Numerical methods in mathematical finance

Tobias Jahnke

Karlsruher Institut für Technologie
Fakultät für Mathematik
Institut für Angewandte und Numerische Mathematik

tobias.jahnke@kit.edu

Version: April 5, 2013

Preface

These notes are the basis of my lecture *Numerical methods in mathematical finance* given at Karlsruhe Institute of Technology in the winter term 2012/13. The purpose of this notes is to help students who have missed parts of the course to fill these gaps, and to provide a service for those students who can concentrate better if they do not have to copy what I write on the blackboard.

It is *not* the purpose of these notes, however, to replace the lecture itself, or to write a text which could compete with the many excellent books about the subject. This is why the style of presentation is rather sketchy. As a rule of thumb, one could say that these notes only cover what I *write* during the lecture, but not everything I *say*.

Since this is the first edition (or should I say "draft"?) of these notes, and since the manuscript is written in parallel to the lecture, there will be many typos and possibly also other mistakes. Of course, I will try to correct any mistake I find as soon as possible, but the reader should be aware of the fact that he or she cannot trust these notes.

Karlsruhe, winter term 2012/13,
Tobias Jahnke

Contents

Part I

Mathematical models in option pricing

Chapter 1

Options and arbitrage

References: [BK04, Sey09]

1.1 European options

Financial markets trade investments into stocks of a company, commodities (e.g. oil, gold), etc.
Stocks and commodities are risky assets, because their future value cannot be predicted. Bonds are considered as riskless assets in this lecture. If $B(t_0)$ is invested at time t_0 into a bond with a risk-free interest rate $r > 0$, then the value of the bond at time $t \geq t_0$ is simply

$$B(t) = e^{r(t-t_0)} B(t_0). \tag{1.1}$$

Simplifying assumption: continuous payment of interest
Spot contract: buy or sell an asset (e.g. a stock, a commodity etc.) with immediate delivery
Financial derivatives: contracts about future payments or deliveries with certain conditions

1. forwards: agreement between two parties to buy or sell an asset at a certain time in the future for a certain delivery price
2. futures: similar to forwards
3. swaps: contracts regulating an exchange of cash flows at different future times (e.g. currency swap, interest rate swaps, credit default swaps)
4. options

Definition 1.1.1 (European option)

- *A **European call option** is a contract which gives the holder (=buyer) of the option the right to buy an underlying risky asset at a future maturity date (expiration time) T at a fixed exercise price (strike) K from the writer (=seller) of the option.*

Typical assets: stocks, parcels of stocks, stock indices, currencies, commodities, ...
Difference to forwards and futures: At maturity the holder can choose if he wants to buy the asset or not.

- ***European put option:*** *Similar to call option, but vice versa, i.e. the holder can sell the underlying to the writer.*

Example: At time $t = 0$ Mr. J. buys 5 European call options. Each of these options gives him the right to buy 10 shares of the company KIT at maturity $T > 0$ at the exercise price of $K = 120€$ per share.

- **Case 1:** At time $t = T$, the market price of KIT is 150€ per share.
  Mr. J. exercises his options, i.e. he buys $5 \cdot 10 = 50$ KIT shares at the price of $K = 120€$ per share and sells the shares on the market for 150€ per share. Hence, he wins $50 \cdot 30 = 1500€$.
- **Case 2:** At time $t = T$, the market price of KIT is 100€ per share.
  Hence, Mr. J. does not exercise his options.

What are options good for?

- Speculation
- Hedging ("insurance" against changing market values)

Since an option gives an advantage to the holder, the option has a certain value.
For given T and K the value $V(t, S)$ of the option must depend on the time t and the current price S of the underlying.
For an European option we know that the value at the maturity T is

$$V(T,S) = \begin{cases} (S-K)^+ := \max\{S-K, 0\} & \text{(European call)} \\ (K-S)^+ := \max\{K-S, 0\} & \text{(European put).} \end{cases}$$

The functions $S \mapsto (S-K)^+$ and $S \mapsto (K-S)^+$ are called the **payoff functions** of a call or put, respectively.

The **goal of this course** is to answer the following question:

> What is the fair price $V(t, S)$ of an option for $t < T$?

Why is this question important? In order to sell/buy an option, we need to know the fair price.
Why is this question non-trivial? Because the value of the risky asset is random. In particular, the price $S(T)$ at the future expiration time T is not yet known when we buy/sell the option at time $t = 0$.

1.2 More types of options

Variations of the basic principle:

- **European options** can be exercised only at the maturity date.
- **American options** can be exercised at any time before and including the maturity date.
- **Bermuda options** can be exercised at a set of times.

The names "European", "American", "Bermuda" etc. have no geographical meaning. American options can be traded in Europe, European options can be traded in the USA, etc.

- **Vanilla options** = standard options, i.e. European, American or Bermuda calls/puts
- **Exotic options** = non-standard options

Examples for exotic options:

- **Path-dependent option:** The payoff function does not only depend on the price $S(T)$ of the underlying at time T, but on the entire path $t \mapsto S(t)$ for $t \in [0, T]$.
  - **Asian options:** The payoff function depends on the average price, e.g.

$$\left(\frac{1}{T} \int_0^T S(t)\, dt - K \right)^+$$

    (payoff of an average price call).
  - **Barrier options:** The payoff depends on the question if the price of the underlying has crossed a certain (upper or lower) barrier.
  - **Lookback options:** The payoff depends on $\max_{t\in[0,T]} S(t)$ or $\min_{t\in[0,T]} S(t)$.
- **Options on several assets:**
  - **Basket options:** The payoff depends on the weighted sum of the prices S_i of several assets, e.g.

$$\left(\sum_{i=1}^d c_i S_i - K \right)^+, \qquad c_i > 0$$

    (payoff of a basket call)
  - **Rainbow options:** The payoff depends on the relation between the assets, e.g. $\max\{S_1, \ldots, S_d\}$.
- **Binary options:** The payoff function has only two possible values
- **Compound options:** Options on options

Remark: There are even more types of options.

1.3 Arbitrage and modelling assumptions

Example. Consider

- a stock with price $S(t)$
- a European call option with maturity $T = 1$, strike $K = 100$, and value $V(t, S(t))$
- a bond with price $B(t)$

Initial data: $S(0) = 100, B(0) = 100, V(0) = 10$.
Assumption: At time $t = 1$, we either have

$$\begin{aligned} &\text{“up”:} \quad B(1) = 110,\ S(1) = 120 \\ \text{or} \quad &\text{“down”:}\ B(1) = 110,\ S(1) = 80 \end{aligned}$$

At $t = 0$, Mrs. C. buys 0.4 bonds, one call option and sells 0.5 stock (“short selling”).
Value of the portfolio at $t = 0$:

$$0.4 \cdot B(0) + 1 \cdot V(0) - 0.5 \cdot S(0) = 0.4 \cdot 100 + 1 \cdot 10 - 0.5 \cdot 100 = 0$$

Value of the portfolio at $t = 1$ is

$$0.4 \cdot B(1) + 1 \cdot \underbrace{V(1, S(1))}_{=(S(1)-K)^+} - 0.5 \cdot S(1)$$

Two cases:

$$\begin{aligned} \text{“up”:} \quad & 0.4 \cdot 110 + 1 \cdot (120 - 100)^+ - 0.5 \cdot 120 = 44 + 20 - 60 = 4 \\ \text{“down”:} \quad & 0.4 \cdot 110 + 1 \cdot (80 - 100)^+ - 0.5 \cdot 80 \quad = 44 + 0 - 40 \ \ = 4 \end{aligned}$$

In both cases, Mrs. C. wins 4€ without any risk or investment!
Why is this possible? Because the price $V(0) = 10$ of the option is too low!

Definition 1.3.1 (Arbitrage)
Arbitrage *is the existence of a portfolio, which*

- *requires no initial investment, and*
- *which cannot cause any loss, but very likely a gain at maturity.*

Remark. A bond will always yield a risk-less gain, but it requires an investment.

Assumptions for modelling an idealized market:

(A1) Arbitrage is impossible (no-arbitrage principle)

(A2) There is a risk-free interest rate $r > 0$ which applies for all credits. Continuous payment of interest according to (1.1).

(A3) No transaction costs, taxes, etc. Trading is possible at any time. Any fraction of an asset can be sold. Liquid market, i.e. selling an asset does not change its value significantly.

(A4) A seller can sell assets he/she does not own yet ("short selling", cf. Mrs. C. above)

(A5) No dividends on the underlying asset are paid.

Remark. Discrete payment of interest: obtain $r \cdot \Delta t \cdot B(0)$ after time Δt. Value at $t = n\Delta t$:

$$\tilde{B}(t) = (1 + r \cdot \Delta t)^n B(0) = (1 + rt/n)^n B(0)$$

For $n \longrightarrow \infty$ and $\Delta t \longrightarrow 0$:

$$\lim_{n\to\infty} \tilde{B}(t) = \lim_{n\to\infty} (1 + rt/n)^n B(0) = e^{rt} B(0) = B(t)$$

(continuous payment of interest)

1.4 Arbitrage bounds

Consider European options with strike $K > 0$ and maturity T on an underlying with price $S(t)$. Let $V_P(t, S)$ and $V_C(t, S)$ be the values of a put option and call option, respectively.

Lemma 1.4.1 (Put-call parity) *Under the assumptions (A1)-(A5) we have*

$$S(t) + V_P(t, S(t)) - V_C(t, S(t)) = e^{-r(T-t)} K$$

for all $t \in [0, T]$.

Proof. Buy one stock, buy a put, write (sell) a call. Then, the value of this portfolio is

$$\phi(t) = S(t) + V_P(t, S(t)) - V_C(t, S(t))$$

and at maturity

$$\phi(T) = S(T) + V_P(T, S(T)) - V_C(T, S(T)) = S(T) + (K - S(T))^+ - (S(T) - K)^+ = K.$$

Hence, the portfolio is risk-less. No arbitrage: The profit of the portfolio must be the same as the profit for investing $\phi(t)$ into a bond at time t:

$$\phi(T) = K \stackrel{!}{=} e^{r(T-t)}\phi(t) \quad \Longrightarrow \quad e^{-r(T-t)} K = \phi(t) = S(t) + V_P(t, S(t)) - V_C(t, S(t)). \quad \blacksquare$$

Lemma 1.4.2 (Bounds for European calls and puts) *Under the assumptions (A1)-(A5), the following inequalities hold for all* $t \in [0, T]$ *and all* $S = S(t) \geq 0$*:*

$$\left(S - e^{-r(T-t)} K\right)^+ \leq V_C(t, S) \leq S \tag{1.2}$$

$$\left(e^{-r(T-t)} K - S\right)^+ \leq V_P(t, S) \leq e^{-r(T-t)} K \tag{1.3}$$

Proof.

- It is obvious that $V_C(t,S) \geq 0$ and $V_P(t,S) \geq 0$ for all $t \in [0,T]$ and $S \geq 0$.
- Assume that $V_C(t,S(t)) > S(t)$ for some $S(t) \geq 0$.
  Write (sell) a call, buy the stock and put the difference $V_C(t,S(t)) - S(t) > 0$ in your pocket.
  At $t = T$, there are two scenarios:
  If $S(T) > K$: Must sell stock at the price K to the owner of the call.
  Gain: $K + V_C(t,S(t)) - S(t) > 0$
  If $S(T) \leq K$: Gain $S(T) + V_C(t,S(t)) - S(t) > 0$
  $\Longrightarrow$ Arbitrage! Contradiction!
- Put-call parity:

$$S - e^{-r(T-t)}K = V_C(t,S) - \underbrace{V_P(t,S)}_{\geq 0} \leq V_C(t,S)$$

  This proves (1.2). The proof of (1.3) is left as an exercise. ■

Remark. Similar inequalities can be shown for American options (exercise).

1.5 A simple discrete model

Consider

- a stock with price $S(t)$
- a European option with maturity T, strike K, and value $V(t,S(t))$
- a bond with price $B(t) = e^{rt}B(0)$

Suppose that the initial data $S(0) = S_0$ and $B(0) = 1$ are known, and that (A1)-(A5) hold.

Goal: Find $V(0,S_0)$.

Simplifying assumption: At time $t = T$, there are only two scenarios

or

"up": $S(T) = u \cdot S_0$ with probability p
"down": $S(T) = d \cdot S_0$ with probability $1-p$

Assumption: $0 < d < u$ and $p \in (0,1)$
In both cases, we have $B(T) = e^{rT}B(0) = e^{rT}$.

Replication strategy: Construct portfolio with c_1 bonds and c_2 stocks such that

$$c_1 B(t) + c_2 S(t) \stackrel{!}{=} V(t,S(t))$$

for $t \in \{0, T\}$. For $t = T$, this means

$$\text{case "up":} \quad c_1 e^{rT} + c_2 u S_0 \stackrel{!}{=} V(T, uS_0) =: V_u$$
$$\text{case "down":} \quad c_1 e^{rT} + c_2 d S_0 \stackrel{!}{=} V(T, dS_0) =: V_d$$

V_u and V_d are known if u and d are known. The unique solution is (check!)

$$c_1 = \frac{uV_d - dV_u}{(u-d)e^{rT}} \qquad\qquad c_2 = \frac{V_u - V_d}{(u-d)S_0}.$$

Hence, the fair price of the option is

$$V(0, S_0) = c_1 \underbrace{B(0)}_{=1} + c_2 S_0 = \frac{uV_d - dV_u}{(u-d)e^{rT}} + \frac{V_u - V_d}{(u-d)}$$

which yields (check!)

$$V(0, S_0) = e^{-rT}\Big(qV_u + (1-q)V_d\Big) \qquad \text{with} \qquad q := \frac{e^{rT} - d}{u - d}. \tag{1.4}$$

Remark: The value of the option does **not** depend on p.

The no-arbitrage assumption (A1) implies $d \le e^{rT} \le u$. Hence, $q \in [0, 1]$ can be seen as a probability. Now, define a new probability distribution $\mathbb{P}_q$ by

$$\mathbb{P}_q\Big(S(T) = uS_0\Big) = q, \qquad\qquad \mathbb{P}_q\Big(S(T) = dS_0\Big) = 1 - q$$

(q instead of p). Then, we have

$$\mathbb{P}_q\Big(V(T, S(T)) = V_u\Big) = q, \qquad\qquad \mathbb{P}_q\Big(V(T, S(T)) = V_d\Big) = 1 - q$$

and hence

$$qV_u + (1-q)V_d = \mathbb{E}_q\Big(V(T, S(T))\Big)$$

can be regarded as the **expectation** of the payoff $V(T, S(T))$ with respect to $\mathbb{P}_q$. In (1.4), this expectation is multiplied by an **discounting** factor e^{-rT}.
Interpretation: In order to have an amount of $B(t)$ at time t, we have to invest $B(0) = e^{-rt}B(t)$ into a bond at time $t = 0$.
The probability q has the property that

$$\mathbb{E}_q(S(T)) = quS_0 + (1-q)dS_0 = \frac{e^{rT} - d}{u - d} uS_0 + \frac{u - e^{rT}}{u - d} dS_0 = e^{rT} S_0.$$

Hence, the expected (with respect to $\mathbb{P}_q$) value of $S(T)$ is exactly the amount we obtain when we invest S_0 into a bond. Therefore, $\mathbb{P}_q$ is called the **risk-neutral probability**.

Moral of the story so far:

Under the risk-neutral probability, the price of a European option is the discounted expectation of the payoff.

Chapter 2

The Itô integral and stochastic differential equations

References: [KP99, Øks03, Shr04, Ste01]

The model considered in 1.5 is clearly too simple: only two discrete times, only two possible prices of $S(T)$.

Goal: Construct a more realistic model for the dynamics of $S(t)$.

Ansatz:

$$\underbrace{\frac{dS}{dt} = f(t,S)}_{\substack{\text{ordinary differential} \\ \text{equation}}} \quad + \quad \underbrace{\text{random noise}}_{?}$$

2.1 Some definitions from probability theory

Definition 2.1.1 (Probability space) *The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**, if the following holds:*

1. *$\Omega \neq \emptyset$ is a set, and $\mathcal{F}$ is a **σ-algebra** (or σ-field) on Ω, i.e. a family of subsets of Ω with the following properties:*
   - *$\emptyset \in \mathcal{F}$*
   - *If $F \in \mathcal{F}$, then $\Omega \setminus F \in \mathcal{F}$*
   - *If $F_i \in \mathcal{F}$ for all $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} F_i \in \mathcal{F}$*

   *The pair $(\Omega, \mathcal{F})$ is called a **measurable space**.*
2. *$\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ is a **probability measure**, i.e.*
   - *$\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$*

- *If $F_i \in \mathcal{F}$ for all $i \in \mathbb{N}$ are pairwise disjoint (i.e. $F_i \cap F_j = \emptyset$ for $i \neq j$), then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(F_i).$$

A probability space is complete if $\mathcal{F}$ contains all subsets G of Ω with $\mathbb{P}$-outer measure zero, i.e. with

$$\mathbb{P}^*(G) := \inf\left\{\mathbb{P}(F) : F \in \mathcal{F} \text{ and } G \subset F\right\} = 0.$$

Any probability space can be completed. Hence, we can assume that every probability space in this lecture is complete.

Definition 2.1.2 (Borel σ-algebra) *If $\mathcal{U}$ is a family of subsets of Ω, then* ***the σ-algebra generated by $\mathcal{U}$*** *is*

$$\mathcal{F}_{\mathcal{U}} = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra of } \Omega \text{ and } \mathcal{U} \subset \mathcal{F}\}.$$

If $\mathcal{U}$ is the collection of all open subsets of a topological space Ω (e.g. $\Omega = \mathbb{R}^d$), then $\mathcal{B} = \mathcal{F}_{\mathcal{U}}$ is called the ***Borel σ-algebra*** *on Ω. The elements $B \in \mathcal{B}$ are called Borel sets.*

For the rest of this section $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition 2.1.3 (Measurable functions, random variables)

- *A function $X : \Omega \longrightarrow \mathbb{R}^d$ is called* ***$\mathcal{F}$-measurable*** *if*

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

  for all Borel sets $B \in \mathcal{B}$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then every $\mathcal{F}$-measurable functions is called a ***random variable****.*

- *Random variables $X_1, \ldots, X_n$ are called* ***independent*** *if*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right) = \prod_{i=1}^{n} \mathbb{P}\big(X_i^{-1}(A_i)\big)$$

  for all $A_1, \ldots, A_n \in \mathcal{B}$.

- *If $X : \Omega \longrightarrow \mathbb{R}^d$ is any function, then the* ***σ-algebra generated by X*** *is the smallest σ-algebra on Ω containing all the sets*

$$X^{-1}(B) \text{ for all } B \in \mathcal{B}.$$

  Notation: $\mathcal{F}^X = \sigma\{X\}$
  $\mathcal{F}^X$ is the smallest σ-algebra where X is measurable.

Definition 2.1.4 (Stochastic process) *Let $\mathcal{T}$ be an ordered set (e.g. $\mathcal{T} = [0, \infty), \mathcal{T} = \mathbb{N}$). A **stochastic process** is a family $X = \{X_t : t \in \mathcal{T}\}$ of random variables*

$$X_t : \Omega \longrightarrow \mathbb{R}^d.$$

Below, we will often simply write X_t instead of $\{X_t : t \in \mathcal{T}\}$.
Equivalent notations: $X(t, \omega)$, $X(t)$, $X(\omega)$, $X_t(\omega)$, X_t, ...
For a fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called a realization (or path or trajectory) of X.

The path of a stochastic process is associated to some $\omega \in \Omega$. As time evolves, more information about ω becomes available.

Example (cf. chapter 2 in [Shr04]). Toss a coin three times. Possible results are:

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
HHH	HHT	HTH	HTT	THH	THT	TTH	TTT

(H = heads, T = tails).

- Before the first toss, we only know that $\omega \in \Omega$.
- After the first toss, we know if the final result will belong to

  $$\{HHH, HHT, HTH, HTT\} \text{ or to } \{THH, THT, TTH, TTT\}.$$

  These sets are "resolved by the information". Hence, we know in which of the sets

  $$\{w_1, w_2, w_3, w_4\}, \{w_5, w_6, w_7, w_8\}$$

  ω is.
- After the second toss, the sets

  $$\{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}$$

  are resolved, and we know in which of the sets

  $$\{w_1, w_2\}, \{\omega_3, w_4\}, \{w_5, w_6\}, \{w_7, w_8\}$$

  ω is.

Definition 2.1.5 (Filtration)

- *A **filtration** is a family $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $t \geq s \geq 0$.*
- *If $\{X_t : t \geq 0\}$ is a family of random variables and X_t is $\mathcal{F}_t$-measurable, then $\{X_t : t \geq 0\}$ is **adapted** to (or **nonanticipating** with respect to) $\{\mathcal{F}_t : t \geq 0\}$.*
  Interpretation: At time t we know for each set $S \in \mathcal{F}_t$ if $\omega \in S$ or not.

- $\{\mathcal{F}_t : t \geq 0\}$ *is called the* ***natural filtration*** *of a stochastic process X_t if $\mathcal{F}_t$ is the smallest σ-algebras which contains $\mathcal{F}^{X_s}$ for all $s \in [0,t]$, i.e. $\mathcal{F}_t = \sigma\{X_s,\ s \in [0,t]\}$. This is the smallest filtration to which X_t is adapted.*

Definition 2.1.6 (Normal distribution) *A random variable $X : \Omega \longrightarrow \mathbb{R}^d$ with $d \in \mathbb{N}$ is* ***normal*** *if it has a multivariate* ***normal (Gaussian) distribution*** *with mean $\mu \in \mathbb{R}^d$ and a symmetric, positive definite covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, i.e. if*

$$\mathbb{P}(X \in B) = \int_B \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)\, dx$$

for all Borel sets $B \subset \mathbb{R}^d$. Notation: $X \sim \mathcal{N}(\mu, \Sigma)$

Remarks:

1. If $X \sim \mathcal{N}(\mu,\Sigma)$, then $\mathbb{E}(X) = \mu$ and $\Sigma = (\sigma_{ij})$ with $\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$.
2. Standard normal distribution $\Leftrightarrow \mu = 0$, $\Sigma = I$. (I identity matrix)
3. If $X \sim \mathcal{N}(\mu,\Sigma)$ and $Y = v + TX$ for some $v \in \mathbb{R}^d$ and a regular matrix $T \in \mathbb{R}^{d\times d}$, then

$$Y \sim \mathcal{N}(v + T\mu, T\Sigma T^T). \tag{2.1}$$

2.2 The Wiener process

A very important stochastic process is the Wiener process. This process will serve as the "source of randomness" in our model of the financial market.

Robert Brown 1827, Louis Bachelier 1900, Albert Einstein 1905, Norbert Wiener 1923

Definition 2.2.1 (Wiener process, Brownian motion)
A continuous-time stochastic process $\{W_t : t \in [0,T)\}$ is called a ***standard Brownian motion*** *or* ***standard Wiener process*** *if it has the following properties:*

1. *$W_0 = 0$ (with probability one)*
2. *Independent increments: For all $0 \leq t_1 < t_2 < \ldots < t_n < T$ the random variables*

$$W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$$

   are independent.
3. *$W_t - W_s \sim \mathcal{N}(0, t-s)$ for any $0 \leq s < t < T$.*
4. *There is a $\tilde{\Omega} \subset \Omega$ with $\mathbb{P}(\tilde{\Omega}) = 1$ such that $t \mapsto W_t(\omega)$ is continuous for all $\omega \in \tilde{\Omega}$.*

If $W_t^{(1)}, \ldots, W_t^{(d)}$ are independent one-dimensional Wiener processes, then $W_t = \left(W_t^{(1)}, \ldots, W_t^{(d)}\right)$ is called a d-dimensional Wiener process, and

$$W_t - W_s \sim \mathcal{N}(0, (t-s)I).$$

Numerical simulation of a Wiener process

Choose step-size $h > 0$, put $t_n = n \cdot h$ and $\tilde{W}_0 = (0, \ldots, 0) \in \mathbb{R}^d$.

For $n = 0, 1, 2, 3, \ldots$
Generate random vector $Z_n \sim \mathcal{N}(0, I)$
$\tilde{W}_{n+1} = \tilde{W}_n + \sqrt{h} Z_n$

For $h \longrightarrow 0$ the interpolation of $\tilde{W}_n$, $n \in \mathbb{N}$ approximates a path of Brownian motion.

How smooth is a path of a Wiener process? Consider only $d = 1$.

Hölder continuity and non-differentiability

A function $f : (a, b) \longrightarrow \mathbb{R}$ is **Hölder continuous of order α** for some $\alpha \in [0, 1]$ if there is a constant C such that

$$|f(t) - f(s)| \leq C|t - s|^\alpha \qquad \text{for all } s, t \in (a, b).$$

If $\alpha = 1$, then f is Lipschitz continuous.
If $\alpha > 0$, then f is uniformly continuous.
If $\alpha = 0$, then f is bounded.

A path of Brownian motion on a bounded interval is Hölder continuous for any $\alpha \in (0, \frac{1}{2})$ with probability one.
For $\alpha \geq \frac{1}{2}$, however, the path is not Hölder continuous with probability one.
In particular, a path of Brownian motion is nowhere differentiable with probability one.

Proofs: [Ste01], chapter 5

Unbounded total variation

Let $[a, b]$ be an interval and let

$$P_N = (t_n)_{n=0}^N, \qquad a = t_0 < t_1 < \ldots < t_N = b$$

be a partition of $[a, b]$ with $|P_N| = \max_n |t_n - t_{n-1}|$.
Example: equidistant partition, $h = (b-a)/N$, $t_n = a + n \cdot h$. Then, the **total variation** of a function $f : (a, b) \longrightarrow \mathbb{R}$ is

$$TV_{a,b}(f) = \lim_{\substack{N \longrightarrow \infty \\ |P_N| \longrightarrow 0}} \sum_{n=1}^{N} |f(t_n) - f(t_{n-1})|.$$

If f is differentiable and f' is integrable, then (exercise)

$$TV_{a,b}(f) = \int_a^b |f'(t)| \, dt$$

Conversely: If a function f has bounded total variation, then its derivative exists for almost all $x \in [a, b]$.
Consequence: A path of the Wiener process has unbounded total variation with probability one.

Quadratic variation

The quadratic variation of a function $f : (a, b) \longrightarrow \mathbb{R}$ is

$$QV_{a,b}(f) = \lim_{\substack{N \longrightarrow \infty \\ |P_N| \to 0}} \sum_{n=1}^{N} \big(f(t_n) - f(t_{n-1})\big)^2.$$

If f is continuously differentiable, then (exercise)

$$QV_{a,b}(f) = 0.$$

For a path $t \mapsto W_t(\omega)$ with $t \in [0, T]$, however, it can be shown (exercise) that

$$\lim_{\substack{N \longrightarrow \infty \\ |P_N| \longrightarrow 0}} \left\| \sum_{n=1}^{N} \big(W_{t_n}(\omega) - W_{t_{n-1}}(\omega)\big)^2 - T \right\|_{L^2(d\mathbb{P})} = 0,$$

where

$$\|X\|_{L^2(d\mathbb{P})} = \sqrt{\mathbb{E}(X^2)} = \left(\int_{\omega \in \Omega} X^2(\omega)\, d\mathbb{P}(\omega) \right)^{\frac{1}{2}}.$$

By choosing a suitable subsequence, it can be concluded that $QV_{0,t}\big(t \mapsto W_t(\omega)\big) = t$ with probability one.

The standard Brownian filtration

The natural filtration of Brownian motion on $[0, T]$ is given by

$$\{\mathcal{F}_t : t \in [0, T]\}, \qquad \mathcal{F}_t = \sigma\{W_s,\ s \in [0, t]\}$$

(cf. Definition 2.1.5). For technical reasons, however, it is more advantageous to use the **augmented** filtration:

- For fixed t let $\mathcal{Z} = \{S \in \sigma\{W_s,\ s \in [0, T]\} : \mathbb{P}(S) = 0\}$.
- Let $\hat{\mathcal{Z}} = \{\hat{S} \subset \Omega : \hat{S} \subset S \text{ for some } S \in \mathcal{Z}\}$.
- Extend $\mathbb{P}$ by defining $\mathbb{P}(\hat{S}) = 0$ for all $\hat{S} \in \hat{\mathcal{Z}}$.
- Re-define $\mathcal{F}_t$ as the smallest σ-algebra which contains $\sigma\{W_s,\ s \in [0, t]\}$ and $\hat{\mathcal{Z}}$.

This filtration is called the **standard Brownian filtration**.

Consequence: $\mathcal{F}$ fulfills "the usual conditions", i.e.

- If $\hat{S} \in \hat{\mathcal{Z}}$, then $\hat{S} \in \mathcal{F}_0$.
- Right continuity: $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$

2.3 The Itô integral

Itô Kiyoshi 1944

Motivation

Goal: Define stochastic differential equations. Naïve Ansatz:

$$\underbrace{\frac{dX}{dt} = f(t,X)}_{\substack{\text{ordinary differential}\\ \text{equation}}} + \underbrace{g(t,X)Z(t)}_{\text{random noise}}, \qquad Z(t) = ?$$

Apply explicit Euler method: Choose $t \geq 0$ and step-size $N \in \mathbb{N}$, let $h = t/N$, $t_n = n \cdot h$ and define approximations $X_n \approx X(t_n)$ by

$$X_{n+1} = X_n + hf(t_n, X_n) + hg(t_n, X_n)Z(t_n) \qquad (n = 0, 1, 2, \ldots).$$

In the special case $f(t,X) = 0$ and $g(t,X) = 1$, we want that $X_n = W(t_n)$ is the Wiener process, i.e. we postulate that

$$W(t_{n+1}) \stackrel{!}{=} W(t_n) + hZ(t_n).$$

This yields

$$X_{n+1} = X_n + hf(t_n, X_n) + g(t_n, X_n)\Big(W(t_{n+1}) - W(t_n)\Big)$$

and after N steps

$$X_N = X_0 + h\sum_{n=0}^{N-1} f(t_n, X_n) + \sum_{n=0}^{N-1} g(t_n, X_n)\Big(W(t_{n+1}) - W(t_n)\Big). \tag{2.2}$$

Keep t fixed, let $N \longrightarrow \infty$, $h = t/N \longrightarrow 0$. Then, (2.2) should somehow converge to

$$X(t) = X(0) + \int_0^t f(s, X(s))\, ds + \underbrace{\int_0^t g(s, X(s))dW(s)}_{(\star)}. \tag{2.3}$$

Problem: We cannot define $(\star)$ as a pathwise Riemann-Stieltjes integral! When $N \longrightarrow \infty$, the sum

$$\sum_{n=0}^{N-1} g(t_n, X_n(\omega))\Big(W(t_{n+1}, \omega) - W(t_n, \omega)\Big)$$

diverges with probability one, because a path of the Wiener process has unbounded total variation with probability one.

New goal: Define the integral

$$\mathcal{I}_t[u](\omega) = \int_0^t u(s,\omega)\, dW_s(\omega)$$

in a "reasonable" way for the following class of functions.

