Intermediate Mathematics for Business and Economics

Homework Assignment 1

Submission deadline: 12:00 on 29.04.2025

Exercise 1 *Simplification of terms and domain of definition* *(10 points)*

Simplify the following expressions and determine the maximum possible domain (as a subset of $\mathbb{R}$) of the occurring parameters, for which the initial expression is defined.

(a) $\dfrac{(4a^2+12a+9)(\sqrt{b})^3}{(2a+3)b}$ (b) $\dfrac{\sqrt{63}\cdot a}{\sqrt{147a}}$ (c) $\ln\left(\dfrac{e^{2a}}{a}\right)$

(d) $\dfrac{\left(\frac{2}{3}\right)^3}{\left(\frac{10}{9}\right)^2}$ (e) $\left(\dfrac{(1+\sin(x))^2+\cos^2(x)-1}{2\sin(x)+1}\right)^2$

Hint: Refer to the literature or `https://en.wikipedia.org` to recall the definition of a function and the domain of a function. The identities $\sin(\pi - x) = \sin(x)$ and $\sin(2\pi + x) = \sin(x)$ might be useful.

Exercise 2 *Matrix multiplication* *(10 points)*

Let $A = \begin{pmatrix} 0 & 0 \\ 2 & 1 \\ -2 & 1 \end{pmatrix}$, $B = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$, $C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and $D = \begin{pmatrix} 0 & 1 & 0 \\ 2 & 3 & 4 \end{pmatrix}$.

Compute:

(a) AB (b) DB (c) AD (d) $DA + C$ (e) $C^3 = C \cdot C \cdot C$

Argue that (f) $CDAB \neq ADA$ and (g) $C^7 \neq ADA$ without explicitly computing the corresponding matrices.

Exercise 3 *Matrices and commutativity* *(10 points)*

(a) Construct $A, B \in \mathbb{R}^{3\times 3}$ such that $AB \neq BA$.

(b) Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix}$. Under which conditions on the parameters c, $d \in \mathbb{R}$ do we have $AB = BA$?

Intermediate Mathematics for Business and Economics

Homework Assignment 2

Submission deadline: 12:00 on 06.05.2025

Exercise 1 *Vectors* *(10 points)*

Let $\vec{w_1} = (1\ 2\ 0)^T$ and $\vec{w_2} = (1\ 2\ 4)^T$. Find a vector $\vec{v_1} \in \mathbb{R}^3 \setminus \{\vec{0}\}$ and a vector $\vec{v_2} \in \mathbb{R}^3 \setminus \{\vec{0}\}$ such that

(a) $\vec{v_1}^T \vec{w_1} = 0$, and (b) $\vec{v_2}^T \vec{w_2} = 0$.

(Note: $\vec{w_1}$, $\vec{w_2}$, and $\vec{0}$ are elements of $\mathbb{R}^3$, $\vec{0}$ denotes the zero vector.)

Exercise 2 *(Semi-)Groups* *(10 points)*

A set S equipped with a binary operation $S \times S \to S$, denoted by $\bullet$, is a *monoid*[1] if it satisfies the following two axioms:

- *(Associativity)* For all a, b and c in S, the equation $(a \bullet b) \bullet c = a \bullet (b \bullet c)$ holds.
- *(Identity element)* There exists an element e in S such that for every element a in S, the equations $e \bullet a = a$ and $a \bullet e = a$ hold.

The set $\mathbb{R}^{2\times 2}$ together with matrix multiplication is a monoid G.

(a) Which matrix is the *identity element* of G? Is it unique?

(b) Determine the *center*[2]

$$Z(G) := \left\{ z \in \mathbb{R}^{2\times 2} \mid zg = gz \quad \forall g \in \mathbb{R}^{2\times 2} \right\}$$

of G.

Exercise 3 *More about matrices* *(10 points)*

(a) Let $A \in \mathbb{R}^{m\times n}$. Prove that $A^T A$ and AA^T are symmetric.

(b) Find matrices A, B, C with $\operatorname{tr}(ABC) \neq \operatorname{tr}(ACB)$ such that both ABC and ACB exist.

---

[1] See `https://en.wikipedia.org/wiki/Monoid` if you *want* to know more.

[2] See `https://en.wikipedia.org/wiki/Center_(group_theory)` if you *want* to know more.

Intermediate Mathematics for Business and Economics

Homework Assignment 3

Submission deadline: 12:00 on 13.05.2025

Exercise 1 *Investment model* *(10 points)*

A state model is given by

$$\vec{y} = A\vec{x}\,, \qquad \text{with } A = (a_{ij}) = \begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 3 & 0 \end{pmatrix}.$$

An agent can decide to invest in two possible assets. Due to uncertainty one of three scenarios can arise, so different states yield different returns of the asset. The value a_{ij} describes the payoff of the asset j in the case that state i occurs. The vector $\vec{x} = \binom{x_1}{x_2}$ is called portfolio and allocates the investment to the assets, hence we have $x_1 + x_2 = 1$.

(a) A portfolio is called *riskless* if it provides the same return in every state. Is there a riskless portfolio?

(b) A portfolio is called *duplicable* if there is a different portfolio with exactly the same return vector. Is there a duplicable portfolio?

Exercise 2 *True or False* *(10 points)*

Let $A \in \mathbb{R}^{n\times n}$ and $\vec{b} \in \mathbb{R}^n$, $n \in \mathbb{N}$. Are the following statements true or false?

- If $A \mid \vec{b}$ is in reduced row echelon form, then A is an identity matrix. ○ True ○ False
- If A is a diagonal matrix, then $A \mid \vec{b}$ is in reduced row echelon form. ○ True ○ False
- Any system of linear equations $A\vec{x} = \vec{0}$ has at least one solution. ○ True ○ False
- If the columns of A are linearly dependent, then the system of linear equations $A\vec{x} = \vec{b}$ has at least one solution. ○ True ○ False
- If the rows of A are linearly independent, then the system of linear equations $A\vec{x} = \vec{b}$ has at least one solution. ○ True ○ False

Exercise 3 *Linear Independence* *(10 points)*

(a) Are the vectors $\vec{a} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}$, and $\vec{c} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ linearly independent?

(b) Are the vectors $\vec{a} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$, and $\vec{c} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ linearly independent?

(c) Suppose $\vec{a}$, $\vec{b}$ and $\vec{c}$ are three linearly independent vectors in $\mathbb{R}^4$. Are the vectors $2\vec{a} - \vec{b}$, $6\vec{b} - 3\vec{c}$ and $\vec{c} - 4\vec{a}$ linearly independent? Is the assumption about linear independence necessary to answer the previous question?

Intermediate Mathematics for Business and Economics

Homework Assignment 4

Submission deadline: 12:00 on 20.05.2025

Exercise 1 *Rank* *(10 points)*

(a) Let

$$B_1 := \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix},\ B_2 := \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix},\ \text{and } B_3 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

be three square matrices from $\mathbb{R}^{2\times 2}$, and let $A \in \mathbb{R}^{2\times 3}$ be a matrix with rank 2.

Show that the rank of B_1A, B_2A, and B_3A is still 2.

(b) Determine the rank of the matrix $A = \begin{pmatrix} 1 & 2 & 1 & 1 & 2 \\ 0 & 0 & 3 & -4 & -1 \\ 2 & 4 & 3 & 1 & 2 \\ 4 & 8 & 2 & 7 & 7 \\ 0 & 0 & -1 & 1 & 2 \end{pmatrix}$.

Exercise 2 *Basis* *(10 points)*

(a) Calculate a basis for the vector space $\operatorname{span}\left(\left\{\begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 2 \\ 3 \end{pmatrix}\right\}\right)$.

(b) Find a basis for the real vector space $\mathbb{R}^{2\times 3}$.

(c) Find a basis for the real vector space of polynomials of degree at most 5 with coefficients in $\mathbb{R}$.

(d) Let p_1, $p_2 : \mathbb{R} \to \mathbb{R}$ be functions of the $\mathbb{R}$-vector space of polynomials of degree at most 5 with $p_1(x) = 2x^2 + 3x^4$, $p_2(x) = 5x^2 + x^4$. Describe the set $\langle p_1, p_2\rangle := \operatorname{span}(\{p_1, p_2\})$.

Exercise 3 *Inverse matrices* *(10 points)*

Determine, if possible, the inverse of the following matrices:

$$\text{(a)} \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} 1 & 2 & 4 & 5 \\ 0 & 4 & 4 & 2 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 2 & 4 \end{pmatrix}.$$

Exercise 4 *Applications of determinants* *(0 points)*

Read the paragraph about applications from Wikipedia's website:

`https://en.wikipedia.org/wiki/Determinant#Applications`

Intermediate Mathematics for Business and Economics

Homework Assignment 5

Submission deadline: 12:00 on 27.05.2025

Exercise 1 *Scalar product and norms* *(10 points)*

(a) Let $\vec{x} = (1\ ,\ 2\ ,\ 3)^T$ and $\vec{y} = (4\ ,\ 5\ ,\ 6)^T$. Compute $\langle \vec{x}, \vec{y} \rangle$, $\langle \vec{y}, \vec{x} \rangle$, $\|\vec{x}\|_2$, and $\|\vec{y}\|_2$.

(b) Verify $\langle Q\vec{x}, Q\vec{x} \rangle = \|\vec{x}\|_2^2$ for an arbitrary orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ and $\vec{x} \in \mathbb{R}^n$.

(c) Compute the Frobenius norm of $A = \begin{pmatrix} 1 & -2 & 1 \\ 5 & 2 & 5 \\ 2 & 4 & -2 \end{pmatrix}$.

Exercise 2 *Eigenvalues and eigenvectors* *(10 points)*

Compute the eigenvalues of the following matrices. For each eigenvalue, find a basis of the vector space spanned by its associated eigenvectors.

$$\text{(a) } A = \begin{pmatrix} 2 & 0.25 \\ 1 & -3 \end{pmatrix}, \quad \text{(b) } B = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Exercise 3 *Elementary matrices* *(10 points)*

The elementary matrix corresponding to multiplying the i-th row by $s \in \mathbb{R} \setminus \{0\}$ is represented by

$$E_i(s) = (e_{kl}) \text{ where } \begin{cases} e_{ii} = s & \\ e_{kk} = 1 & k \neq i \\ e_{kl} = 0 & \text{otherwise.} \end{cases}$$

(a) Show that the inverse of $E_i(s)$ is $E_i(s^{-1})$ using the basic definition of matrix multiplication.

(b) Compute the determinant of $E_i(s)$ using the rules for $\det^{(n)}$ (cf. lecture).

(c) Let $A \in \mathbb{R}^{m\times n}$ be an arbitrary matrix. Show that multiplying the i-th row of A by s and is equivalent to computing the matrix product $E_i(s) \cdot A$.

(d) Show that

$$\det(E_i(s) \cdot B) = \det(E_i(s)) \cdot \det(B)$$

for an arbitrary square matrix B.

