December 2024



Computational Finance

Modeling, numerics, applications

Lecture Notes

Univ.-Prof. Dr. Ansgar Jüngel Institut für Analysis und Scientific Computing

This manuscript is mainly based on the following book: *Finanzderivate mit MATLAB* of M. Günther and A. Jüngel. Further main sources are the books of R. Seydel, *Tools for Computational Finance*, P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, and G. Lord, C. Powell, and T. Shardlow, *An Introduction to Computational Stochastic PDEs*.

Contents

1	Introduction	4
2	Basics 2.1 Financial options 2.2 Stochastic analysis	10 10 14
3	The Black-Scholes model3.1Black-Scholes formulas3.2Risk-neutral valuation3.3Implied volatility3.4Options on dividend paying assets3.5Multi-asset options3.6Stochastic interest rate and volatility3.7Application: Pricing Asian options3.8Beyond Black-Scholes	 25 34 37 39 42 44 47 49
4	Binomial models4.1The Cox-Ross-Rubinstein model4.2Relation with the Black-Scholes formula4.3Binomial method	58 58 61 63
5	Monte-Carlo method5.1Pseudo random numbers5.2Euler-Maruyama method5.3Milstein method5.4Itô-Taylor and Runge-Kutta schemes5.5Systems of stochastic differential equations5.6Quasi Monte-Carlo method and variance reduction5.7Multilevel Monte-Carlo method5.8Application: Pricing Asian options in the Heston model	 67 68 72 78 81 86 88 94 95
6	Finite-difference methods 6.1 The θ-method	98 98 106 108 110 118

References

English	German	
compound interest	Zinseszins	
expiry date/maturity	Verfallstag/Fälligkeit	
fixed-term deposit	Festgeld	
holder	Käufer einer Option	
interest	Zins	
interest rate	Zinssatz	
loan	Kredit	
payoff	Auszahlung	
premium	Optionspreis	
savings account	Sparguthaben	
share	Anteil	
stock	Aktie	
strike/exercise price	Ausübungspreis	
time to expiration	Laufzeit	
underlying (asset)	Basiswert	
volatility	Preisschwankung/Volatilität	
writer	Verkäufer einer Option	

Table 1: English-German translations of financial terms used in the lecture notes.

1 Introduction

A financial derivative is an agreement between two parties based on an underlying asset (or just underlying). It is designed to hedge financial transactions, but they can be also used to speculate on the prices of the underlying. Already in ancient **Greece**, shipping contracts for trading were used, similar to forward contracts (see below). The buyer receives some money upfront, which finances the trading voyage. After return (if successful), the buyer pays back the loan and the required interest. Similar commodity forward contracts were used in the Roman era, in **Italy** from the 10th century on, and in **Japan** in the 18th century.

In the 17th century, financial options were issued in the **Dutch Republic** to hedge the rapidly increasing prices of tulip bulbs, which were very fashionable at that time. In these options, the obligation to purchase tulip bulbs at a fixed price was converted to an opportunity to do so. The tulip market was accompanied by wild speculative transactions, leading eventually in February 1637 to a collapse of the tulip bulb contract prices.

Later, further precursors of today's financial derivatives have been issued and traded, but a regulated market was still missing. In 1848, the Chicago Board of Trade of the **USA** opened, providing a central location as a storage and market place for grain. Some years later, in 1865, the contracts were standardized. The Chicago Board Option Exchange, founded in 1973 as an extension of the Chicago Board of Trade, then became the first exchange to offer standardized options trading.

Derivatives gained widespread use in the 1970s, fostered by the computer technology allowing the complex models and computations to be solved in an efficient way. In the same decade, in 1973, **Fischer Black** and **Myron Scholes** published their seminal paper "The pricing of options and corporate liabilities". **Robert Merton** expanded the mathematical understanding of the option pricing model. (Interestingly, Louis Bachelier already formulated in 1900 a theory of option pricing in his PhD thesis, far before Black, Scholes, and Merton. However, Bachelier did not include a theory of hedging and replication by dynamic strategies; see the discussion in [32].) The Black-Scholes formula led to a boom in option trading. Merton and Scholes received the Nobel Memorial Prize in Economic Sciences in 1997 for their new methodology to value derivatives (Black died in 1995 and could not receive the award). Nowadays, in spite of the financial crisis from 2007, the derivative market is huge – its size is estimated to be more than ten times larger than the total world gross domestic market. Therefore, the understanding and control of financial instruments is of paramount importance for companies and also for the society.

These lecture notes are devoted to the modeling and numerical simulation of financial derivatives. In particular, we derive the Black-Scholes equation and the Black-Scholes formula, discuss its limitations and extensions, and present some numerical techniques

to solve the corresponding financial models. The topic forces us to use several mathematical techniques:

- ▶ mathematical modeling,
- ► stochastic analysis,
- ▶ theory of partial differential equations,
- ▶ numerical analysis.

The lecture notes are organized in such a way that no special knowledge is assumed to follow the arguments – besides of calculus, ordinary differential equations, and a basic understanding of probability theory.

What is a financial derivative? It is a contract whose value at the expiry date is determined by the value of the underlying asset at or up to expiration. The underlying asset can be, for instance, a stock, commodity, index, or currency. Generally, we can distinguish four classes of financial derivatives:

- ► *Forward contracts.* A forward contract is an agreement between two counterparties to sell or purchase an asset for a price agreed upon today with delivery and payment at a fixed future date. Forwards can be used to hedge risks like currency rate risk.
- ► Future contract. Future contracts are very similar to forward contracts. They are standardized contracts between two counterparties to sell or purchase an asset for a price agreed upon today with delivery and payment at a fixed future date. In contrast to forwards, they are exchange-traded and both parties are required to put up an initial amount of cash to mitigate the risk of default of either party.
- Option contracts. Option contracts give the owner the right, but not the obligation, to sell or purchase an asset for the exercise price until or at the expiry date. In contrast to forward or future contracts, the owner has no obligation to carry out the right of selling or buying the underlying asset. This allows the owner to hedge or mitigate risk due to changes in the asset price.
- Swaps. A swap is a contract determining the cash flow exchange of the counterparties. For instance, one party may exchange an uncertain cash flow to a certain one or to swap a fixed interest rate to a floating one. Swaps are used to hedge risks like the interest rate or currency risk, but they may be also used to speculate on changes of the underlying price. Common swaps are interest rate swaps, currency swaps, credit swaps, commodity swaps, and equity swaps.

We see that financial derivatives are defined through the following values:

- ▶ the quantity and class of the underlying asset,
- ▶ the expiry date,
- ▶ the excercise or strike price,
- ▶ the type of contract (the right or obligation to buy or sell the asset).

In the lecture notes, we focus on financial options. The most common ones are *vanilla options*, and we distinguish between *European* and *American options*. The former option class can be exercised only at the expiry date, the latter one can be exercised until or at expiration. Each of these two option classes can be exercised as a *call option* or *put option*. The holder of a call option has the right to buy a fixed amount of the underlying for a fixed price from the option seller (called the writer), while a put option gives the holder the right to sell a fixed amount of the underlying for a fixed price to the writer; see Table 2. A more formal definition of European options is as follows.

	Call	Put
European	Exercise at expiry only	Exercise at expiry only
	Right to buy the underlying	Right to sell the underlying
American	Exercise until or at expiry	Exercise until or at expiry
	Right to buy the underlying	Right to sell the underlying

Table 2: Overview of vanilla options.

Definition 1.1 (European option). A European call (put) option *is a contract with the following conditions: The holder of the option has the right but not the obligation to buy the underlying from the writer (to sell the underlying to the writer) at the expiry date T for the fixed strike or exercise price K.*

The option gives a right, so it has a value. We denote the value of an call option at time *t* by $C_t = C(t)$, the value of a put option by $P_t = P(t)$.¹ Let S_t be the value of the underlying at time *t*. At expiration, we may distinguish two cases for a call option:

- ► $S_T > K$: We exercise the option, purchase the underlying at price K, and sell it immediately on the spot for the price S_T . We realize the profit $S_T K > 0$.
- ► S_T < K: The right to exercise the option is not interesting since we can buy the underlying on the market for a cheaper price than guaranteed by the option contract. The option expires.</p>

This shows that the value C_T of the call option at expiration (the so-called *payoff*) is

$$C_T = \max\{0, S_T - K\} = (S_T - K)^+.$$

We can use similar arguments to determine the payoff of a put option. If $S_T > K$, the put option has no value since we can sell the underlying for a higher price on the market, while for $S_T < K$, we buy the underlying on the spot for the price S_T , exercise the

¹It is common in the theory of stochastic processes to use an index for the time variable. The notation $C(\omega)$ or $P(\omega)$ is used for the stochastic variables ω . In the theory of partial differential equations, an index usually means partial differentiation with respect to that variable. We try to avoid any confusion, which may arise because of these two conventions.

option, and sell the underlying for the price *K*. We realize the profit $K - S_T > 0$. Thus, the payoff P_T of the put option equals

$$P_T = (K - S_T)^+.$$

The holder of a call option bets on rising prices, while the writer speculates on falling prices. The opposite holds for put options. The payoff of European options can be illustrated by the *payoff diagrams* in Figure 1.1.



Figure 1.1: Payoff diagrams of a European call option (left) and put option (right).

- **Remark.** (1) Our arguments only hold when we neglect transaction costs, bid-ask spreads, etc. We specify these assumptions in Section 3.1 and discuss models including transaction costs in Section 3.8.
 - (2) When an option is exercised, usually the underlying will not be delivered physically but just the profit will be paid out (this is called *cash settlement*). As we already mentioned, options may be used to *speculate* on falling or rising prices or to *hedge* the underlying asset against price fluctuations and to mitigate risk.

We denote the value of a (call or put) option by V(S, t). This means that V is a function of time t and *all possible realizations* of the value of the underlying S, modeled by a function $S_t(\omega) := S(\omega, t)$, where the stochastic variable ω varies in the set of all possible events. At time t, exactly one value of the underlying S_t is realized, and the value of the option is $V(S_t, t)$, but the function V is defined for all values S. The key question is: Which is the "fair" price V(S, 0) at time t = 0, when the writer sells the option?

We will show below that, under some assumptions on the financial market, the fair price V(S, t) satisfies the following partial differential equation, the Black-Scholes equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0, \quad S \in (0, \infty), \ t \in (0, T).$$

Here, the parameter $r \ge 0$ is the riskfree interest rate and $\sigma > 0$ the volatility, a measure for the fluctuations of the underlying. At time t = T, the value of the option is known:

$$V(S,T) = \begin{cases} (S-K)^+ & : \text{ call,} \\ (K-S)^+ & : \text{ put.} \end{cases}$$

Boundary conditions at S = 0 and $S \rightarrow \infty$ will be specified in Section 3.1. Thus, the differential equation has to be solved backwards in time, and we are looking for the solution V(S, t) at t = 0. Interestingly, the equation can be transformed to the heat equation and solved explicitly, leading to the Black-Scholes formula. The Black-Scholes theory will be detailed in Section 3.

In contrast to European options, American options can be exercised at or before the expiration date. The formal definition is as follows.

Definition 1.2 (American option). An American call (put) option *is a contract with the following conditions: The holder of the option has the right but not the obligation to purchase the underlying from the writer (to sell the underlying to the writer) at any time before and including the expiry date.*

The value of an American call or put option at time t = T is the same as for the corresponding European option. Since American options can be also exercised *before* the expiration date, its value is at least as large as the value of the corresponding European option. Besides the value of the American option at time t = 0, we also need to determine the optimal exercise date. This leads to a free-boundary problem. We will study American options in Section 6.

Exotic options are all options which are not of vanilla style (European or American options). We mention some of them:

- ► The *Bermudan option* can only exercised on specified dates on or before expiration. The American option is a special case of a Bermudean option.
- ► The payoff of an *Asian option* is determined by the average of the price of the underlying over some time period.
- ► The *barrier option* requires that the price of the underlying must or must not pass a certain level or barrier before it can be exercised.
- ► The *binary option* pays a certain amount if the price of the underlying satisfies some defined condition on expiration. Otherwise, it expires. Therefore, is is an all-or-nothing option.
- The payoff of *lookback options* depends on the minimum or maximum of the price of the underlying.
- Multi-asset options depend on several underlyings, and correspondingly, the payoff depends on the price of these underlyings at expiration.

While the value of European options can be determined explicitly under some conditions on the market, this does not hold true in more general situations or for certain exotic options. This makes it necessary to develop numerical methods to determine approximate values for the option price. In these lecture notes, we will discuss three numerical techniques:

► Binomial methods (Section 4),

- ► Monte-Carlo methods (Section 5),
- ► finite-difference methods (Section 6).

In Section 2, we investigate basic properties of vanilla options and basic elements of stochastic analysis, needed for the derivation and formulation of the Black-Scholes equations presented in Section 3.

2 Basics

2.1 Financial options

The price of a European option can be determined explicitly under some simplifying assumptions on the financial market. The most important condition is the absence of *arbitrage*. Arbitrage is the riskfree profit by exploiting a price difference between two markets or two financial products. To illustrate this notion, let us consider a simple example taken from [16, Beispiel 1.3].

Example 2.1 (Arbitrage). We consider a financial market that allows for only three investment possibilities: a bond, a stock, and a European call option. The bond is a debt security which requires the writer to pay interest (called the coupon) to the holder and to pay the principal at a fixed later date, similar to a fixed-term deposit. We neglect default risk and assume that the bond is riskfree. The values of the bond, stock, and option at time *t* are denoted by B_t , S_t , and C_t , respectively. We assume that at time t = 0, $B_0 = 100$, $S_0 = 100$, and $C_0 = 10$. The call option has the strike price K = 100 and the expiration date *T*. We assume that trading is only possible at t = 0 and t = T and that there exist only two states of the financial market, "high" and "low" with

"high":
$$B_T = 110$$
, $S_T = 120$,
"low": $B_T = 110$, $S_T = 80$.

We construct the following portfolio: We buy 2/5 shares of the bond and one call option and sell 1/2 share of the stock. The portfolio at time t = 0 becomes $\pi_0 = \frac{2}{5} \cdot 100 + 10 - \frac{1}{2} \cdot 100 = 0$, i.e., initially it has no value. At time t = T, we have

"high":
$$\pi_T = \frac{2}{5} \cdot 110 + (120 - 100)^+ - \frac{1}{2} \cdot 120 = 4,$$

"low": $\pi_T = \frac{2}{5} \cdot 110 + (80 - 100)^+ - \frac{1}{2} \cdot 80 = 4.$

Thus, the portfolio has always the value 4 at time t = T. Then the portfolio must have a positive value also at time t = 0, and the holder could sell it at time t = 0 to realize an instantaneous riskfree profit. This is called arbitrage.

Why do we have arbitrage? The reason is that the price for the call option is too low. If any investment has the same chances of profit, there are numbers $c_1, c_2 \in \mathbb{R}$ such that

$$c_1 \cdot B_t + c_2 \cdot S_t = C_t$$
 for $t = 0, T$.

Clearly, the fair price at t = 0 is $C_0 = c_1 \cdot B_0 + c_2 \cdot S_0$. At time t = T, we have

"high":
$$c_1 \cdot 110 + c_2 \cdot 120 = (120 - 100)^+ = 20$$
,
"low": $c_1 \cdot 110 + c_2 \cdot 80 = (80 - 100)^+ = 0$.

This is a linear system with solution $c_1 = -4/11$ and $c_2 = 1/2$. We infer that

$$C_0 = -\frac{4}{11} \cdot 100 + \frac{1}{2} \cdot 100 = \frac{300}{22} \approx 13.64.$$

As we have replicated the call option by means of the bond and stock, this approach is called a *replication strategy*.

The assumption of the absence of an instantaneous and riskfree profit (arbitrage) is fundamental in the theory of financial markets. A savings account is generally riskfree but the interest is paid only after some time; options may allow for an instantaneous profit but the investment if not riskfree. Real markets are not free of arbitrage. We assume that the financial market is efficient (liquid, accessible, and transparent) which rules out arbitrage opportunities. Interestingly, we can determine some bounds on the option prices just from the assumption of absence of arbitrage.

To derive these bounds, we consider a financial market with the following assumptions:

- there is no arbitrage,
- ▶ there are no dividend payments on the underlying,
- ▶ the riskfree interest rate is the same for investments and loans,
- ▶ the market is liquid and accessible (i.e., trading is possible at any time).

The riskfree interest rate is denoted by $r \ge 0$, and we suppose continuous interest payments. This means that the return of an investment $K_0 > 0$ after time t = T equals $K = K_0 e^{rT}$. This formula can be derived from discrete returns at $t = \Delta t, ..., n \Delta t$, where $n \Delta t = T$. Including compound interest, the return equals after n payments:

$$K_n = K_0 (1 + r \triangle t)^n = K_0 (1 + rT/n)^n \rightarrow K_0 e^{rT}$$
 as $n \rightarrow \infty$.

If we invest today the amount Ke^{-rT} (or buy a bond with value Ke^{-rT}), we receive the value *K* after time *T*. The factor e^{-rT} is called the *discount factor*.

We claim the following relationship between bond, option, and stock value.

Proposition 2.2 (Put-call parity). Under the stated assumptions on the financial market, *it holds that*

$$S_t + P_t - C_t = Ke^{-r(T-t)}$$
 for all $0 \le t \le T$,

where $P_t := P(S_t, t)$ and $C_t := C(S_t, t)$.

Proof. Consider the following portfolio: Purchase a stock *S* and a European put option

P with strike *K* and expiry *T* and sell a European call option with strike *K* and expiry *T*. The portfolio has the value $\pi = S + P - C$. At time t = T, its value is

$$\pi_T = S_T + (K - S_T)^+ - (S_T - K)^+ = K.$$

We claim that the value equals $\pi_t = Ke^{-r(T-t)}$ for any t < T. This corresponds to a bond with value $Ke^{-r(T-t)}$ at time t.

Suppose that $\pi_t < Ke^{-r(T-t)}$ at some time t. Then we buy the portfolio, borrow the amount $Ke^{-r(T-t)}$ and put aside $Ke^{-r(T-t)} - \pi_t > 0$. At time t = T, the portfolio has the value K which we use to pay back the loan. This means that we have realized at time t the instantaneous riskfree profit $Ke^{-r(T-t)} - \pi_t > 0$; contradiction.

Suppose that $\pi_t > Ke^{-r(T-t)}$ at some time *t*. Now, we sell the portfolio, invest $Ke^{-r(T-t)}$ in a riskfree bond and put aside $\pi_t - Ke^{-r(T-t)} > 0$. At time t = T, we sell the bond which has the value *K* and buy back the portfolio. We have realized an instantaneous riskfree profit; contradiction.

European options satisfy the following bounds.

Proposition 2.3 (Bounds for European options). It holds for $0 \le t \le T$ that (i) $(S_t - Ke^{-r(T-t)})^+ \le C_t \le S_t$, (ii) $(Ke^{-r(T-t)} - S_t)^+ \le P_t \le Ke^{-r(T-t)}$.

Proof. We only consider the call option. The bounds for the put option follow from those for the call option and the put-call parity.

The lower bound $C_t \ge 0$ is clear since otherwise, a call with $C_t < 0$ would yield an instantaneous riskfree profit. Similarly, $P_t \ge 0$. It holds $C_t \le S_t$ since otherwise, we sell the call and purchase the underlying. At time t = T, we maybe need to sell the underlying to fulfill the obligation of the call contract. Still, because of $C_t - S_t > 0$, we have realized an instantaneous riskfree profit at time t, which is a contradiction.

The remaining bound $C_t \ge S_t - Ke^{-r(T-t)}$ follows from the call-put parity:

$$C_t = S_t + P_t - Ke^{-r(T-t)} \ge S_t - Ke^{-r(T-t)},$$

which finishes the proof.

Similar bounds can be shown for American options with values C_A , P_A . The values of the corresponding European options are denoted by C_E . P_E .

Proposition 2.4 (Bounds for American options). It holds for $0 \le t \le T$:

- (i) $(S_t Ke^{-r(T-t)})^+ \le C_A(S_t, t) \le S_t$,
- (ii) $C_A(S_t, t) = C_E(S_t, t),$

(iii) $Ke^{-r(T-t)} \le S_t + P_A(S_t, t) - C_A(S_t, t) \le K$, (iv) $(Ke^{-r(T-t)} - S_t)^+ \le P_A(S_t, t) \le K$.

In real markets, the value of American and European call options is generally not the same because of dividend payments. Statement (iii) can be seen as a put-call parity for American options.

Proof. (i) This follows as in the proof of Proposition 2.3.

(ii) Assume that we exercise the American option at time t < T. We receive the amount $S_t - K > 0$ (if $S_t - K \le 0$, we would not exercise the option). It follows from (i) that

$$C_A(S_t,t) \ge (S_t - Ke^{-r(T-t)})^+ = S_t - Ke^{-r(T-t)} > S_t - K.$$

This means that it is better to sell the option than to exercise it. Thus, the early exercise it not optimal, but then the option works like a European option.

(iii) Since American put options are more flexible than European ones, we have $P_A \ge P_E$. By the put-call parity and (ii), we infer that

$$C_A - P_A \le C_E - P_E = S_t - Ke^{-r(T-t)},$$

which is the lower bound. For the upper bound, we apply an arbitrage argument. Assume that $S_t - K > C_A(S_t, t) - P_A(S_t, t)$ for some *t*. Consider the portfolio $\pi_t = C_A(S_t, t) - P_A(S_t, t) - S_t + K < 0$. The cash flow gives $-\pi_t > 0$. Let $\tau \leq T$ be the exercise time of the American put option. It holds that $\tau \geq t$ since otherwise, the put option has already been exercised before time *t*. We distinguish two cases:

• If $S_{\tau} \leq K$ then (because of $C_A \geq 0$)

$$\pi_{\tau} = C_A(S_{\tau}, \tau) - (K - S_{\tau}) - S_{\tau} + Ke^{r(\tau - t)} \ge -K + Ke^{r(\tau - t)} \ge 0.$$

• If $S_{\tau} > K$ then (because of $C_A(S_{\tau}, \tau) \ge S_{\tau} - K > 0$)

$$\pi_{\tau} \ge (S_{\tau} - K) - 0 - S_{\tau} + Ke^{r(\tau - t)} = -K + Ke^{r(\tau - t)} \ge 0.$$

We have found a portfolio that allows for arbitrage; contradiction.

(iv) The inequalities in (iii) can be formulated as

$$Ke^{-r(T-t)} - S_t + C_A \le P_A \le K - S_t + C_A.$$

Now, we know from (i), (iii), and Proposition 2.3 (i) that

$$P_A \ge Ke^{-r(T-t)} - S_t + C_E \ge Ke^{-r(T-t)} - S_t + (S_t - Ke^{-r(T-t)})^+$$

= $(Ke^{-r(T-t)} - S_t)^+$,
 $P_A \le K - S_t + C_E \le K - S_t + S_t = K$.

This ends the proof.

Remark. The lower bound for American put options can be strengthened to

$$P_A(S_t, t) \ge (K - S_t)^+$$

Indeed, the case $K \leq S_t$ means that $P_A \geq 0$ which is trivial. Therefore, let $K > S_t$. If $P_A(S_t, t) < (K - S_t)^+ = K - S_t$ for some t, we buy a put option and excercise it immediately to make the riskfree, instantaneous profit $(K - S_t) - P_A(S_t, t) > 0$; contradiction.

The bounds for European and American option values are illustrated in Figure 2.1.



Figure 2.1: Qualitative illustration of the prices of European and American options. It is assumed that there are no dividend payments.

2.2 Stochastic analysis

We give a concise introduction to basic elements of stochastic analysis, which are needed to formulate the stochastic asset price dynamics and to valuate financial options. We only state the needed results and refer for proofs and examples to the literature, for instance, [3, 10, 27].

The key objects are stochastic processes [27, Section 2]. For this, let Ω be an arbitrary set (the set of all possible events). A σ -algebra \mathcal{F} on Ω is a family of subsets of Ω satisfying the following properties:

(i)
$$\emptyset \in \mathcal{F}$$
, (ii) $\forall A \in \mathcal{F} : \Omega \setminus A \in \mathcal{F}$, (iii) $\forall A_n \in \mathcal{F} : \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

For instance, the collection of open sets on a topological space *X* forms a σ -algebra, the Borel σ -algebra $\mathcal{B}(X)$. A *probability measure* $P : \mathcal{F} \to [0, 1]$ is a function satisfying

(i)
$$P(\emptyset) = 0$$
, $P(\Omega) = 1$,
(ii) $\forall A_n \in \mathcal{F}$ pairwise disjoint : $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

We call the triple (Ω, \mathcal{F}, P) a *probability space*.

Let $X : \Omega \to \mathbb{R}$ be any function on (Ω, \mathcal{F}, P) . We call this function \mathcal{F} -measurable if the preimage $X^{-1}(U) = \{\omega \in \Omega : X(\omega) \in U\}$ is an element of \mathcal{F} for all open (or all Borel) sets $U \subset \mathbb{R}$. A random variable is a \mathcal{F} -measurable function $X : \Omega \to \mathbb{R}$. The same definition holds for codomains \mathbb{R}^n instead of \mathbb{R} .

The *expectation* E of a random variable X is defined as the Lebesgue integral

$$\mathbf{E}(X) = \int_{\Omega} X d\mathbf{P} = \int_{\Omega} X(\omega) d\mathbf{P}(\omega).$$

Definition 2.5 (Stochastic process). A family $(X_t)_{t\geq 0}$ of random variables $X_t : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a stochastic process. For fixed $\omega \in \Omega$, we call the function $t \mapsto X_t(\omega)$ a trajectory or path of X_t .

Example. An example of a stochastic process is the value S_t of an asset (stocks, bonds, commodities, etc.). More precisely, the value is given by $S_t(\omega)$, where ω is an element of the event space Ω but usually, we omit the dependence on ω .

In the following, let (Ω, \mathcal{F}, P) be a probability space. A *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is an increasing family of sub- σ -algebras of \mathcal{F} , i.e. $\mathcal{F}_s \subset \mathcal{F}_t$ for all $0 \leq s \leq t$. The filtration \mathcal{F}_t represents the set of events observable up to time *t* and it becomes larger when time progresses. A *natural filtration* $(\mathcal{F}_t^X)_{t\geq 0}$ of a stochastic process $X = (X_t)_{t\geq 0}$ is the smallest σ -algebra on Ω that contains all preimages of measurable sets in \mathbb{R} , i.e.

$$\mathcal{F}_t^X = \sigma\{X_s^{-1}(U) : 0 \le s \le t, \ U \in \mathcal{B}(\mathbb{R})\}.$$
(2.1)

The natural filtration of a stochastic process contains information on all the past history of the process.

We want to consider only those stochastic processes for which, at any time, the associated filtration gives us enough information to find the value of the process. We call such processes adapted.

Definition 2.6 (Adapted stochastic process). We call a stochastic process $(X_t)_{t\geq 0} \mathbb{F}$ -adapted if X_t is \mathcal{F}_t -measurable for any $t \geq 0$.

We need special adapted stochastic processes, which have the property that at time t > 0, the expectation of X_t , given the knowledge until time $s \le t$, equals X_s . Roughly speaking, this means that X_s is the best expected value of X_t . Such a process is called a martingale. For its precise definition, we need the notion of conditional expectation; see [27, Appendix B].

Definition 2.7 (Conditional expectation). Let $X : \Omega \to \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}|X| < \infty$ and let $\mathcal{B} \subset \mathcal{F}$ be a σ -algebra. Then $\mathbb{E}(X|\mathcal{B}) : \Omega \to \mathbb{R}$ is the a.s. (= almost surely) unique function satisfying

(i)
$$E(X|B)$$
 is *B*-measurable,
(ii) $\int_{B} E(X|B) dP = \int_{B} X dP$ for all $B \in B$.

This definition contains an existence and uniqueness result which follows from the Theorem of Radon-Nikodym. Indeed, the finite measure $\mu(B) = \int_B X dP$ for $B \in \mathcal{B}$ is absolutely continuous with respect to $P|_{\mathcal{B}}$ (i.e., $\mu(B) = 0$ for all $B \in \mathcal{B}$ with P(B) = 0), therefore there exists a function $f : \Omega \to \mathbb{R}$ such that $\mu(B) = \int_B f dP$ and it is enough to set $E(X|\mathcal{B}) := f$.

We note the *tower property* of conditional expectations: If $\mathcal{B} \subset \mathcal{C} \subset \mathcal{F}$ then

$$E(E(X|\mathcal{C})|\mathcal{B}) = E(X|\mathcal{B})$$
 P-a.s.

Indeed, by definition,

$$\int_{B} E(X|\mathcal{B})dP = \int_{B} XdP \text{ for all } B \in \mathcal{B},$$
$$\int_{C} E(X|\mathcal{C})dP = \int_{C} XdP \text{ for all } C \in \mathcal{C}.$$

The last identity clearly also holds for all $B \in \mathcal{B} \subset C$. Therefore, combining these identities leads to

$$\int_{B} \mathcal{E}(X|\mathcal{B}) d\mathbf{P} = \int_{B} \mathcal{E}(X|\mathcal{C}) d\mathbf{P} \quad \text{for all } B \in \mathcal{B},$$

which, by uniqueness, shows that E(E(X|C)|B) equals E(X|B).

Definition 2.8 (Martingale). Let $(X_t)_{t\geq 0}$ be a stochastic process on a probability space (Ω, \mathcal{F}, P) and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration. We call $(X_t)_{t\geq 0}$ a martingale if $(X_t)_{t\geq 0}$ is \mathbb{F} -adapted, $\mathbb{E}|X_t| < \infty$ for all $t \geq 0$, and

$$E(X_t|\mathcal{F}_s) = X_s$$
 P-a.s. for all $s \leq t$.

Example. Let *Y* be a random variable which is \mathbb{F} -measurable with $E|Y| < \infty$, and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration. We claim that $X_t = E(Y|\mathcal{F}_t)$ defines a martingale. Indeed, by the tower property, since $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$,

$$\mathrm{E}(X_t|\mathcal{F}_s) = \mathrm{E}(\mathrm{E}(Y|\mathcal{F}_t)|\mathcal{F}_s) = \mathrm{E}(Y|\mathcal{F}_s) = X_s.$$

An important example for a martingale is the Brownian motion or Wiener process (see [29, Section 2.1]).

Definition 2.9 (Wiener process). A stochastic process $(W_t)_{t\geq 0}$ on a probability space (Ω, \mathcal{F}, P) is called a (standard) Wiener process if

- (i) $W_0 = 0$,
- (ii) W has continuous trajectories P-a.s.,
- (iii) W has independent increments, i.e. $W_{t_1} W_0$, $W_{t_2} W_{t_1}$,..., $W_{t_n} W_{t_{n-1}}$ are independent for all $0 \le t_1 \le \cdots \le t_n \le T$ and $n \in \mathbb{N}$,
- (iv) W has Gaussian increments, i.e. $W_t W_s$ is N(0, t s)-distributed for all s < t, i.e. $W_t W_s$ is normally distributed with $E(W_t W_s) = 0$ and $Var(W_t W_s) = t s$.

Example. The Brownian motion allows us to define a model for the asset price dynamics. The first idea is to write the asset value S_t as the sum of the price at time t = 0, S_0 , a premium $a \cdot t$, and a stochastic component Z_t ,

$$S_t = S_0 + a \cdot t + Z_t, \quad t \ge 0.$$

This ansatz may lead to negative values of S_t if Z_t is sufficiently large and negative. We do not allow negative values and prefer another idea. A bond B_t with riskfree interest rate $r \in \mathbb{R}$ evolves according to $B_t = B_0 e^{rt}$ or, equivalently, $\ln B_t = \ln B_0 + rt$. This motivates the following ansatz for S_t :

$$\ln S_t = \ln S_0 + a \cdot t + Z_t, \quad t \ge 0.$$

We assume that the stochastic component is normally distributed with expectation zero and variance $\sigma^2 t$, where $\sigma > 0$. These properties are satisfied by the Brownian motion (times σ),

$$\ln S_t = \ln S_0 + a \cdot t + \sigma W_t, \quad t \ge 0.$$

Taking the exponent leads to $S_t = S_0 \exp(at + \sigma W_t)$.

Although almost every trajectory $t \mapsto W_t(\omega)$ is continuous, it is neither differentiable nor of bounded variation. This has an important consequence for the value of an asset. Indeed, assume that the asset price is modeled by a Brownian motion W_t and that we are allowed to trade the asset only at times $0 = t_0 < t_1 < \cdots < t_n = T$. At time t_k , we can choose to hold X_k shares of the asset but we are only allowed to use information up to time t_k , i.e., we cannot see the future. The change in value in the period $[t_{k-1}, t_k]$ is the change in price times the number of shares we own at time t_{k-1} , i.e. $X_{k-1}(W_{t_k} - W_{t_{k-1}})$. Thus, the change in value from t_0 to t_n equals the sum

$$\sum_{k=1}^{n} X_{k-1}(W_{t_k} - W_{t_{k-1}}).$$

When we allow for trading in continuous time, the sum becomes the stochastic integral

$$\int_0^T X_t dW_t.$$

Since the Brownian motion is not of bounded variation, we cannot associate a measure with the increments to construct a Lebesgue-Stieltjes integral, and it is not clear how to define the integral.

This problem can be solved by defining the integral $\int_0^T X_t dW_t$ first for so-called simple processes, which are piecewise constant in time. More precisely, let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of [0, T] and set $X_t(\omega) = \theta_{k-1}(\omega)$ for $t_{k-1} < t \le t_k$, where θ_{k-1} is a $\mathcal{F}_{t_{k-1}}$ -measurable random variable. We assume that $\int_0^T E|X_t|^2 dt < \infty$. Then $(X_t)_{t\geq 0}$ is called simple, and its stochastic integral is defined by

$$\int_0^T X_t dW_t := \sum_{k=1}^n \theta_{k-1} (W_{t_k} - W_{t_{k-1}}).$$

It can be shown that $I(t, \omega) := \int_0^t X_s(\omega) dW_s(\omega)$ is a martingale with respect to the filtration \mathbb{F} . We wish to extend the definition to square-integrable adapted stochastic processes. Let $(X_t)_{t\geq 0}$ be a stochastic process satisfying the following properties:

- (i) $(t, \omega) \mapsto X_t(\omega)$ is $(\mathcal{B} \times \mathcal{F})$ -measurable, where \mathcal{B} is the Borel σ -algebra of $[0, \infty)$,
- (ii) (X_t) is adapted,
- (iii) $\int_0^T \mathbf{E} |X_t|^2 dt < \infty$ for all T > 0.

It is possible to show that for given $(X_t)_{t\geq 0}$ satisfying these properties, there exists a sequence $(Y^{(n)})$ of simple processes such that

$$\operatorname{E} \int_0^T |X_t - Y_t^{(n)}|^2 dt \to 0 \quad \text{as } n \to \infty.$$

By the Itô isometry (see Theorem 2.10),

$$E\left(\int_{0}^{T}|X_{t}-Y_{t}^{(n)}|dW_{t}\right)^{2}=E\int_{0}^{T}|X_{t}-Y_{t}^{(n)}|^{2}dt,$$

the sequence $(\int_0^T |X_t - Y_t^{(n)}| dW_t)_{n \in \mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbf{P})$, and we can define

$$\int_0^T X_t dW_t := \lim_{n \to \infty} \int_0^T Y_t^{(n)} dW_t.$$

The limit exists in the sense of $L^2(\mathbf{P})$ and is independent of the choice of $(Y^{(n)})$. The integral can be chosen in such a way that $t \mapsto \int_0^t X_s dW_s$ is continuous and $I_t(\omega) = \int_0^t X_s(\omega) dW_s(\omega)$ is a martingale.

It remains to formulate the Itô isometry. (We ignore the inconsistency that we are using the stochastic integral before actually defining it. For a rigorous proof, we prove first the Itô isometry for simple processes, define the Itô integral, and then prove the Itô isometry for general stochastic processes.)

Theorem 2.10 (Itô isometry). Let $(X_t)_{t\geq 0}$ be a stochastic process adapted to the natural filtration (see (2.1)) of the Wiener process. Then

$$\mathbf{E}\left(\int_0^T X_t dW_t\right)^2 = \mathbf{E}\int_0^T X_t^2 dt.$$
 (2.2)

The derivation of the Black-Scholes equation for financial derivatives is based on an integral version of the chain rule, called the Itô formula. For its statement, we need the notion of an Itô process.

Definition 2.11 (Itô process). Let T > 0 and a, b be \mathbb{F} -adapted stochastic processes satisfying $\int_0^T |a_t| dt < \infty$ and $\int_0^T b_t^2 dt < \infty$ P-a.s. Furthermore, let $(W_t)_{t\geq 0}$ be the Brownian motion and $(X_t)_{t\geq 0}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We call $(X_t)_{t\geq 0}$ an Itô process *if*

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t b_s dW_s, \quad 0 \le t \le T.$$

We write formally

 $dX_t = a_t dt + b_t dW_t, \quad 0 \le t \le T.$ (2.3)

The function *a* is called the drift and *b* the diffusion coefficient. Equation (2.3) is a linear stochastic differential equation. We call $(X_t)_{t\geq 0}$ also an Itô process if the differential equation is nonlinear, i.e., if the functions *a* and *b* also depend on X_t :

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t, \quad t > 0.$$

Example. Let $a_t = 0$ and $b_t = 1$. Then $X_t = X_0 + \int_0^t dW_s = X_0 + W_t$, since $W_0 = 0$. Thus, a Wiener process is a special Itô process. Next, let $a_t = rX_t \ge 0$ and $b_t = 0$. Then $X_t = X_0 + r \int_0^t X_s ds$ is the integral formulation of the differential equation $dX_t/dt = rX_t$ with the solution $X_t = X_0e^{rt}$. We can interpret X_t as a bond with interest rate $r \ge 0$.

These two extreme cases can be combined to formulate a stochastic differential equation for the value of an asset S_t . Indeed, the relative return dS_t/S_t , where we

interpret dS_t as the infinitesimal change of S_t , is given by two components. The deterministic, riskfree return equals μdt , where the drift μ is a measure of the average rate of growth of the asset value. The stochastic component is modeled by σdW_t , where the volatility σ measures the standard deviation of the returns. Both μ and σ may depend on t and S_t . Combining these terms, we arrive at the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \quad t \ge 0.$$
(2.4)

Note that S_t is a martingale if the drift vanishes, $\mu = 0$, since W_t is a martingale. We show later (see the next example) that $S_t = S_0 \exp(at + \sigma W_t)$ is the solution to (2.4) for a suitable value of a.