Definition 2.3.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t : t \in [0,T]\}$ be the standard Brownian filtration. Then, we define $\mathcal{H}^2 = \mathcal{H}^2[0,T]$ to be the class of functions*

$$u = u(t,\omega), \qquad u : [0,T] \times \Omega \longrightarrow \mathbb{R}$$

with the following properties:

- *$(t,\omega) \mapsto u(t,\omega)$ is $(\mathcal{B} \times \mathcal{F})$-measurable.*
- *u is adapted to $\{\mathcal{F}_t : t \in [0,T]\}$, i.e. $u(t,\cdot)$ is $\mathcal{F}_t$-measurable.*
- $\mathbb{E}\left(\int_0^T u^2(t,\omega)\, dt\right) < \infty$

Step 1: Itô integral for elementary functions

Definition 2.3.2 (Elementary functions) *A function $\phi \in \mathcal{H}^2$ is called **elementary** if it is a stochastic step function of the form*

$$\begin{aligned}\phi(t,\omega) &= a_0(\omega)\mathbf{1}_{[0,0]}(t) + \sum_{n=0}^{N-1} a_n(\omega)\mathbf{1}_{(t_n,t_{n+1}]}(t) \\ &= a_0(\omega)\mathbf{1}_{[0,t_1]}(t) + \sum_{n=1}^{N-1} a_n(\omega)\mathbf{1}_{(t_n,t_{n+1}]}(t)\end{aligned}$$

where a_n is $\mathcal{F}_{t_n}$-measurable with $\mathbb{E}(a_n^2) < \infty$. Here and below,

$$\mathbf{1}_{[c,d]}(t) = \begin{cases} 1 & \text{if } t \in [c,d] \\ 0 & \text{else} \end{cases} \tag{2.4}$$

is the indicator function of an interval $[c,d]$.

For $0 \le c < d \le T$, the only reasonable way to define the Itô integral of an indicator function $\mathbf{1}_{(c,d]}$ is

$$\mathcal{I}_T[\mathbf{1}_{(c,d]}](\omega) = \int_0^T \mathbf{1}_{(c,d]}(s)\, dW(s,\omega) = \int_c^d dW(s,\omega) = W(d,\omega) - W(c,\omega).$$

Hence, by linearity, we define the Itô integral of an elementary function by

$$\mathcal{I}_T[\phi](\omega) = \sum_{n=0}^{N-1} a_n(\omega)\big(W(t_{n+1},\omega) - W(t_n,\omega)\big).$$

Lemma 2.3.3 (Itô isometry for elementary functions) *For all elementary functions we have*

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \mathbb{E}\left(\int_0^T \phi^2(t,\omega)\,dt\right)$$

or equivalently

$$\|\mathcal{I}_T[\phi]\|_{L^2(d\mathbb{P})} = \|\phi\|_{L^2(dt\times d\mathbb{P})}$$

with

$$\|\phi\|_{L^2(dt\times d\mathbb{P})} = \left(\int_\Omega \int_0^T \phi^2(t,\omega)\,dt\,d\mathbb{P}\right)^{\frac{1}{2}} = \left(\mathbb{E}\left(\int_0^T \phi^2(t,\omega)\,dt\right)\right)^{\frac{1}{2}}.$$

Proof. Since

$$\phi^2(t,\omega) = a_0^2(\omega)\mathbf{1}_{[0,0]}(t) + \sum_{n=0}^{N-1} a_n^2(\omega)\mathbf{1}_{(t_n,t_{n+1}]}(t)$$

we obtain

$$\mathbb{E}\left(\int_0^T \phi^2(t,\omega)\,dt\right) = \sum_{n=0}^{N-1} \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n) \tag{2.5}$$

for the right-hand side. If we let $\Delta W_n = W(t_{n+1}) - W(t_n)$, then

$$\mathcal{I}_T[\phi]^2 = \left(\sum_{n=0}^{N-1} a_n \Delta W_n\right)^2 = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} a_n a_m \Delta W_n \Delta W_m. \tag{2.6}$$

By definition, the Wiener process has independent increments with $\mathbb{E}(\Delta W_n) = 0$ and $\mathbb{E}(\Delta W_n^2) = \mathbb{V}(\Delta W_n) = t_{n+1} - t_n$. Since a_n is independent of ΔW_n, it follows that

$$\mathbb{E}\left(a_n a_m \Delta W_n \Delta W_m\right) = \begin{cases} 0 & \text{if } n \neq m \\ \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n) & \text{if } n = m \end{cases}$$

and taking the expectation of (2.6) gives

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \sum_{n=0}^{N-1} \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n). \tag{2.7}$$

Comparing (2.5) and (2.7) yields the assertion. ∎

Step 2: Itô integral on $\mathcal{H}^2$

Lemma 2.3.4 *For any $u \in \mathcal{H}^2$ there is a sequence $(\phi_k)_{k\in\mathbb{N}}$ of elementary functions $\phi_k \in \mathcal{H}^2$ such that*

$$\lim_{k\to\infty} \|u - \phi_k\|_{L^2(dt\times d\mathbb{P})} = 0$$

Proof: Section 6.6 in [Ste01].

Let $u \in \mathcal{H}^2$ and let $(\phi_k)_{k\in\mathbb{N}}$ be elementary functions such that

$$u = \lim_{k\to\infty} \phi_k \qquad \text{in } L^2(dt \times d\mathbb{P})$$

as in Lemma 2.3.4. The linearity of $\mathcal{I}_T[\cdot]$ and Lemma 2.3.3 yield

$$\|\mathcal{I}_T[\phi_j] - \mathcal{I}_T[\phi_k]\|_{L^2(d\mathbb{P})} = \|\mathcal{I}_T[\phi_j - \phi_k]\|_{L^2(d\mathbb{P})} = \|\phi_j - \phi_k\|_{L^2(dt\times d\mathbb{P})} \longrightarrow 0$$

for $j, k \longrightarrow \infty$. Hence, $(\mathcal{I}_T[\phi_k])_k$ is a Cauchy sequence in the Hilbert space $L^2(d\mathbb{P})$. Thus, $(\mathcal{I}_T[\phi_k])_k$ converges in $L^2(d\mathbb{P})$, and we can define

$$\mathcal{I}_T[u] = \lim_{k\to\infty} \mathcal{I}_T[\phi_k].$$

The choice of the sequence does not matter: If $(\psi_k)_{k\in\mathbb{N}}$ are elementary functions with $u = \lim_{k\to\infty} \psi_k$ in $L^2(dt \times d\mathbb{P})$, then by Lemma 2.3.3 we obtain for $k \longrightarrow \infty$

$$\begin{aligned}\|\mathcal{I}_T[\phi_k] - \mathcal{I}_T[\psi_k]\|_{L^2(d\mathbb{P})} &= \|\mathcal{I}_T[\phi_k - \psi_k]\|_{L^2(d\mathbb{P})} \\ &= \|\phi_k - \psi_k\|_{L^2(dt\times d\mathbb{P})} \\ &\leq \|\phi_k - u\|_{L^2(dt\times d\mathbb{P})} + \|u - \psi_k\|_{L^2(dt\times d\mathbb{P})} \longrightarrow 0.\end{aligned}$$

Theorem 2.3.5 (Itô isometry) *For all $u \in \mathcal{H}^2$ we have*

$$\|\mathcal{I}_T[u]\|_{L^2(d\mathbb{P})} = \|u\|_{L^2(dt\times d\mathbb{P})}.$$

Proof: Exercise.

Step 3: The Itô integral as a process

So far we have defined the Itô integral $\mathcal{I}_T[u](\omega)$ over the interval $[0, T]$ for **fixed** T. For applications in mathematical finance, however, we want to consider $\{\mathcal{I}_t[u](\omega) : t \in [0, T]\}$ as a stochastic process. Therefore, we let

$$m_t(s, \omega) = \begin{cases} 1 & \text{if } s \in [0, t] \\ 0 & \text{else.} \end{cases}$$

If $u \in \mathcal{H}^2$, then $m_t u \in \mathcal{H}^2$. Can we define $\mathcal{I}_t[u](\omega)$ by $\mathcal{I}_T[m_t u](\omega)$? **No!**

Problem: The integral $\mathcal{I}_T[m_t u](\omega)$ is only defined in $L^2(d\mathbb{P})$. Hence, the value $\mathcal{I}_T[m_t u](\omega)$ is arbitrary on sets of $\mathbb{P}$-measure zero. This is the case for every $t \in [0,T]$, and since the set $[0,T]$ is uncountable, the union

$$\bigcup_{t\in[0,T]} \{Z_t \in \mathcal{F}_t : \mathbb{P}(Z_t) = 0\}$$

(i.e. the set where the process is not well-defined) could be "very large"!

Let $(\phi_k)_{k\in\mathbb{N}}$ be elementary functions with $\lim_{k\to\infty} \phi_k = u$. Define a continuous process by

$$X^{(k)}(t,\omega) = \mathcal{I}_T[m_t\phi_k](\omega).$$

It can be shown that there is a sub-sequence $\left(X^{(k_i)}\right)_i$ such that

$$\max_{t\in[0,T]} |X^{(k_i)}(t,\omega) - X^{(k_j)}(t,\omega)| \longrightarrow 0 \qquad \text{for } i,j \longrightarrow \infty$$

with probability one. Hence, $(X^{(k_i)})_i$ converges uniformly on $[0,T]$ to a continuous process X with probability one. The assumption

$$\lim_{k\to\infty} m_t\phi_{k_i} = m_t u \qquad \text{wrt. } \|\cdot\|_{L^2(dt\times d\mathbb{P})}$$

implies

$$\lim_{k\to\infty} \underbrace{\mathcal{I}_T[m_t\phi_{k_i}]}_{X^{(k_i)}} = \mathcal{I}_T[m_t u] \qquad \text{wrt. } \|\cdot\|_{L^2(d\mathbb{P})},$$

and since it also can be shown that

$$\lim_{k\to\infty} X^{(k_i)}(t,\omega) = X(t,\omega) \qquad \text{wrt. } \|\cdot\|_{L^2(d\mathbb{P})},$$

it follows that for each $t \in [0,T]$

$$X(t,\omega) = \mathcal{I}_T[m_t u](\omega)$$

with probability one. Details: Theorem 6.2 in [Ste01].

Step 4: The Itô integral on $\mathcal{L}^2_{loc}$

So far we have defined the Itô integral for functions $u \in \mathcal{H}^2[0,T]$; cf. Definition 2.3.1. Such functions must satisfy

$$\mathbb{E}\left(\int_0^T u^2(t,\omega)\, dt\right) < \infty, \tag{2.8}$$

and this condition is sometimes too restrictive. With some more work, the Itô integral can be extended to all functions

$$u = u(t,\omega), \qquad u : [0,T]\times\Omega \longrightarrow \mathbb{R}$$

with the following properties:

- $(t, \omega) \mapsto u(t, \omega)$ is $(\mathcal{B} \times \mathcal{F})$-measurable.
- u is adapted to $\{\mathcal{F}_t : t \in [0, T]\}$.
- $\mathbb{P}\left(\int_0^T u^2(t, \omega)\, dt < \infty\right) = 1$

This class is called $\mathcal{L}^2_{loc}[0, T]$. The first two conditions are the same as for $\mathcal{H}^2$, but the third condition is weaker than (2.8). If $y : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, then $u(t, \omega) = y\big(W(t, \omega)\big) \in \mathcal{L}^2_{loc}[0, T]$, because $\omega \mapsto y\big(W(t, \omega)\big)$ is continuous with probability one and hence bounded on $[0, T]$.

Details: Chapter 7 in [Ste01].

Notation

The process X constructed above is called the **Itô integral** of $u \in \mathcal{L}^2_{loc}[0, T]$ and is denoted by

$$X(t, \omega) = \int_0^t u(s, \omega)\, dW(s, \omega).$$

The Itô integral over an arbitrary interval $[a, b] \subset [0, T]$ is defined by

$$\int_a^b u(s, \omega)\, dW(s, \omega) = \int_0^b u(s, \omega)\, dW(s, \omega) - \int_0^a u(s, \omega)\, dW(s, \omega).$$

Alternative notations:

$$\int_a^b u(s, \omega)\, dW(s, \omega) = \int_a^b u(s, \omega)\, dW_s(\omega) = \int_a^b u_s(\omega)\, dW_s(\omega) = \int_a^b u_s\, dW_s$$

Properties of the Itô integral

Lemma 2.3.6 *The Itô integral on $[a, b] \subset [0, T]$ has the following properties:*

1. *Linearity: For all $c \in \mathbb{R}$ and $u, v \in \mathcal{L}^2_{loc}$, we have*

$$\int_a^b \Big(cu(s, \omega) + v(s, \omega)\Big)\, dW_s(\omega) = c \int_a^b u(s, \omega)\, dW_s(\omega) + \int_a^b v(s, \omega)\, dW_s(\omega)$$

   with probability one.

2. $\mathbb{E}\left(\int_a^b u(s, \omega)\, dW_s(\omega)\right) = 0$

3. $\int_a^t u(s,\omega)\, dW_s(\omega)$ *is* $\mathcal{F}_t$*-measurable for* $t \geq a$.

4. *Itô isometry on* $[a,b]$*: For all* $u \in \mathcal{L}^2_{loc}$ *we have*

$$\mathbb{E}\left(\left(\int_a^b u(s,\omega)\, dW_s(\omega)\right)^2\right) = \mathbb{E}\left(\int_a^b u^2(s,\omega)\, ds\right)$$

(cf. Theorem 2.3.5).

Proof. Show these properties for elementary functions and pass to the limit. ∎

Important example.

The Itô integral of $u(s,\omega) = W(s,\omega)$ on $[0,t]$ yields (exercise!)

$$\int_0^t W(s,\omega)\, dW(s,\omega) = \frac{1}{2}W^2(t,\omega) - \frac{1}{2}t. \tag{2.9}$$

The term $-\frac{1}{2}t$ is surprising, because for the corresponding Riemann-Stieltjes integral of a continuously differentiable (deterministic) function $v : [0,t] \longrightarrow \mathbb{R}$ with $v(0) = 0$, we obtain

$$\int_0^t v(s)\, dv(s) = \int_0^t v(s)v'(s)\, ds = \frac{1}{2}\int_0^t \frac{d}{ds}\big(v^2(s)\big)\, ds = \frac{1}{2}v^2(t). \tag{2.10}$$

The reason for the "strange" behaviour of the Itô integral will be revealed in the next subsection.

2.4 Stochastic differential equations and the Itô formula

Definition 2.4.1 (SDE) *A* ***stochastic differential equation (SDE)*** *is an equation of the form*

$$X(t) = X(0) + \int_0^t f\big(s, X(s)\big)\, ds + \int_0^t g\big(s, X(s)\big)\, dW(s). \tag{2.11}$$

The functions $f : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ *and* $g : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ *are called drift and diffusion coefficients, respectively. These functions are typically given while* $X(t) = X(t,\omega)$ *is unknown. The solution* $X(t)$ *is called an* ***Itô process****.*

This equation is actually not a **differential** equation, but an **integral** equation! Often people write

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t$$

as a shorthand notation for (2.11). Some people even "divide by dt" in order to make the equation look like a differential equation, but this is more than audacious since "dW_t/dt" does not make sense.

Two special cases:

- If $g\big(t, X(t)\big) \equiv 0$, then (2.11) is reduced to

$$X(t) = X(0) + \int_0^t f\big(s, X(s)\big)\, ds.$$

  If $X(t)$ is differentiable, this is equivalent to the initial value problem

$$\frac{dX(t)}{dt} = f\big(t, X(t)\big), \qquad X(0) = X_0.$$

- For $f\big(t, X(t)\big) \equiv 0$, $g\big(t, X(t)\big) \equiv 1$ and $X(0) = 0$, (2.11) turns into

$$X(t) = \underbrace{X(0)}_{=0} + \underbrace{\int_0^t f\big(s, X(s)\big)\, ds}_{=0} + \int_0^t \underbrace{g\big(s, X(s)\big)}_{=1}\, dW(s) = W(t) - W(0) = W(t).$$

Computing Riemann integrals via the basic definition is usually very tedious. The fundamental theorem of calculus provides an alternative which is more convenient in most cases. For Itô integrals, the situation is similar: The approximation via elementary functions is rarely used to compute the integral. What is the counterpart of the fundamental theorem of calculus for the Itô integral?

Theorem 2.4.2 (Itô formula) *Let X_t be the solution of the SDE*

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t$$

and let $F(t, x)$ be a function with continuous partial derivatives $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x}$, and $\frac{\partial^2 F}{\partial x^2}$. Then, we have for $Y_t := F(t, X_t)$ that

$$\begin{aligned} dY_t &= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}g^2 dt \\ &= \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial x}f + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}g^2\right)dt + \frac{\partial F}{\partial x}g dW_t. \end{aligned} \tag{2.12}$$

with $f = f(t, X_t)$, $g = g(t, X_t)$, $\dfrac{\partial F}{\partial x} = \dfrac{\partial F(t, X_t)}{\partial x}$, and so on.

Notation. From now on, the partial derivatives of some function $u(t,x)$ will be denoted by

$$\partial_t u := \frac{\partial u}{\partial t}, \qquad \partial_x u := \frac{\partial u}{\partial x}, \qquad \partial_x^2 u := \frac{\partial^2 u}{\partial x^2}$$

and so on. Evaluations of the derivatives of F are to be understood in the sense of, e.g.,

$$\partial_x F(s, X_s) := \partial_x F(t,x)\big|_{(t,x)=(s,X_s)}$$

and so on.

Remarks:

1. The Itô formula can be considered as a stochastic chain rule, but the term $\frac{1}{2}\partial_x^2 F \cdot g^2 dt$ is surprising since such a term does not appear in the chain rule for deterministic functions: If X_t and $F(t,x)$ are smooth deterministic fuctions, then the derivative of $t \mapsto F(t, X_t)$ is

$$\partial_t F(t, X_t) + \partial_x F(t, X_t) \cdot \frac{dX_t}{dt}, \qquad \text{i.e. } dF = \partial_t F dt + \partial_x F dX_t.$$

2. Let $f(t, X_t) = 0$, $g(t, X_t) = 1$, $X_t = W_t$ and suppose that $F(t,x) = F(x)$. Then, the Itô formula yields for $Y_t := F(W_t)$ that

$$dY_t = F'(W_t)dW_t + \frac{1}{2}F''(W_t)dt$$

   which is the shorthand notation for

$$F(W_t) = F(W_0) + \int_0^t F'(W_s)dW_s + \frac{1}{2}\int_0^t F''(W_s)ds.$$

   This can be seen as a counterpart of the fundamental theorem of calculus. Again, the last term is surprising, because for a suitable deterministic function $v(t) = v_t$ we obtain

$$F(v_t) = F(v_0) + \int_0^t F'(v_s)dv_s.$$

Sketch of the proof of Theorem 2.4.2.

- Equation (2.12) is the shorthand notation for

$$Y_t = Y_0 + \int_0^t \left(\partial_t F(s, X_s) + \partial_x F(s, X_s) \cdot f(s, X_s) + \frac{1}{2}\partial_x^2 F(s, X_s) \cdot g^2(s, X_s)\right) ds + \int_0^t \partial_x F(s, X_s) \cdot g(s, X_s) dW_s$$

Assume that F is twice continuously differentiable with bounded partial derivatives. (Otherwise F can be approximated by such functions with uniform convergence on compact subsets of $[0,\infty)\times\mathbb{R}$.)
Assume that $(t,\omega)\mapsto f(t,X_t(\omega))$ and $(t,\omega)\mapsto g(t,X_t(\omega))$ are elementary functions. (Otherwise approximate by elementary functions.) Hence, there is a partition $0 = t_0 < t_1 < \ldots < t_N = t$ such that

$$f(t,X_t(\omega)) = f(0,X_0(\omega))\mathbf{1}_{[0,t_1]}(t) + \sum_{n=1}^{N-1} f(t_n,X_{t_n}(\omega))\mathbf{1}_{(t_n,t_{n+1}]}(t)$$

and the same equation with f replaced by g.

- Notation: For the rest of the proof, we define

$$f^{(n)} := f(t_n,X_{t_n}), \qquad F^{(n)} := F(t_n,X_{t_n}),$$
$$g^{(n)} := g(t_n,X_{t_n}), \qquad \partial_t F^{(n)} := \partial_t F(t_n,X_{t_n})$$

and so on, and

$$\Delta t_n = t_{n+1}-t_n, \qquad \Delta X_n = X_{t_{n+1}} - X_{t_n}, \qquad \Delta W_n = W_{t_{n+1}} - W_{t_n}.$$

Since f and g are elementary functions, we have

$$X_{t_n} = X_0 + \int_0^{t_n} f(s,X_s)\,ds + \int_0^{t_n} g(s,X_s)\,dW_s$$
$$X_{t_n} = X_0 + \sum_{k=0}^{n-1} \underbrace{f(t_k,X_{t_k})}_{f^{(k)}}\Delta t_k + \sum_{k=0}^{n-1} \underbrace{g(t_k,X_{t_k})}_{g^{(k)}}\Delta W_k.$$

and hence

$$\Delta X_n = X_{t_{n+1}} - X_{t_n} = f^{(n)}\Delta t_n + g^{(n)}\Delta W_n.$$

- Telescoping sum:

$$Y_t = Y_{t_N} = Y_0 + \sum_{n=0}^{N-1}(Y_{t_{n+1}} - Y_{t_n}) = Y_0 + \sum_{n=0}^{N-1}\left(F^{(n+1)} - F^{(n)}\right)$$

Apply Taylor's theorem:

$$\begin{aligned} &F^{(n+1)} - F^{(n)} \\ &= \partial_t F^{(n)}\cdot\Delta t_n + \partial_x F^{(n)}\cdot\Delta X_n + \frac{1}{2}\partial_t^2 F^{(n)}\cdot(\Delta t_n)^2 + \partial_t\partial_x F^{(n)}\cdot\Delta t_n\Delta X_n \\ &\quad + \frac{1}{2}\partial_x^2 F^{(n)}\cdot(\Delta X_n)^2 + R_n(\Delta t_n,\Delta X_n) \end{aligned}$$

with a remainder term R_n. Insert this into the telescoping sum.

- Consider the limit $N \longrightarrow \infty$, $\Delta t_n \longrightarrow 0$ with respect to $\|\cdot\|_{L^2(d\mathbb{P})}$. For the first two terms, this yields

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t F^{(n)} \cdot \Delta t_n = \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t F(t_n, X_{t_n}) \cdot \Delta t_n \qquad = \int_0^t \partial_t F(s, X_s)\, ds$$

and

$$\begin{aligned}
&\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot \Delta X_n \\
&= \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot f^{(n)} \Delta t_n + \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot g^{(n)} \Delta W_n \\
&= \int_0^t \partial_x F(s, X_s) \cdot f(s, X_s)\, ds + \int_0^t \partial_x F(s, X_s) \cdot g(s, X_s)\, dW_s.
\end{aligned}$$

- Next, we investigate the "$\partial_x^2 F^{(n)}$ term". Since

$$(\Delta X_n)^2 = \left(f^{(n)} \Delta t_n + g^{(n)} \Delta W_n \right)^2$$

we have

$$\begin{aligned}
\frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (\Delta X_n)^2 =& \frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(f^{(n)} \right)^2 (\Delta t_n)^2 && (2.13)\\
&+ \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot f^{(n)} g^{(n)} \Delta t_n \Delta W_n && (2.14)\\
&+ \frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(g^{(n)} \right)^2 (\Delta W_n)^2. && (2.15)
\end{aligned}$$

For the right-hand side of (2.13), we obtain

$$\left\| \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(f^{(n)} \right)^2 (\Delta t_n)^2 \right\|_{L^2(d\mathbb{P})}^2 = \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(f^{(n)} \right)^2 (\Delta t_n)^2 \right)^2 \right] \longrightarrow 0.$$

With the abbreviation $\alpha^{(n)} := \partial_x^2 F^{(n)} \cdot f^{(n)} g^{(n)}$ we obtain for the right-hand side of (2.14) that

$$\begin{aligned}
\left\| \sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right\|_{L^2(d\mathbb{P})}^2 &= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right)^2 \right] \\
&= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}\left(\alpha^{(n)} \alpha^{(m)} \Delta W_n \Delta W_m \right) \Delta t_n \Delta t_m.
\end{aligned}$$

Since

$$\mathbb{E}\left(\alpha^{(n)}\alpha^{(m)}\Delta W_n \Delta W_m\right) = \mathbb{E}\left(\alpha^{(n)}\alpha^{(m)}\Delta W_n\right)\underbrace{\mathbb{E}\left(\Delta W_m\right)}_{=0} = 0$$

for $n < m$ and similar for $m < n$, only the terms with $n = m$ have to be considered, which yields

$$\left\|\sum_{n=0}^{N-1}\alpha^{(n)}\Delta t_n \Delta W_n\right\|^2_{L^2(d\mathbb{P})} = \sum_{n=0}^{N-1}\mathbb{E}\left((\alpha^{(n)})^2\right)(\Delta t_n)^2\underbrace{\mathbb{E}\left[(\Delta W_n)^2\right]}_{=\Delta t_n} \longrightarrow 0.$$

The third term (2.15), however, has a non-zero limit: We show that

$$\lim_{N\to\infty}\frac{1}{2}\sum_{n=0}^{N-1}\partial_x^2 F^{(n)}\cdot\left(g^{(n)}\right)^2(\Delta W_n)^2 = \frac{1}{2}\int_0^t \partial_x^2 F(s,X_s)\cdot\left(g(s,X_s)\right)^2 ds$$

which yields the strange additional term in the Itô formula. With the abbreviation $\beta^{(n)} = \frac{1}{2}\partial_x^2 F^{(n)}\cdot\left(g^{(n)}\right)^2$ we have

$$\begin{aligned}
&\left\|\sum_{n=0}^{N-1}\beta^{(n)}\left((\Delta W_n)^2 - \Delta t_n\right)\right\|^2_{L^2(d\mathbb{P})} \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N-1}\beta^{(n)}\left((\Delta W_n)^2 - \Delta t_n\right)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\beta^{(n)}\beta^{(m)}\left((\Delta W_n)^2 - \Delta t_n\right)\left((\Delta W_m)^2 - \Delta t_m\right)\right].
\end{aligned}$$

For $n < m$ we have

$$\begin{aligned}
&\mathbb{E}\left[\beta^{(n)}\beta^{(m)}\left((\Delta W_n)^2 - \Delta t_n\right)\left((\Delta W_m)^2 - \Delta t_m\right)\right] \\
&= \mathbb{E}\left[\beta^{(n)}\beta^{(m)}\left((\Delta W_n)^2 - \Delta t_n\right)\right]\underbrace{\mathbb{E}\left[\left((\Delta W_m)^2 - \Delta t_m\right)\right]}_{=0} = 0
\end{aligned}$$

and vice versa for $n > m$. Hence, only the terms with $n = m$ have to be considered, and we obtain

$$\begin{aligned}
\left\|\sum_{n=0}^{N-1}\beta^{(n)}\left((\Delta W_n)^2 - \Delta t_n\right)\right\|^2_{L^2(d\mathbb{P})} &= \mathbb{E}\left[\sum_{n=0}^{N-1}\left(\beta^{(n)}\right)^2\left((\Delta W_n)^2 - \Delta t_n\right)^2\right] \\
&= \sum_{n=0}^{N-1}\mathbb{E}\left[\left(\beta^{(n)}\right)^2\right]\underbrace{\mathbb{E}\left[\left((\Delta W_n)^2 - \Delta t_n\right)^2\right]}_{\longrightarrow 0}
\end{aligned}$$

according to Exercise 5.

- With essentially the same arguments, it can be shown that

$$\lim_{N\to\infty} \frac{1}{2} \sum_{n=0}^{N-1} \partial_t^2 F^{(n)} \cdot (\Delta t_n)^2 = 0$$
$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t \partial_x F^{(n)} \cdot \Delta t_n \Delta X_n = 0$$

  and that the remainder term from the Taylor expansion can be neglected when the limit is taken. ■

Example 1. Consider again the integral

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

(cf. 2.3). $X_t := W_t$ solves the SDE with $f(t, X_t) \equiv 0$ and $g(t, X_t) \equiv 1$. For

$$F(t,x) = x^2, \qquad\qquad Y_t = F(t, X_t) = X_t^2 = W_t^2$$

the Itô formula

$$dY_t = \left(\partial_t F + \partial_x F \cdot f + \frac{1}{2} \partial_x^2 F \cdot g^2 \right) dt + \partial_x F \cdot g \, dW_t$$

yields

$$d(W_t^2) = 0 + 0 + \frac{1}{2} \cdot 2 \cdot 1^2 \, dt + 2W_t \cdot 1 \, dW_t = dt + 2W_t \, dW_t$$
$$\implies \qquad W_t \, dW_t = \frac{1}{2} \left(d(W_t^2) - dt \right)$$

This means that

$$\int_0^t W_s \, dW_s = \frac{1}{2} \int_0^t 1 \, d(W_s^2) - \frac{1}{2} \int_0^t 1 \, ds = \frac{1}{2} W_t^2 - \frac{1}{2} t.$$

Hence, we have obtained the same result as before, but much easier.

Example 2. The solution of the SDE

$$dY_t = \mu Y_t \, dt + \sigma Y_t \, dW_t$$

with constants $\mu, \sigma \in \mathbb{R}$ is given by

$$Y_t = Y_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right).$$

This process is called a **geometric Brownian motion** and is often used in mathematical finance to model stock prices (see below).

Proof. Let $f(t, X_t) \equiv 0$, $g(t, X_t) \equiv 1$, $X_t = W_t$ as before, but now with

$$F(t, x) = \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x\right)$$

and derivatives

$$\partial_t F(t,x) = \left(\mu - \frac{\sigma^2}{2}\right)F(t,x), \qquad \partial_x^i F(t,x) = \sigma^i F(t,x), \qquad i \in \{1,2\}.$$

Hence, the Itô formula applied to $Y_t = F(t, X_t) = F(t, W_t)$ yields

$$\begin{aligned} dY_t &= \left(\left(\mu - \frac{\sigma^2}{2}\right)Y_t + 0 + \frac{1}{2}\sigma^2 Y_t \cdot 1\right) dt + \sigma Y_t \cdot 1^2 \, dW_t \\ &= \mu Y_t \, dt + \sigma Y_t \, dW_t. \end{aligned}$$

■

2.5 Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Recall the definition of the expectation:

$$\mathbb{E}(X) = \int_\Omega X(\omega) \, d\mathbb{P}(\omega)$$

Definition 2.5.1 (conditional expectation) *Let X be an integrable random variable, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then, Y is a* ***conditional expectation*** *of X with respect to $\mathcal{G}$ if Y is $\mathcal{G}$-measurable and if*

$$\begin{aligned} & \mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A) && \text{for all } A \in \mathcal{G}, \\ \Leftrightarrow \qquad & \int_A X(\omega) \, d\mathbb{P}(\omega) = \int_A Y(\omega) \, d\mathbb{P}(\omega) && \text{for all } A \in \mathcal{G}. \end{aligned}$$

In this case, we write $Y = \mathbb{E}(X \mid \mathcal{G})$.

"This definition is not easy to love. Fortunately, love is not required."
J.M. Steele in [Ste01], p. 45.

Interpretation. $\mathbb{E}(X \mid \mathcal{G})$ is a random variable on $(\Omega, \mathcal{G}, \mathbb{P})$ and hence on $(\Omega, \mathcal{F}, \mathbb{P})$, too. Roughly speaking, $\mathbb{E}(X \mid \mathcal{G})$ is the best approximation of X detectable by the events in $\mathcal{G}$. The more $\mathcal{G}$ is refined, the better $\mathbb{E}(X \mid \mathcal{G})$ approximates X.

Examples.