Intermediate Mathematics for Business and Economics

Homework Assignment 6

Submission deadline: 12:00 on 03.06.2025

Exercise 1 *Determinants* *(10 points)*

Compute the determinant of

$$\text{(a)}\quad A=\begin{pmatrix}1&0&2&0\\3&1&0&0\\5&0&0&3\\7&0&4&0\end{pmatrix}\quad\text{and}\quad\text{(b)}\quad B=\begin{pmatrix}2&21&5&1\\3&21&5&1\\2&21&7&1\\2&21&5&11\end{pmatrix}.$$

Exercise 2 *Geometric meaning of eigenvectors* *(10 points)*

The matrix

$$A=\begin{pmatrix}0&-1&0\\1&0&0\\0&0&1\end{pmatrix}$$

defines a rotation around the z-axis by 90 degrees in three-dimensional space.

(a) Give an example that illustrates this fact, i.e., find a vector that is easy to rotate geometrically, and then check to see if multiplying the matrix leads to the same result.

(b) Find an eigenvector of A and the corresponding eigenvalue.

Hint: Think about the properties of an eigenvector; in this particular case it is possible to find one without computing anything. (Though you have to explain your approach.)

Exercise 3 *Eigenvectors* *(10 points)*

Let $A=\begin{pmatrix}a&1&p\\b&2&q\\c&-1&r\end{pmatrix}$. Assume that A has eigenvectors

$$\begin{pmatrix}1\\1\\0\end{pmatrix},\ \begin{pmatrix}-1\\0\\2\end{pmatrix},\ \begin{pmatrix}0\\1\\-1\end{pmatrix}.$$

Compute the eigenvalues of A.

Intermediate Mathematics for Business and Economics

Homework Assignment 7

Submission deadline: 12:00 on 24.06.2025

Exercise 1 *Diagonalization of matrices* *(10 points)*

Consider the following matrices:

$$A_1 = \begin{pmatrix} -1 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 4 \end{pmatrix}.$$

If possible, compute a diagonalization for each matrix, i.e., find matrices P_k and D_k, such that $P_k^{-1} A_k P_k = D_k$ for $k = 1, 2, 3$.

Exercise 2 *Comprehension questions* *(5 points)*

Let $A \in \mathbb{R}^{n \times n}$ be an arbitrary square matrix. Answer the following items. (Always give reasons for your answer.)

(a) True or false: The eigenvalues of A never coincide with the eigenvalues of the matrix $2A$?

(b) Is any multiple of an eigenvector of A also an eigenvector?

(c) Is any sum of eigenvectors of A also an eigenvector?

(d) Suppose A is singular. What does this imply for its eigenvalues?

(e) Suppose $A \in \mathbb{R}^{n \times n}$ has the eigenvalue λ. What does this imply for the eigenvalues of A^2?

Exercise 3 *Eigenvalues of invertible matrices* *(5 points)*

Prove the following statement:

> A square matrix A is invertible iff it does not have an eigenvalue equal to zero.

Exercise 4 *Definiteness* *(10 points)*

Are the following matrices positive definite, positive semidefinite, negative definite, negative semidefinite, or indefinite?

(a) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{pmatrix}$, (b) $\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, (c) $\begin{pmatrix} 0 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 0 \end{pmatrix}$, (d) $\begin{pmatrix} -1 \end{pmatrix}$, and (e) $\begin{pmatrix} -2 & 5 \\ 4 & -10 \end{pmatrix}$

Mathematics in Economy
apl. Prof. Dr. Sascha Kurz Summer Semester 2025

Intermediate Mathematics for Business and Economics

Homework Assignment 8

Submission deadline: 12:00 on 01.07.2025

Exercise 1 *A model of employment* *(10 points)*

Consider the following model of employment: If an individual is not employed in a given week, in the next week he or she may either find a job or remain unemployed. With probability $0 < q < 1$ the individual will remain unemployed, and therefore with probability $\overline{q} := 1 - q$ he or she will find a job. Similarly, if an individual is employed in a given week, let $0 < p < 1$ be the probability that he or she will remain employed and $\overline{p} := 1 - p$ the probability of becoming unemployed.

Given the numbers of employed individuals x_t and unemployed individuals y_t in period t the corresponding numbers for the next period are given by:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = A \begin{pmatrix} x_t \\ y_t \end{pmatrix} \qquad (*)$$

(a) Determine A such that $(*)$ represents the model of employment described above.

(b) Compute a diagonalization of A.

(c) Assume that $p = 2/3$ and $q = 1/3$. Compute the long term distribution of employed and unemployed individuals, i.e., compute $\lim_{n\to\infty} A^n$.

Hint: Use the result from the supplement to powers of diagonal matrices and the fact there are matrices P and D, such that $AP = PD$.

Exercise 2 *Stochastic matrices* *(10 points)*

Let $A \in \mathbb{R}^{n\times n}$ be a row stochastic matrix.

(a) Show that 1 is an eigenvalue of A.

(b) Compute an eigenvector corresponding to the eigenvalue 1.

Exercise 3 *Convergence of sequences* *(10 points)*

Do the following sequences converge for $n \to \infty$, $n \in \mathbb{N}$? Find the limit of each convergent sequence.

$$\text{(a) } a_n = \sqrt{\frac{3n^3 + 2n^2 + n}{4n^2 + 2}}, \quad \text{(b) } b_n = \frac{8n+4}{\pi - 9n} + \left(\frac{4}{5}\right)^n, \quad \text{(c) } c_n = \frac{(-1)^n}{-1}\frac{1}{n}$$

Intermediate Mathematics for Business and Economics

Homework Assignment 9

Submission deadline: 12:00 on 08.07.2025

Exercise 1 *Sequences* *(8 points)*

Let the sequence $(X_n)_{n\in\mathbb{N}}$ be recursively defined by

$$X_0 = 1,$$
$$X_{n+1} = \frac{X_n + 2/X_n}{2},$$

i.e., the next term of the sequence is given as a function of its predecessor; the values of the sequence are rational numbers.

(a) Compute X_n such that X_{n-1} and X_n do not differ in the first 9 decimal places. (Use a calculator or similar tool.)

(b) Compute the limit X of the sequence by setting $X_{n+1} = X_n$. (To be able to do this we need to assume that the limit exists.)

(c) Show that the limit found in part (b) is not a rational number (even though every element of the sequence is). Start by assuming that the limit X is a rational number, i.e., $X = p/q$ where $p, q \in \mathbb{N}$ are coprime.

Exercise 2 *Cauchy sequences* *(6 points)*

The sequence (X_n) is defined by $X_n := \sum_{k=0}^{n} 1/(k!)$, where $0! = 1$ and $k! = \prod_{i=1}^{k} i$ for $k \geq 1$. Show that $(X_n)_{n\in\mathbb{N}}$ is a Cauchy sequence for the distance function

$$d\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}_{\geq 0},\ (x, y) \mapsto |x - y|.$$

Hint: You can use either $\sum_{k=0}^{n} \left(\frac{1}{2}\right)^k \leq 2$ or $\sum_{k=1}^{n} \frac{1}{k(k+1)} = 1 - \frac{1}{n+1}$.

Exercise 3 *Continuity* *(4 points)*

Use definition 39 from the lecture to find out whether the following functions are continuous:

(a) $f\colon \mathbb{R} \to \mathbb{R},\ f(x) := \begin{cases} \left(\frac{x}{100}\right)^2 \cdot x & \text{for } x \leq 10 \\ 0.01 \cdot x & \text{otherwise} \end{cases}$

(b) $g\colon \mathbb{R}_{\geq 0} \to \mathbb{R},\ g(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq 1, \\ \ln(\sqrt{x}) & \text{for } 1 < x < e^2, \\ \sin(x - e^2) & \text{for } x \geq e^2 \end{cases}$

Exercise 4 *Accumulation points* *(4 points)*

Find the accumulation points of the sequence $a_n := (1 + \frac{1}{n})(-1)^n$.

Exercise 5 *Sets* *(8 points)*

(a) Let $I_n := (1 - 1/n, 1 + 1/n)$. Is the interval I_n open for each $n \in \mathbb{N}_{>0}$? Determine the set $C := \bigcap_{n \in \mathbb{N}_{>0}} I_n$. Is it open?

(b) Determine the set of boundary points of $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 4\}$.

(c) Draw the closed unit-balls around $(0, 0)$ for $\|\cdot\|_2$, $\|\cdot\|_\infty$, and $\|\cdot\|_1$.

Hint: Unit-balls have a radius of 1, that is, it is equal to the set $\{\vec{x} \in \mathbb{R}^2 \mid \|\vec{x}\| \leq 1\}$.

(d) Give an example of a set in $\mathbb{R}^2$ which is neither open nor closed.

Intermediate Mathematics for Business and Economics

Homework Assignment 10

Submission deadline: 12:00 on 15.07.2025

Exercise 1 *Infimum and supremum* *(4 points)*

Determine the infimum and the supremum of the set $\{\, x \in \mathbb{R} \mid 1 < x^2 < 2 \,\}$.

Exercise 2 *Convergent and Cauchy sequences* *(10 points)*

(a) Show that every convergent sequence in a metric space (M, d) is a Cauchy sequence.

(b) Show that $(x_n) = \sum_{k=1}^{n} \frac{1}{k^2}$ is a convergent sequence.

(*Hint:* Don't bother figuring out what the limit point is.)

Exercise 3 *Surjectivity, injectivity, bijectivity* *(10 points)*

(a) Find the composition $f \circ g$ and simplify.

- $f\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto x^2 \cdot \ln(x)$ and $g\colon \mathbb{R} \to \mathbb{R}_{>0},\ x \mapsto e^{-x}$,
- $f\colon \mathbb{R}^4 \to \mathbb{R}^3,\ \vec{x} \mapsto \begin{pmatrix} 1 & 2 & 4 & 3 \\ 3 & 4 & 2 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix} \cdot \vec{x}$ and $g\colon \mathbb{R}^2 \to \mathbb{R}^4,\ \vec{x} \mapsto \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \vec{x}$.

(b) Are the following functions surjective? Are they injective? Explain.

- $f_1\colon [\,0, 1\,] \to [\,0, 1\,],\ x \mapsto x^2$
- $f_2\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto x^2$
- $f_3\colon \mathbb{R} \to \mathbb{R}_{\geq 0},\ x \mapsto x^2$
- $f_4\colon \mathbb{N} \to \mathbb{N},\ n \mapsto 2n + 1$

(c) Determine a non-empty set $U \subseteq \mathbb{R}$ such that the restriction $f_3|_U$ is bijective.

Exercise 4 *Derivatives* *(6 points)*

Compute the derivative for each of the following functions for all elements of the domain where the derivative exists.

(a) $f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto |x^3 - 1|$

(b) $g\colon \mathbb{R} \to \mathbb{R},\ x \mapsto |x| \cdot x$

(c) $h\colon \mathbb{R} \to \mathbb{R},\ x \mapsto e^{\sin(x^2)}$

Intermediate Mathematics for Business and Economics

Homework Assignment 11

Submission deadline: 12:00 on 22.07.2025

Exercise 1 *Total differentiability and directional derivatives* *(7 points)*

Consider the function

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ (x,y)^T \mapsto f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{if } (x,y) \neq (0,0), \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the Jacobian of f.