Let $(X_t)_{t\geq 0}$ be an Itô process such that dX = adt + bdW and $f : \mathbb{R} \to \mathbb{R}$ be a smooth function. Is $(f(X_t))_{t\geq 0}$ also an Itô process? The answer is yes, and we can compute the stochastic differential equation for $f(X_t)$ by using a stochastic Taylor expansion. Intuitively, we expect that

$$df(X_t) = f_x(X_t)dX_t + \frac{1}{2}f_{xx}(X_t)dX_t^2 + \cdots$$

= $f_x(X_t)(a_tdt + b_tdW_t) + \frac{1}{2}f_{xx}(X_t)(a^2dt^2 + ab(dtdW_t + dW_tdt) + b^2dW_t^2) + \cdots$

The expression $E((W_{t+\triangle t} - W_t)^2) = \triangle t$ motivates the symbolic formula $E(dW^2) = dt$. Thus, dW is of the "order" \sqrt{dt} (in expectation). As the expansion is only needed for terms up to order dt, we expect that dtdW and dWdt can be neglected, leading to

$$df(X_t) = f_x(X_t)(a_t dt + b_t dW_t) + \frac{1}{2}f_{xx}(X_t)b^2 dt.$$

These formal observations can be made rigorous (and slightly generalized) by the Itô formula for Itô processes.

Theorem 2.12 (Itô formula for an Itô process). Let $(X_t)_{t\geq 0}$ be an Itô process and $f : \mathbb{R} \times [0, \infty) \to \mathbb{R}$ be a twice continuously differentiable function. Then $(f(X_t, t))_{t\geq 0}$ is also an Itô process satisfying for $0 \le t \le T$,

$$df(X_t,t) = \left(f_t(X_t,t) + f_x(X_t,t)a_t + \frac{1}{2}f_{xx}(X_t,t)b_t^2\right)dt + f_x(X_t)b_tdW_t,$$

where $f_t = \partial f / \partial t$, $f_x = \partial f / \partial X$, and $f_{xx} = \partial^2 f / \partial X^2$.

Example. We have seen that the asset price can be modeled by

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \quad t \ge 0.$$
(2.5)

We assume that μ and σ are constants. The expression dS_t/S_t looks like the differential of $\ln S_t$. Since S_t is an Itô process, this is not completely true. Indeed, Itô's formula gives for $f(z) = \ln z$, $a = \mu S$, and $b = \sigma S$:

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

Thus, we have the correction $-\frac{1}{2}\sigma^2 dt$ due to the stochastic nature of S_t . In integral form, the previous equation reads as

$$\ln S_t = \ln S_0 + \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) ds + \int_0^t \sigma dW_s = \ln S_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t$$

Taking the exponent leads to the analytic solution to the asset price equation,

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right), \quad t \ge 0.$$

A similar expression, $S_t = S_0 \exp(at + \sigma W_t)$, was already stated previously. This process is called a *geometric Brownian motion*. It is the explicit solution to the linear stochastic differential equation (2.5). Compared to our previous price model, the premium $a \cdot t$ becomes $(\mu - \sigma^2/2)t$.

For later reference, we calculate the expected value and the variance of the geometric Brownian motion.

Lemma 2.13. Let
$$S_t = S_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t)$$
 be a geometric Brownian motion. Then

$$E(S_t) = S_0 e^{\mu t}, \quad Var(S_t) = S_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1).$$
 (2.6)

Proof. Since the Brownian motion is N(0, t)-distributed, its density is given by $(2\pi t)^{-1/2} \exp(-x^2/(2t))$ Hence, we compute

$$\begin{split} \mathbf{E}(S_t) &= S_0 e^{\mu t - \sigma^2 t/2} \mathbf{E}(e^{\sigma W_t}) = \frac{S_0 e^{\mu t - \sigma^2 t/2}}{\sqrt{2\pi t}} \int_{\mathbb{R}} e^{\sigma x} e^{-x^2/(2t)} dx \\ &= \frac{S_0 e^{\mu t}}{\sqrt{2\pi t}} \int_{\mathbb{R}} e^{-(x - \sigma t)^2/(2t)} dx = S_0 e^{\mu t}, \\ \operatorname{Var}(S_t) &= \mathbf{E}(S_t^2) - \mathbf{E}(S_t)^2 = S_0^2 e^{2\mu t - \sigma^2 t} \mathbf{E}(e^{2\sigma W_t}) - \mathbf{E}(S_t)^2 \\ &= S_0^2 e^{2\mu t - \sigma^2 t} \left(e^{-\sigma^2 t/2} \mathbf{E}(e^{2\sigma W_t}) \right) - (S_0 e^{\mu t})^2 \\ &= S_0^2 e^{2\mu t + \sigma^2 t} - (S_0 e^{\mu t})^2 = S_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1), \end{split}$$

showing the claim.

The product rule of stochastic analysis is different than that one in differential calculus since it involves a contribution due to the quadratic variation process. The following result holds for Itô processes.

Lemma 2.14 (Itô's product rule). Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two stochastic processes on the probability space (Ω, \mathcal{F}, P) with

$$dX = a_1 dt + b_1 dW, \quad dY = a_2 dt + b_2 dW.$$

Then

$$d(XY) = XdY + YdX + d[X,Y], \quad where \quad d[X,Y] := b_1b_2dt.$$

To motivate this rule, we write *formally*

$$d(XY) = XdY + YdX + dXdY, \text{ where}$$

$$dXdY = a_1a_2dt^2 + (a_1b_2 + a_2b_1)dtdW + b_1b_2dW^2.$$

Neglecting the higher-order terms dt^2 and dtdW and replacing dW^2 by dt shows that $d[X, Y] := dXdY = b_1b_2dt$. For a more rigorous proof, we apply the two-dimensional Itô formula (see below) to the function $f(x_1, x_2) = x_1x_2$.

Example. Let $(X_t)_{t\geq 0}$ be a stochastic process and $(Y_t)_{t\geq 0}$ be a family of continuously differentiable functions $Y : [0, \infty) \to \mathbb{R}$. Then $dY_t = (dY_t/dt)dt$ and Itô's product rule gives

$$d(X_tY_t) = X_t dY_t + Y_t dX_t,$$

which looks like the product rule from standard calculus. This formula is used to analyze the product of a stock price and a bond value.

We also need a multidimensional version of the Itô formula for Itô processes. For this, we first introduce multidimensional stochastic processes.

Definition 2.15 (Multidimensional processes). *Let* $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ *be a filtered probability space.*

- (1) We call $B_t = (B_1(t), \ldots, B_n(t))$ an n-dimensional (standard) Wiener process (or Brownian motion) if the components $B_i(t)$ are independent one-dimensional Wiener processes and the vector B(t) - B(s) is standard normally distributed. We write $B_t \sim N(0, It)$, where $I \in \mathbb{R}^{n \times n}$ is the unit matrix.
- (2) We call the vector $W_t = (W_1(t), ..., W_n(t))$ correlated Wiener processes if *B* is an *n*-dimensional Wiener process and there exists a matrix $A \in \mathbb{R}^{n \times n}$ such that W = AB. The covariance matrix of *W* equals $\Sigma = AA^{\top}$. We write $W_t \sim N(0, \Sigma t)$.

(3) We call $(X_t)_{t\geq 0}$ with $X = (X_1, \ldots, X_n)$ a multidimensional Itô process if

$$dX_i = a_i dt + \sum_{j=1}^n b_{ij} dW_j,$$

where $a = (a_1, ..., a_n)$, $b = (b_{ij})_{i,j=1}^n$ are adapted stochastic processes, and $(W_1(t), ..., W_n(t))$ is a vector of correlated Wiener processes.

We recall that the correlation between two stochastic processes X_1 and X_2 is defined by

$$\rho(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)}\sqrt{\operatorname{Var}(X_2)}},$$

where the covariance between X_1 and X_2 is given by

$$Cov(X_1, X_2) = E[(X_1 - EX_1)(X_2 - EX_2)].$$

The covariance matrix of $W = (W_1, ..., W_n)$ has the coefficients $Cov(W_i, W_j)$ and the diagonal elements $Cov(W_i, W_i) = EW_i^2$ (*i*, *j* = 1, ..., *n*).

Example. In the two-dimensional case, the covariance matrix of (W_1, W_2) can be formulated as

$$\Sigma = \begin{pmatrix} \sigma_1^2 &
ho \sigma_1 \sigma_2 \\
ho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$
 ,

where $\sigma_1^2 = EW_1^2$, $\sigma_2^2 = EW_2^2$, and $\rho = Cov(W_1, W_2)/(\sigma_1\sigma_2) \in [-1, 1]$ is the correlation coefficient between the two Wiener processes W_1 and W_2 . If $\rho > 0$, we say that W_1 and W_2 are positively correlated, and if $\rho < 0$ that W_1 and W_2 are negatively correlated.

Theorem 2.16 (Multidimensional Itô formula). Let $(X_t)_{t\geq 0}$ be a multidimensional Itô process and $f : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}$ be a twice continuously differentiable function. The correlated Wiener processes are assumed to have the covariance matrix Σ such that $\Sigma_{ii} = 1$ for i = 1, ..., n. Then $(f(X_t, t))_{t\geq 0}$ is also a multidimensional Itô process whose components satisfy for $0 \le t \le T$,

$$df(X_t,t) = \left(\frac{\partial f}{\partial t}(X_t,t) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_t,t)a_i(t) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t,t)(b_t \Sigma b_t^\top)_{ij}\right)dt \\ + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_t,t)b_i(t) \cdot dW_t.$$

Since $\Sigma_{ii} = 1$, the correlation $\rho_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}} = \Sigma_{ij}$ equals the covariance coefficient. As a rule of thumb, one may write $dW_i dW_j = \rho_{ij} dt$ to indicate the correlation

between W_i and W_j .

Example. We compute the multidimensional Itô formula for options on multiple assets. Let $(S_i(t))_{t\geq 0}$ be the value of the *i*th underlying, satisfying the Itô differential equation

$$dS_i = \mu_i S_i dt + \sigma_i S_i dW_i,$$

where W_1, \ldots, W_n are correlated Wiener processes with covariance matrix Σ . We assume that $\Sigma_{ii} = 1$ for $i = 1, \ldots, n$ and $\Sigma_{ij} = \rho_{ij} \ge 0$ for $i \ne j$. Then $a_i = \mu_i S_i$, $b_{ij} = \delta_{ij}\sigma_i S_i$ and consequently, $(b\Sigma b^{\top})_{ij} = \sigma_i \sigma_j S_i S_j \Sigma_{ij}$. This yields

$$df(S_t, t) = \left(\frac{\partial f}{\partial t} + \sum_{i=1}^n \mu_i S_i \frac{\partial f}{\partial S_i} + \frac{1}{2} \sum_{i,j=1}^n \rho_{ij} \sigma_i \sigma_j S_i S_j \frac{\partial^2 f}{\partial S_i \partial S_j}\right) dt + \sum_{i=1}^n \sigma_i S_i \frac{\partial f}{\partial S_i} dW_i.$$
(2.7)

3 The Black-Scholes model

The main goal of this section is to determine the price of an option in the framework of the Black-Scholes model and to discuss some of its extensions. To simplify the presentation, we assume that the occuring stochastic processes are adapted to the corresponding filtration and that their expectation and variance always exist. Moreover, to avoid too many technicalities, we give formal arguments only and refer to the literature for rigorous proofs.

3.1 Black-Scholes formulas

We impose the following simplifying assumptions on the financial market:

The price of the underlying satisfies the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \quad 0 < t < T, \tag{3.1}$$

with constant parameters $\mu \in \mathbb{R}$ and $\sigma \geq 0$.

▶ The constant interest rate $r \ge 0$ is the same for investments and loans. The corresponding bond satisfies the differential equation

$$dB_t = rB_t dt, \quad 0 < t < T. \tag{3.2}$$

- The stochastic processes are continuous and defined on a filtered probability space (Ω, F, F, P). This means that crashes are not modeled.
- ▶ No dividends are paid on the underlying asset.
- The market is arbitrage-free, liquid, and frictionless (no transaction costs, the debit and credit interests are the same, all market parties have the same information, and any asset can be traded any time). In particular, short selling is allowed (i.e., we may sell a security that we do not own).

Let V(S, t) be the value of an option at time t. The aim is to derive an evolution equation for V. To this end, we use a *replication strategy*, i.e., we construct a portfolio that produces the same payoff function as another asset (in our case, a riskless bond). We assume that V is twice continuously differentiable with respect to S and continuously differentiable with respect to t. Let us consider a portfolio that consists of $c_1(t)$ shares of the bond, $c_2(t)$ shares of the underlying, and a short option (this means that we sold the option; a long option means that we bought the option):

$$\pi_t = c_1(t)B_t + c_2(t)S_t - V(S_t, t).$$
(3.3)

When the shares $c_1(t)$ and $c_2(t)$ are negative, we are short selling the corresponding asset. We assume that the portfolio is self-financing, i.e., changes in the shares have to be financed from the portfolio exclusively and not from exogeneous transfer of money.

Given a portfolio $\pi_t = c(t) \cdot S(t) := \sum_{i=1}^n c_i(t)S_i(t)$, its value at $t + \Delta t$ equals its value at time *t* plus a contribution due to the change of prices of the assets:

 $\pi_{t+\triangle t} = \pi_t + c(t) \cdot (S(t+\triangle t) - S(t)),$

where we need to take into account the shares c(t) at time t. In the "infinitesimal limit", we obtain formally a stochastic differential equation, leading to the following definition.

Definition 3.1 (Self-financing portfolio). We call the portfolio $\pi_t = c(t) \cdot S(t)$ self-financing *if* π_t satisfies the following stochastic differential equation for $t \ge 0$,

$$d\pi_t = \sum_{i=1}^n c_i(t) dS_i(t).$$

Assuming that the portfolio (3.3) is riskless and self-financing, we are able to derive the Black-Scholes equation.

Theorem 3.2 (Black-Scholes equation). *Let the assumptions stated at the beginning of this subsection hold. Then the option price is a solution to the partial differential equation*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0, \quad S > 0, \ t > 0, \tag{3.4}$$

with final condition $V(S,T) = V_T(S)$ for $S \ge 0$.

Equation (3.4) is a backward parabolic equation since we have prescribed a final value not an initial datum. Therefore, the sign of the second derivative of V is positive and not negative as in standard parabolic equations. The final datum is given by

$$V_T(S) = \begin{cases} (S-K)^+ & : \text{ call,} \\ (K-S)^+ & : \text{ put.} \end{cases}$$
(3.5)

In the following, we abbreviate a partial derivative by an index, i.e. $V_t = \partial V / \partial t$, $V_{SS} = \partial^2 V / \partial S^2$, etc. However, do not confuse V_t with the value $V_t = V(t)$ of the stochastic process V.

Proof of Theorem 3.2 The proof is based on Itô's formula and a no-arbitrage argument. The idea is to set $c_2(t) = V_S$ in the portfolio (3.3). Since π_t is riskless and the market is assumed to be arbitrage-free, its value cannot be better than a riskless bond, i.e.

$$d\pi = r\pi dt = r(c_1B + c_2S - V)dt.$$

The assumption that π_t is self-financing leads to

$$d\pi = c_1 dB + c_2 dS - dV(S, t) = (c_1 rB + c_2 \mu S) dt + c_2 \sigma S dW - dV(S, t),$$

using the equations for the asset price and bond value (3.2). The equation for dV(S, t) is determined from Itô's formula, using (3.1):

$$dV(S,t) = \left(V_t + V_S \mu S + \frac{1}{2} V_{SS} \sigma^2 S^2\right) dt + V_S \sigma S dW.$$

Inserting this expression in the equation for $d\pi$, we find that

$$\begin{aligned} r(c_1 B + c_2 S - V)dt &= d\pi = (c_1 r B + c_2 \mu S)dt + c_2 \sigma dW - dV(S, t) \\ &= \left(c_1 r B + (c_2 - V_S)\mu S - V_t - \frac{1}{2}\sigma^2 S^2 V_{SS}\right)dt + \sigma S(c_2 - V_S)dW. \end{aligned}$$

Since $c_2 = V_S$, some terms cancel and we end up with

$$r(V_SS-V)dt = -\left(V_t + \frac{1}{2}\sigma^2 S^2\right)dt.$$

This is a differential equation which can be written as (3.4). (Indeed, the differential equation a(t)dt = 0 is written in integral formulation as $\int_0^t a(\tau)d\tau = 0$ for all t > 0. This implies that a(t) = 0 for all t > 0.)

Remark. Interestingly, the choice $c_2 = V_S$ cancels the stochastic component and makes the dynamics of V(S, t) purely deterministic. Moreover, the drift rate μ has been canceled out. The choice $c_2 = V_S$ is called delta hedging strategy, which eliminates the risk due to stochastic fluctuations. Choosing the so-called delta $\Delta := V_S$ makes the portfolio riskless, at least under the assumptions imposed on the market. In real markets, other risks, like volatility risk, model risk, or default risk, exist, and these risks cannot be completely hedged. It is still possible to derive a Black-Scholes-type equation, but it involves an additional function, the market price of risk. We come back to this point in Section 3.6 in the context of volatility risk.

Remark. The identity for *dV* can be obtained *formally* just by the Taylor expansion

$$dV = V_t dt + V_S dS + \frac{1}{2} V_{SS} dS^2 = V_t dt + V_S S(\mu dt + \sigma dW) + \frac{1}{2} V_{SS} S^2 (\mu dt + \sigma dW)^2$$

= $V_t + V_S \mu S dt + V_S \sigma S dW + \frac{1}{2} V_{SS} S^2 \sigma^2 dW^2 = V_t + V_S \mu S dt + \frac{1}{2} V_{SS} S^2 \sigma^2 dt + V_S \sigma S dW$

by neglecting "higher-order" terms and using $dW^2 = dt$. This formal computation may help to memorize the Itô formula.

We solve the Black-Scholes equation (3.4) on the set $(S, T) \in (0, \infty) \times (0, T)$. For unique solvability, we need to prescribe boundary conditions at S = 0 and for $S \to \infty$. Consider first a call, V = C. Then C(0, t) = 0 since the right to purchase a valueless asset has no value too. If the value of the asset is very large (compared to usual assets on the market), it is very likely that the call will be exercised, and its value is $S - Ke^{-r(T-t)}$ (we need to discount the strike *K*). As *S* is very large, we may neglect the second term, which leads to the condition $C(S,t) \sim S$ as $S \rightarrow \infty$, meaning that $C(S,t)/S \rightarrow 1$ as $S \rightarrow \infty$.

Next, let V = P be a put option. When the asset price is very high, it is likely that we do not exercise the put, and its value is negligible: $P(S, t) \rightarrow 0$ as $S \rightarrow \infty$. To determine the value at S = 0, we use the put-call parity (Proposition 2.2):

$$P(0,t) = C(0,t) + Ke^{-r(T-t)} - 0 = Ke^{-r(T-t)}.$$

We summarize:

European call:
$$C(0,t) = 0$$
, $C(S,t) \sim S(S \to \infty)$, (3.6)

European put: $P(0,t) = Ke^{-r(T-t)}, \quad P(S,t) \to 0 \ (S \to \infty).$ (3.7)

The Black-Scholes equation (3.4) with final condition (3.5) and boundary conditions (3.6) or (3.7) can be solved explicitly.

Theorem 3.3 (Black-Scholes formulas). *Problem* (3.4), (3.5), *and either* (3.6) *or* (3.7) *has the solution*

call:
$$V(S,t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$
 (3.8)

put:
$$V(S,t) = Ke^{-r(T-t)}\Phi(-d_2) - S\Phi(-d_1),$$
 (3.9)

where Φ is the distribution function of the standard normal distribution,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} dz, \quad x \in \mathbb{R},$$

and $d_{1/2}$ are the real numbers

$$d_{1/2} = \frac{\ln(S/K) + (r \pm \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}.$$
(3.10)

Proof. Step 1: We transform equation (3.4) to the heat equation. For this, we eliminate first the nonconstant coefficients S and S^2 . Set

$$x = \ln(S/K), \quad \tau = \frac{1}{2}\sigma^2(T-t), \quad v(x,\tau) = V(S,t)/K.$$

For $S \in (0, \infty)$ and $t \in [0, T]$, we have $x \in \mathbb{R}$ and $\tau \in [0, T_1]$, where $T_1 := \frac{1}{2}\sigma^2 T$. Using the chain rule of differential calculus,

$$V_t = K \frac{\partial v}{\partial \tau} \frac{d\tau}{dt} = -\frac{1}{2} K \sigma^2 v_\tau, \quad V_S = K \frac{\partial v}{\partial x} \frac{dx}{dS} = \frac{K}{S} v_x, \quad V_{SS} = -\frac{K}{S^2} v_x + \frac{K}{S^2} v_{xx},$$

it follows that

$$0 = V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV = -\frac{1}{2}K\sigma^2 \left(v_\tau + v_x - v_{xx} - \frac{2r}{\sigma^2}v_x + \frac{2r}{\sigma^2}v\right)$$

and consequently, setting $k = 2r/\sigma^2$,

$$v_{\tau} - v_{xx} + (1 - k)v_x + kv = 0. \tag{3.11}$$

The final condition transforms to the initial datum $v(x, 0) = (S - K)^+ / K = (e^x - 1)^+$.

Next, we eliminate the lower-order terms by making the ansatz $v(x, \tau) = \exp(\alpha x + \beta \tau)u(x, \tau)$ with parameters α , $\beta \in \mathbb{R}$ which will be determined later. Inserting this ansatz into (3.11), using

$$v_{\tau} = e^{\alpha x + \beta \tau} (\beta u + u_{\tau}), \quad v_x = e^{\alpha x + \beta \tau} (\alpha u + u_x), \quad v_{xx} = e^{\alpha x + \beta \tau} (\alpha^2 u + 2\alpha u_x + u_{xx}),$$

and dividing the resulting expression by $\exp(\alpha x + \beta \tau)$, we infer that

$$0 = (\beta u + u_{\tau}) - (\alpha^2 u + 2\alpha u_x + u_{xx}) + (1 - k)(\alpha u + u_x) + ku$$

= $u_{\tau} - u_{xx} + (-2\alpha + 1 - k)u_x + (-\alpha^2 + (1 - k)\alpha + \beta + k)u.$

The lower-order terms are eliminated if we choose $\alpha = \frac{1}{2}(1-k)$ and $\beta = -\frac{1}{4}(k+1)^2$, since

$$-\alpha^{2} + (1-k)\alpha + \beta + k = \frac{1}{4}(1-2k+k^{2}) - \frac{1}{4}(1+2k+k^{2}) + k = 0.$$

This means that the function

$$u(x,\tau) = \exp(-\alpha x - \beta \tau) = \exp\left(\frac{1}{2}(k-1)x + \frac{1}{4}(k+1)^2\tau\right)v(x,\tau)$$

solves the equation

$$u_{\tau} - u_{xx} = 0, \quad x \in \mathbb{R}, \ \tau \in (0, T_1],$$
 (3.12)

with the initial condition

$$u(x,0) = u_0(x) := e^{(k-1)x/2}(e^x - 1)^+ = \left(e^{(k+1)x/2} - e^{(k-1)x/2}\right)^+.$$
 (3.13)

Step 2: We know from the lectures on partial differential equations that the solution to (3.12)-(3.13) equals

$$u(x,\tau) = \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} u_0(z) e^{-(x-z)^2/(4\tau)} dz = \frac{1}{\sqrt{4\pi\tau}} \int_0^{\infty} u_0(z) e^{-(x-z)^2/(4\tau)} dz.$$

We formulate the integral in terms of the standard normal distribution Φ . Using the transformation $y = (z - x)/\sqrt{2\tau}$ and (3.13), we deduce that

$$u(x,\tau) = \frac{1}{\sqrt{2\pi}} \int_{-x/\sqrt{2\tau}}^{\infty} u_0(\sqrt{2\tau}y+x)e^{-y^2/2}dy$$

= $\frac{1}{\sqrt{2\pi}} \int_{-x/\sqrt{2\tau}}^{\infty} \exp\left(\frac{1}{2}(k+1)(x+y\sqrt{2\tau})\right)e^{-y^2/2}dy$

$$-\frac{1}{\sqrt{2\pi}}\int_{-x/\sqrt{2\tau}}^{\infty}\exp\left(\frac{1}{2}(k-1)(x+y\sqrt{2\tau})\right)e^{-y^2/2}dy.$$
 (3.14)

These integrals can be formulated in terms of Φ . Indeed, a computation shows that

$$\frac{1}{\sqrt{2\pi}} \int_{-x/\sqrt{2\tau}}^{\infty} \exp\left(\frac{1}{2}(k\pm 1)(x+y\sqrt{2\tau})\right) e^{-y^2/2} dy$$

= $\exp\left(\frac{1}{2}(k\pm 1)x + \frac{1}{4}(k\pm 1)^2\tau\right) \Phi(d_{1/2}).$

Consequently, we obtain

$$u(x,\tau) = \exp\left(\frac{1}{2}(k+1)x + \frac{1}{4}(k+1)^2\tau\right)\Phi(d_1)$$
$$-\exp\left(\frac{1}{2}(k-1)x + \frac{1}{4}(k-1)^2\tau\right)\Phi(d_2).$$

Step 3: Finally, we transform back:

$$\begin{split} V(S,t) &= Kv(x,\tau) = K\exp\left(-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau\right)u(x,\tau) \\ &= K\exp(x)\Phi(d_1) - K\exp\left(-\frac{1}{4}(k+1)^2\tau + \frac{1}{4}(k-1)^2\tau\right)\Phi(d_2) \\ &= S\Phi(d_1) - Ke^{-k\tau}\Phi(d_2) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2). \end{split}$$

Step 4: It remains to verify the final and boundary conditions. The final condition is fulfilled in the sense of the limit $t \to T$ since the denominator of $d_{1/2}$ is singular at t = T. Indeed, it follows in the limit $t \to T$ from

$$d_{1/2} \to \begin{cases} \infty & : S > K \\ 0 & : S = K \\ -\infty & : S < K \end{cases} \text{ and } \Phi(d_{1/2}) \to \begin{cases} 1 & : S > K \\ \frac{1}{2} & : S = K \\ 0 & : S < K \end{cases}$$

(see Figure 3.1) that

$$V(S,t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \to \left\{ \begin{array}{ccc} S-K & : & S>K \\ 0 & : & S \le K \end{array} \right\} = (S-K)^+$$

We turn to the boundary conditions. We have $d_{1/2} \to -\infty$ as $S \to 0$, thus $\Phi(d_{1/2}) \to 0$ and consequently, $V(S, t) \to 0$ as $S \to 0$. When $S \to \infty$, we find that $\Phi(d_{1/2}) \to 1$ and thus, $\Phi(d_2)/S \to 0$ from which we deduce that

$$\frac{V(S,t)}{S} = \Phi(d_1) - Ke^{-r(T-t)}\frac{\Phi(d_2)}{S} \to 1 \quad \text{as } S \to \infty.$$

Step 5: The formula for put options follows from that one for call options and the put-call parity. This completes the proof. \Box



The option premiums for a European call and put option at various times are illustrated in Figure 3.2. The distribution function Φ is computed by means of the error function erf, which is implemented in mathematical software packages according to



 $\Phi(x) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right), \text{ where } \operatorname{erf}(z) := \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$

Figure 3.2: Values of the European call option (left) and put option (right) at times t = 0, 0.2, 0.4, 0.6, 0.8, 1 with parameters K = 100, T = 1, r = 0.1, and $\sigma = 0.4$.

Remark. (Variational framework of the Black-Scholes equations). The derivation of the Black-Scholes formulas in Theorem 3.3 is based on the assumption that the interest rate r and the volatility σ are constant. It is possible to extend this approach to time-dependent functions r(t) and $\sigma(t)$, which still leads to explicit formulas; see [30]. However, such formulas do not exist in more general situations, e.g., when the coefficients r or σ are given by another equation or when they depend nonlinearly on the option price or its derivatives (which is the case when transaction costs are included; see Section 3.8). Then it is appropriate to work with the original Black-Scholes equation. Its mathematical treatment, however, is delicate since the

coefficient $\frac{1}{2}\sigma^2 S^2$ vanishes at S = 0 such that standard techniques from the theory of partial differential equations do not apply. We discuss briefly the mathematical framework to treat such situations. Details are presented in [1, Section 2.3].

Consider the Black-Scholes equation

$$V_t - \frac{1}{2}\sigma^2 S^2 V_{SS} - rSV_S + rV = 0, \quad S > 0, \ 0 < t < T,$$
(3.15)

where r > 0 and $\sigma > 0$ may depend on (S, t). Observe that we have reversed the time, $t \mapsto -t$, to make it an initial-boundary-value problem. The initial and boundary conditions are

$$V(S,0) = V_T(S), \quad V(0,t) = V_0(t), \quad \lim_{S \to \infty} V(S,t) = 0.$$

The boundary condition for $S \to \infty$ corresponds to a put option. The price of a call option can be obtained from the put-call parity. Since the coefficients of V_{SS} and V_S vanish at S = 0, we cannot expect that the derivatives exist in the standard sense, for instance in the weak sense $V_{SS}(t) \in L^2(\mathbb{R}_+)$ or $V_S(t) \in L^2(\mathbb{R}_+)$, where $L^2(\mathbb{R}_+)$ is the space of square integrable functions and $\mathbb{R}_+ = (0, \infty)$. This problem is overcome by introducing the space

$$X = \left\{ V \in L^2(\mathbb{R}_+) : SV_S \in L^2(\mathbb{R}_+) \right\},\$$

where the derivative is understood in the sense of distributions. This space becomes a Hilbert space endowed with the norm

$$\|V\|_X^2 = \int_0^\infty \left(V(S)^2 + S^2 V_S(S)^2 \right) dS$$

We derive a weak formulation for (3.15), where only first weak derivatives are contained. Let V be a smooth solution to (3.15) and multiply this equation by a test function $U \in C_0^{\infty}(\mathbb{R}_+)$. Then

$$0 = \int_0^\infty \left(V_t - \frac{1}{2} \sigma^2 S^2 V_{SS} - rSV_S + rV \right) U dS$$

=
$$\int_0^\infty \left(V_t U + \frac{1}{2} \sigma^2 S^2 V_S U_S + (\sigma \sigma_S S^2 + \sigma^2 S - rS) V_S U + rVU \right) dS.$$

The boundary integral vanishes since $\frac{1}{2}\sigma^2 S^2 V_S U = 0$ at S = 0 and for $S \to \infty$ (*U* has compact support on \mathbb{R}_+). This can be written as

$$\frac{d}{dt} \int_0^\infty VUdS + a(V, U) = 0, \quad \text{where}$$
$$a(V, U; t) = \int_0^\infty \left(\frac{1}{2}\sigma^2 S^2 V_S U_S + r V U\right) dS + \int_0^\infty \left(S\sigma\sigma_S + \sigma^2 - r\right) S V_S U dS$$

Under suitable assumptions on *r* and σ , it can be shown that the bilinear form *a* is continuous in the sense

 $|a(V, U)| \le C_1 ||V||_X ||U||_X$ for all $V, U \in X$,

and that it satisfies the Gårding inequality,

$$a(V,V) \ge C_2 \|V\|_X^2 - C_3 \|V\|_{L^2(\mathbb{R}_+)}^2$$
 for all $V \in X$,

which is a coercivity property. A general existence result then provides the existence of a solution $V \in C^0([0,T];L^2(\mathbb{R}_+)) \cap L^2(0,T;X)$ with $V_t \in L^2(0,T;X')$, where X' is the dual space of X.

Example. Since the end of 2017, there exist European options on the Bitcoin price. Bitcoin is a decentralized digital currency that is not backed by a central bank or government. Bitcoin-to-Bitcoin transactions are made by digitally exchanging anonymous encrypted codes. The Bitcoin network is designed to mathematically generate no more than 21 million Bitcoins. Currently, about 19 million Bitcoins are in circulation. The Bitcoin-Euro rate is highly volatile, and the annualized volatility has reached 500% in certain periods! Figure 3.3 (left) shows the Bitcoin price from December 2017 until June 2018. Its value changed dramatically over the last years; the first exchange rate for one Bitcoin was 0.08 USD cent only. The Bitcoin value is still very volatile; between 2018 and 2024 it changed in the range between 3000 and 64,000 EUR.

Bitcoins can be seen more as a trading asset than a currency. As popularity in the cryptocurrency grows, the products to trade the underlying asset widens, and Bitcoin futures and Bitcoin option contracts have been launched. European options in Bitcoins are the same as on any other asset, but the volatility is usually much higher compared to the volatility of stocks (which is typically around 15 ... 30%). We present in Figure 3.3 (right) the Black-Scholes values of a European call option for various volatilities. The call price increases significantly with the volatility, which can be interpreted as a measure of the uncertainty of the price at expiration time.



Figure 3.3: Left: Bitcoin values from 05 December 2017 until 03 June 2018. Right: Call option prices according to the Black-Scholes formula for $\sigma = 0.5, 1.0, 1.5, 2.0, r = 0.01, K = 10,000$, and T = 0.5 (6 months).

3.2 Risk-neutral valuation

The price V of an option can be determined by the expected value of the discounted payoff with respect to a modified probability measure, the so-called risk-neutral measure. In this section, we discuss this valuation.

Theorem 3.4 (Risk-neutral pricing formula). Let the assumptions of Theorem 3.2 hold with the exception that the interest rate $(r_t)_{t\geq 0}$ is a deterministic process and the drift $(\mu_t)_{t\geq 0}$ and volatility $(\sigma_t)_{t\geq 0}$ are adapted stochastic processes such that $dS_t = \mu_t S_t dt + \sigma_t S_t dW_t$. Then the arbitrage-free price of the option is

$$V_t = \exp\left(-\int_t^T r_s ds\right) \mathbf{E}^Q(V_T | \mathcal{F}_t), \quad 0 < t < T,$$

where *Q* is the risk-neutral measure and the expected value is taken with respect to *Q*.

According to the Girsanov theorem, the risk-neutral measure is defined by the Radon-Nikodym derivative

$$\left. \frac{dQ}{dP} \right|_t = \exp\left(-\frac{1}{2}\gamma^2 t - \gamma W_t \right), \tag{3.16}$$

where $\gamma := (\mu - r)/\sigma$ is called the *market price of risk*. This allows us to move from the "real-world" asset price dynamics

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

to the risk-neutral price dynamics

$$dS_t = rS_t dt + \sigma S_t dW_t^Q,$$

where W_t^Q is defined by $dW_t^Q = \gamma dt + dW$; see the proof below.

Proof. Let $D_t = \exp(-\int_0^t r_s ds)$ be the discount process. It satisfies $dD_t = -r_t D_r dt$. We rewrite the Itô process $(S_t)_{t\geq 0}$ as

$$dS_t = r_t S_t dt + \sigma_t S_t \left(\frac{\mu_t - r_t}{\sigma_t} dt + dW_t\right) = r_t S_t dt + \sigma_t S_t (\gamma_t dt + dW_t).$$

This motivates us to introduce the new process $dW^Q = \gamma dt + dW$. The interpretation of $\gamma = (\mu - r)/\sigma$ is that it is the amount which the investor expects as a compensation for the risk. Therefore, γ is called the market price of risk. The new process $(W_t^Q)_{t\geq 0}$ is not a Wiener process under P, but it can be shown (by the Girsanov theorem) that $(W_t^Q)_{t\geq 0}$ is a Wiener process under the probability measure Q, which is defined by (3.16).

We define the stochastic process $(X_t)_{t\geq 0}$ by

$$dX_t = \Delta_t dS_t + r_t (X_t - \Delta_t S_t) dt.$$

It can be interpreted as a portfolio with Δ_t shares of the stock and the remaining value $X_t - \Delta_t S_t$ is invested in a bond. This strategy is always self-financing since X_t is by definition the gain from the trade. We claim that $(D_t X_t)_{t\geq 0}$ is a martingale under Q. Indeed, using $dS_t = r_t S_t dt + \sigma_t S_t dW_t^Q$ and the product rule (Lemma 2.14), we find that

$$d(D_tX_t) = D_t dX_t + X_t dD_t = D_t (\Delta_t dS_t + r_t (X_t - \Delta_t S_t) dt) - r_t D_t X_t dt$$

= $D_t \Delta_t (r_t S_t dt + \sigma_t S_t dW_t^Q) - r_t D_t \Delta_t S_t dt = D_t \Delta_t \sigma_t S_t dW_t^Q.$

Suppose that X_t is the portfolio replicating the option. By the martingale representation theorem (guaranteeing that any martingale with respect to the Brownian filtration can be expressed as an Itô integral with respect to W^Q), such a process always exists. Then $V_t = X_t$, and by definition of a martingale, we infer that for $t \leq T$,

$$V_t = \frac{1}{D_t}(D_t X_t) = \frac{1}{D_t} \mathbf{E}^{\mathbb{Q}}(D_T X_T | \mathcal{F}_t) = \frac{D_T}{D_t} \mathbf{E}^{\mathbb{Q}}(V_T | \mathcal{F}_t),$$

which concludes the proof.

Remark. The value V_t in Theorem 3.4 can be computed explicitly when r and σ are constant. For this, we insert definition (3.14) of $u(x, \tau)$ and definition (3.13) of $u_0(x)$ to obtain

$$\begin{split} V(S,t) &= K \exp\left(-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau\right)u(x,\tau) \\ \stackrel{(3.14)}{=} \frac{K}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau\right)\int_{-\infty}^{\infty}u_0(x+\sqrt{2\pi}y)e^{-y^2/2}dy \\ \stackrel{(3.13)}{=} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau\right) \\ &\qquad \times \int_{-\infty}^{\infty} \exp\left(\frac{1}{2}(k-1)(x+\sqrt{2\pi}y) - \frac{y^2}{2}\right)\left(Ke^{x+\sqrt{2\pi}y} - K\right)^+dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{4}(k+1)^2\tau + \frac{1}{2}(k-1)\sqrt{2\tau}y - \frac{y^2}{2}\right)\left(Se^{\sqrt{2\tau}y} - K\right)^+dy. \end{split}$$

Then, transforming back with $y \mapsto S' := Se^{\sqrt{2\tau}y}$, a computation shows that

$$V(S,t) = \mathrm{E}^{Q} \big(e^{-r(T-t)} V_{T}(S) | \mathcal{F}_{t} \big),$$

where $E^Q(V_T(S)|\mathcal{F}_t) = \int_0^\infty f(S'; S, t) V_T(S) dS' / S'$ is the expected value of V_T with respect to the density function

$$f(S'; S, t) = \frac{1}{\sigma\sqrt{2\pi(T-t)}} \exp\left(-\frac{[\ln(S'/S) - (r-\sigma^2/2)(T-t)]^2}{2\sigma^2(T-t)}\right)$$

of the log-normal distribution for $S = S_t$.