1. If $\mathcal{G} = \{\Omega, \emptyset\}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$.
2. If $\mathcal{G} = \mathcal{F}$, then $\mathbb{E}(X \mid \mathcal{G}) = X$.
3. If $F \in \mathcal{F}$ with $\mathbb{P}(F) > 0$ and

$$\mathcal{G} = \{\emptyset, F, \Omega \setminus F, \Omega\}$$

then it can be shown that

$$\mathbb{E}(X \mid \mathcal{G})(\omega) = \begin{cases} \dfrac{1}{\mathbb{P}(F)} \displaystyle\int_F X d\mathbb{P} & \text{if } \omega \in F \\ \dfrac{1}{\mathbb{P}(\Omega \setminus F)} \displaystyle\int_{\Omega \setminus F} X d\mathbb{P} & \text{if } \omega \in \Omega \setminus F \end{cases}$$

(exercise).

Lemma 2.5.2 (Properties of the conditinal expectation) *For all integrable random variables X and Y and all sub-σ-algebras $\mathcal{G} \subset \mathcal{F}$, the conditinal expectation has the following roperties:*

- *Linearity:* $\mathbb{E}(X + Y \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G}) + \mathbb{E}(Y \mid \mathcal{G})$
- *Positivity: If* $X \geq 0$, *then* $\mathbb{E}(X \mid \mathcal{G}) \geq 0$.
- *Tower property: If* $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ *are sub-σ-algebras, then*

$$\mathbb{E}\Big(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}\Big) = \mathbb{E}(X \mid \mathcal{H})$$

- $\mathbb{E}\Big(\mathbb{E}(X \mid \mathcal{G})\Big) = \mathbb{E}(X)$
- *Factorization property: If* Y *is* $\mathcal{G}$*-measurable and* $|XY|$ *and* $|Y|$ *are integrable, then*

$$\mathbb{E}(XY \mid \mathcal{G}) = Y\mathbb{E}(X \mid \mathcal{G})$$

Proof: Exercise.

Definition 2.5.3 (martingale) *Let X_t be a stochastic process which is adapted to to a filtration $\{\mathcal{F}_t : t \geq 0\}$ of $\mathcal{F}$. If*

1. $\mathbb{E}(|X_t|) < \infty$ *for all* $0 \leq t < \infty$, *and*
2. $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ *for all* $0 \leq s \leq t < \infty$,

then X_t is called a ***martingale****. A martingale X_t is called continuous if there is a set $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that the path $t \mapsto X_t(\omega)$ is continuous for all $\omega \in \Omega_0$.*

Interpretation: A martingale models a fair game. Observing the game up to time s does not give any advantage for future times.

Examples. Each of the following processes is a continuous martingale with respect to the standard Brownian filtration:

$$W_t, \qquad W_t^2 - t, \qquad \exp\left(\alpha W_t - \frac{\alpha^2}{2}t\right)$$

Proof: Exercise.

Theorem 2.5.4 (The Itô integral as a martingale) *The Itô integral*

$$X(t,\omega) = \int_0^t u(s,\omega)\, dW(s,\omega).$$

of a function $u \in \mathcal{H}^2$ is a continuous martingale with respect to the standard Brownian filtration.

Proof: Theorem 6.2, p. 83 in [Ste01].

Remark: If $u \in \mathcal{L}^2_{loc}$, then the Itô integral is only a local martingale; cf. Proposition 7.7 in [Ste01].

2.6 The Feynman-Kac formula

Let X_t be the solution of the SDE

$$dX_t = f(t,X_t)dt + g(t,X_t)dW_t, \qquad t \in [t_0, t_{\text{end}}], \qquad X_{t_0} = \xi$$

with suitable functions f and g. Let $u(t,x)$ be the solution of the (deterministic) partial differential equation (PDE)

$$\partial_t u(t,x) + f(t,x)\partial_x u(t,x) + \frac{1}{2}g^2(t,x)\partial_x^2 u(t,x) = 0, \qquad t \in [t_0, t_{\text{end}}], \qquad x \in \mathbb{R}$$

with terminal condition

$$u(t_{\text{end}}, x) = \psi(x)$$

for some $\psi : \mathbb{R} \longrightarrow \mathbb{R}$. Apply the Itô formula (Theorem 2.4.2) to $u(t, X_t)$:

$$du(t,X_t) = \underbrace{\left(\partial_t u(t,x) + f(t,x)\partial_x u(t,x) + \frac{1}{2}g^2(t,x)\partial_x^2 u(t,x)\right)}_{=0} dt + g(t,x)\partial_x u(t,x)\, dW_t$$

Equivalent:

$$\underbrace{u(t_{\text{end}}, X_{t_{\text{end}}})}_{\psi(X_{t_{\text{end}}})} = u(t_0, \underbrace{X_{t_0}}_{\xi}) + \int_{t_0}^{t_{\text{end}}} g(t, X_t)\partial_x u(t, X_t)\, dW_t$$

Taking the expectation and applying Lemma 2.3.6 yields the **Feynman-Kac formula** (Richard Feynman, Mark Kac)

$$\mathbb{E}\Big(\psi(X_{tend})\Big) = u(t_0, \xi).$$

Remark: This derivation is informal, because we have tacitly assumed that all terms exist. Se, e.g., Chapter 15 in [Ste01] for a correct proof.

2.7 Extension to higher dimensions

In order to model options on several underlying assets (e.g. basket options), we have to consider vector-valued Itô integrals and SDEs. A d-dimensional SDE takes the form

$$X_j(t) = X_j(0) + \int_0^t f_j(s, X(s))\, ds + \sum_{k=1}^{m} \int_0^t g_{jk}(s, X(s))\, dW_k(s) \qquad (j = 1, \ldots, d) \tag{2.16}$$

for $d, m \in \mathbb{N}$ and suitable functions

$$f_j : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}, \qquad g_{jk} : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}.$$

$W_1(s), \ldots, W_m(s)$ are one-dimensional scalar Wiener processes which are pairwise independent. (2.16) is equivalent to

$$X(t) = X(0) + \int_0^t f(s, X(s))\, ds + \int_0^t g(s, X(s))\, dW(s) \tag{2.17}$$

with vectors

$$W(t) = \big(W_1(t), \ldots, W_m(t)\big)^T \in \mathbb{R}^m$$
$$f(t, x) = \big(f_1(t, x), \ldots, f_d(t, x)\big)^T \in \mathbb{R}^d$$

and a matrix

$$g(t, x) = \begin{pmatrix} g_{11}(t, x) & \cdots & g_{1m}(t, x) \\ \vdots & & \vdots \\ g_{d1}(t, x) & \cdots & g_{dm}(t, x) \end{pmatrix} \in \mathbb{R}^{d \times m}$$

Theorem 2.7.1 (Multi-dimensional Itô formula) *Let X_t be the solution of the SDE (2.17) and let $F : [0,\infty) \times \mathbb{R}^d \longrightarrow \mathbb{R}^n$ be a function with continuous partial derivatives $\partial_t F$, $\partial_{x_j} F$, and $\partial_{x_j}\partial_{x_k} F$. Then, the process $Y(t) := F(t, X_t)$ satisfies*

$$\begin{aligned} dY_\ell(t) &= \partial_t F_\ell(t, X_t)\, dt \\ &\quad + \sum_{i=1}^{d} \partial_{x_i} F_\ell(t, X_t) \cdot f_i(t, X_t)\, dt \\ &\quad + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \partial_{x_i}\partial_{x_j} F_\ell(t, X_t) \cdot \left(\sum_{k=1}^{m} g_{ik}(t, X_t) g_{jk}(t, X_t) \right) dt \\ &\quad + \sum_{i=1}^{d} \partial_{x_i} F_\ell(t, X_t) \cdot \sum_{k=1}^{m} g_{ik}(t, X_t)\, dW_k \end{aligned}$$

or equivalently

$$dY_\ell = \left\{ \partial_t F_\ell + f^T \nabla F_\ell + \frac{1}{2} \mathrm{tr}\Big(g^T (\nabla^2 F_\ell) g \Big) \right\} dt + (\nabla F_\ell)^T g\, dW(t)$$

where ∇F_ℓ is the gradient and $\nabla^2 F_\ell$ is the Hessian of F_ℓ, and where $\mathrm{tr}(A) = \sum_{j=1}^{m} a_{jj}$ is the trace of a matrix $A = (a_{ij})_{i,j} \in \mathbb{R}^{m\times m}$.

Proof: Similar to the case $d = m = 1$.

Final remarks.

1. **Existence and uniqueness of solutions.** Ordinary differential equations can have multiple solutions with the same initial value, and solutions do not necessarily exist for all times. Hence, we cannot expect that every SDE has a unique solution. As in the ODE case, however, existence and uniqueness can be shown under certain assumptions concerning the coefficients f and g. See 4.5 in [KP99] for details.
2. **Itô vs. Stratonovich.** The Itô integral is not the only stochastic integral, and the Stratonovich integral is a famous alternative. The Stratonovich integral has the advantage that the ordinary chain rule remains valid, i.e. the additional term in the Itô formula does not appear when the Stratonovich integral is used. Their disadvantage is the fact that Stratonovich integrals are not martingales, whereas Itô integrals are. Stratonovich integrals can be transformed into Itô integrals and vice versa. See 3.1, 3.3 in [Øks03] and 3.5, 4.9 in [KP99].
   Actually, the SDEs (2.11) or (2.16) should be called "Itô SDE" or "SDE of the Itô type". Since only Itô SDEs and no Stratonovich SDEs will appear in this lecture, however, we simply use the term "SDE" for "Itô SDE".

Chapter 3

The Black-Scholes equation

References: [BK04, GJ10, Sey09]

Goal: Find equations to determine the value of an option on a single underlying asset. Throughout this chapter, we make the assumptions (A1)-(A5) from 1.3 unless otherwise stated.

3.1 Geometric Brownian motion

First step: Model the price of the underlying by a suitable process S_t.
For the value of a bond with interest rate $r > 0$, we have $B_t = B_0 e^{rt}$. Try to "stochastify" this equation with a Wiener process in order to model the underlying.

First attempt: $S_t = S_0 e^{at} + \sigma W_t$ for some $a, \sigma \in \mathbb{R}$. Problem: S_t can have negative values. Not good.

Second attempt: For the bond we have

$$\ln B_t = \ln B_0 + rt.$$

which motivates the ansatz

$$\ln S_t = \ln S_0 + at + \sigma W_t$$

for some $a, \sigma \in \mathbb{R}$ to model the underlying. The parameter σ is called the **volatility**. Applying $\exp(\dots)$ gives

$$S_t = S_0 \exp(at + \sigma W_t)$$

and hence $S_t \geq 0$ if $S_0 \geq 0$. In fact, S_t is the geometric Brownian motion from 2.4 and solves the SDE

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

with $\mu = a + \sigma^2/2$. Interpretation:

$$\begin{array}{rcccc} \dfrac{dS_t}{S_t} & = & \mu dt & + & \sigma dW_t \\ \text{relative change} & = & \text{deterministic trend} & + & \text{random fluctuations} \end{array}$$

Lemma 3.1.1 (moments of GBM) *The geometric Brownian motion*

$$S_t = S_0 \exp(at + \sigma W_t), \qquad a = \mu - \sigma^2/2$$

with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}$ and fixed (deterministic) initial value X_0 has the following properties:

1. $\mathbb{E}(S_t) = S_0 e^{\mu t}$
2. $\mathbb{E}(S_t^2) = S_0^2 e^{(2\mu+\sigma^2)t}$
3. $\mathbb{V}(S_t) = S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)$

Proof: Exercise.

Definition 3.1.2 (log-normal distribution) *A vector-valued random variable $X(\omega) \in \mathbb{R}^d$ is **log-normal** (=log-normally distributed) if $\ln X = (\ln X_1, \ldots, \ln X_d)^T \in \mathbb{R}^d$ is normally distributed, i.e. $\ln X \sim \mathcal{N}(\xi, \Sigma)$ for some $\xi \in \mathbb{R}^d$ and a symmetric, positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$. The expectation and the covariance matrix have the entries*

$$\begin{aligned} \mathbb{E}_i(X) &= e^{\xi_i + \Sigma_{ii}/2}, \\ \mathbb{V}(X)_{ij} &= \mathbb{E}\Big(\big(\mathbb{E}_i(X) - X_i\big)\big(\mathbb{E}_j(X) - X_j\big)\Big) = e^{\xi_i+\xi_j+\frac{1}{2}(\Sigma_{ii}+\Sigma_{jj})}\left(e^{\Sigma_{ij}} - 1\right). \end{aligned}$$

For $d = 1$ and $\Sigma = \sigma^2$ the corresponding density is

$$\phi(x) = \phi(x, \xi, \sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\dfrac{(\ln x - \xi)^2}{2\sigma^2}\right) & \text{if } x > 0 \\ 0 & \text{else.} \end{cases}$$

Proof: Exercise.

Example. The (one-dimensional) geometric Brownian motion

$$S_t = S_0 \exp(at + \sigma W_t)$$

is log-normal, because

$$\ln S_t = \ln S_0 + at + \sigma W_t \sim \mathcal{N}(\ln S_0 + at, \sigma^2 t)$$

3.2 Derivation of the Black-Scholes equation

Situation: S_t value of an underlying, B_t value of a bond.

Goal: Determine the fair price V_t of an option.

Replication strategy: Consider a portfolio containing $a_t \in \mathbb{R}$ underlyings and $b_t \in \mathbb{R}$ bonds such that

$$V_t = a_t S_t + b_t B_t$$

(cf. 1.5). Assume that the portfolio is self-financing: no cash inflow or outflow, i.e. buying an item must be financed by selling another one. Consequence:

$$\begin{aligned} V_{t+\delta} - V_t &= (a_{t+\delta} S_{t+\delta} - a_t S_t) + (b_{t+\delta} B_{t+\delta} - b_t B_t) \\ &\approx a_t (S_{t+\delta} - S_t) + b_t (B_{t+\delta} - B_t) \end{aligned}$$

for all $t \geq 0$ and small $\delta > 0$. For $\delta \longrightarrow 0$ we obtain (in an integral sense)

$$dV_t = a_t dS_t + b_t dB_t.$$

Now suppose that

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \qquad dB_t = r B_t dt \tag{3.1}$$

with $\mu, \sigma, r \in \mathbb{R}$. This yields

$$\begin{aligned} dV_t &= a_t\big(\mu S_t dt + \sigma S_t dW_t\big) + b_t\big(r B_t dt\big) \\ &= \big(a_t \mu S_t + b_t r B_t\big) dt + a_t \sigma S_t dW_t. \end{aligned} \tag{3.2}$$

Now assume that the value of the option is a function of t and S_t, i.e. $V_t = V(t, S_t)$. Apply the Itô formula:

$$\begin{aligned} dV(t, S_t) = &\left(\partial_t V(t, S_t) + \partial_S V(t, S_t) \cdot \mu S_t + \frac{1}{2} \partial_S^2 V(t, S_t) \cdot \sigma^2 S_t^2\right) dt \\ &+ \partial_S V(t, S_t) \cdot \sigma S_t dW_t \end{aligned} \tag{3.3}$$

Equating the dW_t-terms in (3.2) and (3.3) yields

$$a_t = \partial_S V(t, S_t),$$

while equating the dt-terms yields

$$\begin{aligned} & a_t \mu S_t + b_t r B_t = \partial_t V(t, S_t) + \partial_S V(t, S_t) \cdot \mu S_t + \frac{1}{2} \partial_S^2 V(t, S_t) \cdot \sigma^2 S_t^2 \\ \Longrightarrow \quad & b_t r B_t = \partial_t V(t, S_t) + \frac{1}{2} \partial_S^2 V(t, S_t) \cdot \sigma^2 S_t^2 \\ \Longrightarrow \quad & b_t = \frac{1}{B_t r} \left(\partial_t V(t, S_t) + \frac{1}{2} \partial_S^2 V(t, S_t) \cdot \sigma^2 S_t^2\right) \end{aligned}$$

if we assume that $B_t r \neq 0$. The formulas for a_t and b_t yield

$$V(t,S_t) = a_t S_t + b_t B_t = \partial_S V(t,S_t)\cdot S_t + \frac{1}{B_t r}\left(\partial_t V(t,S_t) + \frac{1}{2}\partial_S^2 V(t,S_t)\cdot \sigma^2 S_t^2\right) B_t.$$

Since this is true for every value of S_t, we can consider $S = S_t$ as a parameter. Multiplying with r yields the **Black-Scholes equation**

$$\partial_t V(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0.$$

Fischer Black and Myron Scholes 1973, Robert Merton 1973
Nobel Prize in Economics 1997

The Black-Scholes equation is a partial differential equation (PDE): It involves partial derivatives with respect to t and S. This PDE must be solved **backwards** in time: instead of an initial condition, we have the **terminal condition**

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

where T is the expiration time and $\psi(S)$ is the payoff function, i.e. $\psi(S) = (S-K)^+$ for a call and $\psi(S) = (K-S)^+$ for a put.
The Black-Scholes equation must be solved for $S \in \mathbb{R}_+ := [0,\infty)$ because only non-negative prices make sense. At the boundary $S = 0$, no boundary conditions are required, because

$$\lim_{S\to 0} \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) = 0, \qquad \lim_{S\to 0} rS\partial_S V(t,S) = 0$$

if V is sufficiently smooth. For $S = 0$, we obtain

$$0 = \partial_t V(t,0) - rV(t,0) \qquad \Longrightarrow \qquad V(t,0) = e^{-r(T-t)} V(T,0). \tag{3.4}$$

This yields $V(t,0) = 0$ for calls and $V(t,0) = e^{-r(T-t)}K$ for puts.

Remark. Surprisingly, the parameter μ from (3.1) does **not** appear in the Black-Scholes equation. A similar observation has been made for the simple discrete model from 1.5.

3.3 Black-Scholes formulas

First goal: Solve the Black-Scholes equation for an European call, i.e.

$$\begin{aligned} &\partial_t V(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0 \qquad t\in[0,T], \quad S>0 \\ &V(T,S) = (S-K)^+ \end{aligned}$$

with parameters $r, \sigma, K, T > 0$.

Step 1: Transformation to the heat equation

Define new variables:

$$\begin{aligned} x(S) &= \ln(S/K) && x : (0,\infty) \longrightarrow (-\infty,\infty) \\ \tau(t) &= \frac{\sigma^2}{2}(T-t) && \tau : [0,T] \longrightarrow [0,\sigma^2 T/2] \\ w(\tau,x) &= \frac{V(t,S)}{K} && w : [0,\sigma^2 T/2]\times(-\infty,\infty) \longrightarrow \mathbb{R} \end{aligned}$$

Derivatives in new variables:

$$\begin{aligned} \partial_t V(t,S) &= K\partial_t w(\tau,x) = K\partial_\tau w(\tau,x)\frac{d\tau}{dt} = -K\frac{\sigma^2}{2}\partial_\tau w(\tau,x) \\ \partial_S V(t,S) &= K\partial_x w(\tau,x)\frac{dx}{dS} = \frac{K}{S}\partial_x w(\tau,x) \qquad \left(\text{because } \frac{dx}{dS} = \frac{1}{S/K}\cdot\frac{1}{K} = \frac{1}{S}\right) \\ \partial_S^2 V(t,S) &= \quad \ldots \quad = \frac{K}{S^2}\left(\partial_x^2 w(\tau,x) - \partial_x w(\tau,x)\right) \end{aligned}$$

Insert into the Black-Scholes equation:

$$\begin{aligned} 0 &= \partial_t V(t,S) + \frac{\sigma^2}{2}S^2\partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) \\ &= -K\frac{\sigma^2}{2}\partial_\tau w(\tau,x) + \frac{\sigma^2}{2}S^2\frac{K}{S^2}\left(\partial_x^2 w(\tau,x) - \partial_x w(\tau,x)\right) + rS\frac{K}{S}\partial_x w(\tau,x) - rKw(\tau,x) \end{aligned}$$

Divide by $K\frac{\sigma^2}{2}$:

$$\partial_\tau w(\tau,x) = \partial_x^2 w(\tau,x) - \partial_x w(\tau,x) + c\partial_x w(\tau,x) - cw(\tau,x)$$

with $c := 2r/\sigma^2$. Next, we eliminate the last three terms. Ansatz:

$$u(\tau,x) = e^{-\alpha x-\beta\tau}w(\tau,x), \qquad \alpha,\beta\in\mathbb{R}$$

Substitute:

$$\begin{aligned} \partial_\tau u(\tau,x) &= -\beta u(\tau,x) + e^{-\alpha x-\beta\tau}\partial_\tau w(\tau,x) \\ &= -\beta u(\tau,x) + e^{-\alpha x-\beta\tau}\left(\partial_x^2 w(\tau,x) + (c-1)\partial_x w(\tau,x) - cw(\tau,x)\right) \end{aligned}$$

Since

$$\begin{aligned} \partial_x w(\tau,x) &= \partial_x\left(e^{\alpha x+\beta\tau}u(\tau,x)\right) = \alpha e^{\alpha x+\beta\tau}u(\tau,x) + e^{\alpha x+\beta\tau}\partial_x u(\tau,x) \\ \partial_x^2 w(\tau,x) &= \alpha^2 e^{\alpha x+\beta\tau}u(\tau,x) + 2\alpha e^{\alpha x+\beta\tau}\partial_x u(\tau,x) + e^{\alpha x+\beta\tau}\partial_x^2 u(\tau,x) \end{aligned}$$

it follows that

$$\begin{aligned} \partial_\tau u(\tau,x) =& -\beta u(\tau,x) + \Big(\alpha^2 u(\tau,x) + 2\alpha\partial_x u(\tau,x) + \partial_x^2 u(\tau,x)\Big) \\ &+ (c-1)\Big(\alpha u(\tau,x) + \partial_x u(\tau,x)\Big) - cu(\tau,x) \\ =& \partial_x^2 u(\tau,x) + \Big(2\alpha + (c-1)\Big)\partial_x u(\tau,x) + \Big(-\beta + \alpha^2 + (c-1)\alpha - c\Big)u(\tau,x) \end{aligned}$$

Hence, the terms including $u(\tau, x)$ and $\partial_x u(\tau, x)$ vanish if

$$-\beta + \alpha^2 + (c-1)\alpha - c = 0 \qquad \text{and} \qquad 2\alpha + (c-1) = 0.$$

The solution is

$$\alpha = -\frac{1}{2}(c-1), \qquad \beta = -\frac{1}{4}(c+1)^2 = -(1-\alpha)^2.$$

With these parameters, $u(\tau, x)$ solves the **heat equation**

$$\partial_\tau u(\tau, x) = \partial_x^2 u(\tau, x), \qquad x \in \mathbb{R}, \tau \in [0, \sigma^2 T/2]$$

with initial condition

$$u(0, x) = e^{-\alpha x} w(0, x) = e^{-\alpha x} \frac{V(T, S)}{K} = e^{-\alpha x} \frac{(S-K)^+}{K} = e^{-\alpha x}(e^x - 1)^+$$

since $S = Ke^x$.

Step 2: Solving the heat equation

Lemma 3.3.1 (solution of the heat equation) *Let $u_0 : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous function which satisifes the growth condition*

$$|u_0(x)| \leq M e^{\gamma x^2}.$$

with constants $M > 0$ and $\gamma \in \mathbb{R}$. Then, the function

$$u(\tau, x) = \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\xi)^2}{4\tau}\right) u_0(\xi)\, d\xi$$

is the unique solution of the heat equation

$$\partial_\tau u(\tau, x) = \partial_x^2 u(\tau, x), \qquad x \in \mathbb{R}, \tau > 0$$

and we have

$$\lim_{\tau \to 0} u(\tau, x) = u_0(x).$$

Proof. The fact that u solves the PDE can be checked by substituting and computing the partial derivatives (exercise). The last assertion can be verified via the transformation $\eta = (\xi - x)/\sqrt{4\tau}$ (exercise). Uniqueness follows from the maximum principle. ■

By a tedious[1] calculation, it can be shown that

$$u(\tau,x) = \exp\left((1-\alpha)x + (1-\alpha)^2\tau\right)\Phi(d_1) - \exp\left(-\alpha x + \alpha^2\tau\right)\Phi(d_2)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-s^2/2}\,ds \tag{3.5}$$

$$d_{1/2} = \frac{\ln\frac{S}{K} + \left(r \pm \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{3.6}$$

Remark: $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

Step 3: Inverse transform

Since $\beta = -(1-\alpha)^2$ and $\beta + \alpha^2 = 2\alpha - 1 = -c$ it follows that

$$\begin{aligned} V(t,S) &= Kw(\tau,x) = K\exp(\alpha x + \beta\tau)u(\tau,x) \\ &= K\exp(\alpha x + \beta\tau)\exp\left((1-\alpha)x + (1-\alpha)^2\tau\right)\Phi(d_1) \\ &\quad - K\exp(\alpha x + \beta\tau)\exp\left(-\alpha x + \alpha^2\tau\right)\Phi(d_2) \\ &= \underbrace{K\exp(x)}_{=S}\Phi(d_1) - K\exp\Big(\underbrace{(\beta+\alpha^2)}_{-c}\tau\Big)\Phi(d_2) \\ &= S\Phi(d_1) - K\exp(-r(T-t))\Phi(d_2) \end{aligned}$$

Check boundary:

$$V(t,0) = -K\exp(-r(T-t))\Phi(d_2) = \exp(-r(T-t))V(T,0) \iff (3.4)\ \checkmark$$

Check terminal condition:

$$V(T,S) = S\Phi(d_1) - K\Phi(d_2)$$

By definition of $d_{1/2} = d_{1/2}(t)$

$$\lim_{t\to T} d_{1/2}(t) = \lim_{t\to T}\frac{\ln\frac{S}{K} + \left(r \pm \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} = \lim_{t\to T}\frac{\ln\frac{S}{K}}{\sigma\sqrt{T-t}} = \begin{cases}\infty & \text{if } S > K\\ 0 & \text{if } S = K\\ -\infty & \text{if } S < K\end{cases}$$

and hence

$$\lim_{t\to T}\Phi(d_{1/2}(t)) = \begin{cases}1 & \text{if } S > K\\ 1/2 & \text{if } S = K\\ 0 & \text{if } S < K\end{cases} \quad\Longrightarrow\quad \lim_{t\to T}V(t,S) = \begin{cases}S-K & \text{if } S > K\\ 0 & \text{if } S = K\\ 0 & \text{if } S < K\end{cases} \quad\checkmark$$

All in all, we have shown the following

[1] ... so tedious that we do not even dare to ask the reader to prove this as an exercise.

Theorem 3.3.2 (Black-Scholes formula for calls) *If $r, \sigma, K, T > 0$, then the* ***Black-Scholes formula***

$$V(t,S) = S\Phi(d_1) - K\exp(-r(T-t))\Phi(d_2)$$

with Φ and $d_{1/2}$ from (3.5) and (3.6), respectively, is the (unique) solution of the Black-Scholes equation for European calls, i.e.

$$\begin{aligned} &\partial_t V(t,S) + \frac{\sigma^2}{2}S^2\partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0 \qquad t \in [0,T], \quad S > 0 \\ &V(T,S) = (S-K)^+. \end{aligned}$$

Corollary 3.3.3 (Black-Scholes formula for puts) *The Black-Scholes equation for a European put*

$$\begin{aligned} &\partial_t V(t,S) + \frac{\sigma^2}{2}S^2\partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0 \qquad t \in [0,T], \quad S > 0 \\ &V(T,S) = (K-S)^+ \end{aligned}$$

with $r, \sigma, K, T > 0$ has the unique solution

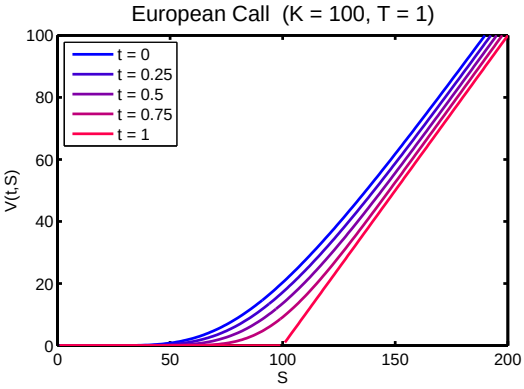
$$V(t,S) = K\exp(-r(T-t))\Phi(-d_2) - S\Phi(-d_1)$$

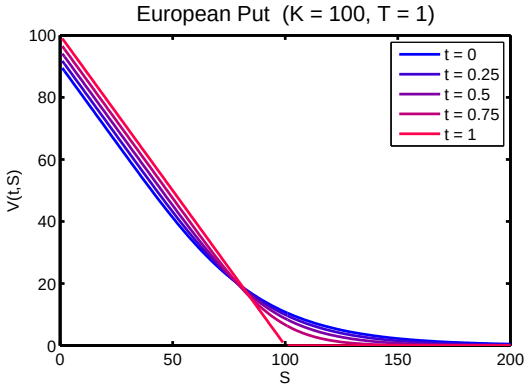
with Φ and $d_{1/2}$ from (3.5) and (3.6), respectively.

Proof. Let $V_C(t,S)$ be the value of a call with the same T and K. The put-call-parity (Lemma 1.4.1) and Theorem 3.3.2 imply

$$\begin{aligned} V(t,S) &= e^{-r(T-t)}K + V_C(t,S) - S \\ &= e^{-r(T-t)}K + S\Phi(d_1) - K\exp(-r(T-t))\Phi(d_2) - S \\ &= e^{-r(T-t)}K\big(1-\Phi(d_2)\big) + S\big(\Phi(d_1)-1\big) \\ &= e^{-r(T-t)}K\Phi(-d_2) - S\Phi(-d_1) \end{aligned}$$

because $\Phi(x) + \Phi(-x) = 1$. ∎





Definition 3.3.4 (Greeks) *For a European option with value $V(t,S)$ we define* ***"the greeks"***

delta:	Δ	$=$	$\partial_S V$	*theta:*	θ	$=$	$\partial_t V$
gamma:	Γ	$=$	$\partial_S^2 V$	*rho:*	ρ	$=$	$\partial_r V$
vega/kappa:	κ	$=$	$\partial_\sigma V$				

These partial derivatives can be considered as "condition numbers" which measure the sensitivity of $V(t,S)$ with respect to the corresponding parameters. This information is important for stock broker.

Remark: Explicit formulas for the greeks can be derived from the Black-Scholes formulas (execise).

3.4 Risk-neutral valuation and equivalent martingale measures

In 1.5 we have seen that in the simplified two-scenario model the value of an option can be priced by replication. The same strategy was applied to the refined model in the previous section. In the simple situation considered in 1.5, the value of an option turned out to be the discounted expectation of the payoff under the risk-neutral probability. In this subsection, we will see that this is also true for the refined model from 3.2.

