat{S}(t)}\partial_2\hat{v}(t, S(t)) \\ &= e^{rt}\partial_2\hat{v}(t, S(t)) \\ &= \partial_2 v(t, S(t)),\end{aligned} \tag{3.29}$$

i.e. the partial derivative of the option price relative to the stock price yields the number of shares of stock in the replicating portfolio. This number is often called *delta* of the option in the literature. The remaining funds in the hedging portfolio are invested in bonds, which yields

$$\begin{aligned}\varphi_0(t) &= \frac{v(t, S(t)) - \varphi_1(t)S(t)}{B(t)} \\ &= \Big(v(t, S(t)) - \partial_2 v(t, S(t))S(t)\Big)e^{-rt}.\end{aligned} \tag{3.30}$$

For call options $f(x) = (x - K)^+$ we obtain

$$\begin{aligned}\varphi_1(t) &= \Phi\left(\frac{\log\frac{S(t)}{K} + r(T-t) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}\right), \\ \varphi_0(t) &= -Ke^{-rT}\Phi\left(\frac{\log\frac{S(t)}{K} + r(T-t) - \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}\right).\end{aligned}$$

Path-dependent options

We turn now to a few exotic options whose payoff depends not only on $S(T)$, but also on the minimum or maximum of the stock price in the interval $[0, T]$. We start with Barrier options whose payoff depends on the fact whether or not the stock price has reached a certain threshold before expiration. The payoff of e.g. a *down-and-out call* with strike K and *knock-out barrier* H equals

$$X = (S(T) - K)^+ 1_{\{\min_{t\le T} S(t) > H\}}. \tag{3.31}$$

Similarly, we could have *up-and-out, down-and-in, up-and-in* options depending on whether the threshold is an upper or lower one and whether the option loses (*knocks out*) or gains (*knocks in*) its value at the barrier. We consider here the contingent claim (3.31). Upper barriers are treated similarly. Knock-in options as e.g. a down-and-in call can be decomposed as

$$(S(T) - K)^+ 1_{\{\min_{t\le T} S(t) \le H\}} = (S(T) - K)^+ - (S(T) - K)^+ 1_{\{\min_{t\le T} S(t) > H\}}$$

and hence reduced to standard and down-and-out calls.

We take the PDE approach to compute the fair price and hedge for the down-and-out call. We assume $S(0) > H$ for the initial stock price because the option is worthless otherwise. It makes sense to assume that the option price $V(t)$ is a function of time t, the present stock price $S(t)$ and the running minimum $M(t) = \min_{s\le t} S(s)$. Of course, $V(t) = 0$ if $M(t) \le H$ because the expected payoff vanishes in this case. If $M(t) > H$, however, the precise value of $M(t)$ should not play a role because it does not affect the probability whether the option will be knocked out in its remaining life time $[t, T]$. These considerations, which can be made precise, lead to the ansatz

$$V(t) = v(t, S(t))1_{\{M(t)>H\}}$$

with some deterministic function v on $[0, T] \times \mathbb{R}_+$. For the discounted price process, this yields

$$\hat{V}(t) = \hat{v}(t, S(t))1_{\{M(t)>H\}}$$

with $\hat{v}(t, x) = e^{-rt}v(t, x)$. This process is a Q-martingale. An application of the martingale representation theorem yields that it can be written as a stochastic integral relative to W^Q, which implies that it is a continuous process. On the other hand, it seems to jump to 0 at the stopping time $\tau = \inf\{t \ge 0 : M(t) \le H\} = \inf\{t \ge 0 : S(t) = H\}$, i.e. at the first time that the stock price touches the barrier. This apparent contradiction is resolved by imposing the boundary condition $\hat{v}(t, H) = 0$ or equivalently $v(t, H) = 0$. Itō's formula yields

$$\begin{aligned} d\hat{v}(t, S(t)) &= \left(\partial_1\hat{v}(t, S(t)) + \partial_2\hat{v}(t, S(t))rS(t) + \frac{1}{2}\partial_{22}\hat{v}(t, S(t))\sigma^2S^2(t)\right)dt \\ &\quad + \partial_2\hat{v}(t, S(t))\sigma S(t)dW^Q(t) \end{aligned}$$

as in (3.24). Since the stopped process $\hat{v}(t \wedge \tau, S(t \wedge \tau)) = \hat{V}^\tau(t) = \hat{V}(t)$ is a Q-martingale, its drift part must vanish before τ. Therefore, we end up with the same PDE (3.26) as for non-path dependent claims. However, it must hold only beyond the barrier, i.e. for $x > H$. For the final value we have $v(T, x) = (x - K)^+$ for $x > H$. Altogether, we need to solve

$$\begin{aligned} \frac{\partial}{\partial t}v(t, x) &= rv(t, x) - rx\frac{\partial}{\partial x}v(t, x) - \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}v(t, x), \\ v(t, H) &= 0 \\ v(T, x) &= (x - K)^+ \end{aligned}$$

for $t \in [0,T]$, $x \geq H$. It is not difficult to verify that this PDE is solved by

$$\begin{aligned} v(t,x) &= x\left(\Phi\left(\frac{\log(\frac{x}{K})+(r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right)\right. \\ &\quad \left.-\left(\frac{H}{x}\right)^{1+\frac{2r}{\sigma^2}}\Phi\left(\frac{\log(\frac{H^2}{Kx})+(r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right)\right) \\ &\quad -e^{-r(T-t)}K\left(\Phi\left(\frac{\log(\frac{x}{K})+(r-\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right)\right. \\ &\quad \left.-\left(\frac{H}{x}\right)^{\frac{2r}{\sigma^2}-1}\Phi\left(\frac{\log(\frac{H^2}{Kx})+(r-\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right)\right). \end{aligned}$$

For the replicating strategy $\varphi = (\varphi_0, \varphi_1)$, the same argument as above yields that (3.30, 3.29) hold up to time τ. If resp. after the barrier has been reached, there is nothing to be hedged any more, i.e. $\varphi = 0$ after τ.

As a second example we consider a *lookback call* with payoff

$$X = S(T) - M(T),$$

where

$$M(t) = \min_{s \leq t} S(s)$$

denotes the running minimum as before. This contingent claim can be interpreted as a call whose strike equals the minimal stock price in the period under consideration. As for the down-and-out call it makes sense to assume that the fair option price is a function of time, the present stock price and the running minimum, i.e. of the form

$$V(t) = v(t, S(t), M(t)).$$

Let us write the discounted option price $\hat{V}(t) = e^{-rt}V(t)$ as $\hat{V}(t) = \hat{v}(t, \hat{S}(t), M(t))$ with

$$\hat{v}(t,x,m) = e^{-rt}v(t, xe^{rt}, m). \tag{3.32}$$

We do not know the Itō process decomposition of M. In fact, it is not even an Itō process because it is a decreasing process which cannot be written as an integral with respect to time. Therefore, we use the more general Itō formula (3.3) to get ahead. Note that all covariation processes involving t or $M(t)$ vanish because these are continuous processes of finite variation. Hence (3.3) yields

$$\begin{aligned} d\hat{V}(t) &= d\hat{v}(t, \hat{S}(t), M(t)) \\ &= \partial_1\hat{v}(t, \hat{S}(t), M(t))dt + \partial_2\hat{v}(t, \hat{S}(t), M(t))d\hat{S}(t) \\ &\quad + \partial_3\hat{v}(t, \hat{S}(t), M(t))dM(t) + \frac{1}{2}\partial_{22}\hat{v}(t, \hat{S}(t), M(t))d[\hat{S}, \hat{S}](t) \end{aligned}$$

Since $d\hat{S}(t) = \sigma\hat{S}(t)dW^Q(t)$, we obtain

$$\begin{aligned} d\hat{V}(t) &= d\hat{v}(t,\hat{S}(t),M(t)) \\ &= \partial_2\hat{v}(t,\hat{S}(t),M(t))d\hat{S}(t) + \partial_3\hat{v}(t,\hat{S}(t),M(t))dM(t) \\ &\quad + \left(\partial_1\hat{v}(t,\hat{S}(t),M(t)) + \frac{1}{2}\partial_{22}\hat{v}(t,\hat{S}(t),M(t))\sigma^2\hat{S}^2(t)\right)dt \end{aligned}$$

The $dM(t)$-term does not contribute anything unless the stock price $S(t)$ happens to coincide with its running minimum $M(t)$. The discounted stock price $\hat{S}(t)$ is a Q-martingale. In order for the sum $\hat{V}(t)$ to be a Q-martingale as well, the drift part must vanish, i.e. we conclude

$$\partial_1\hat{v}(t,x,m) + \frac{1}{2}\partial_{22}\hat{v}(t,x,m)\sigma^2x^2 = 0$$

on $xe^{rt} > m$, i.e. as long as the stock price exceeds its running minimum. By (3.32) and using $\frac{\partial}{\partial x}$-notation this yields

$$\frac{\partial}{\partial t}v(t,x,m) = rv(t,x,m) - rx\frac{\partial}{\partial x}v(t,x,m) - \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}v(t,x,m), \tag{3.33}$$

which is again the Black-Scholes PDE up to the fact that v depends on m as well. In order for $\hat{V}$ to be a Q-martingale, the $\partial_3\hat{v}(t,\hat{S}(t),M(t))dM(t)$-term must vanish as well because it belongs to the drift part in the Doob-Meyer decomposition. This means that the partial derivative must vanish whenever M moves, i.e. whenever the stock price coincides with its running minimum. In other words, we obtain the boundary condition $\partial_3\hat{v}(t,x,m) = 0$ for $x = me^{-rt}$ or

$$\frac{\partial}{\partial m}v(t,x,m) = 0 \quad \text{for } x = m. \tag{3.34}$$

The terminal condition is obviously

$$v(T,x,m) = x - m. \tag{3.35}$$

Hence we need to solve the PDE (3.33, 3.34, 3.35) for $t \in [0,T]$, $x > 0$, $m \in (0,x]$ in order to price the lookback call. It is once more straightforward to verify that this PDE is solved by

$$\begin{aligned} v(t,x,m) &= x\Phi\left(\frac{\log(\frac{x}{m}) + (r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad - e^{-r(T-t)}m\Phi\left(\frac{\log(\frac{x}{m}) + (r-\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad - \frac{x\sigma^2}{2r}\Phi\left(\frac{\log(\frac{m}{x}) - (r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad + e^{-r(T-t)}\frac{x\sigma^2}{2r}\left(\frac{m}{x}\right)^{\frac{2r}{\sigma^2}}\Phi\left(\frac{\log(\frac{m}{x}) - (r-\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right). \end{aligned}$$

For the perfect hedge $\varphi = (\varphi_0, \varphi_1)$ we can apply the earlier arguments and obtain once more

$$\begin{aligned}\varphi_1(t) &= \partial_2 v(t, S(t), M(t)),\\ \varphi_0(t) &= \Big(v(t, S(t), M(t)) - \partial_2 v(t, S(t), M(t))S(t)\Big)e^{-rt}.\end{aligned}$$

For *lookback puts* with payoff $X = \max_{t\in[0,T]} S(t) - S(T)$ one obtains similar results.

American options

We turn now to the valuation of American options in the Black-Scholes model. Suppose that the exercise process is of the form $X(t) = g(t, S(t))$ for some function g of time and stock price. For standard call and put options, the payoff function $g(t, x) = (x - K)^+$ resp. $g(t, x) = (K - x)^+$ does not actually depend on time. Let us denote the fair price process of the option as $V(t)$. As in the European case, it makes sense to assume that it is a deterministic function of time and the present stock price, i.e. of the form $V(t) = v(t, S(t))$. The corresponding discounted price can be written as $\hat{V}(t) = \hat{v}(t, S(t))$ with $\hat{v}(t, x) = v(t, x)e^{-rt}$. As in (3.24) Itō's formula yields

$$\begin{aligned}d\hat{V}(t) &= d\hat{v}(t, S(t))\\ &= \Big(\partial_1\hat{v}(t, S(t)) + \partial_2\hat{v}(t, S(t))rS(t) + \frac{1}{2}\partial_{22}\hat{v}(t, S(t))\sigma^2S^2(t)\Big)\,dt\\ &\quad + \partial_2\hat{v}(t, S(t))\sigma S(t)dW^Q(t).\end{aligned}$$

The American option price dominates the exercise price, i.e.

$$v(t, x) \geq g(t, x).$$

Moreover, general theory tells us that this process is a supermartingale relative to the unique EMM Q. In other words, its drift rate in the Itō process decomposition is less or equal to zero. This leads to the inequality

$$\partial_1\hat{v}(t, x) + \partial_2\hat{v}(t, x)rx + \frac{1}{2}\partial_{22}\hat{v}(t, x)\sigma^2x^2 \leq 0$$

instead of (3.25). Replacing $\hat{v}$ by v as in (3.26) and using $\frac{\partial}{\partial x}$-notation we obtain

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

Recall also that the Q-drift rate of an American option vanishes as long as the option price $V(t)$ is strictly above the exercise price $X(t) = g(t, S(t))$. This leads once more to the familiar Black-Scholes PDE

$$\frac{\partial}{\partial t}v(t, x) - rv(t, x) + rx\frac{\partial}{\partial x}v(t, x) + \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}v(t, x) = 0,$$

which, however, generally holds only for those (t,x) that satisfy $v(t,x) > g(t,x)$. The terminal value is

$$v(T,x) = g(T,x)$$

as in the European case. Putting everything together, we look for a function $v(t,x)$ on $[0,T]\times\mathbb{R}_+$ which satisfies

$$\begin{aligned}
\frac{\partial}{\partial t}v(t,x) - rv(t,x) + rx\frac{\partial}{\partial x}v(t,x) + \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}v(t,x) &\leq 0,\\
v(t,x) - g(t,x) &\geq 0,\\
\left(\frac{\partial}{\partial t}v(t,x) - rv(t,x) + rx\frac{\partial}{\partial x}v(t,x) + \frac{1}{2}\sigma^2x^2\frac{\partial^2}{\partial x^2}v(t,x)\right)(v(t,x) - g(t,x)) &= 0,\\
v(T,x) - g(T,x) &= 0.
\end{aligned} \tag{3.36}$$

This *linear complementarity problem*, *variational inequality*, or *free boundary value problem* does not generally allow for a closed form solution. We will consider its numerical solution in Chapter 6. However, recall from Section 3.3 that the prices of American and European calls coincide unless the interest rate r is strictly negative or dividends are paid on the stock.

Optimal exercise times and hedging strategies can also be deduced from the solution v to (3.36). The earliest optimal stopping time is the first time t where $v(t,x) = g(t,x)$ for $x = S(t)$. Alternatively, one may choose the latest optimal time which is the first time where the first inequality in (3.36) starts being strict. Let us also note that there is some *early-exercise curve* $S_f(t)$ for the American put. If the present stock price $S(t)$ falls below $S_f(t)$, it is time to exercise the option. As far as hedging is concerned, the results (3.29, 3.30) and arguments from the European case apply before the option is exercised.

3.5 Heston model

It has been repeatedly observed in the empirical literature that the assumption of constant volatility σ in the Black-Scholes model seems to be violated in real markets. In this section the parameter $\gamma = \sigma^2$ is replaced by a specific stochastic process. Contrary to the previous section we model the stock directly under the pricing measure from the EMM, i.e. we take the viewpoint of martingale modelling in the language of the previous course. We replace Equation (3.19) for the discounted stock by

$$d\hat{S}(t) = \hat{S}(t)\sigma(t)dW^Q(t) = \hat{S}(t)\sqrt{\gamma(t)}dW^Q(t),$$

with some Q-Wiener process W^Q. The stochastic process $\gamma(t) = \sigma^2(t)$ is assumed to satisfy the SDE

$$d\gamma(t) = (\kappa - \lambda\gamma(t))\,dt + \sqrt{\gamma(t)}\tilde{\sigma}d\widetilde{W}(t), \tag{3.37}$$

where $\kappa, \lambda, \widetilde{\sigma}, \gamma(0)$ are parameters and $\widetilde{W}$ denotes another Q-Wiener process which may or may not be correlated with the above Wiener process W^Q. For ease of exposition we focus on the uncorrelated case, which means that $(W^Q, \widetilde{W})$ is a Q-standard Brownian motion in $\mathbb{R}^2$. The bond is given as before by $B(t) = e^{rt}$ or

$$dB(t) = B(t)rdt, \quad B(0) = 1 \tag{3.38}$$

In undiscounted terms we have

$$dS(t) = S(t)rdt + S(t)\sqrt{\gamma(t)}dW^Q(t) \tag{3.39}$$

by $S(t) = B(t)\hat{S}(t)$ and integration by parts.

We want to determine fair option prices in this joint model (3.38, 3.39, 3.37) for the bond $B(t)$, the stock $S(t)$, and the squared volatility $\gamma(t)$. In principle, this can be done by integrating the payoff relative to the law of $S(T)$ as in (3.21). However, contrary to the Black-Scholes case we do not have a closed-form expression for the density of $S(T)$ or $\log(S(T))$. Therefore we consider the PDE approach here. By contrast to the Black-Scholes model we face an additional state variable, namely the current value of volatility. Therefore, it makes sense to assume that the fair price of a European option with payoff $f(S(T))$ is of the form

$$V(t) = v(t, S(t), \gamma(t))$$

with some deterministic function v. In discounted terms, we write

$$\hat{V}(t) = \hat{v}(t, S(t), \gamma(t))$$

with $\hat{v}(t, x, \gamma) = e^{-rt}v(t, x, \gamma)$. Itō's formula yields

$$\begin{aligned} d\hat{V}(t) &= d\hat{v}(t, S(t), \gamma(t)) \\ &= \Big(\partial_1\hat{v}(t, S(t), \gamma(t)) + \partial_2\hat{v}(t, S(t), \gamma(t))rS(t) + \partial_3\hat{v}(t, S(t), \gamma(t))\left(\kappa - \lambda\gamma(t)\right) \\ &\quad + \frac{1}{2}\partial_{22}\hat{v}(t, S(t), \gamma(t))\gamma(t)S^2(t) + \frac{1}{2}\partial_{33}\hat{v}(t, S(t), \gamma(t))\gamma(t)\widetilde{\sigma}^2\Big)dt \\ &\quad + \partial_2\hat{v}(t, S(t), \gamma(t))\sqrt{\gamma(t)}S(t)dW^Q(t) \\ &\quad + \partial_3\hat{v}(t, S(t), \gamma(t))\sqrt{\gamma(t)}\widetilde{\sigma}d\widetilde{W}(t) \end{aligned} \tag{3.40}$$

similar to (3.24). However, we need to apply the general Itō formula (3.3) because we formulated (3.8) only for one-dimensional X and W. As in the Black-Scholes model, the Q-martingale property leads to a SDE for $\hat{v}$, namely

$$\begin{aligned} \partial_1\hat{v}(t, x, \gamma) + \partial_2\hat{v}(t, x, \gamma)rx + \partial_3\hat{v}(t, x, \gamma)\left(\kappa - \lambda\gamma\right) & \\ + \frac{1}{2}\partial_{22}\hat{v}(t, x, \gamma)\gamma x^2 + \frac{1}{2}\partial_{33}\hat{v}(t, x, \gamma)\gamma\widetilde{\sigma}^2 &= 0 \end{aligned} \tag{3.41}$$

or

$$\begin{aligned}\frac{\partial}{\partial t}v(t,x,\gamma) &= rv(t,x,\gamma) - rx\frac{\partial}{\partial x}v(t,x,\gamma) - (\kappa-\lambda\gamma)\frac{\partial}{\partial\gamma}v(t,x,\gamma)\\ &\quad - \frac{1}{2}\gamma x^2\frac{\partial^2}{\partial x^2}v(t,x,\gamma) - \frac{1}{2}\gamma\tilde{\sigma}^2\frac{\partial^2}{\partial\gamma^2}v(t,x,\gamma),\end{aligned} \tag{3.42}$$

in undiscounted terms and $\frac{\partial}{\partial x}$-notation. For the final value we have

$$v(T,x,\gamma) = f(x) \tag{3.43}$$

as in the Black-Scholes case. The PDE (3.42, 3.43) generalises (3.26, 3.27) in the stochastic volatility model (3.38, 3.39, 3.37).

This can be extended to American options along the same lines as in the Black-Scholes model. Instead of (3.36) we obtain

$$\begin{aligned}\frac{\partial}{\partial t}v(t,x,\gamma) - rv(t,x,\gamma) + rx\frac{\partial}{\partial x}v(t,x,\gamma) + (\kappa-\lambda\gamma)\frac{\partial}{\partial\gamma}v(t,x,\gamma)&\\ + \frac{1}{2}\gamma x^2\frac{\partial^2}{\partial x^2}v(t,x,\gamma) + \frac{1}{2}\gamma\tilde{\sigma}^2\frac{\partial^2}{\partial\gamma^2}v(t,x,\gamma) &\le 0,\\ v(t,x,\gamma) - g(t,x) &\ge 0,\\ \left(\frac{\partial}{\partial t}v(t,x,\gamma) - rv(t,x,\gamma) + rx\frac{\partial}{\partial x}v(t,x,\gamma) + (\kappa-\lambda\gamma)\frac{\partial}{\partial\gamma}v(t,x,\gamma)\right.&\\ \left. + \frac{1}{2}\gamma x^2\frac{\partial^2}{\partial x^2}v(t,x,\gamma) + \frac{1}{2}\gamma\tilde{\sigma}^2\frac{\partial^2}{\partial\gamma^2}v(t,x,\gamma)\right)(v(t,x,\gamma)-g(t,x)) &= 0,\\ v(T,x,\gamma) - g(T,x) &= 0.\end{aligned} \tag{3.44}$$

Coming back to European options, we want to determine perfect hedging strategies. However, the Heston model is driven by a two-dimensional Brownian motion. Therefore it is incomplete according to the rule of thumb. As a way out we consider a liquid European call option as an additional hedging instrument. We denote its price process by $C(t) = c(t,S(t),\gamma(t))$. By (3.40) we have

$$d\hat{C}(t) = \partial_2\hat{c}(t,S(t),\gamma(t))\sqrt{\gamma(t)}S(t)dW^Q(t) + \partial_3\hat{c}(t,S(t),\gamma(t))\tilde{\sigma}\sqrt{\gamma(t)}d\widetilde{W}(t)$$

for the discounted price $\hat{C}(t) = \hat{c}(t,S(t),\gamma(t)) = c(t,S(t),\gamma(t))e^{-rt}$. The function c is just the solution to (3.42, 3.43) for $f(x) = (x-K)^+$.