(b) Show that f is totally differentiable at $(0,0)$.

(c) Compute the derivatives of f along the directions $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}1\\-1\end{pmatrix}$.

Exercise 2 *Inverse mapping* *(5 points)*

Find the inverse g^{-1} of the mapping $g\colon \mathbb{R}^2 \to \mathbb{R}^2$, $\begin{pmatrix}x\\y\end{pmatrix} \mapsto \begin{pmatrix}2x-y\\y-4x\end{pmatrix}$.

Exercise 3 *Taylor polynomials* *(6 points)*

(a) Compute the 6th order Taylor polynomial at $x_0 = 0$ for the functions $\exp(x)$, $\sin(x)$, and $\cos(x)$.

(b) Compute the 2nd order Taylor polynomial at $\vec{x}_0 = (-1, 0.5)$ for $f\colon \mathbb{R}^2 \to \mathbb{R},\ f(x,y) = 0.5y\cdot(x - x^2 + 2y) - 0.75x$.

Exercise 4 *Chain rule in higher dimensions* *(5 points)*

Determine the Jacobi matrix of the function

$$g(x,y) = \begin{pmatrix} \ln(x+y) + \cos\left(x^2+y^2\right) \\ \ln\left(x^2+y^2\right) + \cos(x+y) \end{pmatrix}$$

$(x, y \in \mathbb{R}_{>0})$ using the multidimensional chain rule and extract the partial derivatives.

Exercise 5 *Total differentiability and continuity* *(7 points)*

Consider the function

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ \vec{x} = (x,y)^T \mapsto f(x,y) = \begin{cases} xy^2/(x^2+y^4) & \text{if } x \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Check whether f is continuous.

(b) Check whether f is totally differentiable at $(0,0)$.

Intermediate Mathematics for Business and Economics

Homework Assignment 12

Exercise 1 *Global extrema I*

Determine the global extrema of the function $f\colon [-1,1]^2 \to \mathbb{R},\ f(x,y) = x^2 - y^2 + 2xy + 2x + 4y$.

Exercise 2 *Global extrema II*

Determine the gradient of the function $f\colon \mathbb{R}^2 \to \mathbb{R},\ f(x,y) = x^4 - 4x^2y^2 + y^4$. Identify stationary points and check for global extrema of f.

Exercise 3 *Global extrema III*

Let $U \coloneqq [-1, 0.5] \cup [1, 3]$ be a subset of $\mathbb{R}$. Determine the global extrema of the function $f\colon U \to \mathbb{R},\ x \mapsto x^2 - 10 \cdot \max\{x, 2\}$.

Exercise 4 *Convex and concave functions*

Check whether the following functions are convex or concave:

(a) $f_1\colon \mathbb{R} \to \mathbb{R},\ x \mapsto (x-2)^2$;

(b) $f_2\colon [0,2] \to \mathbb{R},\ x \mapsto 2$;

(c) $f_3\colon \mathbb{R}^3 \to \mathbb{R},\ x \mapsto |x|_2$;

(d) $f_4\colon \mathbb{R}^2 \to \mathbb{R},\ \vec{x} \mapsto \vec{x}^T \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \vec{x}$.

Intermediate Mathematics for Business and Economics

Sascha Kurz

May 28, 2025

Contents

Notation

$A \in \mathbb{R}^{m\times n}$	matrix, plural: matrices
$(a_{ij})_{\substack{i=1\dots m\\ j=1\dots n}}$	alternative way to denote a matrix
$\vec{v} \in \mathbb{R}^n$	column vector
$\wedge$	logical "and"
$\vee$	logical "or"
$\iff$	"is equivalent to"
$\implies$, $\not\Longrightarrow$	"implies", "does not imply"
iff	"if and only if", see $\iff$,
$:=$	"is defined as"
$\pm, \mp$	"plus/minus", "minus/plus"
$I_{n\times n} \in \mathbb{R}^{n\times n}$	identity matrix
$0_{m\times n} \in \mathbb{R}^{m\times n}$	null matrix
$\mapsto$	"maps to"
#	"number of" or "cardinality of"
$\exists$	"exists"
$\exists!$	"exists exactly one"
$\lvert S \rvert$	cardinality of the set S
$\lvert a \rvert$	absolute value of a real number $a \in \mathbb{R}$
$\lVert \vec{v} \rVert$	(Euclidean) norm of a vector $\vec{v} \in \mathbb{R}^n$
$\operatorname{span}(S)$	(linear) span of a set of vectors S
$\langle S \rangle$	alternative notation for the (linear) span of a set of vectors S

Chapter 1

Linear Algebra

Application 1. Assemblies A_1, A_2 are manufactured from components C_1, C_2, C_3, C_4. Finally, A_1, A_2 are put together into products P_1, P_2, P_3.

Compostition tables (*matrix, matrices*)

$$\underbrace{\begin{array}{c|cc} & A_1 & A_2 \\ \hline C_1 & 2 & 1 \\ C_2 & 0 & 3 \\ C_3 & 1 & 1 \\ C_4 & 1 & 2 \end{array}}_{M_{C\to A}} \qquad \underbrace{\begin{array}{c|ccc} & P_1 & P_2 & P_3 \\ \hline A_1 & 4 & 3 & 1 \\ A_2 & 2 & 2 & 2 \end{array}}_{M_{A\to P}}$$

Demand table (*vector*)

$$\underbrace{\begin{array}{c|c} P_1 & 10 \\ P_2 & 20 \\ P_3 & 10 \end{array}}_{D}$$

Production table assemblies

$$M_{A\to P} \cdot D = \begin{array}{c|c} A_1 & 4\cdot 10 + 3\cdot 20 + 1\cdot 10 \\ A_2 & 2\cdot 10 + 2\cdot 20 + 2\cdot 10 \end{array} = \underbrace{\begin{array}{c|c} A_1 & 110 \\ A_2 & 80 \end{array}}_{P_A}$$

Order table components

$$M_{C\to A} \cdot P_A = \begin{array}{c|c} C_1 & 2\cdot 110 + 1\cdot 80 \\ C_2 & 0\cdot 110 + 3\cdot 80 \\ C_3 & 1\cdot 110 + 1\cdot 80 \\ C_4 & 1\cdot 110 + 2\cdot 80 \end{array} = \underbrace{\begin{array}{c|c} C_1 & 300 \\ C_2 & 240 \\ C_3 & 190 \\ C_4 & 270 \end{array}}_{P_C}$$

Manager's view $M_{C\to P}$?

$$M_{C\to A}\cdot M_{A\to P} = \begin{array}{c|ccc} & P_1 & P_2 & P_3 \\ \hline C_1 & 2\cdot 4+1\cdot 2 & 2\cdot 3+1\cdot 2 & 2\cdot 1+1\cdot 2 \\ C_2 & 0\cdot 4+3\cdot 2 & 0\cdot 3+3\cdot 2 & 0\cdot 1+3\cdot 2 \\ C_3 & 1\cdot 4+1\cdot 2 & 1\cdot 3+1\cdot 2 & 1\cdot 1+1\cdot 2 \\ C_4 & 1\cdot 4+2\cdot 2 & 1\cdot 3+2\cdot 2 & 1\cdot 1+2\cdot 2 \end{array}$$

$$= \begin{array}{c|ccc} & P_1 & P_2 & P_3 \\ \hline C_1 & 10 & 8 & 4 \\ C_2 & 6 & 6 & 6 \\ C_3 & 6 & 5 & 3 \\ C_4 & 8 & 7 & 5 \end{array}$$

$M_{C\to P}\cdot D = P_C$ holds.

1.1 Basic Concepts

Definition 1. A rectangular array of numbers $A \in \mathbb{R}^{m\times n}$, where $m, n \in \mathbb{N}_{>0}$, is called a *matrix*:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

The real numbers $a_{ij} \in \mathbb{R} \quad \forall 1 \le i \le m, 1 \le j \le n$ are called *elements* (or *entries*) of A.
For $1 \le i \le m$ the elements $a_{i1}, a_{i2}, \ldots, a_{in}$ are called the ith *row* of A. For $1 \le j \le n$ the elements $a_{1j}, a_{2j}, \ldots, a_{mj}$ are called the jth *column* of A.
By convention, uppercase letters denote matrices, lowercase letters their elements. As an abbreviation we will write $A = (a_{ij})_{\substack{i=1\ldots m\\ j=1\ldots n}}$ or even $A = (a_{ij})$ whenever m, n are clear from the context.
Matrices consisting of only one column are called *(column) vectors*:

$$\vec{v} \in \mathbb{R}^{m\times 1} = \mathbb{R}^m : \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}, \text{ with } v_i \in \mathbb{R} \quad \forall 1 \le i \le m.$$

1.1.1 Matrix Operations

1. Order:
   Two matrices $A \in \mathbb{R}^{m_1\times n_1}, B \in \mathbb{R}^{m_2\times n_2}$ have the same *order* iff $m_1 = m_2 \wedge n_1 = n_2$.

2. Equality:
   Two matrices are equal iff they have the same order and corresponding

elements are equal. For $A = (a_{ij}) \in \mathbb{R}^{m_1 \times n_1}, B = (b_{ij}) \in \mathbb{R}^{m_2 \times n_2}$ we have

$$A = B \iff m_1 = m_2 \wedge n_1 = n_2 \wedge (a_{ij} = b_{ij} \quad \forall 1 \leq i \leq m, 1 \leq j \leq n).$$

3. Sum and difference:
   Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}, B = (b_{ij}) \in \mathbb{R}^{m \times n}$, then

   $$A \pm B := (a_{ij} \pm b_{ij}).$$

   Example:

   $$A = \begin{pmatrix} 2 & 5 & 7 \\ 8 & 9 & 1 \end{pmatrix}, B = \begin{pmatrix} -1 & 6 & 4 \\ 5 & 2 & 0 \end{pmatrix}$$
   $$\rightsquigarrow \quad A + B = \begin{pmatrix} 1 & 11 & 11 \\ 13 & 11 & 1 \end{pmatrix}, A - B = \begin{pmatrix} 3 & -1 & 3 \\ 3 & 7 & 1 \end{pmatrix}$$

4. Scalar multiplication:
   Let $\lambda \in \mathbb{R}$, $A = (a_{ij}) \in \mathbb{R}^{m \times n}$. Then

   $$\lambda \cdot A = \lambda A := (\lambda \cdot a_{ij}) = (a_{ij} \cdot \lambda) = A \cdot \lambda.$$

   Note: Scalar multiplication is commutative.

   Example:
   $$A = \begin{pmatrix} 2 & 3 \\ 7 & 8 \end{pmatrix} \rightsquigarrow 5A = \begin{pmatrix} 10 & 15 \\ 35 & 40 \end{pmatrix}$$

5. Matrix multiplication:
   Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}, B = (b_{ij}) =\in \mathbb{R}^{k \times p}$. If $n = k$ then

   $$A \cdot B = AB := C = (c_{ij}) \in \mathbb{R}^{m \times p} \text{ with } c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}$$
   $$= a_{i1} \cdot b_{1j} + \cdots + a_{in} \cdot b_{nj}.$$

   If $n \neq k$, the matrix product does not exist!