Theorem 3.4.1 (Option price as discounted expectation) *If $V(t,S)$ is the solution of the Black-Scholes equation*

$$\partial_t V(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0 \qquad t \in [0,T], \quad S > 0$$
$$V(T,S) = \psi(S)$$

with payoff function $\psi(S)$, then

$$V(t_\star, S_\star) = e^{-r(T-t_\star)} \int_0^\infty \psi(x)\phi(x,\xi,\beta)\, dx \tag{3.7}$$

for all $t_\star \in [0,T]$ and $S_\star > 0$. The function ϕ is the density of the log-normal distribution (cf. Definition 3.1.2) with parameters

$$\xi = \ln S_\star + \left(r - \frac{\sigma^2}{2}\right)(T - t_\star), \qquad \beta = \sigma\sqrt{T - t_\star}. \tag{3.8}$$

The assertion can be shown by showing that the above representation coincides with the Black-Scholes formulas for puts and calls. Such a proof, however, involves several changes

of variables in the integral representations and rather tedious calculations. We give a shorter and more elegant proof:

Proof. Step 1: In our derivation of the Black-Scholes model, we have assumed that

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

i.e. that the price of the underlying is a geometric Brownian motion with drift μS_t; cf. (3.1). It turned out, however, that the parameter μ does not appear in the Black-Scholes equation. Hence, we can choose $\mu = r$ and consider the SDE

$$\begin{aligned} d\widehat{S}_t &= r\widehat{S}_t dt + \sigma \widehat{S}_t dW_t, \qquad t \in [t_\star, T] \\ \widehat{S}_{t_\star} &= S_\star \end{aligned}$$

as a model for the stock price.
Step 2: The function $u(t,S) := e^{r(T-t)}V(t,S)$ solves the PDE

$$\partial_t u(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 u(t,S) + rS\partial_S u(t,S) = 0, \qquad t \in [0,T]$$

because

$$\begin{aligned} &\partial_t u(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 u(t,S) + rS\partial_S u(t,S) \\ =& -re^{r(T-t)}V(t,S) + e^{r(T-t)}\partial_t V(t,S) + \frac{\sigma^2}{2} S^2 e^{r(T-t)}\partial_S^2 V(t,S) + rSe^{r(T-t)}\partial_S V(t,S) \\ =& e^{r(T-t)} \underbrace{\left(-rV(t,S) + \partial_t V(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) + rS\partial_S V(t,S) \right)}_{=0 \quad \text{(Black-Scholes equation)}} = 0. \end{aligned}$$

Moreover, u satisfies the terminal condition

$$u(T,S) = V(T,S) = \psi(S).$$

Step 3: Applying the Feynman-Kac formula (cf. 2.6) with $f(t,S) = rS$ and $g(t,S) = \sigma S$ yields

$$\mathbb{E}\left(\psi(\widehat{S}_T)\right) = u(t_\star, S_\star) = e^{r(T-t_\star)}V(t_\star, S_\star)$$

and thus

$$V(t_\star, S_\star) = e^{-r(T-t_\star)}\mathbb{E}\left(\psi(\widehat{S}_T)\right).$$

We know that $\widehat{S}_T$ is log-normal, i.e.

$$\mathbb{E}\left(\psi(\widehat{S}_T)\right) = \int_0^\infty \psi(x)\phi(x,\xi,\beta)\, dx$$

with parameters (3.8). ∎

Interpretation. We know from 3.1 that

$$\begin{aligned}\mathbb{E}\left(\widehat{S}_T\right) &= \int_0^\infty x\phi(x,\xi,\beta)\,dx \\ &= \exp\left(\xi + \frac{\beta^2}{2}\right) \\ &= \exp\left(\ln S_\star + \left(r - \frac{\sigma^2}{2}\right)(T-t_\star) + \frac{1}{2}\left(\sigma\sqrt{T-t_\star}\right)^2\right) \\ &= \exp\left(\ln S_\star + r(T-t_\star)\right) \\ &= S_\star \exp\left(r(T-t_\star)\right).\end{aligned}$$

This means that for $\mu = r$ the expected value of the stock is exactly the money obtained by investing $S_\star$ into a bond at time $t_\star$ and waiting until $T - t_\star$. Hence, the log-normal distribution with parameters (3.8) defines the **risk-neutral probability**; cf. 1.5. The integral in (3.7) is precisely the expected payoff under the risk-neutral probability, and (3.7) states that the price of the option is obtained by discounting the expected payoff.

A different perspective. Consider now the geometric Brownian motion

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

with $\mu \neq r$. Since $\mathbb{E}(S_t) = S_0 e^{\mu t}$, an investor expects $\mu > r$ as a compensation for the risk, because otherwise he might prefer to invest into the riskless bond $B_t = B_0 e^{rt}$. The term

$$\gamma = \frac{\mu - r}{\sigma}$$

is called **market price of risk**, and we have

$$dS_t = rS_t dt + \sigma S_t(\gamma\, dt + dW_t) = rS_t dt + \sigma S_t dW_t^\gamma, \qquad \text{with } W_t^\gamma = \gamma\, t + W_t.$$

Problem: W_t^γ is not a Wiener process under the probability measure $\mathbb{P}$, because $\mathbb{E}(W_t^\gamma) = \gamma\, t + \mathbb{E}(W_t) = \gamma\, t \neq 0$ for $t > 0$ and $\mu \neq r$.

Question: Is there another probability measure $\mathbb{Q}$ such that dW_t^γ is a Wiener process under $\mathbb{Q}$?

Definition 3.4.2 (equivalent martingale measure) *A probability measure $\mathbb{Q}$ is called an **equivalent martingale measure** or **risk-neutral probability** on $(\Omega, \mathcal{F}, \mathbb{P})$ if there is a random variable $Y > 0$ such that*

- $\mathbb{Q}(A) = \mathbb{E}(\mathbf{1}_A \cdot Y) = \int_A Y(\omega) d\mathbb{P}(\omega)$ *for all events $A \in \mathcal{F}$, and*
- $e^{-rt} S_t$ *is a martingale under $\mathbb{Q}$.*

Remark. The first property implies that $\mathbb{P}(A) > 0 \iff \mathbb{Q}(A) > 0$ ("equivalent").

Now let

$$Y_T := \exp\left(-\gamma W_T - \frac{\gamma^2}{2}T\right) \qquad \text{and} \qquad \mathbb{Q}(A) = \mathbb{E}(\mathbf{1}_A \cdot Y_T).$$

Then, Girsanov's theorem states that $W_t^\gamma = \gamma\, t + W_t$ is a Wiener process under $\mathbb{Q}$ (see e.g. 4.4 in [Ben04], 8.6 in [Øks03]). Moreover, the Itô formula yields

$$d(e^{-rt}S_t) = -re^{-rt}S_t dt + e^{-rt}\Big(rS_t dt + \sigma S_t dW_t^\gamma\Big) = \sigma e^{-rt}S_t dW_t^\gamma.$$

Hence, $e^{-rt}S_t$ is a martingale under $\mathbb{Q}$, and $\mathbb{Q}$ is an equivalent martingale measure. All in all, we have exchanged

$$\mu \longrightarrow r, \qquad \mathbb{P} \longrightarrow \mathbb{Q}, \qquad W_t \longrightarrow W_t^\gamma.$$

If $\mu = r$, then $\mathbb{P} = \mathbb{Q}$ and $W_t = W_t^\gamma$. Now we are back in the situation of Theorem 3.4.1, and it follows that

$$V(t_\star, S_\star) = e^{-r(T-t_\star)}\mathbb{E}_\mathbb{Q}\left(\psi(S_T)\right) \tag{3.9}$$

where S_t is the solution of

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_t dW_t^\gamma, \qquad t \in [t_\star, T] \\ S_{t_\star} &= S_\star. \end{aligned}$$

General pricing formula. Up to now, we have only considered European options, i.e. options with a payoff that depends only on the value of the underlying at maturity. For Asian or barrier options, the pricing formula (3.9) can be generalized to

$$V_t = e^{-r(T-t_\star)}\mathbb{E}_\mathbb{Q}\left(V_T|\mathcal{F}_t\right)$$

(cf. 5.2.4 in [Shr04]).

Fundamental theorems of option pricing:

- If a market model has at least one equivalent martingale measure, then there is no arbitrage possibility (cf. Theorem 5.4.7 in [Shr04]).
- Consider a market model with at least one equivalent martingale measure. Then, the equivalent martingale measure is unique if and only if the model is complete, i.e. if every derivative security can be replicated (hedged) (cf. Theorem 5.4.9 in [Shr04]).

These theorems can be extended to much more general market models (many underlyings, time-dependent interest rate, etc.).

3.5 Extensions

The "standard" Black-Scholes model can be generalized in several ways:

- Options with $d > 1$ underlyings (e.g. basket options) are modeled by the d-dimensional Black-Scholes equation

$$\partial_t V + \frac{1}{2}\sum_{i,j=1}^{d} \rho_{ij}\sigma_i\sigma_j S_i S_j \partial_{S_i}\partial_{S_j} V + r\sum_{i=1}^{d} S_i \partial_{S_i} V - rV = 0$$

  where $V = V(t, S_1, \ldots, S_d)$, $r, \sigma_i > 0$ and $\rho_{ij} \in [-1, 1]$ are the correlation coefficients.
- Non-constant interest rate and volatility: $r = r(t, S)$, $\sigma = \sigma(t, S)$
- Stochastic volatility: Either $\sigma = \sigma(\omega)$ is a random variable with known distribution or $\sigma = \sigma_t(\omega)$ is a stochastic process.
- Dividends: When a dividend $\delta \cdot S_t$ with $\delta \geq 0$ is paid at time t, the price of the underlying drops by the same amount due to the no-arbitrage assumption. Hence, a **continuous** flow of dividends can be modeled by

$$dS_t = (\mu - \delta)S_t dt + \sigma S_t dW_t,$$

  which yields the Black-Scholes equation

$$\partial_t V(t, S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t, S) + (r - \delta) S \partial_S V(t, S) - rV(t, S) = 0.$$

  Black-Scholes formulas with dividends: 4.5.1 in [GJ10].
- Nonzero transaction costs $\Longrightarrow$ nonlinear Black-Scholes equation
- Discontinuous underlyings: Jump-diffusion models, Black-Scholes PDE with additional integral term

Remark: Some of these extensions will be considered in the lecture *Numerical methods in mathematical finance II* (summer term).

Part II

Numerical methods

Chapter 4

Binomial methods

Situation: Let $S(t)$ be the value of an underlying, and let $V(t,S)$ be the value of an option with maturity $T > 0$.

Assumptions: Assume (A1)-(A5) from 1.3.

Goal: Approximate $V(t,S)$ by a numerical method.

Idea: Refine the simple discrete model from 1.5 such that it approximates the continuous-time Black-Scholes model.

Remark: For European calls/puts, $V(t,S)$ such an approximation is not necessary, because $V(t,S)$ can be computed via the Black-Scholes formula. Nevertheless, such options will serve as a model problem. The numerical method can be extended to other types of options.

4.1 Derivation

Discretize the time-interval $[0,T]$: Choose $N \in \mathbb{N}$, let $\Delta t = T/N$ and $t_n = n \cdot \Delta t$.
Additional assumptions:

1. For a given price $S(t_n)$, the price at $t_{n+1} = t_n + \Delta t$ is

$$S(t_{n+1}) = \begin{cases} u \cdot S(t_n) & \text{with probability } p \\ d \cdot S(t_n) & \text{with probability } 1-p \end{cases}$$

   with (unknown) $u > 1$, $0 < d < 1$, $p \in [0,1]$.

2. The expected profit from investing into the underlying is the same as for the bond:

$$\mathbb{E}\big(S(t_{n+1})|S(t_n)\big) = e^{r\Delta t} S(t_n).$$

3. $\mathbb{E}\big(S^2(t_{n+1})|S(t_n)\big) = e^{(2r+\sigma^2)\Delta t} S^2(t_n)$ with given volatility $\sigma \in \mathbb{R}$.

Remark: In the continuous-time model where S_t is modeled by a geometric Brownian motion, the last two conditions hold if $\mu = r$ (risk-neutral pricing).

Compute u, d, p:

1. $e^{r\Delta t}S(t_n) = \mathbb{E}\big(S(t_{n+1})|S(t_n)\big) = uS(t_n)p + dS(t_n)(1-p) \Longleftrightarrow p = \dfrac{e^{r\Delta t} - d}{u - d}$
   Since $p \in [0,1]$, we have $d \leq e^{r\Delta t} \leq u$.
2. $e^{(2r+\sigma^2)\Delta t}S^2(t_n) = \mathbb{E}\big(S^2(t_{n+1})|S(t_n)\big) = \big(uS(t_n)\big)^2 p + \big(dS(t_n)\big)^2(1-p)$

Only two conditions for three unknowns u, d, p. Choose $u \cdot d \stackrel{!}{=} 1$ as third condition:

$$p \stackrel{!}{=} \frac{e^{r\Delta t} - d}{u - d}, \qquad e^{(2r+\sigma^2)\Delta t} \stackrel{!}{=} u^2 p + d^2(1-p), \qquad u \cdot d \stackrel{!}{=} 1$$

Solution:

$$u = \beta + \sqrt{\beta^2 - 1} \qquad \beta := \frac{1}{2}\left(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t}\right)$$
$$d = \frac{1}{u} = \beta - \sqrt{\beta^2 - 1} \qquad p \stackrel{!}{=} \frac{e^{r\Delta t} - d}{u - d}.$$

Replication strategy: Consider a portfolio containing $a \in \mathbb{R}$ underlyings and $b \in \mathbb{R}$ bonds such that

$$aS(t_n) + bB(t_n) \stackrel{!}{=} V\big(t_n, S(t_n)\big) =: V(t_n)$$

It follows that

$$\begin{aligned}\mathbb{E}\big(V(t_{n+1}) \mid V(t_n)\big) &= a\underbrace{\big(pu + (1-p)d\big)}_{e^{r\Delta t}} S(t_n) + be^{r\Delta t}B(t_n) \\ &= e^{r\Delta t}\big(aS(t_n) + bB(t_n)\big) \\ &= e^{r\Delta t}V(t_n) \end{aligned} \tag{4.1}$$

4.2 Algorithm

Cox, Ross & Rubinstein 1979

1. Forward phase: initialization of the tree. For all $n = 0, \ldots, N$ and $j = 0, \ldots, n$ let

$$S_{jn} = u^j d^{n-j} S(0) = \text{(approximate) price of the underlying at time } t_n \text{ after } j \text{ "ups" and } n-j \text{ "downs"}.$$

The condition $d \cdot u = 1$ implies that

$$\begin{aligned} S(0) = S_{00} = S_{12} = S_{24} = \ldots \\ S_{11} = S_{23} = S_{35} = \ldots \\ S_{01} = S_{13} = S_{25} = \ldots \end{aligned}$$

At t_n there are only $n+1$ possible values $S_{0n}, \ldots, S_{nn}$ of the underlying.

$S_{00} = S(0)$
`for` $n = 0, 1, 2, \ldots, N-1$
    `for` $j = 0, \ldots, n$
        $S_{j+1,n+1} = uS_{j,n}$
        $S_{j,n+1} = dS_{j,n}$
    `end`
`end`

2. Backward phase: compute the option values. Let V_{jn} be the value of the option after j "ups" and $n-j$ "downs" of the underlying. At maturity, we have

$$V_{jN} = \psi(S_{jN}), \qquad \psi(S_{jN}) = \begin{cases} (S_{jN} - K)^+ & \text{(call)} \\ (K - S_{jN})^+ & \text{(put)} \end{cases}$$

Use (4.1):

$$\begin{aligned} & e^{r\Delta t} V_{jn} = \mathbb{E}\big(V(t_{n+1}) \mid V_{jn}\big) = pV_{j+1,n+1} + (1-p)V_{j,n+1} \\ \Longrightarrow \qquad & V_{jn} = e^{-r\Delta t}\Big(pV_{j+1,n+1} + (1-p)V_{j,n+1}\Big) \end{aligned}$$

(a) European options:

`for` $j = 0, \ldots, N$
    $V_{jN} = \psi(S_{jN})$
`end`
`for` $n = N-1, N-2, \ldots, 0$
    `for` $j = 0, \ldots, n$
        $V_{jn} = e^{-r\Delta t}\Big(pV_{j+1,n+1} + (1-p)V_{j,n+1}\Big)$
    `end`
`end`
`Result:` V_{00}

(b) American options: Check in each step if early exercise is advantageous. At time t_n the value of the option must not be less than $\psi(S_{jn})$ for all j.

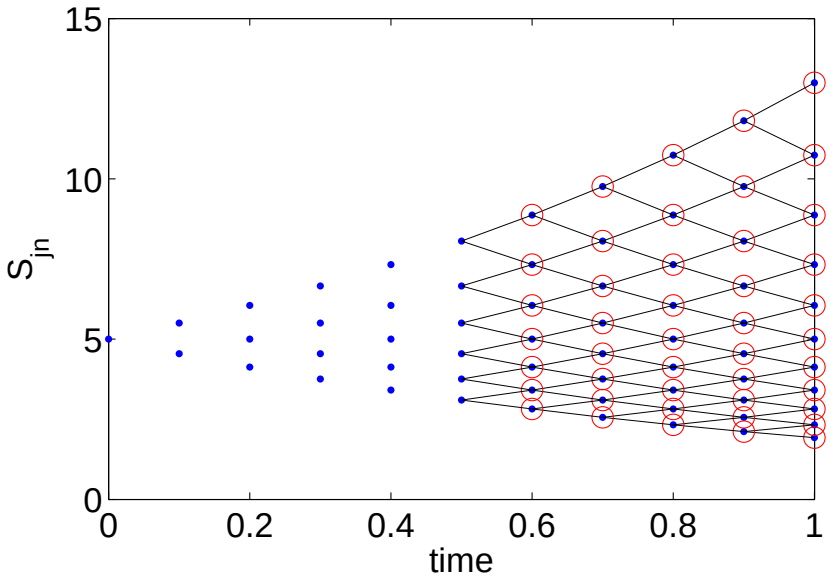


Figure 4.1: Illustration of the binomial method ($S(0) = 5$, $T = 1$, $N = 10$).

```
for j = 0, ..., N
    V_jN = ψ(S_jN)
end
for n = N − 1, N − 2, ..., 0
    for j = 0, ..., n
        Ṽ_jn = e^{−rΔt} ( p V_{j+1,n+1} + (1 − p) V_{j,n+1} )
        V_jn = max{Ṽ_jn, ψ(S_jn)}
    end
end
Result:  V_00
```

Remarks.

1. The result V_{00} is only an **approximation** for the true value $V(0, S(0))$ of the option, because the price process has been approximated.
2. The result $V_{00} \approx V(0, S(0))$ depends on the initial value $S(0)$. For a different value of $S(0)$, the entire computation must be repeated.
3. An efficient implementation of the binomial method requires only $\mathcal{O}(N)$ operations; see [Hig02].

Examples: see slides

4.3 Discrete Black-Scholes formula

Lemma 4.3.1 *Let $V(t,S)$ be the value of a European option with payoff function $\psi(S)$ and maturity $T>0$. Then, the binomial method yields the approximation*

$$V_{00} = e^{-rT} \sum_{j=1}^{N} \mathcal{B}(j,N,p)\psi(S_{jN})$$

where

$$\mathcal{B}(j,N,p) = \binom{N}{j} p^j (1-p)^{N-j}$$

is the binomial distribution.

Proof: exercise.

Remarks:

1. This result explains the name "binomial method".
2. Interpretation: V_{00} is the discounted expected payoff under a suitable probability; cf. 1.5 and 3.4.

Proposition 4.3.2 (Discrete Black-Scholes formula) *Under the conditions of Lemma 4.3.1, an equivalent formula for the option price is*

$$V_{00} = S(0)\Psi(m,N,q) - K\exp(-rT)\Psi(m,N,p)$$

where

$$\begin{aligned} q &= upe^{-r\Delta t} \\ m &= \min\{0 \le j \le N : (S_{jN} - K) \ge 0\}. \\ \Psi(m,N,p) &= \sum_{j=m}^{N} \mathcal{B}(j,N,p) \end{aligned}$$

Proof: exercise.

Question: What happens if we let $N \longrightarrow \infty$ and $\Delta t = T/N \longrightarrow 0$?

Proposition 4.3.3 (Convergence of the discrete Black-Scholes formula) *Let $\hat{\sigma} = (\ln u)/\sqrt{\Delta t}$, i.e. $u = e^{\hat{\sigma}\sqrt{\Delta t}}$ and $d = 1/u = e^{-\hat{\sigma}\sqrt{\Delta t}}$. If $V_{00} = V_{00}^{(N)}$ is the approximation given by the binomial method with $\Delta t = T/N$, then $V_{00}^{(N)}$ converges to the value given by*

the (continuous) Black-Scholes equation:

$$\lim_{N\to\infty} V_{00}^{(N)} = S(0)\Phi(d_1) - K\exp(-rT)\Phi(d_2)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2}\, ds$$

$$d_{1/2} = \frac{\ln\left(\frac{S(0)}{K}\right) + \left(r \pm \frac{\hat{\sigma}^2}{2}\right) T}{\hat{\sigma}\sqrt{T}}$$

Proof: See 3.3 in [GJ10] (use central limit theorem).

Chapter 5

Numerical methods for stochastic differential equations

5.1 Motivation

According to 3.4 the value of a European option is the discounted expected payoff under the risk-neutral probability:

$$V(0, S_0) = e^{-rT}\mathbb{E}_{\mathbb{Q}}\Big(\psi\big(S(T)\big)\Big)$$

For the standard Black-Scholes model:

$$V(0, S_0) = e^{-rT} \int_0^\infty \psi(x)\phi(x, \xi, \beta)\, dx$$

with log-normal density ϕ and parameters

$$\xi = \ln S_0 + \left(r - \frac{\sigma^2}{2}\right) T, \qquad \beta = \sigma\sqrt{T}.$$

Two ways to price the option:

1. **Quadrature formula.** Let $w(x) := \psi(x)\phi(x, \xi, \beta)$.
   Choose $0 \leq x_{min} < x_{\max}$ such that $w(x) \approx 0$ for $x \notin [x_{\min}, x_{\max}]$.
   $x_{\min} = K$ and $x_{\max}$ sufficiently large for calls, $x_{\min} = 0$ and $x_{\max} = K$ for puts.
   Choose large $N \in \mathbb{N}$, let $h = (x_{\max} - x_{\min})/N$ and $x_k = x_{\min} + kh$. Approximate

   $$\int_0^\infty w(x)\, dx \approx \int_{x_{\min}}^{x_{\max}} w(x)\, dx = \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} w(x)\, dx \approx \sum_{k=0}^{N-1} h \sum_{j=1}^{s} b_j w(x_k + c_j h)$$

   with suitable nodes $c_j \in [0, 1]$ and weights b_j.

2. **Monte-Carlo method.** In the Black-Scholes model, $S(t)$ is defined by the SDE

$$dS(t) = rS(t)dt + \sigma S(t)dW(t), \qquad t \in [0,T], \qquad S_0 \text{ given}$$

(risk-neutral, $\mu = r$)
Solution: Geometric Brownian motion

$$S(t) = S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W(t)\right).$$

This is the process which corresponds to $\phi(x, \xi, \beta)$, because $S(T)$ is log-normal with the same parameters. Estimate the expectated payoff as follows:

- Generate many realizations $S(T, \omega_1), \ldots, S(T, \omega_m)$, $m \in \mathbb{N}$ "large".
- Approximate

$$V(0, S_0) \approx e^{-rT} \frac{1}{m} \sum_{j=1}^{m} \psi\big(S(T, \omega_j)\big)$$

Consider now a more complicated price process:

$$dS(t) = rS(t)dt + \sigma S(t)dW^1(t) \tag{5.1a}$$

$$d\sigma^2(t) = \kappa\big(\theta - \sigma^2(t)\big)dt + \nu\left(\rho dW^1(t) + \sqrt{1-\rho^2}dW^2(t)\right) \tag{5.1b}$$

Heston model with parameters $r, \kappa, \theta, \nu > 0$, initial values S_0, σ_0, independent scalar Wiener processes $W^1(t)$, $W^2(t)$, correlation $\rho \in [-1, 1]$
Steven L. Heston 1993
Now the volatility is not a parameter, but a stochastic process defined by a second SDE. We do not have an explicit formula for $S(t)$ and $\sigma(t)$, but the Monte-Carlo approach is still feasible:

- Choose $N \in \mathbb{N}$, define step-size $\tau = T/N$ and $t_n = n\tau$. For each $\omega_1, \ldots, \omega_m$ compute approximations

$$X_n^1(\omega_j) \approx S(t_n, \omega_j), \qquad X_n^2(\omega_j) \approx \sigma^2(t_n, \omega_j), \qquad n = 0, \ldots, N$$

by solving the SDEs (5.1a), (5.1b) numerically.

- Approximate

$$V(0, S_0) \approx e^{-rT} \frac{1}{m} \sum_{j=1}^{m} \psi\big(\underbrace{X_N^1(\omega_j)}_{\approx S(T,\omega_j)}\big)$$

The Monte-Carlo approach even works for other types of options. As an example, consider an Asian option with payoff

$$\psi\big(t \mapsto S(t)\big) = \left(S(T) - \frac{1}{T}\int_0^T S(t)\, dt\right)^+ \qquad \text{(average strike call).}$$

Now the payoff depends on the entire path $t \mapsto S(t)$. We approximate

$$S(T,\omega_j) \approx X_N^1(\omega_j), \qquad \frac{1}{T}\int_0^T S(t,\omega_j)\,dt \approx \frac{1}{N}\sum_{n=1}^N X_n^1(\omega_j)$$

and hence

$$V(0,S_0) \approx e^{-rT}\frac{1}{m}\sum_{j=1}^m \left(X_N^1(\omega_j) - \frac{1}{N}\sum_{n=1}^N X_n^1(\omega_j)\right)^+$$

Remark: In the original paper, Heston derives an explicit Black-Scholes-type formula for European options by means of characteristic functions. Hence, European options in the Heston model can also be priced by quadrature formulas, but for Asian options this is impossible.

Goal: Construct and analyze numerical methods for SDEs.

5.2 Euler-Maruyama method

5.2.1 Derivation

Consider the one-dimensional SDE

$$dX(t) = f\big(t,X(t)\big)dt + g\big(t,X(t)\big)dW(t), \qquad t\in[0,T], \qquad X(0)=X_0$$

with suitable functions f and g and a given initial value X_0. Choose $N\in\mathbb{N}$, define step-size $\tau = T/N$ and $t_n = n\tau$.

$$\begin{aligned} X(t_{n+1}) &= X(t_n) + \int_{t_n}^{t_{n+1}} f\big(s,X(s)\big)\,ds + \int_{t_n}^{t_{n+1}} g\big(s,X(s)\big)dW(s) \\ &\approx X(t_n) + \underbrace{(t_{n+1}-t_n)}_{=\tau} f\big(t_n,X(t_n)\big) + g\big(t_n,X(t_n)\big)\underbrace{\big(W(t_{n+1})-W(t_n)\big)}_{=:\Delta W_n} \end{aligned}$$

Replacing $X(t_n) \longrightarrow X_n$ and "$\approx$" $\longrightarrow$ "$=$" yields the

Euler-Maruyama method (Gisiro Maruyama 1955, Leonhard Euler 1768-70):
For $n=0,\dots,N-1$ let $\Delta W_n = W(t_{n+1})-W(t_n)$ and

$$X_{n+1} = X_n + \tau f\big(t_n,X_n\big) + g\big(t_n,X_n\big)\Delta W_n.$$

Hope that $X_n \approx X(t_n)$.

	SDE		recursion
$X(t)$	exact	$\Longrightarrow$	approx.
X_n	approx.	$\Longleftarrow$	exact

$\Big\}$ $\overset{?}{\Longrightarrow}$ $X(t_n)\approx X_n$

The exact solution $X(t_n)$ and the numerical approximation X_n are random variables. For every path $t \mapsto W(t,\omega)$ of the Wiener process, a different result is obtained. $X(t)$ is called **strong solution** if $t \mapsto W(t,\omega)$ is given, and **weak solution** if $t \mapsto W(t,\omega)$ can be chosen. Approximations of weak solutions: For each n, generate a random number $Z_n \sim \mathcal{N}(0,1)$ and let

$$\Delta W_n = \sqrt{\tau} Z_n.$$

Question: Does X_n really approximate $X(t_n)$? In which sense? How accurately?

5.2.2 Weak and strong convergence

Definition 5.2.1 (strong and weak convergence) *Let $T > 0$, $N \in \mathbb{N}$, $\tau = T/N$ and $t_n = n\tau$. An approximation $X_n(\omega) \approx X(t_n,\omega)$ converges*

- ***strongly** with order $\gamma > 0$, if there is a constant $C > 0$ independent of τ such that*

$$\max_{n=0,\dots,N} \mathbb{E}\Big(|X(t_n) - X_n|\Big) \leq C\tau^\gamma$$

  for all sufficiently small τ, and

- ***weakly** with order $\gamma > 0$ with respect to a function $F : \mathbb{R} \longrightarrow \mathbb{R}$, if there is a constant $C > 0$ independent of τ such that*

$$\max_{n=0,\dots,N} \Big|\mathbb{E}\big[F\big(X(t_n)\big)\big] - \mathbb{E}\big[F(X_n)\big]\Big| \leq C\tau^\gamma$$

  for all sufficiently small τ.

Remarks:

- Strong convergence $\Longrightarrow$ path-wise convergence
  Weak convergence $\Longrightarrow$ convergence of moments (if $F(x) = x^k$) or probabilities (if $F(x) = \mathbf{1}_{[a,b]}$).
- Strong convergence $\leftrightarrow$ Asian and barrier options
  Weak convergence $\leftrightarrow$ European options
- Strong convergence of order γ implies weak convergence of order γ with respect to $F(x) = x$ (exercise).

5.2.3 Existence and uniqueness of solutions of SDEs

Theorem 5.2.2 (existence and uniqueness)
Let $f : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ be functions with the following properties:

- ***Lipschitz condition:** There is a constant $L \geq 0$ such that*

$$|f(t,x) - f(t,y)| \leq L|x-y|, \qquad |g(t,x) - g(t,y)| \leq L|x-y| \tag{5.2}$$

  for all $x, y \in \mathbb{R}$ and $t \geq 0$.

- ***Linear growth condition:*** *There is a constant $K \geq 0$ such that*

$$|f(t,x)|^2 \leq K\left(1+|x|^2\right), \qquad |g(t,x)|^2 \leq K\left(1+|x|^2\right) \tag{5.3}$$

for all $x \in \mathbb{R}$ and $t \geq 0$.

Then, the SDE

$$dX(t) = f\big(t,X(t)\big)dt + g\big(t,X(t)\big)dW(t), \qquad t \in [0,T]$$

with deterministic initial value $X(0) = X_0$ has a continuous adapted solution and

$$\sup_{t\in[0,T]} \mathbb{E}\left(X^2(t)\right) < \infty.$$

If both $X(t)$ and $\widetilde{X}(t)$ are such solutions, then

$$\mathbb{P}\Big(X(t) = \widetilde{X}(t) \text{ for all } t \in [0,T]\Big) = 1.$$

Proof: Theorem 9.1 in [Ste01] or Theorem 4.5.3 in [KP99].

Remark: The assumptions can be weakened.

5.2.4 Strong convergence of the Euler-Maruyama method

For simplicity, we only consider the autonomous SDE

$$dX(t) = f\big(X(t)\big)dt + g\big(X(t)\big)dW(t), \qquad t \in [0,T]$$

and the Euler-Maruyama approximation

$$X_{n+1} = X_n + \tau f\big(X_n\big) + g\big(X_n\big)\Delta W_n.$$

with $X(0) = X_0$, $T > 0$, $N \in \mathbb{N}$, $\tau = T/N$, $t_n = n\tau$.
We assume that $f = f(x)$ and $g = g(x)$ satisfy the Lipschitz condition (5.2). In the autonomous case, this implies the linear growth condition (5.3) (exercise).