Now, we consider an additional option as above with discounted fair price $\hat{V}(t) = \hat{v}(t,S(t),\gamma(t))$. By (3.40) it moves according to

$$d\hat{V}(t) = \partial_2\hat{v}(t,S(t),\gamma(t))\sqrt{\gamma(t)}S(t)dW^Q(t) + \partial_3\hat{v}(t,S(t),\gamma(t))\tilde{\sigma}\sqrt{\gamma(t)}d\widetilde{W}(t). \tag{3.45}$$

The discounted value of a self-financing strategy $\varphi = (\varphi_0, \varphi_1, \varphi_2)$ consisting of φ_0 bonds, φ_1 shares of stock, and φ_2 liquid call options satisfies

$$\begin{aligned} d\hat{V}_\varphi(t) &= \varphi_1(t)d\hat{S}(t) + \varphi_2(t)d\hat{C}(t) \\ &= \left(\varphi_1(t)\sqrt{\gamma(t)}\hat{S}(t) + \varphi_2(t)\partial_2\hat{c}(t, S(t), \gamma(t))\sqrt{\gamma(t)}S(t)\right) dW^Q(t) \\ &\quad + \varphi_2(t)\partial_3\hat{c}(t, S(t), \gamma(t))\tilde{\sigma}\sqrt{\gamma(t)}d\widetilde{W}(t) \end{aligned} \tag{3.46}$$

φ is a perfect hedge for the option with price process V if the right-hand sides of (3.45) and (3.46) match. This can be done by choosing

$$\begin{aligned} \varphi_2(t) &= \frac{\partial_3\hat{v}(t, S(t), \gamma(t))}{\partial_3\hat{c}(t, S(t), \gamma(t))}, \\ \varphi_1(t) &= \partial_2\hat{v}(t, S(t), \gamma(t))e^{rt} - \varphi_2(t)\partial_2\hat{c}(t, S(t), \gamma(t))e^{rt}. \end{aligned}$$

In undiscounted terms, this yields

$$\varphi_2(t) = \frac{\partial_3 v(t, S(t), \gamma(t))}{\partial_3 c(t, S(t), \gamma(t))}, \tag{3.47}$$

$$\varphi_1(t) = \partial_2 v(t, S(t), \gamma(t)) - \varphi_2(t)\partial_2 c(t, S(t), \gamma(t)). \tag{3.48}$$

φ_0 can be obtained by the self-financing condition as usual:

$$\begin{aligned} \varphi_0(t) &= \frac{V(t) - \varphi_1(t)S(t) - \varphi_2(t)C(t)}{B(t)} \\ &= \Big(v(t, S(t), \gamma(t)) - \varphi_1(t)S(t) - \varphi_2(t)c(t, S(t), \gamma(t))\Big)e^{-rt}. \end{aligned} \tag{3.49}$$

Consequently, we can compute hedging strategies if we dispose of numerical solutions for the option prices $c(t, x, \gamma), v(t, x, \gamma)$ and their derivatives. The latter are called *sensitivities* in the literature.

Chapter 4

Integral transforms

In this chapter, we discuss another method for finding expectations. The main tool are integral transformations as known from other fields of mathematics and statistics.

4.1 Option prices via Laplace transform

Closed-form expressions for option prices are available only in rare cases. Often, the probability densities of $S(T)$ or $\log S(T)$ are unknown so that computation by numerical integration of the payoff is not obvious either. We consider here an approach to pricing European non-path dependent options which is based on characteristic functions rather than densities.

We consider a model with bond and stock of the form

$$\begin{aligned} B(t) &= e^{rt}, \\ S(t) &= e^{X(t)} \end{aligned}$$

with some stochastic process X as e.g. Brownian motion. If the payoff of the claim under consideration is of the form $f(X(T))$, the initial fair option price equals

$$\begin{aligned} V(0) &= B(0)E_Q\left(\frac{f(X(T))}{B(T)}\right) \\ &= e^{-rT}E_Q\Big(f(X(T))\Big). \end{aligned} \tag{4.1}$$

For call resp. put options we have $f(x) = (e^x - K)^+$ resp. $f(x) = (K - e^x)^+$ because the payoff is written as a function of the logarithmic stock price. For very simple payoffs, the expectation in (4.1) can in fact be calculated explicitly in many models, namely for $f(x) = e^{zx}$ with some constant z. For the sequel, we even need to take complex-valued z into account, i.e. we consider payoffs of the form

$$f(x) = e^{zx} = e^{\mathrm{Re}(z)x}(\cos(\mathrm{Im}(z)x) + i\sin(\mathrm{Im}(z)x)),$$

where $z \in \mathbb{C}$ and $i = \sqrt{-1}$. Of course, complex-valued options do not make sense from an economic point of view but this fortunately does not affect their mathematical treatment. The corresponding "option price" equals

$$V(0) = e^{-rT} E_Q(e^{zX(T)}) = e^{-rT} \chi(-iz)$$

with

$$\chi(u) = E_Q(e^{iuX(T)}).$$

$\chi(u)$ is called (extended) *characteristic function* of $X(T)$ in u and it is known in closed form for many processes X. We assume it to be given for the time being and turn to particular examples later.

The next step and key idea is to write an arbitrary, more complicated payoff (e.g. a European call) in the form

$$f(x) = \int \varrho(z) e^{zx} dz \tag{4.2}$$

with some function $\varrho(z)$. Such a representation can be viewed as a generalized linear combination of "simple" payoffs e^{zx}. By linearity of the pricing rule (4.1) in the payoff, we end up with

$$\begin{aligned} V(0) &= e^{-rT} \int \varrho(z) E_Q(e^{zX(T)}) dz \\ &= e^{-rT} \int \varrho(z) \chi(-iz) dz, \end{aligned} \tag{4.3}$$

which can be evaluated by numerical integration if both functions ϱ and χ are known.

But how can we determine an integral representation as in (4.2)? To this end, consider the *bilateral Laplace transform*

$$\widetilde{f}(z) = \int_{-\infty}^{\infty} f(x) e^{-zx} dx$$

of f. This integral may not be defined for all $z \in \mathbb{C}$. But if $\widetilde{f}$ exists for some $z = R \in \mathbb{R}$ and satisfies a certain integrability condition, then f can be recovered from $\widetilde{f}$ by the following *inversion formula*:

$$\begin{aligned} f(x) &= \frac{1}{2\pi i} \int_{R-i\infty}^{R+i\infty} \widetilde{f}(z) e^{zx} dz \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \widetilde{f}(R+iu) e^{(R+iu)x} du. \end{aligned}$$

Hence we have found an integral representation as in (4.2) and the correponding pricing formula (4.3) reads as

$$\begin{aligned} V(0) &= \frac{e^{-rT}}{2\pi i} \int_{R-i\infty}^{R+i\infty} \widetilde{f}(z) \chi(-iz) dz \\ &= \frac{e^{-rT}}{2\pi} \int_{-\infty}^{\infty} \widetilde{f}(R+iu) \chi(u - iR) du. \end{aligned}$$

For slight simplification observe that $\widetilde{f}(R-iu) = \overline{\widetilde{f}(R+iu)}$ and $\chi(-u-iR) = \overline{\chi(u-iR)}$. This yields

$$\begin{aligned} V(0) &= \frac{e^{-rT}}{2\pi}\int_0^\infty \Big(\widetilde{f}(R+iu)\chi(u-iR) + \widetilde{f}(R-iu)\chi(-u-iR)\Big)du \\ &= \frac{e^{-rT}}{2\pi}\int_0^\infty \Big(\widetilde{f}(R+iu)\chi(u-iR) + \overline{\widetilde{f}(R+iu)\chi(u-iR)}\Big)du \\ &= \frac{e^{-rT}}{\pi}\int_0^\infty \mathrm{Re}\Big(\widetilde{f}(R+iu)\chi(u-iR)\Big)du. \end{aligned} \tag{4.4}$$

Similarly, we obtain the fair price at time t if we replace the lifetime T by $T-t$ and the characteristic function of $X(T)$ by the characteristic function of $X(T)$ given $\mathscr{F}_t$, i.e. we have

$$V(t) = \frac{e^{-r(T-t)}}{\pi}\int_0^\infty \mathrm{Re}\Big(\widetilde{f}(R+iu)\chi_t(u-iR)\Big)du, \tag{4.5}$$

where χ_t denotes the conditional characteristic function

$$\chi_t(u) = E_Q(e^{iuX(T)}|\mathscr{F}_t).$$

Let us now determine the Laplace transform of the call payoff function $f(x) = (e^x - K)^+$:

$$\begin{aligned} \widetilde{f}(z) &= \int_{-\infty}^\infty (e^x-K)^+ e^{-zx}dx \\ &= \int_{\log K}^\infty (e^{(1-z)x} - Ke^{-zx})dx \\ &= \Big[\frac{1}{1-z}e^{(1-z)x} + \frac{K}{z}e^{-zx}\Big]_{x=\log K}^{x=\infty} \\ &= -\frac{K^{1-z}}{1-z} - \frac{K^{1-z}}{z} \\ &= \frac{K^{1-z}}{z(z-1)}. \end{aligned} \tag{4.6}$$

This holds for all z with $R = \mathrm{Re}(z) > 1$ because otherwise $e^{(1-z)x}$ does not tend to zero for $x \to \infty$. Interestingly, the same formula is obtained for the put $f(x) = (K-e^x)^+$ but now for z with $R = \mathrm{Re}(z) < 0$.

We still need to determine the characteristic function χ in order to use formulas (4.4) resp. (4.5). Let us start with the Black-Scholes model (3.17, 3.18). As noted above,

$$e^{-rT}\chi(u) = e^{-rT}E_Q\Big(e^{iuX(T)}\Big)$$

is just the fair price of an option with payoff $e^{iuX(T)} = e^{iu\log S(T)}$. From Section 3.4 we know that it can be written as $v(0,S(0))$ with some function v satisfying the PDE (3.26). We make the ansatz that the solution is of the form $v(t,x) = e^{\alpha-\beta t+iu\log x}$ with some unknown parameters α,β. Here, x denotes the stock price as opposed to (4.2) etc. where it represented the logarithmic

stock price. Inserting this function shows that it solves the PDE for $\alpha = \beta T$, $\beta = (iu-1)r - (iu+u^2)\sigma^2/2$, i.e. we conclude

$$\chi(u) = \exp\left(iu(\log S(0) + rT) - (iu+u^2)\frac{\sigma^2}{2}T\right).$$

For χ_t we obtain accordingly

$$\chi_t(u) = \exp\left(iu(\log S(t) + r(T-t)) - (iu+u^2)\frac{\sigma^2}{2}(T-t)\right)$$

and hence

$$\begin{aligned} V(t) &= v(t, S(t)) \\ &= \frac{e^{-r(T-t)}}{\pi}\int_0^\infty \mathrm{Re}\Big(\widetilde{f}(R+iu)\eta_t(u-iR, S(t))\Big)du \end{aligned} \tag{4.7}$$

with

$$\eta_t(u,x) := \exp\left(iu(\log x + r(T-t)) - (iu+u^2)\frac{\sigma^2}{2}(T-t)\right).$$

In the Heston model of Section 3.5 we must consider the slightly more complicated PDE (3.41) instead of (3.26). We work directly with the martingale $E_Q(e^{iuX(T)}|\mathscr{F}_t)$ as a process in t and make the ansatz that it can be written as a function

$$\hat{v}(t,x,\gamma) = e^{\psi_0(t)+\psi_1(t)\log x+\psi_2(t)\gamma}$$

with unknown functions ψ_0, ψ_1, ψ_2. Insertion in the PDE leads to the system of *ordinary differential equations (ODE's)*

$$\begin{aligned} \psi_0'(t) &= -\kappa\psi_2(t) - iur, & \psi_0(T) &= 0, \\ \psi_1'(t) &= 0, & \psi_1(T) &= iu, \\ \psi_2'(t) &= -\tfrac{\widetilde{\sigma}}{2}\psi_2(t)^2 + \lambda\psi_2(t) + \tfrac{u(u+i)}{2}, & \psi_2(T) &= 0 \end{aligned}$$

for ψ_0, ψ_1, ψ_2. The equation for ψ_2 is an ODE of *Riccati* type. ψ_0 is obtained by integration from ψ_2. The solution to the system is of the form

$$\begin{aligned} \exp(\psi_0(t)) &= e^{iur(T-t)}\left(\frac{e^{\lambda(T-t)/2}}{\cosh(d(u)(T-t)/2) + \lambda\sinh(d(u)(T-t)/2)/d(u)}\right)^{2\kappa/\widetilde{\sigma}^2}, \\ \psi_1(t) &= iu, \\ \psi_2(t) &= -\frac{(iu+u^2)\sinh(d(u)(T-t)/2)/d(u)}{\cosh(d(u)(T-t)/2) + \lambda\sinh(d(u)(T-t)/2)/d(u)} \end{aligned}$$

with

$$d(u) = \sqrt{\lambda^2 + \widetilde{\sigma}^2(iu+u^2)}.$$

These formulas hold for $\mathrm{Re}(u)$ in a neighbourhood of 0 whose size depends on the parameters. Hence we have

$$\begin{aligned}\chi_t(u) &= e^{iu(\log S(t)+r(T-t))}\left(\frac{e^{\lambda(T-t)/2}}{\cosh(d(u)(T-t)/2)+\lambda\sinh(d(u)(T-t)/2)/d(u)}\right)^{2\kappa/\widetilde{\sigma}^2} \\ &\quad\times\exp\left(-\gamma(t)\frac{(iu+u^2)\sinh(d(u)(T-t)/2)/d(u)}{\cosh(d(u)(T-t)/2)+\lambda\sinh(d(u)(T-t)/2)/d(u)}\right) \qquad (4.8)\end{aligned}$$

in the Heston model (3.38, 3.39, 3.37), which leads to

$$\begin{aligned}V(t) &= v(t,S(t),\gamma(t)) \\ &= \frac{e^{-r(T-t)}}{\pi}\int_0^\infty \mathrm{Re}\Big(\widetilde{f}(R+iu)\eta_t(u-iR,S(t),\gamma(t))\Big)du \qquad (4.9)\end{aligned}$$

with

$$\begin{aligned}\eta_t(u,x,\gamma) &= e^{iu(x+r(T-t))}\left(\frac{e^{\lambda(T-t)/2}}{\cosh(d(u)(T-t)/2)+\lambda\sinh(d(u)(T-t)/2)/d(u)}\right)^{2\kappa/\widetilde{\sigma}^2} \\ &\quad\times\exp\left(-\gamma\frac{(iu+u^2)\sinh(d(u)(T-t)/2)/d(u)}{\cosh(d(u)(T-t)/2)+\lambda\sinh(d(u)(T-t)/2)/d(u)}\right). \qquad (4.10)\end{aligned}$$

Therefore, option prices can now be determined from (4.9) by numerical integration. The constant R must be chosen such that the exponential moment $E(e^{RX(T)})$ is finite, which is true for arbitrary R in the Black-Scholes model but not e.g. in the Heston case. Moreover, the integrand of (4.4) tends to vary a lot near the poles of $\widetilde{f}$, i.e. near 1 and 0 in the case (4.6) of calls resp. puts. This complicates numerical integration. The question how to pick R optimally is considered in detail in [Lee04]. As a reasonable rule of thumb one may want to stay away from both the poles of $\widetilde{f}$ and the maximal resp. minimal R with $E(e^{RX(T)})<\infty$.

4.2 Option pricing by fast Fourier transform

For calibration purposes one often needs the prices of many calls resp. puts simultaneously. Let us denote by $V_\kappa(0)$ the price of a call with strike $K=e^\kappa$, which is of the form (4.5) with $\widetilde{f}=\widetilde{f}_\kappa$ given by

$$\widetilde{f}_\kappa(z)=\frac{e^{(1-z)\kappa}}{z(z-1)}. \qquad (4.11)$$

Since $\widetilde{f}_\kappa(z)=e^{(1-z)\kappa}\widetilde{f}_0(z)=e^{(1-R-iu)\kappa}\widetilde{f}_0(z)$, (4.5) can be rewritten as

$$V_\kappa(t) = \frac{e^{-r(T-t)+(1-R)\kappa}}{\pi}\int_0^\infty \mathrm{Re}\Big(\widetilde{f}_0(R+iu)\chi_t(u-iR)e^{-iu\kappa}\Big)du.$$

This can be interpreted in terms of a Fourier transform. The *Fourier transform* of a function g is defined as

$$\hat{g}(\kappa)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty g(u)e^{-iu\kappa}du$$

or a closely related expression. If $g(-u) = \overline{g(u)}$, this equals

$$\hat{g}(\kappa) = \sqrt{\frac{2}{\pi}} \int_0^\infty \mathrm{Re}\Big(g(u)e^{-iu\kappa}\Big)du. \tag{4.12}$$

Therefore,

$$V_\kappa(t) = \frac{e^{-r(T-t)+(1-R)\kappa}}{\sqrt{2\pi}} \hat{g}(\kappa)$$

with

$$g(u) = \widetilde{f}_0(R+iu)\chi_t(u-iR). \tag{4.13}$$

Efficient numerical algorithms have been developed which compute $\hat{g}(\kappa)$ simultaneously for many values κ. These, however, refer to the *discrete* rather than continuous Fourier transform. In order to use them, we must approximate the integral in (4.12) by a finite sum. To this end, we approximate the improper integral by a proper one:

$$\int_0^\infty \mathrm{Re}\Big(g(u)e^{-iu\kappa}\Big)du \approx \int_0^M \mathrm{Re}\Big(g(u)e^{-iu\kappa}\Big)du$$

with some sufficiently large upper limit M. The error of this approximation depends on how quickly the integrand and in particular $\chi_t(u-iR)$ decays to 0 for $u \to \infty$. This in turns depends on the smoothness of the probability density of $X(T)$: the smoother the density, the faster the decay of its Fourier transform. In a second step, we approximate the integral by a finite sum. We use the midpoint rule and obtain

$$\int_0^M \mathrm{Re}\Big(g(u)e^{-iu\kappa}\Big)du \approx \sum_{n=1}^N \mathrm{Re}\Big(g((n-\tfrac{1}{2})\Delta)e^{-i(n-\frac{1}{2})\Delta\kappa}\Big)\Delta \tag{4.14}$$

for some sufficiently large N and corresponding mesh size $\Delta = M/N$. The quality of this approximation depends on the smoothness of g and in particular of χ_t. This is determined by the existence of high moments of $X(T)$ or, put differently, by the thickness of the tails of the law of $X(T)$.

If we set

$$\kappa_m := \kappa_1 + (m-1)\frac{2\pi}{M}, \quad m = 1, \dots, N \tag{4.15}$$

with arbitrary κ_1, the right-hand side of (4.14) equals

$$\mathrm{Re}\left(\sum_{n=1}^N e^{-2\pi i\frac{(m-1)(n-1)}{N}} g((n-\tfrac{1}{2})\Delta)e^{-i(n-1)\Delta\kappa_1}\Delta e^{-\frac{i}{2}\Delta\kappa_m}\right)$$

for $\kappa = \kappa_m$. The *discrete Fourier transfom (DFT)* of a vector $x = (x_1, \dots, x_N)$ is often defined as $\hat{x} = (\hat{x}_1, \dots, \hat{x}_N)$ with

$$\hat{x}_m = \sum_{n=1}^N e^{-2\pi i\frac{(m-1)(n-1)}{N}} x_n, \quad m = 1, \dots, N. \tag{4.16}$$

If we set

$$x_n := g((n-\tfrac{1}{2})\Delta)\Delta e^{-i(n-1)\Delta\kappa_1}, \tag{4.17}$$

the right-hand side of (4.14) equals $\mathrm{Re}(\hat{x}_m e^{-\frac{i}{2}\Delta\kappa_m})$ for $\kappa = \kappa_m$. Consequently,

$$V_{\kappa_m}(t) \approx \frac{\exp(-r(T-t)+(1-R)\kappa_m)}{\pi}\mathrm{Re}\Big(\hat{x}_m e^{-\frac{i}{2}\Delta\kappa_m}\Big), \quad m = 1,\dots,N. \tag{4.18}$$

A good approximation is obtained if $M = N\Delta$ is large and Δ is small. For error bounds we refer again to [Lee04]. The discrete Fourier transform allows for an efficient implementation if N is a power of two, i.e. of the form 2^p for some integer p. The corresponding *fast Fourier transform (FFT)* algorithm computes $\hat{x} = (\hat{x}_1,\dots,\hat{x}_N)$ simultaneously in only $O(N\log N)$ steps, rather than in $O(N^2)$ steps as one would expect. If option prices are needed for some κ with $\kappa_{m-1} < \kappa < \kappa_m$, the simplest approach is linear interpolation, i.e. the approximation

$$V_\kappa(t) \approx \frac{(\kappa_m-\kappa)V_{\kappa_{m-1}}(t) + (\kappa-\kappa_{m-1})V_{\kappa_m}(t)}{\kappa_m - \kappa_{m-1}}.$$

4.3 Hedging strategies and calibration

From Sections 3.4 and 3.5 we know that hedging strategies can often be expressed in terms of derivatives of the pricing function. Let us consider the Black-Scholes case first. We know from Section 3.4 and (4.7) that the fair option price can be written as a function of t and the present stock price $S(t)$. By (3.29, 3.30) the perfect hedge $\varphi = (\varphi_0,\varphi_1)$ of the option is obtained by calculating the derivative of (4.7) with respect to the stock price. Under weak regularity conditions, the derivative can be moved inside the integral and the real part, which yields

$$\varphi_1(t) = \frac{e^{-r(T-t)}}{\pi}\int_0^\infty \mathrm{Re}\Big(\tilde{f}(R+iu)\delta_t(u-iR,S(t))\Big)du,$$

with

$$\begin{aligned}\delta_t(u,x) &= \frac{\partial}{\partial x}\eta_t(u,x)\\ &= iu\exp\left((iu-1)\log x + iur(T-t) - (iu+u^2)\frac{\sigma^2}{2}(T-t)\right).\end{aligned}$$

In other words, replacing η_t by δ_t in the above approach leads to hedging strategies instead of option prices.