- **Remark.** There exists a close relationship between the risk-neutral measure and the market model. For this, let the market model consist of stocks S_t and risk-free bonds B_t . We say that a contingent claim X (a nonnegative, \mathcal{F}_T measurable random variable) is *attainable* if there exists a self-financing trading strategy, i.e., there exist (a_t, b_t) such that $a_T B_T + b_T S_T = X$. An arbitrage-free market model is *complete* if every contingent claim is attainable. Then the following two results hold:
 - First fundamental theorem of asset pricing: A market model on a discrete probability space (Ω, \mathcal{F}, P) is arbitrage-free if and only if there exists at least one risk-neutral measure that is equivalent to P.
 - Second fundamental theorem of asset pricing: An arbitrage-free market model is complete if and only if there exists a unique risk-neutral measure equivalent to P.

We refer to [11, 14, 28] for details and proofs.

The fact that the option price can be determined from either the Black-Scholes formula or the expectation is not a surprise. Indeed, both formulations are related by means of the Feynman-Kac formula.

Theorem 3.5 (Feynman-Kac formula). Let $u : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$ be a solution to the backward parabolic equation

$$\partial_t u + a(x,t)u_x + \frac{1}{2}b(x,t)^2 u_{xx} - r(x,t)u = 0, \quad 0 < t < T, \quad u(T) = g,$$
 (3.17)

where *a*, *b*, *r*, *g* are suitable smooth functions with at most linear growth in x. Furthermore, let $(X_t)_{t \in [0,T]}$ be an Itô process satisfying

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t, \quad 0 < t < T, \quad X_t = x.$$
(3.18)

Then u can be written as the expectation

$$u(x,t) = \mathbb{E}\bigg(\exp\bigg(-\int_t^T r(X_s,s)ds\bigg)g(X_T)\bigg|X_t = x\bigg).$$
(3.19)

The converse is also true: Any Itô process satisfying (3.18) *defines via* (3.19) *a solution to the backward parabolic equation* (3.17).

This theorem provides a link between partial differential equations and diffusion processes. Solutions to partial differential equations can be interpreted as expectations of suitable transformations of solutions to stochastic differential equations and vice versa.

Remark. Equation (3.17) is backward in time. For forward equations, we transform $t \mapsto T - t$ and obtain the following version of the Feynman-Kac formula: Let $(X_t)_{t \in [0,T]}$ solve (3.18)
with $X_0 = x$. Then

$$u(x,t) = \mathbb{E}\left(\exp\left(\int_0^t r(X_s,s)ds\right)g(X_t)\right)$$

solves $\partial_t u - a(x,t)u_x - \frac{1}{2}b(x,t)^2u_{xx} + r(x,t)u = 0$ for $0 < t < T$ and $u(0) = g$

3.3 Implied volatility

While the interest rate of a riskfree bond can be rather easily determined from market data, it is more involved to compute the volatility σ . One approach is to consider historical data and to compute the *historical volatility* σ_{hist} . It is the annualized standard deviation of the logarithmic asset values, $Y_i = \ln(S_{i+1}/S_i)$, where S_i is the price of the asset at day t_i . Then

$$\sigma_{\text{hist}} = \sqrt{N} \left(\frac{1}{n-1} \sum_{i=1}^{n-1} (Y_i - \overline{Y})^2 \right)^{1/2}, \text{ where } \overline{Y} = \frac{1}{n-1} \sum_{i=1}^{n-1} Y_i,$$

where *n* is the number of sample points and *N* is the average number of banking days (often, N = 252 is taken). Other definitions take into account that older data are less relevant and introduce a weight function.

Another approach to determine σ in the Black-Scholes formula is to compute the *implied volatility* implicit in the option value. We write $d_{1/2}(\sigma)$ to highlight that these parameters, defined in (3.10), depend on σ . For a given option value C_0 , we need to solve the equation

$$C_0 = C(\sigma) := S\Phi(d_1(\sigma)) - Ke^{-r(T-t)}\Phi(d_2(\sigma)).$$

There exists a unique solution, at least as long as the arbitrage bounds $(S - Ke^{-r(T-t)})^+ \le C_0 \le S$ are fulfilled, and we call the solution σ_{impl} , i.e. $C(\sigma_{\text{impl}}) = C_0$. This statement follows from the intermediate value theorem applied to the function $\sigma \mapsto C(\sigma)$, since $\partial C / \partial \sigma > 0$ and consequently, $C(\sigma) - C_0 \le 0$ for $\sigma = 0$ and $C(\sigma) - C_0 \ge 0$ as $\sigma \to \infty$. The unique zero of $f(\sigma) := C(\sigma) - C_0$ can be calculated iteratively by using the Newton method. Let $\sigma_0 > 0$ be an initial guess and define

$$\sigma_{k+1} = \sigma_k - rac{f(\sigma_k)}{f'(\sigma_k)}, \quad k \in \mathbb{N} \cup \{0\}.$$

The derivative can be computed: $f'(\sigma_k) = (\partial C/\partial \sigma)(\sigma_k) = S\sqrt{T-t}\Phi'(d_1(\sigma_k))$. It is possible to show that (σ_k) converges as $k \to \infty$, and the limit is the implied volatility.

When we calculate the implied volatility from market data on an underlying with the same expiration date but different strike prices, we observe that σ_{impl} is not constant. It turns out that the function $K \mapsto \sigma_{impl}$ forms a "smile', i.e., the implied volatility is smallest *at the money* and it becomes larger *in the money* or *out of the money*. These notions mean the following:

- ► At-the-money: The strike price is identical (or close to) the price of the underlying.
- In-the-money: The strike price is lower (higher) than the price of the underlying when we consider call (put) options. Being in the money means that the option is worth exercising.
- Out-of-the-money: The strike price is higher (lower) than the price of the underlying when we consider call (put) options. An option that is out of the money may likely have no value at expiration date.

This observation shows that the standard Black-Scholes model – strictly speaking – fails in real markets. Yet, the Black-Scholes formulas are still used extensively in practice. More elaborated models have been developed using a stochastic volatility instead of a constant volatility (see Section 3.6) or using jump models instead of the Brownian motion (see Section 3.8).

As an example of the *volatility smile*, we present in Figure 3.4 the values of the implied volatility versus the strike price of a call option on the DAX index, issued by the Deutsche Bank, with expiration time of three months. We clearly see that the volatility is large in-the-money and out-of-the-money. Interestingly, the volatility is smallest (at $K \approx 14,000$) not exactly at the money (the DAX index was about 13,000).



Remark. The volatility can be an important measure of the state of the financial market. Large values of the volatility come from large fluctuations of the price of the underlying. This increases the risk of valuating the asset, and it is reasonable that this increases the call option price. The change of the call price with respect to the volatility is called *vega* or *kappa*, and it can be computed explicitly from the Black-Scholes formula:

$$\kappa = \frac{\partial C}{\partial \sigma} = \frac{\partial P}{\partial \sigma} = S \sqrt{\frac{T-t}{2\pi}} \exp\left(-\frac{d_1^2}{2}\right),$$

where d_1 is given in (3.10). We see that V(S, t) increases with the volatility since $\kappa > 0$.

3.4 Options on dividend paying assets

One of our market model assumptions was that there are no dividend payments. We extend the Black-Scholes model to assets with continuous or discrete dividend payments.

• **Continuous dividend payments.** Options on an index involve many stocks with dividend payments which are approximately distributed all along the year. Therefore, an easy approach is to assume that these payments are continuous in time. Furthermore, we suppose that the dividend yield δ is proportional to the price of the asset *S*, i.e., we receive the dividend $\delta S \Delta t$ after time Δt . By the no-arbitrage principle, the price of the asset decreases by exactly this value. Otherwise, we would make an instantaneous risk-free profit of amount $\delta S \Delta t$ by purchasing the asset at time *t* and selling it immediately after having received the dividend. Thus, the drift μ in the asset price model needs to be changed to $\mu - \delta$:

$$dS = (\mu - \delta)Sdt + \sigma SdW,$$

and we need to include the dividend payment in the variation of the self-financing portfolio (3.3):

$$d\pi = c_1 dB + c_2 (dS + \delta S dt) - dV(S, t).$$

Choosing $c_2 = V_S$, applying Itô's formula, and using dB = rBdt, the same computation as in Section 3.1 leads to

$$d\pi = \left(c_1 r B + \delta S V_S - V_t - \frac{1}{2}\sigma^2 S^2 V_{SS}\right) dt$$

As the portfolio is riskless, we have

$$d\pi = r\pi dt = r(c_1B + SV_S - V)dt,$$

and equating these identities, we arrive at the modified Black-Scholes equation

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + (r - \delta)SV_S - rV = 0, \quad S \in (0, \infty), \ t \in (0, T).$$
(3.20)

The final condition is as before, but we need to modify the boundary conditions. We only consider call options since the corresponding expressions for put options can be derived from the modified put-call parity

$$Se^{-\delta(T-t)} + P_t - C_t = Ke^{-r(T-t)},$$

which takes into account the discount factor for the asset price. When S = 0, the call option has no value, while as $S \rightarrow \infty$, V(S, t) approaches the discounted asset price:

$$V(0,t) = 0, \quad V(S,t) \sim Se^{-\delta(T-t)} \quad \text{as } S \to \infty.$$
 (3.21)

This problem can be solved explicitly.

Proposition 3.6 (Call option price with continuous dividends). *The solution to problem* (3.20)-(3.21) *with final condition* $C(S,T) = (S - K)^+$ *reads as*

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

where

$$d_{1/2}^{\delta} = \frac{\ln(S/K) + (r - \delta \pm \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}.$$

Remark. It can be seen that the call option on an asset *with* dividend payments is always cheaper than the corresponding one on an asset *without* dividend payments. The corresponding price for a put option equals

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

Proof. The idea is to define the new variable $C^*(S,t) := e^{\delta(T-t)}C(S,t)$. It fulfills the Black-Scholes equation

$$C_t^* + \frac{1}{2}\sigma^2 S^2 C_{SS}^* + (r-\delta)SC_S^* - rC^*$$

= $e^{\delta(T-t)} \left(C_t - \delta C + \frac{1}{2}\sigma^2 S^2 C_{SS} + (r-\delta)SC_S - rC \right)$
= $-\delta e^{\delta(T-t)}C = -\delta C^*$,

with the final condition $C^*(S, T) = (S - K)^+$ and the initial conditions

 $C^*(0,t) = 0$, $C^*(S,t) \sim S$ $(S \to \infty)$.

This is exactly the Black-Scholes problem with interest rate $r - \delta$. Its solution is given by

$$C(S,t) = e^{-\delta(T-t)}C^{*}(S,t) = e^{-\delta(T-t)} \left(S\Phi(\delta_{1}^{\delta}) - Ke^{-(r-\delta)(T-t)}\Phi(d_{2}^{\delta}) \right),$$

which yields the statement of the proposition.

• **Discrete dividend payments.** When the underlying of the option is a stock, the dididend is usually paid once per year. In this situation we need to consider discrete dividend payments. Let this payment happen at time $t = t_d$ during the lifetime of the option. The following arguments can be extended to a finite number of payments, but we restrict ourselves to a single payment to simplify the notation.

At time $t = t_d$, the holder of the stock receives the dividend $d \cdot S$, where $d \in [0, 1)$ is the dividend rate and *S* is the stock value just before the dividend payment. As before,

because of the no-arbitrage principle, the asset value must decrease right after $t = t_d$ by the amount $d \cdot S$,

$$S(t_d^+) = (1-d)S(t_d^-),$$

where

$$S(t_d^+) = \lim_{t \searrow t_d} S_t, \quad S(t_d^-) = \lim_{t \nearrow t_d} S_t$$

are the limits from above and below, respectively. Clearly, we assume that such limits exist almost surely. Note that if d > 0, the asset price is discontinuous. However, the option price is continuous in time since the option holder does not receive any dividend payment. This means that

$$V(S(t_d^{-}), t_d^{-}) = V(S(t_d^{+}), t_d^{+}) = V((1-d)S(t_d^{-}), t_d^{+}).$$

Since we consider all possible realizations of the stochastic process S_t , we require that

$$V(S, t_d^-) = V((1 - d)S, t_d^+) \quad \text{for all } S > 0.$$
(3.22)

The option value can be computed from the Black-Scholes equation in the intervals $(t_d, T]$ and $[0, t_d)$ except at the point $t = t_d$. At $t = t_d$, we use the jump condition (3.22). More precisely, the algorithm is as follows:

- ► Solve the Black-Scholes equation in $[t_d, T]$ with final condition $V(S, T) = V_T(S)$. This gives the value $V(S, t_d^+)$.
- ▶ Define $V(S, t_d^-)$ according to (3.22).
- ▶ Solve the Black-Scholes equation in $[0, t_d]$ with final condition $V(S, t_d^-)$.

Example. We determine the premium on a European call option with expiration date *T* and strike *K* on a stock with a single dividend payment at time t_d . We denote by $C_d(S, t)$ the call price and by $C_E(S, t; K)$ the call price on a European call option with the same specification but on a stock without dividend payment. We know that

$$C_d(S,t) = C_E(S,t;K)$$
 for $t_d < t \le T$.

At time $t = t_d$, the interface condition (3.22) reads as

$$C_d(S, t_d^-) = C_d((1-d)S, t_d^+) = C_E((1-d)S, t_d^+; K).$$

It remains to determine the option value for $0 \le t < t_d$. We claim that

$$C_d(S,t) = (1-d)C_E(S,t;K/(1-d))$$
 for $0 \le t < t_d$.

To show this statement, we consider the function $C^*(S,t) := C_E((1-d)S,t;K)$. It solves the Black-Scholes equation since the factor 1 - d cancels. We compute the final and initial conditions:

$$C^*(S,T) = ((1-d)S - K)^+ = (1-d)\left(S - \frac{K}{1-d}\right)^+,$$

$$C^*(0,t) = 0, \quad C^*(S,t) \sim (1-d)S \quad \text{as } S \to \infty.$$

As the Black-Scholes equation is uniquely solvable, this shows that $C^*(S,t)$ equals 1 - d times the price of a European option with strike K/(1 - d), i.e. $C^*(S,t) = (1 - d)C_E(S,t;K/(1 - d))$, which is our claim. The solution is illustrated in Figure 3.5 for a call option with dividend rate d = 0.05 and dividend payment at $t_d = 0.6$.

Figure 3.5: Values of the European call option C_d at times t = 0, 0.2, 0.4 (solid lines), 0.6, 0.8, 1 (dashed lines) with parameters $K = 100, T = 1, r = 0.1, \sigma = 0.4$. The dividend with rate d = 0.05 is paid at $t_d = 0.6$.



3.5 Multi-asset options

The Black-Scholes equation derived in Section 3.1 determines the price of a European option on a single asset. In this subsection, we consider options on several underlyings S_i . The processes $(S_i(t))_{t>0}$ are assumed to satisfy

$$dS_i = \mu_i S_i dt + \sigma_i S_i dW_i, \quad i = 1, \dots, n,$$

where $W_1, ..., W_n$ are correlated Wiener processes with covariance matrix Σ satisfying $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \rho_{ij}$ for $i \neq j$. To derive a multidimensional Black-Scholes equation, we consider the riskless and self-financing portfolio

$$\pi_t = c_0(t)B_t + \sum_{i=1}^n c_i(t)S_i(t) - V(S_t, t),$$

where $c_0, ..., c_n$ are the shares of the corresponding assets, the bond process solves dB = rBdt, and $V(S_t, t)$ is the value of a European option. We repeat the arguments used for the derivation of the standard Back-Scholes equation. Since the portfolio is assumed to be riskless and self-financing, we obtain

$$d\pi = r\pi dt = r \left(c_0 B + \sum_{i=1}^n c_i S_i - V \right) dt,$$
(3.23)

$$d\pi = c_0 dB + \sum_{i=1}^n c_i dS_i - dV = c_0 rB dt + \sum_{i=1}^n c_i (\mu_i S_i dt + \sigma_i S_i dW_i) - dV.$$
(3.24)

By the multidimensional Itô formula, we can compute dV (see (2.7)):

$$dV(S,t) = \left(V_t + \sum_{i=1}^n \mu_i S_i V_{S_i} + \frac{1}{2} \sum_{i,j=1}^n \rho_{ij} \sigma_i \sigma_j S_i S_j V_{S_i S_j}\right) dt + \sum_{i=1}^n \sigma_i S_i V_{S_i} dW_i.$$

Inserting this expression into (3.24) and equating (3.23) and (3.24), it follows that

$$c_{0}rBdt + \sum_{i=1}^{n} rc_{i}S_{i}dt - rVdt = c_{0}rBdt + \sum_{i=1}^{n} c_{i}\mu_{i}S_{i}dt + \sum_{i=1}^{n} \sigma_{i}S_{i}(c_{i} - V_{S_{i}})dW_{i} - \left(V_{t} + \sum_{i=1}^{n} \mu_{i}S_{i}V_{S_{i}} + \frac{1}{2}\sum_{i,j=1}^{n} \rho_{ij}\sigma_{i}\sigma_{j}S_{i}S_{j}V_{S_{i}}S_{j}\right)dt$$

Then choosing $c_i = V_{S_i}$, the stochastic part cancels, and we end up with a deterministic equation. We summarize this result in the following theorem.

Theorem 3.7 (Multidimensional Black-Scholes equation). Let the assumptions at the beginning of this section hold. Then the option price is a solution to the partial differential equation

$$V_t + \frac{1}{2} \sum_{i,j=1}^n \rho_{ij} \sigma_i \sigma_j S_i S_j V_{S_i S_j} + r \sum_{i=1}^n S_i V_{S_i} - rV = 0$$

for $S_i \in (0, \infty)$, t > 0, and the final datum is given by $V(S, T) = V_T(S)$ for $S \in [0, \infty)^n$.

Example (Basket option). As as application, we consider *basket options*, which are contracts on *n* assets with the payoff function $V_T(S) = V_T(S_1, ..., S_n)$. Similarly as explained in the remark on page 35, the option price according to the Black-Scholes equation can be formulated as an integral with respect to some density function:

$$V(S,t) = \int_{(0,\infty)^n} f(S';S,t) V_T(S) \frac{dS'}{S'_1 \cdots S'_n},$$
(3.25)

where the density function reads as

$$f(S'; S, t) = \frac{e^{-r(T-t)}}{(\sigma_1 \cdots \sigma_n)\sqrt{(2\pi(T-t))^n \det \Sigma}} \exp\left(-\frac{1}{2}z^\top \Sigma^{-1}z\right),$$

where $z_i = \frac{\ln(S'_i/S_i) - (r - \sigma_i^2/2)(T-t)}{\sigma_i\sqrt{T-t}}, \quad z = (z_1, \dots, z_n)^\top.$

The payoff of the basket option may be defined by the payoff of a vanilla option

associated to the weighted value of the basket. For instance, in case of a call option, we may set

$$V_T(S) = \left(\sum_{i=1}^n \alpha_i S_i - K\right)^+,$$

where $\alpha_i \ge 0$ are some weights. Another example of a payoff is given by the maximum of the payoffs of the single-asset options, i.e.

$$V_T(S) = \max\{(S_i - K)^+ : i = 1, ..., n\}$$

The difficulty in the valuation of basket options is the integration of the highdimensional Black-Scholes equation or the high-dimensional integral (3.25). Indeed, when the basket consists of the DAX index with 30 stocks, we need to calculate a 30-dimensional integral. Even if we have only three base points in each of the 30 directions, we need to perform $3^{30} \approx 2 \cdot 10^{14}$ evaluations, which is extremely time consuming. A possible way out is the use of Monte-Carlo or quasi Monte-Carlo simulations, which are discussed in Section 5.

The determination of the correlations ρ_{ij} is usually a delicate task. A possibility is to estimate them from historical data. For instance, consider the case of two assets with correlation $\rho := \rho_{12} = \rho_{21}$. Let $S_i(t_0), \ldots, S_i(t_m)$ be observed prices of the *i*th asset at the times t_0, \ldots, t_m . The correlation between S_1 and S_2 is estimated by the asset log return instead of the price:

$$\widehat{\rho} = \frac{\sum_{k=1}^{n} Y_1^k Y_2^k}{\sqrt{\sum_{k=1}^{n} (Y_1^k)^2 \sum_{k=1}^{n} (Y_2^k)^2}}, \text{ where}$$
$$Y_i^k = \ln \frac{S_i(t_k)}{S_i(t_{k-1})} - \frac{1}{m} \sum_{j=1}^{m} \ln \frac{S_i(t_j)}{S_i(t_{j-1})}, \quad i = 1, 2, \ k = 1, \dots, m.$$

3.6 Stochastic interest rate and volatility

We have assumed that the interest rate and the volatility are given functions. Generally, these functions are not deterministic, and it is reasonable to model them as stochastic processes. Let us first consider stochastic interest rate models (with deterministic volatility). One class of models is given by

$$dr_t = \kappa(\theta - r_t)dt + \sigma_t r_t^{\alpha} dW_t, \quad t > 0,$$

where $\kappa > 0$, $\theta \ge 0$, and $\alpha \ge 0$ are constant model parameters. The Wiener process \widetilde{W}_t may be different from the Wiener process W_t in the equation for the asset price, and both Wiener processes may or may not be correlated. When the interest rate r_t is smaller (larger) than θ , the drift $\kappa(\theta - r_t)$ is positive (negative), and we expect increasing

(decreasing) interest rates. This means that there is a tendency to return to the mean value θ (disregarded stochastic fluctuations). This effect is called *mean reversion*. Two examples for α are

- $\alpha = 0$: Vasicek model [34] (also called Ornstein-Uhlenbeck process),
- $\alpha = \frac{1}{2}$: Cox-Ingersoll-Ross (CIR) model [9].

Due to the stochastic component that may be negative, the solution to the Vasicek model may become negative. Although negative interest rates have been observed in the course of the financial crisis after 2007, one may prefer to avoid such an effect. Negative values of r_t are avoided in the CIR model since the stochastic term vanishes when $r_t = 0$. More precisely, the interest rate stays positive if $\kappa \theta > \sigma_t^2/2$ [35, Section 40.8.2].

Next, we consider stochastic volatility models. Usually, they are also of mean-reversion type:

$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \lambda_t \sigma_t^{2lpha} d\widetilde{W}_t.$$

The mean value θ may be constant or time-dependent. Examples are

- ► $\alpha = 1$: Hull-White model [19],
- $\alpha = \frac{1}{2}$: Heston model [17].

We wish to derive a Black-Scholes equation taking into account the Heston model. The dynamics is given by

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t, \quad dU_t = \kappa (\theta - U_t) dt + \lambda_t \sqrt{U_t} dW_t, \quad (3.26)$$

where $U_t := \sigma_t^2$. The option price *V* depends on the asset price *S*, the variance *U*, and the time *t*.

Theorem 3.8 (Heston-Black-Scholes equation). Let the assumptions stated at the beginning of Section 3.1 hold and consider the Heston market model (3.26) with the correlation ρ between W_t and \widetilde{W}_t (i.e. formally $dWd\widetilde{W} = \rho dt$). Then there exists a function $\gamma = \gamma(S, U, t)$ such that the option price is a solution to the modified Black-Scholes equation

$$V_t + \frac{1}{2}US^2V_{SS} + \rho\lambda UV_{SU} + \frac{1}{2}\lambda^2 UV_{UU} + rSV_S - rV = \left(\kappa(\theta - U) - \gamma\lambda\sqrt{U}\right)V_U,$$

where $S \in (0, \infty)$ *,* $U \in (0, \infty)$ *, and* $t \in (0, T)$ *.*

The function $\gamma(S, U, t)$ is called the market price of volatility risk. It needs to be determined from additional information on the market.

Proof. We derive the Black-Scholes equation for V(S, U, t) by hedging a suitable riskless self-financing portfolio π . Since we have two sources of randomness, we need to hedge the option with two other contracts, the stock *S* and another option \tilde{V} depending on *S*, *U*, and *t*:

$$\pi = c_1 S + c_2 \tilde{V} - V.$$

By assumption, the portfolio is self-financing, so its change is given by

$$d\pi = c_1 dS + c_2 dV - dV. (3.27)$$

By Itô's lemma,

$$dV = V_t dt + V_S dS + V_U dU + \frac{1}{2} V_{SS} dS^2 + V_{SU} dS dU + \frac{1}{2} V_{UU} dU^2.$$

Squaring the changes of the processes S_t and U_t , we obtain formally

$$dS^{2} = (\mu Sdt + \sigma SdW)^{2} = \sigma^{2}S^{2}dt,$$

$$dU^{2} = (\kappa(\theta - U)dt + \lambda\sqrt{U}d\widetilde{W})^{2} = \lambda^{2}Udt,$$

$$dSdU = \sigma\lambda S\sqrt{U}dWd\widetilde{W} = \rho\sigma\lambda S\sqrt{U}dt.$$

Consequently, it follows that

$$dV = V_t dt + V_S dS + V_U dU + \left(\frac{1}{2}\sigma^2 S^2 V_{SS} + \rho\sigma\lambda S\sqrt{U}V_{SU} + \frac{1}{2}\lambda^2 UV_{UU}\right)dt,$$

$$d\widetilde{V} = \widetilde{V}_t dt + \widetilde{V}_S dS + \widetilde{V}_U dU + \left(\frac{1}{2}\sigma^2 S^2 \widetilde{V}_{SS} + \rho\sigma\lambda S\sqrt{U}\widetilde{V}_{SU} + \frac{1}{2}\lambda^2 U\widetilde{V}_{UU}\right)dt.$$

Inserting these equations into (3.27) gives

$$d\pi = -\left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + \rho\sigma\lambda S\sqrt{U}V_{SU} + \frac{1}{2}\lambda^2 UV_{UU}\right)dt$$

+ $c_2\left(\widetilde{V}_t + \frac{1}{2}\sigma^2 S^2 \widetilde{V}_{SS} + \rho\sigma\lambda S\sqrt{U}\widetilde{V}_{SU} + \frac{1}{2}\lambda^2 U\widetilde{V}_{UU}\right)dt$
+ $(c_1 + c_2\widetilde{V}_S - V_S)dS + (c_2\widetilde{V}_U - V_U)dU.$ (3.28)

We eliminate the randomness appearing in *dS* and *dU* by setting

$$c_1+c_2\widetilde{V}_S-V_S=0, \quad c_2\widetilde{V}_U-V_U=0,$$

which determines c_1 and c_2 :

$$c_2 = rac{V_U}{\widetilde{V}_U}, \quad c_1 = V_S - rac{V_U}{\widetilde{V}_U}\widetilde{V}_S.$$

Since the portfolio is riskless, we also have

$$d\pi = r\pi dt = r(c_1 S + c_2 \widetilde{V} - V)dt$$

Equating this expression with (3.28) and inserting the relations for c_1 and c_2 leads to

$$\frac{1}{V_U}\left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + \rho\sigma\lambda S\sqrt{U}V_{SU} + \frac{1}{2}\lambda^2 UV_{UU} + rSV_S - rV\right)$$

$$=\frac{1}{\widetilde{V}_{U}}\bigg(\widetilde{V}_{t}+\frac{1}{2}\sigma^{2}S^{2}\widetilde{V}_{SS}+\rho\sigma\lambda S\sqrt{U}\widetilde{V}_{SU}+\frac{1}{2}\lambda^{2}U\widetilde{V}_{UU}+rS\widetilde{V}_{S}-r\widetilde{V}\bigg).$$

The left-hand side only depends on V, the right-hand side only on \tilde{V} . We can choose any option \tilde{V} . Therefore, this identity can only hold if both sides are independent of the contract type and in some sense "constant". Still, this "constant" may depend on the independent variables *S*, *U*, and *t*. Denoting this "constant" by $\gamma_0 = \gamma_0(S, U, t)$, it follows that

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + \rho \sigma \lambda S \sqrt{U} V_{SU} + \frac{1}{2}\lambda^2 U V_{UU} + r S V_S - r V = \gamma_0 V_U.$$

Analogous to the introduction of the risk-neutral measure in the proof of Theorem 3.4, we introduce the market price of volatility risk as

$$\gamma = \frac{\widetilde{\mu} - \gamma_0}{\widetilde{\sigma}}$$
, where $\widetilde{\mu} = \kappa(\theta - U)$, $\widetilde{\sigma} = \lambda \sqrt{U}$,

which yields $\gamma_0 = \tilde{\mu} - \tilde{\sigma}\gamma = \kappa(\theta - U) - \lambda\sqrt{U}\gamma$. Inserting this expression in the previous Black-Scholes equation concludes the proof.

3.7 Application: Pricing Asian options

Asian options are contracts whose payoff depends on the average of the values of the underlying. They are useful if, for instance, price fluctuations of the underlying are high due to low liquidity in the market or due to price manipulations. There are various possibilites how to compute the average (discrete or continuous, arithmetic or geometric):

Arithmetic averages:

$$\overline{S} = \frac{1}{n} \sum_{i=1}^{n} S(t_i), \quad \overline{S} = \frac{1}{T} \int_0^T S_t dt,$$

► Geometric averages:

$$\overline{S} = \left(\prod_{i=1}^{n} S(t_i)\right)^{1/n}, \quad \overline{S} = \exp\left(\frac{1}{T}\int_0^T \ln S_t dt\right).$$

The payoff function involves such an average and either the strike or the final asset price, which distinuishes two types of Asian options:

- Fixed strike: call $C_T = (\overline{S} K)^+$, put $P_T = (K \overline{S})^+$,
- ▶ Floating strike: call $C_T = (S_T \overline{S})^+$, put $P_T = (\overline{S} S_T)^+$.

We wish to derive a Black-Scholes equation for Asian options of European type whose payoff is a function of S_t and the average $I_t = \int_0^t f(S_\tau, \tau) d\tau$, i.e. $V_T = V_T(S, I)$. For instance, we have f(S,t) = S and $V_T = (S - I/T)^+$ in the case of an *arithmeticaverage floating strike call option*. The dynamics of the asset price is as usual given by

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t, \quad t > 0.$$

Note that the average I_t can be interpreted also as a stochastic process since

$$dI_t = f(S_t, t)dt, \quad t > 0.$$

Thus, the pair (S_t, I_t) is an Itô process.

Theorem 3.9 (Black-Scholes equation for Asian options). Let the assumptions stated at the beginning of Section 3.1 hold. Then the option price is a solution to the modified Black-Scholes equation

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + f(S,t)V_I + rSV_S - rV = 0, \quad S > 0, \ I > 0, \ t \in (0,T).$$
(3.29)

Proof. We consider the portfolio $\pi = c_1B + c_2S - V(S, I, t)$, where the bond *B* has the riskless interest rate $r \ge 0$. Assuming that the portfolio is riskless and self-financing, we deduce from Itô's formula, after a similar computation as in the proof of Theorem 3.8, that

$$d\pi = c_1 dB + c_2 dS - dV(S, I, t)$$

= $c_1 rBdt + (c_2 - V_S)dS - \left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + f(S, t)V_I\right)dt$,
 $d\pi = r\pi dt = r(c_1 B + c_2 S - V)dt$.

Equating both equations and setting $c_2 = V_S$, it follows that

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + f(S,t)V_I + rSV_S - rV = 0,$$

which ends the proof.

Equation (3.29) is a partial differential equation. It is generally not of parabolic type since the derivative V_{II} is missing. Therefore, its (numerical) solution is delicate. However, when we consider arithmetic averages and a special payoff, we can reduce the Black-Scholes equation for the two variables (S, I) to a parabolic equation of *one* variable by a similarity reduction technique.

Proposition 3.10 (Arithmetic-average Asian option). Let *V* be a solution to (3.29) with f(S,t) = S and let the final datum given by $V_T(S,I) = S^{\alpha}F(I/S)$, where $\alpha \in \mathbb{R}$ and the function F = F(R) with R = I/S are given. Then $H(R,t) = S^{-\alpha}V$ solves the equation

$$H_t + \frac{1}{2}\sigma^2 R^2 H_{RR} + \left(1 - \sigma^2(\alpha - 1)R - rR\right)H_R + (\alpha - 1)\left(\frac{\alpha}{2}\sigma^2 + r\right)H = 0, \quad (3.30)$$

where R > 0, $t \in (0, T)$, and the final condition is H(R, T) = F(R).

Proof. The result follows by differentiating $V(S, I, t) = S^{\alpha}H(I/S, t)$,

$$\begin{split} V_t &= S^{\alpha} H_t, \quad V_I = S^{\alpha - 1} H_R, \\ V_S &= \alpha S^{\alpha - 1} H - S^{\alpha - 2} I H_R = S^{\alpha - 1} (\alpha H - R H_R), \\ V_{SS} &= \left(\alpha (\alpha - 1) S^{\alpha - 2} H - \alpha S^{\alpha - 3} I H_R \right) - \left((\alpha - 2) S^{\alpha - 3} I H_R - S^{\alpha - 4} I^2 H_{RR} \right) \\ &= S^{\alpha - 2} \left(\alpha (\alpha - 1) H - 2 (\alpha - 1) R H_R + R^2 H_{RR} \right), \end{split}$$

and inserting these expressions into (3.29).

Example. Equation (3.30) simplifies for an *arithmetic-average floating strike call option*. Indeed, when $V_T(S, I) = (S - I/T)^+ = S(1 - (I/S)/T)^+$, we have $\alpha = 1$ and $F(R) = (1 - R/T)^+$, and the option price is V = SH, where H solves

$$H_t + \frac{1}{2}\sigma^2 R^2 H_{RR} + (1 - rR)H_R = 0, \quad R > 0, \ t \in (0, T).$$

The final condition is $H(R, T) = (1 - R/T)^+$. We can specify boundary conditions at R = 0 and for $R \to \infty$. When $R = I/S \to \infty$ and I is fixed, it follows that $S \to 0$. Then the call will not be exercised and has no value at S = 0, i.e.

 $H(R,t) \to 0$ as $R \to \infty$.

Assume that *H* is twice differentiable at $R \searrow 0$ and for $t \in (0, T)$. Passing to the limit $R \searrow 0$ in the differential equation for *H*, we obtain the boundary condition on $R \searrow 0$:

$$H_t + H_R = 0$$
 on $R \searrow 0, t \in (0, T)$.

The option price is now given as a solution to this parabolic problem, V(S, I, t) = SH(I/S, t). We solve this problem numerically in Section 5.8.

3.8 Beyond Black-Scholes

The Black-Scholes equation in Section 3.1 was derived under a number of assumptions some of which are not realistic in real-world markets. Nevertheless, the Black-Scholes

equations are widely used in practice because the Black-Scholes formulas are easy to calculate and it is a good basis for more refined models.

It is possible to relax most of the conditions, and in this section, we will discuss some generalizations. To simplify the presentation and to highlight the ideas, we only sketch the extensions and use sometimes quite heuristic arguments. Let us review the assumptions imposed in Section 3.1:

- ► There are no dividend payments: Dividend yields can be included in the Black-Scholes model; see Section 3.4.
- The interest rate *r* and volatility σ are constant: We have already discussed models with stochastic interest rates and stochastic volatilities in Section 3.6.
- ► The stock price is governed by the Wiener process: The analysis of real-market data has shown that there might be heavy-tailed returns which cannot be reproduced by the Brownian motion. Thus, other stochastic processes need to be used. An example is the fractional Brownian motion.
- There are no market crashes: This means that the underlying stochastic process is continuous. We know from real markets that crashes may happen. This may be modeled, for instance, by jump-diffusion equations.
- The market is frictionless: This includes that there are no transaction costs. Again, this is a strong simplification and may lead to wrong prices when delta-hedging is expensive. Models including transcation costs typically lead to nonlinear Black-Scholes equations.

In the following, we discuss the last three extensions in more detail.

• **Fractional Brownian motion**. The fractional Brownian motion is a generalization of the standard Brownian motion. It can be defined by the properties that it is a continuous-time stochastic process $(B_t^H)_{t\geq 0}$ with $B_0^H = 0$, the expectation $EB_t^H = 0$ and the covariance

$$\mathcal{E}(B_s^H B_t^H) = \frac{1}{2} (s^{2H} + t^{2H} - |s - t|^{2H}) \text{ for all } s, t \ge 0.$$

The number $H \in (0,1)$ is called the Hurst parameter. When H = 1/2, we recover the standard Brownian motion $(W_t)_{t\geq 0}$ with covariance $E(W_sW_t) = \min\{s, t\}$, i.e. $B_t^{1/2} = W_t$. Figure 3.6 illustrates some paths of the fractional Brownian motion for various values of H. We see that the trajectories are becoming smoother for increasing values of H.

There are some important differences between the fractional and the standard Brownian motion:

- ▶ The increments of the fractional Brownian motion with $H \neq 1/2$ are not independent; they are independent for the Wiener process (H = 1/2) only.
- Almost all trajectories are locally Hölder continuous of order $\alpha < H$.



Figure 3.6: Trajectories of the fractional Brownian motion for H = 0.15 (top), H = 0.5 (middle), and H = 0.85 (bottom).

The fractional Brownian motion has a long-range dependence (or long memory) in the sense of

$$\sum_{n=1}^{\infty} \mathbb{E} \left[B_1^H (B_{n+1}^H - B_n^H) \right] = \infty, \quad H > 1/2.$$

Long-range dependence means that the coupling between values at different times does not decrease rapidly (i.e. exponential) for increasing time differences. The sum is zero in the case of the Wiener process H = 1/2, which means that the Wiener process has a short-range dependence.

The fractional Brownian motion can be computed as a stochastic integral involving the Wiener process (obtained by Mandelbrot and Van Ness):

$$B_t^H = B_0^H + \frac{1}{\Gamma(H+1/2)} \left(\int_{-\infty}^0 \left((t-s)^{H-1/2} - (-s)^{H-1/2} \right) dW_s + \int_0^t (t-s)^{H-1/2} dW_s \right), \quad t > 0,$$

where Γ is the Gamma function and the definition for t < 0 is similar.