Theorem 5.2.3 (strong error of the Euler-Maruyama method) *Under these conditions, there is a constant $\hat{C}$ such that*

$$\max_{n=0,\dots,N} \mathbb{E}\Big(|X(t_n) - X_n|\Big) \leq \hat{C}\tau^{1/2}$$

for all sufficiently small τ. $\hat{C}$ does not depend on τ.

For the proof we need the following

Lemma 5.2.4 (Gronwall) *Let $\alpha : [0,T] \longrightarrow \mathbb{R}_+$ be a positive integrable function. If there are constants $a > 0$ and $b > 0$ such that*

$$0 \leq \alpha(t) \leq a + b \int_0^t \alpha(s)\, ds$$

for all $t \in [0,T]$, then $\alpha(t) \leq ae^{bt}$.

Proof: exercise.

Proof of Theorem 5.2.3.
Strategy:

- Define the step function

$$Y(t) = \sum_{n=0}^{N-1} \mathbf{1}_{[t_n, t_{n+1})}(t) X_n \qquad \text{for } t \in [0,T), \qquad Y(T) := X_N.$$

  For $n = 0, \ldots, N-1$ this means that

$$Y(t) = X_n \quad \Longleftrightarrow \quad t \in [t_n, t_{n+1}).$$

- Define $\alpha(s) := \sup_{r \in [0,s]} \mathbb{E}\big(|Y(r) - X(r)|^2\big)$ and prove the Gronwall inequality

$$0 \leq \alpha(t) \leq C\tau + b \int_0^t \alpha(s)\, ds. \tag{5.4}$$

- Apply Gronwall's lemma. This yields $\alpha(t) \leq \tau \hat{C}^2$ with $\hat{C}^2 = Ce^{bt}$.
- Since[1] $\mathbb{E}(Z) \leq \sqrt{\mathbb{E}(Z^2)}$ for random variables Z, it follows that

$$\begin{aligned} \max_{n=0,\ldots,N} \mathbb{E}\big(|X_n - X(t_n)|\big) &\leq \sup_{t \in [0,T]} \mathbb{E}\big(|Y(t) - X(t)|\big) \\ &\leq \sup_{t \in [0,T]} \sqrt{\mathbb{E}\big(|Y(t) - X(t)|^2\big)} \\ &= \sqrt{\alpha(T)} \leq \sqrt{\tau}\hat{C} \end{aligned}$$

Main challenge: Prove Gronwall inequality (5.4). Choose fixed $t \in [0,T]$ and let n be the index with $t \in [t_n, t_{n+1})$.

[1] Elementary calculation: $0 \leq \mathbb{V}(Z) = \mathbb{E}\left[(Z - \mathbb{E}(Z))^2\right] = \mathbb{E}\left[Z^2 - 2Z\mathbb{E}(Z) + \mathbb{E}(Z)^2\right] = \mathbb{E}\left(Z^2\right) - \big(\mathbb{E}(Z)\big)^2$ and hence $\big(\mathbb{E}(Z)\big)^2 \leq \mathbb{E}\left(Z^2\right)$.

Derive integral representation of the error:

$$\begin{aligned} Y(t) = X_n &= X_0 + \sum_{k=0}^{n-1} (X_{k+1} - X_k) \\ &= X_0 + \sum_{k=0}^{n-1} \Big(\tau f(X_k) + g(X_k) \Delta W_k \Big) \\ &= X_0 + \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} f(X_k)\, ds + \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} g(X_k)\, dW(s) \\ &= X_0 + \int_0^{t_n} f\left(Y(s)\right)\, ds + \int_0^{t_n} g\left(Y(s)\right)\, dW(s) \end{aligned}$$

Comparing with the exact solution

$$X(t) = X(0) + \int_0^t f\big(X(s)\big) ds + \int_0^t g\big(X(s)\big) dW(s)$$

yields the error representation

$$\begin{aligned} Y(t) - X(t) = &\underbrace{\int_0^{t_n} \Big[f\left(Y(s)\right) - f\big(X(s)\big) \Big]\, ds}_{=:\mathcal{T}_1} + \underbrace{\int_0^{t_n} \Big[g\left(Y(s)\right) - g\big(X(s)\big) \Big]\, dW(s)}_{=:\mathcal{T}_2} \\ &- \underbrace{\int_{t_n}^{t} f\big(X(s)\big) ds}_{=:\mathcal{T}_3} - \underbrace{\int_{t_n}^{t} g\big(X(s)\big) dW(s)}_{=:\mathcal{T}_4} \\ = &\ \mathcal{T}_1 + \mathcal{T}_2 - \mathcal{T}_3 - \mathcal{T}_4. \end{aligned}$$

The Cauchy-Schwarz inequality gives

$$(\mathcal{T}_1 + \mathcal{T}_2 - \mathcal{T}_3 - \mathcal{T}_4)^2 = \big((1,1,-1,-1)\mathcal{T}\big)^2 \le 4\|\mathcal{T}\|_2^2 = 4 \cdot (\mathcal{T}_1^2 + \mathcal{T}_2^2 + \mathcal{T}_3^2 + \mathcal{T}_4^2)$$

and hence

$$\mathbb{E}\big(|Y(t) - X(t)|^2\big) \le 4 \cdot \mathbb{E}\left(\mathcal{T}_1^2 + \mathcal{T}_2^2 + \mathcal{T}_3^2 + \mathcal{T}_4^2\right).$$

First term: For functions $u \in L^2([0, t_n])$ the Cauchy-Schwarz inequality yields

$$\left(\int_0^{t_n} u(s) \cdot 1\, ds \right)^2 \le \int_0^{t_n} |u(s)|^2\, ds \cdot \underbrace{\int_0^{t_n} 1^2\, ds}_{=t_n}. \tag{5.5}$$

Using the Lipschitz bound (5.2), we obtain

$$\begin{aligned}
\mathbb{E}\left(\mathcal{T}_1^2\right) &= \mathbb{E}\left[\left(\int_0^{t_n}\Big[f\left(Y(s)\right)-f\big(X(s)\big)\Big]\,ds\right)^2\right] \\
&\leq t_n\mathbb{E}\left(\int_0^{t_n}\Big|f\left(Y(s)\right)-f\big(X(s)\big)\Big|^2\,ds\right) \\
&\leq TL^2\int_0^{t_n}\mathbb{E}\left(\big|Y(s)-X(s)\big|^2\right)\,ds \\
&\leq TL^2\int_0^{t}\alpha(s)\,ds \qquad (t \text{ instead of } t_n)
\end{aligned}$$

because $t \geq t_n$ by assumption.

Second term: It follows from the Itô isometry (Theorem 2.3.5) and the Lipschitz bound (5.2) that

$$\begin{aligned}
\mathbb{E}\left(\mathcal{T}_2^2\right) &= \mathbb{E}\left[\left(\int_0^{t_n}\Big[g\left(Y(s)\right)-g\big(X(s)\big)\Big]\,dW(s)\right)^2\right] \\
&= \mathbb{E}\left(\int_0^{t_n}\Big|g\left(Y(s)\right)-g\big(X(s)\big)\Big|^2\,ds\right) \\
&\leq L^2\int_0^{t_n}\mathbb{E}\left(\big|Y(s)-X(s)\big|^2\right)\,ds \\
&\leq L^2\int_0^{t}\alpha(s)\,ds
\end{aligned}$$

because $t \geq t_n$ by assumption.

Third term: Equation (5.5) and the linear growth bound (5.3) yield

$$\begin{aligned}\mathbb{E}\left(\mathcal{T}_3^2\right) &= \mathbb{E}\left[\left(\int_{t_n}^{t} f\big(X(s)\big)ds\right)^2\right] \\ &\le (t-t_n)\mathbb{E}\left(\int_{t_n}^{t}\big|f\big(X(s)\big)\big|^2 ds\right) \\ &\le \tau K\cdot\mathbb{E}\left(\int_{t_n}^{t}\big(1+|X(s)|^2\big)\,ds\right) \le c\tau^2\end{aligned}$$

because Theorem 5.2.2 states that $\mathbb{E}\,(1+|X(s)|^2)$ remains bounded on $[t_n,t]$.

Last term: Using the Itô isometry and the linear growth bound (5.3) it follows that

$$\begin{aligned}\mathbb{E}\left(\mathcal{T}_4^2\right) &= \mathbb{E}\left[\left(\int_{t_n}^{t} g\big(X(s)\big)dW(s)\right)^2\right] \\ &\le \mathbb{E}\left(\int_{t_n}^{t}\big|g\big(X(s)\big)\big|^2 ds\right) \\ &\le K\cdot\mathbb{E}\left(\int_{t_n}^{t}\big(1+|X(s)|^2\big)\,ds\right) \le c\tau\end{aligned}$$

These bounds yield the Gronwall inequality (5.4) with $b = 4(T+1)L^2$ and with C depending on K and $\sup_{s\in[0,T]}\mathbb{E}\,(1+|X(s)|^2)$. ∎

5.2.5 Weak convergence of the Euler-Maruyama method

Theorem 5.2.5 (weak error of the Euler-Maruyama method) *Under the conditions of 5.2.4, there is a constant $\hat{C}$ such that*

$$\max_{n=0,\dots,N}\Big|\mathbb{E}\big[F\big(X(t_n)\big)\big]-\mathbb{E}\big[F(X_n)\big]\Big| \le \hat{C}\tau$$

for all sufficiently small τ and all smooth functions F. $\hat{C}$ does not depend on τ.

Proof. Define piecewise linear interpolation: In addition to the piecewise constant $Y(t)$, we define the piecewise linear interpolation

$$Z(t) = X_n + (t-t_n)f\big(X_n\big) + g\big(X_n\big)\big(W(t)-W(t_n)\big) \qquad \text{for } t\in[t_n,t_{n+1}).$$

Since $Y(t) = X_n$ for $t \in [t_n, t_{n+1})$, this is equivalent to

$$Z(t) = X(0) + \int_0^t f\big(Y(s)\big)ds + \int_0^t g\big(Y(s)\big)dW(s).$$

Properties:

- $Z(t_n) = X_n = Y(t_n)$ for all $n = 0, \dots, N$.
- For every $\delta \in [0, \tau]$, $Z(t_n + \delta)$ is the Euler-Maruyama approximation after one step with step-size δ and initial value $Z(t_n) = Y_n$.
- $t \mapsto Z(t, \omega)$ is continuous with probability 1.

Apply the Feynman-Kac formula: Let $u(t, x)$ be the solution of the PDE

$$\partial_t u(t,x) + f(x)\partial_x u(t,x) + \frac{1}{2}g^2(x)\partial_x^2 u(t,x) = 0, \qquad t \in [0,T], \qquad x \in \mathbb{R}$$

with terminal condition $u(T, x) = F(x)$. Apply the Itô formula to $u(t, Z(t))$:

$$\begin{aligned} du(t,Z) &= \Big(\underbrace{\partial_t u(t,Z)}_{= \text{ ... (PDE)}} + f(Y)\partial_x u(t,Z) + \frac{1}{2}g^2(Y)\partial_x^2 u(t,Z) \Big)\, dt \\ &\quad + g(Y)\partial_x u(t,Z)\, dW(t) \\ &= \Big(\big[f(Y) - f(Z)\big]\partial_x u(t,Z) + \frac{1}{2}\big[g^2(Y) - g^2(Z)\big]\partial_x^2 u(t,Z) \Big)\, dt \\ &\quad + g(Y)\partial_x u(t,Z)\, dW(t) \end{aligned}$$

Equivalent:

$$\begin{aligned} u(T,Z(T)) = u(0,Z(0)) &+ \int_0^T \big[f(Y) - f(Z)\big]\partial_x u(t,Z)\, dt \\ &+ \frac{1}{2}\int_0^T \big[g^2(Y) - g^2(Z)\big]\partial_x^2 u(t,Z)\, dt \\ &+ \int_0^T g(Y)\partial_x u(t,Z)\, dW(t) \end{aligned}$$

By construction: $u(T, Z(T)) = u(T, X_N) = F(X_N)$
Feynman-Kac (see 2.6): $u(0, Z(0)) = u(0, X(0)) = \mathbb{E}\big[F(X(T))\big]$

This yields

$$\big|\mathbb{E}[F(X_N)] - \mathbb{E}\big[F(X(T))\big]\big| \le \underbrace{\int_0^T \Big|\mathbb{E}\Big(\big[f(Y) - f(Z)\big]\partial_x u(t,Z)\Big)\Big|\ dt}_{=:\mathcal{T}_1} + \underbrace{\frac{1}{2}\int_0^T \Big|\mathbb{E}\Big(\ \big[g^2(Y) - g^2(Z)\big]\ \partial_x^2 u(t,Z)\Big)\Big|\ dt}_{=:\mathcal{T}_2}$$

Bounds for $\mathcal{T}_1$ and $\mathcal{T}_2$: Define

$$G(t,x) = \big[f(Y) - f(x)\big]\partial_x u(t,x)$$

and apply the Itô formula to $G(t,Z)$:

$$dG(t,Z) = \left(\partial_t G(t,Z) + \partial_x G(t,Z)\cdot f(Y) + \frac{1}{2}\partial_x^2 G(t,Z)\cdot g^2(Y)\right)dt + \partial_x G(t,Z)\cdot g(Y)dW(t)$$

Equivalent:

$$G(t,Z(t)) = \underbrace{G(t_n,Z(t_n))}_{=0\ \text{(Def.)}} + \int_{t_n}^t \partial_t G(s,Z)\ ds + \int_{t_n}^t \partial_x G(s,Z)\cdot f(Y)\ ds + \frac{1}{2}\int_{t_n}^t \partial_x^2 G(t,Z)\cdot g^2(Y)\ ds + \int_{t_n}^t \partial_x G(s,Z)\cdot g(Y)\ dW(s)$$

where $Z = Z(s)$ and $Y = Y(s)$. Consider $\mathbb{E}(\ldots)$:

$$\mathbb{E}\big[G(t,Z(t))\big] = \int_{t_n}^t \mathbb{E}\Big(\partial_t G(s,Z)\Big)\ ds + \int_{t_n}^t \mathbb{E}\Big(\partial_x G(s,Z)\cdot f(Y)\Big)\ ds + \frac{1}{2}\int_{t_n}^t \mathbb{E}\Big(\partial_x^2 G(t,Z)\cdot g^2(Y)\Big)\ ds\ \ +\ \ 0$$

It can be shown that all three integrands remain bounded. It follows that

$$\big|\mathbb{E}\big[G(t,Z(t))\big]\big| \le C\cdot(t-t_n) \le C\tau.$$

Consequence:

$$\mathcal{T}_1 = \int_0^T \Big|\mathbb{E}\Big(\big[f(Y) - f(Z)\big]\partial_x u(t,Z)\Big)\Big|\ dt = \int_0^T \big|\mathbb{E}\big[G(t,Z(t))\big]\big|\ dt \le CT\tau.$$

In a similar way, it can be shown that $\mathcal{T}_2 \leq CT\tau$. This proves that

$$\Big|\mathbb{E}\big[F\big(X(t_N)\big)\big] - \mathbb{E}\big[F(X_N)\big]\Big| \leq \hat{C}\tau.$$

Replacing N by n and $T = t_N$ by t_n in the above arguments proves the assertion. ∎

5.3 Higher-order methods

Consider again the one-dimensional SDE

$$dX(t) = f\big(X(t)\big)dt + g\big(X(t)\big)dW(t), \qquad t \in [0,T], \qquad X(0) = X_0$$

with suitable functions f and g and a given initial value X_0.
Goal: Construct numerical methods with higher order.

Caution! Numerical methods for ordinary differential equations can in general not be extended to stochastic differential equations!

Example: Heun's method.

Heun's method for the ODE $\dot{y}(t) = f(y)$ with initial value $y(0) = y_0$ takes the form

$$\begin{aligned}\widetilde{y}_{n+1} &= y_n + \tau f(y_n)\\ y_{n+1} &= y_n + \frac{\tau}{2}\left(f(y_n) + f(\widetilde{y}_{n+1})\right).\end{aligned}$$

Similar to trapezoidal rule, but explicit. The natural modification of this method for SDEs is

$$\begin{aligned}\widetilde{X}_{n+1} &= X_n + \tau f(X_n) + g(X_n)\Delta W_n\\ X_{n+1} &= X_n + \frac{\tau}{2}\Big(f(X_n) + f(\widetilde{X}_{n+1})\Big) + \frac{1}{2}\Big(g(X_n) + g(\widetilde{X}_{n+1})\Big)\Delta W_n.\end{aligned}$$

Consider the special case $f(x) \equiv 0$, $g(x) = x$. For the exact solution $X(t)$, it follows that

$$\mathbb{E}(X(t)) = \mathbb{E}(X_0) + \int_0^t \mathbb{E}\big[\underbrace{f\big(X(s)\big)}_{=0}\big]\,ds + \underbrace{\mathbb{E}\left[\int_0^t g\big(s, X(s)\big)dW(s)\right]}_{=0},$$

i.e. that $\mathbb{E}(X(t))$ is constant. The method simplifies to

$$\begin{aligned}\widetilde{X}_{n+1} &= X_n + X_n\Delta W_n\\ X_{n+1} &= X_n + \frac{1}{2}\Big(X_n + \widetilde{X}_{n+1}\Big)\Delta W_n.\end{aligned}$$

or equivalently

$$X_{n+1} = X_n + X_n \Delta W_n + \frac{1}{2} X_n (\Delta W_n)^2.$$

This yields

$$\mathbb{E}(X_{n+1}) = \mathbb{E}(X_n) + \underbrace{\mathbb{E}(X_n \Delta W_n)}_{=0} + \frac{1}{2}\mathbb{E}\left(X_n (\Delta W_n)^2\right) = \mathbb{E}(X_n) + \frac{\tau}{2}\mathbb{E}(X_n)$$

and thus for $N \longrightarrow \infty$ and $\tau = T/N \longrightarrow 0$

$$\lim_{\tau \to 0} \mathbb{E}(X_N) = \lim_{\tau \to 0} (1 + \tau/2)^N X_0 = \lim_{N \to \infty} \left(1 + \frac{T}{2N}\right)^N X_0 = e^{T/2} X_0.$$

Hence, the method is not consistent! No convergence!

Stochastic Taylor expansions

Important tool for the constrution of higher-order methods.
For smooth $F : \mathbb{R} \longrightarrow \mathbb{R}$, the Itô formula yields

$$dF(X) = \underbrace{\left(F'(X) \cdot f(X) + \frac{1}{2} F''(X) \cdot g^2(X)\right)}_{=:\mathcal{L}_0 F(X)} dt + \underbrace{F'(X) \cdot g(X)}_{=:\mathcal{L}_1 F(X)} dW(t)$$

(no time derivative, because $F = F(x)$ does not depend on t) or equivalently

$$F\big(X(s)\big) = F\big(X(t_n)\big) + \int_{t_n}^{s} \mathcal{L}_0 F\big(X(r)\big)\, dr + \int_{t_n}^{s} \mathcal{L}_1 F\big(X(r)\big)\, dW(r) \tag{5.6}$$

Let $F(x) = f(x)$ and $F(x) = g(x)$, respectively, and substitute into the SDE:

$$\begin{aligned} X(t) &= X(t_n) + \int_{t_n}^{t} f\big(X(s)\big) ds + \int_{t_n}^{t} g\big(X(s)\big) dW(s) \\ &= X(t_n) + \underbrace{\int_{t_n}^{t} f\big(X(t_n)\big)\, ds}_{=(t-t_n) f\left(X(t_n)\right)} + \underbrace{\int_{t_n}^{t} \int_{t_n}^{s} \mathcal{L}_0 f\big(X(r)\big)\, dr\, ds}_{=\mathcal{T}_{11}} + \underbrace{\int_{t_n}^{t} \int_{t_n}^{s} \mathcal{L}_1 f\big(X(r)\big)\, dW(r)\, ds}_{=\mathcal{T}_{12}} \\ &\quad + \underbrace{\int_{t_n}^{t} g\big(X(t_n)\big)\, dW(s)}_{g\left(X(t_n)\right)\left[W(t)-W(t_n)\right]} + \underbrace{\int_{t_n}^{t} \int_{t_n}^{s} \mathcal{L}_0 g\big(X(r)\big)\, dr\, dW(s)}_{=\mathcal{T}_{21}} + \underbrace{\int_{t_n}^{t} \int_{t_n}^{s} \mathcal{L}_1 g\big(X(r)\big)\, dW(r)\, dW(s)}_{=\mathcal{T}_{22}} \end{aligned} \tag{5.7}$$

with

$$\mathcal{L}_0 f = f' \cdot f + \frac{1}{2} f'' \cdot g^2 \qquad \mathcal{L}_1 f = f' \cdot g$$
$$\mathcal{L}_0 g = g' \cdot f + \frac{1}{2} g'' \cdot g^2 \qquad \mathcal{L}_1 g = g' \cdot g$$

If all double integrals $\mathcal{T}_{ij}$ are ignored and $t = t_{n+1}$, we obtain the Euler-Maruyama method.

The Milstein method

Since

$$\mathbb{E}\left[\big(W(t_{n+1}) - W(t_n)\big)^2\right] = t_{n+1} - t_n,$$

we conjecture that for $t \longrightarrow t_n$ the dominant integral term is

$$\mathcal{T}_{22} = \int_{t_n}^{t} \int_{t_n}^{s} \mathcal{L}_1 g\big(X(r)\big)\, dW(r) dW(s).$$

Ignoring $\mathcal{T}_{11}, \mathcal{T}_{12}, \mathcal{T}_{21}$ yields the approximation

$$X(t) \approx X(t_n) + (t - t_n) f\big(X(t_n)\big) + g\big(X(t_n)\big)\big(W(t) - W(t_n)\big) + \mathcal{T}_{22}.$$

In order to approximate $\mathcal{T}_{22}$, we apply (5.6) with $F(x) = \mathcal{L}_1 g(x)$ and ignore higher-order terms:

$$\begin{aligned}
\mathcal{T}_{22} &= \int_{t_n}^{t} \int_{t_n}^{s} \mathcal{L}_1 g\big(X(r)\big)\, dW(r) dW(s) \\
&\approx \mathcal{L}_1 g\big(X(t_n)\big) \int_{t_n}^{t} \int_{t_n}^{s} dW(r) dW(s) \\
&= g'\big(X(t_n)\big) \cdot g\big(X(t_n)\big) \frac{1}{2} \left[\big(W(t) - W(t_n)\big)^2 - (t - t_n)\right] \qquad (5.8)
\end{aligned}$$

(exercise). For $t = t_{n+1} = t_n + \tau$, this yields the

Milstein method (Grigori N. Milstein 1974):
For $n = 0, \ldots, N-1$ let $\Delta W_n = W(t_{n+1}) - W(t_n)$ and

$$X_{n+1} = X_n + \tau f\big(X_n\big) + g\big(X_n\big)\Delta W_n + g'(X_n) \cdot g(X_n) \frac{1}{2} \left[\big(\Delta W_n\big)^2 - \tau\right].$$

Strong order 1, weak order 1 (cf. 10.3 in [KP99])

Remark: Milstein = Euler-Maruyama + additional term
If $g'(x) = 0$ ("additive noise") then Milstein = Euler-Maruyama

Problem: Have to compute derivative g'. Difficult if g is complicated or not explicitly given (i.e. no formula for g).
Idea: Avoid g' by using a (sufficiently accurate) approximation. Let

$$\widetilde{X}_{n+1} = X_n + \tau f\big(X_n\big) + \sqrt{\tau} g\big(X_n\big).$$

(similar to Euler-Maruyama, but with $\sqrt{\tau}$ instead of ΔW_n). Taylor yields

$$\begin{aligned} g(\widetilde{X}_{n+1}) &= g(X_n) + g'(X_n)\left[\widetilde{X}_{n+1} - X_n\right] + \mathcal{O}\left(\left|\widetilde{X}_{n+1} - X_n\right|^2\right) \\ &= g(X_n) + g'(X_n)\left[\tau f\big(X_n\big) + g\big(X_n\big)\sqrt{\tau}\right] + \mathcal{O}(\tau) \\ &= g(X_n) + g'(X_n) g(X_n)\sqrt{\tau} + \mathcal{O}(\tau) \end{aligned}$$

and hence

$$g'(X_n) g(X_n) = \frac{g(\widetilde{X}_{n+1}) - g(X_n)}{\sqrt{\tau}} + \mathcal{O}\big(\sqrt{\tau}\big)\,.$$

This yields the

Stochastic Milstein-Runge-Kutta method:
For $n = 0, \ldots, N-1$ let $\Delta W_n = W(t_{n+1}) - W(t_n)$ and

$$\begin{aligned} \widetilde{X}_{n+1} &= X_n + \tau f\big(X_n\big) + \sqrt{\tau} g\big(X_n\big) \\ X_{n+1} &= X_n + \tau f\big(X_n\big) + g\big(X_n\big)\Delta W_n + \frac{g(\widetilde{X}_{n+1}) - g(X_n)}{\sqrt{\tau}} \cdot \frac{1}{2}\left[\big(\Delta W_n\big)^2 - \tau\right]. \end{aligned}$$

Strong order 1, weak order 1 (in spite of the additional approximation).

Higher weak order

Go back to (5.7)

$$\begin{aligned} X(t) = X(t_n) + (t - t_n) f\big(X(t_n)\big) + \mathcal{T}_{11} + \mathcal{T}_{12} \\ + g\big(X(t_n)\big)\big[W(t_n) - W(t)\big] + \mathcal{T}_{21} + \mathcal{T}_{22} \end{aligned}$$

and approximate each double integral by freezing the integrand at t_n:

$$\mathcal{T}_{11} = \int_{t_n}^{t}\int_{t_n}^{s} \mathcal{L}_0 f\big(X(r)\big)\, dr\, ds \;\approx\; \int_{t_n}^{t}\int_{t_n}^{s} \mathcal{L}_0 f\big(X(t_n)\big)\, dr\, ds = \frac{(t-t_n)^2}{2}\mathcal{L}_0 f\big(X(t_n)\big)$$

$$\mathcal{T}_{12} = \int_{t_n}^{t}\int_{t_n}^{s} \mathcal{L}_1 f\big(X(r)\big)\, dW(r)\, ds \;\approx\; \mathcal{L}_1 f\big(X(t_n)\big)\mathcal{I}_{(1,0)}(t)$$

$$\mathcal{T}_{21} = \int_{t_n}^{t}\int_{t_n}^{s} \mathcal{L}_0 g\big(X(r)\big)\, dr\, dW(s) \;\approx\; \mathcal{L}_0 g\big(X(t_n)\big)\mathcal{I}_{(0,1)}(t)$$

$$\mathcal{T}_{22} \;\approx\; \mathcal{L}_1 g\big(X(t_n)\big)\frac{1}{2}\Big(\big[W(t)-W(t_n)\big]^2 - (t-t_n)\Big) \qquad \text{(cf. 5.8)}$$

with

$$\mathcal{I}_{(1,0)}(t) = \int_{t_n}^{t}\int_{t_n}^{s} dW(r)\, ds = \int_{t_n}^{t} W(s) - W(t_n)\, ds$$

$$\mathcal{I}_{(0,1)}(t) = \int_{t_n}^{t}\int_{t_n}^{s} dr\, dW(s) = \int_{t_n}^{t} (s-t_n)\, dW(s).$$

With integration by parts (cf. exercise 4, sheet 3) it can be shown (exercise) that

$$\mathcal{I}_{(0,1)}(t) = (t-t_n)\big[W(t)-W(t_n)\big] - \mathcal{I}_{(1,0)}(t).$$

Hence, only $\mathcal{I}_{(1,0)}(t)$ has to be computed. For weak convergence all random variables can be replaced by other random variables with the same moments. It can be shown (exercise) that

$$\mathbb{E}\left(\mathcal{I}_{(1,0)}(t)\right) = 0, \qquad \mathbb{E}\left(\mathcal{I}^2_{(1,0)}(t)\right) = \frac{(t-t_n)^3}{3}, \qquad \mathbb{E}\Big(\mathcal{I}_{(1,0)}(t)\big[W(t)-W(t_n)\big]\Big) = \frac{(t-t_n)^2}{2}.$$

Let $Z_1 \sim \mathcal{N}(0,1)$ and $Z_2 \sim \mathcal{N}(0,1)$ be independent and $W(t)-W(t_n) = \sqrt{t-t_n}Z_1$. If we let

$$Y_n := \frac{\tau^{3/2}}{2}\left(Z_1 + \frac{1}{\sqrt{3}}Z_2\right),$$

then Y_n has the same properties as $\mathcal{I}_{(1,0)}(t_{n+1})$, i.e.

$$\mathbb{E}\left(Y_n\right) = 0, \qquad \mathbb{E}\left(Y_n^2\right) = \frac{\tau^3}{3}, \qquad \mathbb{E}\Big(Y_n \Delta W_n\Big) = \frac{\tau^2}{2}$$

(exercise). For $t = t_{n+1}$, this yields the following method:

$$X_{n+1} = X_n + \tau f\big(X_n\big) + \underbrace{\frac{\tau^2}{2}\mathcal{L}_0 f\big(X_n\big)}_{\approx \mathcal{T}_{11}} + \underbrace{\mathcal{L}_1 f\big(X_n\big)Y_n}_{\approx \mathcal{T}_{12}}$$
$$+ g\big(X_n\big)\Delta W_n + \underbrace{\mathcal{L}_0 g(X_n)\big[\tau \Delta W_n - Y_n\big]}_{\approx \mathcal{T}_{21}} + \underbrace{\mathcal{L}_1 g\big(X_n\big)\frac{1}{2}\left(\Delta W_n^2 - \tau\right)}_{\approx \mathcal{T}_{22}}$$

Weak order 2 (no proof).

Simplification: The weak order is not reduced if ΔW_n is replaced by a "cheaper" random variable with the same moments. Let $\Delta V_n \in \big\{\sqrt{3\tau}, -\sqrt{3\tau}, 0\big\}$ with

$$\mathbb{P}\left(\Delta V_n = \sqrt{3\tau}\right) = \mathbb{P}\left(\Delta V_n = -\sqrt{3\tau}\right) = \frac{1}{6}, \quad \mathbb{P}(\Delta V_n = 0) = \frac{2}{3}.$$

It can be checked (exercise) that

$$\mathbb{E}\left((\Delta V_n)^k\right) = \mathbb{E}\left((\Delta W_n)^k\right) \qquad \text{for } k = 1, \ldots, 5.$$

Moreover, we can replace Y_n by

$$\widetilde{Y}_n = \frac{\tau}{2}\Delta V_n.$$

because

$$\mathbb{E}\left(\widetilde{Y}_n\right) = 0, \qquad \mathbb{E}\left(\widetilde{Y_n}^2\right) = \frac{\tau^3}{4} = \frac{\tau^3}{3} + \mathcal{O}(\tau^3), \qquad \mathbb{E}\Big(\widetilde{Y}_n \Delta V_n\Big) = \frac{\tau^2}{2}.$$

5.4 Numerical methods for systems of SDEs

Consider now the vector-valued SDE

$$X(t) = X(0) + \int_0^t f(s, X(s))\, ds + \int_0^t g(s, X(s))\, dW(s)$$

with

$$X(t) \in \mathbb{R}^d, \quad W(t) \in \mathbb{R}^m, \quad f : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}^d, \quad \text{and} \quad g : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times m}.$$

Notation as in 2.7.
The **Euler-Maruyama method** can be readily extended to vector-valued SDEs: For $n = 0, \ldots, N-1$ let $\Delta W_n = W(t_{n+1}) - W(t_n)$ and

$$X_{n+1} = X_n + \tau f\big(t_n, X_n\big) + g\big(t_n, X_n\big)\Delta W_n.$$

Strong order 1/2, weak order 1, similar proof.
What about the Milstein method?