   Example:

   The stars $*$ denote irrelevant (possibly different) real numbers.

   $$A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}, B = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix} \rightsquigarrow AB = \begin{pmatrix} 6 & 13 \\ -3 & -4 \end{pmatrix} \neq BA = \begin{pmatrix} 2 & * \\ * & * \end{pmatrix}$$

   Note: Matrix multiplication is not commutative.

German: „Spur“

6. Trace of a matrix:
   Suppose $A \in \mathbb{R}^{n \times n}$, i.e. A is a *square matrix*. $tr(A) := \sum_{i=1}^{n} a_{ii}$, i.e. the *trace* of a matrix is the sum of its *main diagonal elements*.

   Important properties: Given matrices A, B, C the equalities

$$\text{tr}(AB) = \text{tr}(BA)$$
$$\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$$

hold (assuming all matrix products exist). ABC, CAB and BCA are *cyclic* permutations. In general $\text{tr}(ABC) \neq \text{tr}(ACB)$.

Example:

$$A = \begin{pmatrix} 3 & 123 & * \\ * & 4 & * \\ 7 & * & -5 \end{pmatrix} \rightsquigarrow tr(A) = 3 + 4 + (-5) = 2$$

7. Transpose of a matrix
   Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, then $A^T = A' := (a_{ji}) \in \mathbb{R}^{n \times m}$. This effectively swaps the rows and columns of the matrix.

   Example:

$$A = \begin{pmatrix} 1 & 11 \\ 2 & 22 \\ 3 & 33 \end{pmatrix} \rightsquigarrow A^T = \begin{pmatrix} 1 & 2 & 3 \\ 11 & 22 & 33 \end{pmatrix}$$

Definition 2. If $A = A^T$ then A is called *symmetric*. Note that A necessarily has to be square for this property to hold, i.e. $A = A^T \implies A$ square, but in general A square $\not\Longrightarrow A = A^T$.
Rules:

(i) $(A^T)^T = A$

(ii) $(A + B)^T = A^T + B^T$

(iii) $(AB)^T = B^T A^T$

Proof of (iii): Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}, B = (b_{ij}) \in \mathbb{R}^{n \times p}$. Then $AB = (c_{ij})$ with $c_{ij} = \sum_{l=1}^{n} a_{il} \cdot b_{lj}$ and $(AB)^T = (c_{ji}) = \sum_{l=1}^{n} a_{jl} \cdot b_{li}$, $B^T = (b_{ji})$, $A^T = (a_{ji})$. Finally $B^T \cdot A^T = (b_{ji}) \cdot (a_{ji}) = \sum_{l=1}^{n} b_{li} \cdot a_{jl} = \sum_{l=1}^{n} a_{jl} \cdot b_{li} = (AB)^T$. □

1.1.2 Special Matrices

Identity matrix $I_{n \times n} := (\delta_{ij})$, where the *Kronecker delta* is defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

and $n \in \mathbb{N}_{>0}$.

Recall that $1 \cdot a = a \cdot 1 = a$ for all $a \in \mathbb{R}$.

Rule: $I_{m \times m} \cdot A = A \cdot I_{n \times n} = A \quad \forall A \in \mathbb{R}^{m \times n}, m, n \in \mathbb{N}_{>0}$.

Diagonal matrix $D = (d_{ij}) \in \mathbb{R}^{m \times n}$ is a *diagonal matrix* iff $m = n$ and $d_{ij} = 0$ $\forall\, 1 \leq i, j \leq n$ with $i \neq j$.

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Null matrix $0_{m\times n} \in \mathbb{R}^{m\times n}$

$$m \text{ rows} \left\{ \underbrace{\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}}_{n \text{ columns}} \right. = 0_{m\times n}$$

Rules: Let m, n, and k be positive integers and $A \in \mathbb{R}^{m\times n}$. Then

– $A + 0_{m\times n} = 0_{m\times n} + A = A$,
– $A \cdot 0_{n\times k} = 0_{m\times k}$ and $0_{k\times m} \cdot A = 0_{k\times n}$.

Upper (right) triangular matrix $U \in \mathbb{R}^{n\times n}$, with $u_{ij} = 0$ for $i > j$.

$$U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \ddots & u_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{pmatrix}$$

Lower (left) triangular matrix $L \in \mathbb{R}^{n\times n}$, with $l_{ij} = 0$ for $i < j$.

$$L = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}$$

Definition 3. Let A be a square matrix. If $A \cdot A = A$, then A is called *idempotent*.

Example.

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \text{ is idempotent.}$$

1.2 Linear equation systems

Application 2 (Warehouse clearance)**.** Recall that $M_{C\to P} \cdot D = P_C$, where $M_{C\to P}$ and P_C are known and D is not.

Linear equation system Find $\vec{x}$ such that

$$A\vec{x} = \vec{b},$$

where $A \in \mathbb{R}^{m\times n}, \vec{x} \in \mathbb{R}^n, \vec{b} \in \mathbb{R}^m$.

Example.

$$A = \begin{pmatrix} 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 11 \\ 5 \\ 3 \end{pmatrix}$$

$$\rightsquigarrow \left| \begin{array}{l} x_1 + 2x_3 + x_4 = 11 \\ x_2 + x_3 + 3x_4 = 5 \\ x_5 \qquad\qquad\quad = 3 \end{array} \right| \iff \left| \begin{array}{l} x_1 = 11 - 2x_3 - x_4 \\ x_2 = 5 - x_3 - 3x_4 \\ x_5 = 3 \end{array} \right|$$

Adding trivial constraints we have:

$$\left| \begin{array}{l} x_1 = 11 + (-2)x_3 + (-1)x_4 \\ x_2 = 5 + (-1)x_3 + (-3)x_4 \\ x_3 = 0 + 1x_3 + 0x_4 \\ x_4 = 0 + 0x_3 + 1x_4 \\ x_5 = 3 + 0x_3 + 0x_4 \end{array} \right| \rightsquigarrow \vec{x} = \underbrace{\begin{pmatrix} 11 \\ 5 \\ 0 \\ 0 \\ 3 \end{pmatrix}}_{=: \vec{s_0}} + \underbrace{\begin{pmatrix} -2 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{=: \vec{s_1}} \tilde{x}_3 + \underbrace{\begin{pmatrix} -1 \\ -3 \\ 0 \\ 1 \\ 0 \end{pmatrix}}_{=: \vec{s_2}} \tilde{x}_4$$

Therefore the solution space is:

$$\mathbb{L} = \{\vec{s_0} + \lambda_1 \vec{s_1} + \lambda_2 \vec{s_2} \mid \lambda_1, \lambda_2 \in \mathbb{R}\}$$

Linear equation systems can be manipulated by *elementary row operations* (or *feasible row operations*), i.e. operations for which the solution space stays the same:

– multiply an equation (a row) by $\lambda \in \mathbb{R} \setminus \{0\}$

– swap two equations (rows)

– add a multiple of one equation (row) to another equation (row)

Definition 4 (row echelon form)**.** Let $A \in \mathbb{R}^{m \times n}$. The first nonzero number of a row from the left is called *pivot* or *leading coefficient*. A is in *row echelon form* if

– the pivot of any row is always strictly to the right of the pivot of the row above it,

– rows without a pivot are below those with a pivot.

Definition 5 (reduced row echelon form)**.** $A \in \mathbb{R}^{m \times n}$ is in *reduced row echelon form* if

– it is in row echelon form,

– every leading coefficient is 1 and the only nonzero entry in its column.

Algorithm (Gauss-Jordan Elimination of $A\vec{x} = \vec{b}$)**.**

Step 1: Convert

$$A \mid \vec{b} := \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array} \right)$$

into row echelon form $A' \mid \vec{b}'$ using elementary row operations.

Example.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{array} \right) \xrightarrow[R3 \,||\, R3-R1]{R2 \,||\, R2-R1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & -2 & -2 \end{array} \right)$$

Step 2: If the last column ($\vec{b}'$) contains a pivot, the system has no solution. Stop.

Step 3: Convert $A' \mid \vec{b}'$ into reduced echelon form using elementary row operations.

Example.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & -2 & -2 \end{array}\right) \xhookrightarrow[R3 \,||\, -\frac{1}{2}R3]{R1 \,||\, R1+\frac{1}{2}R3} \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & 1 & 1 \end{array}\right)$$

$$\xhookrightarrow[R2 \,||\, -\frac{1}{2}R2]{R1 \,||\, R1+\frac{1}{2}R2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right)$$

Step 4: Read off the solution.

Example. $\mathbb{L} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Definition 6 (Linear independence)**.** $n \in \mathbb{N}_{>0}$ column vectors $\vec{a}^1, \ldots, \vec{a}^n \in \mathbb{R}^m$ are *linearly independent* iff $A\vec{x} = \vec{0}$ has the unique solution $\vec{x} = \vec{0}$, where $A = \left(\vec{a}^1 \mid \cdots \mid \vec{a}^n\right)$. Otherwise $\vec{a}^1, \ldots, \vec{a}^n$ are *linearly dependent*.

Example.

– $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix}$ are linearly independent.

– $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}$ are linearly dependent, since

$$2 \cdot \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} \iff 2 \cdot \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Definition 7. The *rank* of a matrix A, written $\operatorname{rank}(A)$, is the maximum number of linearly independent column vectors of A.

Observation. $\operatorname{rank}(0_{m \times n}) = 0$.

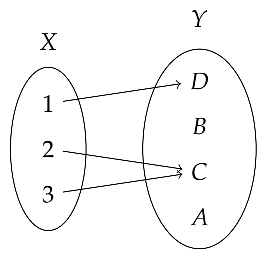
Observation. Elementary row operations do not change the rank of a matrix. (cf. Exercises)

Observation. If A is in row echelon form, then $\operatorname{rank}(A)$ equals the *number of pivots*.

Corollary. *$A\vec{x} = \vec{b}$ has a solution iff* $\operatorname{rank}(A) = \operatorname{rank}(A \mid \vec{b})$.

1.2.1 Vector spaces

Definition 8 (Function)**.** A *function* f from X to Y is a subset G_f of the Cartesian product $X \times Y := \{(x, y) \mid x \in X \wedge y \in Y\}$ such that every element of X is in the first component of exactly one pair in the subset.



$G_f = \{(1, D), (2, C), (3, C)\}$

Notation. $f\colon X \to Y$, X is called *domain*, Y *codomain*. $(x, y) \in G_f \iff f(x) = y$, where $x \in X$, $y \in Y$.

Example.