In the slightly more involved Heston model the situation is similar. According to Section 3.5 and (4.9), the option price is a function of t, $S(t)$, and $\gamma(t)$. We need the partial derivatives with respect to both state variables in order to compute hedging strategies $\varphi = (\varphi_0,\varphi_1,\varphi_2)$ as in (3.47 – 3.49). Again, differentiation can be moved inside the integral sign and the real part.

As in the Black-Scholes model, hedging strategies are obtained by replacing η_t in (4.9) by the corresponding derivatives of (4.10). We leave their calculation to the reader.

Let us come back to martingale modelling as it was mentioned in the introduction and Section 3.5. We consider the Heston model as an example. Its parameters $\kappa, \lambda, \widetilde{\sigma}$ and also the initial squared volatility $\gamma(0)$ must be determined in practice. This is typically done by calibration to liquid options. Suppose that the prices $\pi_1, \ldots, \pi_n$ of n European call resp. put options are observed at time 0. Their model prices $V_1(0), \ldots, V_n(0)$ depend on the above unknown parameters. Calibration means choosing $\kappa, \lambda, \widetilde{\sigma}, \gamma(0)$ such that model and observed prices coincide, i.e.

$$\pi_i = V_i(0), \quad i = 1, \ldots, n.$$

Since the number of options n typically exceed the number of parameters (here: 4), these equations probably cannot be solved exactly. A way out is to minimize some form of distance as e.g.

$$\sum_{i=1}^{n} (\pi_i - V_i(0))^2 \tag{4.19}$$

as a function of the unknown parameters. This distance can be modified in many ways. The options could e.g. be weighted by their liquidity or bid-ask spread. Or the distance could be modified such that the contribution of an option vanishes as long as its model price lies inside that observed bid-ask-bounds.

Many algorithms have been developed in order to minimise expressions as in (4.19). They typically proceed by some form of successive approximation. In any case, one needs a fast implementation for the computation of model prices $V_i(0)$ because this step is repeated many times in the optimisation procedure for various parameter vectors. Some algorithms rely on knowing the derivative of the objective function (4.19) with respect to the parameters. The integral transform method is well suited to provide such information. As in the above discussion on hedging, the dependence on parameters is hidden in the characteristic function, cf. (4.10). The relevant derivatives of $V_i(0)$ can be calculated easily by differentiation under the integral sign.

4.4 Algorithms and discussion

As an example we provide an algorithm which computes a range of European call resp. put prices in the Heston model.

Input Model parameters: $r, \kappa, \lambda, \widetilde{\sigma}, S(0), \gamma(0)$

Option parameters: T, call/put

Further parameters: $R > 1$ for calls, $R < 0$ for puts, N, M, κ_1

Compute Define function $\widetilde{f}_0$ as in (4.11).

Define function $\chi = \chi_0$ as in (4.8).

Define function g as in (4.13).

Set $\Delta = M/N$.

Define vector $x = (x_1, \dots, x_N)$ as in (4.17).

Compute the discrete Fourier transform (4.16) of x.

Compute κ_m as in (4.15) for $m = 1, \dots, N$.

Compute $K_m = \exp(\kappa_m)$ for $m = 1, \dots, N$.

Compute $V_m := V_{\kappa_m}(0)$ as in (4.18) for $m = 1, \dots, N$.

Output (K_m, V_m) for $m = 1, \dots, N$.

The output of this algorithm are N pairs of a strike and the respective call/put price at time 0 which corresponds to this strike.

Let us also consider an algorithm which calibrates parameters to observed option prices. We consider a model class with parameters $\vartheta_1, \dots, \vartheta_p$, e.g. $\kappa, \lambda, \widetilde{\sigma}, \gamma(0)$ in the Heston model. Moreover, we suppose that we have at our disposal a numerical algorithm for the computation of option prices in this model. This algorithm may depend on parameters $\eta_1, \dots, \eta_m$ e.g. R for (4.5) is used or R, N, M, κ_1 for the above FFT-based algorithm.

Input Fixed model parameters: $r, S(0)$

Option data: T_i, K_i, call/put, π_i for $i = 1, \dots, n$

Parameters for numerical algorithms: $\eta_1, \dots, \eta_m$

Starting values: $\vartheta_{1,0}, \dots, \vartheta_{p,0}$.

Compute Define the option pricing function $V_0(r, S(0), \vartheta_1, \dots, \vartheta_p, T, K, \text{call/put})$, e.g. using (4.4). It may depend on parameters $(\eta_1, \dots, \eta_m)$ needed for the algorithm.

Define the loss function $\ell(\vartheta_1, \dots, \vartheta_p)$ as in (4.19).

Minimize $\ell(\vartheta_1, \dots, \vartheta_p)$ as a function of $\vartheta_1, \dots, \vartheta_p$. The minimisation procedure may require starting values $\vartheta_{1,0}, \dots, \vartheta_{p,0}$. It may also require the derivatives of ℓ relative to $\vartheta_1, \dots, \vartheta_p$, in which case these should be provided as well.

Output Calibrated parameters $\vartheta_1, \dots, \vartheta_p$

The output of this algorithm are model parameters $\vartheta_1, \ldots, \vartheta_p$ such that theoretical and observed option prices match as closely as possible.

Option pricing by integral transforms is by now a standard approach in both theory and practice. It is fast, relatively simple, and flexible. It can be applied to models with jumps, stochastic volatility, many state variables etc. as long as the characteristic function of the underlying is known in closed form. However, the approach cannot be applied immediately to American options. A way out is to approximate the latter by *Bermudan options*, which can be exercised only at finitely many instances. These can be priced by a recursive procedure similar as in Section 1.4 for the Cox-Ross-Rubinstein model. Each step in the recursion corresponds to the computation of a European option which can be done as in the previous sections. For details we refer to [LFBO07]. In any case, calculating American options using integral transforms requires higher effort than in the European case, both as far as implementation and computing time is concerned. Treating path-dependent options is not obvious either but it has been considered in particular cases.

4.5 A few facts on complex numbers

Complex numbers are of the form $z = a + ib$ with $a, b \in \mathbb{R}$, where the imaginary number $i = \sqrt{-1}$ denotes one of the two square roots of -1, i.e. $i^2 = -1$. $\mathrm{Re}(z) = a$ resp. $\mathrm{Im}(z) = b$ are called *real part* and *imaginary part* of z. Calculations with complex numbers follow largely the same rules as those with real numbers. The exponential function for complex numbers is defined as

$$e^z = e^a(\cos b + i \sin b)$$

for $z = a + ib$. The *complex conjugate* of $z = a + ib$ is $\overline{z} = a - ib$. We have $z + \overline{z} = 2\mathrm{Re}(z)$, $\overline{z + z'} = \overline{z} + \overline{z}'$, $\overline{zz'} = \overline{z}\overline{z'}$, $\overline{e^z} = e^{\overline{z}}$ Moreover, $\overline{z} = z$ if z is real and $\overline{z} = -z$ if z is purely imaginary, i.e. if $z = ib$ with real b. The analysis of complex functions obeys similar rules as in the real-valued case. At this point we recall the definition of the hyperbolic sine and cosine functions:

$$\begin{aligned} \sinh(z) &= \frac{e^z - e^{-z}}{2}, \\ \cosh(z) &= \frac{e^z + e^{-z}}{2}. \end{aligned}$$

Chapter 5

More on Monte Carlo techniques

Based on the extended knowledge from Chapter 3, we now revisit Monte Carlo methods for financial models.

5.1 Simulation of sensitivities

Hedging strategies do often involve derivatives of the option price relative to state variables as e.g. the current stock price; see the previous chapter. Put in a general context, we are facing the problem how to compute the derivative $z'(\vartheta)$ of an expectation $z(\vartheta) = E(Z(\vartheta))$ which depends on a parameter ϑ.

As an example, consider once more Equation (2.10) for the fair price of a European call, here written as $z(s) = E(Z(s))$, where

$$Z(s) = e^{-rT}\left(s\exp\left((r-\sigma^2/2)T + \sigma\sqrt{T}X\right) - K\right)^+,$$

X denotes some standard normal random variable, and s the initial stock price. According to (3.29) the derivative $z'(s)$ stands for the number of shares of the stock in the hedging portfolio.

Finite difference method For a sufficiently smooth function z we have the Taylor expansion

$$z(\vartheta + h/2) = z(\vartheta) + z'(\vartheta)h/2 + z''(\vartheta)h^2/8 + z'''(\vartheta)h^3/48 + O(h^4),$$

where $O(h^4)$ stands for an expression such that $O(h^4)/h^4$ is bounded in a neighbourhood of $h = 0$. Similarly,

$$z(\vartheta - h/2) = z(\vartheta) - z'(\vartheta)h/2 + z''(\vartheta)h^2/8 - z'''(\vartheta)h^3/48 + O(h^4),$$

which yields

$$\begin{aligned} z'(\vartheta) &= \frac{z(\vartheta+h/2)-z(\vartheta-h/2)}{h}+O(h^2) \\ &= E\left(\frac{Z(\vartheta+h/2)-Z(\vartheta-h/2)}{h}\right)+O(h^2) \end{aligned}$$

The quantities $E(Z(\vartheta+h/2))$ and $E(Z(\vartheta-h/2))$ can be computed by Monte Carlo simulation as usual. Choosing a small h reduces the bias due to the $O(h^2)$ term. However, it increases the variance of the estimator, at least if $E(Z(\vartheta+h/2))$ and $E(Z(\vartheta-h/2))$ are computed using separate simulations.

Let us analyse how large h should be chosen in order to minimise the mean square error $E((\hat{z}'(\vartheta)-z'(\vartheta))^2)$ of the Monte Carlo estimate

$$\hat{z}'(\vartheta) := \frac{1}{N}\sum_{n=1}^{N}\frac{Z_n(\vartheta+h/2)-Z_n(\vartheta-h/2)}{h}$$

of $z'(\vartheta)$. For the time being let us assume that all $2N$ random variables $Z_n(\vartheta\pm h/2)$ are independent and distributed as $Z(\vartheta\pm h/2)$. If we assume that $\mathrm{Var}(Z_n(\vartheta\pm h/2))\approx\mathrm{Var}(Z(\vartheta))$ for sufficiently small h, we have

$$\mathrm{Var}(\hat{z}'(\vartheta))\approx\frac{1}{N^2}N\frac{1}{h^2}2\mathrm{Var}(Z(\vartheta))=\frac{2}{Nh^2}\mathrm{Var}(Z(\vartheta)).$$

The Taylor expansions from above yield

$$\frac{z(\vartheta+h/2)-z(\vartheta-h/2)}{h} = z'(\vartheta)+\frac{h^2}{24}z'''(\vartheta)+O(h^4).$$

This means that the bias equals approximately

$$E(\hat{z}'(\vartheta))-z'(\vartheta)\approx\frac{h^2}{24}z'''(\vartheta)$$

for small h. The mean squared error is the sum of squared bias and variance:

$$\begin{aligned} E((\hat{z}'(\vartheta)-z'(\vartheta))^2) &= (E(\hat{z}'(\vartheta))-z'(\vartheta))^2+\mathrm{Var}(\hat{z}'(\vartheta)) \\ &\approx \left(\frac{h^2}{24}z'''(\vartheta)\right)^2+\frac{2}{Nh^2}\mathrm{Var}(Z(\vartheta)) \end{aligned}$$

This is to be minimised as a function of h or, equivalently, h^2. A function $f(x)=bx^2+a/x$ attains its minimal value at $x_0=(a/2b)^{1/3}$, namely $f(x_0)=(ba^2)^{1/3}(2^{1/3}+(1/4)^{1/3})$. In our case, we have $x=h^2$; $a=2\mathrm{Var}(Z(\vartheta))/N$; $b=(z'''(\vartheta)/24)^2$. This yields that the optimal h is approximately

$$h_N=\left(\frac{576\mathrm{Var}(Z(\vartheta))}{z'''(\vartheta)^2}\right)^{1/6}N^{-1/6},$$

with a *root mean square error*

$$\sqrt{E((\hat{z}'(\vartheta) - z'(\vartheta))^2)} = O(1/\sqrt[3]{N}).$$

This reduced convergence rate means that we need eight times as many simulations to halve the error. Note that it in practice typically impossible to compute the optimal constant $\left(\frac{576\mathrm{Var}(Z(\vartheta))}{z'''(\vartheta)^2}\right)^{1/6}$ exactly because $\mathrm{Var}(Z(\vartheta))$ and $z'''(\vartheta)$ are often unknown as well. One solution is to first do preliminary Monte Carlo simulations for these quantities, with a small N and using similar methods as before, and then to base the choice of h_N on these estimators for $\mathrm{Var}(Z(\vartheta))$ and $z'''(\vartheta)$.

Let us come back to the Black-Scholes example, in which case we need to simulate

$$\begin{aligned}
&\frac{Z_n(s+h/2) - Z_n(s-h/2)}{h} \\
&= \frac{e^{-rT}}{h}\Bigg(\left((s+h/2)\exp\left((r-\sigma^2/2)T + \sigma\sqrt{T}X_+\right) - K\right)^+ \\
&\quad - \left((s-h/2)\exp\left((r-\sigma^2/2)T + \sigma\sqrt{T}X_-\right) - K\right)^+ \Bigg)
\end{aligned}$$

with independent standard normal random variables X_+, X_-. In this case, we may alternatively want to use the same rather than independent random variables $X = X_+ = X_-$. As before we have

$$E\left(\frac{Z(s+h/2) - Z(s-h/2)}{h}\right) = z'(s) + O(h^2)$$

because dependence does not play any role for the expectation. Moreover, we have

$$\begin{aligned}
0 \le &\frac{Z(s+h/2) - Z(s-h/2)}{h} \\
\le &\frac{e^{-rT}}{h}\Bigg(\left((s+h/2)\exp\left((r-\sigma^2/2)T + \sigma\sqrt{T}X\right) - K\right) \\
&\quad - \left((s-h/2)\exp\left((r-\sigma^2/2)T + \sigma\sqrt{T}X\right) - K\right) \Bigg) \\
= &\exp(-\sigma^2 T/2 + \sigma\sqrt{T}X)
\end{aligned}$$

because the function $f(x) = (x-K)^+$ is Lipschitz with Lipschitz constant 1. Hence we get for the variance

$$\begin{aligned}
\mathrm{Var}\left(\frac{Z(s+h/2) - Z(s-h/2)}{h}\right) &\le E\left(\left(\frac{Z(s+h/2) - Z(s-h/2)}{h}\right)^2\right) \\
&\le E\left(\exp(-\sigma^2 T + 2\sigma\sqrt{T}X)\right),
\end{aligned}$$

which now does not depend on h at all. Therefore there is no reason not to choose a very small h.

As a result we see that the use of common random variables in the difference quotient may considerably reduce the variance, to the effect that h can be chosen very small.

Infinitesimal perturbation A second approach is to simulate $Z'(\vartheta)$. If differentiation and expectation can be interchanged, we have $E(\frac{d}{d\vartheta}Z(\vartheta)) = \frac{d}{d\vartheta}E(Z(\vartheta)) = \frac{d}{d\vartheta}z(\vartheta) = z'(\vartheta)$, which means that the estimate

$$\hat{z}'(\vartheta) = \frac{1}{N}\sum_{n=1}^{N} Z'_n(\vartheta)$$

makes sense, where $Z'_1(\vartheta), \ldots, Z'_N(\vartheta)$ denote independent random variables with the same law as $Z'(\vartheta)$. It is indeed possible to interchange differentiation and expectation if e.g. $Z(\vartheta)$ is Lipschitz in a neighbourhood of ϑ with a random Lipschitz constant whose expected value is finite. In this case we obtain the usual Monte Carlo convergence rate of $1/\sqrt{N}$ instead of the $1/\sqrt[3]{N}$ of the finite difference approach.

In the above example of a European call we simulate

$$Z'(s) = 1_{\{s\exp((r-\sigma^2/2)T+\sigma\sqrt{T}X)>K\}} \exp\Big(-\sigma^2 T/2 + \sigma\sqrt{T}X\Big)$$

with standard normal X. The use of the finite difference method with common random variables leads in the limit $h \to 0$ to the infinitesimal perturbation approach.

Likelihood ratio method Sometimes the dependence of $z(\vartheta) = E(Z(\vartheta))$ on the parameter ϑ can be shifted from the random variable Z to the probability measure P, i.e. we are actually considering $z(\vartheta) = E_\vartheta(Z)$, where the probability measure P_ϑ instead of Z depends on ϑ. More specifically, suppose that $Z = f(X)$ where X has probability density function g_ϑ under P_ϑ, which implies

$$z(\vartheta) = E_\vartheta(f(X)) = \int f(x)g_\vartheta(x)dx.$$

If differentiation and integration can be interchanged, we have

$$\begin{aligned} z'(\vartheta) &= \int f(x)\frac{dg_\vartheta(x)}{d\vartheta}dx \\ &= \int f(x)\frac{\frac{dg_\vartheta(x)}{d\vartheta}}{g_\vartheta(x)}g_\vartheta(x)dx \\ &= \int f(x)\frac{d\log(g_\vartheta(x))}{d\vartheta}g_\vartheta(x)dx \\ &= E_\vartheta\left(f(X)\frac{d\log(g_\vartheta(X))}{d\vartheta}\right) \end{aligned}$$

Consequently, we may use the Monte Carlo estimate

$$\hat{z}'(\vartheta) = \frac{1}{N}\sum_{n=1}^{N} f(X_n)\frac{d\log(g_\vartheta(X_n))}{d\vartheta},$$

where $X_1, \dots, X_N$ are independent random variables with density g_ϑ. As with infinitesimal perturbation we obtain an unbiased estimate under conditions that allow to interchange differentiation and integration. Here, however, these hold more frequently than in the infinitesimal perturbation case.

As an example let us turn back to the European call with

$$\begin{aligned} z(s) &= E\left(e^{-rT}\left(s\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}X\right)-K\right)^+\right) \\ &= E\left(e^{-rT}\left(\exp\left(\log s+(r-\sigma^2/2)T+\sigma\sqrt{T}X\right)-K\right)^+\right) \\ &= E_s\left(e^{-rT}\left(\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}X\right)-K\right)^+\right), \end{aligned}$$

where X is $N(0,1)$-distributed under P and $N(\log(s)/\sqrt{\sigma^2 T},1)$-distributed under P_s. In the above notation (but with s instead of ϑ) we need

$$\begin{aligned} f(x) &= e^{-rT}\left(\exp\left((r-\sigma^2/2)T+\sigma\sqrt{T}x\right)-K\right)^+, \\ g_s(x) &= \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-\left(x-\frac{\log(s)}{\sqrt{\sigma^2 T}}\right)^2}{2}\right), \\ \frac{d\log(g_s(x))}{ds} &= \frac{x}{s\sqrt{\sigma^2 T}}-\frac{\log s}{s\sigma^2 T}. \end{aligned}$$

In contrast to the infinitesimal perturbation method this works also for discontinuous payoffs as e.g. digital options. One may also consult [FLL+99] for this approach in a more general context.

5.2 Simulation of stochastic integrals

In some applications we need to simulate the whole path $(X(t))_{t\in[0,T]}$ of a stochastic process rather than just its terminal value $X(T)$. This happens e.g. if we want to compute prices of path-dependent claims as e.g. lookback, barrier or Asian options. But even if we are interested only in $X(T)$, it may be necessary to simulate the past as well because we do not know the law of $X(T)$ in closed form (e.g. in the Heston model).

In this section we focus on diffusion-type processes, i.e. solutions to stochastic differential equations of the form

$$dX(t) = a(X(t),t)dt + b(X(t),t)dW(t) \tag{5.1}$$

with some Wiener process W, some starting value $X(0)$, and given deterministic functions a, b. It is obviously impossible to simulate infinitely many numbers on a real computer. The goal is therefore to generate random paths of X on some equidistant time grid $0 = t_0 < t_1 < \cdots < t_m = T$ with $t_i = i\Delta t$ and $\Delta t = T/m$, i.e. $X(t_0), \dots, X(t_m)$.

Simulation of Brownian motion and geometric Brownian motion In rare cases, the stochastic differential equation (5.1) has an explicit solution in terms of the driving Brownian motion W. This is obviously the case if a and b are constants, i.e.

$$X(t) = X(0) + at + bW(t). \tag{5.2}$$

Note that $\Delta W_i = W(t_i) - W(t_{i-1}), i = 1, \dots, m$ are independent normal random variables with mean 0 and variance Δt. Moreover, we have $X(t_i) = X(0) + \sum_{j=1}^{i}(a\Delta t + b\Delta W_j)$. In order to simulate $X(t_0), \dots, X(t_m)$, start with m independent standard normal random variables $Z_1, \dots, Z_m$ and set $Y_i := X(0) + \sum_{j=1}^{i}(a\Delta t + b\sqrt{\Delta t}Z_j)$ for $i = 0, \dots, m$. Then $(Y_0, Y_1, \dots, Y_m)$ has the same law and hence can be used as a simulation for the random vector $(X(t_0), \dots, X(t_m))$.