- **Case $d \geq 1$ and $m = 1$:** Straightforward extension.
  For $n = 0, \ldots, N-1$ let $\Delta W_n = W(t_{n+1}) - W(t_n)$ and

$$X_{n+1} = X_n + \tau f\big(t_n X_n\big) + g\big(t_n, X_n\big)\Delta W_n + J_g(t_n, X_n) g(t_n, X_n) \frac{1}{2}\Big[\big(\Delta W_n\big)^2 - \tau\Big]$$

  where $J_g = \big[\partial_{x_k} g_j\big]_{j,k}$ is the Jacobian. Strong order 1, weak order 1.

- **Case $d \geq 1$ and $m > 1$:** More complicated. Adapting the derivation via stochastic Taylor expansions yields

$$X_{n+1}^{(j)} = X_n^{(j)} + \tau f\big(t_n, X_n\big) + \sum_{k=1}^{m} g_{jk}\big(t_n, X_n\big)\Delta W_n^{(k)} + \sum_{i,k=1}^{m}\sum_{l=1}^{d} \partial_{x_l} g_{jk}\big(t_n, X_n\big) \cdot g_{li}\big(t_n, X_n\big) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_i(\theta) dW_k(s)$$

  where $X_n^{(j)}$ and $\Delta W_n^{(j)}$ denote the j-th entry of X_n and ΔW_n, respectively. Similar to the scalar case, the derivatives of g can be avoided by a Runge-Kutta-type approach. The stochastic integrals can not be computed analytically. These integrals are solutions of small systems of SDEs, which have to be approximated numerically. Details: 5.3 in [GJ10].

Chapter 6

Pseudo-random numbers and Monte Carlo simulation

6.1 Uniform pseudo-random numbers

Stochastic simulations are based on random variables. In order to approximate the weak solution of an SDE, for example, the Wiener increment $\Delta W_n = W(t_{n+1}) - W(t_n)$ is simulated by drawing a random number $Z_n \sim \mathcal{N}(0,1)$ and letting $\Delta W_n = \sqrt{\tau} Z_n$; cf. 5.2.

Question: What does it mean to "draw a random number"? How can a computer generate a random number?

Computers can only generate **pseudo-random numbers**, i.e. sequences of numbers which seem to be random, but which are actually generated by a deterministic algorithm. Hence, simulation results can be reproduced if necessary. Every such sequence is periodic, but with a very large period.

First goal: generate uniform pseudo-random numbers $X_i \in [0,1]$.
MATLAB command: `rand(...)`

Method 1: Linear congruential generator

Choose $M \in \mathbb{N}$ and $a, b, X_0 \in \{0, 1, \ldots, M-1\}$ and let

$$X_i = (aX_{i-1} + b) \mod M, \qquad U_i = \frac{X_i}{M} \qquad (i = 1, 2, 3, \ldots)$$

Reminder: $x \mod y = z \Leftrightarrow x = ny + z$ for some $n \in \mathbb{N}$ and $z \in \{0, 1, \ldots, y-1\}$.
The entire sequence depends on the "seed" X_0.
"Bad" parameters must be avoided:

- $a \neq 0$
- If $b = 0$, then $X_0 \neq 0$

- $a \neq 1$ (too predictable)

MATLAB: $a = 7^5$, $b = 0$, $M = 2^{31} - 1$.
Since $X_i \in \{0, 1, \ldots, M-1\}$ the sequence $(X_i)_i$ is periodic with period $\leq M$.
MATLAB: period $\approx 2 \cdot 10^9$. Too small!

Method 2: Fibonacci generator

Choose $k, l, M \in \mathbb{N}$, let $m = \max\{k, l\}$, generate $X_1, \ldots, X_{m-1}$ with Method 1

$$X_i = (X_{i-k} + X_{i-l}) \mod M, \qquad U_i = \frac{X_i}{M} \qquad (i = m, m+1, m+2, \ldots)$$

MATLAB: $k = 31$, $l = 63$, $M = 2^{64}$, period $\approx 2^{124}$

Method 3: Combined multiple recursive generator

Choose $M_1, M_2, M \in \mathbb{N}$ very large, $a, b, c, d \in \mathbb{N}$ large

$$X_i = (aX_{i-2} - bX_{i-3}) \mod M_1$$
$$Y_i = (cY_{i-1} - dY_{i-3}) \mod M_2$$
$$Z_i = (X_i - Y_i) \mod (M-1) \qquad U = \begin{cases} Z_i/M & \text{if } Z_i \neq 0 \\ (M-1)/M & \text{if } Z_i = 0 \end{cases}$$

MATLAB parameters: p. 108 in [GJ10].
Remark: There are many more methods.

6.2 Normal pseudo-random numbers

Idea: Transform uniform pseudo-random numbers to obtain normal pseudo-random numbers.

Method 1: Inversion

Let U be uniformly distributed on $[0, 1]$, i.e. $\mathbb{P}(U \leq x) = x$ for all $x \in [0, 1]$.
Let $F : \mathbb{R} \longrightarrow [0, 1]$ be a strictly increasing probability distribution. Hence, $F^{-1} : [0, 1] \longrightarrow \mathbb{R}$ exists, and if $X := F^{-1}(U)$, then

$$\mathbb{P}(X \leq x) = \mathbb{P}\big(U \leq F(x)\big) = F(x).$$

Hence, F is the distribution of X. Apply this to the normal distribution

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2}\, ds.$$

Problem: No explicit formula for F or F^{-1}. Numerical inversion with Newton's method is ill-conditioned: If $u \approx 0$ or $u \approx 1$, then small perturbations cause large perturbations of $F^{-1}(u)$.

Method 2: Box-Muller method

Let $X \in \mathbb{R}^d$ be a random variable with density $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, and let $A := \{x \in \mathbb{R}^d : f(x) > 0\}$. Let $g : A \longrightarrow B := g(A) \subset \mathbb{R}^d$ be invertible with continuously differentiable inverse g^{-1}. If $Y = g(X)$, then

$$\mathbb{P}(Y \in C) = \mathbb{P}\big(g(X) \in C\big) = \mathbb{P}\big(X \in g^{-1}(C)\big) = \int\limits_{g^{-1}(C)} f(x)\, dx$$

$$= \int\limits_C f\Big(g^{-1}(y)\Big) \cdot |\det J_{g^{-1}}(y)|\ dy$$

for all Borel sets $C \subset B$. $J_{g^{-1}}$ denotes the Jacobian of g^{-1}. Hence, the function

$$y \mapsto f\Big(g^{-1}(y)\Big) \cdot |\det J_{g^{-1}}(y)|$$

is the density of $Y = g(X)$.

Use this to transform the uniform distribution to the normal distribution. Let $d = 1$, $A = [0,1]$, $f(x) = \mathbf{1}_A(x)$ and seek g such that for all $y \in B$

$$\underbrace{f\Big(g^{-1}(y)\Big)}_{=1 \text{ since } g^{-1}(y) \in A} \cdot |\det J_{g^{-1}}(y)| = \left|\frac{dg^{-1}}{dy}(y)\right| \stackrel{!}{=} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

Problem: No explicit formula for g. Idea: Transform in $\mathbb{R}^2$ instead of $\mathbb{R}$.
Let $A = (0,1) \times (0,1)$, $f(x) = \mathbf{1}_A$ and

$$g(x) = \begin{pmatrix} \sqrt{-2\ln x_1}\cos(2\pi x_2) \\ \sqrt{-2\ln x_1}\sin(2\pi x_2) \end{pmatrix}, \qquad x = (x_1, x_2) \in A.$$

The inverse is (exercise)[1]

$$g^{-1}(y) = \begin{pmatrix} \exp\left(-\frac{1}{2}(y_1^2 + y_2^2)\right) \\ \frac{1}{2\pi} \arctan\left(\frac{y_2}{y_1}\right) \end{pmatrix}$$

and it can be shown (exercise) that

$$|\det J_{g^{-1}}(y)| = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(y_1^2 + y_2^2)\right)$$

is the density of the standard normal distribution in $\mathbb{R}^2$. Hence:

[1] The arctan cannot yield negative values because y_1 and y_2 are not arbitrary – these values are coupled according to the definition of g.

$$g_1(X), g_2(X) \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad X_1, X_2 \text{ uniformly distributed on } (0,1).$$

Box-Muller algorithm: Generate uniformly distributed random numbers $U_1, U_2 \in (0,1)$ and let

$$Z_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2) \sim \mathcal{N}(0,1), \qquad Z_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2) \sim \mathcal{N}(0,1)$$

G. E. P. Box and M. E. Muller 1958

Method 3: Polar method

Goal: Avoid trigonometric functions.

If U_i be uniformly distributed on $(0,1)$, then $V_i = 2U_i - 1$ uniformly distributed on $(-1,1)$. Reject (V_1, V_2) if $V := V_1^2 + V_2^2 \geq 1$. The accepted samples are uniformly distributed on the unit circle (with density $f(x) = 1/\pi$), and it can be shown (exercise) that

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} V \\ \frac{1}{2\pi}\arctan\left(\frac{V_2}{V_1}\right) \end{pmatrix}$$

is uniformly distributed on $(0,1)\times(0,1)$. Hence

$$Z_1 = \sqrt{-2\ln W_1}\cos(2\pi W_2) \sim \mathcal{N}(0,1), \qquad Z_2 = \sqrt{-2\ln W_1}\sin(2\pi W_2) \sim \mathcal{N}(0,1)$$

and by definition

$$\cos(2\pi W_2) = \frac{V_1}{\sqrt{V}}, \qquad \sin(2\pi W_2) = \frac{V_2}{\sqrt{V}}.$$

Polar method: For $i \in \{1,2\}$ generate uniform random numbers $U_i \in (0,1)$ and let $V_i = 2U_i - 1$.

- If $V := V_1^2 + V_2^2 \geq 1$: reject and start again.
- Else: Let

$$Z_1 = \frac{V_1}{\sqrt{V}}\sqrt{-2\ln V} \sim \mathcal{N}(0,1), \qquad Z_2 = \frac{V_2}{\sqrt{V}}\sqrt{-2\ln V} \sim \mathcal{N}(0,1).$$

G. Marsaglia

The probability that $V < 1$ is $\pi/4$. Hence, about 21.46% of the random tuples (V_1, V_2) are rejected. Nevertheless, the polar method is usually more efficient than the standard Box-Muller method.

Correlated normal random vectors

Let $X(\omega) \in \mathbb{R}^d$, $\mu \in \mathbb{R}^d$ and let $\Sigma \in \mathbb{R}^{d\times d}$ be symmetric and positive definite.

Goal: Generate random vectors $X \sim \mathcal{N}(\mu, \Sigma)$, i.e.

$$\mathbb{P}(X \in B) = \int\limits_B \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)\, dx$$

for all Borel sets $B \subset \mathbb{R}^d$; cf. 2.1. The matrix

$$\rho \in \mathbb{R}^{d\times d} \quad \text{with entries} \quad \rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma ii \Sigma jj}}$$

is called the correlation matrix.
Reminder: Every symmetric, positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$ has a **Cholesky decomposition**

$$\Sigma = LL^T, \qquad L = \begin{pmatrix} \star & 0\cdots & \cdots & \cdots & 0 \\ \star & \star & 0 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ \star & \cdots & \cdots & \cdots & \star \end{pmatrix}, \qquad L_{ij} = 0 \text{ if } i < j.$$

Proof by induction (exercise). If $z \in \mathbb{R}^d$ and $x = Lz$, then

$$z^Tz = (L^{-1}x)^T(L^{-1}x) = x^T(LL^T)^{-1}x = x^T\Sigma^{-1}x.$$

For $A \subset \mathbb{R}^d$ and $B := \{z = L^{-1}x, x \in A\}$ we have

$$\begin{aligned}\int\limits_B \frac{1}{\sqrt{(2\pi)^d}} \exp\Big(-\frac{1}{2}z^Tz\Big)\, dz &= \int\limits_A \frac{1}{\sqrt{(2\pi)^d}} \cdot \frac{1}{|\det(L)|} \exp\Big(-\frac{1}{2}x^T\Sigma^{-1}x\Big)\, dx \\ &= \int\limits_A \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(-\frac{1}{2}x^T\Sigma^{-1}x\Big)\, dx\end{aligned}$$

because $(\det(L))^2 = \det(L)\cdot\det(L^T) = \det(\Sigma)$. Consequence:

$$Z \sim \mathcal{N}(0, I) \implies X = LZ \sim \mathcal{N}(0, \Sigma) \implies X + \mu \sim \mathcal{N}(\mu, \Sigma).$$

6.3 Monte-Carlo integration and variance reduction

Example: European basket call with $d \in \mathbb{N}$ underlyings modelled by geometric Brownian motion

$$dS(t) = rSdt + \mathrm{diag}(\sigma)\mathrm{diag}(S)LdW(t), \qquad S(t_\star) = S_\star$$

with $\sigma = (\sigma_1, \ldots, \sigma_d)$, $r > 0$, $S = (S_1, \ldots, S_d)$, $dW = (dW_1, \ldots, dW_d)$ and a lower triangular matrix $L \in \mathbb{R}^{d \times d}$.
Payoff function (cf. 1.2):

$$\psi(x) = \left(\sum_{i=1}^{d} c_i x_i - K \right)^+, \qquad c_i > 0$$

As in the scalar case (cf. 3.4), it can be shown that the value of the option is the discounted expected payoff

$$V(t_\star, S_\star) = e^{-r(T-t_\star)} \mathbb{E}_{\mathbb{Q}} \left(\psi(S_T) \right) = e^{-r(T-t_\star)} \int_0^\infty \cdots \int_0^\infty \underbrace{\psi(x) \phi(x, \xi, \beta)}_{=:g(x)} \, dx_1 \cdots dx_d$$

where ϕ is the multivariate log-normal distribution. The parameters ξ and β depend on $S_\star$, $T - t_\star$, and on the covariance matrix $\Sigma = LL^T$. In order to price the option, we thus have to approximate the d-dimensional integral

$$\int_0^\infty \cdots \int_0^\infty g(x) \, dx_1 \cdots dx_d$$

Approximation by quadrature as in 5.1:

- Truncation: $0 \leq x_{min}^{(i)} < x_{\max}^{(i)}$ such that

$$\int_{x_{min}^{(1)}}^{x_{\max}^{(1)}} \cdots \int_{x_{min}^{(d)}}^{x_{\max}^{(d)}} g(x) \, dx_1 \cdots dx_d \approx \int_0^\infty \cdots \int_0^\infty g(x) \, dx_1 \cdots dx_d$$

- Discretization: Choose large $N \in \mathbb{N}$, let $h^{(i)} = (x_{\max}^{(i)} - x_{\min}^{(i)})/N$ and $x_k^{(i)} = x_{\min}^{(i)} + k h^{(i)}$
- Approximate by quadrature (here: midpoint rule):

$$\int_{x_{min}^{(1)}}^{x_{\max}^{(1)}} \cdots \int_{x_{min}^{(d)}}^{x_{\max}^{(d)}} g(x) \, dx_1 \cdots dx_d$$

$$\approx h^{(1)} \cdot \ldots \cdot h^{(d)} \sum_{k_1=1}^{N} \ldots \sum_{k_d=1}^{N} g\left(x_{k_1}^{(1)} + 0.5 h^{(1)}, \ \ldots \ , x_{k_d}^{(d)} + 0.5 h^{(d)} \right)$$

Problem: Need function evaluations at N^d points. Exponential growth for $d \longrightarrow \infty$, "curse of dimension". Very expensive or impossible for $d \gg 1$.
Solutions? Sparse grids ($\longrightarrow$ summer term) or (Quasi-)Monte-Carlo integration.

Monte Carlo integration

Consider bounded domain $\mathcal{D} \subset \mathbb{R}^d$, function $f : \mathcal{D} \longrightarrow \mathbb{R}$, density $\phi : \mathcal{D} \longrightarrow \mathbb{R}$. As in 5.1 approximate

$$\mathbb{E}_\phi(f) := \int_{\mathcal{D}} f(x)\phi(x)\, dx \approx \frac{1}{m}\sum_{j=1}^{m} f(X_j) \tag{6.1}$$

where X_j are random numbers with

$$\mathbb{P}(X_j \in A) = \int_A \phi(x)\, dx \qquad \text{for all } A \subset \mathcal{D} \text{ measurable.}$$

Question: How accurate is this approximation?

Lemma 6.3.1 (Chebyshev's inequality) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\delta > 0$. If $Z : \Omega \longrightarrow \mathbb{R}$ is square integrable, i.e. if $\int_\Omega |Z|^2 d\mathbb{P}$ exists, then*

$$\mathbb{P}\big(|Z - \mathbb{E}(Z)| \geq \delta\big) \leq \frac{\mathbb{V}(Z)}{\delta^2}$$

where $\mathbb{V}(Z) = \mathbb{E}(|Z - \mathbb{E}(Z)|^2)$ is the variance of Z.

Proof. Define

$$\chi_\delta(\omega) = \begin{cases} 1 & \text{if } |Z(\omega) - \mathbb{E}(Z)| \geq \delta \\ 0 & \text{else.} \end{cases}$$

Then $\chi_\delta(\omega) \leq \frac{|Z(\omega)-\mathbb{E}(Z)|}{\delta}$ for all $\omega \in \Omega$ by construction, and hence

$$\begin{aligned} \mathbb{P}\big(|Z - \mathbb{E}(Z)| \geq \delta\big) &= \int_\Omega \chi_\delta(\omega)\, d\mathbb{P}(\omega) = \int_\Omega \chi_\delta^2(\omega)\, d\mathbb{P}(\omega) \\ &\leq \frac{1}{\delta^2}\int_\Omega |Z(\omega) - \mathbb{E}(Z)|^2\, d\mathbb{P}(\omega) = \frac{\mathbb{V}(Z)}{\delta^2}. \end{aligned}$$

■

Now let $Y_m := \frac{1}{m}\sum_{j=1}^m f(X_j) \approx \mathbb{E}_\phi(f)$ and

$$\mathbb{E}(f(X_j)) = \mathbb{E}_\phi(f), \qquad \mathbb{V}(f(X_j)) = \sigma^2$$

for all $j = 1, \ldots, m$ and some $\sigma > 0$. Consequence:

$$\begin{aligned} \mathbb{E}(Y_m) &= \mathbb{E}_\phi(f) \\ \mathbb{V}(Y_m) &= \mathbb{V}\left(\frac{1}{m}\sum_{j=1}^m f(X_j)\right) = \frac{1}{m^2}\sum_{j=1}^m \underbrace{\mathbb{V}\big(f(X_j)\big)}_{=\sigma^2} = \frac{\sigma^2}{m}. \end{aligned}$$

Applying Lemma 6.3.1 to Y_m yields for all $\delta > 0$

$$\mathbb{P}\big(|Y_m - \mathbb{E}_\phi(f)| \geq \delta\big) \leq \frac{\mathbb{V}(Y_m)}{\delta^2} = \frac{\sigma^2}{\delta^2 m}.$$

Now choose $\varepsilon > 0$ and let $\delta := \frac{\sigma}{\sqrt{\varepsilon m}}$:

$$\mathbb{P}\left(|Y_m - \mathbb{E}_\phi(f)| \geq \frac{\sigma}{\sqrt{\varepsilon m}}\right) \leq \varepsilon$$

or equivalently

$$\mathbb{P}\left(|Y_m - \mathbb{E}_\phi(f)| < \frac{\sigma}{\sqrt{\varepsilon m}}\right) > 1 - \varepsilon.$$

Interpretation:

$$\text{Good approximation with high probability} \quad \Leftrightarrow \quad \varepsilon \text{ small, } \frac{\sigma}{\sqrt{\varepsilon m}} \text{ small.}$$

Slow convergence: For fixed ε and σ, reducing the error by a factor of 10 comes at the cost of increasing the number of samples by a factor of 100.

Variance reduction

Idea: Try to decrease σ to improve the accuracy.

Method 1: Decomposition

Let $g : \Omega \longrightarrow \mathbb{R}$ be a function such that

$$\mathbb{E}_\phi(g) = \int_\Omega g(x)\phi(x)\,dx \approx \int_\Omega f(x)\phi(x)\,dx = \mathbb{E}_\phi(f)$$

and such that $\mathbb{E}_\phi(g)$ can be computed analytically. Let

$$Y_m = \frac{1}{m}\sum_{j=1}^{m} f(X_j) \approx \mathbb{E}_\phi(f) \qquad \text{(as before)}$$

$$Y_m^\star = \frac{1}{m}\sum_{j=1}^{m} g(X_j) \approx \mathbb{E}_\phi(g)$$

and approximate

$$\mathbb{E}_\phi(f) \approx \widehat{Y}_m := Y_m - Y_m^\star + \mathbb{E}_\phi(g).$$

Let $\mathrm{Cov}(Z_1, Z_2)$ the covariance of two random variables Z_1 and Z_2, i.e.

$$\mathrm{Cov}(Z_1, Z_2) = \mathbb{E}\Big(\big(Z_1 - \mathbb{E}(Z_1)\big)\big(Z_2 - \mathbb{E}(Z_2)\big)\Big).$$

Since

$$0 \le \mathbb{V}(Z_1 \pm Z_2) = \mathbb{V}(Z_1) + \mathbb{V}(Z_2) \pm 2\mathrm{Cov}(Z_1, Z_2) \tag{6.2}$$

it follows that

$$\mathrm{Cov}(Z_1, Z_2) \le \frac{1}{2}\big(\mathbb{V}(Z_1) + \mathbb{V}(Z_2)\big). \tag{6.3}$$

Idea: If $g \approx f$, then $Y_m^\star \approx Y_m$ and

$$\mathrm{Cov}(Y_m, Y_m^\star) \approx \frac{1}{2}\big(\mathbb{V}(Y_m) + \mathbb{V}(Y_m^\star)\big).$$

Hence, the new estimator $\widehat{Y}_m$ has a smaller variance:

$$\mathbb{V}(\widehat{Y}_m) \;=\; \mathbb{V}\left(Y_m - Y_m^\star\right) + \underbrace{\mathbb{V}\left(\mathbb{E}_\phi(g)\right)}_{=0} \overset{(6.2)}{=} \mathbb{V}\left(Y_m\right) + \mathbb{V}\left(Y_m^\star\right) - 2\mathrm{Cov}(Y_m, Y_m^\star) \;\approx\; 0.$$

Method 2: Antithetic variates

Assumption: Y_m is generated with normal random variables $X_j \sim \mathcal{N}(0,1)$.
Since $-X_j \sim \mathcal{N}(0,1)$, too, we define $Y_m^- := \frac{1}{m}\sum_{j=1}^m f(-X_j) \approx \mathbb{E}_\phi(f)$ and put $\widehat{Y}_m = \frac{1}{2}(Y_m + Y_m^-)$. Then, we obtain

$$\mathbb{V}(\widehat{Y}_m) = \frac{1}{4}\mathbb{V}(Y_m + Y_m^-) \overset{(6.2)}{=} \frac{1}{4}\Big(\mathbb{V}(Y_m) + \underbrace{\mathbb{V}(Y_m^-)}_{=\mathbb{V}(Y_m)} + 2\mathrm{Cov}(Y_m, Y_m^-)\Big).$$

If $\mathrm{Cov}(Y_m, Y_m^-) > 0$, then (6.3) yields

$$\mathrm{Cov}(Y_m, Y_m^-) \le \frac{1}{2}\big(\mathbb{V}(Y_m) + \mathbb{V}(Y_m^-)\big) = \mathbb{V}(Y_m)$$

and hence

$$\mathbb{V}(\widehat{Y}_m) = \frac{1}{2}\mathbb{V}(Y_m) + \frac{1}{2}\mathrm{Cov}(Y_m, Y_m^-) \le \mathbb{V}(Y_m) \qquad (\Longrightarrow \text{at least not worse}).$$

If $\mathrm{Cov}(Y_m, Y_m^-) \le 0$, then

$$\mathbb{V}(\widehat{Y}_m) \le \frac{1}{2}\mathbb{V}(Y_m) \qquad (\Longrightarrow \text{smaller variance}).$$

Variance reduction by antithetic variates for SDEs

Approximate the solution of the SDE

$$dX(t) = f\big(t, X(t)\big)dt + g\big(t, X(t)\big)dW(t)$$

with Euler-Maruyama with antithetic variates:

$$\begin{aligned} X^+_{n+1} &= X^+_n + \tau f\big(t_n, X^+_n\big) + g\big(t_n, X^+_n\big)\Delta W_n \\ X^-_{n+1} &= X^-_n + \tau f\big(t_n, X^-_n\big) - g\big(t_n, X^-_n\big)\Delta W_n, \qquad n = 0, \dots, N-1, \end{aligned}$$

and then use the values $X_n = \frac{1}{2}\big(X^+_n + X^-_n\big)$.

Remark. For SDE-based Monte Carlo methods, there are (at least) two sources of error:

- The **sampling error** caused by approximating the integral with a finite number of points. This error decreases when the number of realizations is increased.
- The **bias**, i.e. the error cause by solving the SDE with a numerical method (Euler-Maruyama, Milstein etc.). This error decreases when the step-size is decreased.

In application, both error sources have to be balanced ($\longrightarrow$ summer term).

6.4 Quasi Monte Carlo methods

Let $\mathcal{C} = [0,1]^d$ be the d-dimensional unit cube. Monte Carlo integration[2]:

$$\mathbb{E}(f) := \int_{\mathcal{C}} f(x)\, dx \approx \frac{1}{m}\sum_{j=1}^{m} f(X_j)$$

with uniformly distributed random vectors $X_j \in \mathcal{C}$.

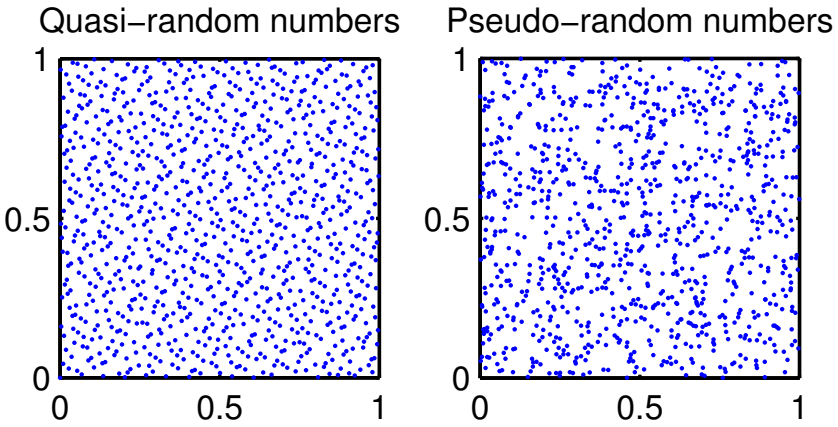
Problem: Tuples of uniform random numbers are usually not homogeneously distributed in space.
Quasi Monte Carlo methods use the same formula, but replace the random vectors X_j by deterministic low-discrepancy point sequences.

Definition 6.4.1 (Discrepancy) *1. Let $\mathcal{R}$ be the set of all axially parallel d-dimensional rectangles $R \subset \mathcal{C}$. The **discrepancy** of the points $x_1, \dots, x_m \in \mathcal{C} \subset \mathbb{R}^d$ is*

$$D_m := \sup_{R\in\mathcal{R}} \left| \frac{\#\text{of } x_i \text{ in } R}{m} - vol(R) \right|$$

where $vol(R) = \int_R 1\, dx$ denotes the volume of R.

[2]This corresponds to (6.1) with $\phi(x) = \mathbf{1}_C(x)/\mathrm{vol}(\mathcal{C})$. Since $\mathrm{vol}(\mathcal{C}) = 1$ and $\mathbf{1}_C(x) = 1$ for all $x \in \mathcal{C}$, the density ϕ can be omitted.



2. *The* ***star discrepancy*** $D_m^\star$ *is defined as* D_m*, but the supremum is only taken over those* R *for which* $(0,\dots,0)$ *is one of the corners.*

3. *A sequence* $(x_k)_{k\in\mathbb{N}}$ *of points* $x_k \in \mathcal{C}$ *is called* ***low-discrepancy sequence*** *if*

$$D_m = \mathcal{O}\left(\frac{(\log m)^d}{m}\right).$$

In this case, the x_k *are called* ***quasi-random numbers****.*

Properties (without proofs):

- $D_m^\star \le D_m \le 2^d D_m^\star$
- The Koksma-Hlawka-Theorem provides the deterministic error bound

$$\left|\mathbb{E}(f) - \frac{1}{m}\sum_{j=1}^{m} f(X_j)\right| \le TV(f)\cdot D_m^\star$$

where $TV(f)$ is the total variation of f; cf. 2.2. Numerical tests show that this bound is often too pessimistic.

- It can be shown that

$$\mathbb{E}(D_m) = \mathcal{O}\left(\sqrt{\frac{\log\log m}{m}}\right)$$

for *randomly* chosen sequences.

Examples:

For $d = 1$ the sequence with

$$x_j = \frac{2j-1}{2m}$$

has $D_m^\star = 1/(2m)$. This value is optimal. But: This sequence can only be used if m is known *a priory*, and if m is changed, then all values change.

Van der Corput sequence:

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \dots$$

Algorithm: Represent the index $j \in \mathbb{N}$ as a binary number

$$j = \sum_{k=0}^{L} d_k 2^k = (d_L, d_{L-1}, \dots, d_1, d_0)_2, \qquad d_k \in \{0, 1\}$$

and define

$$\eta_2(j) = \sum_{k=0}^{L} d_k 2^{-k-1} = (.d_0, d_1, \dots, d_L)_2, \qquad d_k \in \{0, 1\}.$$

Interpretation: Reverse binary digits and put the radix point in front of the sequence. Example: $j = 6$ yields $d_2 = d_1 = 1$, $d_0 = 0$ and hence $\eta_2(6) = \frac{1}{8} + \frac{1}{4} + 0 = \frac{3}{8}$.

Generalization: For an arbitrary base $b \in \mathbb{N}$ define the **radical-inverse function**

$$\eta_b(j) = \sum_k d_k b^{-k-1}$$

where $d_k \in \{0, 1, \dots, b-1\}$ are the coefficients from the representation $j = \sum_k d_k b^k$. The **Halton sequence** generates quasi-random numbers in the hypercube $\mathcal{C} = [0, 1]^d$ by letting

$$x_j = \big(\eta_{p_1}(j), \dots, \eta_{p_d}(j)\big)$$

where $p_1, \dots, p_d$ are prime numbers with $p_i \neq p_j$ for $i \neq j$.
Other possibility: Sobol sequence

Chapter 7

Finite-difference methods for parabolic differential equations

7.1 Motivation and model problem

Reminder (cf. 3.2): The value of an European option is the solution of the Black-Scholes equation

$$\partial_t V(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0$$

with terminal condition

$$V(T,S) = \psi(S) \qquad \text{(payoff function)}.$$

Notation: $T > 0$ maturity, $r > 0$ interest rate, $\sigma \in \mathbb{R}$ volatility, S price of the underlying.

More complicated market models (i.e. with volatility $\sigma = \sigma(t,S)$ leads to similar PDEs for which no solution formulas are available; cf. 3.5.