The price of the underlying is assumed to satisfy the stochastic differential equation

$$dS_t = \mu S_t dS_t + \sigma S_t dB_t^H, \quad 0 < t < T,$$

which can be interpreted – as for Itô processes – as the integral formulation

$$S_t = S_0 + \int_0^t \mu S_\tau d\tau + \int_0^t \sigma S_\tau dB_\tau^H.$$

The definition of the stochastic integral depends on the value of *H*. Indeed, when 1/2 < H < 1, we can define the integral using pathwise Riemann-Stieltjes integrals since the fractional Brownian motion has Hölder continuous trajectories, and the corresponding integral is called a *Young integral*; see, e.g., the textbook [4]. When H < 1/2, there exist

several integral notions, and the integrals may be different. In the case 1/4 < H < 1/2, the definition is based on the theory of rough paths analysis introduced by Lyons [24]. Statistical analysis indicates that the Hurst parameter in financial markets is between 1/2 and 1, such that it is sufficient to consider this parameter range only.

In the Black-Scholes model (H = 1/2), there exists an equivalent probability measure Q under which $\mu = r$ and the discounted price process $e^{-rt}S_t$ is a martingale. When $H \neq 1/2$, there still exists an equivalent probability measure Q under which $\mu = r$, but $e^{-rt}S_t$ is not a martingale under Q. However, using the theory of forward/Wock integrals, it is possible to define the self-financing property and to derive a formula for the call option price [13, 18]:

$$C_t = S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2), \text{ where} d_{1/2} = \frac{\ln(S/K) + r(T-t) \pm \sigma^2 (T^{2H} - t^{2H})/2}{\sigma \sqrt{T^{2H} - t^{2H}}}.$$

Observe that we recover the classical Black-Scholes formula for H = 1/2.

Remark. There are several methods to simulate fractional Brownian motion. Here we report a technique using the Cholesky decomposition:

- Compute the variance-covariance matrix $\bar{S} = (\bar{S}_{ij})$ with $\bar{S}_{ij} = \frac{1}{2}(t_i^{2H} + t_j^{2H} |t_i t_j|^{2H})$ for i, j = 1, ..., n.
- Compute the square root matrix Σ , the solution to $\Sigma^2 = \overline{S}$, using Cholesky decomposition.
- ▶ Construct a vector *v* of *n* numbers from independent standard normal distributions.
- The vector $u = \Sigma v$ yields a sample of fractional Brownian motions.

• Jump-diffusion models. The Wiener process is continuous in time, and jumps in prices cannot be modeled. Jumps can be described by so-called jump-diffusion models, containing the Brownian motion (the diffusion part) and the Poisson process (the jump part).

A Poisson process $(N_t)_{t\geq 0}$ represents the number of events that have happened up to and including time t. The waiting time between events is exponentially distributed (i.e. a random variable with density function $\phi_{\lambda}(x) = \lambda e^{-\lambda t}$ for some $\lambda > 0$), while the total number of events up to time t is a Poisson process. The Poisson process has the Poisson distribution with parameter λt (and $\lambda > 0$), i.e., it is integer-valued and

$$\mathbf{P}(N_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbb{N}.$$

The Poisson process satisfies $N_0 = 0$ and it has independent increments $N_t - N_s$ for t > s (i.e. for any $s, t \ge 0$ with $s \le t$, the increment $N_t - N_s$ has the same distribution as N_{t-s}). It holds that $E(N_t) = \lambda t$ and $Var(N_t) = \lambda t$.

For financial applications, a process with a single jump size is not sufficient; the jump size may change from occurrence to occurrence. For this reason, we introduce *compound*

Poisson processes. Here, the waiting times between jumps are exponentially distributed but the jump sizes can have any distribution. More precisely, we call the process

$$X_t = \sum_{i=1}^{N_t} Y_i$$

a compound Poisson process if $(N_t)_{t\geq 0}$ is a Poisson process and $(Y_i)_{i\in\mathbb{N}}$ is a sequence of independent random variables with some law ϕ . The trajectories are piecewise constant with random jump sizes distributed according to the law ϕ .

A compound Poisson process can be simulated from the following algorithm:

- Simulate N_T from the Poisson distribution with parameter λT . This can be done, for instance, using the MATLAB command poissrnd(lambda*T).
- Simulate N_T uniformly distributed random variables U_1, \ldots, U_{N_T} on [0, T].
- Simulate N_T independent variables Y_1, \ldots, Y_{N_T} with law ϕ .
- Then the compound Poisson process is given by

$$X_t = \sum_{i=1}^{N_T} Y_i \chi_{\{U_i \le t\}},$$

where χ_A is the characteristic function on the set *A*.

We set $J_i = \exp Y_i$. We interpret $J_i - 1$ as the relative price jump size (in percent). For instance, if there are no jumps, i.e. $Y_i = 0$, then $J_i - 1 = 0$. The value of the underlying changes to (J - 1)S. The Merton jump-diffusion model [26] is then defined by

$$dS = \mu Sdt + \sigma SdW + (J-1)SdN,$$

where *S* is evaluated at t^- just before a jump happens at time *t*. Whenever a jump occurs $(dN \neq 0)$, the relative price of the underlying dS/S changes by the amount (J-1)dN. We can solve this stochastic differential equation by using the following variant of the Itô lemma. Let (X_t) solve

$$dX_t = a_t dt + b_t dW_t + \triangle X_t dN_t,$$

where $\triangle X_t$ is another process, and let $f : (0, \infty) \rightarrow \mathbb{R}$ be a twice differentiable function. Then

$$df(X) = \left(a_t f'(X) + \frac{1}{2}b_t^2 f''(X)\right) dt + b_t f'(X) dW_t + \left(f(X + \triangle X) - f(X)\right).$$

We apply this formula to $f = \ln, X = S$, and $\triangle X = (J - 1)S$, which gives

$$d\ln S = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW + \left(\ln(S + (J-1)S) - \ln S\right)$$

$$= \left(\mu - \frac{1}{2}\sigma^{2}\right)dt + \sigma dW + \ln J = \left(\mu - \frac{1}{2}\sigma^{2}\right)dt + \sigma dW + Y$$

Integrating this expression in (0, t) leads to

$$\ln S_t - \ln S_0 = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t + \sum_{i=1}^{N_t} Y_{i,i}$$

since we have N_t jumps of height Y_i , and the integration of the last term becomes a sum. We infer that

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t + \sum_{i=1}^{N_t} Y_i\right)$$
$$= S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right) \prod_{i=1}^{N_t} \exp(Y_i)$$

Thus, between the jumps, the process evolves like a geometric Brownian motion, while after each jump, the value of S_t is multiplied by $J_i = \exp(Y_i)$. The additional contribution $\sum_{i=1}^{N_t} Y_i$ is a compound Poisson process.

Next, we wish to derive the price of an option V(S, t). The jump-diffusion version of Itô's lemma becomes here

$$dV(S,t) = \left(V_t + \mu SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS}\right)dt + \sigma SV_S dW + \left(V(JS,t) - V(S,t)\right)dN,$$

as V(S + (J - 1)S, t) - V(S, t) = V(JS, t) - V(S, t). Assuming that the portfolio $\pi = V - \Delta S$ consists of one option and $-\Delta = -V_S$ shares of the underlying and is self-financing, the change of the portfolio becomes

$$d\pi = dV - \Delta dS = \left(V_t + \mu SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS}\right)dt + \sigma SV_S dW + \left(V(JS, t) - V(S, t)\right)dN_t - \Delta(\mu Sdt + \sigma SdW + (J-1)SdN).$$

Choosing $\Delta = V_S$, some terms cancel and we end up with

$$d\pi = \left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS}\right)dt + \left(V(JS, t) - V(S, t) - (J - 1)SV_S\right)dN_t.$$
 (3.31)

Merton argued that the jump component of the asset price should not be priced in the option. We used a similar argument in the derivation of the option price on dividend-paying assets in Section 3.4. Then the expectation of the change of the portfolio is the same as the profit from a riskless investment, $E(d\pi) = r\pi dt$. This argument is not completely convincing, but a common assumption when the risk cannot be fully hedged. We also have $E(dN) = \lambda dt$. Therefore, taking the expectation in (3.31) leads to

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + \lambda \left(V(JS, t) - V(S, t) - (J - 1)SV_S \right) = r\pi = r(V - V_S S).$$
(3.32)

The concept of compound processes can be generalized, leading to *Lévy processes*. We say that a stochastic process $(L_t)_{t\geq 0}$ is a *Lévy process* if $L_0 = 0$, L_t has independent and stationary increments, and L_t is stochastically continuous, i.e. for all $\varepsilon > 0$ and $t \ge 0$, it holds that

$$\lim_{h\to 0} \mathbb{P}(|L_{t+h} - L_t| > \varepsilon) = 0.$$

(We say that the increments are independent if for all $t_1 < \cdots < t_n$, the increments $L_{t_1}, L_{t_2} - L_{t_1}, \ldots, L_{t_n} - L_{t_{n-1}}$ are independent.) Other definitions replace the stochastic continuity by right continuity with left limits for the paths of L_t . Wiener and Poisson processes are special Lévy processes. In contrast, general Lévy processes may have infinitely many jumps in each finite interval.

The price dynamics can be given by an exponential Lévy model, $S_t = S_0 \exp(\mu t + L_t)$. It can be shown that in the risk-neutral world, the price of a European option V(S, T) is given by a partial-integro differential equation (called PIDE),

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV + \int_{\mathbb{R}} \left(V(Se^y, t) - V(S, t) - S(e^y - 1)V_S(S, t) \right) \nu(dy) = 0,$$
(3.33)

where ν is the Lévy measure (i.e. $\nu(0) = 0$ and $\int_{\mathbb{R}} \min\{1, |y|^2\}\nu(dy) < \infty$). This equation is nonlocal, since the integral term needs the solution for all values. We refer to [8, Chapter 12] for details on the derivation and numerical solution. We just remark that in the Merton jump-diffusion model (3.32), we have $\nu(\mathbb{R}) < \infty$. In this case, we can compute each of the three integrals in (3.33), which leads to a model like (3.32).

• **Transaction costs.** The Black-Scholes model is based on the assumption that hedging the portfolio does not involve any costs. In real markets, however, transaction costs (or bid-ask spreads) need to be paid. Here, we present the approach of Leland [22]. He assumes that rebalancing can happen only at discrete times with time step Δt . Hence, the price dynamics is given by the discrete equation

$$\triangle S = \mu S \triangle t + \sigma S \triangle W.$$

We suppose that the transaction cost is proportional to the trading volume, i.e., trading *c* shares of an asset *S* costs γcS , where $\gamma > 0$ is the transaction cost rate.

Similarly as in Section 3.1, we derive the price V(S, t) of an option by hedging the portfolio $\pi = cS - V(S, t)$. The change of the (self-financing) portfolio equals

$$\Delta \pi = c \Delta S - \Delta V(S, t) - \gamma S |\Delta c|,$$

where the additional term $-\gamma S|\triangle c|$ is the cost of trading $|\triangle c|$ assets. As usual, we assume that the portfolio change cannot be better than a change of a riskless bond with interest rate r, $\triangle \pi = r \pi \triangle t$. It remains to determine $\triangle V(S, t)$. For this, we use a discrete Itô lemma:

$$\triangle V(S,t) = \left(V_t + \mu S V_S + \frac{1}{2}\sigma^2 S^2 V_{SS}\right) \triangle t + \sigma S V_S \triangle W.$$

Combining these expressions, it follows that

$$\begin{aligned} r(cS - V) \triangle t &= r\pi \triangle t = \triangle \pi = c \triangle S - \triangle V(S, t) - \gamma S |\triangle c| \\ &= c(\mu S \triangle t + \sigma S \triangle W) - \left(V_t + \mu S V_S + \frac{1}{2}\sigma^2 S^2 V_{SS}\right) \triangle t \\ &- \sigma S V_S \triangle W - \gamma S |\triangle c|. \end{aligned}$$

Choosing $c = V_S$, the stochastic terms and the μ -terms cancel:

$$r(SV_S-V) \triangle t = -\left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS}\right) \triangle t - \gamma S |\triangle c|.$$

It remains to determine $|\triangle c|$. Since $\partial c/\partial S = V_{SS}$, we may expand as

$$\Delta c = \frac{\partial c}{\partial S} \Delta S + \frac{\partial c}{\partial t} \Delta t + O((\Delta S)^2) = V_{SS} \Delta S + V_{St} \Delta t + O((\Delta S)^2)$$

= $V_{SS} \Delta S + O(\Delta t) = V_{SS}(\mu S \Delta t + \sigma S \Delta W) + O(\Delta t)$
= $\sigma S V_{SS} \Delta W + O(\Delta t).$

Thus, $|\triangle c| = \sigma S |V_{SS}| |\triangle W| + O(\triangle t)$. It is possible to show that $E(|\triangle W|) = \sqrt{2 \Delta t / \pi}$, so we replace $|\triangle c|$ by $\sigma S |V_{SS}| \sqrt{2 \Delta t / \pi}$ and obtain

$$r(SV_S - V) \triangle t = -\left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS}\right) \triangle t - \gamma \sigma S^2 |V_{SS}| \sqrt{\frac{2}{\pi}} \sqrt{\triangle t}.$$

Note that we have neglected a term of order $O(\triangle t)$, which is questionable. Thus, the replacement of $|\triangle c|$ can be seen as a further assumption. Division by $\triangle t$ now leads to

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + \frac{\sqrt{2\gamma\sigma}}{\sqrt{\pi\Delta t}}S^2 |V_{SS}| + rSV_S - rV = 0.$$

Setting

$$\widetilde{\sigma}(V_{SS})^2 = \sigma^2 \bigg(1 + \frac{2\sqrt{2\gamma}}{\sigma\sqrt{\pi\Delta t}} \operatorname{sign}(V_{SS}) \bigg),$$

this equation can be formulated as

$$V_t + \frac{1}{2}\tilde{\sigma}(V_{SS})^2 S^2 V_{SS} + rSV_S - rV = 0.$$
(3.34)

Equation (3.34) is a *nonlinear* partial differential equation. The effective volatility $\tilde{\sigma}$ depends on the gamma $\Gamma = V_{SS}$, which can be seen as a measure of the degree of mishedging the portfolio. In the absence of transaction costs, the price of a European

call option is convex with respect to *S*, so $V_{SS} > 0$. We expect that this still holds true for sufficiently small transaction costs. Then we can drop the modulus sign and obtain

$$\widetilde{\sigma}^2 = \sigma^2 + \frac{\gamma\sigma}{\sqrt{2\pi\Delta t}}.$$

Note that the limit $\Delta t \rightarrow 0$ is not possible in this model. Moreover, as long as the option price stays convex, it gives the *linear* Black-Scholes equation with an increased volatility. For other options (for instance, barrier options), $\Gamma = V_{SS}$ may change its sign, and we need to solve the nonlinear equation.

We have seen that discrete hedging leads to an error in the replication (which we ignored by approximating $|\triangle c|$). Strictly speaking, this makes the no-arbitrage argument invalid. An alternative approach to the no-arbitrage framework is to introduce preferences of the investors to evaluate the option price. In arbitrage pricing, the risk tolerance of an investor is not relevant, since she/he is never exposed to any risk. Here, an investor follows a trading strategy that optimizes her/his preferences. Transaction costs will influence the trading strategy and change the expected cost of the replication. The preferences of an investor can be modeled by a utility function, and the investor aims to maximize the utility. Based on this approach, Barles and Soner [2] derived equation (3.34) with the following adjusted volatility:

$$\widetilde{\sigma}(V_{SS})^2 = \sigma^2 \left(1 + g(e^{r(T-t)}a^2 S^2 V_{SS}) \right),$$

where $a = \gamma \sqrt{\mu N}$, μ is the risk aversion factor, N is the number of options to be sold, and g is the unique solution to

$$\frac{dg}{dz} = \frac{g(z)+1}{2\sqrt{zg(z)}-z}, \quad z \neq 0, \quad g(0) = 0.$$
(3.35)

The solution to this ordinary differential equation is illustrated in Figure 3.7. It is a smooth variant of the sign function for negative arguments, but in contrast to the sign function, g is unbounded for positive arguments.



Figure 3.7: Solution to the differential equation (3.35).

4 Binomial models

4.1 The Cox-Ross-Rubinstein model

Often, financial markets are modeled by simple time-discrete models, so-called binomial or tree models or Cox-Ross-Rubinstein models. In this section, we review these models and show their relation to the Black-Scholes formula. The financial market is assumed to consist of the following products:

- ▶ Bond B_t : At time t = 0, the bond has the value $B_0 = 1$. Assuming that the riskfree interest rate is $r \ge 0$ and that the return is continuous, its value at $t = \triangle t$ is $B_{\triangle t} = e^{r \triangle t}$.
- ▶ Stock S_t : The value at time t = 0 equals $S_0 = S$. At $t = \triangle t$, the market has exactly two states: up or down, and we assume that $S_u = uS$ (up) with probability p > 0 and $S_d = dS$ (down) with probability 1 p > 0, where u > d > 0.
- Call option C_t : The call option has the strike K > 0 and expiration date t > 0.

We suppose as in Section 3.1 that the market is arbitrage-free, liquid, and frictionless.

Consider first one period of time. At time $t = \Delta t$, the value of the call option is $C_u = (uS - K)^+$ (up) with probability p or $C_d = (dS - K)^+$ (down) with probability 1 - p. We wish to determine its value at time t = 0.

Lemma 4.1. The value of the call option at time t = 0 in the one-period model equals

$$C_0 = e^{-r \triangle t} \left(p^* C_u + (1 - p^*) C_d \right), \quad \text{where } p^* = \frac{e^{r \triangle t} - d}{u - d}. \tag{4.1}$$

Proof. The proof is based on a replication strategy, similar as in the derivation of the Black-Scholes equations. We introduce the portfolio

$$\pi_t = c_1 B_t + c_2 S_t - C_t, \quad t = 0, \Delta t.$$

We assume that this portfolio has no value at time t = 0 and $t = \Delta t$:

$$0 = c_1 B_0 + c_2 S - C_0,$$

$$0 = c_1 e^{r \triangle t} + c_2 u S - C_u \quad \text{(up),}$$

$$0 = c_1 e^{r \triangle t} + c_2 dS - C_d \quad \text{(down).}$$

This is a linear system for the unknowns c_1 and c_2 with the solution

$$c_1 = \frac{uC_d - dC_u}{(u-d)e^{r\triangle t}}, \quad c_2 = \frac{C_u - C_d}{(u-d)S}$$

Then the option price at time t = 0 is

$$C_0 = c_1 B_0 + c_2 S = \frac{u C_d - d C_u}{(u - d)e^{r \Delta t}} + \frac{C_u - C_d}{u - d},$$

and a reformulation gives the conclusion.

- **Remark.** (1) We claim that our assumptions on the financial market imply that $d \le e^{r \triangle t} \le u$ and thus $0 \le p^* \le 1$. Indeed, a stock must earn a return at least as large as the interest rate *r* in the up state and at most as low as *r* in the down state, $dS \le e^{r \triangle t}S \le uS$.
 - (2) The option premium C_0 can be interpreted like in the Black-Scholes model as the discounted expectation value with respect to p^* . Indeed, let X be a discrete stochastic process such that $X = X_u$ with probability p^* and $X = X_d$ with probability $1 p^*$. Then $E_{p^*}(X) = p^*X_u + (1 p^*)X_d$, and (4.1) can be written as

$$C_0 = e^{-r \triangle t} \mathbf{E}_{p^*} \big((S_{\triangle t} - K)^+ \big).$$

This expression corresponds to the option price from Theorem 3.4. The risk-neutral measure is represented here just by the so-called risk-neutral probability p^* . Because of

$$\mathbf{E}_{p^*}(S_{\triangle t}) = p^* u S + (1-p^*) dS = \frac{e^{r \triangle t} - d}{u-d} u S + \frac{u - e^{r \triangle t}}{u-d} dS = e^{r \triangle t} S_{A}$$

the expected value of the asset equals the return of the bond. This is the viewpoint of a risk-neutral investor who is indifferent between a risky asset and a riskless investment.

Next, we consider an *n*-period financial market with times $t = 0, \Delta t, ..., n \Delta t$, where $n \Delta t = T$. In each period $[t, t + \Delta t]$, the asset price changes by the factor u (up) with probability p or d (down) with probability 1 - p; see Figure 4.1. Thus, after k up states and n - k down states, the asset value becomes $S_k^n := u^k d^{n-k}S$.

Again, we wish to determine the call price at time t = 0. When n = 2, we have, according to Lemma 4.1,

$$C_0 = e^{-r \triangle t} (p^* C_u + (1 - p^*) C_d).$$

where C_u and C_d are given by

$$C_u = e^{-r \triangle t} \left(p^* (S_2^2 - K)^+ + (1 - p^*) (S_1^2 - K)^+ \right),$$

 \square



Figure 4.1: Binomial tree with *n* periods.

$$C_d = e^{-r \Delta t} \left(p^* (S_1^2 - K)^+ + (1 - p^*) (S_0^2 - K)^+ \right).$$

This gives

$$C_0 = e^{-2r \triangle t} \Big((p^*)^2 (S_2^2 - K)^+ + 2p^* (1 - p^*) (S_1^2 - K)^+ + (1 - p^*)^2 (S_0^2 - K)^+ \Big).$$

We can guess the option price for the *n*-period model:

$$C_0 = e^{-nr\Delta t} \sum_{k=0}^n \binom{n}{k} (p^*)^k (1-p^*)^{n-k} (S_k^n - K)^+.$$
(4.2)

This formula can be written in different ways as shown in the following proposition.

Proposition 4.2 (Call option price). The call price of the *n*-period financial market at time t = 0 equals

$$C_0 = e^{-rT} E_{p^*} ((S^n_{\cdot} - K)^+), \qquad (4.3)$$

where

$$\mathbf{E}_{p^*}(X) = \sum_{k=0}^n \binom{n}{k} (p^*)^k (1-p^*)^{n-k} X_k$$

is the expected value of a discrete stochastic process $(X_k)_{0 \le k \le n}$. Furthermore, it holds that

$$C_0 = S\Phi(m, p') - Ke^{-rT}\Phi(m, p^*),$$
(4.4)

where $p' = p^* u e^{-r \triangle t}$, $p^* = (e^{r \triangle t} - d)/(u - d)$, and

$$\Phi(m,p) = \sum_{k=m}^{n} \binom{n}{k} p^{k} (1-p)^{n-k}$$

with $m = \min\{0 \le k \le n : S_k^n - K \ge 0\}$ is the binomial cumulative distribution function.

Equation (4.3) is the discounted expectation value of the payoff with respect to the risk-neutral probability p^* . Formula (4.4) is a discrete version of the Black-Scholes formula (3.8) for t = 0. We observe that $\Phi(m, p)$ is the probability that the binomial process X_p with parameter p has values larger or equal m:

$$\Phi(m,p) = \mathrm{P}(X_p \ge m).$$

Proof. We only need to show (4.4) since (4.3) follows immediately from (4.2). Using the definition $S_k^n = u^k d^{n-k} S$, (4.2) becomes

$$C_{0} = e^{-rn\Delta t} \sum_{k=m}^{n} \binom{n}{k} (p^{*})^{k} (1-p^{*})^{n-k} u^{k} d^{n-k} S - e^{-rn\Delta t} \sum_{k=m}^{n} \binom{n}{k} (p^{*})^{k} (1-p^{*})^{n-k} K$$

= $S \sum_{k=m}^{n} \binom{n}{k} (p^{*} u e^{-r\Delta t})^{k} ((1-p^{*}) d e^{-r\Delta t})^{n-k} - K e^{-rT} \sum_{k=m}^{n} \binom{n}{k} (p^{*})^{k} (1-p^{*})^{n-k}.$

The definition of p^* implies that $p^*u + (1 - p^*)d = e^{r\Delta t}$ or $e^{-r\Delta t}(1 - p^*)d = 1 - p^*ue^{-r\Delta t}$ and thus $1 - p' = 1 - p^*ue^{-r\Delta t} = (1 - p^*)de^{-r\Delta t}$. We deduce that

$$C_0 = S \sum_{k=m}^n \binom{n}{k} (p')^k (1-p')^{n-k} - K e^{-rT} \sum_{k=m}^n \binom{n}{k} (p^*)^k (1-p^*)^{n-k},$$

and taking into account the definition of Φ , we conclude the proof.

4.2 Relation with the Black-Scholes formula

By the central limit theorem, we expect that the binomial distribution on $\{0, ..., n\}$ converges in some sense to the standard normal distribution as $n \to \infty$. Hence, we may conjecture that the option price of the discrete model "converges" to the price computed from the Black-Scholes formula. Before we make this statement precise, we recall the central limit theorem.

Theorem 4.3 (Central limit theorem). Let Y_n be a binomial random variable with parameter p on $\{0, ..., n\}$. Then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right) = \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz.$$

Recall that a binomial random variable *Y* with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$ is defined via $P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.

We claim that the option price (4.4) converges to the Black-Scholes price (3.8) for a certain choice of *u* and *d*.

Theorem 4.4. Let $u = \exp(\sigma\sqrt{\Delta t})$ and $d = \exp(-\sigma\sqrt{\Delta t})$ and let C_0 be the option price (4.4). Then

$$\lim_{\Delta t\to 0} C_0 = S\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where the limit $\triangle t \rightarrow 0$ means that also $n \rightarrow \infty$ such that $n \triangle t = T$ is constant, and $d_{1/2}$ are defined in (3.10).

Proof. Since

$$C_0 = SP(X_{p'} \ge m) - Ke^{-rT}P(X_{p^*} \ge m),$$

it is sufficient to show that $P(X_{p'} \ge m) \to \Phi(d_1)$ and $P(X_{p^*} \ge m) \to \Phi(d_2)$ as $\Delta t \to 0$. We prove only the latter relation as the proof of the former one is similar.

First, we make a Taylor expansion for p^* with respect to $\triangle t$:

$$p^* = \frac{e^{r \bigtriangleup t} - e^{-\sigma \sqrt{\bigtriangleup t}}}{e^{\sigma \sqrt{\bigtriangleup t}} - e^{-\sigma \sqrt{\bigtriangleup t}}}$$

=
$$\frac{(1 + r \bigtriangleup t) - (1 - \sigma \sqrt{\bigtriangleup t} + \sigma^2 \bigtriangleup t/2) + O(|\bigtriangleup t|^{3/2})}{(1 + \sigma \sqrt{\bigtriangleup t} + \sigma^2 \bigtriangleup t/2) - (1 - \sigma \sqrt{\bigtriangleup t} + \sigma^2 \bigtriangleup t/2) + O(|\bigtriangleup t|^{3/2})}$$

=
$$\frac{\sigma + (r - \sigma^2/2)\sqrt{\bigtriangleup t} + O(\bigtriangleup t)}{2\sigma + O(\bigtriangleup t)} \quad \text{as } \bigtriangleup t \to 0.$$

This shows that

$$\lim_{\Delta t \to 0} p^* = \frac{1}{2}, \quad \lim_{\Delta t \to 0} \frac{2p^* - 1}{\sqrt{\Delta t}} = \frac{r - \sigma^2/2}{\sigma}.$$

We infer that

$$\lim_{\Delta t \to 0} np^{*}(1-p^{*}) \left(\ln \frac{u}{d} \right)^{2} = \lim_{\Delta t \to 0} \frac{T}{\Delta t} p^{*}(1-p^{*}) (2\sigma\sqrt{\Delta t})^{2}$$
$$= \lim_{\Delta t \to 0} 4p^{*}(1-p^{*})\sigma^{2}T = \sigma^{2}T,$$
(4.5)

$$\lim_{\Delta t \to 0} n \left(p^* \ln \frac{u}{d} + \ln d \right) = \lim_{\Delta t \to 0} \frac{T}{\Delta t} \left(p^* \cdot 2\sigma \sqrt{\Delta t} - \sigma \sqrt{\Delta t} \right)$$
$$= \lim_{\Delta t \to 0} \frac{T}{\sqrt{\Delta t}} (2p^* - 1)\sigma = \left(r - \frac{\sigma^2}{2} \right) T.$$
(4.6)

We need to normalize X_p in order to apply the central limit theorem:

$$P(X_{p^*} \ge m) = 1 - P(X_{p^*} < m) = 1 - P\left(\frac{X_{p^*} - np^*}{\sqrt{np^*(1 - p^*)}} < \frac{m - np^*}{\sqrt{np^*(1 - p^*)}}\right).$$
(4.7)

By definition of *m* (see the proof of Proposition 4.2), $S_m^n - K = Su^m d^{n-m} - K \ge 0$ which is equivalent to $\ln(u^m d^{n-m}S) \ge \ln K$ or $m \ln u + (n-m) \ln d \ge -\ln(S/K)$. We solve

this inequality for *m*:

$$m \ge -\frac{\ln(S/K) + n\ln d}{\ln(u/d)}.$$

Again by definition of *m*, there exists a number $\gamma \in [0, 1)$ such that

$$m = -\frac{\ln(S/K) + n\ln d}{\ln(u/d)} + \gamma.$$

Hence, taking into account the limits (4.5), (4.6), and $\ln(u/d) = 2\sigma\sqrt{\Delta t} \to 0$ as $\Delta t \to 0$, we obtain

$$\frac{m - np^*}{\sqrt{np^*(1 - p^*)}} = \frac{-\ln(S/K) - n\ln d - np^*\ln(u/d) + \gamma\ln(u/d)}{\ln(u/d)\sqrt{np^*(1 - p^*)}}$$
$$\rightarrow \frac{-\ln(S/K) - (r - \sigma^2/2)T}{\sigma\sqrt{T}} \quad \text{as } \Delta t \rightarrow 0.$$

We perform the limit $\triangle t \rightarrow 0$ or, equivalently, $n \rightarrow \infty$ in (4.7) to infer from the central limit theorem that

$$\lim_{\Delta t \to 0} \mathbb{P}(X_{p^*} \ge m) = 1 - \Phi\left(\frac{-\ln(S/K) - (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right).$$

Finally, because of $1 - \Phi(-x) = \Phi(x)$, we deduce that

$$\lim_{\Delta t \to 0} \mathbb{P}(X_{p^*} \ge m) = \Phi\left(\frac{\ln(S/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right) = \Phi(d_2),$$

concluding the proof.

4.3 Binomial method

The Cox-Ross-Rubinstein model allows us to design a numerical algorithm to determine the option price. The advantage of the algorithm, compared to the explicit formula (4.4) is that we may consider exotic options for which no explicit formula is available.

As in the previous subsections, we divide the time interval [0, T] into n equal subintervals of length $\Delta t = T/n$. Our assumptions on the financial market are similar to those in Section 4.1:

- ▶ The asset price S_{i+1} at time $t_{i+1} = (i+1) \triangle t$ can take only two values: either $S_{i+1} = uS_i$ (up state) with probability p > 0 or $S_{i+1} = dS_i$ (down state) with probability 1 p > 0.
- ► The expected return after time $\triangle t$ is the same as for a riskless bond. This means that we set $\mu = r$ in (2.6) of the geometric Brownian motion:

$$E(S_{i+1}) = S_i e^{r \triangle t}, \quad Var(S_{i+1}) = S_i^2 e^{2r \triangle t} (e^{\sigma^2 \triangle t} - 1).$$
(4.8)

► The expected value of the time-continuous option price V_t is the same as for the riskless bond, $E(V_{i+1}) = V_i e^{r \triangle t}$.

Furthermore, we suppose that the market is arbitrage-free, liquid, and frictionless. The model includes three parameters: p, u, and d, which need to be determined. The idea is to equate the expectation and variance of S_i from (4.8) with

$$E(S_{i+1}) = p \cdot uS_i + (1-p) \cdot dS_i,$$

$$Var(S_i) = E(S_{i+1}^2) - E(S_{i+1})^2 = p(uS_i)^2 + (1-p)(dS_i)^2 - (puS_i + (1-p)dS_i)^2.$$

This yields

$$S_{i}e^{r\Delta t} = E(S_{i+1}) = puS_{i} + (1-p)dS_{i},$$

$$S_{i}^{2}e^{2r\Delta t}(e^{\sigma^{2}\Delta t} - 1) = \operatorname{Var}(S_{i+1}) = p(uS_{i})^{2} + (1-p)(dS_{i})^{2} - (puS_{i} + (1-p)dS_{i})^{2}.$$
(4.9)

Inserting the first equation into the second one and dividing by S_i^2 gives

$$e^{2r \triangle t}(e^{\sigma^2 \triangle t} - 1) = pu^2 + (1 - p)d^2 - e^{2r \triangle t}.$$

We add $e^{2r \triangle t}$ to find that

$$e^{(2r+\sigma^2)\triangle t} = pu^2 + (1-p)d^2.$$
(4.10)

Equations (4.9) (after division by S_i) and (4.10) are two equations for the three parameters p, u, and d. We need a third equation to determine these parameters uniquely. Since we obtain the same asset price when it increases once and decreases once, $S_i = d(uS_i) = u(dS_i)$, it is reasonable to assume that

$$ud = 1.$$
 (4.11)

The nonlinear system (4.9)-(4.11) can be solved explicitly:

$$u = s + \sqrt{s^2 - 1}, \quad d = s - \sqrt{s^2 - 1}, \quad p = \frac{e^{r \triangle t} - d}{u - d},$$

where

$$s := \frac{1}{2} (e^{-r \bigtriangleup t} + e^{(r+\sigma^2) \bigtriangleup t}).$$

Remark. We discuss the binomial method.

- (1) There are many possible choices for the third equation. For instance, we may set p = 1/2 or equate the third moments $E(S_i^3)$.
- (2) Observe that the value for p corresponds to the riskneutral value p^* ; see (4.1).

(3) We have argued in Section 4.1 that $d \le e^{r \triangle t} \le u$. These inequalities are satisfied since

$$s - \sqrt{s^2 - 1} \le e^{r \bigtriangleup t} \le s + \sqrt{s^2 - 1}$$

is equivalent to $|e^{r riangle t} - s| \le \sqrt{s^2 - 1}$ or

$$\frac{1}{2}(e^{-r\bigtriangleup t}+e^{r\bigtriangleup t})\le s=\frac{1}{2}(e^{-r\bigtriangleup t}+e^{(r+\sigma^2)\bigtriangleup t}),$$

and this inequality is satisfied (since $\sigma \ge 0$).

(4) Often, an approximated binomial method is used in the literature:

$$u = e^{\sigma\sqrt{\Delta t}}, \quad d = e^{-\sigma\sqrt{\Delta t}}, \quad p = \frac{e^{r\Delta t} - d}{u - d}$$

This choice follows from an expansion of $u = s + \sqrt{s^2 - 1} = e^{\sigma\sqrt{\Delta t}} + O(|\Delta t|^{3/2})$ as $\Delta t \rightarrow 0$. However, the inequalities $d \leq e^{r\Delta t} \leq u$ are only satisfied if $\Delta t \leq (\sigma/r)^2$, which means that the time step cannot be chosen arbitrarily large.

We are now able to formulate the algorithm: Let S_0 the asset price at time t = 0 and set $S_{ji} = u^j d^{i-j}S_0$, which is the asset price after j up states and i - j down states (this requires that $i \ge j$). The set $\{S_{ji} : i = 0, ..., n, j = 0, ..., i\}$ contains all possible asset prices at time $t_i = i \triangle t$.

- ▶ Initialization of the binomial tree: $S_{jn} = u^j d^{n-j} S_0$ for j = 0, ..., n.
- ► Computation of the final option price:

$$V_{jn} = \begin{cases} (S_{jn} - K)^+ & : \text{ call,} \\ (K - S_{jn})^+ & : \text{ put,} \end{cases} \quad j = 0, \dots, n.$$

For the third step, we need a recursion formula for the option values V_{ji} . The first equation (4.9) of the nonlinear system can be written as

$$S_{ji}e^{r\Delta t} = puS_{ji} + (1-p)dS_{ji} = pS_{j+1,i+1} + (1-p)S_{j,i+1}.$$

Recall that we have assumed that the expected value of V_{i+1} equals $V_i e^{r \Delta t}$. Thus, supposing that the evolution of the option price follows the up and down states as for the asset price, we may propose

$$V_{ji}e^{r \bigtriangleup t} = pV_{j+1,i+1} + (1-p)V_{j,i+1}$$

Backward iteration: Compute in case of European options

$$V_{ji} = e^{-r \triangle t} \left(p V_{j+1,i+1} + (1-p) V_{j,i+1} \right),$$

and in case of American options

$$Z_{ji} = e^{-r \triangle t} (pV_{j+1,i+1} + (1-p)V_{j,i+1}),$$

$$V_{ji} = \begin{cases} \max\{(S_{ji} - K)^+, Z_{ji}\} & : \text{ call,} \\ \max\{(K - S_{ji})^+, Z_{ji}\} & : \text{ put.} \end{cases}$$

Then V_{00} is an approximation of $V(S_0, 0)$.

Observe that, technically, we have evaluated a Bermudean option, which is an option that can only be exercised at certain specified dates. For an increasing number of time steps, we approach the American option price.

Figure 4.2 shows the values of a European put option computed by the binomial method using n = 5 and n = 50 time steps. For comparison, the prices computed from the Black-Scholes formula are also plotted. We see that in case n = 5, the prices differ from the Black-Scholes values at the money, while the prices with n = 50 cannot be distinguished from the Black-Scholes values in the figure.



Remark. In case of dividend payments, we need to modify the previous algorithm. When the divident payment occurs continuously with rate $\delta > 0$, we just replace *r* in the expression for *p* by $r - \delta$:

$$u = s + \sqrt{s^2 - 1}, \quad d = s - \sqrt{s^2 - 1}, \quad p = \frac{e^{(r-\delta) \triangle t} - d}{u - d},$$

A more realistic case is modeled when the dividend payment is discrete, for instance at time t_D . Then we may use the same values for p, u, and d as for the no-dividend case, but at time t_D , the asset value jumps by the dividend payment D, and we replace S_{jD} by $S_{jD} - D$.

5 Monte-Carlo method

The fair price of exotic options generally can only be determined by means of numerical methods. Often, we need to solve stochastic differential equations or to integrate high-dimensional integrals. In this section, we detail the Monte-Carlo method for the solution of such situations.