Let us consider the slightly more involved case of geometric Brownian motion, which is of the form

$$dX(t) = X(t)(\mu dt + \sigma dW(t)),$$

i.e. $a(X(t), t) = \mu X(t)$, $b(X(t), t) = \sigma X(t)$. Recall that the solution to this stochastic differential equation is

$$X(t) = X(0)\exp((\mu - \sigma^2/2)t + \sigma W(t)).$$

The process in the exponent is of the form (5.2) with $a = \mu - \sigma^2/2$, $b = \sigma$. If $(Y_0, Y_1, \dots, Y_m)$ denotes a simulation of this process at $t_0, \dots, t_m$, then $(X(0)e^{Y_0}, \dots, X(0)e^{Y_m})$ has the same law as the random vector $(X(t_0), \dots, X(t_m))$.

Euler method In general, exact simulation as in the previous paragraph is out of reach. In order to obtain a decent approximation, we proceed step by step. This means, that we always suppose that $X(t_0), \dots, X(t_{i-1})$ have already been simulated, and we are then interested in $X(t_i)$. Note that

$$X(t_i) = X(t_{i-1}) + \int_{t_{i-1}}^{t_i} a(X(t), t)dt + \int_{t_{i-1}}^{t_i} b(X(t), t)dW(t).$$

If the mesh size of the grid is small and the coefficients a, b are sufficiently smooth, we may approximate

$$a(X(t), t) \approx a(X(t_{i-1}), t_{i-1}), \quad b(X(t), t) \approx b(X(t_{i-1}), t_{i-1})$$

for $t_{i-1} < t \leq t_i$, which implies that

$$\begin{aligned} X(t_i) &\approx X(t_{i-1}) + \int_{t_{i-1}}^{t_i} a(X(t_{i-1}), t_{i-1})dt + \int_{t_{i-1}}^{t_i} b(X(t_{i-1}), t_{i-1})dW(t) \\ &= X(t_{i-1}) + a(X(t_{i-1}), t_{i-1})(t_i - t_{i-1}) + b(X(t_{i-1}), t_{i-1})(W(t_i) - W(t_{i-1})) \\ &= X(t_{i-1}) + a(X(t_{i-1}), t_{i-1})\Delta t + b(X(t_{i-1}), t_{i-1})\Delta W_i. \end{aligned}$$

with $\Delta W_i = W(t_i) - W(t_{i-1})$. Since $\Delta W_1, \ldots, \Delta W_m$ are independent normal random variables with mean 0 and variance Δt, we have found a simple method to simulate an approximation of X on the grid. The algorithm of this *Euler method* in pseudo code reads as follows.

Initialise $\Delta t = T/m$, $t_0 = 0$, $Y_0 = X(0)$,

Loop For $i = 1, 2, \ldots, m$ compute

$t_i = t_{i-1} + \Delta t$

$\Delta W = Z\sqrt{\Delta t}$ with new standard normal random variable Z

$Y_i = Y_{i-1} + a(Y_{i-1}, t_{i-1})\Delta t + b(Y_{i-1}, t_{i-1})\Delta W$

Output $Y_0, \ldots, Y_m$

The output $Y_0, \ldots, Y_m$ stands for the desired simulation of $X(t_0), \ldots, X(t_m)$.

Milshtein method and refinements

Stochastic Taylor expansion We analyse the Euler method using the so-called *stochastic Taylor expansion*. For simplicity, we suppose that a and b above do not depend on t, i.e.

$$dX(t) = a(X(t))dt + b(X(t))dW(t). \tag{5.3}$$

Itō's formula yields

$$\begin{aligned} df(X(t)) &= \left(f'(X(t))a(X(t)) + \frac{1}{2}f''(X(t))b(X(t))^2\right)dt + f'(X(t))b(X(t))dW(t) \\ &= L_0 f(X(t))dt + L_1 f(X(t))dW(t) \end{aligned}$$

for smooth functions f and

$$\begin{aligned} L_0 f(x) &= a(x)f'(x) + \frac{1}{2}b(x)^2 f''(x), \\ L_1 f(x) &= b(x)f'(x). \end{aligned}$$

Letting $f = a$ resp. b in (5.3) yields

$$\begin{aligned} X(t_i) &= X(t_{i-1}) + \int_{t_{i-1}}^{t_i} \left(a(X(t_{i-1})) + \int_{t_{i-1}}^{t} L_0 a(X(s))ds + \int_{t_{i-1}}^{t} L_1 a(X(s))dW(s)\right)dt \\ &\quad + \int_{t_{i-1}}^{t_i} \left(b(X(t_{i-1})) + \int_{t_{i-1}}^{t} L_0 b(X(s))ds + \int_{t_{i-1}}^{t} L_1 b(X(s))dW(s)\right)dW(t) \end{aligned}$$

with

$$\begin{aligned} L_0 a &= aa' + \frac{1}{2}b^2 a'', \quad L_1 a = ba' \\ L_0 b &= ab' + \frac{1}{2}b^2 b'', \quad L_1 b = bb'. \end{aligned}$$

Consequently,

$$X(t_i) = X(t_{i-1}) + a(X(t_{i-1}))(t_i - t_{i-1}) + b(X(t_{i-1}))(W(t_i) - W(t_{i-1})) + R_i$$

with

$$\begin{aligned} R_i &= \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} L_0 a(X(s))dsdt + \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} L_1 a(X(s))dW(s)dt \\ &\quad + \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} L_0 b(X(s))dsdW(t) + \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} L_1 b(X(s))dW(s)dW(t). \end{aligned}$$

The Euler method from above amounts to the approximation $R_i \approx 0$. But we can do better. Let us approximate R_i by assuming $X(s) \approx X(t_{i-1})$, which yields

$$R_i \approx L_0 a(X(t_{i-1}))I_{00} + L_1 a(X(t_{i-1}))I_{10} + L_0 b(X(t_{i-1}))I_{01} + L_1 b(X(t_{i-1}))I_{11}$$

with

$$\begin{aligned} I_{00} &= \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} dsdt = \int_{t_{i-1}}^{t_i}(t - t_{i-1})dt = \frac{1}{2}(t_i - t_{i-1})^2 = \frac{1}{2}(\Delta t)^2 = O((\Delta t)^2), \\ I_{10} &= \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} dW(s)dt = \int_{t_{i-1}}^{t_i}(W(t) - W(t_{i-1}))dt \\ &= \int_{t_{i-1}}^{t_i}(t_i - t)dW(t) \sim N(0, (t_i - t_{i-1})^3/3) = N(0, (\Delta t)^3/3), \\ I_{01} &= \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} dsdW(t) = \int_{t_{i-1}}^{t_i}(t - t_{i-1})dW(t) \\ &\quad \sim N(0, (t_i - t_{i-1})^3/3) = N(0, (\Delta t)^3/3), \\ I_{11} &= \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} dW(s)dW(t) = \int_{t_{i-1}}^{t_i}(W(t) - W(t_{i-1}))dW(t) \\ &= \frac{1}{2}\left((W(t_i) - W(t_{i-1}))^2 - (t_i - t_{i-1})\right) = \frac{1}{2}\left((\Delta W_i)^2 - \Delta t\right). \end{aligned}$$

The equality of the two integrals in I_{10} follows from integration by parts:

$$\begin{aligned} \int_{t_{i-1}}^{t_i}(t_i - t_{i-1})dW(t) &= (t_i - t_{i-1})(W(t_i) - W(t_{i-1})) \\ &= \int_{t_{i-1}}^{t_i}(t - t_{i-1})dW(t) + \int_{t_{i-1}}^{t_i}(W(t) - W(t_{i-1}))dt. \end{aligned}$$

Moreover, recall from Section 3.2 that integrals $\int_0^t \sigma(s)dW(s)$ with deterministic σ are normally distributed with mean 0 and variance $\int_0^t \sigma(s)^2 ds$. Since ΔW_i ist normally distributed with mean 0 and variance Δt, we have $E((\Delta W_i)^2) = \Delta t$ and $\mathrm{Var}((\Delta W_i)^2) = 2(\Delta t)^2$. Consequently, $E(I_{11}) = 0$ and $\mathrm{Var}(I_{11}) = (\Delta t)^2/2$.

Milshtein method The standard deviation of the I_{11}-term is of the order Δt and hence larger than $\sqrt{E(I^2)}$ for the other integrals $I = I_{00}, I_{10}, I_{01}$. Therefore, it seems most worthwhile to

include I_{11} in the approximation:

$$\begin{aligned} X(t_i) &\approx X(t_{i-1}) + a(X(t_{i-1}))(t_i - t_{i-1}) + b(X(t_{i-1}))(W(t_i) - W(t_{i-1}) \\ &\quad + \frac{1}{2} b(X(t_{i-1}))b'(X(t_{i-1})) \left((W(t_i) - W(t_{i-1}))^2 - (t_i - t_{i-1}) \right) \end{aligned}$$

The algorithm of the corresponding *Milshtein method* in pseudo code reads as follows.

Initialise $\Delta t = T/m$, $t_0 = 0$, $Y_0 = X(0)$,

Loop For $i = 1, 2, \ldots, m$ compute

$t_i = t_{i-1} + \Delta t$

$a = a(Y_{i-1})$

$b = b(Y_{i-1})$

$b' = b'(Y_{i-1})$

$\Delta W = Z\sqrt{\Delta t}$ with new standard normal random variable Z

$Y_i = Y_{i-1} + a\Delta t + b\Delta W + \frac{1}{2} bb'((\Delta W)^2 - \Delta t)$

Output $Y_0, \ldots, Y_m$

Again, the output $Y_0, \ldots, Y_m$ stands for the desired simulation of $X(t_0), \ldots, X(t_m)$.

A more refined method One can go even one step further and incorporate all four terms $I_{00}, I_{10}, I_{01}, I_{11}$ in the approximation:

$$\begin{aligned} X(t_i) &\approx X(t_{i-1}) + a(X(t_{i-1}))(t_i - t_{i-1}) + b(X(t_{i-1}))(W(t_i) - W(t_{i-1})) \\ &\quad + L_0 a(X(t_{i-1}))I_{00} + L_1 a(X(t_{i-1}))I_{10} + L_0 b(X(t_{i-1}))I_{01} + L_1 b(X(t_{i-1}))I_{11} \end{aligned}$$

If we set $\Delta Y_i = \int_{t_{i-1}}^{t_i} (t - t_{i-1}) dW(t)$, we can write

$$\begin{aligned} I_{00} &= \frac{1}{2}(\Delta t)^2, \\ I_{10} &= \Delta W_i \Delta t - \Delta Y_i, \\ I_{01} &= \Delta Y_i, \\ I_{11} &= \frac{1}{2}\left((\Delta W_i)^2 - \Delta t \right), \end{aligned}$$

hence

$$\begin{aligned} X(t_i) &\approx X(t_{i-1}) + a(X(t_{i-1}))\Delta t + b(X(t_{i-1}))\Delta W_i \\ &\quad + L_0 a(X(t_{i-1}))\frac{1}{2}(\Delta t)^2 + L_1 a(X(t_{i-1}))(\Delta W_i \Delta t - \Delta Y_i) \\ &\quad + L_0 b(X(t_{i-1}))\Delta Y_i + L_1 b(X(t_{i-1}))\frac{1}{2}\left((\Delta W_i)^2 - \Delta t \right). \end{aligned}$$

For a simulation of this expression we need the joint law of ΔW_i and ΔY_i. They are jointly normally distributed because they are both obtained by a linear operation from the same Wiener process W. Moreoever, both have expectation 0 because they are integrals relative to W. Their variances equal $\mathrm{Var}(\Delta W_i) = \Delta t$ and $\mathrm{Var}(\Delta Y_i) = (\Delta t)^3/3$ has already been calculated. The covariance of the two equals

$$\mathrm{Cov}(\Delta W_i, \Delta Y_i) = \int_{t_{i-1}}^{t_i} (t - t_{i-1})dt = \frac{1}{2}(\Delta t)^2$$

by (3.10). Altogther, we obtain that $(\Delta W_i, \Delta Y_i)$ is a Gaussian random vector with mean $(0, 0)$ and covariance matrix

$$\begin{pmatrix} \Delta t & \frac{1}{2}(\Delta t)^2 \\ \frac{1}{2}(\Delta t)^2 & \frac{1}{3}(\Delta t)^3 \end{pmatrix}.$$

Approximation error Let us briefly discuss the error of the above methods. To this end we denote the approximation of the solution $X(T)$ to the stochastic differential equation (5.3) based on the mesh Δt by $Y_{\Delta t}$.

The *absolute error* is

$$\varepsilon(\Delta t) = E(|X(T) - Y_{\Delta t}|).$$

One says that the method *converges strongly with order* $\gamma > 0$ if $\varepsilon(\Delta t) = O((\Delta t)^\gamma)$ for $\Delta t \to 0$. Relative to this criterion, the Euler method converges strongly with order $1/2$ whereas the inclusion of the I_{11}-term in the Milshtein method leads to an improved order of 1.

Note, however, that the absolute error is not the only reasonable. For the purposes of Monte-Carlo simulation, it is more important whether $E(f(X(T)))$ is close to $E(f(Y_{\Delta t}))$ for the function f under consideration. One says that the method converges weakly relative to f with convergence order $\beta > 0$ if

$$|E(f(X(T))) - E(f(Y_{\Delta t}))| = O((\Delta t)^\beta).$$

For sufficiently regular functions a, b, f both the Euler and the Milshtein method have weak convergence order 1. Therefore, the inclusion of the I_{11}-term does not improve the weak convergence order. This may be different if path dependent claims are to be computed. On the other hand, the inclusion of all four terms $I_{00}, I_{10}, I_{01}, I_{11}$ in the more refined method leads to an improved weak convergence order 2. Alternatively, the following *Richardson extrapolation* may be used to obtain an improved approximation.

Richardson extrapolation Since the Euler method has weak convergence order 1, it makes sense to expect

$$E(f(Y_{\Delta t})) = E(f(X(T))) + c\Delta t + O((\Delta t)^2)$$

with some constant c, at least if the coefficients a, b of the stochastic differential equation and the function f are sufficiently regular. If we repeat the simulation with mesh size $2\Delta t$ instead of Δt, we obtain

$$E(f(Y_{2\Delta t})) = E(f(X(T))) + 2c\Delta t + O((\Delta t)^2).$$

Let us now consider the linear combination

$$Z = 2f(Y_{\Delta t}) - f(Y_{2\Delta t})$$

of both simulations. We obtain

$$\begin{aligned} E(Z) &= E(f(X(T))) + 2c\Delta t - 2c\Delta t + O((\Delta t)^2) \\ &= E(f(X(T))) + O((\Delta t)^2), \end{aligned}$$

i.e. this simple modification yields a method of weak convergence order 2 under sufficient regularity.

Let us also consider the variance of this modified method. If $Y_{\Delta t}, Y_{2\Delta t}$ are simulated with independent random variables, we obtain

$$\mathrm{Var}(Z) = 4\mathrm{Var}(f(Y_{\Delta t})) + \mathrm{Var}(f(Y_{2\Delta t})).$$

As in the simulation of derivatives with the finite difference approach, the use of common random variables may reduce the variance. If $Y_{\Delta t}, Y_{2\Delta t}$ are not simulated independently, we obtain the expression

$$\mathrm{Var}(Z) = 4\mathrm{Var}(f(Y_{\Delta t})) + \mathrm{Var}(f(Y_{2\Delta t})) - 4\mathrm{Cov}(f(Y_{\Delta t}), f(Y_{2\Delta t}))$$

for the variance. We should therefore try to increase the correlation of $Y_{\Delta t}$ and $Y_{2\Delta t}$ as much as possible. To this end, observe that the sum

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

of two neighbouring increments corresponds to one increment on the grid with double mesh size $2\Delta t$. In the above approaches, the increments $W_{t_{i-1}} - W_{t_{i-2}}$, $W_{t_i} - W_{t_{i-1}}$ are simulated as $\sqrt{\Delta t} Z_{i-1}$ resp. $\sqrt{\Delta t} Z_i$ with independent standard normal random variables Z_1, Z_2. Consequently, one should use $\sqrt{\Delta t}(Z_{i-1} + Z_i)$ for the simulation of the corresponding increment on the coarse grid.

5.3 American options with regression-based Monte Carlo

Since American options prices involve Snell envelopes rather than straightforward expectations, their computation with Monte Carlo methods is not obvious. In this section we discuss the

popular Longstaff-Schwartz [**?**] approach, which can be applied in the Black-Scholes model and to similar Markov processes.

We use the notation of Section 3.4, in particular concerning American options. We specifically consider the American put with strike K and assume the interest rate to be nonnegative. Recall that the fair price process V of the American option is of the form

$$V(t) = \max\Big\{E_Q(e^{-r(\tau-t)}X(\tau)|\mathscr{F}_t) : \tau \text{ stopping time in } [t,T]\Big\}$$

with $X(t) = (K - S(t))^+$. In particular, we have

$$V(0) = E_Q(e^{-r\tau^\star}X(\tau^\star)) \tag{5.4}$$

for the optimal stopping time $\tau^\star$. The idea of the Longstaff-Schwartz approach is

- to provide an algorithm for approximating $\tau^\star$ by some $\bar\tau$ and, given $\bar\tau$,
- to replace the expected value $E_Q(e^{-r\bar\tau}X(\bar\tau))$ by a Monte Carlo sample mean as usual.

We consider the price process on a grid of times $0 = t_0 < t_1 < \cdots < t_m = T$ as in Section 5.2 with $t_i - t_{i-1} = T/m$. As a *first approximation* we replace the American option by the correponding *Bermudan* one which can only be exercised at these discrete time points, i.e.

$$V(t_i) := \max\Big\{E_Q(e^{-r(\tau-t_i)}X(\tau)|\mathscr{F}_{t_i}) : \tau \text{ stopping time in } \{t_i, t_{i+1}, \dots, t_m\}\Big\}. \tag{5.5}$$

The sequence of optimal stopping time $\tau_i^\star$, $i = 0, \dots, m$ in (5.5) satisfies the recursion $\tau_m^\star = T$ and

$$\tau_i^\star = \begin{cases} t_i & \text{if } X(t_i) \geq E_Q(e^{-r(\tau_{i+1}^\star - t_i)}X(\tau_{i+1}^\star)|\mathscr{F}_{t_i}), \\ \tau_{i+1}^\star & \text{otherwise.} \end{cases} \tag{5.6}$$

Indeed, this follows from the fact that $V(t_i) = E_Q(e^{-r(\tau_i^\star - t_i)}X(\tau_i^\star)|\mathscr{F}_{t_i})$ and

$$\begin{aligned}\tau_i^\star &= \inf\Big\{j \in \{t_i, t_{i+1}, \dots, t_{m-1}\} : X(t_j) \geq E_Q(e^{-(t_{j+1}-t_j)}V(t_{j+1})|\mathscr{F}_{t_j})\Big\} \wedge T \\ &= \inf\Big\{j \in \{t_i, t_{i+1}, \dots, t_{m-1}\} : X(t_j) \geq E_Q(e^{-r(\tau_{j+1}^\star - t_j)}X(\tau_{j+1}^\star)|\mathscr{F}_{t_j})\Big\} \wedge T.\end{aligned}$$

The idea now is to compute approximate solutions $\bar\tau_i$, $i = 1, \dots, m$ to (5.6), starting recursively with $\bar\tau_m = T$. In our Markovian context there exists some bounded deterministic function $f_i : \mathbb{R}_+ \to \mathbb{R}$ with $E_Q(e^{-r(\tau_{i+1}^\star - t_i)}X(\tau_{i+1}^\star)|\mathscr{F}_{t_i}) = f_i(S(t_i))$. This function belongs to the space $L^2(\mathbb{R}_+, \mathscr{B}_+, \mu)$ for any probability measure on $\mathbb{R}_+$. Hilbert space theory tells us that

$$f_i(s) = \sum_{k=0}^{\infty} a_{i,k}^\star \ell_k(s)$$

for some coefficients $a^\star_{i,0}, a^\star_{i,1}, \dots$ in $\mathbb{R}$, where $\ell_0, \ell_1, \dots$ denotes a family of basis functions of $L^2(\mathbb{R}_+, \mathscr{B}_+, \mu)$. We may for example use the sequence of Laguerre polynomials, which are orthogonal with respect to the scalar product on $L^2(\mathbb{R}_+, \mathscr{B}_+, \mu)$ if μ equals the exponential distribution with parameter 1.

Recall that $E(Y|\mathscr{C})$ minimizes $Z \mapsto E((Y-Z)^2)$ among all $\mathscr{C}$-measurable random variables Z. In our context $(a^\star_{i,k})_{k=0,1,\dots}$ minimizes

$$(a_k)_{k=0,1,\dots} \mapsto E\left(\left(e^{-r(\tau^\star_{i+1}-t_i)}X(\tau^\star_{i+1}) - \sum_{k=0}^{\infty} a_k \ell_k(S(t_i))\right)^2\right) \tag{5.7}$$

among all sequences $(a_k)_{k=0,1,\dots}$ such that the series inside the parentheses converges.