Question: Numerical methods?

Basic types of PDES

- **Elliptic PDEs:** Poisson equation

$$\Delta u(x) = f(x), \qquad f(x) \text{ given}, \Delta u = \sum_{k=1}^{d} \partial_{x_k}^2 u(x) \text{ Laplace operator}$$

- **Parabolic PDEs:** Heat equation

$$\partial_t u(t,x) = \Delta u(t,x)$$

- **Hyperbolic PDEs:** Wave equation

$$\partial_t^2 u(t,x) = \Delta u(t,x)$$

The Black-Scholes equation is a parabolic PDE and can be transformed to the heat equation; cf. 3.3. Therefore, we will use the one-dimensional heat equation as a model problem.

Model problem

$$\begin{aligned}
&\partial_t u(t,x) = \partial_x^2 u(t,x) && t \in [0, t_{\text{end}}], x \in [a,b] && \text{PDE} && (7.1a)\\
&u(t,a) = g_a(t), \quad u(t,b) = g_b(t) && t \in [0, t_{\text{end}}] && \text{boundary conditions} && (7.1b)\\
&u(0,x) = u_0(x) && x \in [a,b] && \text{initial condition} && (7.1c)
\end{aligned}$$

The parameters a, b, t_{end} and the boundary values $g_a(t)$ and $g_b(t)$ are given.

Lemma 7.1.1 (Solution formula) *Let $a = 0$, $b = \pi$ and $g_0(t) = g_\pi(t) = 0$. If u_0 is continuous and piecewise continuously differentiable, then the solution of* (7.1a), (7.1b), (7.1c) *is given by the Fourier series*

$$u(t,x) = \sum_{k=1}^{\infty} c_k \sin(kx) e^{-k^2 t} \quad \textit{with} \quad c_k = \frac{2}{\pi} \int_0^\pi u_0(x) \sin(kx)\, dx. \tag{7.2}$$

Proof. It can be shown that the Fourier series converges uniformly (cf. §6, 2.4 in volume 2 of [FK08]). Hence, the derivatives can be applied to each term of the sum, and it can be checked that each term solves (7.1a) and (7.1b). (7.1c) follows from the fact that

$$\int_0^\pi \sin(jx)\sin(kx)\, dx = \begin{cases} 0 & \text{if } j \neq k \\ \frac{\pi}{2} & \text{if } j = k. \end{cases}$$

See exercises for details. ∎

Remark:

1. The function

$$\tilde{u}(t,x) = g_0 + \frac{g_\pi - g_0}{\pi} x + u(t,x)$$

solves (7.1a), (7.1b), (7.1c) in the special case $a = 0$, $b = \pi$ with arbitrary but constant g_0 and g_π. Solutions on arbitrary intervals can be constructed by rescaling.

2. Since (7.1a) involves $\partial_x^2 u(t,x)$, one may expect that all solutions of the PDE are twice continuously differentiable with respect to x. Lemma 7.1.1 shows, however, that solutions with lower regularity of the initial data exist. This is important for applications in mathematical finance, because the payoff function of an European put or call is not differentiable, but only piecewise smooth.

   In fact, solutions of parabolic problems such as the heat equation typically become smoother as time evolves. All terms in the series representation (7.2) oscillate in space due to the term $\sin(kx)$, and the higher k, the faster the oscillations. The k-th term, however, is multiplied with $c_k e^{-k^2 t}$, i.e. it decays exponentially as time evolves, and the larger k, the faster the decay.

Notation: We say that $g \in C^j([a,b])$ if and only if $x \mapsto g(x)$ is j times continuously differentiable on (a,b), and all derivatives can be extended to $[a,b]$. Moreover, let $\|g\|_\infty = \max_{x\in[a,b]} |g(x)|$ denotes the maximum norm on $[a,b]$.

7.2 Space discretization with finite differences

Choose $1 < m \in \mathbb{N}$, let $h = (b-a)/m$ and $x_k = a + k \cdot h$ for $k = 0, \ldots, m$.
Goal: Find $v : [0, t_{\text{end}}] \longrightarrow \mathbb{R}^{m-1}$ such that the entries $v_k(t)$ of the vector $v(t)$ approximate the values of the exact solution at the inner grid points, i.e.

$$v_k(t) \approx u(t, x_k) \qquad\qquad (k = 0, \ldots, m)$$

for all $t \in [0, t_{\text{end}}]$. Approximate spatial derivatives by difference quotients.

Lemma 7.2.1 (difference quotients) *The derivatives of a function $y : [a,b] \longrightarrow \mathbb{R}$ can be approximated as follows:*

- *If $y \in C^2([a,b])$, then*

$$\max_{k=1,\ldots,m-1} \left| y'(x_k) - \frac{y(x_k+h) - y(x_k)}{h} \right| \le Ch\|y''\|_\infty \tag{7.3}$$

- *If $y \in C^2([a,b])$, then*

$$\max_{k=1,\ldots,m-1} \left| y'(x_k) - \frac{y(x_k) - y(x_k-h)}{h} \right| \le Ch\|y''\|_\infty \tag{7.4}$$

- *If $y \in C^3([a,b])$, then*

$$\max_{k=1,\ldots,m-1} \left| y'(x_k) - \frac{y(x_k+h) - y(x_k-h)}{2h} \right| \le Ch^2 \left\| \frac{d^3 y}{dx^3} \right\|_\infty \tag{7.5}$$

- *If $y \in C^4([a,b])$, then*

$$\max_{k=1,\dots,m-1} \left| y''(x_k) - \frac{y(x_k+h) - 2y(x_k) + y(x_k-h)}{h^2} \right| \leq Ch^2 \left\| \frac{d^4 y}{dx^4} \right\|_\infty \tag{7.6}$$

The constant C is independent of h and can have different values in each case.

Proof: Use Taylor's theorem (exercise).

Since $y(x_k \pm h) = y(x_{k\pm 1})$, applying (7.6) to the heat equation yields

$$\partial_t u(t,x_k) = \partial_x^2 u(t,x_k) \approx \frac{u(t,x_{k+1}) - 2u(t,x_k) + u(t,x_{k-1})}{h^2} \tag{7.7}$$

for all $k = 1,\dots,m-1$ and $t \in [0,t_{\text{end}}]$. The boundary values $u(t,x_0) = g_a(t)$ and $u(t,x_m) = g_b(t)$ are known from (7.1b).

Reformulation in matrix-vector notation: Define

$$\bar{u}(t) = \Big(u(t,x_1),\dots,u(t,x_{m-1})\Big)^T \in \mathbb{R}^{m-1} \tag{7.8}$$

$$f^{(m)}(t) = \Big(g_a(t),0,0,\dots,0,0,g_b(t)\Big)^T \in \mathbb{R}^{m-1} \tag{7.9}$$

and the matrix

$$A^{(m)} = \begin{pmatrix} -2 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & -2 & 1 & 0 & & & \vdots \\ 0 & 1 & -2 & 1 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & 1 & -2 & 1 & 0 \\ \vdots & & & 0 & 1 & -2 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -2 \end{pmatrix} \in \mathbb{R}^{(m-1)\times(m-1)}. \tag{7.10}$$

Then (7.7) is equivalent (check!) to

$$\bar{u}'(t) \approx \frac{1}{h^2} A^{(m)} \bar{u}(t) + \frac{1}{h^2} f^{(m)}(t).$$

The approximation $v(t) \approx \bar{u}(t)$ is now defined as the solution of the initial value problem

$$v'(t) = \frac{1}{h^2} A^{(m)} v(t) + \frac{1}{h^2} f^{(m)}(t) \tag{7.11a}$$

$$v(0) = \bar{u}(0). \tag{7.11b}$$

Roughly speaking, we have turned a PDE into an ODE.

Properties of the discretization matrix

Lemma 7.2.2 *For $k = 1, \dots, m-1$, the vectors*

$$\nu_k^{(m)} := \left(\sin\left(\frac{k\pi}{m}\right), \sin\left(\frac{2k\pi}{m}\right), \dots, \sin\left(\frac{(m-1)k\pi}{m}\right) \right)^T \in \mathbb{R}^{m-1}$$

are eigenvectors of the matrix $A^{(m)}$, and the corresponding eigenvalues are

$$\lambda_k^{(m)} := 2\left(\cos\left(\frac{k\pi}{m}\right) - 1\right) \in (-4, 0).$$

Proof: Exercise.

Definition 7.2.3 (scaled norm) *Let $|\cdot|_m$ be the scaled vector norm defined by*

$$|v|_m = \left(\frac{1}{m-1} \sum_{i=1}^{m-1} v_i^2 \right)^{1/2} = \frac{\sqrt{v^T v}}{\sqrt{m-1}}$$

Remark: If $g \in L^2([a,b])$ is square integrable with $g(a) = g(b) = 0$ and $v_k = g(x_k)$ for $k = 1, \dots, m-1$, then

$$\|g\|_{L^2([a,b])} = \left(\int_a^b |g(x)|^2 \, dx \right)^{1/2} \approx \left(\frac{b-a}{m} \sum_{k=1}^{m-1} |g(x_k)|^2 \, dx \right)^{1/2} \approx \sqrt{b-a} \cdot |v|_m$$

for sufficiently large m.

Lemma 7.2.4 *For any $s \geq 0$, $1 < m \in \mathbb{N}$ and $h = (b-a)/m$ we have*

$$\left| e^{sA^{(m)}} \right|_m \leq 1, \qquad \left| \left(I - sA^{(m)}\right)^{-1} \right|_m \leq 1$$

where $e^M = \sum_{k=0}^{\infty} \frac{M^k}{k!}$ is the matrix exponential function and I is the identity matrix.

Proof. Choose fixed $m \in \mathbb{N} \setminus \{1\}$ and write A instead of $A^{(m)}$. Since A is symmetric, there is a orthogonal matrix Q such that $A = Q\Lambda Q^T$ with

$$\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_{m-1}), \qquad \lambda_k = 2\left(\cos\left(\frac{k\pi}{m}\right) - 1\right) < 0.$$

Hence,

$$A^2 = Q\Lambda Q^T Q\Lambda Q^T = Q\Lambda^2 Q^T, \qquad A^k = Q\Lambda^k Q^T$$

and as a consequence

$$e^{sA} = \sum_{k=0}^{\infty} Q \frac{(s\Lambda)^k}{k!} Q^T = Q e^{s\Lambda} Q^T.$$

This yields the bound

$$\left|e^{sA}\right|_m \leq |Q|_m \cdot \left|e^{s\Lambda}\right|_m \cdot |Q^T|_m \leq 1,$$

because

$$\begin{aligned} |Q|_m^2 &= \sup_{v \neq 0} \frac{\frac{1}{m-1} v^T Q^T Q v}{\frac{1}{m-1} v^T v} = 1 && \text{(since } Q \text{ is orthogonal)} \\ \left|e^{s\Lambda}\right|_m &= \sup_{k=1,\ldots,m-1} |e^{s\lambda_k}| \leq 1 && \text{(since } s\lambda_k \leq 0\text{).} \end{aligned}$$

In a similar way, it can be checked that

$$(I - sA)^{-1} = Q \left(I - s\Lambda\right)^{-1} Q^T$$

which yields the bound

$$\left|(I - sA)^{-1}\right|_m \leq |Q|_m \cdot \left|(I - s\Lambda)^{-1}\right|_m \cdot |Q^T|_m = \max_{k=1,\ldots,m-1} \frac{1}{1 - s\lambda_k} \leq 1$$

since $s\lambda_k \leq 0$. ∎

Theorem 7.2.5 (error of the space discretization) *Let $u = u(t,x)$ be the solution of the model problem (7.1a), (7.1b), (7.1c). Assume that $t \mapsto u(t,x) \in C^1([0,t_{end}])$ and that $x \mapsto u(t,x) \in C^4([a,b])$ for all $t \in [0,t_{end}]$ with*

$$\sup_{s \in [0,t_{end}]} \left\|\partial_x^4 u(s,\cdot)\right\|_\infty < \infty.$$

Then, there is a constant C such that

$$|\bar{u}(t) - v(t)|_m \leq C t_{end} h^2 \sup_{s \in [0,t_{end}]} \left\|\partial_x^4 u(s,\cdot)\right\|_\infty$$

for all $t \in [0,t_{end}]$.

Proof. By definition of $\bar{u}$ we have

$$\bar{u}_k'(t) = \partial_t u(t,x_k) = \partial_x^2 u(t,x_k) = \frac{u(t,x_{k+1}) - 2u(t,x_k) + u(t,x_{k-1})}{h^2} + r_k(t)$$

for $k = 1, \dots, m-1$. According to Lemma 7.2.1, the remainder term $r_k(t)$ can be bounded by

$$|r_k(t)| \le Ch^2 \left\|\partial_x^4 u(t,\cdot)\right\|_\infty .$$

Setting $r := (r_1, \dots, r_{m-1})^T$, this is equivalent to

$$\bar{u}'(t) = \frac{1}{h^2} A^{(m)} \bar{u}(t) + \frac{1}{h^2} f^{(m)}(t) + r(t).$$

Comparing with (7.11a) shows that the error solves the ODE

$$\frac{d}{dt}\big(\bar{u}(t) - v(t)\big) = \frac{1}{h^2} A^{(m)} \big(\bar{u}(t) - v(t)\big) + r(t).$$

with initial value $\bar{u}(0) - v(0) = 0$. The solution is given by the variation-of-constants formula

$$\big(\bar{u}(t) - v(t)\big) = e^{tA^{(m)}/h^2} \underbrace{\big(\bar{u}(0) - v(0)\big)}_{=0} + \int_0^t e^{(t-s)A^{(m)}/h^2} r(s)\, ds,$$

and hence

$$\begin{aligned} \big|\bar{u}(t) - v(t)\big|_m &\le \int_0^t \underbrace{\left|e^{(t-s)A^{(m)}/h^2}\right|_m}_{\le 1} |r(s)|_m \, ds \le \int_0^t Ch^2 \left\|\partial_x^4 u(s,\cdot)\right\|_\infty \, ds \\ &\le Ct_{\text{end}} h^2 \sup_{s\in[0,t_{\text{end}}]} \left\|\partial_x^4 u(s,\cdot)\right\|_\infty . \end{aligned}$$

■

Remark. The regularity assumptions are very strong and exclude payoff functions as initial data.

7.3 Time discretization with Runge-Kutta methods

(a) Runge-Kutta methods: Examples

Consider the initial value problem

$$y'(t) = f(t,y), \qquad t \in [t_0, t_{\text{end}}], \qquad y(0) = y_0$$

with appropriate function $f : [t_0, t_{\text{end}}] \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ (e.g. $y \mapsto f(t,y)$ Lipschitz continuous).

Choose $N \in \mathbb{N}$, let $\tau = (t_{\text{end}} - t_0)/N$ and $t_n = t_0 + n\tau$.
Goal: Find approximation $y_n \approx y(t_n)$

Ansatz:

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} y'(s)\,ds = y(t_n) + \int_{t_n}^{t_{n+1}} f\big(s, y(s)\big)\,ds$$

Approximate the integral by quadrature:

$$y(t_n + \tau) \approx y(t_n) + \tau \sum_{i=1}^{s} b_i f\big(t_n + c_i\tau, y(t_n + c_i\tau)\big)\,ds$$

with $s \in \mathbb{N}$ and coefficients b_i and c_i. Approximate the unknown function evaluations by

$$y(t_n + c_i\tau) \approx y(t_n) + \tau \sum_{j=1}^{s} a_{ij} f\big(t_n + c_j\tau, y(t_n + c_j\tau)\big)\,ds, \qquad i = 1, \dots, s$$

with coefficients a_{ij}. This yields the **Runge-Kutta method**: For each $n = 1, \dots, N-1$ solve the nonlinear system

$$Y_i = y_n + \tau \sum_{j=1}^{s} a_{ij} f\big(t_n + c_j\tau, Y_j\big)\,ds, \qquad i = 1, \dots, s \tag{7.12a}$$

(e.g. by a version of Newton's method) and let

$$y_{n+1} = y_n + \tau \sum_{i=1}^{s} b_i f\big(t_n + c_i\tau, Y_i\big)\,ds. \tag{7.12b}$$

Each Runge-Kutta method is characterized by its coefficients a_{ij}, b_i, c_j. These are represented in the Butcher tableau:

$$\begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array}$$

Examples:

- Explicit Euler method: $y_{n+1} = y_n + \tau f(t_n, y_n)$
- Implicit Euler method: $y_{n+1} = y_n + \tau f(t_{n+1}, y_{n+1})$
- Trapezoidal rule: $y_{n+1} = y_n + \frac{\tau}{2}\big(f(t_n, y_n) + f(t_{n+1}, y_{n+1})\big)$

0	0
	1

explicit Euler

1	1
	1

implicit Euler

0	0	0
1	1/2	1/2
	1/2	1/2

trapezoidal rule

The Runge-Kutta method (7.12a), (7.12b) is explicit if $a_{ij} = 0$ for all $j \geq i$.
Order conditions:

- Order 1:

$$\sum_{j=1}^{s} b_j = 1, \qquad \sum_{j=1}^{s} a_{ij} = c_i \quad \text{for all } i = 1, \dots, s$$

- Order 2:

$$\sum_{j=1}^{s} b_j c_j = \frac{1}{2}$$

- Order 3:

$$\sum_{j=1}^{s} b_j c_j^2 = \frac{1}{3} \qquad \sum_{i=1}^{s} b_i \sum_{j=1}^{s} a_{ij} c_j = \frac{1}{6}$$

More complicated conditions for higher order. Both Euler methods have order 1, the trapezoidal rule has order 2.

(b) A-stability

Consider the heat equation after space discretization (cf. (7.11a))

$$v'(t) = \frac{1}{h^2} A^{(m)} v(t) + \frac{1}{h^2} f^{(m)}(t)$$

For simplicity: Assume homogeneous boundary conditions in (7.1b), i.e. $u(t,a) = u(t,b) = 0$ and hence $f^{(m)}(t) = 0$.
As in the proof of Lemma 7.2.4 we consider the eigendecomposition

$$A^{(m)} = Q^{(m)} \Lambda^{(m)} (Q^{(m)})^T, \qquad \Lambda^{(m)} = \operatorname{diag}(\lambda_1^{(m)}, \dots, \lambda_{m-1}^{(m)}), \qquad \lambda_k^{(m)} < 0 \tag{7.13}$$

with $|Q^{(m)}|_m \le 1$ and $\lambda_k^{(m)} \in (-4, 0)$. The exact solution

$$v(t) = \exp\left(tA^{(m)}/h^2\right) v(0) = Q^{(m)} \exp\left(t\Lambda^{(m)}/h^2\right) (Q^{(m)})^T v(0)$$

remains bounded for all $t \ge 0$ because

$$|v(t)|_m \le \underbrace{\left|\exp\left(t\Lambda^{(m)}/h^2\right)\right|_m}_{\le 1} |v(0)|_m.$$

Explicit Euler:

$$v_{n+1} = v_n + \frac{\tau}{h^2} A^{(m)} v_n = \left(I + \frac{\tau}{h^2} A^{(m)}\right)^{n+1} v_0$$

(We write $v_n \in \mathbb{R}^m$ for the n-th approximation and $v_{n,k}$ for the k-th entry of v_n.)

$$|v_n|_m \leq \left|I + \frac{\tau}{h^2} A^{(m)}\right|_m^n |v_0|_m \leq \left|I + \frac{\tau}{h^2} \Lambda^{(m)}\right|_m^n |v_0|_m$$

Hence, the numerical solution remains bounded for all $n \in \mathbb{N}$ if

$$\left|1 + \frac{\tau}{h^2} \lambda_k^{(m)}\right| \leq 1 \quad \Longleftrightarrow \quad \tau \leq \frac{2h^2}{|\lambda_k^{(m)}|}$$

for all $k = 1, \ldots, m-1$. Since $\max_{k=1,\ldots,m-1} |\lambda_k^{(m)}| \approx 4$, we obtain the **stability condition**

$$\tau \leq \frac{h^2}{2}.$$

This is a severe restriction, because $h \ll 1$ must be small to ensure an acceptable accuracy of the spatial approximation. For larger step-sizes, the norm of the numerical solution may tend to ∞ whereas the exact solution remains bounded. Reducing the step size τ, however, increases the number of steps and hence the numerical costs. Inefficient!

Definition 7.3.1 (A-stability) *The initial value problem*

$$y' = \lambda y, \quad t \geq 0, \qquad\qquad y(0) = y_0, \tag{7.14}$$

is called ***Dahlquist's test equation****. A one-step method (e.g. Runge-Kutta) is called* ***A-stable*** *if the numerical solution $(y_n)_{n\in\mathbb{N}}$ of (7.14) with arbitrary $\lambda \in \mathbb{C}$, $Re(\lambda) \leq 0$ and arbitrary step-size $\tau > 0$ remains bounded for all $n \in \mathbb{N}$.*

Remark. If ν_k is the k-th eigenvector of $A^{(m)}$, then

$$v'(t) = \frac{1}{h^2} A^{(m)} v(t), \qquad\qquad v(0) = \nu_k$$

is equivalent to (7.14) with $\lambda := \lambda_k^{(m)}/h^2 < 0$.
The explicit Euler method (and every other explicit Runge-Kutta method) is not A-stable.

Implicit Euler:

The implicit Euler

$$v_{n+1} = v_n + \frac{\tau}{h^2} A^{(m)} v_{n+1} \qquad \text{or equivalently} \qquad \left(I - \frac{\tau}{h^2} A^{(m)}\right) v_{n+1} = v_n$$

is the simplest A-stable method. Hence, we expect that the corresponding numerical approximation of

$$v'(t) = \frac{1}{h^2} A^{(m)} v(t)$$

remains bounded without any step-size restrictions. This is indeed the case:

$$|v_n|_m \leq \left|I - \frac{\tau}{h^2} A^{(m)}\right|_m^{-1} |v_{n-1}|_m \leq |v_{n-1}|_m$$

due to Lemma 7.2.4, and hence $|v_n|_m \leq |v_0|_m$ for all $n \in \mathbb{N}$.

Trapezoidal rule:

The trapezoidal rule is A-stable (exercise).

(c) Error bound for the implicit Euler method

Let $\langle z, \tilde{z}\rangle = z^T\tilde{z}$ denote the Euclidean scalar product on $\mathbb{R}^d$ with norm $|z| = \sqrt{\langle z, z\rangle}$.

Theorem 7.3.2 (error bound for the implicit Euler method) *Let $f : [t_0, t_{end}]\times\mathbb{R}^d \longrightarrow \mathbb{R}^d$ be C^1 and assume that f satisfies the one-sided Lipschitz condition*

$$\Big\langle f(t,z) - f(t,\tilde{z}), z - \tilde{z}\Big\rangle \le \ell|z-\tilde{z}|^2 \tag{7.15}$$

with a constant $\ell \in \mathbb{R}$ for all $t \in [t_0, t_{end}]$ and $z, \tilde{z} \in \mathbb{R}^d$. Let $y(t)$ be the exact solution of the initial-value problem

$$y' = f(t,y), \quad t \in [t_0, t_{end}], \qquad y(t_0) = y_0,$$

and let $y_n \approx y(t_n)$ be the approximations computed with the implicit Euler method with step-size $\tau = (t_{end} - t_0)/N$. If $\tau\ell < 1$, then the error is bounded by

$$\max_{n=0,\dots,N} |y_n - y(t_n)| \le \frac{C}{2} \max_{t\in[t_0,t_{end}]} |y''(t)| \cdot \tau \tag{7.16}$$

with

$$C := \begin{cases} 1/|\ell| & \text{if } \ell < 0 \\ (t_{end} - t_0) & \text{if } \ell = 0 \\ \dfrac{e^{\ell(t_{end}-t_0)/(1-\tau\ell)} - 1}{\ell} & \text{if } \ell > 0. \end{cases}$$

Remarks:

- The assumption $\tau\ell < 1$ guarantees that the nonlinear problem which has to be solved in each time-step has indeed a solution; cf. Satz 75.1, p. 561, in [HB09].
- For large positive $\ell \gg 1$, the implicit Euler method is usually not suited, because then the condition $\tau\ell < 1$ imposes very small time-steps. The error bound also shows that this condition is not sufficient for an accurate approximation: If $\tau\ell < 1$ but $\tau\ell \approx 1$, then $e^{\ell(t_{\rm end}-t_0)/(1-\tau\ell)} \gg 1$ and hence $C \gg 1$.

  In many applications, however, we have $\ell \le 0$. In this case, the assumption $\tau\ell < 1$ is not a restriction.
- For Dahlquist's test equation (7.14) with real $\lambda < 0$, we can choose $\ell = \lambda$.

Proof. For simplicity consider only autonomous ODEs, i.e. $y' = f(y)$. Implicit Euler:

$$y_{n+1} = \Phi_\tau(y_n) = y_n + \tau f(y_{n+1}) = y_n + \tau f\Big(\Phi_\tau(y_n)\Big). \tag{7.17}$$

Step 1: Local error. Taylor expansion of the exact solution:

$$y(t_n) = y(t_{n+1}) - \tau \underbrace{y'(t_{n+1})}_{=f(y(t_{n+1}))} + d_{n+1} \tag{7.18}$$

$$\text{with} \quad |d_{n+1}| \leq \underbrace{\frac{1}{2} \max_{t\in[t_0,t_{\text{end}}]} |y''(t)|}_{=:\hat{C}} \tau^2.$$

(7.17), (7.18) and the Cauchy-Schwarz inequality yield

$$\begin{aligned}
&|y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big)|^2 \\
&= \left\langle \underbrace{y(t_{n+1})}_{=\dots\ (7.18)} - \underbrace{\Phi_\tau\big(y(t_n)\big)}_{=\dots\ (7.17)}, y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big) \right\rangle \\
&= \Big\langle y(t_n) + \tau f\big(y(t_{n+1})\big) - y(t_n) - \tau f\Big(\Phi_\tau\big(y(t_n)\big)\Big), y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big) \Big\rangle \\
&\quad - \big\langle d_{n+1}, y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big) \big\rangle \\
&\overset{(7.15)}{\leq} \tau\ell \left|y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big)\right|^2 + |d_{n+1}| \cdot \left|y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big)\right|
\end{aligned}$$

and hence

$$(1-\tau\ell) \left|y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big)\right|^2 \leq |d_{n+1}| \cdot \left|y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big)\right|.$$

Since $1 - \tau\ell > 0$ by assumption, it follows that

$$\left|y(t_{n+1}) - \Phi_\tau\big(y(t_n)\big)\right| \leq \frac{1}{(1-\tau\ell)} |d_{n+1}| \leq \frac{\hat{C}}{(1-\tau\ell)} \tau^2.$$

Step 2: Stability. For any $z, \hat{z} \in \mathbb{R}^d$ we have

$$\begin{aligned}
&|\Phi_\tau(z) - \Phi_\tau(\hat{z})|^2 \\
&= \big\langle z + \tau f\big(\Phi_\tau(z)\big) - \hat{z} - \tau f\big(\Phi_\tau(\hat{z})\big), \Phi_\tau(z) - \Phi_\tau(\hat{z}) \big\rangle \\
&\overset{(7.15)}{\leq} |z - \hat{z}| \cdot |\Phi_\tau(z) - \Phi_\tau(\hat{z})| + \tau\ell |\Phi_\tau(z) - \Phi_\tau(\hat{z})|^2.
\end{aligned}$$

Using again that $1 - \tau\ell > 0$ by assumption, this yields

$$|\Phi_\tau(z) - \Phi_\tau(\hat{z})| \leq \frac{1}{1-\tau\ell} |z - \hat{z}|.$$

For the error after k steps, this implies

$$|\Phi_\tau^k(z) - \Phi_\tau^k(\hat{z})| \leq \frac{1}{(1-\tau\ell)^k} |z - \hat{z}|.$$

Step 3: Error accumulation and global error. Let $y_n = \Phi_\tau^n(y_0)$ be the approximation after n steps. We represent the global error by the telescope sum ("Lady Windemere's fan")

$$\begin{aligned} y(t_n) - y_n &= \Phi_\tau^0\big(y(t_n)\big) - \Phi_\tau^n(y_0) \\ &= \sum_{k=0}^{n-1} \big(\Phi_\tau^k\big(y(t_{n-k})\big) - \Phi_\tau^{k+1}\big(y(t_{n-k-1})\big)\big)\,. \end{aligned} \tag{7.19}$$

From step 1 and 2, we know that

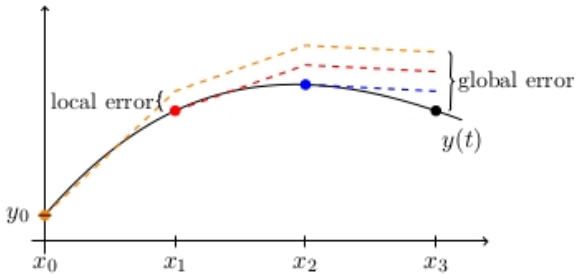


Figure 7.1: Lady Windemere's fan. The term was coined in [HNrW10] as an allusion to a play of the same title by Oscar Wilde. Thanks to Lukas Baron for providing this picture.

$$\begin{aligned} \big|\Phi_\tau^k\big(y(t_{n-k})\big) - \Phi_\tau^{k+1}\big(y(t_{n-k-1})\big)\big| &\le \frac{1}{(1-\tau\ell)^k}\,\big|y(t_{n-k}) - \Phi_\tau\big(y(t_{n-k-1})\big)\big| \\ &\le \frac{\hat{C}}{(1-\tau\ell)^{k+1}}\tau^2. \end{aligned}$$

Taking norms in (7.19) and applying the triangle inequality thus gives

$$\begin{aligned} |y(t_n) - y_n| &\le \frac{\hat{C}}{1-\tau\ell}\tau^2 \sum_{k=0}^{n-1} \frac{1}{(1-\tau\ell)^k} = \frac{\hat{C}}{1-\tau\ell}\tau^2 \cdot \frac{(1-\tau\ell)^{-n}-1}{(1-\tau\ell)^{-1}-1} \\ &= \hat{C}\tau \cdot \frac{(1-\tau\ell)^{-n}-1}{\ell} \end{aligned}$$

- If $\ell < 0$, then

$$\frac{(1-\tau\ell)^{-n}-1}{\ell} = \frac{1-(1+\tau|\ell|)^{-n}}{|\ell|} \le \frac{1}{|\ell|}$$

because $(1+\tau|\ell|)^{-n} > 0$.

- If $\ell > 0$, then $1 - \tau\ell < 1$. By assuption, we also know that $1 - \tau\ell > 0$, i.e. $1 - \tau\ell \in (0, 1)$ and hence $(1 - \tau\ell)^{-n} > 1$. This yields

$$(1 - \tau\ell)^{-n} = \left(1 + \frac{\tau\ell}{1 - \tau\ell}\right)^n \leq \left(e^{\tau\ell/(1-\tau\ell)}\right)^n = e^{n\tau\ell/(1-\tau\ell)} \leq e^{\ell(t_{\text{end}}-t_0)/(1-\tau\ell)}$$

  because $1 + \xi \leq e^{\xi}$ for all $\xi \in \mathbb{R}$ and $n\tau \leq t_{\text{end}} - t_0$ for all $n = 0, \ldots, N$.