- $f\colon \mathbb{R} \to \mathbb{R}$, $\underbrace{f(x) = x^2}_{\text{alternatively } x \mapsto f(x) := x^2}$.
- $g\colon \mathbb{R}^2 \to \mathbb{R}^3$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} xy \\ x + y \\ x^2 + y^2 \end{pmatrix}$

Theorem 1. *Let $V = \mathbb{R}^{m \times n}$, $A, B \in V$ and $r, s \in \mathbb{R}$. There are functions*

$$+ : V \times V \to V,\ (A, B) \mapsto +(A, B) := A + B \tag{VS+}$$
$$\cdot : \mathbb{R} \times V \to V,\ (r, A) \mapsto \cdot(r, A) := r \cdot A \tag{VS$\cdot$}$$

satisfying

$$A + B = B + A \tag{VS1}$$
$$(A + B) + C = A + (B + C) \tag{VS2}$$
$$r(A + B) = rA + rB \tag{VS3}$$
$$(r + s)A = rA + sA \tag{VS4}$$
$$(rs)A = r(sA) \tag{VS5}$$
$$1 \cdot A = A \tag{VS6}$$

for all $A, B, C \in V$, $r, s \in \mathbb{R}$ and

$$\exists!\, 0 \in V \text{ with } A + 0 = A\ \forall A \in V \tag{VS7}$$
$$\text{For each } A \in V\ \exists!\, X \in V \text{ with } A + X = 0. \tag{VS8}$$

Here: $0 = 0_{m \times n}$.
$X = -A := (-1) \cdot A$

Definition 9. Let V be a set with functions "+" and "·" satisfying VS1–8 from Theorem 1, then $(V, +, \cdot)$ or simply V is called a *real vector space*.

Example.

- $\mathbb{R}^{m \times n}$ $\forall m, n \in \mathbb{N}_{>0}$
- the set of functions $: D \to \mathbb{R}$ (D arbitrary)
- polynomials with degree at most n $\forall n \in \mathbb{N}_{>0}$.

$(f + g)(w) = f(w) + g(w)$;
$(r \cdot f)(w) = r \cdot f(w)$

A polynomial is given by $\sum_{i=0}^{n} a_i \cdot x^i$; its degree is $\max\{i \mid a_i \neq 0, 0 \leq i \leq n\}$.

Elements of V are called "vectors". Therefore Definiton 6 (linear independence) can be applied to all real vector spaces.

Definition 10. Let V be a real vector space and $S := \{v_1, \dots, v_k\}$, $k \in \mathbb{N}$, be a set of elements of V. Then

$$\operatorname{span}(S) := \left\{ \sum_{i=1}^{k} \lambda_i v_i \mid k \in \mathbb{N}, v_i \in S, \lambda_i \in \mathbb{R} \right\}$$

is called the (linear) *span of S* and is itself a vector space.

Example.

– $\operatorname{span}\left(\left\{ \begin{pmatrix}2\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}\right) = \mathbb{R}^3$

– $\operatorname{span}\left(\left\{ \operatorname{span}\left(\left\{ \begin{pmatrix}0\\2\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}\right), \begin{pmatrix}0\\5\\3\end{pmatrix} \right\}\right)$
$= \operatorname{span}\left(\left\{ \begin{pmatrix}0\\2\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}\right) = \operatorname{span}\left(\left\{ \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}\right)$

Remark. Another common notation for the span of a set of vectors S is to enclose the set in angle brackets, that is, $\operatorname{span}(S) = \langle S \rangle$.

Definition 11. A *basis* of a (real) vector space V is a linearly independent subset of V that spans V.

Abbreviated as $e_1, \dots, e_n$ whenever the dimension of the surrounding space is clear.

Example. The *unit vectors* $e_1^{(n)}, \dots, e_n^{(n)}$, where

$$\mathbb{R} \ni e_i^{(n)} = \begin{pmatrix}0\\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0\end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow i\text{th position} \\ \\ \\ \\ \end{matrix}$$

form a basis of $\mathbb{R}^n$.

1.3 Inverse matrix

Given $A \in \mathbb{R}^{n \times n}$, does $X \in \mathbb{R}^{n \times n}$ with

$$AX = I_{n \times n} = XA$$

exist? If so, X is called *inverse matrix* and denoted by A^{-1}.

Computation. Solve the n equation systems $Ax_i = e_i$ simultaneously.

Example.

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 0 \end{pmatrix} \hookrightarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xhookrightarrow[R3||R3-2\cdot R1]{R2||R2-2\cdot R1} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & -3 & 0 & -2 & 0 & 1 \end{array}\right) \xhookrightarrow{R2\leftrightarrow R3} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & -3 & 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{array}\right)$$

$$\xhookrightarrow[R2||-\frac{1}{3}\cdot R2]{R1||R1+\frac{2}{3}\cdot R2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/3 & 0 & 2/3 \\ 0 & 1 & 0 & 2/3 & 0 & -1/3 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{array}\right)$$

Note that at the end we have the identity matrix ($I_{n\times n}$) on the left hand side of the vertical bar and the inverse of A, i.e. A^{-1}, on the right hand side.

Definition 12. Let $A \in \mathbb{R}^{n\times n}$. If A^{-1} exists A is called *regular* and *singular* otherwise.

1.4 Multilinear maps

Definition 13. Suppose $U_1, \dots, U_k, V$ are $\mathbb{R}$-vector spaces. A map $\Phi\colon U_1 \times \dots \times U_k \to V$ is called *k-linear* if $\forall u_i, v_i \in U_i, \lambda \in \mathbb{R}, 1 \le j \le k$ we have

$$\begin{aligned} &\Phi(u_1, \dots, u_{j-1}, u_j + v_j, u_{j+1}, \dots, u_k) \\ &= \Phi(u_1, \dots, u_k) + \Phi(u_1, \dots, u_{j-1}, v_j, u_{j+1}, \dots, u_k) \end{aligned}$$

and

$$\Phi(u_1, \dots, u_{j-1}, \lambda \cdot u_j, u_{j+1}, \dots, u_k) = \lambda \cdot \Phi(u_1, \dots, u_k).$$

If $U_1 = \dots = U_k, V = \mathbb{R}$ then Φ is called a *k-linear form on U* (bilinear form for $k = 2$).

Definition 14. Let U be a real vector space and $\Phi\colon U^k \to \mathbb{R}$ a k-linear form on U, then Φ is called *symmetric* if

$$\begin{aligned} &\Phi(u_1, \dots, u_{i-1}, u_i, u_{i+1}, \dots, u_{j-1}, u_j, u_{j+1}, \dots, u_k) \\ =&\Phi(u_1, \dots, u_{i-1}, u_j, u_{i+1}, \dots, u_{j-1}, u_i, u_{j+1}, \dots, u_k) \end{aligned}$$

for all $1 \le i < j \le k$ and $u_k \in U$. It is called *skew symmetric* if

$$\begin{aligned} &\Phi(u_1, \dots, u_{i-1}, u_i, u_{i+1}, \dots, u_{j-1}, u_j, u_{j+1}, \dots, u_k) \\ =& -\Phi(u_1, \dots, u_{i-1}, u_j, u_{i+1}, \dots, u_{j-1}, u_i, u_{j+1}, \dots, u_k) \end{aligned}$$

Definition and Proposition 1. For $n \in \mathbb{N}_{>0}$ $\det^{(n)}\colon (\mathbb{R}^n)^n \to \mathbb{R}$ is the unique skew-symmetric n-linear form on $\mathbb{R}^n$ such that $\det^{(n)}(e_1^{(n)}, \dots, e_n^{(n)}) = 1$. Let $A = (\vec{a}_1 | \cdots | \vec{a}_n) \in \mathbb{R}^{n\times n}$ (with column vectors $\vec{a}_k \in \mathbb{R}^n$ for $k = 1, \dots, n$), then the *determinant of A* is denoted by

$$|A| := \det(A) := \det{}^{(n)}(\vec{a}_1, \dots, \vec{a}_n).$$

The following rules apply:

- $\det^{(n)}(*, \vec{0}, *) = 0$
- $\det^{(n)}(*, \lambda\vec{u}, *) = \lambda \det^{(n)}(*, \vec{u}, *)$ for $\lambda \in \mathbb{R}, \vec{u} \in \mathbb{R}^n$
- $\det^{(n)}(*, \vec{u}, *, \vec{u}, *) = 0$ for $\vec{u} \in \mathbb{R}^n$
- $\det^{(n)}(*, \vec{u}, *, \vec{v}, *) = -\det^{(n)}(*, \vec{v}, *, \vec{u}, *)$ for $\vec{u}, \vec{v} \in \mathbb{R}^n$
- $\det^{(n)}(*, \vec{u}, *, \vec{v}, *) = \det^{(n)}(*, \vec{u}, *, \vec{v} + \lambda\vec{u}, *)$ for $\lambda \in \mathbb{R}, \vec{u}, \vec{v} \in \mathbb{R}^n$
- $\det(A) = \det(A^T)$
- $\det(I_{n\times n}) = 1$
- $\det\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} = \det\begin{pmatrix} d_1 & & * \\ & \ddots & \\ 0 & & d_n \end{pmatrix} = \det\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ * & & d_n \end{pmatrix} = \prod_{i=1}^n d_i$
- $\det\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right) = \det(A) \cdot \det(B)$, where $B \in \mathbb{R}^{m\times m}$ is also a square matrix.

Computation. Use elementary row operations to obtain row echelon form, then compute the product of the diagonal elements.

Example.

$$\begin{vmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 0 \end{vmatrix} \overset{R2||R2-2\cdot R1}{\underset{R3||R3-2\cdot R1}{=}} \begin{vmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & -3 & 0 \end{vmatrix} \overset{R2\leftrightarrow R3}{=} -\begin{vmatrix} 1 & 2 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 3$$

Definition 15. The *scalar product* of $\vec{x}, \vec{y} \in \mathbb{R}^n$ is defined as $\langle \vec{x}, \vec{y} \rangle := \vec{x}^T\vec{y}$.

$\vec{x}^T\vec{y} = (\vec{x}^T\vec{y})^T = \vec{y}^T\vec{x}$

Remark. $\langle \cdot, \cdot \rangle$ is a symmetric bilinear form on $\mathbb{R}^n$.

Definition 16. The (Frobenius) scalar product of $A, B \in \mathbb{R}^{m\times n}$ is defined as $\langle A, B \rangle := \operatorname{tr}(A^T B)$.

Remark. $\langle \cdot, \cdot \rangle$ is a symmetric bilinear form on $\mathbb{R}^{m\times n}$.

Definition 17. For an $\mathbb{R}$-vectorspace V a *norm* is a function $\|\cdot\| \colon V \to \mathbb{R}$ satisfying

(1) $\|a \cdot \vec{v}\| = |a| \cdot \|\vec{v}\|$ (absolute homogenity)

(2) $\|\vec{u} + \vec{v}\| \le \|\vec{u}\| + \|\vec{v}\|$ (triangle inequality)

(3) $\|\vec{v}\| = 0 \implies \vec{v} = 0$ (separate points)

for all $a \in \mathbb{R}, \vec{u}, \vec{v} \in V$.

Observation. $\|0\| = 0$ (from (3)), $\|-\vec{v}\| = \|\vec{v}\|$ (from (1)), (2) implies $\|\vec{v}\| \ge 0$ (positivity).

Remark. The scalar product induces a norm on $\mathbb{R}^n$: $\|\cdot\|_2 \colon \mathbb{R}^n \to \mathbb{R}$, $\vec{x} \mapsto \sqrt{\langle \vec{x}, \vec{x} \rangle}$.