The *second approximation* in our approach is to consider only the first $1+M$ basis functions, i.e. to project on the smaller space

$$\left\{\sum_{k=0}^{M} a_k \ell_k(S_{t_i}) : a_0, \dots, a_M \in \mathbb{R}\right\}$$

of random variables. As a *third approximation* we replace the unknown stopping time $\tau^\star_{t_{i+1}}$ in (5.7) by its estimated counterpart $\bar\tau_{t_{i+1}}$. The final *fourth approximation* is to compute the expected value in (5.7) by Monte Carlo. To this end, we simulate N independent realizations of $(S(t_0), \dots, S(t_m))$ on the computer, the simulations denoted as $(S_{n,0}, \dots, S_{n,m})$ with $n = 1, \dots, N$. Since $X(t) = g(S(t))$ for the payoff function $g(x) = (K-x)^+$, we obtain corresponding realizations $(X_{n,0}, \dots, X_{n,m}) := (g(S_{n,0}), \dots, g(S_{n,m}))$. In the context of our recursion, suppose that the nearly optimal $\bar\tau_{i+1}$ has been computed already on each of these realized paths, the N realizations denoted as $\bar\tau_{n,i+1}$, $n = 1, \dots, N$. The payoff process at this time is denoted as $X_{n,\bar\tau_{i+1}}$, $n = 1, \dots, N$. Note that $\bar\tau_{n,i} = \bar\tau_{n,i+1}$ should obviously hold if $X_{n,i} = 0$ because stopping at t_i does not make sense if $X(t_i) = 0$. Therefore we consider only the in-the-money paths in the Monte-Carlo simulation, i.e. the subset $I_i \subset \{1, \dots, N\}$ with $X_{n,i} > 0$ for $n \in I_i$.

In view of (5.7), the task is now to minimize

$$(a_k)_{k=0,\dots,M} \mapsto \frac{1}{|I_i|}\sum_{n\in I_i}\left(e^{-r(\bar\tau_{n,i+1}-t_i)}X_{\bar\tau_{n,t_{i+1}}} - \sum_{k=0}^{\infty} a_k \ell_k(S_{n,i})\right)^2, \tag{5.8}$$

the minimizer denoted as $\hat a_i = (\hat a_{i,k})_{k=0,\dots,M} \in \mathbb{R}^{1+M}$. This amounts to a linear regression problem of the form

$$\min_{a\in\mathbb{R}^{1+M}} \sum_{n\in I_i}\left(y_n - \sum_{k=0}^{M} a_k x_{n,k}\right)^2$$

whose solution is of the form $a = (x^\top x)^{-1}x^\top y$ if $x^\top x$ is invertible. Applied to our situation we conclude that $\hat a_i = (x^\top x)^{-1}x^\top y$ with $y_n = e^{-r(\bar\tau_{n,i+1}-t_i)}X_{n,\bar\tau_{t_{i+1}}}$ and $x_{n,k} = \ell_k(S_{n,i})$. In view of

(5.6) we now set

$$\bar{\tau}_{n,i} = \begin{cases} t_i & \text{if } e^{-r(\bar{\tau}_{n,i}-t_i)} X_{n,i} \geq \sum_{k=0}^{M} \bar{a}_{i,k} \ell_k(S_{n,i}) \text{ and } X_{n,i} > 0, \\ \bar{\tau}_{n,i+1} & \text{otherwise.} \end{cases}$$

The final Monte Carlo approximation of V_0 is now given by $\frac{1}{N} \sum_{n=1}^{N} e^{-r\bar{\tau}_{n,1}} X_{\bar{\tau}_1,n}$ unless immediate stopping yields a higher value, i.e. we obtain

$$V_0 \approx \max \left\{ (K - S(0))^+, \frac{1}{N} \sum_{n=1}^{N} e^{-r\bar{\tau}_{n,1}} X_{\bar{\tau}_1,n} \right\}.$$

An algorithm for the computation of American put prices with the Longstaff-Schwartz approach reads as follows.

Input Model parameters: r, σ, S_0

Option parameters: K (strike), T (maturity)

Further parameters: m, N, basis functions $\ell_0, \ldots, \ell_M$

Compute $\Delta t = T/m$

$t_i = i\Delta t$, $i = 1, \ldots, m$

On the grid $0 = t_0 < t_1 < \cdots < t_m = T$, simulate N paths of the geometric Brownian motion $S(t) = S_0 \exp((r - \sigma^2/2)t + \sigma W(t))$, as explained in Section 5.2: $S_{n,i} = S_n(t_i)$ for $i = 0, \ldots, m, \quad n = 1, \ldots, N$

$\tau_n = T$ and $g_n := (K - S_{n,m})^+$ for $n = 1, \ldots, N$

t-loop For $i = m - 1, \ldots, 1$:

For the subset I_i of in-the-money paths (i.e. for the n satisfying $K - S_{i,n} > 0$) set $(x_n, y_n) := (\ell_k(S_{n,i})_{k=0,\ldots,M}, e^{-r(\tau_n - i\Delta t)} g_n), \quad n \in I_i$.

Linear regression: compute the coefficients $\hat{a}_0, \ldots, \hat{a}_M$ minimizing $(a_0, \ldots, a_M) \mapsto \sum_{n \in I_i} \left(y_n - \sum_{k=0}^{M} a_k x_{n,k} \right)^2$.

For the $n \in I_i$ with $(K - S_{n,i})^+ \geq \sum_{k=0}^{M} \hat{a}_k \ell_k(S_{n,i})$ update $\tau_n := t_i$, $g_n = (K - S_{n,i})^+$.

Output $V_0 := \max\{(K - S_0)^+, \frac{1}{N} \sum_{n=1}^{N} e^{-r\tau_n} g_n\}$

The output of this algorithm is the American put price at time 0.

5.4 Outlook and discussion

Let us briefly discuss the multivariate case where both the Brownian motion and the solution X of (5.1) are vector-valued. This happens e.g. for the Heston model (3.37, 3.39). The Euler method and Richardson extrapolation allow for straightforward extensions to this slightly more general setup. Higher order methods relying on the stochastic Taylor expansion, on the other hand, may require a detailed analysis because of involved multiple integrals.

Altogether, the Monte Carlo method provides a universal tool which can be implemented relatively easily. Variance reduction and similar refinements may be necessary to achieve a reasonable accuracy.

Chapter 6

Finite differences

As we have seen in Section 3.4, option pricing and hedging often boils down to solving partial differential equations (PDE's). In this chapter we discuss methods to solve such equations. Recall that the fair price at time t of a European option with payoff $f(S(T))$ is of the form $v(t, S(t))$, where v solves (3.26, 3.27). The goal is now to solve this Black-Scholes PDE numerically.

6.1 Discretising the Black-Scholes PDE: an explicit scheme

Transformation to the heat equation We begin with some transformations which lead to a simpler equation and help avoiding some numerical problems. To this end write $\widetilde{x} = \log(x/K)$, $\widetilde{t} = \frac{\sigma^2}{2}(T-t)$, $q = 2r/\sigma^2$,

$$y(\widetilde{t}, \widetilde{x}) = v\left(T - \frac{2\widetilde{t}}{\sigma^2}, Ke^{\widetilde{x}}\right) K^{-1} \exp\left(\frac{1}{2}(q-1)\widetilde{x} + \left(\frac{1}{4}(q-1)^2 + q\right)\widetilde{t}\right).$$

Straightforward calculations show that $y(\widetilde{t}, \widetilde{x})$ solves the PDE

$$\frac{\partial y}{\partial \widetilde{t}} = \frac{\partial^2 y}{\partial \widetilde{x}^2} \tag{6.1}$$

with initial condition

$$y(0, \widetilde{x}) = \begin{cases} \max\{e^{\frac{\widetilde{x}}{2}(q+1)} - e^{\frac{\widetilde{x}}{2}(q-1)}, 0\} & \text{for a call,} \\ \max\{e^{\frac{\widetilde{x}}{2}(q-1)} - e^{\frac{\widetilde{x}}{2}(q+1)}, 0\} & \text{for a put.} \end{cases} \tag{6.2}$$

We look for a solution for $\widetilde{t} \in [0, \frac{1}{2}\sigma^2 T]$, $\widetilde{x} \in \mathbb{R}$. The solution to the original problem can be recovered from $y(\widetilde{t}, \widetilde{x})$ by

$$\begin{aligned} & v(t,x) \\ &= y\left(\frac{\sigma^2}{2}(T-t), \log\frac{x}{K}\right) K \exp\left(-\frac{1}{2}(q-1)\log\frac{x}{K} - \left(\frac{1}{4}(q-1)^2 + q\right)\frac{\sigma^2}{2}(T-t)\right). \end{aligned}$$

Equation (6.1) is called *heat equation* and it plays an important role in physics.

Discrete approximation In the finite difference approach to solving PDE's numerically, the derivatives are approximated by difference quotients. Similarly as in Section 5.1 we consider Taylor expansions of a general smooth function f. Its first order approximation reads as

$$f(x+h) = f(x) + f'(x)h + O(h^2),$$

which implies

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h).$$

Recall that $O(h)$ means that $O(h)/h$ is bounded for small h, i.e. the approximation error depends roughly linearly on h. A better approximation is obtained if we consider a symmetric difference quotient. Considering second order Taylor approximations we get

$$\begin{aligned} f(x+h) - f(x-h) &= f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + O(h^3) \\ &\quad - f(x) + f'(x)h - \frac{1}{2}f''(x)h^2 + O(h^3), \end{aligned}$$

which implies

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2). \tag{6.3}$$

Note that the error is now quadratic in h and hence much smaller if h is small enough.

For the numerical treatment of (6.1) we need to consider second order derivatives as well. As before we prefer symmetric expressions. Using third order Taylor expansions we obtain

$$\begin{aligned} &f(x+h) - 2f(x) + f(x-h) \\ &= f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(x)h^3 + O(h^4) \\ &\quad - 2f(x) + f(x) - f'(x)h + \frac{1}{2}f''(x)h^2 - \frac{1}{6}f'''(x)h^3 + O(h^4) \end{aligned}$$

which yields

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2).$$

We are now ready to discretise PDE (6.1). Instead of solving it on the whole area $\widetilde{t} \in [0, \frac{1}{2}\sigma^2 T]$, $\widetilde{x} \in \mathbb{R}$, we compute an approximation of y on some grid. More specifically we look for

$$y_{\nu i} = y(\widetilde{t}_\nu, \widetilde{x}_i), \quad \nu = 0, 1, \ldots, \nu_{\max}, \quad i = 0, 1, \ldots, m,$$

where $\widetilde{t}_\nu = \nu\Delta\widetilde{t}$; $\Delta\widetilde{t} = \widetilde{t}_{\max}/\nu_{\max}$; $\widetilde{t}_{\max} = \frac{\sigma^2}{2}T$; $a < 0 < b$; $\widetilde{x}_i = a + i\Delta\widetilde{x}$; $\Delta\widetilde{x} = (b-a)/m$. Let us denote the numerical approximation as $w_{\nu i}$, in contrast to the true solution which is denoted by $y_{\nu i}$.

Explicit scheme In the explicit method we use a one-sided approximation for the derivative in time and a symmetric one for the derivative in space:

$$\frac{\partial y_{\nu i}}{\partial \widetilde{t}} = \frac{y_{\nu+1,i} - y_{\nu i}}{\Delta \widetilde{t}} + O(\Delta \widetilde{t}) \tag{6.4}$$

$$\frac{\partial^2 y_{\nu i}}{\partial \widetilde{x}^2} = \frac{y_{\nu,i+1} - 2y_{\nu i} + y_{\nu,i-1}}{\Delta \widetilde{x}^2} + O(\Delta \widetilde{x}^2). \tag{6.5}$$

A numerical approximation $w_{\nu i}$ of the solution $y_{\nu i}$ to (6.1) is obtained by equating (6.4, 6.5) and ignoring the error terms:

$$\frac{w_{\nu+1,i} - w_{\nu i}}{\Delta \widetilde{t}} = \frac{w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1}}{\Delta \widetilde{x}^2}. \tag{6.6}$$

We proceed step by step in time. Suppose that $w_{\widetilde{\nu} i}$ has already been computed for $\widetilde{\nu} \leq \nu$ and all i. Then $w_{\nu+1,i}$ is obtained as

$$w_{\nu+1,i} = w_{\nu i} + \frac{\Delta \widetilde{t}}{\Delta \widetilde{x}^2}(w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1})$$

or

$$w_{\nu+1,i} = \lambda w_{\nu,i-1} + (1 - 2\lambda) w_{\nu i} + \lambda w_{\nu,i+1} \tag{6.7}$$

with $\lambda = \frac{\Delta \widetilde{t}}{\Delta \widetilde{x}^2}$. The starting values $w_{0i} = y(0, \widetilde{x}_i)$ are obtained from the initial condition (6.2). It is less obvious how to choose the boundary values for $i = 0$ resp. m. For the time being we let $w_{\nu 0} = w_{\nu m} = 0$. A better choice will be discussed later. In vector notation we write the approximation at time $\widetilde{t}_\nu$ as $w^{(\nu)} = (w_{\nu 1}, \ldots, w_{\nu,m-1})^\top$. Equation (6.7) now reads as

$$w^{(\nu+1)} = A w^{(\nu)}, \quad \nu = 0, 1, 2, \ldots \tag{6.8}$$

with

$$A = A_{\text{expl}} = \begin{pmatrix} 1-2\lambda & \lambda & 0 & \ldots & 0 \\ \lambda & 1-2\lambda & \lambda & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \lambda & 1-2\lambda & \lambda \\ 0 & \ldots & 0 & \lambda & 1-2\lambda \end{pmatrix}.$$

An algorithm for the computation of option prices using the explicit scheme reads as follows.

Input Model parameters: r, σ

Option parameters: K, T, call/put

Further parameters: $a, b, m, \nu_{\max}$ (with $\sigma^2 T/(2\nu_{\max}) < (b-a)^2/(2m^2)$, see below)

Compute $q = 2r/\sigma^2$

$\Delta \widetilde{x} = (b-a)/m$

$\Delta\widetilde{t} = \sigma^2 T/(2\nu_{\max})$

$\widetilde{x}_i = a + i\Delta\widetilde{x}$ for $i = 0, \ldots, m$

$\lambda = \Delta\widetilde{t}/\Delta\widetilde{x}^2$

$w = (w_1, \ldots, w_{m-1})$ with $w_i = \begin{cases} \max\{e^{\frac{\widetilde{x}_i}{2}(q+1)} - e^{\frac{\widetilde{x}_i}{2}(q-1)}, 0\} & \text{for a call} \\ \max\{e^{\frac{\widetilde{x}_i}{2}(q-1)} - e^{\frac{\widetilde{x}_i}{2}(q+1)}, 0\} & \text{for a put} \end{cases}$

$\widetilde{t}$-loop For $\nu = 1, \ldots, \nu_{\max}$:

$v_1 = (1 - 2\lambda)w_1 + \lambda w_2$

$v_{m-1} = \lambda w_{m-2} + (1 - 2\lambda)w_{m-1}$

$v_i = \lambda w_{i-1} + (1 - 2\lambda)w_i + \lambda w_{i+1}$ for $i = 2, \ldots, m-2$

$w = (v_1, \ldots, v_{m-1})$

Compute for $i = 1, \ldots, m-1$:

$S_i = Ke^{\widetilde{x}_i}$

$v(0, S_i) = Kw_i \exp(-\frac{\widetilde{x}_i}{2}(q-1) - \frac{1}{2}\sigma^2 T((q-1)^2/4 + q))$

Output $(S_i, v(0, S_i))_{i=1,\ldots,m-1}$

The output of this algorithm are $m-1$ pairs of initial stock prices and the corresponding option prices.

Stability The approximation error in (6.4) resp. (6.5) is small if the grid is very fine. On the other hand, choosing a fine grid means repeating (6.7) many times. It may not be a priori clear whether these many iterations do not ultimately lead to a large error even if the error of any single computation is very small.

If the mesh size of the grid tends to zero, we would like the difference between the numerical and the exact solution to converge to 0 as well. As a general rule in numerical analysis, this *convergence* of the method follows from *consistency* and *stability*. A method is called *consistent* if the local approximation error tends to 0. Roughly speaking, this local error corresponds to the $O(\Delta\widetilde{t})$ and $O(\Delta\widetilde{x}^2)$ terms in (6.4) and (6.5). *Stability* prevents that small errors blow up in the process of the computation. Rather than giving precise definitions and theorems we illustrate the phenomenon heuristically. The explicit scheme above is based on the iteration (6.8). Up to rounding errors, $w^{(\nu)}$ coincides with the exact solution $y^{(\nu)}$ for $\nu = 0$. This is no longer the case for $\nu \geq 1$. Let us denote the difference between the two by $e^{(\nu)}$. How does this initial error from step 1 contribute to the result at a later time ν? To this end note that $A^\nu w^{(1)} - A^\nu y^{(1)} = A^\nu e^{(1)}$.

In order for errors not to blow up, we would like this error to tend to 0 for $\nu \to \infty$. In other words, we require

$$A^\nu z \overset{\nu\to\infty}{\to} 0 \text{ for any } z \in \mathbb{R}^{m-1}. \tag{6.9}$$

According to a result from numerical analysis, (6.9) holds if and only if $(A^\nu)_{ij} \to 0$ for any i, j which in turn holds if and only if the maximal absolute eigenvalue $\varrho(A)$ of A satisfies $\varrho(A) < 1$. This maximal absolute eigenvalue is called *spectral radius* of A.

We have the following lemma, which can be verified by straightforward insertion.

Lemma 6.1 (Eigenvalues of simple tridiagonal matrices) *The eigenvalues and eigenvectors of*

$$G = \begin{pmatrix} \alpha & \beta & 0 & \dots & 0 \\ \gamma & \alpha & \beta & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \gamma & \alpha & \beta \\ 0 & \dots & 0 & \gamma & \alpha \end{pmatrix} \in \mathbb{R}^{n\times n}$$

are of the form $\mu^{(k)} = \alpha + 2\beta\sqrt{\gamma/\beta}\cos(k\pi/(n+1))$ *and*

$$v^{(k)} = \left(\sqrt{\frac{\gamma}{\beta}}\sin\left(\frac{k\pi}{n+1}\right), \dots, \left(\sqrt{\frac{\gamma}{\beta}}\right)^n \sin\left(\frac{nk\pi}{n+1}\right)\right)$$

for $k = 1, \dots, n$.

Note that matrix $A = A_{\text{expl}}$ above is of the form $A = 1_n - \lambda G$ with

$$G = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}. \tag{6.10}$$

Applying Lemma 6.1 yields eigenvalues $2 - 2\cos(k\pi/m) = 4\sin^2(k\pi/(2m))$, $k = 1, \dots, m-1$ of G and hence $1 - 4\lambda\sin^2(k\pi/(2m))$ of A. Now $|1 - 4\lambda\sin^2(k\pi/(2m))| < 1$ holds if and only if $\lambda\sin^2((m-1)\pi/(2m)) < 1/2$. Note that $\sin^2((m-1)\pi/(2m))$ converges to 1 for $m \to \infty$. In order for A to satisfy the stability criterion above regardless of m, we require $\lambda < 1/2$ or, equivalently, $\Delta\widetilde{t} < \Delta\widetilde{x}^2/2$.

Altogether, this implies that the time steps $\Delta\widetilde{t}$ must be chosen very small if we want to attain a fine resolution $\Delta\widetilde{x}$ in space, This restriction leads to possibly high computation costs. We will now consider a modified method without this restriction.

6.2 Implicit and Crank-Nicolson schemes

Implicit scheme (6.4) is not the only way to discretise the time derivative. We could also use the approximation

$$\frac{\partial y_{\nu i}}{\partial \widetilde{t}} = \frac{y_{\nu i} - y_{\nu-1,i}}{\Delta \widetilde{t}} + O(\Delta \widetilde{t}), \tag{6.11}$$

which leads to

$$\frac{w_{\nu,i} - w_{\nu-1,i}}{\Delta \widetilde{t}} = \frac{w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1}}{\Delta \widetilde{x}^2}. \tag{6.12}$$

or

$$w_{\nu-1,i} = -\lambda w_{\nu,i-1} + (1+2\lambda) w_{\nu i} - \lambda w_{\nu,i+1} \tag{6.13}$$

instead of (6.7), where $\lambda = \frac{\Delta \widetilde{t}}{\Delta \widetilde{x}^2}$. On first glance, this does not seem to lead to anything useful because three unknowns $w_{\nu,i-1}, w_{\nu i}, w_{\nu,i+1}$ now have to be computed from only one known variable $w_{\nu-1,i}$. However, in vector notation along the same lines as (6.8), (6.13) reads as

$$A w^{(\nu)} = w^{(\nu-1)}, \quad \nu = 1, 2, \ldots \tag{6.14}$$

with

$$A = A_{\text{impl}} = \begin{pmatrix} 1+2\lambda & -\lambda & 0 & \ldots & 0 \\ -\lambda & 1+2\lambda & -\lambda & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -\lambda & 1+2\lambda & -\lambda \\ 0 & \ldots & 0 & -\lambda & 1+2\lambda \end{pmatrix}.$$

The solution of this equation is given by

$$w^{(\nu)} = A^{-1} w^{(\nu-1)}.$$

As in the explicit case, we tacitly assumed the boundary conditions $w_{\nu 0} = w_{\nu m} = 0$.

What about stability of this *implicit scheme*? Note that $A = A_{\text{impl}}$ is of the form $A = 1_n + \lambda G$ with G as in (6.10). Similar as before, we have that its eigenvalues are of the form $1 + 4\lambda \sin^2(k\pi/(2m))$, $k = 1, \ldots, m-1$. Consequently, the eigenvalues of the inverse matrix A^{-1} equal $1/(1 + 4\lambda \sin^2(k\pi/(2m)))$, $k = 1, \ldots, m-1$. It is easy to see that the modulus of these eigenvalues is dominated by 1 regardless of λ. Consequently, we need not impose any restriction on $\Delta \widetilde{t}$ in order for the implicit scheme to be stable.

But for an efficient implementation we need to invert matrix A_{impl} quickly. In general, inversion of an $n \times n$ matrix e.g. by Gaussian elimination requires $O(n^3)$ operations. However, for a tridiagonal matrix this can be done in $O(n)$ steps as we will see below.