- The case $\ell = 0$ is left as an exercise. ■

General principle: Consistency + stability $\Longrightarrow$ convergence

7.4 Approximation of the heat equation in time and space

Back to the heat equation:

$$\begin{aligned} \partial_t u(t,x) &= \partial_x^2 u(t,x) && t \in [0, t_{\text{end}}], x \in [a,b] \\ u(t,a) = g_a(t), \quad u(t,b) &= g_b(t) && t \in [0, t_{\text{end}}] \\ u(0,x) &= u_0(x) && x \in [a,b], \end{aligned}$$

cf. (7.1a)-(7.1c). Semi-discretization in space leads to the initial-value problem

$$\begin{aligned} v'(t) &= \frac{1}{h^2} A^{(m)} v(t) + \frac{1}{h^2} f^{(m)}(t) \\ v(0) &= \bar{u}(0) \end{aligned}$$

with $v_k(t) \approx u(t, x_k)$ for $k = 1, \ldots, m-1$; cf. (7.11a)-(7.11b).
Apply the implicit Euler method:

$$w_{n+1} = w_n + \frac{\tau}{h^2} A^{(m)} w_{n+1} + \frac{\tau}{h^2} f^{(m)}(t_{n+1})$$

Remark: To avoid confusion, we point out that $v_k(t) \in \mathbb{R}$ is a **scalar** for each t (namely the k-th entry of $v(t)$), whereas $w_n \in \mathbb{R}^{m-1}$ is a **vector** for each n. The k-th entry of w_n is denoted by $w_{n,k}$. By definition, we expect that

$$w_{n,k} \approx v_k(t_n) \approx u(t_n, x_k).$$

In order to apply Theorem 7.3.2, we need the following

Lemma 7.4.1 *For every $1 < m \in \mathbb{N}$ and $h = (b-a)/m$ there is a constant $\ell < 0$ such that the function*

$$F(t,w) = \frac{1}{h^2} A^{(m)} w + \frac{1}{h^2} f^{(m)}(t)$$

satisfies the one-sided Lipschitz condition

$$\Big\langle F(t,w) - F(t,\tilde{w}), w - \tilde{w} \Big\rangle \leq \ell |w - \tilde{w}|^2$$

for all $t \in [t_0, t_{end}]$ and $w, \tilde{w} \in \mathbb{R}^{m-1}$.

Proof. The eigendecomposition $A^{(m)} = Q^{(m)} \Lambda^{(m)} (Q^{(m)})^T$ from (7.13) yields

$$\begin{aligned}
\Big\langle F(t,w) - F(t,\tilde{w}), w - \tilde{w} \Big\rangle &= \frac{1}{h^2} \Big\langle A^{(m)}(w - \tilde{w}), w - \tilde{w} \Big\rangle \\
&= \frac{1}{h^2} \Big\langle \Lambda^{(m)} (Q^{(m)})^T (w - \tilde{w}), (Q^{(m)})^T (w - \tilde{w}) \Big\rangle \\
&\leq \underbrace{\frac{1}{h^2} \max_{k=1,\ldots,m-1} \lambda_k^{(m)}}_{=:\ell<0} \Big| (Q^{(m)})^T (w - \tilde{w}) \Big|^2 \\
&= \ell \big| w - \tilde{w} \big|^2 .
\end{aligned}$$

■

Remark. The constant ℓ chosen in the proof depends on m, but we can simply choose $\ell = 0$ in order to obtain a one-sided Lipschitz bound with a constant independent of m.

Corollary 7.4.2 (total error in time and space) *Under the assumptions of Theorem 7.2.5 the approximation obtained with the implicit Euler method in time and finite differences in space is bounded by*

$$\max_{n=1,\ldots,N} |\bar{u}(t_n) - w_n|_m \leq K t_{end} (h^2 + \tau)$$

where

$$\bar{u}(t) = \Big(u(t,x_1), \ldots, u(t,x_{m-1}) \Big)^T \in \mathbb{R}^{m-1}. \tag{7.20}$$

The constant K does not depend on h or τ, but on the regularity of the solution.

Proof. For all $n \in \mathbb{N}$ and $1 < m \in \mathbb{N}$ Theorem 7.2.5 and 7.3.2 imply

$$\begin{aligned}
|\bar{u}(t_n) - w_n|_m &\leq |\bar{u}(t_n) - v(t_n)|_m + |v(t_n) - w_n|_m \\
&\leq C_1 t_{\text{end}} h^2 \sup_{t \in [0, t_{\text{end}}]} \big\| \partial_x^4 u(t,\cdot) \big\|_\infty + t_{\text{end}} \frac{C_2}{2} \max_{t \in [t_0, t_{\text{end}}]} |v''(t)| \cdot \tau
\end{aligned}$$

■

If the initial-value problem for $v(t)$ is solved with the trapezoidal rule instead of the implicit Euler, we obtain the error bound

$$\max_{n=1,\dots,N} |\bar{u}(t_n) - w_n|_m \leq K t_{\text{end}}(h^2 + \tau^2)$$

because the order of the trapezoidal rule is 2. (The combination of the trapezoidal rule in time and finite differences in space is called the Crank-Nicolson in the literature.) Nevertheless, the results obtained with the implicit Euler method are sometimes better because this method is L-stable such that errors committed in previous steps are damped; cf. IV.3 in [HW10]

If a Runge-Kutta method of order p is applied, then one would expect that the corresponding total error could be bounded by $Kt_{\text{end}}(h^2+\tau^p)$. Unfortunately, this is in general **not** the case! The reason is that the error analysis for ODEs is based on regularity assumptions of the right-hand side which are not given in case of the heat equation. Roughly speaking, this is due to the fact that for $m \longrightarrow \infty$ we have $|A^{(m)}| \longrightarrow \infty$, and the problem becomes "infinitely stiff". This leads to **order reduction**; cf. IV.15 in [HW10].

Solving the linear systems

A disadvantage of implicit methods is the fact that in each step a system of (typically nonlinear) equations has to be solved. In case of the heat equation, this system is linear, namely

$$M w_{n+1} = w_n + \frac{\tau}{h^2} f^{(m)}(t_{n+1}) =: b \qquad \text{with} \quad M := I - \frac{\tau}{h^2} A^{(m)}.$$

Solving linear systems for an *arbitrary* $M \in \mathbb{R}^{\tilde{m}\times\tilde{m}}$ via the Cholesky decomposition requires $\mathcal{O}(\tilde{m}^3)$ operations. In the special case $M = I - \frac{\tau}{h^2}A^{(m)}$, however, the structure of M (sparse, symmetric) makes things easier. It can be shown (exercise) that there is a decomposition $M = LDL^T$ with

$$D = \text{diag}(d_1, \dots, d_{m-1}), \qquad L = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ l_2 & 1 & 0 & & & \vdots \\ 0 & l_3 & 1 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & l_{m-2} & 1 & 0 \\ 0 & \cdots & \cdots & 0 & l_{m-1} & 1 \end{pmatrix}$$

Hence, the solution of $Mw_{n+1} = b$ can be computed by solving $Ly = b$ and then $DL^T w_{n+1} = y$ by simple backward iteration.

7.5 Application to the Black-Scholes equation

Consider a European capped symmetric power call, i.e. a European call option with payoff

$$V(T,S) = \min\Big(L, ((S-K)^+)^p\Big) \tag{7.21}$$

and maturity $T > 0$, strike $K > 0$ and parameters $p, L > 0$.

- "capped": $V(T,S) \le L$ bounded
- "power": $((S-K)^+)^p$ instead of $(S-K)^+$
- "symmetric": $((S-K)^+)^p$ instead of $(S^p-K)^+$

For $p = 1$ and $L = \infty$: standard European call
The value $V(t,S)$ of the option evolves according to the Black-Scholes equation

$$\partial_t V(t,S) + \frac{\sigma^2}{2} S^2 \partial_S^2 V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0, \qquad S \in [0,\infty), t \in [0,T]$$

with volatility $\sigma > 0$, interest rate $r > 0$ and terminal condition (7.21).

(a) Truncation of the domain

Replace $[0,\infty)$ by $[0,\bar{S}]$ with a sufficiently large $\bar{S} > K + L$ and boundary condition

$$V(t,\bar{S}) = Le^{-r(T-t)}.$$

(b) Time inversion

Let $u(t,S) := V(T-t,S)$. Then, the problem reads

$$\begin{aligned}
\partial_t u &= \frac{\sigma^2}{2} S^2 \partial_S^2 u + rS\partial_S u - ru, && S \in [0,\bar{S}], t \in [0,T] \\
u(0,S) &= \min\Big(L, ((S-K)^+)^p\Big) && S \in [0,\bar{S}] \\
u(t,\bar{S}) &= Le^{-rt} && t \in [0,T]
\end{aligned}$$

No boundary conditions for $u(t,0)$ are required.

(c) Space discretization

Choose $1 < m \in \mathbb{N}$, let $h = \bar{S}/m$ and $S_k = k \cdot h$ for $k = 0, \dots, m$. Approximate

$$\begin{aligned}
\partial_S^2 u(t,S_k) &\approx \frac{u(t,S_{k+1}) - 2u(t,S_k) + u(t,S_{k-1})}{h^2} \\
\partial_S u(t,S_k) &\approx \frac{u(t,S_{k+1}) - u(t,S_{k-1})}{2h}
\end{aligned}$$

for all $k = 1, \dots, m-1$ and $t \in [0, T]$. This yields the ODE

$$v'(t) = M^{(m)} v(t) + f^{(m)}(t)$$
$$M^{(m)} := \frac{\sigma^2}{2} (D^{(m)})^2 \frac{1}{h^2} A^{(m)} + r D^{(m)} \frac{1}{2h} B^{(m)} - rI$$

with $A^{(m)}$ as before and

$$D^{(m)} = \operatorname{diag}(S_1, \dots, S_{m-1}) \in \mathbb{R}^{(m-1)\times(m-1)}$$
$$B^{(m)} = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ -1 & 0 & 1 & \ddots & & \vdots \\ 0 & -1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & 0 \\ \vdots & & \ddots & -1 & 0 & 1 \\ 0 & \cdots & \cdots & 0 & -1 & 0 \end{pmatrix} \in \mathbb{R}^{(m-1)\times(m-1)}$$
$$f^{(m)}(t) = \big(0, \dots, 0, f^{(m)}_{m-1}(t)\big)^T, \qquad f^{(m)}_{m-1}(t) = \left(\frac{\sigma^2}{2h^2} + \frac{r}{2h}\right) \bar{S} u(t, \bar{S})$$

The initial condition is

$$v(0) = \big(u(0, S_1), \dots, u(0, S_{m-1})\big)^T.$$

(d) Time discretization

Solve the ODE with the implicit Euler method

$$(I - \tau M^{(m)}) w_{n+1} = w_n + \tau f^{(m)}(t_{n+1})$$

or the trapezoidal rule.

$$(I - \tau M^{(m)}/2) w_{n+1} = (I + \tau M^{(m)}/2) w_n + \frac{\tau}{2}\Big(f^{(m)}(t_n) + f^{(m)}(t_{n+1})\Big)$$

Final result: $w_{n,k} \approx u(t_n, S_k)$ for all $n = 0, \dots, N$ and $k = 1, \dots, m-1$.

(e) Numerical experiments

See slides.

Chapter 8

Finite-difference methods for American options

8.1 Modelling American options

American options can be exercised **before** maturity. Mathematical model?

Properties of American options: If $V_C^{Am}(t,S)$ is the value of an American call, $V_P^{Eu}(t,S)$ is the value of an European put, etc., then

$$\begin{aligned}
&(K-S)^+ \le V_P^{Am}(t,S) \le K \\
&V_P^{Eu}(t,S) \le V_P^{Am}(t,S) \\
&V_C^{Eu}(t,S) = V_C^{Am}(t,S) \quad \text{if no dividends are paid} \\
&V_C^{Eu}(t,S) \le V_C^{Am}(t,S) \quad \text{if dividends are paid}
\end{aligned}$$

Proof: exercise 1 on sheet 2 or Proposition 2.7 and remark 2.9 in [GJ10].

Picture

In the entire chapter, we consider American puts with no dividends and drop the indices, i.e. $V(t,S) = V_P^{Am}(t,S)$. American calls (with dividends) can be treated in a similar way.

For every $t \in [0,T)$, there is a unique $0 \le S_\star(t) < K$ such that

$$V(t,S) > (K-S)^+ \quad \text{for } S > S_\star(t) \quad \Longrightarrow \text{ no early exercise} \tag{8.1a}$$
$$V(t,S) = (K-S)^+ \quad \text{for } S \le S_\star(t) \quad \Longrightarrow \text{ early exercise} \tag{8.1b}$$

Sketch of the proof. The above bounds imply $V(t,0) = K = (K-0)^+$. Since $V(t,S) > 0$ for $t < T$, it follows that $V(t,S) > (K-S)^+$ for $S \ge K$. A monotonicity argument yields the existence and uniqueness of $0 \le S_\star(t) < K$. ∎

For fixed t, $S_\star(t)$ is called the **contact point**, and the function $t \mapsto S_\star(t)$ is called the **early-exercise curve**, because the option should be exercised before time T if $S \leq S_\star(t)$.

For $S \leq S_\star(t)$, the value of the option is known. For $S > S_\star(t)$ the option is not exercised and can thus be modeled by the Black-Scholes equation. If $S \mapsto V(t,S)$ is C^1 for all $t \in [0,T)$, then it follows that

$$\partial_S V\big(t, S_\star(t)\big) = -1.$$

Hence, we have to solve

$$\begin{aligned}
\partial_t V(t,S) + \mathcal{A}V(t,S) &= 0 && \text{for } S > S_\star(t),\ t \in [0,T) && \text{(PDE)} && \text{(8.2a)}\\
V(T,S) &= (K-S)^+ && \text{for } S \geq 0 && \text{(terminal cond.)} && \text{(8.2b)}\\
V\big(t, S_\star(t)\big) &= \big(K - S_\star(t)\big)^+ && \text{for } t \in [0,T) && \text{(Dirichlet b.c.)} && \text{(8.2c)}\\
\partial_S V\big(t, S_\star(t)\big) &= -1 && \text{for } t \in [0,T) && \text{(Neumann b.c.)} && \text{(8.2d)}
\end{aligned}$$

where $\mathcal{A}$ denotes the operator

$$\mathcal{A}V = \frac{\sigma^2}{2} S^2 \partial_S^2 V + rS\partial_S V - rV.$$

This is a **free boundary value problem**: The boundary $S_\star(t)$ changes in time and depends on the solution. Goal: Reformulate the problem without $S_\star(t)$. We know that for $S \leq S_\star(t) < K$, the value of the option is $V(t,S) = (K-S)^+ = K - S$. This function, however, does **not** solve the Black-Scholes equation (8.2a), because then

$$\begin{aligned}
\partial_t V(t,S) + \mathcal{A}V(t,S) &= \underbrace{\partial_t(K-S)}_{=0} + \mathcal{A}(K-S)\\
&= \frac{\sigma^2}{2} S^2 \underbrace{\partial_S^2(K-S)}_{=0} + rS \underbrace{\partial_S(K-S)}_{=-1} - r(K-S) = -rK < 0
\end{aligned}$$

for all $S \leq S_\star(t)$. This yields the **Black-Scholes inequality**

$$\partial_t V(t,S) + \mathcal{A}V(t,S) \leq 0 \qquad \text{for } S \geq 0,\ t \in [0,T],$$

and we know:

$$\begin{array}{lllllll}
S > S_\star(t) & \Leftrightarrow & V(t,S) > (K-S)^+ & \Leftrightarrow & \partial_t V(t,S) + \mathcal{A}V(t,S) = 0 & \Leftrightarrow & \text{hold}\\
S \leq S_\star(t) & \Leftrightarrow & V(t,S) = (K-S)^+ & \Leftrightarrow & \partial_t V(t,S) + \mathcal{A}V(t,S) < 0 & \Leftrightarrow & \text{exercise}
\end{array}$$

Hence, the free boundary value problem (8.2) is equivalent to the **linear complementary problem**

$$\begin{aligned}
\big(V(t,S) - (K-S)^+\big)\big(\partial_t V(t,S) + \mathcal{A}V(t,S)\big) &= 0 && S \geq 0, t \in [0,T]\\
-\big(\partial_t V(t,S) + \mathcal{A}V(t,S)\big) &\geq 0\\
V(t,S) - (K-S)^+ &\geq 0
\end{aligned}$$

with terminal condition

$$V(T,S) = (K-S)^+$$

and boundary condition

$$V(t,0) = K.$$

8.2 Discretisation

(a) Transformation to the heat equation.

As in (3.3) we use the following transformation:

$$\begin{aligned} x(S) &= \ln(S/K), \qquad \theta(t) = \frac{\sigma^2}{2}(T-t), \qquad c = \frac{2r}{\sigma^2} \\ E(\theta,x) &= \exp\left(\frac{1}{2}(c-1)x + \frac{1}{4}(c+1)^2\theta\right) \\ u(\theta,x) &= \frac{1}{K}E(\theta,x)V(t,S) \\ \psi(\theta,x) &= E(\theta,x)(1-e^x)^+ \end{aligned}$$

The new function $u(\theta,x)$ solves the transformed complementary problem

$$\big(u(\theta,x)-\psi(\theta,x)\big)\big(\partial_\theta u(\theta,x) - \partial_x^2 u(\theta,x)\big) = 0 \qquad x\in\mathbb{R}, \theta\in[0,\sigma^2T/2] \tag{8.3a}$$

$$\partial_\theta u(\theta,x) - \partial_x^2 u(\theta,x) \geq 0 \tag{8.3b}$$

$$u(\theta,x) - \psi(\theta,x) \geq 0 \tag{8.3c}$$

with initial condition

$$u(0,x) = \psi(0,x)$$

and "boundary conditions"

$$\begin{aligned} \lim_{x\longrightarrow-\infty}\big(u(\theta,x)-\psi(\theta,x)\big) &= 0 \\ \lim_{x\longrightarrow\infty} u(\theta,x) &= 0 \end{aligned}$$

Proof: exercise

(b) Truncation and discretisation in time and space.

Truncate the computational domain: Consider $x \in [x_{\min}, x_{\max}]$ instead of $x \in \mathbb{R}$.
Choose $1 < m \in \mathbb{N}$, let $h = (x_{\max} - x_{\min})/m$ and $x_k = x_{\min} + kh$.
Choose $N \in \mathbb{N}$, let $\tau = \frac{\sigma^2 T}{2N}$ and $\theta_n = n\tau$.
Goal: Compute approximation $w_{n,k} \approx u(\theta_n, x_k)$.
Reminder: The Crank-Nicolson-discretization of the heat equation

$$\partial_\theta u(\theta, x) - \partial_x^2 u(\theta, x) = 0$$

is

$$\begin{aligned} 0 = w_{n+1,k} - w_{n,k} &- \frac{\tau}{2h^2}\Big(w_{n+1,k+1} - 2w_{n+1,k} + w_{n+1,k-1}\Big) \\ &- \frac{\tau}{2h^2}\Big(w_{n,k+1} - 2w_{n,k} + w_{n,k-1}\Big) \end{aligned}$$

for $n = 0, \dots, N-1$ and $k = 1, \dots, m-1$. In vector notation:

$$0 = \Big(I - \frac{\tau}{2h^2}A\Big)\, w_{n+1} - \Big(I + \frac{\tau}{2h^2}A\Big)\, w_n - \frac{\tau}{2h^2}\Big(g_{n+1} + g_n\Big)$$

with $A = A^{(m)}$ defined in (7.10) and

$$\begin{aligned} w_n &= (w_{n,1}\ ,\ \dots\ ,\ w_{n,m-1})^T \\ g_n &= \Big(\psi(\theta_n, x_{\min}), 0, \dots, 0\Big)^T \end{aligned}$$

(We omit the subscript (m) in this chapter.)
For vectors $(y_1, \dots, y_d)^T$ and $(z_1, \dots, z_d)^T$ with nonnegative entries, we have

$$y_k \cdot z_k = 0 \text{ for all } k = 1, \dots, d \quad \Longleftrightarrow \quad y^T z = 0$$

This motivates the discretization

$$\Big(w_{n+1} - \bar{\psi}_{n+1}\Big)^T \Big(\Big(I - \frac{\tau}{2h^2}A\Big)\, w_{n+1} - \Big(I + \frac{\tau}{2h^2}A\Big)\, w_n - \frac{\tau}{2h^2}\Big(g_{n+1} + g_n\Big)\Big) = 0 \quad \text{(8.4a)}$$

$$w_{n+1} - \bar{\psi}_{n+1} \geq 0 \quad \text{(8.4b)}$$

$$\Big(I - \frac{\tau}{2h^2}A\Big)\, w_{n+1} - \Big(I + \frac{\tau}{2h^2}A\Big)\, w_n - \frac{\tau}{2h^2}\Big(g_{n+1} + g_n\Big) \geq 0 \quad \text{(8.4c)}$$

with $\bar{\psi}_n = \big(\psi(\theta_n, x_1), \dots, \psi(\theta_n, x_{m-1})\big)$ for the transformed linear complementary problem (8.3). This has to be solved for $n = 0, \dots, N-1$.

8.3 An iterative method for linear complementary problems

Consider the linear complementary problem

$$\begin{aligned} (w - v)^T (Mw - b) &= 0 \\ w - v &\geq 0 \\ Mw - b &\geq 0 \end{aligned}$$

with given $M \in \mathbb{R}^{d\times d}$, given $v, b \in \mathbb{R}^d$ and unknown $w \in \mathbb{R}^d$.
The problem (8.4) is obtained for $w = w_{n+1}$, $M := I - \frac{\tau}{2h^2}A$, $v = \bar{\psi}_{n+1}$ and

$$b = \Big(I + \frac{\tau}{2h^2}A\Big)\, w_n + \frac{\tau}{2h^2}\Big(g_{n+1} + g_n\Big).$$

Numerical method?

(a) Iterative methods for linear systems

First, consider only the linear system $Mw = b$ with $M = (M_{ij})_{ij} \in \mathbb{R}^{d\times d}$. Instead of direct methods (e.g. Gauss elimination) we consider iterative methods based on the decomposition

$$M = D - L - U$$

with

$$D = \begin{pmatrix} M_{11} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & M_{dd} \end{pmatrix}, \qquad L = -\begin{pmatrix} 0 & \cdots & \cdots & 0 \\ M_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ M_{d1} & \cdots & M_{d,d-1} & 0 \end{pmatrix},$$

$$U = -\begin{pmatrix} 0 & M_{12} & \cdots & M_{1d} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & M_{d-1,d} \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

Assume that $D_{kk} = M_{kk} > 0$. This is the case, e.g., if M is symmetric and positive definite. By definition:

$$Mw = b \quad \Longleftrightarrow \quad Dw = (L + U)w + b$$

Idea: Turn this into an fixed-point iteration. This yields the **Jacobi iteration**

$$w^{(j+1)} = D^{-1}\Big((L + U)w^{(j)} + b\Big), \qquad j = 0, 1, 2, \ldots$$

Hope that the sequence $(w^{(j)})_{j\in\mathbb{N}}$ converges to a fixed-point.

Sometimes a better convergence rate is obtained with the **Gauss-Seidel iteration:**

$$w^{(j+1)} = D^{-1}\Big(Lw^{(j+1)} + Uw^{(j)} + b\Big), \qquad j = 0, 1, 2, \ldots$$

This method seems to be implicit, since $w^{(j+1)}$ appears on the right-hand side. A closer look reveals, however that the Gauss-Seidel iteration is explicit because the entries of

$w^{(j+1)}$ can be computed one after the other:

$$w_1^{(j+1)} = \frac{1}{D_{11}}\Big(0 + \sum_{k=2}^{d} U_{1k} w_k^{(j)} + b_k\Big)$$

$$w_2^{(j+1)} = \frac{1}{D_{22}}\Big(L_{21} w_1^{(j+1)} + \sum_{k=3}^{d} U_{2k} w_k^{(j)} + b_2\Big)$$

$$w_i^{(j+1)} = \frac{1}{D_{ii}}\Big(\sum_{k=1}^{i-1} L_{ik} w_k^{(j+1)} + \sum_{k=i+1}^{d} U_{ik} w_k^{(j)} + b_i\Big) \qquad i = 1, \ldots, d.$$

Generalization: **Successive overrelaxation method (SOR method)**:

```
For j = 1, 2, ...
    For k = 1, ..., d
        w̃_k^(j) = 1/D_kk [L w^(j+1) + U w^(j) + b]_k        k-th entry
        w_k^(j+1) = w_k^(j) + r(w̃_k^(j) − w_k^(j))
    end
end
```

with relaxation parameter $r \in \mathbb{R}$. If $r = 1$, then $w^{(j+1)} = \tilde{w}^{(j)}$, and we obtain the Gauss-Seidel method. For $r \in (0, 1)$, $w^{(j+1)}$ is an interpolation between $w^{(j)}$ and $\tilde{w}^{(j)}$.
In practice, however, typically $r > 1$ is chosen. It can be shown that the SOR method converges for $r \in (1, 2)$ to the solution of $Mw = b$. If r is chosen in an optimal way, then the SOR method needs less iterations than the Gauss-Seidel method to achieve a given accuracy.

(b) The projected SOR method for linear complementary problems

Back to the linear complementary problem:

$$(w - v)^T (Mw - b) = 0 \tag{8.5a}$$

$$w - v \geq 0 \tag{8.5b}$$

$$Mw - b \geq 0 \tag{8.5c}$$

In general, we cannot expect that $Mw = b$. There are some entries k where $[Mw - b]_k > 0$. The complementary problem is equivalent to

$$\min\Big\{[Mw - b]_k, [w - v]_k\Big\} = 0 \qquad \text{for all } k = 1, \ldots, d$$

$$\iff \quad \min\Big\{\Big[\underbrace{D^{-1}\big((D - L - U)w - b\big)}_{w - D^{-1}\left((L+U)w+b\right)}\Big]_k, [w - v]_k\Big\} = 0 \qquad \text{for all } k = 1, \ldots, d$$

$$\iff \quad \max\Big\{\Big[D^{-1}\big((L + U)w + b\big)\Big]_k, v_k\Big\} = w_k \qquad \text{for all } k = 1, \ldots, d$$

This motivates the **projected SOR method**:

```
For j = 0, 1, 2, ...
    For k = 1, ..., d
        w̃_k^(j) = 1/D_kk [L w^(j+1) + U w^(j) + b]_k
        w_k^(j+1) = max{ w_k^(j) + r(w̃_k^(j) − w_k^(j)) , v_k }
    end
end
```

Theorem 8.3.1 (Cryer) *Let $v, b \in \mathbb{R}^d$, $r \in (1,2)$ and assume that $M \in \mathbb{R}^{d\times d}$ is symmetric and positive definite. Then, the linear complementary problem* (8.5) *has a unique solution, and the iterates $w^{(j)}$ of the projected SOR method converge to the solution.*

Proof: Chapter 7.3 in [GJ10]

8.4 Summary: Pricing American options with the projected SOR method

- Start: Free boundary problem (8.2) with solution $V(t,S)$.
- Reformulation as a linear complementary problem.
- Transformation:

  $V(t,S) \longrightarrow u(\theta,x)$, Black-Scholes inequality $\longrightarrow$ heat inequality

  $\Longrightarrow$ transformed linear complementary problem (8.3).
- Truncation: Restrict $x \in \mathbb{R}$ to $x \in [x_{\min}, x_{\max}]$, choose boundary conditions.
- Discretize time and space: $w_{n,k} \approx u(\theta_n, x_k)$.
- Algorithm:

```
For n = 0, 1, ..., N − 1                                  (time points)
    Solve linear complementary problem (8.4) with the
    projected SOR method
    For j = 0, 1, 2, ...                                  (iteration number)
        For k = 1, ..., m − 1                               (entry number)
            ...
        end
    end
end
```

- Transform back.

8.5 Penalty methods for American options

Consider again the linear compementary problem

$$\begin{aligned}(w-v)^T(Mw-b) &= 0\\ w-v &\geq 0\\ Mw-b &\geq 0\end{aligned}$$

but replace the projected SOR method by a different approach. Replace the inequality $Mw-b$ by an equality with a penalty term. Define

$$q_\delta(y) = \frac{(v-y)^+}{\delta}, \qquad \text{i.e.} \qquad [q_\delta(y)]_k = \max\left\{\frac{1}{\delta}[v-y]_k,\ 0\right\}.$$

Hence, $q_\delta(y)=0$ implies $y \geq v$. Now, find $w^\delta \in \mathbb{R}^d$ such that

$$Mw^\delta - b - q_\delta(w^\delta) = 0 \tag{8.6}$$

or equivalently

$$\delta(Mw^\delta - b) = (v-w^\delta)^+$$

For every $\delta > 0$, we have $Mw^\delta - b \geq 0$, and $\delta \longrightarrow 0$ implies $(v-w^\delta)^+ \longrightarrow 0$ and hence $w^\delta \geq v$. It can be shown that w^δ converges indeed to the solution w of the linear complementary problem, cf. 7.4 in [GJ10]

— FIN —

Bibliography

[Ben04] Fred Espen Benth. *Option theory with stochastic analysis : an introduction to mathematical finance*. Springer, 2004.

[BK04] N. H. Bingham and Rüdiger Kiesel. *Risk-neutral valuation. Pricing and hedging of financial derivatives*. Springer Finance. Springer, London, 4nd ed. edition, 2004.

[FK08] Helmut Fischer and Helmut Kaul. *Mathematik für Physiker. Band 2: Gewöhnliche und partielle Differentialgleichungen, mathematische Grundlagen der Quantenmechanik*. Teubner Studienbücher Mathematik. Teubner, Stuttgart, 3rd revised ed. edition, 2008.

[GJ10] Michael Günther and Ansgar Jüngel. *Finanzderivate mit MATLAB. Mathematische Modellierung und numerische Simulation.* Vieweg, Wiesbaden, 2nd ed. edition, 2010.

[HB09] Martin Hanke-Bourgeois. *Grundlagen der numerischen Mathematik und des wissenschaftlichen Rechnens*. Vieweg+Teubner, Wiesbaden, 3rd revised ed. edition, 2009.

[Hig02] Desmond J. Higham. Nine ways to implement the binomial method for option valuation in MATLAB. *SIAM Rev.*, 44(4):661–677, 2002.

[HNrW10] Ernst Hairer, Syvert P. Nø rsett, and Gerhard Wanner. *Solving ordinary differential equations. I: Nonstiff problems*. Number 8 in Springer Series in Computational Mathematics. Springer, Berlin, 2nd revised ed., 3rd corrected printing. edition, 2010.

[HW10] Ernst Hairer and Gerhard Wanner. *Solving ordinary differential equations. II: Stiff and differential-algebraic problems*. Number 14 in Springer Series in Computational Mathematics. Springer, Berlin, reprint of the 1996 2nd revised ed. edition, 2010.

[KP99] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations.* Number 23 in Applications of Mathematics. Springer, Berlin, corr. 3rd printing edition, 1999.

[Øks03] Bernt Øksendal. *Stochastic differential equations. An introduction with applications*. Universitext. Springer, Berlin, 6th ed. edition, 2003.

[Sey09] Rüdiger U. Seydel. *Tools for computational finance. 4th revised and extended ed.* Universitext. Springer, Berlin, 4th revised and extended ed. edition, 2009.

[Shr04] Steven E. Shreve. *Stochastic calculus for finance. II: Continuous-time models.* Springer Finance. Springer, 2004.

[Ste01] J. Michael Steele. *Stochastic calculus and financial applications.* Number 45 in Applications of Mathematics. Springer, New York, NY, 2001.