Example. $\|\begin{pmatrix} 4 & 4 & 7 \end{pmatrix}^T\|_2 = \sqrt{4^2 + 4^2 + 7^2} = 9$

$\|\cdot\|_F \colon \mathbb{R}^{m\times n} \to \mathbb{R},\ A \mapsto \sqrt{\langle A, A\rangle}$ is called the Frobenius norm.

Definition 18. A distance function on a set M is a mapping $d\colon M \times M \to \mathbb{R}$ satisfying

(1) $d(x,y) = 0 \iff x = y$

(2) $d(x,y) = d(y,x)$ (symmetry)

(3) $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality)

for all $x, y, z \in M$.

Observation. $0 \overset{(1)}{=} d(x,x) \overset{(3)}{\leq} d(x,y) + d(y,x) \overset{(2)}{=} 2 \cdot d(x,y) \implies d(x,y) \geq 0$ $\forall x, y \in M$.

Remark. Any norm $\|\cdot\|$ on V induces a distance function: $d(x,y) := \|x - y\|$.

1.5 Eigenvalues and -vectors

Definition 19. Let $A \in \mathbb{R}^{n\times n}$. If $A\vec{x} = \lambda\vec{x}$, where $\lambda \in \mathbb{R}$ and $\vec{x} \in \mathbb{R}^n \setminus \{0\}$, then λ is called *eigenvalue* and $\vec{x}$ *eigenvector* (associated with λ) of A.

How can eigenvalues be found?

$$\begin{aligned} A\vec{x} = \lambda\vec{x} &\iff (A - \lambda I)\vec{x} = \vec{0} \\ \exists \text{ solution } \vec{x} \neq \vec{0} &\iff |A - \lambda I| = 0 \end{aligned}$$

Once an eigenvalue λ has been determined its associated eigenvector is found by solving the homogeneous system of linear equations $(A - \lambda I)\vec{x} = \vec{0}$.

Definition 20. Let $A \in \mathbb{R}^{n\times n}$, $1 \leq i, j \leq n$. The (i,j)*-minor of* A is $|A_{ij}|$, where A_{ij} arises from A by deleting the ith row and the jth column.

Example.

$$A = \begin{pmatrix} 1 & 12 & 13 \\ 10 & 20 & 30 \\ 4 & 15 & 6 \end{pmatrix} \rightsquigarrow A_{12} = \begin{pmatrix} 10 & 30 \\ 4 & 6 \end{pmatrix}$$

The (i,j)-*cofactor* is the (i,j)-minor multiplied by $(-1)^{i+j}$.

Theorem 2 (Laplace formula)**.** *Let* $A \in \mathbb{R}^{n\times n}$.

$$\det(A) = \underbrace{\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot |A_{ij}|}_{\text{expansion by ith row}} = \underbrace{\sum_{i=1}^{n} (-1)^{i+j} a_{ij} \cdot |A_{ij}|}_{\text{expansion by jth column}}$$

Corollary (Characteristic polynomial). $p(\lambda) := |A - \lambda I|$ *is a polynomial of degree n in* λ *for each* $A \in \mathbb{R}^{n \times n}$ *and*

$$p(\lambda) = (-\lambda)^n + b_{n-1}(-\lambda)^{n-1} + \cdots + b_1(-\lambda) + b_0 \tag{1.1}$$

for some $b_0, b_1, \ldots, b_{n-1} \in \mathbb{R}$.

Remark. The roots of the characteristic polynomial are precisely the eigenvalues of A. Denoting the eigenvalues by $\lambda_1, \ldots, \lambda_n$ we have

$$p(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdot \cdots \cdot (\lambda - \lambda_n) = (-1)^n \prod_{i=1}^{n}(\lambda - \lambda_i). \tag{1.2}$$

Theorem 3. *Using above notation:*

(a) $|A| = \prod_{i=1}^n \lambda_i$

(b) $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$

Proof. From (1.1) we have $p(0) = b_0 = |A|$. Plugging in $\lambda = 0$ into (1.2) we have $p(0) = (-1)^n(-1)^n \prod_{i=1}^n \lambda_i$. Putting both together we have $|A| = \prod_{i=1}^n \lambda_i$.
Now consider $\prod_{i=1}^n (a_{ii} - \lambda)$ (product of main diagonal) and determine b_{n-1} in (1.1). The Laplace formula yields $b_{n-1} = \sum_{i=1}^n a_{ii} = \mathrm{tr}(A)$. Finally, consider the coefficient of $(-\lambda)^{n-1}$ in the expansion of (1.2): $b_{n-1} = \sum_{i=1}^n \lambda_i$. This implies $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$. □

Definition 21. $A \in \mathbb{R}^{n \times n}$ is called *orthogonal* if $A^T A = A A^T = I_{n \times n}$.

Definition 22. Each $E \in \mathbb{R}^{n \times n}$ resulting from an elementary row operation on $I_{n \times n}$ is called an *elementary* $n \times n$*-matrix* .

Lemma 4. *Let E be an elementary* $m \times m$ *matrix obtained by performing a particular row operation on* $I_{m \times m}$. *For each* $A \in \mathbb{R}^{m \times n}$ *applying the same row operation on A yields* $E \cdot A$.

Lemma 5. *Let E be an elementary* $n \times n$ *matrix and* $B \in \mathbb{R}^{n \times n}$ *be arbitrary. Then*

$$\det(E \cdot B) = \det(E) \cdot \det(B).$$

Lemma 6. *Each elementary matrix is invertible. Its inverse is an elementary matrix.*

$$\underbrace{E_k \cdot \ldots \cdot E_2 \cdot E_1}_{\text{elementary matrices}} \cdot A = U \iff A = E_1^{-1} \cdot E_2^{-1} \cdot \ldots \cdot E_k^{-1} \cdot U$$

where U is in (reduced) row echelon form. If A is invertible then there exist elementary matrices with $A = E_1 \cdot E_2 \cdot \ldots \cdot E_k$.

Theorem 7. *For* $A, B \in \mathbb{R}^{n \times n}$, $|A \cdot B| = |A| \cdot |B|$.

Proof. (a) A invertible $\implies \exists$ elementary matrices E_i with $A = E_1 \cdot \ldots \cdot E_k$. Now,

$$\begin{aligned}\det(A \cdot B) &= \det(E_1 \cdot \ldots \cdot E_k \cdot B)\\ &= \det(E_1) \cdot \det(E_2 \cdot \ldots \cdot E_k \cdot B)\\ &\vdots\\ &= \det(E_1) \cdot \det(E_2) \cdot \ldots \cdot \det(E_k) \cdot \det(B)\\ &= \det(E_1 \cdot \ldots \cdot E_k) \cdot \det(B)\\ &= \det(A) \cdot \det(B).\end{aligned}$$

(b) A singular $\implies \exists$ elementary matrices E_i with $A = E_1 \cdot \ldots \cdot E_k \cdot R$, where R is in row echelon form so its bottom rows contain zeroes only. As before

$$\det(A \cdot B) = \det(E_1 \cdot \ldots \cdot E_k) \cdot \det(R \cdot B).$$

Since the bottom rows of R contain zeroes only $\det(R \cdot B) = 0$. Therefore $\det(A \cdot B) = 0$, $\det(A) = 0$ and the statement follows.

□

$1 = \det(I) = \det(A \cdot A^{-1}) = \det(A) \cdot \det(A^{-1})$

Corollary. *If $A \in \mathbb{R}^{n\times n}$ is regular, then $\det(A^{-1}) = \det(A)^{-1}$.*

1.6 Diagonalization

Lemma 8. *Let $A, P \in \mathbb{R}^{n\times n}$, P regular. Then A and $P^{-1}AP$ have the same eigenvalues.*

Proof. Consider the characteristic polynomial:

$$\begin{aligned}|P^{-1}AP - \lambda I| = |P^{-1}AP - P^{-1}\lambda IP| &= |P^{-1}(A - \lambda I)P| = |P^{-1}||A - \lambda I||P|\\ &= |A - \lambda I|.\end{aligned}$$

□

Definition 23. $A \in \mathbb{R}^{n\times n}$ is *diagonalizable* if there exists a regular matrix $P \in \mathbb{R}^{n\times n}$ and diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that $P^{-1}AP = D$.

Theorem 9. *$A \in \mathbb{R}^{n\times n}$ is diagonalizable iff it has a set of n linearly independent eigenvector $\vec{x}_1, \ldots \vec{x}_n$. Further*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D$$

where $P = (\vec{x}_1|\cdots|\vec{x}_n)$ and $\lambda_1, \ldots, \lambda_n$ are the corresponding eigenvalues.

Proof. "$\Longleftarrow$": Suppose A has n linearly independent eigenvectors $\vec{x}_1, \ldots, \vec{x}_n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. (This implies that P is regular.) Then

$$AP = A(\vec{x}_1|\cdots|\vec{x}_n) = (A\vec{x}_1|\cdots|A\vec{x}_n) = (\lambda_1\vec{x}_1|\cdots|\lambda_n\vec{x}_n) = P\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

"$\Longrightarrow$": A diagonalizable $\Longrightarrow \exists\ \bar{P}, \bar{D} : \bar{P}^{-1}A\bar{P} = \bar{D}$ where $\bar{D}$ is diagonal. $\Longrightarrow A\bar{P} = \bar{P}\bar{D}$. Let $\bar{P} = (\vec{y_1}|\cdots|\vec{y_n})$ then

$$A\bar{P} = \bar{P}\bar{D} \iff A(\vec{y_1}|\cdots|\vec{y_n}) = (\vec{y_1}|\cdots|\vec{y_n})\begin{pmatrix} \bar{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \bar{\lambda_n} \end{pmatrix} = (\bar{\lambda_1}\vec{y_1}|\cdots|\bar{\lambda_n}\vec{y_n}).$$

This implies that $\vec{y_1}, \ldots, \vec{y_n}$ must be eigenvectors of A and the diagonal elements of $\bar{D}$ the corresponding eigenvalues. □

1.7 Subspaces attached to a matrix

Definition 24. Let $A = (\vec{a_1}|\cdots|\vec{a_n})$, where $\vec{a_i} \in \mathbb{R}^m$. Then $\mathrm{Col}(A) := \langle \vec{a_1}, \ldots, \vec{a_n} \rangle$ is called the *column space* (or *image*) of A.

Observation. $\mathrm{Col}(A) = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\}$. Elementary column operations on A, i.e. elementary row operations on A^T do not change the column space.

Definition 25. The dimension of an $\mathbb{R}$-vector space V is the cardinality, i.e. the number of vectors, of a basis of V. Notation: $\dim(V)$.

Example. $\left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle$ has dimension 3.

Definition 26. Let $\vec{a_1}, \ldots, \vec{a_m} \in \mathbb{R}^{1\times n}$ be the rows of $A \in \mathbb{R}^{m\times n}$, i.e. $A^T = (\vec{a_1}^T|\cdots|\vec{a_m}^T)$. Then $\mathrm{Row}(A) := \langle \vec{a_1}, \ldots, \vec{a_m} \rangle$ is called the *row space* of A.

Observation. $\mathrm{Row}(A) = \{\vec{x}A \mid \vec{x}^T \in \mathbb{R}^m\}$. Elementary row operations do not change the row space.