The key steps of a Monte-Carlo simulation (for a call option) are as follows:

► Simulation of the asset values: Compute *m* paths $S_t^{(k)}$, k = 1, ..., m, of the stochastic process $(S_t)_{t>0}$ solving the stochastic differential equation

$$dS_t = rS_t dt + \sigma S_t dW_t, \quad 0 < t < T.$$
(5.1)

- Calculation of the payoff function: Compute, for each path k = 1, ..., m, $V_T^{(k)} = (S_T^{(k)} K)^+$.
- \blacktriangleright Estimator for the expectation: The expected value of V_T may be approximated by

$$\widehat{\mathrm{E}}(V_T) = \frac{1}{m} \sum_{k=1}^m V_T^{(k)}, \text{ where } V_T = (V_T^{(1)}, \dots, V_T^{(m)}).$$

► Calculation of the option price: An approximation of the option price is given by $\hat{V}_0 = e^{-rT} \hat{E}(V_T)$.

The last three steps are easy to compute. The third step is based on the law of large numbers according to which the arithmetic average of uniformly distributed and independent random variables converges almost surely to the expected value. The challenge is the first step, which consists of two tasks:

- the simulation of *m* independent realizations of a Wiener process $W_t^{(k)}$ and
- the approximate solution of (5.1) associated to the corresponding path of the Wiener process.

A simple approximation of the stochastic differential equation (5.1) is

$$\triangle S_t = rS_t \triangle t + \sigma S_t \triangle W_t, \tag{5.2}$$

where $\triangle S_t = S_{t+\triangle t} - S_t$ and $\triangle W_t = W_{t+\triangle t} - W_t$. We expect that this approximation is only of low order, and one may ask for better discretizations (which are discussed below). We need to determine realizations of $\triangle W_t$, which are $N(0, \triangle t)$ -distributed. For instance, we may compute realizations of an N(0, 1)-distributed random variable *Z* and then define $\triangle W_t = Z\sqrt{\triangle t}$. In Figure 5.1, we illustrate 20 trajectories of S_t obtained from (5.2).

In the following sections, we explain how normally distributed random variables can be computed and how (5.1) can be approximated up to a certain order.



5.1 Pseudo random numbers

All realizations of random numbers with computer algorithms do not give real random numbers, since the algorithms are deterministic. Therefore, we call them *pseudo random numbers*. It is possible to produce true random numbers from physical processes with statistically random noise signals, like thermal noise, atmospheric noise, or quantum phenomena, but we are discussing only pseudo random numbers.

The idea is to produce uniformly distributed (pseudo) random numbers, and then to transform them into random numbers that are distributed according to the desired density function. We recall that the random variable X is uniformly distributed on [0, 1], written as $X \sim U[0, 1]$, if X possesses the density f(x) = 1 for $x \in [0, 1]$, i.e.

$$P(a \le X \le b) = \int_{a}^{b} f(x)dx = b - a \text{ for } 0 \le a \le b \le 1.$$

We say that a sequence of random numbers is *F*-distributed if the numbers are independent realizations of random variables with cumulative distribution function *F*, where $F(x) = \int_{-\infty}^{x} f(z) dz$. Clearly, pseudo random numbers are never independent since they are calculated from a deterministic algorithm, but we are satisfied with random numbers that satisfy the statistical properties approximately. In the following, we detail some selected random number generators.

• Uniformly distributed random numbers. The oldest random number generators are based on linear congruence: Let $X_0 \in \{0, ..., m-1\}$ with $m \in \mathbb{N}$ be given; compute

$$X_{i+1} = (aX_i + b) \mod m, \quad i \in \mathbb{N},$$

where $a, b \in \{0, ..., m-1\}$. The operation "*a* mod *m*" computes the remainder of the division a/m. in the mathematical software MATLAB, the numbers $a = 7^5 = 16807$, b = 0, and $m = 2^{31} - 1$ are chosen. The numbers X_i repeat after some iterations, but the period is maximal, $p = 2^{31} - 1 \approx 2 \cdot 10^9$. This algorithm is called mcg16807 (mcg stands for "multiplicative congruential generator").

An extremely large period is obtained in the so-called *Mersenne-Twister algorithm*, developed by Makoto Matsumoto and Takuji Nishimura in 1997 [25], namely $p = 2^{19937} - 1 \approx 4 \cdot 10^{6001}$ (this number is a Mersenne prime number). The algorithm is as follows: Let X_1, \ldots, X_n with n = 624 be given; compute for i > n,

$$\begin{split} h &= X_{i-n} - (X_{i-n} \bmod 2^{31}) + (X_{i-n+1} \bmod 2^{31}), \\ Z_i &= Z_{i-227} \oplus \lfloor h/2 \rfloor \oplus ((h \bmod 2) \cdot 9908B0DF_{hex}), \\ a &= Z_i \oplus \lfloor X_i/2^{11} \rfloor, \\ b &= a \oplus ((a \cdot 2^7) \land 9D2C5680_{hex}), \\ c &= b \oplus ((b \cdot 2^{15}) \land EFC60000_{hex}), \\ X_i &= c \oplus \lfloor c/2^{18} \rfloor, \end{split}$$

where \oplus is the bitwise XOR operation, \land the AND operation, $\lfloor x \rfloor = \max\{m \in \mathbb{N} : m \le x\}$ is the floor function (e.g. $\lfloor 2.7 \rfloor = 2$ and $\lfloor -2.4 \rfloor = -3$), and the index "hex" means that the corresponding number is a hexidecimal number. This algorithm is implemented in MATLAB as the function mt19937ar.

• Normally distributed random numbers. Normally distributed random numbers can be computed from uniformly distributed numbers. We discuss two methods:

- inversion of the distribution function,
- transformation of random numbers.

For the first technique, let $X \sim U[0, 1]$ be a uniformly distributed random variable and F be a continuous, strictly monotone function. We claim that the random variable $F^{-1}(X)$ is F-distributed. Indeed, the assumptions guarantee that the inverse F^{-1} exists. Since X is uniformly distributed, we have $P(X \le x) = x$ for $x \in [0, 1]$. Consequently,

$$P(F^{-1}(X) \le x) = P(X \le F(x)) = F(x),$$

showing the claim. This result is applicable to the standard normal cumulative distribution function $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-y^2/2} dy$, but Φ^{-1} is not explicit. We need to invert $\Phi(x) = y$ numerically, using for instance the Newton method. Unfortunately, this problem is ill-conditioned around y = 1, meaning that small changes in y yield large changes in x. Therefore, one may replace Φ^{-1} by a rational function R and set x = R(y). We refer to [16, Section 5.2.2] and [33, Appendix D2] for further details.

The second technique is based on the following result.

Lemma 5.1. Let X be a random variable with density function f on the set $\Omega = \{x \in \mathbb{R}^n : f(x) > 0\}$. The mapping $g : \Omega \to g(\Omega)$ is assumed to be invertible with continuously differentiable inverse $g^{-1} : g(\Omega) \to \Omega$ (this is called a diffeomorphism). Then the random variable Y = g(X) has the density function

$$y \mapsto f(g^{-1}(y)) \left| \det \frac{dg^{-1}}{dy}(y) \right|, \quad y \in g(\Omega).$$

Proof. We give only a motivation of the result; a full proof can be found in [12, Theorem 4.2]. By a change of variables, for any set $\omega \subset \mathbb{R}^n$,

$$\mathbf{P}(g(X) \in \omega) = \mathbf{P}(X \in g^{-1}(\omega)) = \int_{g^{-1}(\omega)} f(z)dz = \int_{\omega} f(g^{-1}(y)) \left| \det \frac{dg^{-1}}{dy}(y) \right| dy,$$

which shows the claim.

This transformation is not useful in one dimension but surprisingly in two dimensions. Indeed, we have f(x) = 1 (uniform distribution) and we are looking for a function g such that y = g(x) and the transformed density function is the one-dimensional standard normal distribution:

$$\left|\frac{dg^{-1}}{dy}(y)\right| = \frac{e^{-y^2/2}}{\sqrt{2\pi}}.$$

This differential equation for g^{-1} has no explicit solution. However, in two dimensions, we choose the transformation y = g(x) with

$$g(x) = \begin{pmatrix} \sqrt{-2\ln x_1} \cos(2\pi x_2) \\ \sqrt{-2\ln x_1} \sin(2\pi x_2) \end{pmatrix}, \quad x = (x_1, x_2)^\top \in [0, 1]^2.$$

Its inverse is explicit,

$$g^{-1}(y) = \begin{pmatrix} \exp(-|y|^2/2) \\ \arctan(y_2/y_1)/(2\pi) \end{pmatrix}, \quad y = (y_1, y_2)^\top,$$

and the determinant becomes, for $x = g^{-1}(y)$,

$$\left|\det\frac{dg^{-1}}{dy}\right| = \left|\det\begin{pmatrix}-y_1x_1 & -y_2x_1\\\frac{1}{2\pi}\frac{-y_2/y_1^2}{1+y_2^2/y_1^2} & \frac{1}{2\pi}\frac{1/y_1}{1+y_2^2/y_1^2}\end{pmatrix}\right| = \frac{x_1}{2\pi} = \frac{e^{-|y|^2/2}}{2\pi},$$

which is the two-dimensional standard normal density function of two independent one-dimensional random variables. We infer that if $X \sim U[0,1]$ then $g(X) \sim N(0,I)$, where *I* is the unit matrix. This gives the *algorithm of Box-Muller*:

• Generate $X_1, X_2 \sim U[0, 1]$.

• Compute
$$Y_1 = \sqrt{-2 \ln X_1} \cos(2\pi X_2)$$
 and $Y_2 = \sqrt{-2 \ln X_1} \sin(2\pi X_2)$

Then Y_1 and Y_2 are standard normally distributed.

The Box-Muller algorithm needs three function calls (sqrt, ln, and cos/sin) to generate two random numbers. This effort can be reduced by the polar rejection algorithm that only needs two function calls (sqrt, ln). However, it produces random numbers which need to be rejected if they are not in the domain of definition. More efficient is the *Ziggurat algorithm* which does not need sqrt or ln as function calls, but uses only multiplications. As in the polar rejection algorithm, the algorithm produces random numbers which possibly have to be rejected. The name originates from covering the distribution function with stacked rectangular segments. This algorithm is much faster than the Box-Muller or polar rejection method. As the Ziggurat algorithm is more complex to implement, it is best used when many random numbers have to be computed.

• **Correlated normally distributed random numbers.** When we simulate a multidimensional Brownian motion, we often need correlated random variables. They can be generated from standard normally distributed random variables.

Proposition 5.2. Let $X = (X_1, ..., X_n)^{\top}$ be a vector of (independent) standard normally distributed random variables X_i with density function $f(x) = (2\pi)^{-n/2}e^{-|x|^2/2}$. Furthermore, let $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix with the Cholesky decomposition $\Sigma = LL^{\top}$ with a lower triangular matrix L. Then $Y = \mu + LX$ is $N(\mu, \Sigma)$ -distributed.

Proof. It is sufficient to show that Z = LX is $N(0, \Sigma)$ -distributed. This follows with z = Lx and $dz = |\det L| dx$ from

$$f(x)dx = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{x^{\top}x}{2}\right)dx$$

= $\frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{(L^{-1}z)^{\top}(L^{-1}z)}{2}\right)dx$
= $\frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{z^{\top}(LL^{\top})^{-1}z}{2}\right)dx$
= $\frac{1}{(2\pi)^{n/2}|\det L|} \exp\left(-\frac{z^{\top}\Sigma^{-1}z}{2}\right)dz$
= $\frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{z^{\top}\Sigma^{-1}z}{2}\right)dz.$

Thus, $Y = \mu + Z$ is $N(\mu, \Sigma)$ -distributed.

This yields the following algorithm:

- Determine the Cholesky decomposition $\Sigma = LL^{\top}$.
- Compute independent $X_i \sim N(0, 1)$ for i = 1, ..., n and set $X = (X_1, ..., X_n)$.
- Compute $Y = \mu + LX \sim N(\mu, \Sigma)$.

Example 5.3. We wish to determine a two-dimensional $N(0, \Sigma)$ -distributed random variable (Y_1, Y_2), where

$$\Sigma = egin{pmatrix} \sigma_1^2 &
ho\sigma_1\sigma_2 \
ho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

We say that (Y_1, Y_2) is correlated with correlation coefficient $\rho \in [-1, 1]$. Recall that if $\rho = 0$, the random variables Y_1 and Y_2 are uncorrelated; $\rho < 0$ means that the random variables are negatively correlated, and if $\rho > 0$, they are positively correlated.

The ansatz

$$L = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$$

leads to

$$\begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix},$$

and by equating coefficients, $a^2 = \sigma_1^2$, $ab = \rho\sigma_1\sigma_2$, $b^2 + c^2 = \sigma_2^2$. Solving for *a*, *b*, and *c* gives

$$L = \begin{pmatrix} \sigma_1 & 0\\ \rho \sigma_2 & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix}.$$

Consequently, if Z_1 and Z_2 are independent and N(0, 1)-distributed then

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = L \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) \end{pmatrix}$$

is $N(0, \Sigma)$ -distributed.

5.2 Euler-Maruyama method

We wish to approximate stochastic differential equations of the form

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t, \quad 0 < t < T.$$
(5.3)

We are looking for approximations Y_i of X_{t_i} , where $t_i = ih$, h = T/n > 0 is the (uniform) time step, and $n \in \mathbb{N}$. Non-uniform time steps may also be considered, but we restrict ourselves to uniform ones to simplify the presentation. The stochastic process $(X_t)_{t\geq 0}$ is an Itô process, and we can write, according to Definition 2.11,

$$X_{t_{i+1}} = X_0 + \int_0^{t_{i+1}} a(X_s, s) ds + \int_0^{t_{i+1}} b(X_s, s) dW_s,$$

$$X_{t_i} = X_0 + \int_0^{t_i} a(X_s, s) ds + \int_0^{t_i} b(X_s, s) dW_s.$$
Subtracting both equations gives

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} a(X_s, s) ds + \int_{t_i}^{t_{i+1}} b(X_s, s) dW_s.$$

Depending on the approximation of the integrals, we obtain different methods. When we take *a* and *b* constant over $[t_i, t_{i+1})$, we can compute the stochastic integral, giving

$$\triangle W_i := \int_{t_i}^{t_{i+1}} dW_s = W_{t_{i+1}} - W_{t_i},$$

and this expression is N(0, h)-distributed. This leads to the Euler-Maruyama method:

Definition 5.4 (Euler-Maruyama scheme). $Y_{i+1} = Y_i + a(Y_i, t_i)h + b(Y_i, t_i) \triangle W, \quad i = 0, \dots, n-1, \quad (5.4)$

where $\triangle W = \sqrt{h}Z$ and Z is standard normally distributed.

This method corresponds to the explicit Euler method for ordinary differential equations. Note that we employ the same realizations of the Wiener process W_t in (5.3) and (5.4).

Example. The Euler-Maruyama scheme for the asset price dynamics (5.1) reads as

$$S_{i+1}^{(k)} = S_i^{(k)} (1 + rh + \sigma \triangle W^{(k)}), \quad i = 0, \dots, n-1, \ k = 1, \dots, m,$$

where $\triangle W^{(k)} = Z\sqrt{h}$, *Z* is *N*(0,1)-distributed, and *k* is the index of the Monte-Carlo simulation. The put option price is approximated by

$$P_m = e^{-rT} \frac{1}{m} \sum_{k=1}^m (K - S_n^{(k)})^+.$$

The exact solution P^* is given by the Black-Scholes formula (3.8). Figure 5.2 shows the relative error $|P_m - P^*|/P^*$ versus the number of simulations *m* using the parameters K = 100, r = 0.03, $\sigma = 0.3$, T = 1, $S_0 = 70$, and h = 0.02. We observe that the error decays rather slowly. This behavior will be discussed in more detail in Section 5.6.

Let us introduce the piecewise constant function $X_t^h = Y_i$ for $t \in (t_i, t_{i+1}]$, i = 0, ..., n - 1. We are interested in the quality of the approximation X_t^h , i.e. in the error $|X_t^h - X_t|$. This error is different for each realization of the Wiener process, so it is reasonable to average it. We define:

Figure 5.2: Relative error of the put option price versus number of the Monte-Carlo simulations, computed from the Euler-Maruyama scheme.



Definition 5.5 (Strong convergence). Let X_t be a solution to (5.3) (i.e. an Itô process) and let X_t^h be an approximation of X_t . We say that X_T^h converges strongly to X_T with order m > 0 at time T if there exists C > 0 such that for all (sufficiently small) h > 0,

$$\mathbf{E}|X_T^h - X_T| \le Ch^m.$$

Furthermore, X_T^h *is called* strongly convergent to X_T *if* $\lim_{h\to 0} E|X_T^h - X_T| = 0$.

Remark. We discuss how the expectation of $|X_T^h - X_T|$ is determined in practice. If X_1, \ldots, X_m are independent samples of a random variable *X* with the same expectation and variance, we may use the estimator

$$\varepsilon_m = \frac{1}{m} \sum_{i=1}^m X_i$$

It has the property of being unbiased and with converging variance, i.e.

$$E(\varepsilon_m) = \frac{1}{m} \sum_{i=1}^m E(X_i) = \frac{1}{m} \sum_{i=1}^m E(X) = E(X),$$
$$Var(\varepsilon_m) = \frac{1}{m^2} \sum_{i=1}^m Var(X_i) = \frac{1}{m} Var(X) \to 0 \quad \text{as } m \to \infty$$

Now, determine *m* realizations $W_t^{(1)}, \ldots, W_t^{(m)}$ of a Wiener process, let $X_t^{(k)}$ for $k = 1, \ldots, m$ be solutions to (5.3), and let $X_t^{h,(k)}$ for $k = 1, \ldots, m$ be defined by (5.4). Then we can approximate the error $E|X_T - X_T^h|$ by

$$\widehat{\varepsilon}_n = \frac{1}{m} \sum_{k=1}^m |X_T^{(k)} - X_T^{h,(k)}|.$$

The convergence rate of the explicit Euler method is one. As the difference $W_{t_{i+1}} - W_{t_i}$ has standard deviation $h^{1/2}$, we cannot expect the same convergence order for the Euler-Maruyama method. The following theorem shows that the order is only 1/2; see [20, Theorem 10.2.2].

Theorem 5.6 (Strong convergence rate of Euler-Maruyama). Let X_t be a solution to (5.3) and let X_t^h be given by the Euler-Maruyama scheme (5.4). Furthermore, assume that

there exists C > 0 independent of h such that for all $X, Y \in \mathbb{R}$ and $t \in [0, T]$,

$$\begin{split} & \mathsf{E}X_0^2 < \infty, \quad \left(\mathsf{E}|X_0 - X_0^h|^2\right)^{1/2} \le Ch^{1/2}, \\ & |a(X,t)| + |b(X,t)| \le C(1+|X|), \\ & |a(X,t) - a(Y,t)| + |b(X,t) - b(Y,t)| \le C|X - Y|, \\ & |a(X,t) - a(X,s)| + |b(X,t) - b(X,s)| \le C(1+|X|)|t - s|^{1/2} \end{split}$$

Then, for another constant C > 0 *which does not depend on* h*, such that for all* $t \in [0, T]$ *,*

$$\mathbf{E}|X_t - X_t^h| \le Ch^{1/2}.$$

Proof. We show first that EX_t^2 is bounded uniformly in $t \in [0, T]$. Equation (5.3) is written as

$$X_t = X_0 + \int_0^t a(X_s, s) ds + \int_0^t b(X_s, s) dW_s.$$

Taking the square, using the elementary inequality $(a + b + c)^2 \le 3(a^2 + b^2 + c^2)$, taking the expectation value, and applying the Cauchy-Schwarz inequality and the Itô isometry (Theorem 2.10), we obtain

$$\begin{aligned} \mathsf{E}X_{t}^{2} &\leq 3\mathsf{E}X_{0}^{2} + 3\mathsf{E}\bigg(\int_{0}^{t}a(X_{s},s)ds\bigg)^{2} + 3\mathsf{E}\bigg(\int_{0}^{t}b(X_{s},s)dW_{s}\bigg)^{2} \\ &\leq 3\mathsf{E}X_{0}^{2} + \mathsf{C}\mathsf{E}\int_{0}^{t}a(X_{s},s)^{2}ds + \mathsf{C}\mathsf{E}\int_{0}^{t}b(X_{s},s)^{2}ds. \end{aligned}$$

Since *a* and *b* grow at most linearly with X_s , by assumption, it follows that

$$EX_t^2 \le CEX_0^2 + CE\int_0^t (1+X_s^2)ds \le C(X_0,T) + \int_0^t EX_s^2ds$$

By Gronwall's lemma, we find that

$$\mathrm{EX}_t^2 \le C(X_0, T) \quad \text{for all } 0 \le t \le T.$$
(5.5)

In the following, we proceed by iteration. Assume that, for some i = 0, ..., n - 1, $E|X_{t_i} - X_{t_i}^h|^2 \le Ch$. If i = 0, this estimate holds by assumption. Next, for $t \in (t_i, t_{i+1}]$, we take the difference of

$$X_{t} = X_{t_{i}} + \int_{t_{i}}^{t} a(X_{s}, s)ds + \int_{t_{i}}^{t} b(X_{s}, s)dW_{s},$$

$$X_{t}^{h} = Y_{i} + \int_{t_{i}}^{t} a(Y_{i}, t_{i})ds + \int_{t_{i}}^{t} b(Y_{i}, t_{i})dW_{s}$$

to obtain similarly as above

$$\begin{split} \mathsf{E}|X_{t} - X_{t}^{h}|^{2} &\leq 3\mathsf{E}|X_{t_{i}} - Y_{i}|^{2} + 3\mathsf{E}\bigg(\int_{t_{i}}^{t} \big|a(X_{s},s) - a(Y_{i},t_{i})\big|ds\bigg)^{2} \\ &+ 3\mathsf{E}\bigg(\int_{t_{i}}^{t} \big|b(X_{s},s) - b(Y_{i},t_{i})\big|dW_{s}\bigg)^{2} \\ &\leq 3\mathsf{E}|X_{t_{i}} - Y_{i}|^{2} + \mathsf{C}\mathsf{E}\int_{t_{i}}^{t} \big|a(X_{s},s) - a(Y_{i},t_{i})\big|^{2}ds \\ &+ \mathsf{C}\mathsf{E}\int_{t_{i}}^{t} \big|b(X_{s},s) - b(Y_{i},t_{i})\big|^{2}ds. \end{split}$$

By the local Lipschitz continuity of *a*, we have

$$\begin{aligned} |a(X_s,s) - a(Y_i,t_i)| &\leq |a(X_s,s) - a(X_s,t_i)| + |a(X_s,t_i) - a(Y_i,t_i)| \\ &\leq C(1+|X_s|)h^{1/2} + C|X_s - Y_i|, \end{aligned}$$

and a similar estimate holds for *b*. Therefore, using the induction hypothesis and (5.5),

$$\begin{split} \mathsf{E}|X_{t} - X_{t}^{h}|^{2} &\leq 3\mathsf{E}|X_{t_{i}} - Y_{i}|^{2} + Ch\mathsf{E}\int_{t_{i}}^{t}(1 + X_{s}^{2})ds + C\mathsf{E}\int_{t_{i}}^{t}|X_{s} - Y_{i}|^{2}ds \\ &\leq Ch + Ch\int_{t_{i}}^{t}ds + Ch\int_{t_{i}}^{t}\mathsf{E}(X_{s}^{2})ds + C\mathsf{E}\int_{t_{i}}^{t}|X_{s} - Y_{i}|^{2}ds \\ &\leq Ch + C\int_{t_{i}}^{t}\mathsf{E}|X_{s} - X_{s}^{h}|^{2}ds. \end{split}$$

Gronwall's lemma implies that $E|X_t - X_t^h|^2 \leq C(X_0, T)h$. Finally, we use the Cauchy-Schwarz inequality to conclude that $E|X_t - X_t^h| \leq (E|X_t - X_t^h|^2)^{1/2} \leq Ch^{1/2}$.

Instead of estimating the expectation of the difference $|X_T - X_T^h|$, we may also consider the difference of the expectations, $|EX_T - EX_T^h|$. This leads to the concept of weak convergence.

Definition 5.7 (Weak convergence). Let X_t be a solution to (5.3) (i.e. an Itô process), let X_t^h be an approximation of X_t , and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. We say that X_T^h converges weakly to X_T with order m > 0 with respect to g at time T > 0 if there exists a constant C > 0 such that for all (sufficiently small) h > 0,

$$\left| \mathsf{E}(g(X_T)) - \mathsf{E}(g(X_T^h)) \right| \le Ch^m. \tag{5.6}$$

If g(x) = x, we simply say that X_T^h converges weakly to X_T with order m.

In this definition, we are not interested in the pathwise convergence, so we may use different paths for each time step $i \mapsto i + 1$ in the Monte-Carlo algorithm. The weak convergence concept is interesting for European options which are path-independent. Strong convergence is needed if the whole path plays a role, for instance, for path-dependent derivatives like Asian options.

Remark. (1) In the literature, sometimes the weak convergence is defined differently: X_T^h converges weakly if (5.6) holds for *all* smooth functions *g* with bounded derivatives.

(2) The weak convergence rate is always at least as large as the strong convergence rate. Indeed, let X_T^h be strongly convergent to X_T with order m and let $g \in C^1(\mathbb{R})$ be such that $g' \leq K$. Then, by the mean-value theorem,

$$\left| \mathsf{E}(g(X_T)) - \mathsf{E}(g(X_T^h)) \right| \le \mathsf{E}\left| g(X_T) - g(X_T^h) \right| \le K\mathsf{E}|X_T - X_T^h| \le KCh^m$$

where *K* depends on g', proving the claim.

While the strong convergence rate of the Euler-Marumyama scheme is only 1/2, the weak convergence rate is one; see [20, Theorem 14.1.5].

Theorem 5.8 (Weak convergence rate for Euler-Maruyama). Let X_t be a solution to (5.3) and let X_t^h be given by the Euler-Maruyama scheme (5.4). Furthermore, assume that a and b are smooth functions (at least C^4) with bounded derivatives and g is another smooth function. Then there exists a constant C > 0 which does not depend on h such that for h > 0,

$$\left| \mathsf{E}(g(X_T)) - \mathsf{E}(g(X_T^h)) \right| \le Ch.$$

Proof. The proof is based on the Faynman-Kac formula and Itô's lemma; see Theorems 3.5 and 2.12. Since it is rather technical, we present only some ideas. For a full proof, we refer to [23, Theorem 8.45].

Let $\varepsilon = E(g(X_T)) - E(g(X_T^h))$. We need to prove that $|\varepsilon| \leq Ch$ for some constant C > 0. By the Feynman-Kac formula, $u(x,t) = E(g(X_t))$ solves $\partial_t u - \frac{1}{2}b(x,t)^2 u_{xx} - a(x,t)u_x = 0$ with u(x,0) = g(x) and $dX_t = a(X_t,t)dt + b(X_t,t)dW_t$, $X_0 = x$. In particular, $u(X_0,T) = E(g(X_T))$. We consider another stochastic process,

$$dZ_t = a(Z_{t^*}, t^*)dt + b(Z_{t^*}, t^*)dW_t, \quad Z_0 = X_0,$$

and $t^* = t_i$ if $t \in [t_i, t_{i+1})$ (i.e., t^* is a function of t). Since the drift and diffusion are constant on the intervals $[t_i, t_{i+1})$, the process Z_t interpolates the numerical method, which means that $Z_{t_i} = X_{t_i}^h$. In particular, $u(Z_T, 0) = E(g(Z_T)) = E(g(X_T^h))$. Itô's formula gives

$$\varepsilon = u(X_0, T) - u(Z_T, 0) = u(Z_0, T) - u(Z_T, 0) = -E \int_0^T du(Z_t, T - t) du($$

where

$$du(Z_t, T-t) = \left(-\partial_t u + au_x + \frac{1}{2}b^2 u_{xx}\right)(Z_t, T-t)dt + bu_x(Z_t, T-t)dW_t$$

This expression is inserted in the expression for ε , taking into account that the expectation of the stochastic term vanishes:

$$\varepsilon = -\mathbf{E} \int_0^T \left(-\partial_t u + a u_x + \frac{1}{2} b^2 u_{xx} \right) (Z_t, T - t) dt.$$

The goal now is to show that

$$\varepsilon = \int_0^T \int_{t_i}^t EQ(Z_s, s) ds dt, \quad t \in [t_i, t_{i+1}),$$

for some process Q, where $EQ(Z_s, s)$ is bounded on (0, T). Since $|t - t_i| \le h$, this implies that $|\varepsilon| \le C|t - t_i| \le Ch$.

5.3 Milstein method

The strong convergence rate of the Euler-Maruyama method is only 1/2. This motivates the derivation of numerical schemes that are of higher order. To this end, we consider an Itô process $(X_t)_{t\geq 0}$ solving $dX_t = a(X_t)dt + b(X_t)dW_t$ and choose a smooth function $g : \mathbb{R} \to \mathbb{R}$. We proceed as in [33, Section 3.2].

Let $t \in (t_i, t_{i+1}]$. Then, by Itô's lemma,

$$g(X_t) = g(X_{t_i}) + \int_{t_i}^t \left(a(X_s)g'(X_s) + \frac{1}{2}b(X_s)^2g''(X_s) \right) ds + \int_{t_i}^t b(X_s)g'(X_s)dW_s.$$

The choice g = a and g = b gives, respectively,

$$a(X_t) = a(X_{t_i}) + \int_{t_i}^t \left(a(X_s)a'(X_s) + \frac{1}{2}b(X_s)^2 a''(X_s) \right) ds + \int_{t_i}^t b(X_s)a'(X_s) dW_s,$$

$$b(X_t) = b(X_{t_i}) + \int_{t_i}^t \left(a(X_s)b'(X_s) + \frac{1}{2}b(X_s)^2 b''(X_s) \right) ds + \int_{t_i}^t b(X_s)b'(X_s) dW_s.$$

Replacing $a(X_t)$ and $b(X_t)$ in

$$X_t = X_{t_i} + \int_{t_i}^t a(X_s) ds + \int_{t_i}^t b(X_s) dW_s$$

by these expressions, it follows that

$$X_{t} = X_{t_{i}} + \int_{t_{i}}^{t} \left(a(X_{t_{i}}) + \int_{t_{i}}^{s} \left(aa' + \frac{1}{2}b^{2}a'' \right)(X_{\tau})d\tau + \int_{t_{i}}^{s} (ba')(X_{\tau})dW_{\tau} \right) ds$$

$$+\int_{t_i}^t \left(b(X_{t_i}) + \int_{t_i}^s \left(ab' + \frac{1}{2}b^2b'' \right)(X_{\tau})d\tau + \int_{t_i}^s (bb')(X_{\tau})dW_{\tau} \right) dW_s$$

= $X_{t_i} + a(X_{t_i})(t - t_i) + b(X_{t_i})(W_t - W_{t_i}) + R,$ (5.7)

where the remainder *R* equals

$$R = \int_{t_i}^t \left(\int_{t_i}^s \left(aa' + \frac{1}{2}b^2 a'' \right) (X_{\tau}) d\tau + \int_{t_i}^s (ba')(X_{\tau}) dW_{\tau} \right) ds + \int_{t_i}^t \left(\int_{t_i}^s \left(ab' + \frac{1}{2}b^2 b'' \right) (X_{\tau}) d\tau + \int_{t_i}^s (bb')(X_{\tau}) dW_{\tau} \right) dW_s.$$

If we neglect *R*, we obtain the Euler-Maruyama method. We can derive a more precise scheme by approximating

$$R = b(X_{t_i})b'(X_{t_i})\int_{t_i}^t \int_{t_i}^s dW_{\tau}dW_s + O(h^{3/2}),$$

since the other terms are of order $O(h^2)$ (double deterministic integral) and $O(h^{3/2})$ (double deterministic-stochastic integrals). The error made is expected to be of order O(h), as dW_t is of "order" $O(h^{1/2})$. We have approximated $(bb')(X_\tau)$ by $(bb')(X_{t_i})$ since the double stochastic integral can be computed explicitly:

$$\int_{t_i}^t \int_{t_i}^s dW_{\tau} dW_s = \int_{t_i}^t (W_s - W_{t_i}) dW_s = \int_{t_i}^t W_s dW_s - W_{t_i} (W_t - W_{t_i}).$$

To determine the remaining stochastic integral, we apply Itô's formula (Theorem 2.12) to $f(x) = x^2$ and a = 0, b = 1:

$$dW_t^2 = df(W_t) = \left(f_x(W_t)a + \frac{1}{2}f_{xx}(W_t)b^2\right)dt + f_x(W_t)bdW_t = dt + 2W_t dW_t,$$

or in integral form,

$$\int_{t_i}^t dW_s^2 = \int_{t_i}^t ds + 2 \int_{t_i}^t W_s dW_s$$

This gives

$$\int_{t_i}^t W_s dW_s = \frac{1}{2} (W_t^2 - W_{t_i}^2) - \frac{t - t_i}{2},$$

and we infer that

$$\int_{t_i}^t \int_{t_i}^s dW_\tau dW_s = \frac{1}{2} \big((W_t - W_{t_i})^2 - (t - t_i) \big).$$
(5.8)

This leads to the *Milstein method*.

Definition 5.9 (Milstein scheme).

$$Y_{i+1} = Y_i + a(Y_i, t_i)h + b(Y_i, t_i) \triangle W + \frac{1}{2}b(Y_i, t_i)\frac{\partial b}{\partial x}(Y_i, t_i)\big((\triangle W)^2 - h\big),$$
(5.9)

where i = 0, ..., n - 1, $\triangle W = \sqrt{hZ}$, and Z is standard normally distributed.

Example. Compared to the Euler-Maruyama scheme, the Milstein scheme just contains an additional term. The scheme becomes for the asset price dynamics (5.1):

$$S_{i+1}^{(k)} = S_i^{(k)} \left(1 + rh + \sigma \triangle W + \frac{1}{2} \sigma^2 ((\triangle W)^2 - h) \right),$$

where i = 0, ..., n - 1, k = 1, ..., m. We show in Figure 5.3 the relative errors of the Monte-Carlo put price computed from the Euler-Maruyama and Milstein schemes. For a small number of simulations, the solution from the Milstein scheme does not show a significant improvement; this becomes more apparent for large simulation numbers, here for m > 80,000. By the way: Both curves are not just shifted; their difference is of the order of 10^{-4} which is too small to be visible.



Figure 5.3: Relative errors of the put option price versus number of the Monte-Carlo simulations, computed from the Euler-Maruyama and Milstein schemes.

The Milstein method is weakly and strongly convergent with order m = 1. When the diffusion term $b(X_t, t)$ is constant in the first argument, the Milstein scheme reduces to the Euler-Maruyama method. In this special case, the Euler-Maruyama method is strongly convergent with order m = 1 instead of the general order m = 1/2.

Theorem 5.10 (Strong convergence of Milstein scheme). Let X_t be a solution to (5.3) and let X_t^h be given by the Milstein scheme (5.9). Furthermore, let a and b be smooth functions

with bounded derivatives, $E|X_0 - X_0^h| \le Ch$, and let bb' be globally Lipschitz continuous. Then there exists a constant C > 0 independent of h such that

 $\mathbf{E}|X_T - X_T^h| \le Ch.$

Since the weak convergence rate is always as large as the strong convergence rate, the Milstein scheme is weakly *and* strongly convergent with order one.

Proof. We only sketch the proof; for details we refer to [23, Theorem 8.32]. The calculation (5.7) at the beginning of this section reveals that

$$X_{t} = X_{t_{i}} + a(X_{t_{i}}, t_{i})(t - t_{i}) + b(X_{t_{i}}, t_{i})(W_{t} - W_{t_{i}}) + \int_{t_{i}}^{t} \int_{t_{i}}^{s} (bb')(X_{\tau}, \tau)dW_{\tau}dW_{s} + \widetilde{R},$$

$$X_{t}^{h} = Y_{i} + a(Y_{i}, t_{i})(t - t_{i}) + b(Y_{i}, t_{i})(W_{t} - W_{t_{i}}) + \frac{1}{2}(bb')(Y_{i}, t_{i})((W_{t} - W_{t_{i}})^{2} - (t - t_{i})),$$

where the prime means derivation with respect to the first variable and the remainder \widetilde{R} is given by

$$\widetilde{R} = \int_{t_i}^t \left(\int_{t_i}^s \left(aa' + \frac{1}{2}b^2 a'' \right) (X_\tau) d\tau + \int_{t_i}^s (ba')(X_\tau) dW_\tau \right) ds$$
$$+ \int_{t_i}^t \int_{t_i}^s \left(ab' + \frac{1}{2}b^2 b'' \right) (X_\tau) d\tau dW_s.$$

Then $\mathbb{E}|X_t - X_t^h| \leq I_1 + \cdots + I_4 + |\widetilde{R}|$, where

$$\begin{split} I_{1} &= E|X_{t_{i}} - Y_{i}|, \\ I_{2} &= E|a(X_{t_{i}}, t_{i}) - a(Y_{i}, t_{i})||t - t_{i}|, \\ I_{3} &= E|b(X_{t_{i}}, t_{i}) - b(Y_{i}, t_{i})||W_{t} - W_{t_{i}}|, \\ I_{4} &= E\left|\int_{t_{i}}^{t}\int_{t_{i}}^{s} (bb')(X_{\tau}, \tau)dW_{\tau}dW_{s} - \frac{1}{2}(bb')(Y_{i}, t_{i})((W_{t} - W_{t_{i}})^{2} - h)\right|. \end{split}$$

We proceed by induction. Assume that $E|X_{t_i} - Y_i| \le Ch$ for some *i*. This means that $I_1 \le Ch$. Since *a* and *b* have bounded derivatives, these functions are Lipschitz continuous, and it follows that $I_2 \le Ch^2$ and $I_3 \le Ch^{3/2}$. It remains to show that $I_4 \le Ch$ and $\tilde{R} \le Ch$ (this follows from the approximation of the stochastic integrals). Assuming these estimates, we conclude that $E|X_t - X_t^h| \le Ch$ for $t = t_{i+1}$ and consequently $E|X_{t_{i+1}} - Y_{i+1}| \le Ch$. This finishes the proof.

5.4 Itô-Taylor and Runge-Kutta schemes

The Euler-Maruyama and Milstein schemes are derived from an Itô-Taylor expansion, which means that they are derived from a stochastic Taylor expansion of the solution in

the integral representation. A disadvantage of this approach is that partial derivatives need to be evaluated, which decreases the efficiency of the scheme. This problem can be solved by replacing the partial derivatives by a Taylor expansion, leading to derivativefree schemes or so-called stochastic Runge-Kutta methods. We present two examples.