Inversion of tridiagonal matrices We are looking for the solution $x \in \mathbb{R}^n$ of the equation $Ax = b$ for tridiagonal $A \in \mathbb{R}^{n\times n}$, i.e.

$$\begin{pmatrix} \alpha_1 & \beta_1 & 0 & \dots & 0 \\ \gamma_2 & \alpha_2 & \beta_2 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \gamma_{n-1} & \alpha_{n-1} & \beta_{n-1} \\ 0 & \dots & 0 & \gamma_n & \alpha_n \end{pmatrix} \begin{pmatrix} x_1 \\ \\ \vdots \\ \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \\ \vdots \\ \\ b_n \end{pmatrix}. \tag{6.15}$$

We start by eliminating the γ_i by starting in the second row and successively subtracting an appropriate multiple of row $i-1$ from row i. This yields the new equation

$$\begin{pmatrix} \hat{\alpha}_1 & \beta_1 & 0 & \dots & 0 \\ 0 & \hat{\alpha}_2 & \beta_2 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \hat{\alpha}_{n-1} & \beta_{n-1} \\ 0 & \dots & 0 & 0 & \hat{\alpha}_n \end{pmatrix} \begin{pmatrix} x_1 \\ \\ \vdots \\ \\ x_n \end{pmatrix} = \begin{pmatrix} \hat{b}_1 \\ \\ \vdots \\ \\ \hat{b}_n \end{pmatrix}. \tag{6.16}$$

with

$$\begin{aligned} \hat{\alpha}_1 &= \alpha_1, \\ \hat{b}_1 &= b_1, \\ \hat{\alpha}_i &= \alpha_i - \frac{\gamma_i}{\hat{\alpha}_{i-1}}\beta_{i-1}, \\ \hat{b}_i &= b_i - \frac{\gamma_i}{\hat{\alpha}_{i-1}}\hat{b}_{i-1}, \quad i = 2, \dots, n. \end{aligned}$$

This simpler equation can now be solved moving from the bottom to the top. We obtain

$$x_n = \frac{\hat{b}_n}{\hat{\alpha}_n}, \tag{6.17}$$

$$x_i = \frac{\hat{b}_i - \beta_i x_{i+1}}{\hat{\alpha}_i}, \quad i = n-1, \dots, 1. \tag{6.18}$$

Altogether, we end up with the following algorithm in $O(n)$ operations.

Input $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_{n-1}, \gamma_2, \dots, \gamma_n, b_1, \dots, b_n$.

Compute $\hat{\alpha}_1 = \alpha_1, \hat{b}_1 = b_1$

For $i = 2, \dots, n$: $\hat{\alpha}_i = \alpha_i - \frac{\gamma_i}{\hat{\alpha}_{i-1}}\beta_{i-1}, \hat{b}_i = b_i - \frac{\gamma_i}{\hat{\alpha}_{i-1}}\hat{b}_{i-1}$.

$x_n = \frac{\hat{b}_n}{\hat{\alpha}_n}$

For $i = n-1, \dots, 1$: $x_i = \frac{1}{\hat{\alpha}_i}(\hat{b}_i - \beta_i x_{i+1})$.

Output $x_1, \ldots, x_n$

Of course, some care has to be taken in order not to divide by zero. We could e.g. require that $\beta_i, \gamma_i \leq 0$ for all i and

$$\min\{\alpha_1, \ldots, \alpha_n\} + \max\{\beta_1, \ldots, \beta_{n-1}\} + \max\{\gamma_2, \ldots, \gamma_n\} > 0,$$

cf. the arguments in [JLL90, Lemmas 5.9 and 5.5].

Crank-Nicolson scheme So far, the asymmetric discretisation of the time derivative leads to an error of the order $\Delta\widetilde{t}_i$. By considering instead a symmetric version, we can reduce this to $\Delta\widetilde{t}_i^2$. To this end, we average (6.6) and (6.12), the latter with $\nu+1$ instead of ν, which yields

$$\frac{w_{\nu+1,i} - w_{\nu i}}{\Delta\widetilde{t}} = \frac{w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1} + w_{\nu+1,i+1} - 2w_{\nu+1,i} + w_{\nu+1,i-1}}{2\Delta\widetilde{x}^2}. \tag{6.19}$$

The left-hand side of (6.19) can be interpreted as a symmetric discretisation (6.3) of the derivative at $\widetilde{t}_{\nu+1/2} = (\widetilde{t}_\nu + \widetilde{t}_{\nu+1})/2$. The right-hand side corresponds to the average of the discretisation of the $\widetilde{x}$-derivatives at times $\widetilde{t}_\nu$ and $\widetilde{t}_{\nu+1}$, respectively. This average replaces the discretisation of the $\widetilde{x}$-derivative at time $\widetilde{t}_{\nu+1/2}$, which would involve values that are not on the grid.

(6.19) can be written as

$$-\frac{\lambda}{2}w_{\nu+1,i-1} + (1+\lambda)w_{\nu+1,i} - \frac{\lambda}{2}w_{\nu+1,i+1} = +\frac{\lambda}{2}w_{\nu,i-1} + (1-\lambda)w_{\nu i} + \frac{\lambda}{2}w_{\nu,i+1} \tag{6.20}$$

with $\lambda = \frac{\Delta\widetilde{t}}{\Delta\widetilde{x}^2}$. As in the implicit case, this makes more sense in vector notation. (6.20) means

$$Aw^{(\nu+1)} = Bw^{(\nu)}, \quad \nu = 0, 1, 2, \ldots \tag{6.21}$$

with

$$A = A_{\text{CN}} = \begin{pmatrix} 1+\lambda & -\lambda/2 & 0 & \ldots & 0 \\ -\lambda/2 & 1+\lambda & -\lambda/2 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -\lambda/2 & 1+\lambda & -\lambda/2 \\ 0 & \ldots & 0 & -\lambda/2 & 1+\lambda \end{pmatrix},$$

$$B = B_{\text{CN}} = \begin{pmatrix} 1-\lambda & \lambda/2 & 0 & \ldots & 0 \\ \lambda/2 & 1-\lambda & \lambda/2 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \lambda/2 & 1-\lambda & \lambda/2 \\ 0 & \ldots & 0 & \lambda/2 & 1-\lambda \end{pmatrix},$$

where we implicitly used the boundary conditions $w_{\nu 0} = w_{\nu m} = 0$ as before. Since A is a tridiagonal matrix, the solution $w^{(\nu+1)}$ to (6.21) can be computed in $O(n)$ (here: $O(m-1)$)

operations using the algorithm from above. Let us have a look at the stability of this method. Note that $w^{(\nu+1)} = (4C^{-1} - 1_n)w^{(\nu)}$ with $C = 2(1 + \frac{\lambda}{2}G)$ and G as in (6.10). Analysing the eigenvalues similarly as before yields that the stability condition concerning the the spectral radius of $4C^{-1} - 1$ is satisfied for any choice of λ.

The Crank-Nicolson scheme can be implemented as easily as the implicit scheme and it shares the unconditional stability property. But compared to both the explicit and the implicit method, it converges faster. Due to the symmetric discretisation of the derivatives both in space and time, rate of convergence is typically $O(\Delta\tilde{t}^2) + O(\Delta\tilde{x}^2)$. We end by stating the algorithm of the Crank-Nicolson scheme in pseudo code.

Input $a, b, \nu_{\max}, m$

Compute $\Delta\tilde{x} = (b-a)/m$

$\Delta\tilde{t} = \frac{\sigma^2 T}{2\nu_{\max}}$

$w_i^{(0)} = y(0, \tilde{x}_i)$, $i = 0, \ldots, m$ with y from (6.2).

Loop For $\nu = 0, \ldots, \nu_{\max} - 1$:

Solve

$$Ax = Bw^{(\nu)} \tag{6.22}$$

for x.

Set $w^{(\nu+1)} = x$.

Output $w^{(\nu_{\max})}$

In the implicit case we must use instead $A = A_{\text{impl}}$, $B = 1$. The explicit case, on the other hand, corresponds to $A = 1$, $B = A_{\text{expl}}$ where matrix inversion is not needed.

Boundary conditions So far we imposed the naive boundary conditions $w_{\nu 0} = w_{\nu m} = 0$. These can be improved by considering the true asymptotics of the option price for very small and very large stock prices. Let us consider call options first. If the present stock price is very small, it is very unlikely that the option ends up in the money. This motivates that the fair option price satisfies $v(t, S(t)) \approx 0$ for $S(t) \ll K$. If, on the other hand, the present stock price is very high, a final stock price below the strike is very unlikely. The fair price of an option with payoff $S(T) - K$ equals $S(t) - Ke^{-r(T-t)}$ at time t. This motivates the approximation $v(t, S(t)) \approx S(t) - Ke^{-r(T-t)}$ for $S(t) \gg K$. By the put-call parity, the put price can be approximated as $v(t, S(t)) \approx Ke^{-r(T-t)} - S(t)$ for $S(t) \ll K$ and $v(t, S(t)) \approx 0$ for $S(t) \gg K$.

Changing variables back to $\widetilde{t}, \widetilde{x}$ etc., we end up with the approximations

$$\begin{aligned} y(\widetilde{t},\widetilde{x}) &\approx r_1(\widetilde{t},\widetilde{x}) \text{ for } \widetilde{x} \to -\infty, \\ y(\widetilde{t},\widetilde{x}) &\approx r_2(\widetilde{t},\widetilde{x}) \text{ for } \widetilde{x} \to \infty \end{aligned}$$

with

$$\begin{aligned} r_1(\widetilde{t},\widetilde{x}) &= 0 \\ r_2(\widetilde{t},\widetilde{x}) &= \exp\left(\frac{1}{2}(q+1)\widetilde{x} + \frac{1}{4}(q+1)^2\widetilde{t}\right) - \exp\left(\frac{1}{2}(q-1)\widetilde{x} + \frac{1}{4}(q-1)^2\widetilde{t}\right) \end{aligned}$$

for the European call and

$$\begin{aligned} r_1(\widetilde{t},\widetilde{x}) &= \exp\left(\frac{1}{2}(q-1)\widetilde{x} + \frac{1}{4}(q-1)^2\widetilde{t}\right) - \exp\left(\frac{1}{2}(q+1)\widetilde{x} + \frac{1}{4}(q+1)^2\widetilde{t}\right) \\ r_2(\widetilde{t},\widetilde{x}) &= 0 \end{aligned}$$

for the European put. This suggests modifying the numerical schemes by setting $w_{\nu 0} = r_1(\widetilde{t}_\nu, a)$, $w_{\nu m} = r_2(\widetilde{t}_\nu, b)$. The modified boundary conditions imply that e.g. the Crank-Nicolson algorithm now reads as

$$Aw^{(\nu+1)} = Bw^{(\nu)} + d^{(\nu)}$$

with

$$d^{(\nu)} = \begin{pmatrix} \frac{\lambda}{2}(r_1(\widetilde{t}_{\nu+1}, a) + r_1(\widetilde{t}_\nu, a)) \\ 0 \\ \vdots \\ 0 \\ \frac{\lambda}{2}(r_2(\widetilde{t}_{\nu+1}, b) + r_2(\widetilde{t}_\nu, b)) \end{pmatrix}.$$

Of course $Bw^{(\nu)}$ is to be replaced by $Bw^{(\nu)} + d^{(\nu)}$ in the above algorithm.

6.3 American options

Transformation to simpler system of (in)equalities We now turn to the computation of American put option prices using finite differences. Recall that the partial differential equation (3.26, 3.27) is to be replaced by the linear complementarity problem (3.36). PDE (3.26, 3.27) turned into (6.1, 6.2) after transformation. Using the same change of variables, (3.36) can be rewritten as linear complementarity problem

$$\begin{aligned} \left(\frac{\partial y}{\partial \widetilde{t}} - \frac{\partial^2 y}{\partial \widetilde{x}^2}\right)(y-g) &= 0 \\ \frac{\partial y}{\partial \widetilde{t}} - \frac{\partial^2 y}{\partial \widetilde{x}^2} &\geq 0, \quad y - g \geq 0 \\ y(0,\widetilde{x}) &= g(0,\widetilde{x}) \end{aligned} \tag{6.23}$$

with

$$g(\widetilde{t},\widetilde{x}) = \begin{cases} \exp((q+1)^2\widetilde{t}/4)\max\{e^{\frac{\widetilde{x}}{2}(q+1)} - e^{\frac{\widetilde{x}}{2}(q-1)}, 0\} & \text{for a call,} \\ \exp((q+1)^2\widetilde{t}/4)\max\{e^{\frac{\widetilde{x}}{2}(q-1)} - e^{\frac{\widetilde{x}}{2}(q+1)}, 0\} & \text{for a put} \end{cases} \tag{6.24}$$

and $q = 2r/\sigma^2$. Similarly as in the previous section, the behaviour at the boundaries can be approximated as

$$\begin{aligned} y(\widetilde{t},\widetilde{x}) &\approx g(\widetilde{t},\widetilde{x}) \text{ for } \widetilde{x} \to -\infty, \\ y(\widetilde{t},\widetilde{x}) &\approx g(\widetilde{t},\widetilde{x}) \text{ for } \widetilde{x} \to \infty. \end{aligned}$$

Discretisation using finite differences Recall that we discretised the heat equation (6.1, 6.2) as

$$\frac{w_{\nu+1,i} - w_{\nu i}}{\Delta\widetilde{t}} - \vartheta\frac{w_{\nu+1,i+1} - 2w_{\nu+1,i} + w_{\nu+1,i-1}}{\Delta\widetilde{x}^2} - (1-\vartheta)\frac{w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1}}{\Delta\widetilde{x}^2} = 0$$

with $\vartheta = 0$ for the explicit, $\vartheta = 1$ for the implicit, and $\vartheta = \frac{1}{2}$ for the Crank-Nicolson scheme. Accordingly, the first inequality in (6.23) naturally turns into

$$\begin{aligned} & w_{\nu+1,i} - \lambda\vartheta(w_{\nu+1,i+1} - 2w_{\nu+1,i} + w_{\nu+1,i-1}) \\ & - w_{\nu i} - \lambda(1-\vartheta)(w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1}) \quad \geq \quad 0 \end{aligned} \tag{6.25}$$

with $\lambda = \Delta\widetilde{t}/\Delta\widetilde{x}^2$ after discretisation. As in Section 6.1 vector notation helps to solve the problem. To this end let $b^{(\nu)} = (b_{\nu 1}, \ldots, b_{\nu,m-1})^\top$ with

$$b_{\nu i} = w_{\nu i} + \lambda(1-\vartheta)(w_{\nu,i+1} - 2w_{\nu i} + w_{\nu,i-1}), \quad i = 2, \ldots m-2 \tag{6.26}$$

$b_{\nu 1}$ and $b_{\nu,m-1}$ will be chosen later in order to incorporate the boundary conditions. Moreover, set $w^{(\nu)} = (w_{\nu 1}, \ldots, w_{\nu,m-1})^\top$ as before and $g^{(\nu)} = (g_{\nu 1}, \ldots, g_{\nu,m-1})^\top$ with

$$g_{\nu i} = g(\widetilde{t}_\nu, \widetilde{x}_i). \tag{6.27}$$

Finally setting

$$A = \begin{pmatrix} 1+2\lambda\vartheta & -\lambda\vartheta & 0 & \ldots & 0 \\ -\lambda\vartheta & 1+2\lambda\vartheta & -\lambda\vartheta & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -\lambda\vartheta & 1+2\lambda\vartheta & -\lambda\vartheta \\ 0 & \ldots & 0 & -\lambda\vartheta & 1+2\lambda\vartheta \end{pmatrix} \in \mathbb{R}^{(m-1)\times(m-1)} \tag{6.28}$$

(6.25) can be written as the componentwise inequality

$$Aw^{(\nu+1)} \geq b^{(\nu)}, \quad \nu = 0, 1, \ldots$$

Inequality $y - g \geq 0$ in (6.23) turns into $w^{(\nu)} \geq g^{(\nu)}$. Accordingly,

$$\left(\frac{\partial y}{\partial \widetilde{t}} - \frac{\partial^2 y}{\partial \widetilde{x}^2}\right)(y - g) = 0$$

is expressed as

$$\left(Aw^{(\nu+1)} - b^{(\nu)}\right)^\top \left(w^{(\nu+1)} - g^{(\nu+1)}\right) = 0$$

after discretisation. The initial condition $y(0, \widetilde{x}) = g(0, \widetilde{x})$ turns into $w^{(0)} = g^{(0)}$, which means $w_{0i} = g_{0i}$ for $i = 1, \ldots, m-1$. The approximative behaviour at the boundaries translates into $w_{\nu 0} = g_{\nu 0}$ and $w_{m\nu} = g_{m\nu}$ for $\nu = 0, 1, 2, \ldots$ Similarly as in the previous section, these boundary conditions are taken care of by choosing $b_{\nu 1}, b_{\nu,m-1}$ appropriately, namely as

$$\begin{aligned} b_{\nu 1} &= w_{\nu 1} + \lambda(1-\vartheta)(w_{\nu 2} - 2w_{\nu 1} + g_{\nu 0}) + \lambda\vartheta g_{\nu+1,0}, \\ b_{\nu,m-1} &= w_{\nu,m-1} + \lambda(1-\vartheta)(g_{\nu m} - 2w_{\nu,m-1} + w_{\nu,m-2}) + \lambda\vartheta g_{\nu+1,m}. \end{aligned} \tag{6.29}$$

In the explicit scheme we have $\vartheta = 0$ and hence $A = 1$. In order to compute American option prices, we may simply replace the assignment $w = (v_1, \ldots, v_n)$ in the algorithm in Section 6.1 by

$$w_i = \max(v_i, g_{\nu+1,i}), \quad i = 1, \ldots, m-1,$$

ideally modified by the corrected boundary conditions as discussed at the end of Section 6.2.

For general ϑ we obtain instead the following algorithm for American options, which replaces the loop in the algorithm close to the end of Section 6.2.

Loop For $\nu = 0, \ldots, \nu_{\max} - 1$:

Compute $g = g^{(\nu+1)}$ according to (6.27).

Compute $b = b^{(\nu)}$ according to (6.26, 6.29).

Solve

$$Aw - b \geq 0, \quad w \geq g, \quad (Aw - b)^\top (w - g) = 0 \tag{6.30}$$

for w.

Set $w^{(\nu+1)} = w$.

In the sequel we turn to the numerical solution of the linear complementary problem (6.30), which replaces (6.22).

Brennan-Schwartz algorithm In order to solve (6.30) for $\vartheta \neq 0$, we can use the algorithm of [BS77]. In [JLL90] it is shown to provide correct results for call and put options. It relies on the following result.

Lemma 6.2 *Suppose that we want to solve*

$$Ax - b \geq 0, \quad x \geq g, \quad (Ax - b)^\top (x - g) = 0 \tag{6.31}$$

for a tridiagonal matrix $A \in \mathbb{R}^{n\times n}$ as in (6.15) and vectors $b, g \in \mathbb{R}^n$ such that $\beta_i, \gamma_i \leq 0$ for all i and

$$\min\{\alpha_1, \ldots, \alpha_n\} + \min\{\beta_1, \ldots, \beta_{n-1}\} + \min\{\gamma_2, \ldots, \gamma_n\} > 0 \tag{6.32}$$

In addition we assume that there is some $k \in \{1, \ldots, n\}$ such that the solution x to (6.31) satisfies $x_i > g_i$ for $i = 1, \ldots, k-1$ and $x_i = g_i$ for $i = k, \ldots, n$. Then the solution to (6.31) is obtained as pointwise maximum of g and x from (6.17, 6.18).

More specifically, setting

$$\begin{aligned}
\hat{\alpha}_1 &= \alpha_1, \\
\hat{b}_1 &= b_1, \\
\hat{\alpha}_i &= \alpha_i - \frac{\gamma_i}{\hat{\alpha}_{i-1}}\beta_{i-1}, \quad i = 2, \ldots, n, \\
\hat{b}_i &= b_i - \frac{\gamma_i}{\hat{\alpha}_{i-1}}\hat{b}_{i-1}, \quad i = 2, \ldots, n, \\
x_n &= \max\left\{\frac{\hat{b}_n}{\hat{\alpha}_n}, g_n\right\}, \qquad (6.33) \\
x_i &= \max\left\{\frac{\hat{b}_i - \beta_i x_{i+1}}{\hat{\alpha}_i}, g_i\right\}, \quad i = n-1, \ldots, 1, \qquad (6.34)
\end{aligned}$$

we have that $x = (x_1, \ldots, x_n)$ is the unique solution to (6.31).

Proof. Condition (6.32) warrants that $x^\top Ax \geq 0$ for any $x \in \mathbb{R}^n$, which in turn implies that (6.31) has a unique solution, cf. [JLL90, Lemma 5.9 and Theorem 5.1]. Moreover, this condition allows to verify $\hat{\alpha}_i \geq \alpha_i/2 > 0$, $i = 1, \ldots, n$ by induction. Again by induction we obtain that $x = (x_1, \ldots, x_n)$ in (6.33, 6.34) is the unique solution to

$$\hat{A}x - \hat{b} \geq 0, \quad x \geq g, \quad (\hat{A}x - \hat{b})^\top (x - g) = 0, \tag{6.35}$$

where $\hat{A}$ is the matrix in (6.16) and $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_n)$.

Note from the construction that there are nonnegative numbers $\lambda_{i,j}$ such that

$$\begin{aligned}
\hat{b}_i &= b_i + \lambda_{i,i-1}b_{i-1} + \cdots + \lambda_{i,1}b_1, \\
(\hat{A}y)_i &= (Ay)_i + \lambda_{i,i-1}(Ay)_{i-1} + \cdots + \lambda_{i,1}(Ay)_1
\end{aligned}$$

for $i = 2, \ldots, n$ and any $y \in \mathbb{R}^n$. Hence, $(Ax)_i \geq b_i$ implies $(\hat{A}x)_i \geq \hat{b}_i$ for any i. Moreover, $(Ax)_i = b_i$ for $i = 1, \ldots, k-1$ implies $(\hat{A}x)_i = \hat{b}_i$ for $i = 1, \ldots, k-1$. Therefore the solution

to (6.31) solves (6.35) as well. □

For the American call option it is known that there is some threshold $x_0(t)$ such that the PDE in (3.36) holds for $x \leq x_0(t)$ whereas $v(t,x) = g(t,x)$ is satisfied for $x \geq x_0(t)$. This property can be shown to carry over to function $y(\widetilde{t}, \widetilde{x})$ in (6.23) and eventually to its discretised version (6.31), cf. [JLL90]. Therefore we can proceed as indicated above.