Corollary. *Let $A \in \mathbb{R}^{m\times n}$ and R be a row echelon form of A. Then*

- *the nonzero row vectors of R are a basis of* $\mathrm{Row}(A)$,
- $\dim(\mathrm{Row}(A)) = \mathrm{rank}(A)$,
- *columns of A whose corresponding columns in R contain a pivot form a basis of* $\mathrm{Col}(A)$, *and*
- $\dim(\mathrm{Col}(A)) = \dim(\mathrm{Row}(A)) = \mathrm{rank}(A)$.

Definition 27. For $A \in \mathbb{R}^{m\times n}$ the *kernel* $\ker(A)$ (or null space) is given by

$$\{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}.$$

Theorem 10. *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$\dim(\mathrm{Col}(A)) + \dim(\ker(A)) = n.$$

1.8 Quadratic forms

Definition 28. A *quadratic form* on $\mathbb{R}^n$ is a function $Q\colon \mathbb{R}^n \to \mathbb{R}$, $\vec{x} \mapsto \vec{x}^T A\vec{x}$, where the associated matrix $A \in \mathbb{R}^{n\times n}$ is symmetric.

Definition 29. A quadratic form as well as its associated matrix A is

(a) *positive definite* if $\vec{x}^T A\vec{x} > 0\ \forall x \neq 0$,

(b) *positive semi-definite* if $\vec{x}^T A\vec{x} \geq 0\ \forall x \neq 0$,

(c) *negative definite* if $\vec{x}^T A\vec{x} < 0\ \forall x \neq 0$,

(d) *negative semi-definite* if $\vec{x}^T A\vec{x} \leq 0\ \forall x \neq 0$,

(e) *indefinite* if $\exists\ \vec{x}, \vec{y}$ with $\vec{x}^T A\vec{x} > 0$ and $\vec{y}^T A\vec{y} < 0$.

Definition 30. Two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ with $\langle \vec{x}, \vec{y}\rangle = 0$ are called *orthogonal*.

Theorem 11 (Spectral theorem for symmetrical matrices). *Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then:*

(a) All eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are real numbers.

(b) Eigenvectors corresponding to different eigenvalues are orthogonal.

(c) A is diagonalizable. Moreover, the corresponding matrix P (such that $P^{-1}AP$ is diagonal) can be chosen to be orthogonal.

Theorem 12. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then:*

(a) A is positive definite. $\iff \lambda_1, \ldots, \lambda_n \in \mathbb{R}_{>0}$.

(b) A is positive semi-definite. $\iff \lambda_1, \ldots, \lambda_n \in \mathbb{R}_{\geq 0}$.

(c) A is negative definite. $\iff \lambda_1, \ldots, \lambda_n \in \mathbb{R}_{<0}$.

(d) A is negative semi-definite. $\iff \lambda_1, \ldots, \lambda_n \in \mathbb{R}_{\leq 0}$.

(e) A is indefinite. $\iff$ A has eigenvalues of opposite signs.

Proof. Choose $P \in \mathbb{R}^{n\times n}$ orthogonal such that

$$P^T AP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D$$

and set $\vec{y} = P^T\vec{x}$. Then $\vec{x}^T A\vec{x} = \vec{x}^T PDP^T\vec{x} = \vec{y}^T D\vec{y} = \sum_{i=1}^n \lambda_i y_i^2$, which implies the statements. □

Definition 31. Let $A \in \mathbb{R}^{n\times n}$. A $k \times k$ submatrix of A formed by deleting the same $n-k$ rows and columns, say $i_1 < i_2 < \cdots < i_{n-k}$, is called a k-th order *principal submatrix* of A. The corresponding determinant is called a k-th order *principal minor*. The *leading* principal submatrix/minor is the one where the last $n-k$ rows and columns are deleted.

Theorem 13. *Let $A \in \mathbb{R}^{n \times n}$ be symmetric, D_k be a leading principal minor of order k, and A_k a principal minor of order k, then:*

(a) *A is positive definite.* $\iff D_k > 0$ *for* $k = 1, \ldots, n$.

(b) *A is positive semi-definite.* $\iff A_k \geq 0$ *for* $k = 1, \ldots, n$.

(c) *A is negative definite.* $\iff (-1)^k D_k > 0$ *for* $k = 1, \ldots, n$.

(d) *A is negative semi-definite.* $\iff (-1)^k A_k \geq 0$ *for* $k = 1, \ldots, n$.

1.9 Miscellaneous

Definition and Proposition. Let $A \in \mathbb{R}^{n \times n}$ be a matrix with real eigenvalues only (e.g., a symmetric matrix). Let $\lambda_1, \ldots, \lambda_k$ be the pairwise different eigenvalues of A with multiplicity $m_j^{(alg)}$ for $j = 1, \ldots, k$, i.e., $\sum_{j=1}^{k} m_j^{(alg)} = n$. Then the characteristic polynomial can be written as:

$$p(\lambda) = \prod_{j=1}^{k} (\lambda - \lambda_j)^{m_j^{(alg)}}$$

The multiplicity $m_j^{(alg)}$ is called the *algebraic multiplicity* of λ_j.
The number of linearly independent eigenvectors corresponding to λ_j is called its *geometric multiplicity* and is denoted by $m_j^{(geo)}$.
The following inequality holds: For each $j = 1, \ldots, k$

$$m_j^{(alg)} \geq 1 \text{ and } 1 \leq m_j^{(geo)} \leq m_j^{(alg)}$$

In other words: There is always at least one eigenvector corresponding to any eigenvalue. Its algebraic multiplicity tells us how many linearly independent eigenvectors to expect *at most*. For example, if the algebraic multiplicity is one we know that there is only one corresponding eigenvector; if it's two there *might* be two. In fact, for a symmetric matrix we then know that there have to be two.

Definition 32. $A \in \mathbb{R}_{\geq 0}^{n \times n}$ is called a (row) *stochastic matrix* if its rows sum to 1, i.e. $\sum_{j=1}^{n} a_{ij} = 1$ for $i = 1, \ldots, n$.

Lemma 14. *If λ is an eigenvalue of a (row) stochastic matrix A, then $|\lambda| \leq 1$.*

Proof. Let $\vec{x}$ be an eigenvector of A corresponding to eigenvalue λ. Choose k such that $|x_k| = \max_{1 \leq j \leq n} |x_j|$, i.e., x_k is the largest component of $\vec{x}$. Then:

$$\begin{aligned}
|\lambda| \cdot |x_k| = |\lambda \cdot x_k| = \left| \sum_{j=1}^{n} a_{kj} x_j \right| &\leq \sum_{j=1}^{n} |a_{kj} x_j| \\
&= \sum_{j=1}^{n} |a_{kj}| \cdot |x_j| \leq \sum_{j=1}^{n} |a_{kj}| \cdot |x_k| = |x_k| \cdot \sum_{j=1}^{n} |a_{kj}| \\
&= |x_k| \underbrace{\sum_{j=1}^{n} a_{kj}}_{=1} = |x_k|
\end{aligned}$$

This implies $|\lambda| \leq 1$. $\square$

Definition 33. Let $A \in \mathbb{R}^{n\times n}$ and $C = (c_{ij}) \in \mathbb{R}^{n\times n}$, where c_{ij} is the (i,j)-cofactor of A. $\operatorname{adj} A := C^T$ is called the *adjoint* of A.

Theorem 15. *If $A \in \mathbb{R}^{n\times n}$ is regular, then $A^{-1} = \det(A)^{-1} \cdot \operatorname{adj} A$.*

Observation.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = +\,ad - bc$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{aligned}[t] &+ a_{11}\cdot a_{22}\cdot a_{33} - a_{11}\cdot a_{23}\cdot a_{32} \\ &+ a_{12}\cdot a_{23}\cdot a_{31} - a_{12}\cdot a_{21}\cdot a_{33} \\ &+ a_{13}\cdot a_{21}\cdot a_{32} - a_{13}\cdot a_{22}\cdot a_{31}\end{aligned}$$

Example. Let $A = \begin{pmatrix} 2 & 4 & 5 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. The (i,j)-cofactors, $1 \leq i,j \leq 3$, are:

$$\begin{array}{lll} c_{11} = +\left|\begin{smallmatrix} 3 & 0 \\ 0 & 1 \end{smallmatrix}\right| = 3 & c_{12} = -\left|\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}\right| = 0 & c_{13} = +\left|\begin{smallmatrix} 0 & 3 \\ 1 & 0 \end{smallmatrix}\right| = -3 \\ c_{21} = -\left|\begin{smallmatrix} 4 & 5 \\ 0 & 1 \end{smallmatrix}\right| = -4 & c_{22} = +\left|\begin{smallmatrix} 2 & 5 \\ 1 & 1 \end{smallmatrix}\right| = -3 & c_{23} = -\left|\begin{smallmatrix} 2 & 4 \\ 1 & 0 \end{smallmatrix}\right| = 4 \\ c_{31} = +\left|\begin{smallmatrix} 4 & 5 \\ 3 & 0 \end{smallmatrix}\right| = -15 & c_{32} = -\left|\begin{smallmatrix} 2 & 5 \\ 0 & 0 \end{smallmatrix}\right| = 0 & c_{33} = +\left|\begin{smallmatrix} 2 & 4 \\ 0 & 3 \end{smallmatrix}\right| = 6 \end{array}$$

So, $\operatorname{adj} A = \begin{pmatrix} 3 & -4 & -15 \\ 0 & -3 & 0 \\ -3 & 4 & 6 \end{pmatrix}$. Further,

$$\begin{vmatrix} 2 & 4 & 5 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{vmatrix} \begin{matrix} 2 & 4 \\ 0 & 3 \\ 1 & 1 \end{matrix} = 2\cdot 3\cdot 1 + 4\cdot 0\cdot 1 + 5\cdot 0\cdot 1 - 5\cdot 3\cdot 1 - 2\cdot 0\cdot 0 - 4\cdot 0\cdot 1$$

$$= 6 - 15 = -9$$

Finally, $A^{-1} = -\frac{1}{9}\begin{pmatrix} 3 & -4 & -15 \\ 0 & -3 & 0 \\ -3 & 4 & 6 \end{pmatrix}$.

Theorem 16 (Cramer's rule)**.** *Let $A \in \mathbb{R}^{n\times n}$ be regular. The unique solution of $A\vec{x} = \vec{b}$ is given by $x_i = \frac{\det(B_i)}{\det(A)}$ $\forall\, 1 \leq i \leq n$, where B_i arises from A by replacing the ith column of A by $\vec{b}$.*

Example.

$$\underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 12 & 2 & -3 \\ 3 & 4 & 1 \end{pmatrix}}_{=A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ 5 \\ -4 \end{pmatrix}}_{=\vec{b}}, \quad B_3 = \begin{pmatrix} 1 & 1 & 0 \\ 12 & 2 & 5 \\ 3 & 4 & -4 \end{pmatrix}.$$

Then $\det(A) = 35$ and $\det(B_3) = 35$, so

$$x_3 = \frac{\det(B_3)}{\det(A)} = \frac{35}{35} = 1.$$

Chapter 2

Calculus

2.1 Limits

Definition 34. A pair (M, d) is called a *metric space* if M is a set and d a distance funtion on M. A pair $(V, \|\cdot\|)$ is called normed ($\mathbb{R}$-)vector space if V is an ($\mathbb{R}$-)vector space and $\|\cdot\|$ a norm on V.