• **Strong order 1.0 Runge-Kutta method.** The Milstein method requires the evaluation of the partial derivative $\partial b / \partial x$. This can be avoided by replacing the derivative by its first-order Taylor expansion:

$$b(X + \triangle X) - b(X) = b'(X) \triangle X + O(\triangle X^2).$$

Using $\triangle X = a(X)h + b(X) \triangle W$ (this is the Euler-Maruyama approximation of dX = a(X)dt + b(X)dW) and consequently $O(\triangle X^2) = O(\triangle W^2) = O(h)$, we obtain

$$b(X + \triangle X) - b(X) = b'(X) (a(X)h + b(X)\triangle W) + O(h) = b'(X)b(X)\triangle W + O(h).$$

The expectation of $\triangle W$ is \sqrt{h} , so we may write

$$b(X + \triangle X) - b(X) = b'(X)b(X)\sqrt{h} + O(h)$$

or, replacing $\triangle X$ in this expression by $a(X)h + b(X)\sqrt{h}$ and solving for b'(X)b(X),

$$b'(X)b(X) = \frac{1}{\sqrt{h}}(b(\widetilde{X}) - b(X)) + O(\sqrt{h}), \text{ where } \widetilde{X} = X + a(X)h + b(X)\sqrt{h}.$$

This gives the following variant of the Milstein scheme.

Definition 5.11 (Strong order 1.0 Runge-Kutta scheme).

$$Y_{i+1} = Y_i + a(Y_i, t_i)h + b(Y_i, t_i) \triangle W + \frac{1}{2\sqrt{h}} (b(\widetilde{Y}_i, t_i) - b(Y_i, t_i)) ((\triangle W)^2 - h),$$

where $\widetilde{Y}_i = Y_i + a(Y_i, t_i)h + b(Y_i, t_i)\sqrt{h}, \quad \triangle W = \sqrt{h}Z, \quad Z \sim N(0, 1).$

The strong convergence order is the same as for the Milstein scheme, but we only need to evaluate functions and not its derivatives. Instead of evaluating $b(Y_i, t_i)$ and $(\partial b/\partial x)(Y_i, t_i)$, we need to evaluate only $b(Y_i, t_i)$ and $b(\tilde{Y}_i, t_i)$.

This approach can be extended to general Runge-Kutta schemes, but the strong order 1.5 cannot be surpassed if we use only the increments $\triangle W$ of the Wiener process. This order limit can be bypassed by introducing additional random variables to approximate the iterated stochastic integrals like $\int_{t_i}^t \int_{t_i}^s dW_{\tau} dW_s$. This idea was used in [5] to construct strong order 2.0 Runge-Kutta methods in which the deterministic component of the method is given by the classical (explicit four-stages) Runge-Kutta method. Higher-order methods can also be constructed using Itô-Taylor expansions, and in the following, we give an example. • **Strong order 1.5 Taylor method.** We derive the scheme by taking into account the remainder in (5.7). Instead of approximating

$$R_{1} := \int_{t_{i}}^{t} \int_{t_{i}}^{s} (bb')(X_{\tau}) dW_{\tau} dW_{s} \approx (bb')(X_{t_{i}}) \int_{t_{i}}^{t} \int_{t_{i}}^{s} dW_{\tau} dW_{s}$$
$$= \frac{1}{2} (bb')(X_{t_{i}}) ((W_{t} - W_{t_{i}})^{2} - (t - t_{i})),$$

which is only of first order, we apply Itô's formula to compute a better approximation:

$$(bb')(X_t) = (bb')(X_{t_i}) + \int_{t_i}^t \left(a(bb')' + \frac{1}{2}b^2(bb')'' \right)(X_s)ds + \int_{t_i}^t b(bb')'(X_s)dW_s.$$

The second term is of order O(h). Then, integrating the previous equation twice with respect to the Wiener process,

$$R_{1} = (bb')(X_{t_{i}}) \int_{t_{i}}^{t} \int_{t_{i}}^{s} dW_{\tau} dW_{s} + O(h) \int_{t_{i}}^{t} \int_{t_{i}}^{s} dW_{\tau} dW_{s}$$
$$+ \int_{t_{i}}^{t} \int_{t_{i}}^{s} \int_{t_{i}}^{\tau} (b(bb')')(X_{z}) dW_{z} dW_{\tau} dW_{s}.$$

Using $\int_{t_i}^t \int_{t_i}^s dW_{\tau} dW_s = \frac{1}{2}((W_t - W_{t_i})^2 - (t - t_i))$ (see (5.8)), it follows that

$$R_{1} = \frac{1}{2} (bb') (X_{t_{i}}) ((W_{t} - W_{t_{i}})^{2} - (t - t_{i})) + O(h^{2}) + \int_{t_{i}}^{t} \int_{t_{i}}^{s} \int_{t_{i}}^{\tau} (b(bb')') (X_{z}) dW_{z} dW_{\tau} dW_{s}.$$

Finally, we approximate $(b(bb')')(X_z)$ by $(b(bb')')(X_{t_i})$ (which introduces an error of order $O(h^{3/2})$) and use the formula

$$\int_{t_i}^t \int_{t_i}^s \int_{t_i}^\tau dW_z dW_\tau dW_s = \frac{1}{2} \left(\frac{1}{3} (W_t - W_{t_i})^3 - (t - t_i) \right) (W_t - W_{t_i})$$

to find that

$$R_{1} = \frac{1}{2} (bb') (X_{t_{i}}) \left((W_{t} - W_{t_{i}})^{2} - (t - t_{i}) \right) + \frac{1}{2} (b(bb')') (X_{t_{i}}) \left(\frac{1}{3} (W_{t} - W_{t_{i}})^{3} - (t - t_{i}) \right) (W_{t} - W_{t_{i}}) + O(h^{3/2}).$$

Then the remainder *R* in (5.7) becomes at $t = t_{i+1}$:

$$R = \int_{t_i}^t \int_{t_i}^s \left(aa' + \frac{1}{2}b^2a''\right)d\tau ds + \int_{t_i}^t \int_{t_i}^s (ba')dW_\tau ds$$

$$\begin{split} &+ \int_{t_i}^t \int_{t_i}^s \left(ab' + \frac{1}{2}b^2b'' \right) d\tau dW_s + R_1 \\ &= \frac{1}{2} \left(aa' + \frac{1}{2}b^2a' \right) (X_{t_i})h^2 + (ba')(X_{t_i}) \int_{t_i}^t \int_{t_i}^s dW_\tau ds \\ &+ \left(ab' + \frac{1}{2}b^2b'' \right) (X_{t_i}) \int_{t_i}^t \int_{t_i}^s d\tau dW_s + \frac{1}{2}(bb')(X_{t_i}) \left((W_t - W_{t_i})^2 - (t - t_i) \right) \\ &+ \frac{1}{2} (b(bb')')(X_{t_i}) \left(\frac{1}{3} (W_t - W_{t_i})^3 - (t - t_i) \right) (W_t - W_{t_i}) + O(h^{3/2}). \end{split}$$

The following lemma shows that the two double integrals are related.

Lemma 5.12. It holds that

$$\int_{t_i}^t \int_{t_i}^s dW_{\tau} ds + \int_{t_i}^t \int_{t_i}^s d\tau dW_s = (W_t - W_{t_i})(t - t_i)$$

Proof. We apply the Itô formula to the Itô process $X_t = W_t$ (i.e. a = 0 and b = 1 in the formula) with the function g(x, t) = xt:

$$g(W_t,t) = g(W_{t_i},t_i) + \int_{t_i}^t \left(\partial_t g + ag_x + \frac{1}{2}b^2 g_{xx}\right)(W_s,s)ds + \int_{t_i}^t bg_x(W_s,s)dW_s.$$

This gives

$$tW_t = t_i W_{t_i} + \int_{t_i}^t W_s ds + \int_{t_i}^t s dW_s$$

or, solving for the last integral,

$$\int_{t_i}^t s dW_s = tW_t - t_i W_{t_i} - \int_{t_i}^t W_s ds.$$

Next, we compute, using the previous identity,

$$\int_{t_i}^t \int_{t_i}^s dW_\tau ds = \int_{t_i}^t (W_s - W_{t_i}) ds = \int_{t_i}^t W_s ds - W_{t_i}(t - t_i),$$

$$\int_{t_i}^t \int_{t_i}^s d\tau dW_s = \int_{t_i}^t (s - t_i) dW_s = \int_{t_i}^t s dW_s - t_i (W_t - W_{t_i})$$

$$= W_t (t - t_i) - \int_{t_i}^t W_s ds.$$

The lemma follows after adding these expressions.

Therefore, the double integrals can be formulated in terms of a single stochastic process, $\triangle Z = \int_{t_i}^t \int_{t_i} dW_{\tau} ds$. In view of (5.7), this leads to the following scheme which is of strong convergence order 1.5:

$$\begin{split} Y_{i+1} &= Y_i + a(Y_i, t_i)h + b(Y_i, t_i) \triangle W + \frac{1}{2}(bb')(Y_i, t_i) \big((\triangle W)^2 - h\big) \\ &+ (ba')(Y_i, t_i) \triangle Z + \frac{1}{2} \bigg(aa' + \frac{1}{2}b^2a''\bigg)(Y_i, t_i)h^2 + \bigg(ab' + \frac{1}{2}b^2b''\bigg)(\triangle Wh - \triangle Z) \\ &+ \frac{1}{2}b(Y_i, t_i)\big(bb'' + (b')^2\big)(Y_i, t_i)\bigg(\frac{1}{3}(\triangle W)^3 - h\bigg) \triangle W, \end{split}$$

where i = 0, ..., n - 1, the prime means differentiation with respect to the x-variable, and $\triangle Z$ represents the double integral $\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} dW_{\tau} ds$.

When we are interested in weak convergence only, we can replace the random variables $\triangle W$ and $\triangle Z$ by simpler ones with the same moments. A computation shows that $\triangle Z$ has the expectation, variance, and covariance

$$E(\triangle Z) = 0$$
, $E(\triangle Z)^2 = \frac{h^3}{3}$, $E(\triangle W \triangle Z) = \frac{h^2}{2}$.

One can see that the pair of correlated normally distributed random variables (ΔW , ΔZ) can be realized from two independent standard normally distributed random variables (Y_1, Y_2) via

$$\triangle W = \sqrt{h}Y_1, \quad \triangle Z = \frac{h^{3/2}}{2}\left(Y_1 + \frac{Y_2}{\sqrt{3}}\right).$$

This scheme has the weak convergence order two.

- **Remark.** (1) We can also construct weakly convergent schemes of higher order, but because of the approximation of the multiple integrals, the computational effort becomes high. Moreover, in contrast to higher-order schemes for ordinary differential equations, higher-order solvers for stochastic differential equations are often less compelling. The reason is that often the initial condition is chosen from a probability distribution and higher-order methods do not necessarily improve the accuracy of the solution.
 - (2) All presented schemes are explicit. Explicit schemes for stiff discretizations (arising from stochastic partial differential equations, for instance), suffer from severe stability issues. Therefore, implicit approximations seem to be preferable. As outlined in [20, Section 9.8], implicit schemes may lead to some difficulties, because they involve reciprocals of Gaussian random variables not having finite absolute moments. This issue may be avoided by employing semi-implicit methods that are explicit in the diffusion term and implicit in the drift term. A simple example is the semi-implicit Euler-Maruyama scheme

$$Y_{i+1} = Y_i + a(Y_{i+1}, t_{i+1})h + b(Y_i, t_i) \triangle W, \quad i = 0, \dots, n-1.$$

This scheme has the strong convergence order 1/2 like the explicit Euler-Maruyama method. We refer to [20, Chapter 12] for more details, in particular on the stability properties of implicit schemes.

5.5 Systems of stochastic differential equations

We have considered so far only scalar equations. However, the modeling of options with stochastic volatility or of basket options leads to systems of stochastic differential equations. In this section, we extend the Milstein scheme to systems of dimension n. We proceed as in [16, Section 5.3.4]. More precisely, we consider

$$dX^{(j)}(t) = a^{(j)}(X_t, t)dt + \sum_{k=1}^m b^{(jk)}(X_t, t)dW^{(k)}(t), \quad j = 1, \dots, n,$$

where $X_t = (X^{(1)}(t), \dots, X^{(n)}(t))$. This system can be written in compact form as

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t, \quad t > 0,$$

where $a : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^n$, $b : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^{n \times m}$, and $W_t = (W^{(1)}(t), \dots, W^{(m)}(t))$ is an *m*-dimensional Wiener process.

In case m = 1, we can interpret the term bb' in the Milstein scheme as the product of the Jacobian $b' = (\partial b^{(i)} / \partial x_j)$ and the vector $b = (b^{(1)}, \dots, b^{(n)})^{\top}$, and the Milstein scheme becomes

$$Y_{i+1}^{(j)} = Y_i^{(j)} + a^{(j)}(Y_i, t_i)h + b^{(j)}(Y_i, t_i) \triangle W + \frac{1}{2} \sum_{k=1}^m \frac{\partial b^{(j)}}{\partial x_k} (Y_i, t_i) b^{(k)}(Y_i, t_i) ((\triangle W)^2 - h).$$

The general case m > 1 is more involved, and we report only the result. Repeating the derivation of the Milstein scheme from Section 5.3, it holds that

$$Y_{i+1}^{(j)} = Y_i^{(j)} + a^{(j)}(Y_i, t_i)h + \sum_{k=1}^n b^{(jk)}(Y_i, t_i) \triangle W^{(k)} + \sum_{p,q=1}^m \sum_{\ell=1}^n \frac{\partial b^{(jq)}}{\partial x_\ell}(Y_i, t_i) b^{(\ell p)}(Y_i, t_i) \int_{t_i}^{t_{i+1}} \int_{t_i}^s dW_{\tau}^{(p)} dW_s^{(q)}$$

We face two problems:

- How can we determine efficiently the double integrals $I_{pq} = \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} dW_{\tau}^{(p)} dW_{s}^{(q)}$?
- ▶ How can we compute efficiently the derivatives $(\partial b^{(jq)} / \partial x_{\ell})(Y_i, t_i)$?

Let us consider the first problem. It is sufficient to compute the integral I_{pq} up to O(h), since the Milstein scheme is of first order. The idea is to formulate the integral as the solution of a system of stochastic differential equations with one-dimensional Wiener processes and to discretize this system using the Euler-Maruyama method. For notational simplicity, let p = 2 and q = 1.

Lemma 5.14. The solution to the stochastic differential system

$$dX_{t} = \begin{pmatrix} X_{t}^{(2)} & 0\\ 0 & 1 \end{pmatrix} dW_{t}, \quad t_{i} \le t \le t_{i+1}, \quad X_{t_{i}} = \begin{pmatrix} 0\\ 0 \end{pmatrix},$$
(5.10)

at $t = t_{i+1}$ reads as

$$X_{t_{i+1}} = \begin{pmatrix} I_{21} \\ riangle W^{(2)} \end{pmatrix}$$
,

where $\triangle W^{(2)} = W^{(2)}_{t_{i+1}} - W^{(2)}_{t_i}$.

Proof. The second equation $dX_t^{(2)} = dW_t^{(2)}$ is the Itô process

$$X_s^{(2)} = X_{t_i}^{(2)} + \int_{t_i}^s dW_{\tau}^{(2)} = \int_{t_i}^s dW_{\tau}^{(2)}.$$

Inserting this expression into the first equation $dX_t^{(1)} = X_t^{(2)} dW_t^{(1)}$ then gives

$$X_{t_{i+1}}^{(1)} = X_{t_i}^{(1)} + \int_{t_i}^{t_{i+1}} X_s^{(2)} dW_s^{(1)} = \int_{t_i}^{t_{i+1}} \int_{t_i}^s dW_{\tau}^{(2)} dW_s^{(1)}.$$

This finishes the proof.

For the approximation of (5.10), we divide the interval $[t_i, t_{i+1}]$ into N subintervals of length $\delta = h/N$. Let $Z_k = (Z_k^{(1)}, Z_k^{(2)})$ be an approximation of the intermediate values $X_{t_i+k\delta}$. Then $Z_0 = X_{t_i} = 0$ and the Euler-Maruyama approximation of (5.10) is

$$Z_{k+1} = Z_k + \begin{pmatrix} Z_k^{(2)} & 0 \\ 0 & 1 \end{pmatrix} \triangle W_k, \text{ where } \triangle W_k = W_{t_i + (k+1)\delta} - W_{t_i + k\delta}$$

and k = 0, ..., N - 1. The Euler-Maruyama scheme has the strong convergence order 1/2, so one may ask the question whether this leads to an approximation of I_{21} that is of first order. This is the case if we choose $N \ge 1/h$. Indeed, because of $\delta = h/N$,

$$\mathbb{E}|Z_N^{(1)} - I_{21}| \le C\delta^{1/2} = C\left(\frac{h}{N}\right)^{1/2} \le Ch.$$

Next, we turn to the second problem. As in the previous subsection, we can avoid the computation of the derivatives $\partial b / \partial x_{\ell}$ by using a stochastic Runge-Kutta scheme. For this, we replace the partial derivative by a suitable finite difference.

Lemma 5.15. It holds that

$$\sum_{\ell=1}^{n} \left(\frac{b^{(jq)}}{\partial x_{\ell}} b^{(\ell p)} \right) (Y_i, t_i) = \frac{1}{\sqrt{h}} \left(b^{(jq)} (\widetilde{Y}_i^{(p)}, t_i) - b^{(jq)} (Y_i, t_i) \right) + O(\sqrt{h}),$$

where

$$\widetilde{Y}_i^{(p)} = Y_i + a(Y_i, t_i)h + b^{(p)}(Y_i, t_i) \cdot \sqrt{H},$$

 $b^{(p)}$ are vectors and $\sqrt{H} = (\sqrt{h}, \dots, \sqrt{h}) \in \mathbb{R}^n$.

Proof. The result follows from a Taylor expansion:

$$\begin{split} \frac{1}{\sqrt{h}} & \left(b^{(jq)}(\widetilde{Y}_{i}^{(p)}, t_{i}) - b^{(jq)}(Y_{i}, t_{i}) \right) = \frac{1}{\sqrt{h}} \left(Db^{(jq)}(Y_{i}, t_{i}) \cdot \left(\widetilde{Y}_{i}^{(p)} - Y_{i} \right) + O(h) \right) \\ &= \frac{1}{\sqrt{h}} \left(Db^{(jq)}(Y_{i}, t_{i}) \cdot \left(a(Y_{i}, t_{i})h + b^{(p)}(Y_{i}, t_{i}) \cdot \sqrt{H} \right) + O(\sqrt{h}) \\ &= Db^{(jq)}(Y_{i}, t_{i}) \cdot b^{(p)}(Y_{i}, t_{i}) + O(\sqrt{h}), \end{split}$$

where $Db^{(jq)}$ is the gradient of $b^{(jq)}$ with respect to (x_1, \ldots, x_n) .

We summarize the strong order 1.0 Runge-Kutta scheme for systems:

Definition 5.16 (Strong order 1.0 Runge-Kutta scheme for systems). $Y_{i+1} = Y_i + a(Y_i, t_i)h + b(Y_i, t_i) \cdot \Delta W$ $+ \frac{1}{\sqrt{h}} \sum_{p,q=1}^m \left(b^{(q)} (\widetilde{Y}_i^{(p)}, t_i) - b^{(q)} (Y_i, t_i) \right) I_{pq},$ where $\Delta W = \sqrt{h}Z$, $Z \sim N(0, 1)$, and $\widetilde{Y}_i^{(p)} = Y_i + a(Y_i, t_i)h + b^{(p)}(Y_i, t_i) \cdot \sqrt{H}, \quad I_{pq} = \int_{t_i}^{t_{i+1}} \int_{t_i}^s dW_{\tau}^{(p)} dW_s^{(q)},$ recalling that $\sqrt{H} = (\sqrt{h}, \dots, \sqrt{h}) \in \mathbb{R}^n$.

Remark. The double integrals I_{pq} can be explicitly calculated in a few cases. Indeed, if p = q, we obtain the double integral already computed for the Milstein scheme, leading to

$$I_{pp} = \frac{1}{2} ((\bigtriangleup W^{(p)})^2 - h), \text{ where } \bigtriangleup W^{(p)} = W^{(p)}_{t_{i+1}} - W^{(p)}_{t_i}.$$

When *b* does not depend on *x* (called additive noise), the derivative $\partial b / \partial x_{\ell}$ vanishes, and the Milstein and Runge-Kutta schemes reduce to the Euler-Maruyama method. Finally, if *b* is a diagonal matrix and n = m (called diagonal noise), we only need to approximate $b_j(\partial b_j/\partial x_j)$, which leads to the double integrals I_{ij} .

5.6 Quasi Monte-Carlo method and variance reduction

In the previous sections, we have explained how to simulate realizations of Wiener processes and to approximate the corresponding stochastic differential equations. We explore in the following the accuracy of the Monte-Carlo method and present some techniques to accelerate the algorithm. We consider an abstract framework. The goal is to approximate the expected value $\theta = E(\phi(X))$, where $\phi(X)$ is a functional of a random variable *X*. Let $X_h^{(k)}$ for $k = 1, \ldots, m$ be sample paths from Monte-Carlo simulations with time step *h*. We assume that the approximation has the weak convergence order one and that the variances of $\phi(X_h^{(k)})$ and $\phi(X)$ are the same,

$$\operatorname{Var}(\phi(X_h^{(k)})) = \operatorname{Var}(\phi(X)) =: \sigma^2.$$

The expectation $\theta = E(\phi(X))$ is approximated by the estimator

$$\theta_m = \frac{1}{m} \sum_{k=1}^m \phi(X_h^{(k)}),$$

and we expect that $\theta_m \to E(\phi(X))$ as $h \to 0$ and $m \to \infty$. We wish to estimate $\phi(X)$ with accuracy ε in the mean square error, MSE = $O(\varepsilon^2)$, where

$$MSE = E[(\theta_m - E(\phi(X)))^2].$$

We split the mean square error into two parts:

$$MSE = E[(\theta_m - E(\theta_m) + E(\theta_m) - E(\phi(X)))^2]$$

= $E[(\theta_m - E(\theta_m))^2] + E[(E(\theta_m) - E(\phi(X)))^2]$
+ $2E[(\theta_m - E(\theta_m))(E(\theta_m) - E(\phi(X)))]$
= $E[(\theta_m - E(\theta_m))^2] + (E(\theta_m) - E(\phi(X)))^2$
+ $2E[\theta_m - E(\theta_m)] (E(\theta_m) - E(\phi(X)))$
= $E[(\theta_m - E(\theta_m))^2] + (E(\theta_m) - E(\phi(X)))^2.$

The first term on the right-hand side is the Monte-Carlo variance, $Var(\theta_m)$, the second term is the bias of the approximation. As we assumed that the weak convergence order equals one, we have $|E(\theta_m) - E(\phi(X))| = O(h)$ as $h \to 0$. The Monte-Carlo variance is proportional to 1/m since

$$\operatorname{Var}(\theta_m) = \frac{1}{m^2} \operatorname{Var}\left(\sum_{k=1}^m \phi(X_h^{(k)})\right) = \frac{1}{m^2} \sum_{k=1}^m \sigma^2 = \frac{\sigma^2}{m}.$$

We infer that the mean square error of the Monte-Carlo method is

$$MSE = O\left(\frac{1}{m}\right) + O(h^2)$$
 as $m \to \infty$ and $h \to 0$.

In the following, we write a = O(b) to indicate that both a and b are of the same order, i.e. $C_1a \le b \le C_2a$ for some constants $C_1 \le C_2$. To ensure that MSE is proportional

to ε^2 , we need that $1/m = O(\varepsilon^2)$ and $h^2 = O(\varepsilon^2)$, which means that $m = O(\varepsilon^{-2})$ and $h = O(\varepsilon)$. The computational cost of the Monte-Carlo method is proportional to the number of paths *m* times the cost of generating one path, which is the number of time steps in each path, 1/h. Thus, the cost of the algorithm is proportional to $m/h = O(\varepsilon^{-3})$. This means that if we want to reduce the error by one order of magnitude (i.e. by the factor 10), the cost is increased by the factor 100!

Another question concerns the absolute error $|\theta_m - E(\phi(X))|$. By the Chebychev inequality,

$$\mathbf{P}(|\theta_m - \mathbf{E}(\boldsymbol{\phi}(X))| \ge \delta) \le \frac{1}{\delta^2} \mathbf{E}[(\theta_m - \mathbf{E}(\boldsymbol{\phi}(X)))^2] = \frac{\mathbf{MSE}}{\delta^2}.$$

We know from the previous computation that MSE = $Var(\theta_m) + O(h^2) = \sigma^2/m + O(h^2)$. Then, choosing $\delta = \sqrt{MSE/\varepsilon}$, we find that

$$P\left(|\theta_m - E(\phi(X))| \ge \sqrt{\frac{MSE}{\varepsilon}}\right) \le \varepsilon$$

or, since $\sqrt{\text{MSE}} = \sigma / \sqrt{m} + O(h)$ as $h \to 0$,

$$P\left(\left|\theta_m - E(\phi(X))\right| < \frac{\sigma}{\sqrt{\varepsilon}\sqrt{m}} + O\left(\frac{h}{\sqrt{\varepsilon}}\right)\right) > 1 - \varepsilon.$$
(5.11)

This means that the reduction of the absolute error by one order of magnitude requires to decrease the time step by the same factor *and* to increase the number of sample paths by the factor 100.

Therefore, it is reasonable to devise strategies to decrease the error of the Monte-Carlo method. In the following, we discuss two strategies:

- quasi Monte-Carlo methods,
- variance reduction techniques.

• Quasi Monte-Carlo method. The idea of the quasi Monte-Carlo method is to use random numbers that are more equidistributed than pseudo random numbers. Random numbers are independent and identically distributed (at least approximately in case of pseudo random numbers). Often, the independence is not crucial to the computation. If this property can be discarded, one may create so-called *low-discrepancy numbers*. The discrepancy is a measure for the equidistribution of the random numbers. An easy example is the *van der Corput sequence* with base 2:

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \dots$$

The *i*th element of the sequence is computed by bit inversion:

$$i = \sum_{k=0}^{j} d_k 2^k \quad \mapsto \quad x_i = \sum_{k=0}^{j} d_k 2^{-k-1}.$$

For instance, $i = 5 = 2^2 + 2^0$ gives $d_2 = 1$, $d_1 = 0$, $d_0 = 1$ and consequently, $x_5 = 1/2^1 + 0/2^2 + 1/2^3 = 5/8$. This ansatz can be generalized to an arbitrary base $b \ge 2$. The van der Corput sequence uniformly fills out the unit interval. We compare the van der Corput numbers with uniformly distributed pseudo random numbers in Figure 5.4.



Figure 5.4: Left: The first 5000 van der Corput numbers with base 2 (*x*-axis) and 3 (*y*-axis). Right: 5000 uniformly distributed pseudo random numbers, generated by the Mersenne-Twister generator.

It is possible to show that the error of the approximation by the quasi Monte-Carlo method is $O((\log m)^s/m)$, where s > 0, whereas the Monte-Carlo error $\sqrt{\text{MSE}}$ is of the order $O(1/\sqrt{m})$ (see (5.11)). Thus, the accuracy of the quasi Monte-Carlo method increases generally faster than that of the Monte-Carlo method. However, this is guaranteed only if the number of paths *m* is large.

The second approach is to decrease the variance σ in (5.11). The idea is to calculate the expectation more accurately with fewer calls to the random-number generator. We discuss two techniques:

- antithetic variables,
- control variates.

• Antithetic variables. Let $\theta = EY$, where $Y = \phi(X)$, $\phi : \mathbb{R} \to \mathbb{R}$ is some function and *X* a random variable. The goal is to estimate the expectation θ . Let Y_1 and Y_2 be two samples of *Y* with the same expectation and same variance as *Y*. The estimator

$$\widehat{Y} = \frac{1}{2}(Y_1 + Y_2)$$

is unbiased since $E\hat{Y} = \frac{1}{2}(EY_1 + EY_2) = EY$. In the following, we use the identity

$$0 \le \operatorname{Var}(Y_1 \pm Y_2) = \operatorname{E}\left[\left((Y_1 - EY_1) \pm (Y_2 - EY_2)\right)^2\right]$$

=
$$\operatorname{E}\left[(Y_1 - EY_1)^2 + (Y_2 - EY_2)^2 \pm 2(Y_1 - EY_1)(Y_2 - EY_2)\right]$$

$$= \operatorname{Var}(Y_1) + \operatorname{Var}(Y_2) \pm 2\operatorname{Cov}(Y_1, Y_2), \tag{5.12}$$

from which we deduce (taking the minus sign) that

$$Cov(Y_1, Y_2) \le \frac{1}{2} (Var(Y_1) + Var(Y_2)).$$
 (5.13)

This holds for any random variables Y_1 and Y_2 (with finite variance). If Y_1 and Y_2 are negatively correlated, i.e. $Cov(Y_1, Y_2) < 0$, we obtain a variance reduction since, using (5.12) with the plus sign,

$$\operatorname{Var}(\widehat{Y}) = \frac{1}{4} \operatorname{Var}(Y_1 + Y_2) \stackrel{(5.12)}{=} \frac{1}{4} \left(\operatorname{Var}(Y_1) + \operatorname{Var}(Y_2) + 2 \operatorname{Cov}(Y_1, Y_2) \right) \\ < \frac{1}{4} \left(\operatorname{Var}(Y_1) + \operatorname{Var}(Y_2) \right) = \frac{1}{2} \operatorname{Var}(Y).$$

In fact, even if the covariance is positive, the variance is potentially reduced since by (5.13), it follows that

$$\operatorname{Var}(\widehat{Y}) = \frac{1}{4} \left(\operatorname{Var}(Y_1) + \operatorname{Var}(Y_2) + 2\operatorname{Cov}(Y_1, Y_2) \right)$$
$$\leq \frac{1}{2} \left(\operatorname{Var}(Y_1) + \operatorname{Var}(Y_2) \right) = \operatorname{Var}(Y).$$

Example. Consider the variable θ created from the standard normally distributed variable *Z* and let θ^- be computed as θ but from -Z such that $Var(\theta) = Var(\theta^-)$. For instance, θ and θ^- may be the payoff function of an option associated to the stock prices S_i and S_i^- , respectively, defined through the Euler-Maruyama approximation of $dS = S(rdt + \sigma dW)$,

$$S_{i+1} = S_i(1 + rh + \sigma Z\sqrt{h}), \quad S_{i+1}^- = S_i^-(1 + rh - \sigma Z\sqrt{h})$$

for i = 0, ..., n - 1, where $Z \sim N(0, 1)$. Then the payoff of a put option using antithetic variables is given by

$$\widehat{\theta} = \frac{1}{2} \big((K - S_n)^+ + (K - S_n^-)^+ \big).$$

Figure 5.5 illustrates the relative error of the put option price computed from the Euler-Maruyama scheme with and without antithetic variables. We observe a significant reduction of the error even after a small number of simulations.

• **Control variates.** We want to estimate $\theta = EY$, where $Y = \phi(X)$ and X is some random variable. Suppose that there exists another random variable Z (the so-called control variate) such that the expectation $\tilde{\theta} = EZ$ can be computed explicitly (or easily)



0.1 0.08 Relative error 0.06 Euler-Maruyama 0.04 0.02 antithetic 0 2000 3000 4000 5000 0 1000 Number of simulations

and that it is closely related to $\theta = EY$. Thus, we expect that the covariance Cov(Y, Z) is close to its maximal value, Var(Y); see (5.13). We claim that if $Cov(Y, Z) \neq 0$, the new random variable

$$\widehat{Y} = Y + \beta(Z - EZ)$$

has a smaller variance than \hat{Y} for a suitable parameter β . Note that this gives an unbiased estimator since $E\hat{Y} = EY$. To show the claim, we compute the variance, using Var(EZ) = 0 and (5.12) with the plus sign:

$$Var(\hat{Y}) = Var(Y) + Var(\beta Z) + 2 Cov(Y, \beta Z)$$

= Var(Y) + \beta^2 Var(Z) + 2\beta Cov(Y, Z)
= Var(Y) + Var(Z) \left(\beta + \frac{Cov(Y, Z)}{Var(Z)}\right)^2 - \frac{Cov(Y, Z)^2}{Var(Z)}

Choosing the minimizing value $\beta = -\operatorname{Cov}(Y, Z) / \operatorname{Var}(Z)$, it follows that

$$\operatorname{Var}(\widehat{Y}) = \operatorname{Var}(Y) - \frac{\operatorname{Cov}(Y, Z)^2}{\operatorname{Var}(Z)} < \operatorname{Var}(Y).$$

In concrete cases, the computation of β may be not obvious since the covariance is usually not known explicitly. A way out is to replace the covariance Cov(Y, Z) and the variance Var(Z) by the values coming from a first Monte-Carlo simulation.

Example. Consider an arithmetic-average fixed strike Asian call option with the payoff $Y = (\overline{S} - K)^+$, where $\overline{S} = (1/n) \sum_{i=1}^n S(t_i)$. An easy choice for Z is the stock itself, $Z = S_T$, since Y and Z are positively correlated. Another choice is the payoff of a European call option, $Z = (S_T - K)^+$, since its expectation is known exactly from the Black-Scholes formula. An even better choice is $Z = [(\prod_{i=1}^n S(t_i))^{1/n} - K]^+$, which is the payoff of an Asian option with discrete geometric average, since its expectation is also known exactly. As this choice incorporates all discrete values $S(t_i)$ of the underlying, we expect that the correlation with Y is rather high and the reduction

of the variance is significant.

5.7 Multilevel Monte-Carlo method

The multilevel Monte-Carlo method uses a number of levels of resolution with the aim to reduce the computational cost from $O(\varepsilon^{-3})$ (see the beginning of Section 5.6) to $O(\varepsilon^{-2}(\log \varepsilon)^2)$. It can be seen as a generalization of the control variate technique. Indeed, let X_1 be some sample computed from a fine grid and X_0 be a sample computed from a coarse grid. We want to estimate EX₁. As X_0 is much cheaper to simulate than X_1 , we write

$$EX_1 = EX_0 + E(X_1 - X_0),$$

where EX_0 is the control variate and $E(X_1 - X_0)$ is a correction.

To extend this idea to *L* grids, we proceed as in [23, Section 8.6]. Let Y_{ℓ} for $\ell = 0, ..., L$ be Euler-Maruyama approximations of the random variable X_T , being, for instance, the payoff of an option at time *T*. We introduce the time step sizes $h_{\ell} = \kappa^{-\ell}T$ on the interval [0, T], where $\kappa \in \{2, 3, ...\}$ is some number. Then $\ell = 0$ corresponds to the coarest grid and $\ell = L$ to the finest grid. The multilevel Monte-Carlo method is based on the decomposition

$$E(\phi(Y_L)) = E(\phi(Y_0)) + \sum_{\ell=1}^{L} E(\phi(Y_\ell) - \phi(Y_{\ell-1})).$$

The first term on the right-hand side is the lowest level expectation, and the second term contains the corrections.

We need to find estimators for these expectations. For this, let $Y_{\ell}^{(k)}$ for $k = 1, ..., m_{\ell}$ be independent identically distributed samples of Y_{ℓ} , $\ell = 0, ..., L$. An estimator for $E(\phi(Y_0))$ is the average

$$\mu_0 = \frac{1}{m_0} \sum_{k=1}^{m_0} \phi(Y_0^{(k)}),$$

and the correction at level ℓ can be estimated by

$$\mu_{\ell} = \frac{1}{m_{\ell}} \sum_{k=1}^{m_{\ell}} \left(\phi(Y_{\ell}^{(k)}) - \phi(Y_{\ell-1}^{(k)}) \right), \quad \ell = 1, \dots, L.$$

Then we can estimate $E(\phi(X_T))$ by $\widehat{\mu} = \sum_{\ell=0}^{L} \mu_{\ell}$.

We divide the error $E(\phi(X_T)) - \hat{\mu} = \eta_{WD} + \eta_{MC}$ into two parts,

the weak discretization error

$$\eta_{WD} = E(\phi(X_T)) - E(\phi(Y_L))$$
, and

the multilevel Monte-Carlo error,

$$\eta_{\rm MC} = {\rm E}(\phi(Y_L)) - \widehat{\mu}$$

The error η_{WD} is of order $O(h_L)$ since the Euler-Maruyama scheme has the weak convergence order one. To achieve $\eta_{WD} = O(\varepsilon)$, we identify $O(h_L) = O(\kappa^{-L}T) = O(\varepsilon)$, i.e., we need $\kappa^{-L}T \leq \varepsilon$ (up to some factor). This is the case if $L \geq \log(T\varepsilon^{-1}) / \log \kappa$.

The error η_{MC} can be expanded according to

$$\eta_{\rm MC} = {\rm E}(\phi(Y_0)) - \mu_0 + \sum_{\ell=1}^L \left({\rm E}[\phi(Y_\ell) - \phi(Y_{\ell-1})] - \mu_\ell \right).$$

In order to achieve $\eta_{MC} = O(\varepsilon)$, we compute μ_0 , which estimates $E(\phi(Y_0))$, with a large number m_0 of samples, while the corrections are calculated with smaller sample numbers. Since μ_0 is related to the coarsest grid, the samples are cheap to compute. The samples on the finer grids are more costly, but less samples are sufficient. In contrast, a standard Monte-Carlo simulation needs a large number of samples on the finest grid. The number m_ℓ of samples can be chosen in such a way that $Var(\hat{\mu}) = O(\varepsilon^2)$ (use $m_0 \ge \varepsilon^{-2}$ and $m_\ell \ge \varepsilon^{-2} Lh_\ell$; see [23, Lemma 8.46]). This guarantees that $\eta_{MC} = O(\varepsilon)$ (use the Chebychev inequality).

We still need to determine the computational cost of finding $\hat{\mu}$. We measure the cost as the number of steps needed in the numerical method to compute $\hat{\mu}$. Using $m_{\ell} = O(\varepsilon^{-2}Lh_{\ell})$ and $L = O(\log(T\varepsilon^{-1})) = O(\log\varepsilon)$, we find that

$$\cot = \sum_{\ell=0}^{L} m_{\ell} h_{\ell}^{-1} = \sum_{\ell=0}^{L} O(\varepsilon^{-2} L h_{\ell}) h_{\ell}^{-1} = O(L^{2} \varepsilon^{-2}) = O\left(\varepsilon^{-2} (\log \varepsilon)^{2}\right).$$

Therefore, the multilevel Monte-Carlo method reduces the computational cost by almost one order of magnitude.