For the American put we have instead that, for some $x_0(t)$, the PDE in (3.36) holds for $x \geq x_0(t)$ whereas $v(t,x) = g(t,x)$ is satisfied for $x \leq x_0(t)$. Therefore, a modification of the algorithm is needed where the variables β_i instead of γ_i are removed in the first step. It is proved along the same lines a Lemma 6.2.

Lemma 6.3 *In the same setup as in Lemma 6.2 we suppose that there is some $k \in \{1, \ldots, n\}$ such that the solution x to (6.31) satisfies $x_i = g_i$ for $i = 1, \ldots, k$ and $x_i > g_i$ for $i = k-1, \ldots, n$. Set*

$$\begin{aligned}
\hat{\alpha}_n &= \alpha_n, \\
\hat{b}_n &= b_n, \\
\hat{\alpha}_i &= \alpha_i - \frac{\beta_i}{\hat{\alpha}_{i+1}}\gamma_{i+1}, \quad i = n-1, \ldots, 1, \\
\hat{b}_i &= b_i - \frac{\beta_i}{\hat{\alpha}_{i+1}}\hat{b}_{i+1}, \quad i = n-1, \ldots, 1, \\
x_1 &= \max\left\{\frac{\hat{b}_1}{\hat{\alpha}_1}, g_1\right\}, \\
x_i &= \max\left\{\frac{\hat{b}_i - \gamma_i x_{i-1}}{\hat{\alpha}_i}, g_i\right\}, \quad i = 2, \ldots, n.
\end{aligned}$$

Then $x = (x_1, \ldots, x_n)$ is the unique solution to (6.31).

Algorithm in pseudo code Altogether, we have the following algorithm for computing American put prices in the Black-Scholes model.

Input Model parameters: r, σ

Option parameters: K, T

Further parameters: ϑ (e.g. $\vartheta = 1/2$ for Crank-Nicolson), $a, b, m, \nu_{\max}$ (e.g. $a = -5$, $b = 5$, $m = \nu_{\max} = 100$),

Compute $\Delta\widetilde{x} = (b-a)/m$

$\Delta\widetilde{t} = \sigma^2 T/(2\nu_{\max})$

$\widetilde{x}_i = a + i\Delta\widetilde{x}$ for $i = 0, \ldots, m$

$\lambda = \Delta\widetilde{t}/\Delta\widetilde{x}^2$

Set $w = g^{(0)} = (g(0, \widetilde{x}_1), \ldots, g(0, \widetilde{x}_{m-1}))$ with g fron (6.24).

$\widetilde{t}$-loop For $\nu = 0, 1, \ldots, \nu_{\max} - 1$:

$\widetilde{t}_\nu = \nu\Delta\widetilde{t}$

$g_{\nu i} = g(\widetilde{t}_\nu, \widetilde{x}_i), g_{\nu+1,i} = g(\widetilde{t}_{\nu+1}, \widetilde{x}_i)$ for $i = 0, \ldots, m$

$b_i = w_i + \lambda(1-\vartheta)(w_{i+1} - 2w_i + w_{i-1})$ for $2 \le i \le m-2$

$b_1 = w_1 + \lambda(1-\vartheta)(w_2 - 2w_1 + g_{\nu 0}) + \lambda\vartheta g_{\nu+1,0}$

$b_{m-1} = w_{m-1} + \lambda(1-\vartheta)(g_{\nu m} - 2w_{m-1} + w_{m-2}) + \lambda\vartheta g_{\nu+1,m}$

Let A be as in (6.28). Solve

$$Ax - b \ge 0, \quad x \ge g^{(\nu+1)}, \quad (Ax-b)^\top (x - g^{(\nu+1)}) = 0$$

for x, as indicated in Lemma 6.3.

Set $w^{(\nu+1)} = x$.

Compute for $i = 1, \ldots, m-1$:

$S_i = Ke^{\widetilde{x}_i}$

$v(0, S_i) = Kw_i \exp(-\frac{\widetilde{x}_i}{2}(q-1) - \frac{1}{2}\sigma^2 T((q-1)^2/4 + q))$

Output $(S_i, v(0, S_i))_{i=1,\ldots,m-1}$

The output of this algorithm are $m-1$ pairs of initial stock prices and the corresponding American put prices.

6.4 Accuracy and extrapolation

The computation of option prices is subject to many possible errors which can be estimated only partially. First of all the Black-Scholes model provides only a very rough approximation to real markets. This is more or less true for any model. Moreover, parameters are subject to estimation error. In the finite difference method, discretisation of the derivatives and truncation of the domain lead to numerical errors. These are also involved in the iterative solution of linear (in)equalities and due to roundoff errors.

Elementary error estimates Here we try to provide elementary estimates for the discretisation error. Let us distinguish the "true" solution η^* and its approximation $\eta = \eta(\Delta)$, which depends on a mesh size or some other crucial parameter Δ. The goal is to determine Δ such that $|\eta(\Delta) - \eta^*| < \varepsilon$ for some given error bound ε. To this end, we need to determine $\eta(\Delta) - \eta^*$ as a function of Δ. Let us assume that the numerical scheme is of order p in the sense that $\eta(\Delta) - \eta^* \approx \gamma \Delta^p$ for small Δ, where γ denotes some constant. If we run the numerical schemes with two different parameters Δ_1, Δ_2, we obtain

$$\begin{aligned} \eta_1 = \eta(\Delta_1) &\approx \eta^* + \gamma \Delta_1^p \\ \eta_2 = \eta(\Delta_2) &\approx \eta^* + \gamma \Delta_2^p, \end{aligned} \tag{6.36}$$

which leads to

$$\gamma \approx \frac{\eta_1 - \eta_2}{\Delta_1^p - \Delta_2^p}.$$

For $\Delta_2 = \Delta_1/2$ we have

$$\gamma \Delta_2^p \approx \frac{\eta_1 - \eta_2}{2^p - 1}, \tag{6.37}$$

which yields

$$\gamma \Delta_2^2 \approx \frac{\eta_1 - \eta_2}{3}$$

for $p = 2$ as in the Crank-Nicolson scheme. In particular, we obtain the error estimate

$$|\eta_2 - \eta^*| \approx \frac{|\eta_1 - \eta_2|}{3}.$$

Richardson extrapolation As in Section 5.1, we may use this approximation in order to improve our estimate. (6.36) and (6.37) yield

$$\eta^* \approx \eta_2 - \gamma \Delta_2^p \approx \eta_2 - \frac{\eta_1 - \eta_2}{\Delta_1^p - \Delta_2^p} \Delta_2^p$$

and in particular

$$\eta^* \approx \eta_2 - \frac{\eta_1 - \eta_2}{3} = \frac{4\eta_2 - \eta_1}{3}$$

for $p = 2$ and $\Delta_2 = \Delta_1/2$. This often leads to a considerably improved approximation. However, we lose the information on its precision.

6.5 The Heston model

In this section, we consider European option pricing in the slightly more involved Heston model. Instead of PDE (3.26, 3.27), we must consider (3.42, 3.43), which has an additional state variable γ. In order to obtain a slightly simpler equation, we change variables to $\widetilde{x} = \log(x) - rt$, $\widetilde{t} = T - t$, $\widetilde{\gamma} = \gamma$, and

$$y(\widetilde{t}, \widetilde{x}, \widetilde{\gamma}) = e^{-rt} v(t, x, \gamma) = e^{-r(T-\widetilde{t})} v(T - \widetilde{t}, e^{\widetilde{x} - r(T-\widetilde{t})}, \widetilde{\gamma})$$

Straightforward calculations yield that PDE (3.42, 3.43) turns into

$$\begin{aligned}\frac{\partial}{\partial\widetilde{t}}y(\widetilde{t},\widetilde{x},\widetilde{\gamma}) &= \frac{1}{2}\widetilde{\gamma}\frac{\partial^2}{\partial\widetilde{x}^2}y(\widetilde{t},\widetilde{x},\widetilde{\gamma}) - \frac{1}{2}\widetilde{\gamma}\frac{\partial}{\partial\widetilde{x}}y(\widetilde{t},\widetilde{x},\widetilde{\gamma}) \\ &\quad + \frac{1}{2}\widetilde{\sigma}^2\frac{\partial^2}{\partial\widetilde{\gamma}^2}y(\widetilde{t},\widetilde{x},\widetilde{\gamma}) + \widetilde{\sigma}^2\frac{\partial}{\partial\widetilde{\gamma}}y(\widetilde{t},\widetilde{x},\widetilde{\gamma}), \qquad (6.38)\\ y(0,\widetilde{x},\widetilde{\gamma}) &= f(e^{\widetilde{x}-rT}). \qquad (6.39)\end{aligned}$$

which must be solved for $0 \le \widetilde{t} \le T$, $-\infty < \widetilde{x} < \infty$, $0 < \widetilde{\gamma} < \infty$. Our goal is to obtain a useful approximation of $y_{k,(i,j)} = y(\widetilde{t}_k, \widetilde{x}_i, \widetilde{\gamma}_j)$ on a grid $k = 0,\dots,k_{\max}$, $i = 1,\dots,i_{\max}$, $j = 1,\dots,j_{\max}$. By discretising the derivatives as before one can formulate an explicit scheme which, however, faces stability restrictions as in the Black-Scholes case. Implicit and Crank-Nicolson schemes can be deduced as well. However, since the respective matrices are no longer tridiagonal, numerical solution of the corresponding linear equations is more involved.

ADI scheme

As a way out we discuss the so-called *ADI (alternating directions implicit)* scheme, which proceeds by introducing an intermediate time step $\widetilde{t}_{k+\frac{1}{2}}$. The step $\widetilde{t}_k \to \widetilde{t}_{k+\frac{1}{2}}$ is carried out implictly in $\widetilde{x}$ and explicitly in $\widetilde{\gamma}$. The second half step $\widetilde{t}_{k+\frac{1}{2}} \to \widetilde{t}_{k+1}$, on the other hand, is explicit in $\widetilde{x}$ and implicit in $\widetilde{\gamma}$. To be more specific, the $\widetilde{t}$-derivative $\frac{\partial y}{\partial\widetilde{t}}$ is discretised as

$$\frac{y_{k+\frac{1}{2},(i,j)} - y_{k,(i,j)}}{\Delta\widetilde{t}/2} \quad \text{for } \widetilde{t}_k \to \widetilde{t}_{k+\frac{1}{2}}$$

and as

$$\frac{y_{k+1,(i,j)} - y_{k+\frac{1}{2},(i,j)}}{\Delta\widetilde{t}/2} \quad \text{for } \widetilde{t}_{k+\frac{1}{2}} \to \widetilde{t}_{k+1}.$$

The $\widetilde{x}$-derivatives are discretised only at the intermediate time steps, namely as

$$\begin{aligned}\frac{\partial^2 y}{\partial\widetilde{x}^2}(\widetilde{t}_{k+\frac{1}{2}},\widetilde{x}_i,\widetilde{\gamma}_j) &\approx \frac{y_{k+\frac{1}{2},(i+1,j)} - 2y_{k+\frac{1}{2},(i,j)} + y_{k+\frac{1}{2},(i-1,j)}}{\Delta\widetilde{x}^2}, \qquad (6.40)\\ \frac{\partial y}{\partial\widetilde{x}}(\widetilde{t}_{k+\frac{1}{2}},\widetilde{x}_i,\widetilde{\gamma}_j) &\approx \frac{y_{k+\frac{1}{2},(i+1,j)} - y_{k+\frac{1}{2},(i-1,j)}}{2\Delta\widetilde{x}}. \qquad (6.41)\end{aligned}$$

The $\widetilde{\gamma}$-derivatives, on the other hand, are considered only at integer time steps $\widetilde{t}_k$, with discretisations

$$\begin{aligned}\frac{\partial^2 y}{\partial\widetilde{\gamma}^2}(\widetilde{t}_k,\widetilde{x}_i,\widetilde{\gamma}_j) &\approx \frac{y_{k,(i,j+1)} - 2y_{k,(i,j)} + y_{k,(i,j-1)}}{\Delta\widetilde{\gamma}^2}, \qquad (6.42)\\ \frac{\partial^2 y}{\partial\widetilde{\gamma}^2}(\widetilde{t}_{k+1},\widetilde{x}_i,\widetilde{\gamma}_j) &\approx \frac{y_{k+1,(i,j+1)} - 2y_{k+1,(i,j)} + y_{k+1,(i,j-1)}}{\Delta\widetilde{\gamma}^2}, \qquad (6.43)\\ \frac{\partial y}{\partial\widetilde{\gamma}}(\widetilde{t}_k,\widetilde{x}_i,\widetilde{\gamma}_j) &\approx \frac{y_{k,(i,j+1)} - y_{k,(i,j-1)}}{2\Delta\widetilde{\gamma}}, \qquad (6.44)\\ \frac{\partial y}{\partial\widetilde{\gamma}}(\widetilde{t}_{k+1},\widetilde{x}_i,\widetilde{\gamma}_j) &\approx \frac{y_{k+1,(i,j+1)} - y_{k+1,(i,j-1)}}{2\Delta\widetilde{\gamma}}. \qquad (6.45)\end{aligned}$$

In the first half step $\widetilde{t}_k \to \widetilde{t}_{k+\frac{1}{2}}$, we use (6.40, 6.41) for the $\widetilde{x}$- and (6.42, 6.44) for the $\widetilde{\gamma}$-derivatives. Similarly, we use we use (6.40, 6.41) for the $\widetilde{x}$- and (6.43, 6.45) for the $\widetilde{\gamma}$-derivatives in the second half step $\widetilde{t}_{k+\frac{1}{2}} \to \widetilde{t}_{k+1}$. If we denote the numerical approximation of $y_{k,(i,j)}$ by $w_{k,(i,j)}$ similarly as before, we obtain

$$\begin{aligned}\frac{w_{k+\frac{1}{2},(i,j)} - w_{k,(i,j)}}{\Delta\widetilde{t}/2} &= \frac{1}{2}\widetilde{\gamma}_j \frac{w_{k+\frac{1}{2},(i+1,j)} - 2w_{k+\frac{1}{2},(i,j)} + w_{k+\frac{1}{2},(i-1,j)}}{\Delta\widetilde{x}^2} \\ &\quad - \frac{1}{2}\widetilde{\gamma}_j \frac{w_{k+\frac{1}{2},(i+1,j)} - w_{k+\frac{1}{2},(i-1,j)}}{2\Delta\widetilde{x}} \\ &\quad + \frac{1}{2}\widetilde{\sigma}^2 \frac{w_{k,(i,j+1)} - 2w_{k,(i,j)} + w_{k,(i,j-1)}}{\Delta\widetilde{\gamma}^2} \\ &\quad + \widetilde{\sigma}^2 \frac{w_{k,(i,j+1)} - w_{k,(i,j-1)}}{2\Delta\widetilde{\gamma}} \end{aligned} \tag{6.46}$$

for the first half step $\widetilde{t}_k \to \widetilde{t}_{k+\frac{1}{2}}$ and

$$\begin{aligned}\frac{w_{k+1,(i,j)} - w_{k+\frac{1}{2},(i,j)}}{\Delta\widetilde{t}/2} &= \frac{1}{2}\widetilde{\gamma}_j \frac{w_{k+\frac{1}{2},(i+1,j)} - 2w_{k+\frac{1}{2},(i,j)} + w_{k+\frac{1}{2},(i-1,j)}}{\Delta\widetilde{x}^2} \\ &\quad - \frac{1}{2}\widetilde{\gamma}_j \frac{w_{k+\frac{1}{2},(i+1,j)} - w_{k+\frac{1}{2},(i-1,j)}}{2\Delta\widetilde{x}} \\ &\quad + \frac{1}{2}\widetilde{\sigma}^2 \frac{w_{k+1,(i,j+1)} - 2w_{k+1,(i,j)} + w_{k+1,(i,j-1)}}{\Delta\widetilde{\gamma}^2} \\ &\quad + \widetilde{\sigma}^2 \frac{w_{k+1,(i,j+1)} - w_{k+1,(i,j-1)}}{2\Delta\widetilde{\gamma}} \end{aligned} \tag{6.47}$$

for the second half step $\widetilde{t}_{k+\frac{1}{2}} \to \widetilde{t}_{k+1}$.

Both (6.46) and (6.47) can be expressed in terms of linear equations with tridiagonal matrices. For fixed j we have

$$A^{(1)}(w_{k+\frac{1}{2},(i,j)})_{i=1,\dots,i_{\max}} = (b^{(1)}_{(i,j)})_{i=1,\dots,i_{\max}}$$

with

$$A^{(1)} = \begin{pmatrix} 1+\lambda\widetilde{\gamma}_j/2 & -\lambda\widetilde{\gamma}_j(1-\frac{\Delta\widetilde{x}}{2})/4 & 0 & \dots & 0 \\ -\lambda\widetilde{\gamma}_j(1+\frac{\Delta\widetilde{x}}{2})/4 & 1+\lambda\widetilde{\gamma}_j/2 & -\lambda\widetilde{\gamma}_j(1-\frac{\Delta\widetilde{x}}{2})/4 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -\lambda\widetilde{\gamma}_j(1+\frac{\Delta\widetilde{x}}{2})/4 & 1+\lambda\widetilde{\gamma}_j/2 & -\lambda\widetilde{\gamma}_j(1-\frac{\Delta\widetilde{x}}{2})/4 \\ 0 & \dots & 0 & -\lambda\widetilde{\gamma}_j(1+\frac{\Delta\widetilde{x}}{2})/4 & 1+\lambda\widetilde{\gamma}_j/2 \end{pmatrix}$$

and $\lambda = \Delta\widetilde{t}/\Delta\widetilde{x}^2$. The matrix $b^{(1)}$ is obtained as

$$(b^{(1)}_{(i,j)})_{j=1,\dots,j_{\max}} = B^{(1)}(w_{k,(i,j)})_{j=1,\dots,j_{\max}}$$

for fixed i, where

$$B^{(1)} = \begin{pmatrix} 1-\mu\widetilde{\sigma}^2/2 & \mu\widetilde{\sigma}^2(1+\Delta\widetilde{\gamma})/4 & 0 & \dots & 0 \\ \mu\widetilde{\sigma}^2(1-\Delta\widetilde{\gamma})/4 & 1-\mu\widetilde{\sigma}^2/2 & \mu\widetilde{\sigma}^2(1+\Delta\widetilde{\gamma})/4 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \mu\widetilde{\sigma}^2(1-\Delta\widetilde{\gamma})/4 & 1-\mu\widetilde{\sigma}^2/2 & \mu\widetilde{\sigma}^2(1+\Delta\widetilde{\gamma})/4 \\ 0 & \dots & 0 & \mu\widetilde{\sigma}^2(1-\Delta\widetilde{\gamma})/4 & 1-\mu\widetilde{\sigma}^2/2 \end{pmatrix}$$

with $\mu = \Delta\widetilde{t}/\Delta\widetilde{\gamma}^2$. The second half step works similarly. For fixed i we have

$$A^{(2)}(w_{k+1,(i,j)})_{j=1,\dots,j_{\max}} = (b^{(2)}_{(i,j)})_{j=1,\dots,j_{\max}}$$

with

$$A^{(2)} = \begin{pmatrix} 1+\mu\widetilde{\sigma}^2/2 & -\mu\widetilde{\sigma}^2(1+\Delta\widetilde{\gamma})/4 & 0 & \dots & 0 \\ -\mu\widetilde{\sigma}^2(1-\Delta\widetilde{\gamma})/4 & 1+\mu\widetilde{\sigma}^2/2 & -\mu\widetilde{\sigma}^2(1+\Delta\widetilde{\gamma})/4 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -\mu\widetilde{\sigma}^2(1-\Delta\widetilde{\gamma})/4 & 1+\mu\widetilde{\sigma}^2/2 & -\mu\widetilde{\sigma}^2(1+\Delta\widetilde{\gamma})/4 \\ 0 & \dots & 0 & -\mu\widetilde{\sigma}^2(1-\Delta\widetilde{\gamma})/4 & 1+\mu\widetilde{\sigma}^2/2 \end{pmatrix}$$

The matrix $b^{(2)}$ is obtained as

$$(b^{(2)}_{(i,j)})_{i=1,\dots,i_{\max}} = B^{(2)}(w_{k+\frac{1}{2},(i,j)})_{i=1,\dots,i_{\max}}$$

for fixed j, where

$$B^{(2)} = \begin{pmatrix} 1-\lambda\widetilde{\gamma}_j/2 & \lambda\widetilde{\gamma}_j(1-\frac{\Delta\widetilde{x}}{2})/4 & 0 & \dots & 0 \\ \lambda\widetilde{\gamma}_j(1+\frac{\Delta\widetilde{x}}{2})/4 & 1-\lambda\widetilde{\gamma}_j/2 & \lambda\widetilde{\gamma}_j(1-\frac{\Delta\widetilde{x}}{2})/4 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \lambda\widetilde{\gamma}_j(1+\frac{\Delta\widetilde{x}}{2})/4 & 1-\lambda\widetilde{\gamma}_j/2 & \lambda\widetilde{\gamma}_j(1-\frac{\Delta\widetilde{x}}{2})/4 \\ 0 & \dots & 0 & \lambda\widetilde{\gamma}_j(1+\frac{\Delta\widetilde{x}}{2})/4 & 1-\lambda\widetilde{\gamma}_j/2 \end{pmatrix}.$$

We do not discuss the incorporation of boundary conditions at this point, which may be based on limit considerations as in Section 6.2. Likewise, the ADI scheme can be modified for the solution of linear complementarity problems (3.44), which occur in the treatment of American options.