5.8 Application: Pricing Asian options in the Heston model

We want to combine some techniques presented in the previous sections to simulate an Asian option in the Heston model. We proceed as in [16, Section 5.5]. The goal is to determine the premium of an arithmetic-average floating-strike Asian call option with payoff

$$V_T(S_T) = \left(S_T - \frac{1}{T}\int_0^T S_\tau d\tau\right)^+,$$

where the price of the underlying S_{τ} is computed from the Heston model

$$dS_t = r_t S_t dt + \sigma_t S_t dW_1(t),$$

$$d\sigma_t^2 = \kappa (\theta - \sigma_t^2) dt + \nu \sigma_t dW_2(t), \quad 0 < t < T,$$

where κ , θ , ν are positive numbers, the interest rate r_t is time-dependent but given, and the Wiener process (W_1, W_2) is $N(0, \Sigma)$ -distributed with covariance matrix

$$\Sigma = \begin{pmatrix} 1 &
ho \\
ho & 1 \end{pmatrix}$$
 .

We have explained the computation of a two-dimensional Wiener process in the example on page 72. With the independent one-dimensional standard normally distributed random variables Z_1 and Z_2 , the vector

$$\binom{W_1}{W_2} = \binom{Z_1}{\rho Z_1 + \sqrt{1 - \rho^2} Z_2}$$

is $N(0, \Sigma)$ -distributed. The option price is estimated according to

$$\widehat{V}_{0} = \exp\left(-\int_{0}^{T} r_{t} dt\right) \frac{1}{m} \sum_{k=1}^{m} \left(S_{T}^{(k)} - \frac{1}{n} \sum_{i=1}^{n} S_{t_{i}}^{(k)}\right)^{+},$$

where $S_t^{(k)}$ are the samples of S_t , *m* is the number of Monte-Carlo simulations and *n* is the number of time steps. This gives the following algorithm:

► Calculate the increments of the two-dimensional Wiener process from

$$\triangle W_1^{(k)} = Z_1^{(k)}\sqrt{h}, \quad \triangle W_2^{(k)} = \rho Z_1^{(k)}\sqrt{h} + \sqrt{1-\rho^2}Z_2^{(k)}\sqrt{h}.$$

Solve the Heston model using the Euler-Maruyama scheme:

$$(\sigma_{i+1}^{(k)})^2 = (\sigma_i^{(k)})^2 + \kappa \left(\theta - (\sigma_i^{(k)})^2\right)h + \nu \sigma_i^{(k)} \triangle W_2^{(k)},$$

$$S_{i+1}^{(k)} = S_i^{(k)} \left(1 + r_{t_i}h + \sigma_i^{(k)} \triangle W_1^{(k)}\right), \quad i = 0, \dots, n-1.$$

• Compute the approximation of the option price:

$$\widehat{V}_{0} = \exp\left(-\int_{0}^{T} r_{t} dt\right) \frac{1}{m} \sum_{k=1}^{m} \left(S_{T}^{(k)} - \overline{S}^{(k)}\right)^{+}, \quad \overline{S}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} S_{t_{i}}^{(k)}.$$

An alternative is the use of the Milstein scheme with the functions

$$a(Y_i, t_i) = \begin{pmatrix} r_{t_i} S_i \\ \kappa(\theta - \sigma_i^2) \end{pmatrix}, \quad b(Y_i, t_i) = \begin{pmatrix} \sigma_i S_i & 0 \\ 0 & \nu \sigma_i \end{pmatrix}.$$

As detailed in Section 5.5, the Milstein scheme needs the derivatives of *b* with respect to the first variable, which is, in the present case, the vector $Y_i = (S_i, \sigma_i^2)$. It follows that

$$\frac{\partial b_{11}}{\partial S_i} = \sigma_i, \quad \frac{\partial b_{11}}{\partial \sigma_i^2} = \frac{S_i}{2\sigma_i}, \\ \frac{\partial b_{22}}{\partial S_i} = 0, \quad \frac{\partial b_{22}}{\partial \sigma_i^2} = \frac{\nu}{2\sigma_i}.$$

A computation shows that

$$S_{i+1} = S_i(1 + r_{t_i}h + \sigma_i \triangle W_1(t_i)) + \frac{1}{2}\sigma_i^2 S_i((\triangle W_1(t_i))^2 - h) + \frac{\nu}{2}S_i I_{21},$$

$m \setminus \rho$	-0.9	-0.5	-0.2	0.2	0.5	0.9
10,000	9.607	9.623	9.630	9.631	9.625	9.612
50,000	9.734	9.818	9.881	9.968	10.035	10.132
100,000	9.740	9.817	9.876	9.956	10.015	10.093
500,000	9.856	9.946	10.015	10.108	10.179	10.279
2,000,000	9.793	9.881	9.947	10.039	10.108	10.204
5,000,000	9.810	9.898	9.964	10.054	10.122	10.216

Table 3: Prices of the Asian option for various values of the correlation ρ (horizontal axis) and the number *m* of Monte-Carlo simulations (vertical axis).

$$\sigma_{i+1}^2 = \sigma_i^2 + \kappa(\theta - \sigma_i^2)h + \nu\sigma_i \triangle W_2(t_i) + \frac{\nu^2}{2} \big((\triangle W_2(t_i))^2 - h \big).$$

The integral I_{21} can be approximated by solving the stochastic differential equation from Lemma 5.14. More precisely, if $Z_0 = (Z_0^{(1)}, Z_0^{(2)}) = (0, 0)$, we iterate

$$Z_{k+1}=Z_k+egin{pmatrix} Z_k^{(2)}&0\0&1 \end{pmatrix} riangle W_k,\quad k=0,\ldots,N-1,$$

and obtain $I_{21} = Z_N^{(1)}$. If $N \ge 1/h$, this scheme is weakly converging with order one.

We compute the value of the Asian option with the parameters

$$S_0 = 80, \quad T = 1, \quad \kappa = 3, \quad \theta = 0.3, \quad \nu = 0.5,$$

the time-dependent interest rate $r_t = 0.01(\sin(2\pi t) + t) + 0.03$, and the time step is h = 0.01. In Table 3, we illustrate the prices V_0 of the Asian option as a function of the correlation parameter ρ and the number m of Monte-Carlo simulations from the Euler-Maruyama scheme. The dependency on ρ is rather weak; the premium increases with the correlation. Furthermore, we observe that the number of Monte-Carlo simulations need to be very large to obtain accurate results. This motivates the introduction of variance-reduction techniques as explained in Section 5.6.

6 Finite-difference methods

Although the price of European options can be computed explicitly from the Black-Scholes formulas, this generally does not hold true for exotic options like Asian options. Since Monte-Carlo simulations may be too time consuming, another approach is to solve the corresponding partial differential equation if it exists. In this section, we explain the discretization using finite differences, which is in particular suitable in one or two dimensions. Alternative discretization techniques are finite element, finite volume, or discontinuous Galerkin methods, which are not discussed in this course (since there exist specialized courses for these topics). We follow the approach of [16, Section 6.2].

6.1 The θ -method

The idea of the finite-difference method is to replace derivatives by finite differences. To explain this technique, we first consider the diffusion equation

$$\partial_t u - u_{xx} = 0, \quad 0 < t < T, \quad u(x,0) = u_0(x), \quad x \in \mathbb{R},$$
(6.1)

which follows from the Black-Scholes equation after transformation (see Section 3.1). This problem can be solved explicitly, but we consider it as a prototype example to highlight the ideas. More complex equations will be considered below.

For the numerical approximation, we restrict the problem to a bounded interval and prescribe

$$u(-L,t) = b_1(t), \quad u(L,t) = b_2(t), \quad 0 < t < T,$$

where b_1 and b_2 approximate the solution at -L and L, respectively, and L > 0 is so large that the restriction to the finite interval does not change the solution significantly. Clearly, this is a rather vague statement, and a precise analysis has to be made for the problem at hand.

Let *m*, *n* \in \mathbb{N} and set *h* = 2*L*/*n* and τ = *T*/*m*. We introduce the grid (*x*_{*i*}, *t*_{*j*}) with

$$x_i = -L + ih$$
, $t_j = j\tau$ for $i = 0, ..., n$, $j = 0, ..., m$.

By Taylor expansion around *x* (assuming that *u* is sufficiently smooth), we find that

$$u(x \pm h, t) = u(x, t) \pm u_x(x, t)h + \frac{1}{2}u_{xx}(x, t)h^2 \pm \frac{1}{6}u_{xxx}(x, t)h^3 + O(h^4).$$

Adding both expressions, the terms of order *h* and h^3 cancel and we end up, after division by h^2 , with

$$\frac{1}{h^2}(u(x+h,t)-2u(x,t)+u(x-h,t)) = u_{xx}(x,t)+O(h^2).$$
(6.2)

We replace u_{xx} in (6.1) by these approximations to find that

$$\partial_t u(x,t) = \frac{1}{h^2} (u(x+h,t) - 2u(x,t) + u(x-h,t)) + O(h^2),$$

$$\partial_t u(x,t+\tau) = \frac{1}{h^2} (u(x+h,t+\tau) - 2u(x,t+\tau) + u(x-h,t+\tau)) + O(h^2).$$

The time derivative $\partial_t u$ can be discretized in different ways,

$$\partial_t u(x,t) = \frac{1}{\tau} \big(u(x,t+\tau) - u(x,t) \big) + O(\tau),$$

$$\partial_t u(x,t+\tau) = \frac{1}{\tau} \big(u(x,t+\tau) - u(x,t) \big) + O(\tau).$$

We combine the previous approximations:

$$\begin{split} &\frac{1}{\tau} \big(u(x,t+\tau) - u(x,t) \big) = \frac{1}{h^2} \big(u(x+h,t) - 2u(x,t) + u(x-h,t) \big) + O(h^2 + \tau), \\ &\frac{1}{\tau} \big(u(x,t+\tau) - u(x,t) \big) = \frac{1}{h^2} \big(u(x+h,t+\tau) - 2u(x,t+\tau) + u(x-h,t+\tau) \big) \\ &+ O(\tau+h^2). \end{split}$$

Multiplying the first equation by $1 - \theta$, the second equation by θ (where $\theta \in [0, 1]$) and adding both equations, it follows that

$$\frac{1}{\tau} (u(x,t+\tau) - u(x,t)) = \frac{1-\theta}{h^2} (u(x+h,t) - 2u(x,t) + u(x-h,t))
+ \frac{\theta}{h^2} (u(x+h,t+\tau) - 2u(x,t+\tau) + u(x-h,t+\tau)) (6.3)
+ O(\tau+h^2).$$

Equation (6.3) motivates the following scheme for the approximations w_i^j of $u(x_i, t_i)$:

$$\frac{1}{\tau}(w_i^{j+1} - w_i^j) = \frac{1 - \theta}{h^2}(w_{i+1}^j - 2w_i^j + w_{i-1}^j) + \frac{\theta}{h^2}(w_{i+1}^{j+1} - 2w_i^{j+1} + w_{i-1}^{j+1})$$
(6.4)

for i = 1, ..., n - 1 and j = 0, ..., m - 1. This scheme is called the θ -method. Some values of θ are of special interest:

- $\theta = 0$: We can solve w_i^{j+1} for i = 1, ..., n-1 directly from w_i^j . This approximation is called the *explicit Euler scheme*.
- $\theta = 1$: The values for w_i^{j+1} can only be obtained by solving a linear system. This method is called the *implicit Euler scheme*. Note that the solution of a linear system is needed for any $\theta > 0$.

► $\theta = 1/2$: The corresponding discretization is called the *Crank-Nicolson scheme*. It is again an implicit scheme but it has a better convergence rate with respect to time than the implicit Euler method, as shown below.

We wish to formulate (6.4) as a matrix problem. For this, we reformulate (6.4), setting $\gamma = \tau / h^2$:

$$\begin{split} &-\gamma\theta w_{i+1}^{j+1} + (2\gamma\theta+1)w_i^{j+1} - \gamma\theta w_{i-1}^{j+1} \\ &= \gamma(1-\theta)w_{i+1}^j - \big(2\gamma(1-\theta)-1\big)w_i^j + \gamma(1-\theta)w_{i-1}^j. \end{split}$$

The initial and boundary conditions become

$$w_0^j = b_1(t_j), \quad w_n^j = b_2(t_j), \quad w_i^0 = u_0(x_i)$$

for j = 1, ..., m and i = 1, ..., n - 1. These equations can be written in matrix form as

$$Aw^{j+1} = Bw^j + d^j, \quad j = 0, \dots, m-1,$$
 (6.5)

for the unknowns $w^j = (w_1^j, \dots, w_{n-1}^j)^\top$. The vector $d^j \in \mathbb{R}^{n-1}$ is given by

$$d^{j} = \begin{pmatrix} \gamma(1-\theta)b_{1}(t_{j}) + \gamma\theta b_{1}(t_{j+1}) \\ 0 \\ \vdots \\ 0 \\ \gamma(1-\theta)b_{2}(t_{j}) + \gamma\theta b_{2}(t_{j+1}) \end{pmatrix}$$

The tridiagonal matrices $A, B \in \mathbb{R}^{(n-1) \times (n-1)}$ are defined by

$$A = \operatorname{diag} \left(-\gamma \theta, 2\gamma \theta + 1, -\gamma \theta \right),$$

$$B = \operatorname{diag} \left(\gamma (1 - \theta), -2\gamma (1 - \theta) + 1, \gamma (1 - \theta) \right).$$

Here, we have used the notation

diag
$$(a, b, c) = \begin{pmatrix} b & c & 0 & \cdots & 0 \\ a & b & c & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & a & b & c \\ 0 & \cdots & 0 & & a & b \end{pmatrix}.$$

Interestingly, linear systems with tridiagonal matrices can be solved efficiently using the LU decomposition in a lower triangular matrix *L* and an upper triangular matrix *U*. Since *A* is tridiagonal, we can make the ansatz

$$A = L \cdot U$$
 with $L = \operatorname{diag}(\ell, 1, 0), U = \operatorname{diag}(0, d, u),$

where $\ell = (\ell_i)$, 1, $d = (d_i)$, and $u = (u_i)$ are vectors in \mathbb{R}^{n-1} . By inspection, the coefficients d_i and ℓ_i are given recursively by

$$d_1 = 2\gamma\theta + 1,$$

for $i = 1, ..., n - 1$: $\ell_i = -\gamma\theta/d_i, \quad d_{i+1} = 2\gamma\theta + 1 + \ell_i/(\gamma\theta).$

Then the matrix equation (6.5), $LUw^{j+1} = Bw^j + d^j$, can be formulated in terms of *L* and *U* as

$$Lz = Bw^j + d^j, \quad Uw^{j+1} = z,$$

and these linear systems can be easily solved by forward and backward iteration.

There are two mathematical questions that we need to investigate:

- ▶ Has the linear system (6.5) a unique solution?
- ▶ Does the solution converge to a solution to (6.1) when $(h, \tau) \rightarrow 0$?

The first question is not difficult to answer. The answer is affirmative when *A* is invertible. To show this, we use the Gerschgorin circle theorem, which provides a bound of the spectrum of a square matrix.

Theorem 6.1 (Gerschgorin). Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ and let $r_i = \sum_{j=1, j \neq i}^n |a_{ij}|$. Then all eigenvalues λ of A lie within the union of the balls around a_{ii} with radius r_i , i.e.

$$\lambda \in \bigcup_{i=1}^n B_{r_i}(a_{ii}) := \bigcup_{i=1}^n \left\{ z \in \mathbb{C} : |z - a_{ii}| \le r_i \right\}$$

The matrix *A* is strictly diagonally dominant, i.e.

$$|a_{ii}|=2\gamma heta+1>2\gamma heta\geq\sum_{j
eq i}|a_{ij}|=r_i.$$

This shows that $0 \notin B_{r_i}(a_{ii})$ (see Figure 6.1) and consequently, $\lambda = 0$ cannot be an eigenvalue. This implies that *A* is invertible and the linear system (6.5) can be uniquely solved.



The answer to the second question on the convergence is more involved. The goal is to show that $w^j - u(t_j)$ converges in some norm with order $O(\tau + h^2)$. The idea is to prove concistency and stability and then to deduce convergence.

To define *consistency*, we set $u_i^j = u(x_i, t_j)$ and

$$L(u_i^j) = \frac{1}{\tau}(u_i^{j+1} - u_i^j) - \frac{1 - \theta}{h^2}(u_{i+1}^j - 2u_i^j + u_{i-1}^j) - \frac{\theta}{h^2}(u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}).$$

This means that we have inserted the solution at the grid points in the scheme. We expect from Taylor expansion that $L(u_i^j)$ is of order $O(\tau + h^2)$. This is confirmed, and refined, by the following lemma.

Lemma 6.2 (Consistency). Let u_{tt} be continuous. Then

$$L(u_i^j) = \begin{cases} O(\tau + h^2) & \text{if } \theta \neq 1/2, \\ O(\tau^2 + h^2) & \text{if } \theta = 1/2. \end{cases}$$

Denoting by $L(u^j)$ the vector of all $L(u_i^j)$, we can write $L(u^j) = Au^{j+1} - Bu^j - d^j$. It holds that $L(w^j) = 0$ since w^j is the solution to the numerical scheme (6.5), but $e^j := L(u^j)$ generally does not vanish, and we call e^j the truncation error. The lemma says that the truncation error is at least of order $O(\tau + h^2)$ and improves to $O(\tau^2 + h^2)$ for the Crank-Nicolson scheme.

Proof. We already know from the Taylor expansion at the beginning of the section that

$$\frac{1-\theta}{h^2} (u_{i+1}^j - 2u_i^j + u_{i-1}^j) + \frac{\theta}{h^2} (u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1})
= (1-\theta)u_{xx}(x_i, t_j) + \theta u_{xx}(x_i, t_{j+1}) + O(h^2)
= u_{xx}(x_i, t_j) + \theta (u_{xx}(x_i, t_{j+1}) - u_{xx}(x_i, t_j)) + O(h^2)
= u_{xx}(x_i, t_j) + \theta \tau u_{xxt}(x_i, t_j) + O(\tau^2 + h^2).$$

Note that by assumption, $u_{xxt} = (u_{xx})_t = u_{tt}$ is continuous. Furthermore, we have

$$\frac{1}{\tau}(u_i^{j+1} - u_i^j) = u_t(x_i, t_j) + \frac{\tau}{2}u_{tt}(x_i, t_i) + O(\tau^2).$$

Combining these expansions and recalling that $u_{xxt} = u_{tt}$, it follows that

$$\begin{split} L(u_i^j) &= u_t(x_i, t_j) + \frac{\tau}{2} u_{tt}(x_i, t_i) - u_{xx}(x_i, t_j) - \theta \tau u_{xxt}(x_i, t_i) + O(\tau^2 + h^2) \\ &= \frac{\tau}{2} \big(u_{tt}(x_i, t_i) - 2\theta \tau u_{xxt}(x_i, t_i) \big) + O(\tau^2 + h^2) \\ &= \frac{\tau}{2} (1 - 2\theta) u_{tt}(x_i, t_i) + O(\tau^2 + h^2). \end{split}$$

If $\theta = 1/2$, the first term on the right-hand side vanishes and $L(u_i^j) = O(\tau^2 + h^2)$. Otherwise, we observe that this term is of order τ , yielding $L(u_i^j) = O(\tau + h^2)$.

By *stability*, we mean that the spectral radius of $A^{-1}B$ is smaller than one. We recall that the spectral radius $\rho(A)$ of a matrix $A = (A_{ij})$ is defined by

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

The meaning of the notion of stability is that the matrix in $w^{j+1} = A^{-1}Bw^j + A^{-1}d^j$ (which is a reformulation of (6.5)) leads to a stable behavior in the sense that previous errors are damped.

Lemma 6.3 (Stability). It holds $\rho(A^{-1}B) < 1$ if $0 < \gamma \le 1/(2-4\theta)$ for $0 \le \theta < 1/2$ and $\gamma > 0$ if $1/2 \le \theta \le 1$, where $\gamma = \tau/h^2$.

If $1/2 \le \theta \le 1$, we say that the scheme is *unconditionally stable* since no condition on γ is required. If $0 \le \theta < 1/2$, the bound on γ gives a restriction on the choice of the time step when the spatial grid size is given.

Proof. We write $A = I + \gamma \theta F$ and $B = I - \gamma (1 - \theta) F$, where $I \in \mathbb{R}^{(n-1) \times (n-1)}$ is the unit matrix and $F = \text{diag}(-1, 2, -1) \in \mathbb{R}^{(n-1) \times (n-1)}$. The eigenvalues of F can be computed explicitly:

$$\lambda_k = 4\sin^2\left(\frac{k\pi}{2n}\right), \quad k = 1, \dots, n-1.$$

Thus, the eigenvalues of A^{-1} are given by $\mu_k = 1/(1 + \gamma \theta \lambda_k)$. The matrices A and B are related through $B = I/\theta - (1/\theta - 1)A$. Multiplying this expression by A^{-1} yields

$$A^{-1}B = \frac{1}{\theta}A^{-1} - \left(\frac{1}{\theta} - 1\right)I.$$
 (6.6)

Hence, the eigenvalues of $A^{-1}B$ are $\mu_k/\theta - (1/\theta - 1)$ and the condition $\rho(A^{-1}B) < 1$ holds if and only if for all k = 1, ..., n - 1,

$$1 > \left| \frac{\mu_k}{\theta} - \left(\frac{1}{\theta} - 1 \right) \right| = \left| \frac{1}{\theta} \frac{1}{1 + \gamma \theta \lambda_k} - \frac{1}{\theta} + 1 \right| = \left| -\frac{\gamma \lambda_k}{1 + \gamma \theta \lambda_k} + 1 \right|.$$

This is equivalent to

$$-1 < -rac{\gamma\lambda_k}{1+\gamma heta\lambda_k} + 1 < 1 \quad ext{or} \quad 0 < rac{\gamma\lambda_k}{1+\gamma heta\lambda_k} < 2.$$

Inserting the expression for λ_k , we find that the latter inequality is equivalent to

$$(2-4\theta)\gamma\sin^2\left(\frac{k\pi}{2n}\right) < 1$$

This is always satisfied if $\theta \ge 1/2$. When $\theta < 1/2$, we need the condition $(2 - 4\theta)\gamma \le 1$ or $\gamma \le 1/(2 - 4\theta)$.

Consistency and stability imply convergence of the scheme, as shown in the following theorem.

Theorem 6.4 (Convergence). Let $u_i^j = u(x_i, t_j)$ be the solution to (6.1) and let $w^j = (w_1^j, \ldots, w_{n-1}^j)^\top$ be the finite-difference solution to (6.4) with initial datum $w_i^0 = u_0(x_i)$. Set $u^j = (u_1^j, \ldots, u_{n-1}^j)^\top$. We assume that $u_{tt} \in C^0$ and $\gamma \leq 1/(2-4\theta)$ if $\theta < 1/2$. Then there exists C > 0 depending on A, B, and u and its derivatives such that for all (sufficiently *small*) τ , h > 0,

$$\max_{j=1,\dots,m} \|w^j - u^j\|_2 \le C(\tau + h^2).$$

If $\theta = 1/2$, we can replace the right-hand side by $C(\tau^2 + h^2)$.

The norm $\|\cdot\|_2$ is the Euclidean vector norm defined by $\|v\|_2^2 = \sum_{i=1}^{n-1} v_i^2 / (n-1)$ for $v \in \mathbb{R}^{n-1}$. It induces the matrix norm

$$||A||_2 = \sup_{\|v\|_2 \neq 0} \frac{||Av||_2}{\|v\|_2} \text{ for } A \in \mathbb{R}^{(n-1) \times (n-1)},$$

which is related to the spectral radius by $||A||_2 = \rho(A^{\top}A)^{1/2}$. If *A* is symmetric, it follows that $||A||_2 = \rho(A)$. Actually, because of (6.6), $A^{-1}B$ is symmetric, so $||A^{-1}B||_2 = \rho(A^{-1}B) < 1$.

Proof. The vectors w^j and u^j satisfy the relations

$$Aw^{j} = Bw^{j-1} + d^{j-1}, \quad Au^{j} = Bu^{j-1} + d^{j-1} + e^{j-1}$$

where, by Lemma 6.2, the truncation error e^{j-1} fulfills the estimate

$$|e^{j-1}||_2 \le C \max_{i=1,\dots,n-1} |e_i^{j-1}| \le C_{cons}(\tau+h^2), \quad j=1,\dots,m.$$

We estimate the difference $w^j - u^j = A^{-1}B(w^{j-1} - u^{j-1}) - A^{-1}e^{j-1}$ by recursion and using $w^0 = u^0$:

$$\begin{split} \|w^{j} - u^{j}\|_{2} &\leq \|A^{-1}B\|_{2} \|w^{j-1} - u^{j-1}\|_{2} + \|A^{-1}\|_{2} \|e^{j-1}\|_{2} \\ &\leq \|A^{-1}B\|_{2}^{j} \|w^{0} - u^{0}\|_{2} + \sum_{k=0}^{j-1} \|A^{-1}B\|_{2}^{j+k-1} \|A^{-1}\|_{2} \|e^{k}\|_{2} \\ &\leq \frac{1 - \|A^{-1}B\|_{2}^{j}}{1 - \|A^{-1}B\|_{2}} \|A^{-1}\|_{2} \max_{k=0,\dots,j-1} \|e^{k}\|_{2} \\ &\leq \frac{\|A^{-1}\|_{2}}{1 - \|A^{-1}B\|_{2}} \max_{k=0,\dots,j-1} \|e^{k}\|_{2} \leq C(\tau + h^{2}), \end{split}$$

where $C = C_{\text{cons}} ||A^{-1}||_2 / (1 - ||A^{-1}B||_2)$ and we have used the stability estimate of Lemma 6.3 yielding $||A^{-1}B||_2 < 1$. When $\theta = 1/2$, Lemma 6.2 shows that the convergence actually improves to $\tau^2 + h^2$.

Remark. (Higher-order convergence in *h*.) A higher convergence rate can be achieved by using multipoint formulas. The idea is to approximate u_{xx} up to an order higher than two. Using Taylor expansion, we can show, for instance, that

$$u_{xx}(x) = \frac{1}{h^2} \left(-\frac{1}{12}u(x+2h) + \frac{4}{3}u(x+h) - \frac{5}{2}u(x) + \frac{4}{3}u(x-h) - \frac{1}{12}u(x-2h) \right) + O(h^4).$$

Thus, we may replace the three-point stencil $w_{i+1} - 2w_i + w_{i-1}$ in (6.4) by the five-point stencil

$$\frac{1}{h^2}\left(-\frac{1}{12}w_{i+2}+\frac{4}{3}w_{i+1}-\frac{5}{2}w_i+\frac{4}{3}w_{i-1}-\frac{1}{12}w_{i-2}\right).$$

This yields the convergence rate $O(h^4)$. The disadvantage is that the bandwidth of the matrix is increased, which decreases the efficiency of the solution to the associated linear system. An alternative is to use so-called *higher-order compact schemes*. They keep the three-point stencil but allow for a higher-order convergence rate. The idea is to evaluate the differential equation at the points x_{i-1} , x_i , and x_{i+1} and to add a suitable linear combination. For instance, a fourth-order compact scheme for $u_{xx} = f$ is given by

$$\frac{1}{h^2}(w_{i+1} - 2w_i + w_{i-1}) = \frac{1}{12}(f(x_{i+1}) + 10f(x_i) + f(x_{i-1})).$$

The price to pay is the increased number of evaluations of the right-hand side.

Remark. (Higher-order convergence in τ .) The θ -scheme (6.4) can be written as

$$w^{j+1} = w^j + \tau ((1-\theta)L(w^j) + \theta L(w^{j+1})),$$

where $L(w^j)$ is the finite-difference approximation of $u_{xx}(t_j)$. Higher-order approximations can be derived from the general ansatz

$$w^{j+1} = w^j + \tau \sum_{\ell=1}^{s} b_\ell k_\ell, \quad k_\ell = L \left(w^j + \tau \sum_{m=1}^{s} a_{\ell m} k_m \right),$$

where the weights b_{ℓ} satisfy $\sum_{\ell=1}^{s} b_{\ell} = 1$. This gives the family of *Runge-Kutta methods*, which may be explicit or implicit, depending on the choice of the coefficients $a_{\ell m}$. In particular, if $a_{mm} \neq 0$, we need to solve a system of nonlinear equations. By decoupling the equations, we still keep the nonlinearity, but only need to solve nonlinear equations instead of a nonlinear system, whose numerical solution is more involved. For this, we assume that $a_{\ell m} = 0$ for $m > \ell$ and write the scheme as

$$k_{\ell} = L\left(w^{j} + \tau \sum_{m=1}^{\ell-1} a_{\ell m} k_{m} + \tau a_{\ell \ell} k_{\ell}\right) =: F(k_{\ell}).$$

The Newton method for $k_{\ell} - F(k_{\ell}) = 0$ becomes

$$\left(I - \frac{\partial F}{\partial k_{\ell}}(k_{\ell}^{i})\right)\left(k_{\ell}^{i+1} - k_{\ell}^{i}\right) = F(k_{\ell}^{i}) - k_{\ell}^{i},$$

where the superindex *i* denotes the iteration number. Approximating the derivative $(\partial F / \partial k_{\ell})$ $(k_{\ell}) = \tau a_{\ell\ell} (\partial L / \partial k_{\ell}) (k_{\ell})$ by $\tau a_{\ell\ell} (\partial L / \partial k_{\ell}) (w^{j})$, we obtain

$$\left(I - \tau a_{\ell\ell} \frac{\partial L}{\partial k_{\ell}}(w^j)\right) \left(k_{\ell}^{i+1} - k_{\ell}^i\right) = L\left(w^j + \tau \sum_{m=1}^{\ell-1} a_{\ell m} k_m^i + \tau a_{\ell\ell} k_{\ell}^i\right) - k_{\ell}^i.$$

If we perform only one iteration step, this leads to the ROW (Rosenbrock-Wanner) scheme. The solution of one iteration only requires an LU decomposition which is computationally cheap. This scheme (for s = 2) is implemented in the MATLAB solver ode23s and it is of order two. The solver has nice stability properties, and in particular, it is sensitive to stiff problems where stability conditions are complicated.

6.2 Application: Pricing arithmetic-average floating-strike calls

We discretize the arithmetic-average floating-strike call option from Section 3.7 using finite differences. We recall the equations:

$$\begin{aligned} H_t + \frac{1}{2}\sigma^2 R^2 H_{RR} + (1 - rR)H_R &= 0, \quad R > 0, \ 0 < t < T, \\ H(R,T) &= (1 - R/T)^+, \quad R > 0, \\ H(R,t) &\to 0 \text{ as } R \to \infty, \quad H_t(0,t) + H_R(0,t) = 0, \quad 0 < t < T. \end{aligned}$$

We expect that for sufficiently large values of R, H(R, t) is close to zero. Therefore, we choose some L > 0 and set H(L, t) = 0 to solve the equation in the bounded interval [0, L]. We choose the grid points $R_i = ih$ and $t_j = j\tau$ and set $H_i^j = H(R_i, t_j)$.

The second derivative H_{xx} is discretized by the central finite differences (6.2):

$$H_{RR}(x_i, t_j) = \frac{1}{h^2} (H_{i+1}^j - 2H_i^j + H_{i-1}^j) + O(h^2)$$

The first derivative H_R can be approximated by the one-sided finite differences

$$H_R(x_i, t_j) = \frac{1}{h}(H_{i+1}^j - H_i^j) + O(h)$$
 or $H_R(x_i, t_j) = \frac{1}{h}(H_i^j - H_{i-1}^j) + O(h)$,

but we prefer a second-order discretization to match the discretization order of H_{RR} . Therefore, we choose the central discretization

$$H_R(x_i, t_j) = \frac{1}{2h} (H_{i+1}^j - H_{i-1}^j) + O(h^2),$$

which can be derived from the one-sided approximations by addition. A disadvantage of the central discretization of H_R is its instability in case of large coefficients |1 - rR| (i.e. if the equation is convection dominant). This can be solved by using an upwind discretization, which is based on one-sided approximations but taking care of the sign of 1 - rR, combined with a modification of the diffusion coefficient (Iljin scheme). Since we only wish to present a simple numerical scheme, we do not use such refined approximations. For details, we refer to [31, Section 2.1].

Furthermore, we replace H_t by the explicit Euler discretization. Then the numerical scheme reads as

$$\frac{1}{\tau}(w_i^j - w_i^{j-1}) + \frac{\sigma^2 R_i^2}{2h^2}(w_{i+1}^j - 2w_i^j + w_{i-1}^j) + \frac{1 - rR_i}{2h}(w_{i+1}^j - w_{i-1}^j) = 0$$

for i = 1, ..., n - 1, j = 1, ..., m. The final and boundary conditions can be chosen as follows:

$$w_i^m = (1 - R_i / t_m)^+, \quad w_n^j = 0, \quad \frac{1}{\tau} (w_0^j - w_0^{j-1}) + \frac{1}{h} (w_1^j - w_0^j) = 0.$$

Observe that the discrete boundary condition at R = 0 is only of order one. In order to match the second-order approximation of the differential equation, we aim to find a discretization that is also of order $O(h^2)$. For this, we subtract the expansions

$$4H_1^j = 4H_0^j + 4hH_R(0, t_j) + 2h^2H_{RR}(0, t_j) + O(h^3),$$

$$H_2^j = H_0^j + 2hH_R(0, t_j) + 2h^2H_{RR}(0, t_j) + O(h^3)$$

to find that

$$H_R(0,t_j) = \frac{1}{2h} (4H_1^j - H_2^j - 3H_0^j) + O(h^2).$$

Thus, an improved version of the discrete boundary condition at R = 0 is

$$\frac{1}{\tau}(w_0^j - w_0^{j-1}) + \frac{1}{2h}(4w_1^j - w_2^j - 3w_0^j) = 0.$$

This scheme can be written more compactly in matrix form. For this, we set $w^j = (w_0^j, \ldots, w_{n-1}^j)^\top$. Then our explicit scheme reads as

$$(I - \tau M)w^{j} = w^{j-1}$$
, where $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$

and

$$M_{11} = \frac{3}{2h}, \quad M_{12} = \frac{1}{2h}(-4, 1, 0, \dots, 0),$$

$$M_{21} = \left(-\frac{\sigma^2}{2h^2}R_1^2 + \frac{1}{2h}(1 - rR_1)\right)(1, 0, \dots, 0)^{\top},$$

$$M_{22} = \frac{\sigma^2}{2h^2} \begin{pmatrix} R_1^2 & 0 \\ & \ddots \\ 0 & R_{n-1}^2 \end{pmatrix} \times \operatorname{diag}(-1, 2, 1)$$

$$-\frac{1}{2h} \begin{pmatrix} 1 - rR_1^2 & 0 \\ & \ddots \\ 0 & 1 - rR_{n-1}^2 \end{pmatrix} \times \operatorname{diag}(-1, 0, -1).$$

Figure 6.2 shows H(R,t) as a function of time and the variable R = I/S, where *S* is the stock price and $I = \int_0^T S_t dt/T$ is the average price. The option price *V* can be computed from V(S, I, t) = SH(I/S, t). At t = T, we recover the payoff function $H(R, T) = (1 - R/T)^+$.

In contrast to the above mentioned discretization, we have approximated only the direction *R* by using central finite differences and then solved the resulting system of ordinary differential equations by using the MATLAB function ode23s. The technique of discretizing only the "non-time" variable and to obtain a system of differential equations is called the *method of lines*. Since the system matrix *M* is sparse (most of its elements are zero), it is preferable to store only the nonzero elements. This is realized by the MATLAB command sparse.



average floating-strike call option with 9 months to expiration, interest rate r = 0.05, and volatility $\sigma = 0.3$.

Figure 6.2: Premium of an arithmetic-

6.3 Relation with the binomial method

We have shown in Section 4.2 that the price of a European option computed from the binomial method is close to the Black-Scholes price, at least for small time step sizes. In this section, we discuss the relation between the binomial and the finite-difference method, i.e., we show that an explicit Euler finite-difference scheme can be interpreted as a binomial method. We proceed as in [16, Section 6.2.5].