6.6 Outlook and discussion

Derivative pricing by PDE methods constitutes a standard tool for derivative pricing, hedging, and calibration. In low dimensions the method is fast and yields prices simultaneously for many

initial stock/strike price ratios. In contrast to numerical integration and Monte Carlo methods, American options can be treated with essentially the same effort. On the other hand, the approach becomes increasingly involved and costly in the case of multiple state variables and hence dimensions. This curse of dimensionality partly explains the popularity of Monte Carlo methods in Finance. Asset price models with jumps lead to *partial integro-differential equations (PIDE)* rather than PDE's. They contain integral and hence nonlocal operators. The matrices in the discretisation are no longer tridiagonal, which again complicates their numerical solution.

Finally, we want to mention an alternative, generally much more flexible and powerful approach to solving PDE's and also PIDE's, namely the *finite element method (FEM)*. Let us briefly explain the key idea by considering a simple toy problem. Suppose that we look for a numerical approximation of the solution $u : [0,1] \to \mathbb{R}$ to the boundary value problem

$$u''(x) = f(x), \tag{6.48}$$

$$u(0) = u(1) = 0 \tag{6.49}$$

The finite element method is based on the *weak formulation* or *variational formulation* of the equation. After multiplication with a test function $v : [0,1] \to \mathbb{R}$ and integration, equation (6.48) turns into

$$\int_0^1 u''(x)v(x)dx = \int_0^1 f(x)v(x)dx. \tag{6.50}$$

If (6.50) holds for all functions v from a sufficiently large space, any solution u to (6.50) solves (6.48) as well. We focus on test functions satisfying $v(0) = v(1) = 0$. The left-hand side of (6.50) can then be modified using integration by parts:

$$\begin{aligned}\int_0^1 u''(x)v(x)dx &= u'(1)v(1) - u'(0)v(0) - \int_0^1 u'(x)v'(x)dx \\ &= -\int_0^1 u'(x)v'(x)dx.\end{aligned}$$

This yields

$$-\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx$$

or, equivalently,

$$\varphi(u,v) = \langle f, v\rangle \tag{6.51}$$

if we define

$$\varphi(g,h) = -\int_0^1 g'(x)h'(x)dx$$

and

$$\langle g, h\rangle = \int_0^1 g(x)h(x)dx.$$

In order to discretise the problem we consider only functions from a finite-dimensional subspace, e.g. the n-dimensional space V of piecewise linear functions with grid points $x_1, \ldots, x_n$.

On the other hand, we require (6.51) to hold only for test functions from an n-dimensional space, e.g. the same space V. In other words, the *discrete weak solution* $w \in V$ is defined here as the solution to

$$\varphi(w, v) = \langle f, v \rangle, \quad v \in V.$$

If the dimension n of the subspace tends to ∞, we can hope for convergence of this discrete weak solution to the solution of (6.50). Denote by $w_1, \ldots, w_n$ basis functions of V. Then the coefficients $c_1, \ldots, c_n$ of the discrete weak solution $w = \sum_{i=1}^{n} c_i w_i$ satisfy

$$\sum_{i=1}^{n} c_i \varphi(w_i, w_j) = \langle f, w_j \rangle, \quad j = 1, \ldots, n. \tag{6.52}$$

(6.52) is a linear system of n equations with n unknowns. Consequently, we need to solve linear equations similarly as for the finite difference method. But even if the two approaches ultimately lead to similar equations in this simple example, the finite element method is generally much more flexible and powerful than the finite difference approach. Since the approximation of derivatives is avoided, it is possible to work with irregular grids, which can be chosen adaptively depending on the local discretisation error.

Chapter 7

Machine Learning

In recent years, machine learning techniques have been increasingly applied to questions in financial mathematics. In this chapter, we will give a brief insight into such approaches. More specifically, we will deal with issues of data-driven hedging. Many other fields, such as portfolio optimization, can also be covered. However, this goes too far for this course.

7.1 Artificial Neural Networks

Many machine learning methods are based on the use of so-called artificial neural networks, which we will introduce below.

7.1.1 Motivation from the human brain

Artificial Neural Networks (ANNs) are inspired by the structure of the human brain (although they are not a suitable model for the brain from a biological perspective).

In biology, a *neuron* is a is an electrically excitable cell that communicates with other cells via specialized connections called synapses.

There are

$10^{10} - 10^{11}$ neurons in the brain,
$5000 - 50.000$ synapses per neuron,
overall $> 10^{14}$ connections.

A neuron works with electrical potential that depends on the input signals. When a certain threshold is exceeded, the neuron fires and sends an amount of charge along the axon.

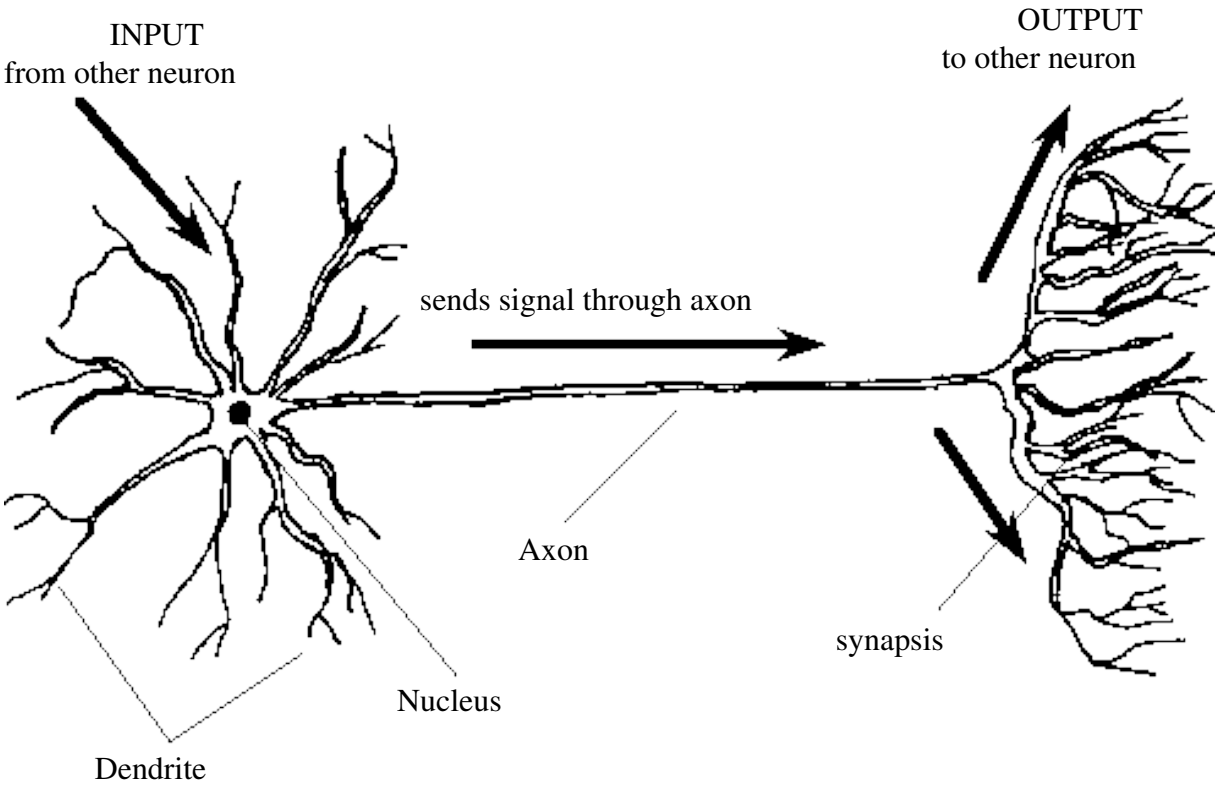


Figure 7.1: Schematic representation of a neuron

We may interpret it as a directed graph where

- knots are associated to neurons,
- edges are associated to connections.

7.1.2 Survey on neural network structures and deep learning

Perceptron The idea of perceptron goes back to [**?**] and [**?**]. The corresponding function class comprises functions of the form

$$g\colon \mathbb{R}^d \to \{-1,1\}, \quad g(\mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x} + b), \quad \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}, \quad \text{with } \sigma(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0 \end{cases},$$

that is, one considers separation by a halfspace.

Shallow neural networks A shallow neural network with one output is a function $f\colon \mathbb{R}^d \to \mathbb{R}$ of the form

$$g(\mathbf{x}) = \sum_{j=1}^{p} c_j \sigma(\mathbf{w}_j^\top \mathbf{x} + b_j), \quad \mathbf{w}_j \in \mathbb{R}^d,\ c_j, b_j \in \mathbb{R},$$

where p is also called number of *hidden units*, $\mathbf{w}_j \in \mathbb{R}^d$ are weight vectors, b_j are called *shifts/biases* and c_j are the coefficients. While a shallow neural network can be seen as a linear combination of perceptron functions, the involved *activation function* $\sigma\colon \mathbb{R} \to \mathbb{R}$ does not need to be the sign function. Few activation functions are widely used in practice:

- 80s/90s: *sigmoidal activations*, e.g., *logistic* $\sigma(x) = 1/(1+e^{-x})$
- more recent is the *ReLU* (*Rectifier Linear Unit*) activation function $\sigma(x) := x^+ := \max\{x, 0\}$
- variations include the *leaky ReLU* $\sigma(x) = ax1\{x < a\} + x1\{x \geq 0\}$, *softmax* $\sigma(x) = \log(1+e^x)$

The activation function σ and p will be fixed, while $\mathbf{w}_j, b_j, c_j$ are considered as free parameters that will be estimated from the data. The associated function class is *non-convex* in its parameters, and the question of shallow network fitting can be viewed as a dictionary learning problem.

Definition 7.1 For $\mathbf{v} = (v_1, \ldots, v_r)^\top$, $\mathbf{y} = (y_1, \ldots, y_r)^\top \in \mathbb{R}^r$, denote by $\sigma_{\mathbf{v}}\colon \mathbb{R}^r \to \mathbb{R}^r$ the *shifted activation function*, defined as

$$\sigma_{\mathbf{v}} \;=\; (\sigma(y_1 - v_1), \ldots, \sigma(y_r - v_r))^\top .$$

Fix a positive integer L (number of hidden layers/depth) and a width vector $\mathbf{p} = (p_0, \ldots, p_{L+1}) \in \mathbb{N}^{L+2}$. A *neural network* with *network architecture* $(L, \mathbf{p})$ is specified as

$$g(\mathbf{x}) \;=\; \varrho W_L \sigma_{\mathbf{v}_L} W_{L-1} \sigma_{\mathbf{v}_{L-1}} \;\cdots\; W_1 \sigma_{\mathbf{v}_1} W_0 \mathbf{x}, \tag{7.1}$$

where W_i is a $p_{i+1} \times p_i$ weight matrix, $\mathbf{v}_i \in \mathbb{R}^{p_i}$ is a shift vector and $\varrho\colon \mathbb{R}^r \to \mathbb{R}^s$ is an output activation function.

- A *shallow neural network* is a neural network with network architecture $(1, \mathbf{p})$ and output activation function $\varrho = \mathbf{id}$.
- If $L > 1$, the network is called *deep network*.

Neural networks are thus compositions of linear mappings and the nonlinear activation function.

Neural networks form a very flexible class of functions. With them, if the width of the hidden layer is allowed to become large, any continuous function can be approximated with arbitrary accuracy. For those interested, the exact result is given:

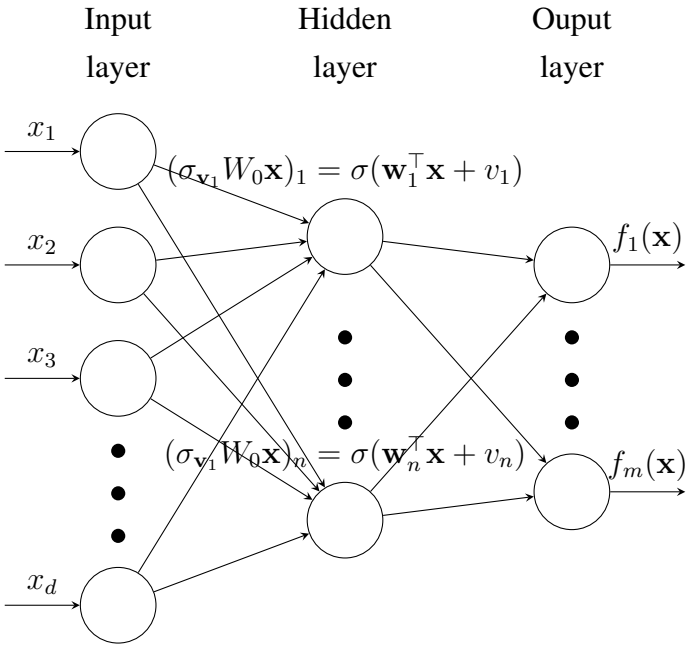


Figure 7.2: ANN with architecture $(L = 1, \mathbf{p} = (d, n, m))$

Theorem 7.2 (Universal approximation theorem) *Shallow networks with continuous and non-polynomial activation function $\sigma\colon \mathbb{R} \to \mathbb{R}$ have the* universal approximation property*, i.e., for any $\varepsilon > 0$ and any continuous function f on some compact set K, there exists an integer m such that*

$$\inf_{g \in F_m} \|f - g\|_\infty \leq \varepsilon,$$

where F_m denotes the set of functions representable by a shallow network with m neurons in the hidden layer.

While the Universal approximation theorem is a powerful result, it is important to note its limits. Firstly, the theorem is based on a single hidden layer network, and it is not always clear how many hidden layers and neurons are required to approximate a particular function. Additionally, the theorem does not guarantee that a neural network will learn the correct underlying function, see next subsection.

7.1.3 Training of ANN

Learning an ANN now consists of having data, usually in the form of pairs (x_j, y_j), $j = 1, ..., n$, and assuming that there is a relationship in the form of an unknown function f between the x_i and y_i such that $f(x_j) \approx y_j$. Now, a typical idea is to approximate f by neural nets of a certain structure. More precisely, one tries to minimize, for example, the expression $\sum_{i=1}^n (g(x_i) - y_i)^2$

over the class of neural networks g. More precisely, the optimization is performed over the parameters $\vartheta = (W_i, \mathbf{v}_i)$ in the representation of $g = g^\vartheta$, i.e., one has to minimize

$$\sum_{i=1}^{n} (g^\vartheta(x_i) - y_i)^2 \tag{7.2}$$

over all parameters of the neural net. This high-dimensional optimization task is naturally challenging. The great success of neural networks is now, partly, based on the fact that in them the optimization task can be performed reasonably efficiently. Keywords for this are *Backpropagation* and *Stochastic Gradient Decent*. However, we will not go into details here and refer to...

7.2 Deep Hedging

Since neural networks can be used very effectively to solve optimization problems, it is not surprising that they have also been increasingly used in the field of financial mathematics in recent years. We explain this here on the basis of hedging.

We follow the ideas presented in [**?**] and consider here an arbitrage-free market (perhaps also incomplete, with transaction costs, etc.) and want to hedge a claim X. To do this, one has to decide on the choice of the hedging portfolio at each point in time. The idea now is to consider this as a function f of some factors (current price, previous prices, volatility...) which we w.l.o.g. assume to be contained in the process S and to approximate f by neural networks g. As a criterion for the approximation quality we use the quadratic loss function. You already know this from the Mathematical Finance course. Here is the definition again as a reminder:

Definition 7.3 We call φ^* *variance-optimal hedging strategy* for X if $\varphi = \varphi^*$ minimizes the expected squared hedging error

$$E\Big((\hat{V}_N(\varphi) - \hat{X})^2\Big)$$

over all self-financing trading strategies φ.

The value V_N can be represented as $V_N(\varphi) = \pi + \varphi \bullet \hat{S}$, $\pi = V_0$ for the hedging strategy φ. So, we just consider φ_t, $t = 1, ..., N$, as a (shallow or deep) NN $g_t = g_t^\vartheta$ of the factors given above, i.e., restrict ourselves to

$$\varphi_t = g_t^\vartheta(S_{t-1}).$$

In contrast to the original problem, the restricted problem of minimizing

$$E\Big((\pi + g^\vartheta(S_{\cdot-1}) \bullet \hat{S}_N - \hat{X})^2\Big)$$

over all ϑ is a parametric problem.

Now an MC approach is used. By simulation one generates n paths $s_1^j, ..., s_N^j$, $j = 1, ..., n$, with corresponding discounted claim-payoffs $\hat{x}^j$ and replaces the expectation by the empirical counterpart: We minimize

$$\sum_{j=1}^{n} (\pi + g^{\vartheta}(s_{\cdot-1}^j) \bullet \hat{s}_N^j - \hat{x}^j)^2 \tag{7.3}$$

over all parameters ϑ of the neural nets, which is structurally the same as the optimization in (7.2) and therefore can be performed using standard algorithms.

This leads to the following algorithm:

Input number of time steps N,
process parameters,
specification of the claim and π,
number n of simulated paths,
architecture of the NN (Number of hidden layers L and width vector p),

Simulation Simulate n paths $s_1^j, ..., s_N^j$, $j = 1, ..., n$ (along the lines of Chapter 5)

Optimization Find minimizer ϑ^* in (7.3)

Output g^{ϑ^*}

Discussion

In this course, we have discussed different methods for pricing and hedging, each with its own advantages and disadvantages. The choice of which method to use for pricing and hedging depends on several factors, such as:

- The type and complexity of the derivative contract and its underlying asset
- The availability and reliability of data and market information
- The computational resources
- The mathematical and computational skills available
- The level of accuracy and stability required
- The trade-off between cost and benefit of using advanced methods

There is no definitive answer to which method is the best in all circumstances. Here is a brief overview:

Tree Methods

Tree methods are easy to implement and can handle various types of contracts, such as American options, exotic options, and interest rate derivatives. However, they also have some drawbacks, such as:

- They require a large number of nodes to achieve high accuracy, which increases the computational cost and memory usage.
- Curse of dimensionality
- They may suffer from numerical instability or convergence issues due to the discretization error or the choice of the branching probabilities.

- They may not be able to capture complex features of the underlying asset price dynamics, such as jumps, stochastic volatility, or correlation.

Integral Transform Methods

Integral transform methods are very elegant and efficient and can handle various types of contracts, such as European options, Asian options, and variance swaps. They can also incorporate complex features of the underlying asset price dynamics, such as jumps, stochastic volatility, or correlation. However, they also have some drawbacks, such as:

- They require a high level of mathematical sophistication and programming skills to implement and invert the integral transform.
- They may suffer from numerical instability or convergence issues due to the inversion error or the choice of the inversion method.
- They may not be able to handle early exercise features, such as American options, without using additional techniques.

Monte Carlo Methods

Monte Carlo methods are very flexible and can handle various types of contracts, such as path-dependent options, basket options, and credit derivatives. They can also incorporate complex features of the underlying asset price dynamics, such as jumps, stochastic volatility, or correlation. However, they also have some drawbacks, such as:

- They require a large number of paths to achieve high accuracy, which increases the computational cost and time.
- They may suffer from sampling error or bias due to the randomness of the simulation or the choice of the random number generator.
- They may not be able to handle early exercise features, such as American options, without using additional techniques.

PDE Methods

PDE methods are very accurate and stable and can handle various types of contracts, such as European options, barrier options, and convertible bonds. However, they also have some drawbacks, such as:

- They require a high level of mathematical sophistication and programming skills to implement and solve the PDE.
- Curse of dimensionality
- They may suffer from numerical instability or convergence issues due to the discretization error or the choice of the grid size.
- They may not be able to capture complex features of the underlying asset price dynamics, such as jumps, stochastic volatility, or correlation.

Machine Learning Methods

Machine learning methods are very powerful and can handle various types of contracts, such as exotic options, credit derivatives, and energy derivatives. They can also incorporate complex features of the underlying asset price dynamics, such as jumps, stochastic volatility, or correlation. However, they also have some drawbacks, such as:

- They require a large amount of data to train and validate the machine learning model, which may not be available or reliable.
- They may suffer from overfitting or underfitting issues due to the complexity or simplicity of the machine learning model or the choice of the hyperparameters.
- They may not be able to explain or interpret the logic or rationale behind their predictions.

Bibliography

[BS77] Michael J Brennan and Eduardo S Schwartz. Savings bonds, retractable bonds and callable bonds. *Journal of Financial Economics*, 5(1):67–88, 1977.

[FLL+99] E. Fournié, J. Lasry, J. Lebuchoux, P. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.

[Gla03] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, Berlin, 2003.

[JLL90] Patrick Jaillet, Damien Lamberton, and Bernard Lapeyre. Variational inequalities and the pricing of american options. *Acta Applicandae Mathematica*, 21(3):263–289, 1990.

[Lee04] R. Lee. Option pricing by transform methods: extensions, unification and error control. *Journal of Computational Finance*, 7:51–86, 2004.

[LFBO07] R. Lord, F. Fang, F. Bervoets, and K. Oosterlee. *A Fast Method for Pricing Early-Exercise Options with the FFT*, volume 4488 of *Lecture Notes in Computer Science*. Springer, Berlin, 2007.

[Rai00] S. Raible. *Lévy Processes in Finance: Theory, Numerics, and Empirical Facts*. Dissertation Universität Freiburg i. Br., 2000.

[Sey06] R. Seydel. *Tools for Computational Finance*. Springer, Berlin, third edition, 2006.

[Vet21] M. Vetter. *Mathematical Finance*. Lecture Notes, CAU Kiel, 2021.

Index