Consider the Black-Scholes equation (3.4) after having performed the transformation $x = \ln(S/K)$ and v(x,t) = V(S,t). We deduce from

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV = 0$$

by using

$$SV_S = S \frac{\partial v}{\partial x} \frac{dx}{dS} = v_x, \quad S^2 V_{SS} = S(SV_S)_S - SV_S = v_{xx} - v_x$$

that

$$v_t + \frac{\sigma^2}{2}v_{xx} + \left(r - \frac{\sigma^2}{2}\right)v_x - rv = 0, \quad x \in \mathbb{R}, \ 0 < t < T.$$

We introduce the grid points $x_i = ih$ and $t_j = \tau j$ and the approximations w_i^j of $v(x_i, t_j)$. Replacing v_t by the explicit Euler scheme and the derivatives v_{xx} , v_x by central finite differences (as in the previous subsection), we obtain the following scheme:

$$\frac{1}{\tau}(w_i^j - w_i^{j-1}) + \frac{\sigma^2}{2h^2}(w_{i+1}^j - 2w_i^j + w_{i-1}^j) + \left(r - \frac{\sigma^2}{2}\right)\frac{1}{2h}(w_{i+1}^j - w_{i-1}^j) - rw_i^{j-1} = 0.$$
(6.7)

This scheme is implicit in the last term. We compute backwards in time, i.e., w_i^j is given and we determine w_i^{j-1} . This can be done explicitly by combining the terms w_i^{j-1}/τ and rw_i^{j-1} , which leads to

$$w_i^{j-1} = \frac{1}{1+r\tau} \left(a_1 w_{i+1}^j + b w_i^j + a_2 w_{i-1}^j \right), \tag{6.8}$$
where
$$a_{1/2} = \frac{\sigma^2 \tau}{2h^2} \mp \frac{\sigma^2 \tau}{4h} \pm \frac{r\tau}{2h}, \quad b = 1 - \frac{\sigma^2 \tau}{h^2}.$$

The identification of the finite-difference scheme with the binomial method is based on the condition

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

We have proved in Lemma 6.3 that the explicit Euler scheme with $\theta = 0$ is stable when $\gamma = \tau/h^2 \le 1/2$. Taking into account the diffusion coefficient $\sigma^2/2$, this inequality becomes $(\sigma^2/2)(\tau/h^2) \le 1/2$. Choosing the maximal value for the time step, we obtain $\sigma^2 \tau/(2h^2) = 1/2$, which is exactly the aforementioned condition. As a byproduct, *b* vanishes and (6.8) can be formulated as

$$w_i^{j-1} = \frac{a_1 w_{i+1}^j + a_2 w_{i-1}^j}{1 + r\tau}, \quad a_{1/2} = \frac{1}{2} \pm \frac{r - \sigma^2/2}{2\sigma} \sqrt{\tau}.$$
(6.9)

Recall the price (4.1) of a European call option in the one-period binomial model:

$$V_i^{j-1} = e^{-r\tau} \left(p^* V_{i+1}^j + (1-p^*) V_{i-1}^j \right), \quad \text{where} \quad p^* = \frac{e^{r\tau} - d}{u - d}. \tag{6.10}$$

The up-state factor u and down-state factor d are chosen as in Section 4.2, namely $u = \exp(\sigma\sqrt{\tau})$ and $d = 1/u = \exp(-\sigma\sqrt{\tau})$. In the proof of Theorem 4.4, we have approximated p^* according to

$$p^* = \frac{\sigma + (r - \sigma^2/2)\sqrt{\tau} + O(\tau)}{2\sigma} + O(\tau) = \frac{1}{2} + \frac{r - \sigma^2/2}{2\sigma}\sqrt{\tau} + O(\tau).$$

Taking into account the definition of $a_{1/2}$, we can write

$$p^* = a_1 + O(\tau), \quad 1 - p^* = a_2 + O(\tau) \quad \text{as } \tau \to 0.$$

Then, using the expansion $\exp(-r\tau) = 1/(1+r\tau) + O(\tau^2)$, formula (6.10) becomes

$$V_i^{j-1} = \frac{1}{1+r\tau} \left(a_1 V_{i+1}^j + a_2 V_{i-1}^j \right) + O(\tau).$$

This corresponds to the finite-difference scheme (6.9) up to the error $O(\tau)$. We summarize these computations.

Proposition 6.5. Let $\sigma^2 \tau / h^2 = 1$, $u = \exp(\sigma \sqrt{\tau})$, and $d = \exp(-\sigma \sqrt{\tau})$. Then the finite-difference scheme (6.7) can be interpreted up to the error $O(\tau)$ as a binomial method.

6.4 Pricing American options

In Section 4.3, we have shown how the price of American options can be determined from the binomial method. The goal of this section is to derive a continuous model for American options. Recall that American and European call options have the same value when no dividend is paid (see Section 2.1). Therefore, we consider American options with continuous dividend payments $\delta > 0$. We proceed similarly as in [16, Chapter 7].

• Free-boundary problems. We claim that there exists a stock price S_f such that it is worth to exercise the put option for $S < S_f$ but one should wait if $S \ge S_f$. First, the bound $S_f < K$ holds. Indeed, otherwise, there exists $S \in [K, S_f]$ such that we excercise the option. But then $P = (K - S)^+ = 0$, and the exercise did not make sense. Second, consider the portfolio $\pi = P + S$, consisting of the put option and the underlying. We should exercise the option as soon as $P = (K - S)^+ = K - S$, since we may invest $\pi = (K - S) + S = K$ until time *T*. However, if $P > (K - S)^+$, we will not exercise the option, since the portfolio has the value $\pi > (K - S)^+ + S \ge K$ before the exercise but $\pi = (K - S) + S = K$ after the exercise. This shows that there exists a price $S_f = S_f(t)$ which distinguishes the two cases. We call $S_f(t)$ a free boundary and summarize:

$$S \le S_f(t):$$
 $P(S,t) = (K-S)^+ = K-S,$
 $S > S_f(t):$ $P(S,t) > (K-S)^+.$

This behavior is illustrated in Figure 6.3. Similarly, there exists for American call options (with dividend payments) a number $S_f = S_f(t) > K$ such that

$$S \ge S_f(t): \quad C(S,t) = (S-K)^+ = S-K,$$

 $S < S_f(t): \quad C(S,t) > (K-S)^+.$

The free boundary S_f complicates the pricing problem, since it needs to be determined together with the option price. Such problems are called *free-boundary problems*.



Figure 6.3: Qualitative behavior of the put price and free boundary S_f .

When we know the free boundary S_f , we can in principle compute the option value. Indeed, we know the option price for $S \leq S_f(t)$, namely P(S,t) = K - S, while for $S > S_f(t)$, the put option fulfills the Black-Scholes equation. The Black-Scholes equation is solved in the interval $[S_f(t), \infty)$, and we need to find suitable boundary conditions. Clearly, $P(S,t) \rightarrow 0$ as $S \rightarrow \infty$, since the option becomes worthless for very large values of the underlying. Assuming that the option price depends continuously on S (otherwise, there are arbitrage opportunities), we have $P(S_f(t), t) = K - S_f(t)$. Similarly, for a call option, we have C(0, t) = 0 and $C(S_f(t), t) = S_f(t) - K$.

These conditions are not sufficient to determine the option price uniquely. The reason is that we also need to determine the free boundary $S_f(t)$, which requires another condition. By an arbitrage argument, one can justify that also the mapping $S \mapsto \partial P/\partial S$ is continuous; see [36, Section 7.4]. Since $P_S = \partial P/\partial S = \partial(K - S)/\partial S = -1$ for $S < S_f(t)$, we infer that $P_S(S, t) = -1$ also at $S = S_f(t)$. In a similar way, we obtain $C_S(S, t) = 1$ at $S = S_f(t)$. We summarize:

Proposition 6.6 (Free-boundary problem for American options). Let $\delta > 0$ be the continuous dividend yield. The value P(S,t) of an American put option is determined by

$$\begin{split} P_t &+ \frac{1}{2}\sigma^2 S^2 P_{SS} + (r-\delta)P_S - rP = 0 \quad \text{for } S > S_f(t), \\ P(S,t) &= K-S \quad \text{for } 0 \le S \le S_f(t), \\ \text{final condition:} \quad P(S,T) &= (K-S)^+, \\ \text{boundary conditions:} \quad \lim_{S \to \infty} P(S,t) &= 0, \ P(S_f(t),t) = K - S_f(t), \ P_S(S_f(t),t) = -1. \end{split}$$

The value C(S, t) *of an American call option is computed from*

$$\begin{aligned} C_t &+ \frac{1}{2}\sigma^2 S^2 C_{SS} + (r - \delta)C_S - rC = 0 \quad for \ 0 < S < S_f(t), \\ C(S,t) &= S - K \quad for \ S \ge S_f(t), \\ final \ condition: \quad C(S,T) &= (S - K)^+, \\ boundary \ conditions: \quad C(0,t) &= 0, \ C(S_f(t),t) = S_f(t) - K, \ C_S(S_f(t),t) = 1. \end{aligned}$$

Remark. The free boundary S_f is the optimal point for the writer of the option if she/he is delta-hedging (i.e. hedging the position by trading the underlying). Therefore, S_f is also called the optimal exercise point. However, it is not the writer but the holder who exercises the option or not. Thus, the writer is exposed to the exercise strategy of the holder. The option price is calculated under the assumption that the holder exercises at the optimal point S_f although the holder may exercise at any point. In the worst case (worst for the writer), the holder exercises at the optimal time, but this has already been priced into the option premium.

• Linear complementarity problems. The solution of the free-boundary problem is challenging since it is unclear how to determine the free boundary. Therefore, we derive another formulation in which the free boundary is eliminated. This leads to so-called complementarity problems. In Section 3, we have derived the Black-Scholes equation by investigating the riskless and self-financing portfolio $\pi = c_1(t)B + c_2(t)S - V(S, t)$, consisting of $c_1(t)$ shares of a bond B, $c_2(t)$ shares of the underlying S, and one short American option with value V(S, t). Choosing $c_2 = V_S$, we obtain

$$d\pi = \left(c_1 r B + \delta S V_S - V_t - \frac{1}{2} \sigma^2 S^2 V_{SS}\right) dt,$$

where $V_t = \frac{\partial V}{\partial t}$ (see Section 3.4). According to the previous remark, the holder of the option may exercise in an optimal way or not. If not, the writer may make a better profit compared to a riskless investment. Therefore, $d\pi \ge r\pi dt$. We deduce that

$$r(c_1B + SV_S - V)dt = r\pi dt \le d\pi = \left(c_1rB + \delta SV_S - V_t - \frac{1}{2}\sigma^2 S^2 V_{SS}\right)dt$$

and hence,

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + (r - \delta)SV_S - rV \le 0, \quad S \in (0, \infty), \ t \in (0, T).$$
(6.11)

We have seen in Proposition 6.6 that equality holds for all $S > S_f(t)$ (in case of put options). We claim that we need to put the inequality sign "<" for all $S < S_f(t)$. To simplify the argument, we assume that there are no dividend payments, $\delta = 0$. The argument for $\delta > 0$ is more complicated and needs some additional bounds for $S_f(t)$; we refer to [21, Theorem 9.4] for details. Now, if $\delta = 0$ and $S < S_f(t)$, the put value is P = K - S. We insert this expression into the Black-Scholes equation:

$$P_t + \frac{1}{2}\sigma^2 S^2 P_{SS} + rSP_S - rP = -rS - r(K - S) = -rK < 0.$$

We infer that if $P(S, t) = (K - S)^+$, we put the less sign "<", and if $P(S, t) > (K - S)^+$, we put the equality sign "=" in (6.11). This means that

$$(P - (K - S)^{+})\left(P_t + \frac{1}{2}\sigma^2 S^2 P_{SS} + rSP_S - rP\right) = 0$$

for all S > 0. A similar argument holds for American call options (with dividend payments). We summarize these statements.

Proposition 6.7 (Linear complementarity problem for American options). *The value* V(S,t) *of an American option is the solution of the system*

$$(V - V_T(S))\left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + (r - \delta)SV_S - rV\right) = 0$$

$$-\left(V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + (r-\delta)SV_S - rV\right) \ge 0, \quad V - V_T(S) \ge 0,$$

together with the final and boundary conditions

put:
$$V(0,t) = K$$
, $\lim_{S \to \infty} V(S,t) = 0$, $V_T(S) = (K-S)^+$,
call: $V(0,t) = 0$, $\lim_{S \to \infty} (V(S,t) - S) = 0$, $V_T(S) = (S-K)^+$.

Remark. (Obstacle problems and variational inequality.) Free-boundary problems are related to obstacle problems, where the solution u to a partial differential equation satisfies the constraint $u \ge f$. The function f can be interpreted as an obstacle, and the solution u lies above this obstacle. For instance, let us look for a solution $u : [-1,1] \rightarrow \mathbb{R}$ to the linear complementarity problem

$$-u'' \ge 0, \quad u - f \ge 0, \quad u''(u - f) = 0 \quad \text{in} (-1, 1)$$
 (6.12)

with the boundary conditions $u(\pm 1) = 0$. Generally, we cannot expect that u is twice differentiable, so we need to interpret $-u'' \ge 0$ in the sense of distributions or in the weak sense. For this, consider the Sobolev space $H^1(-1,1)$, consisting of square integrable functions whose weak derivative is square integrable, and choose a test function $v - f \in K$, where

$$v \in K = \{v \in H^1(-1,1) : v \ge f \text{ in } (-1,1), v(\pm 1) = 0\}$$

This means that we multiply $-u'' \ge 0$ by $v - f \ge 0$, integrate over (-1, 1), add u''(u - f) = 0, and integrate by parts:

$$0 \leq -\int_{-1}^{1} u''(v-f)dx + \int_{-1}^{1} u''(u-f)dx = -\int_{-1}^{1} u''(v-u)dx = \int_{-1}^{1} u'(v-u)'dx.$$

Thus, we are looking for a solution $u \in K$ to the *variational inequality*

$$\int_{-1}^{1} u'(v-u)' dx \ge 0 \quad \text{for all } v \in K.$$

One can show that conversely, if $u \in C^2(-1,1) \cap C^0([-1,1])$ solves the variational inequality, then also the linear complementarity problem (6.12).

• Numerical solution. We wish to solve the linear complementarity problem in Proposition 6.7 numerically using finite differences. For this, we first transform the problem using, as in Section 3, the variables $x = \ln(S/K)$ and $\tau = \sigma^2(T - t)/2$. Then the transformed option price

$$u(x,\tau) = \frac{1}{K} \exp\left(\frac{1}{2}(k_{\delta}-1)x + \frac{1}{4}(k_{\delta}-1)^{2}\tau + k_{0}\tau\right) V(S,t),$$

where $k_{\delta} = 2(r - \delta) / \sigma^2$ for $\delta \ge 0$, solves

$$(u_{\tau} - u_{xx})(u - f) = 0, \quad u_{\tau} - u_{xx} \ge 0, \quad u - f \ge 0$$
 (6.13)

for $x \in \mathbb{R}$ and $t \in (0, T)$. The function

$$f(x,\tau) = \exp\left(\frac{1}{2}(k_{\delta}-1)x + \frac{1}{4}(k_{\delta}-1)^{2}\tau + k_{0}\tau\right)\frac{V_{T}(Ke^{x})}{K}$$
(6.14)

is the transformed constraint. The initial and boundary conditions read as

$$\begin{split} u(x,0) &= f(x,0), \quad x \in \mathbb{R}, \\ \text{put:} \quad \lim_{x \to -\infty} (u(x,\tau) - f(x,\tau)) = 0, \quad \lim_{x \to \infty} u(x,\tau) = 0, \\ \text{call:} \quad \lim_{x \to -\infty} u(x,\tau) = 0, \quad \lim_{x \to \infty} (u(x,\tau) - f(x,\tau)) = 0. \end{split}$$

For the numerical discretization, we restrict ourselves to the interval [-L, L] instead of the whole line \mathbb{R} and require that the boundary conditions are satisfied at $x = \pm L$ instead at $x \to \pm \infty$. Let $x_i = -L + ih$ for i = 0, ..., n and $t_j = j\Delta t$ for j = 0, ..., m with $h = 2L/n, \Delta t = T/m$ be the grid points. We approximate $u_{\tau} - u_{xx} \ge 0$ as in Section 6.1 by the θ -method, i.e., the approximations w_i^j of $u(x_i, t_j)$ solve

$$\frac{1}{\Delta t}(w_i^{j+1} - w_i^j) - \frac{1 - \theta}{h^2}(w_{i+1}^j - 2w_i^j + w_{i-1}^j) - \frac{\theta}{h^2}(w_{i+1}^{j+1} - 2w_i^{j+1} + w_{i-1}^{j+1}) \ge 0.$$

Setting $\gamma = \Delta t / h^2$, we can formulate this inequality more compactly in matrix form:

$$Aw^{j+1} \ge Bw^j + d^j,$$

where

$$A = \operatorname{diag}(-\gamma\theta, 1 + 2\gamma\theta, -\gamma\theta), B = \operatorname{diag}(\gamma(1-\theta), 1 - 2\gamma(1-\theta), \gamma(1-\theta)),$$
(6.15)

and the vector d^{j} contains the boundary conditions:

$$d^{j} = \gamma \begin{pmatrix} (1-\theta)u(-L,t_{j}) + \theta u(-L,t_{j+1}) \\ 0 \\ \vdots \\ 0 \\ (1-\theta)u(L,t_{j}) + \theta u(L,t_{j+1}) \end{pmatrix}.$$
 (6.16)

Let $b^j = Bw^j + d^j$ and $f^j = (f(x_1, t_j), \dots, f(x_{n-1}, t_j))^\top$. Then we need to solve the following discrete linear complementarity problem

$$(Aw^{j+1} - b^j) \cdot (w^{j+1} - f^{j+1}) = 0, \quad Aw^{j+1} - b^j \ge 0, \quad w^{j+1} - f^{j+1} \ge 0.$$
 (6.17)

In the following, we omit the indices j and j + 1. For later use, we remark that the complementarity problem can be written as a minimum problem.

Lemma 6.8. The problem

$$(Aw - b) \cdot (w - f) = 0, \quad Aw - b \ge 0, \quad w - f \ge 0$$
 (6.18)

is equivalent to $\min\{Aw - b, w - f\} = 0$.

Proof. Let $\min\{Aw - b, w - f\}_i = 0$ for all *i*. This implies that $(Aw - b)_i \ge 0$ and $(w - f)_i \ge 0$ for all *i*. Moreover, $(Aw - b)_i = 0$ or $(w - f)_i = 0$, which gives $(Aw - b) \cdot (w - f) = 0$.

Conversely, let (6.18) hold. Fix some index *i*. Then either $(Aw - b)_i > 0$ or $(Aw - b)_i = 0$. In the former case, by the first equation in (6.18), we must have $(w - f)_i = 0$ and hence, min $\{Aw - b, w - f\}_i = 0$. In the latter case, min $\{Aw - b, w - f\}_i = 0$.

We infer that for given $i \in \{1, ..., n-1\}$, either $w_i = f_i$ or $(Aw)_i = b_i$. Assume for the moment that $(Aw)_i = b_i$ for all *i*. Then, writing this equation as

$$a_{ii}w_i = -\sum_{j < i} a_{ij}w_j - \sum_{j > i} a_{ij}w_j + b_i,$$

we may use the following iteration (called the Gauß-Seidel iteration):

$$w_i^{(k+1)} = \frac{1}{a_{ii}} \bigg(-\sum_{j < i} a_{ij} w_j^{(k+1)} - \sum_{j > i} a_{ij} w_j^{(k)} + b_i \bigg), \quad k \in \mathbb{N},$$

where $w_i^{(0)}$ is given. This iteration is motivated from the fact that at step *i*, we have already calculated the iterations $w_1^{(k+1)}, \ldots, w_{i-1}^{(k+1)}$, thus it makes sense to use them when computing $w_i^{(k+1)}$. A more compact formulation is given as follows: We decompose A = D - L - U, where $D = \text{diag}(a_{11}, \ldots, a_{n-1,n-1}), -L$ is the lower triangular matrix, and -U is the upper triangular matrix. Then

$$w^{(k+1)} = D^{-1}(Lw^{(k+1)} + Uw^{(k)} + b).$$

This idea can be refined by employing the *SOR method* (SOR = successive overrelaxation): Compute for i = 1, ..., n - 1,

$$z_i^{(k)} = a_{ii}^{-1} (Lw^{(k+1)} + Uw^{(k)} + b)_i,$$

$$w_i^{(k+1)} = w_i^{(k)} + \omega(z_i^{(k)} - w_i^{(k)}).$$

We recover the Gauß-Seidel iteration when $\omega = 1$. When $0 < \omega < 1$, the new iteration is a linear combination between the old and new Gauß-Seidel iteration. This corresponds to a damped iteration. The idea of the SOR method is to choose $\omega > 1$ and thus to "overshoot". This works well for $1 < \omega < 2$ in the sense that the iterations converge to

a solution to Aw = f and that for an optimally chosen ω , less iteration steps are needed than for the Gauß-Seidel method.

Generally, we do not have the equality $(Aw)_i = f_i$ but we can exploit the property $\min\{Aw - b, w - f\} = 0$. Indeed, decomposing again A = D - L - U, we obtain

$$0 = \min\{Aw - b, w - f\} = \min\{Dw - (Lw + Uw + b), w - f\}$$

= min {w - D⁻¹(Lw + Uw + b), w - f} = w - max {D⁻¹(Lw + Uw + b), f}

The third equality is possible, since we only distinguish between positivity and zero, so we can multiply the first argument of the minimum componentwise by the matrix elements $(D^{-1})_{ii} = a_{ii}^{-1}$. This motivates the *projection SOR method*: Compute for i = 1, ..., n-1,

$$z_i^{(k)} = a_{ii}^{-1} (Lw^{(k+1)} + Uw^{(k)} + b)_{i'},$$

$$w_i^{(k+1)} = \max \{ w_i^{(k)} + \omega(z_i^{(k)} - w_i^{(k)}), f_i \}.$$
(6.19)

Theorem 6.9 (Cryer). Let $A \in (n-1) \times (n-1)$ be a symmetric, positive definite matrix and let $b \in \mathbb{R}^{n-1}$, $f \in \mathbb{R}^{n-1}$, and $1 < \omega < 2$. Let $w^{(k)}$ for $k \in \mathbb{N}$ be a solution to (6.19). Then

$$\lim_{k\to\infty}w_i^{(k)}=w_i,\quad i=1,\ldots,n-1,$$

where w is a the unique solution to the linear complementarity problem (6.18).

Proof. As the proof is rather technical, we give only a sketch. The idea is first to show that the problem (6.18) is equivalent to the minimization problem: Find $w \in M$ such that

$$J(w) = \min_{v \in M} J(v)$$
, where $J(v) = \frac{1}{2}v^{\top}Av - b^{\top}v$

and $M = \{v \in \mathbb{R}^{n-1} : v_i \ge f_i \text{ for all } i\}$; see [16, Lemma 7.6] for a proof. Second, define the sequence

$$w^{(k,i)} = (w_1^{(k+1)}, \dots, w_i^{(k+1)}, w_{i+1}^{(k)}, \dots, w_{n-1}^{(k)})^{ op}$$

and show that $J_j = J(w^{(k,i)})$ with j = (n-1)(k-1) + i is decreasing (for this step, we need that $\omega < 2$). Since A is symmetric and positive definite, $(J_j)_{j \in \mathbb{N}}$ is bounded from below. Therefore, $(J_j)_{j \in \mathbb{N}}$ is convergent. Third, using the convergence of $(J_j)_{j \in \mathbb{N}}$, show that $(w_i^{(k)})_{k \in \mathbb{N}}$ is convergent to some w_i . Finally, the limit $k \to \infty$ in (6.19) gives

$$z_{i} = \lim_{k \to \infty} z_{i}^{(k)} = a_{ii}^{-1} (Lw + Uw + b)_{i} = a_{ii}^{-1} (Dw - Aw + b)_{i} = w_{i} - a_{ii}^{-1} (Aw - b)_{i},$$

$$w_{i} = \max \{w_{i} + \omega(z_{i} - w_{i}), f_{i}\} = \max \{w_{i} - \omega a_{ii}^{-1} (Aw - b)_{i}, f_{i}\}.$$

This is equivalent to

$$\min\left\{(Aw-b)_i, w_i-f_i\right\}=0$$

and thus to the linear complementarity problem (6.18).

The algorithm for the computation of the price of an American option reads as follows. Let w_i^{j-1} be given from the previous time step. For all time steps j = 1, ..., m,

- Define $f_i^j = f(x_i, t_j)$ according to (6.14).
- Compute the boundary values $w_0 = f_0^j$, $w_n = f_n^j$.
- Define $b_i = (Bw + d)_i$ according to (6.15) and (6.16).
- Compute z_i^j and w_i^j according to

$$z_i^j = \frac{1}{1+2\gamma\theta} \left(\gamma\theta(w_{i+1}^j + w_{i-1}^j) + b_i \right)$$
$$w_i^j = \max\left\{ w_i^j + \omega(z_i^j - w_i^j), f_i^j \right\}$$

until convergence is reached.

Transform back to the original variables.

Figure 6.4 shows the prices of an American put option in comparison with a European put with the same parameters. It is clear that the price for the American option is larger than the corresponding one for the European option, since the American option allows for early exercise, thus giving more rights. The free boundary can be estimated from the numerical values yielding $S_f \approx 67$.





Remark. So far, we have not discussed the convergence of the solution to the discrete linear complementarity problem (6.18) towards the solution to the continuous complementarity problem (6.13). The reason is that solutions to such problems generally do not possess sufficient regularity to estimate the truncation error by means of a Taylor expansion, as done in Section 6.1. A way out is the discretization of the complementarity problem using finite elements. Then the convergence can be proved by means of Sobolev space techniques; see, for instance, [15, Chapter 8].

6.5 Application: Pricing swing options in electricity markets

Electricity prices are subject to the principle of supply and demand. The risk of very high electricity prices may be hedged by derivatives. One exampe are so-called *swing options* which are supply contracts for power. They give the holder the right to exercise a certain right multiple times over a specified period but only one right per time interval (like once per day). The right might be to receive the payoff of a call option or a forward contract which supplies the holder with a certain amount of energy to a fixed predetermined price. In this section, we explain the spot price dynamics of the electricity price and present a free-boundary problem to determine the swing option price. Interestingly, swing options can be interpreted as a portfolio of American options with a waiting period between two exercises and consequently, they can be mathematically formulated as a sequence of free-boundary problems, one for each right to exercise. We proceed as in [7].

We assume that the electricity price S_t is given by the stochastic process

$$S_t = \exp\left(f(t) + \tilde{X}_t + Y_t\right),$$

where f(t) is a deterministic function modeling the seasonality, \tilde{X}_t is the zero mean-reverting process

$$dX_t = -\alpha X_t dt + \sigma dW_t,$$

where $\alpha > 0$, $\sigma > 0$, and W_t is a Wiener process. The process $(\tilde{X}_t)_{t\geq 0}$ describes the fluctuations in the energy demand. Combining the seasonality and stochastic component, we set $X_t = f(t) + \tilde{X}_t$. Finally, $(Y_t)_{t>0}$ is another zero mean-reverting process, given by

$$dY_t = -\beta Y_t dt$$

for some $\beta > 0$.

Remark. Often, Y_t includes jump components to incorporate price peaks, $dY_t = -\beta Y_t dt + J_t dN_t$, where J_t are the jump sizes and $(N_t)_{t\geq 0}$ is a Poisson process. Here, we only consider models without jumps. Jump-diffusion models lead to partial integro-differential equations (PIDE), which are nonlocal partial differential equations. The following considerations are still valid, we just have to include the nonlocal integral in the differential operator *L* defined below. For details, we refer to [6].

With these definitions, we can write $S_t = M_t N_t$, where $M_t = \exp(X_t)$ and $N_t = \exp(Y_t)$. The spot price dynamics is then defined by the following set of stochastic differential equations.

Lemma 6.10. The process $(S_t)_{t>0}$ is an Itô process satisfying

$$dS_t = M_t dN_t + N_t dM_t, \quad where dM_t = \alpha (\mu(t) - \ln M_t) M_t dt + \sigma M_t dW_t, \quad dN_t = -\beta N_t \ln N_t dt,$$

and $\mu(t) = f(t) + (\frac{1}{2}\sigma^2 + f'(t))/\alpha$.

Proof. The result follows from Itô's formula (Theorem 2.12) and Itô's product rule (Lemma 2.14). Indeed, applying Itô's formula to $M_t = F(\tilde{X}_t, t) := \exp(f(t) + \tilde{X}_t)$ and recalling that $\tilde{X}_t = X_t - f(t)$, we find that

$$dM_{t} = \left(\frac{\partial F}{\partial t} - \alpha \widetilde{X}_{t} \frac{\partial F}{\partial X} + \frac{1}{2}\sigma^{2} \frac{\partial^{2} F}{\partial X^{2}}\right) dt + \sigma \frac{\partial F}{\partial X} dW_{t}$$

$$= \left(f'(t) - \alpha (\underbrace{X_{t}}_{=\ln M_{t}} - f(t)) + \frac{1}{2}\sigma^{2}\right) M_{t} dt + \sigma M_{t} dW_{t}$$

$$= \left(f'(t) + \alpha (f(t) - \ln M_{t}) + \frac{1}{2}\sigma^{2}\right) M_{t} dt + \sigma M_{t} dW_{t},$$

$$dN_{t} = \exp(Y_{t}) dY_{t} = -\exp(Y_{t})\beta Y_{t} dt = -\beta N_{t} \ln N_{t} dt.$$

By definition of $\mu(t)$, this finishes the proof.

Next, let $V_t = V(M_t, N_t, t)$ be a function of the stochastic factors M_t and N_t . The function V is considered here as a generic function, but we may interpret it as the value of an option. The goal is to derive a partial differential equation for V(M, N, T) with the variables M, N, and t using a dynamic hedging argument. Since we cannot directly hedge the risk, we proceed as in Section 3.6.

Proposition 6.11. Neglecting the market price of risk, the function V(M, N, T) solves the equation

$$V_t + \frac{1}{2}\sigma^2 M^2 V_{MM} + \alpha (\mu(t) - \ln M) M V_M - \beta N \ln N V_N - rV = 0$$

for $M, N \in (0, \infty)$ and $t \in (0, T)$. The final condition is given by $V(M, N, T) = V_T(MN)$.

Proof. We build the riskfree and self-financing portfolio $\pi = V - \Delta \tilde{V}$ by buying one asset *V* with maturity T_1 and selling Δ shares of the asset \tilde{V} with maturity T_2 . Applying Itô's formula to *V* gives

$$dV = \left(V_t + \alpha(\mu(t) - \ln M)MV_M - \beta N \ln NV_N + \frac{1}{2}\sigma^2 M^2 V_{MM}\right)dt + \sigma M V_M dW,$$

and the analogous equation for \tilde{V} . Since the portfolio is assumed to be self-financing, we infer that

$$d\pi = dV - \Delta dV$$

= $\left(V_t + \alpha (\mu(t) - \ln M) M V_M - \beta N \ln N V_N + \frac{1}{2} \sigma^2 M^2 V_{MM} \right) dt$

$$-\Delta \left(\widetilde{V}_t + \alpha (\mu(t) - \ln M) M \widetilde{V}_M - \beta N \ln N \widetilde{V}_N + \frac{1}{2} \sigma^2 M^2 \widetilde{V}_{MM} \right) dt + \sigma M (V_M - \Delta \widetilde{V}_M) dW.$$

We eliminate the stochastic component by setting $\Delta = V_M / \tilde{V}_M$. With this choice, the terms with the factor V_M cancel, and we end up with

$$d\pi = \left(V_t - \beta N \ln N V_N + \frac{1}{2}\sigma^2 M^2 V_{MM}\right) dt$$
$$-\Delta \left(\widetilde{V}_t - \beta N \ln N \widetilde{V}_N + \frac{1}{2}\sigma^2 M^2 \widetilde{V}_{MM}\right) dt$$

Since the portfolio is riskless, we have $d\pi = r\pi dt = r(V - \Delta \tilde{V})dt$, and we find that

$$\frac{1}{V_M} \left(V_t - \beta N \ln N V_N + \frac{1}{2} \sigma^2 M^2 V_{MM} - rV \right)$$
$$= \frac{1}{\widetilde{V}_M} \left(\widetilde{V}_t - \beta N \ln N \widetilde{V}_N + \frac{1}{2} \sigma^2 M^2 \widetilde{V}_{MM} - r \widetilde{V} \right).$$

The left-hand side only depends on V, while the right-hand side only depends on \tilde{V} . Since we can choose any assets V and \tilde{V} , both sides must be independent of the contract type:

$$\gamma_0(M,N,t) = \frac{1}{V_M} \left(V_t - \beta N \ln N V_N + \frac{1}{2} \sigma^2 M^2 V_{MM} - r V \right).$$

We set $\gamma = \alpha(\mu(t) - \ln M)M + \gamma_0$. Then

$$V_t + \frac{1}{2}\sigma^2 M^2 V_{MM} - \beta N \ln N V_N - rV = -\alpha(\mu(t) - \ln M)MV_M + \gamma V_M$$

Similarly as in the derivation of the Heston-Black-Scholes model (see Theorem 3.8), the function γ is not specified. It may be interpreted as the market price of risk, and we set it to zero, $\gamma = 0$.

Swing options can be modeled as financial products with multiple exercises like for American options. However, after one exercise one has to wait for the constant refracting period $\delta > 0$. This period avoids the exercise of all rights at once, which would be optimal when $\delta = 0$. In other words, if $\delta = 0$, the pricing of swing option can be reduced to the valuation of multiple American options. We have seen that American options can be modeled by linear complementarity problems; see Section 6.4. We will formulate the swing option problem as a number of such problems. For this, we reverse the time by introducing $\tau = T - t$ and $u(M, N, \tau) = V(M, N, T - \tau)$. Furthermore, we define the differential operator

$$L(V) = V_t + \frac{1}{2}\sigma^2 M^2 V_{MM} + \alpha(\mu(t) - \ln M) M V_M - \beta N \ln N V_N - rV.$$

Let $p \ge 1$ denote the number of exercise rights. We start from $V_0(M, N, t) = 0$ for $t \in (0, \delta)$. We can write the model as p complementarity problems for the price $u_p(M, N, \tau) = V_p(M, N, T - \tau)$ of a swing option with p exercise rights:

$$-L(u_p)(u_p - g_p) = 0, \quad -L(u_p) \ge 0, \quad u_p \ge g_p,$$

 $u_p(0) = g_p(0) \quad \text{in } \mathbb{R}^2_+,$

where the reward obstacle function g_p is given by (recall that S = MN)

$$g_p(M, N, \tau) = \begin{cases} V_T(S, T - \tau) + w_{\tau, p-1}(M, N, \delta) & \text{for } \tau \in [\delta, T], \\ V_T(S, T - \tau) & \text{for } \tau \in [0, \delta). \end{cases}$$

The function $w_{\tau,p-1}$ is the value of the swing option with one exercise right less, and it is determined for p = 1 by $w_{\tau,0}(M, N, t) = 0$ for $t \in [0, \delta]$ and for $p \ge 2$ by

$$L(w_{\tau,p-1}) = 0$$
 for $t \in (0,\delta)$, $w_{\tau,p-1}(0) = u_{p-1}(\tau - \delta)$ in \mathbb{R}^2_+ .

References

- [1] Y. Achdou and O. Pironneau. *Computational Methods for Option Pricing*. Society for Industrial and Applied Mathematics, Providence, 2005.
- [2] G. Barles and H. Soner. Option pricing with transaction costs and a nonlinear Black-Scholes equation. *Finance Stoch.* 2 (1998), 369-397.
- [3] R. Bass. Stochastic Processes. Cambridge University Press, Cambridge, 2012.
- [4] F. Baudoin. *Diffusion Processes and Stochastic Calculus*. EMS Textbooks in Math., Zürich, 2014.
- [5] K. Burrage und P. Burrage. High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations. *Appl. Numer. Math.* 22 (1996), 81-101.
- [6] M. Calvo-Garrido, M. Ehrhardt, and C. Vázquez. Jump-diffusion models with two stochastic factors for pricing swing options in electricity markets with partialintegro differential equations. Preprint BUW-IMACM 17/08, Universität Wuppertal, 2017.
- [7] M. Calvo-Garrido, M. Ehrhardt, and C. Vázquez. Pricing swing options in electricity markets with two stochastic factors using a partial differential equations approach. J. Comput. Finance 20 (2017), 81-107.
- [8] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall, Boca Raton, 2004.
- [9] J. Cox, J. Ingersoll und S. Ross. A theory of the term structure of interest rates. *Econometrica* 53 (1985), 385-467.
- [10] G. Da Prato and J. Zabczyk. Stochastic Equations in Infinite Dimensions. Second edition. Cambridge University Press, Cambridge, 2014.
- [11] F. Delbaen and W. Schachermayer. *The Mathematics of Arbitrage*. Springer, Berlin, 2006.
- [12] L. Devroye. Non-Uniform Random Variate Generation. Springer, New York, 1986.
- [13] R. Elliott and J. van der Hoek. A general fractional white noise theory and applications to finance. *Math. Finance* 13 (2003), 301-330.
- [14] H. Föllmer and A. Schied. *Stochastic Finance. An Introduction in Discrete Time*. 2nd edition. De Gruyter, Berlin, 2004.

- [15] C. Großmann and H.-G. Roos. Numerik partieller Differentialgleichungen. Teubner, Stuttgart, 1994.
- [16] M. Günther and A. Jüngel. *Finanzderivate mit MATLAB*. Second edition. Vieweg+Teubner, Wiesbaden, 2010.
- [17] S. Heston. A closed-form solution for options with stochastic volatility with application to bond and currency options. *Rev. Financial Studies* 6 (1993), 327-343.
- [18] Y. Hu and B. Øksendal. Fractional white noise calculus and applications to finance. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 6 (2003), 1-32.
- [19] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. J. Finance 42 (1987), 281-300.
- [20] P. Kloeden and E. Platen. Numerical Solution of Stochastic Differential Equations. Second edition. Springer, Berlin, 1995.
- [21] L. Jiang. *Mathematical Theory of Financial Derivatives*. Tongji University, Shanghai, China, 2001.
- [22] H. Leland. Option pricing and replication with transaction costs. J. Finance 40 (1985), 1283-1301.
- [23] G. Lord, C. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge University Press, Cambridge, 2014.
- [24] T. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* 14 (1998), 215-310.
- [25] M. Matsumoto and T. Nishimura. Mersenne twister. A 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Transactions Model. Computer Simul. 8 (1998), 3-30.
- [26] R. Merton. Option pricing when underlying stock returns are discontinuous. J. Financial Econ. 3 (1976), 125-144.
- [27] B. Øksendal. *Stochastic Differential Equations*. Fifth edition. Springer, New York, 2000.
- [28] A. Pascucci. PDE and Martingale Methods in Option Pricing. Springer, Mailand, 2011.
- [29] C. Prévôt and M. Röckner. A Concise Course on Stochastic Partial Differential Equations. Springer, Berlin, 2007.

- [30] M. Rodrigo and R. Mamon. An alternative approach to solving the Black-Scholes equation with time-varying parameters. *Appl. Math. Lett.* 19 (2006), 398-402.
- [31] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust Numerical Methods for Singularly Perturbed Differential Equations*. Second edition. Berlin, Springer, 2008.
- [32] W. Schachermayer and J. Teichmann. How close are the option pricing formulas of Bachelier and Black-Merton-Scholes? *Math. Finance* 18 (2008), 155-170.
- [33] R. Seydel. *Tools for Computational Finance*. Fifth edition. Springer, London, 2012.
- [34] O. Vasicek. An equilibrium characterization of the term structure. *J. Financ. Econom.* 5 (1977), 177-188.
- [35] P. Wilmott. Paul Wilmott on Quantitative Finance. John Wiley, Chichester, 2000.
- [36] P. Wilmott, S. Howison, and J. Dewynne. *The Mathematics of Financial Derivatives*. Cambridge University Press, Cambridge, 1996.
- [37] E. Zeidler. Nonlinear Functional Analysis and Applications. Volume II/A. Springer, New York, 1990.