Definition 35. A *sequence* in a metric space (M, d) is a function $a : \mathbb{N} \to M, k \mapsto a_k (\in M)$. Notation: $(a_k)_{k\in\mathbb{N}}$, (a_k).
A *subsequence* is a sequence of the form $(a_{n_k})_{k\in\mathbb{N}}$ where $(n_k)_{k\in\mathbb{N}}$ is a strictly increasing sequence of positive integers, i.e., $n_1 < n_2 < n_3 < \ldots$.

Example. For the metric space $\mathbb{R}^2$, where the distance function is induced by $\|\cdot\|_2$, $a : \mathbb{N}_{>0} \to \mathbb{R}^2, k \mapsto \begin{pmatrix} 1-1/k \\ 1/k^2 \end{pmatrix}$ or $\begin{pmatrix} 1-1/n \\ 1/n^2 \end{pmatrix}_{n\in\mathbb{N}_{>0}}$ is a sequence.

Definition 36. A point x of a metric space (M, d) is the *limit* of a sequence (x_n) (in M) if, for all $\varepsilon \in \mathbb{R}_{>0}$, there is an $N \in \mathbb{N}$ such that for every $n \geq N$, we have $d(x_n, x) < \varepsilon$. Notation: $\lim_{n\to\infty} x_n = x$. We say that (x_n) *converges* to x.

Example (cont.). $\lim_{n\to\infty} \begin{pmatrix} 1-1/n \\ 1/n^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Definition 37. Let V be an $\mathbb{R}$-vector space. Two norms $\|\cdot\|$, $\|\cdot\|'$ on V are called *equivalent* if there exist $\underline{M}, \overline{M} \in \mathbb{R}_{>0}$ such that $\underline{M} \cdot \|x\| \leq \|x\|' \leq \overline{M} \cdot \|x\|$ for all $x \in V$.

Theorem 17. *For each $\mathbb{R}$-vector space of finite dimension all norms are equivalent.*

Notation. Whenever we speak of a limit or convergence in $\mathbb{R}^n$ or $\mathbb{R}^{m\times n}$ we assume a distance function induced by *any* norm.

Lemma 18. *For convergent sequences (a_n), (b_n) in a normed $\mathbb{R}$-vector space we have*

- $\lim\limits_{n\to\infty} (a_n \pm b_n) = \lim\limits_{n\to\infty} a_n \pm \lim\limits_{n\to\infty} b_n$, *and*
- $\lim\limits_{n\to\infty} c \cdot a_n = c \cdot \lim\limits_{n\to\infty} a_n$ *for all* $c \in \mathbb{R}$.

Lemma 19. *For convergent sequences (a_n), (b_n) in $\mathbb{R}$ and all $p \in \mathbb{R}_{>0}$ we have*

– $\lim_{n\to\infty}(a_n \cdot b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$,

– $\lim_{n\to\infty}\left(\frac{a_n}{b_n}\right) = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$, *provided* $\lim_{n\to\infty} b_n \neq 0$,

– $\lim_{n\to\infty}\left(a_n^p\right) = \left(\lim_{n\to\infty} a_n\right)^p$,

– *if* $a_n \leq b_n$ *for all* $n \geq N$, *then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$,

– *if* $a_n \leq c_n \leq b_n$ *for all* $n \geq N$ *and* $\lim_{n\to\infty} a_n = L = \lim_{n\to\infty} b_n$, *then* $\lim_{n\to\infty} c_n = L$, *(this is known as the* Squeeze theorem*), and*

– *a sequence is convergent if and only if all of its subsequences are convergent.*

Definition 38. Let (A, d_A) and (B, d_B) be two metric spaces, $M \subseteq A$, $N \subseteq B$ and $f : M \to N$ be a function. For a limit point p of M and $L \in N$ we say that the *limit of f as x approaches p is L* if:

Note that p need not be in the domain of f. Also $\lim_{x\to p} f(x) \neq f(p)$ is possible if $p \in M$.

For all $\varepsilon \in \mathbb{R}_{>0}$ there is $\delta \in \mathbb{R}_{>0}$ such that $d_B(f(x), L) < \varepsilon$ whenever $0 < d_A(x, p) < \delta$.

Notation: $\lim_{x\to p} f(x) = L$.

Lemma 20. *Using the same notation:*

$$\lim_{x\to p} f(x) = L \iff \lim_{n\to\infty} f(x_n) = L$$

for all sequences (x_n) *with* $\lim_{n\to\infty} x_n = p$ *and* $x_i \neq p\ \forall i \in \mathbb{N}_{>0}$.

Definition 39. A function f is *continuous at* $\overline{x} \in M$ if for every sequence $(x_n) \subseteq M$ that converges to $\overline{x}$, the sequence $(f(x_n))$ converges to $f(\overline{x})$, i.e. $\lim_{x\to\overline{x}} f(x) = f(\overline{x})$.
f is called *continuous* if it is continuous at every point in M.

Definition 40. A sequence (x_n) in a metric space (M, d) is a *Cauchy sequence* if for all $\varepsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{N}_{>0}$ such that for all $i, j \geq N : d(x_i, x_j) < \varepsilon$.

Theorem 21. *Let* (x_n) *be a sequence in* $\mathbb{R}^n$.

$$(x_n) \text{ is a Cauchy sequence.} \iff (x_n) \text{ has a limit point.}$$

Example. Let $a_n = 1 - \frac{1}{2} + \frac{1}{3} - \cdots + \frac{(-1)^{n+1}}{n}$ for all $n \in \mathbb{N}_{>0}$. Does $\lim_{n\to\infty} a_n$ exist?
For $m \geq n$ and $m - n$ even we have

$$\begin{aligned}
|a_m - a_n| &= \left| \overbrace{\frac{1}{n+1} - \frac{1}{n+2}}^{>0} + \overbrace{\frac{1}{n+3} - \frac{1}{n+4}}^{>0} + \cdots + \overbrace{\frac{1}{m-1} - \frac{1}{m}}^{>0} \right| \\
&= \frac{1}{n+1} \underbrace{- \frac{1}{n+2} + \frac{1}{n+3}}_{<0} - \cdots \underbrace{- \frac{1}{m-2} + \frac{1}{m-1}}_{<0} - \frac{1}{m} \\
&\leq \frac{1}{n+1} - \frac{1}{m} \leq \frac{1}{n}.
\end{aligned}$$

For $m > n$ and $m - n$ odd we have

$$|a_m - a_n| = |a_m - a_{m-1} + a_{m-1} - a_n| \leq |a_m - a_{m-1}| + |a_{m-1} - a_n| \\ \leq \frac{1}{m} + \frac{1}{n} \leq \frac{2}{n}.$$

Therefore the limit exists. ($\lim_{n\to\infty} a_n = \ln 2$.)

Definition 41. A metric space (M, d) in which every Cauchy sequence converges to an element of M is called *complete*.

Example. $\mathbb{R}^n$ is complete for all $n \in \mathbb{N}_{>0}$.

Definition 42. The vector $\vec{x} \in \mathbb{R}^n$ is called an *accumulation point* of the sequence $(\vec{x}_k)$ (over $\mathbb{R}^n$) if for any given $\varepsilon > 0$ there are infinitely many integers l such that $\|\vec{x}_l - \vec{x}\| < \varepsilon$.

2.2 Classification of sets

Definition 43. A subset U of a metric space (M, d) is called *open*, if there exists $\varepsilon \in \mathbb{R}_{>0}$ for any point $x \in U$ such that $B_\varepsilon(x) \subseteq U$, where $B_\varepsilon(x) := \{y \in M \mid d(x, y) < \varepsilon\}$ (open ball around x).

Example. The interval $(0, 1) = \{x \in \mathbb{R} \mid 0 < x < 1\}$ is open (in $\mathbb{R}$). Open balls are open sets.

Observation.

(a) Any union of open sets is open.

(b) The finite intersection of open sets is open.

Definition 44. A subset U of a metric space (M, d) is called *closed*, if its *complement* $U^c = M \setminus U$ is open.

Example. $\{(x, y) \mid x^2 + y^2 \leq 1\} \subseteq \mathbb{R}^2$ (disc with radius 1) and $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\} \subseteq \mathbb{R}$ are closed.

Observation.

(a) Any intersection of closed sets is closed.

(b) The finite union of closed sets is closed.

Definition 45. Let U be a subset of a metric space (M, d). The *closure* $\text{cl}(U)$ (or $\overline{U}$) of U is the intersection of all closed sets (in M) containing U.

$\text{cl}(U)$ is closed.

Example. $\overline{B}_\varepsilon(x) = \{y \in M \mid d(x, y) \leq \varepsilon\}$ is a closed ball around x.

Theorem 22. *A set U in $\mathbb{R}^n$ is closed if, whenever $(x_n) \subseteq U$ is a convergent sequence, its limit is also contained in U.*

Example. $U = [0,1) = \{x \in \mathbb{R} \mid 0 \le x < 1\}$, $x_n = 1 - 1/n \in U\ \forall n \in \mathbb{N}_{>0}$ but $\lim_{n\to\infty} x_n = 1 \notin U$.

Definition 46. Let U be a subset of a metric space (M, d). The *interior* $\operatorname{int}(U)$ of U is the union of all open sets contained in U.

Definition 47. Given a metric space. A point x is in the *boundary* of S if every open ball around x contains both points in S and points in the complement of S.

Example. The boundary of $[0,1]$ is $\{0,1\}$. As is the boundary of $[0,1)$.

Theorem 23. *The set of boundary points of a set S equals* $\operatorname{cl}(S) \cap \operatorname{cl}(S^c)$.

Definition 48. A subset U of a normed $\mathbb{R}$-vector space $(V, \|\cdot\|)$ is *bounded* if there exists a constant $B \in \mathbb{R}$ such that $\|x\| \le B$ for all $x \in U$. (In other words, U is contained in some ball (of finite radius).)

Definition 49. A subset U of a normed $\mathbb{R}$-vector space $(V, \|\cdot\|)$ is *compact* if it is closed and bounded.

Example. The n-dimensional box $\times_{i=1}^{n} [a_i, b_i] \subseteq \mathbb{R}^n$ is compact.

Theorem 24 (Bolzano-Weierstrass)**.** *Let C be a compact subset in $\mathbb{R}^n$ and let (x_n) be any sequence in C. Then, (x_n) has a convergent subsequence whose limit lies in C.*

Corollary. *Let C be a compact subset in $\mathbb{R}^n$, $(V, \|\cdot\|)$ be a normed $\mathbb{R}$-vector space, and $f : C \to V$ be a continuous function. Then, f is bounded, i.e. there exists a constant $\Delta \in \mathbb{R}$ such that $\|f(x)\| \le \Delta$ for all $x \in C$.*

Proof. Assume that f is not bounded, i.e. for each $n \in \mathbb{N}_{>0}$ there exists a point $x_n \in C$ with $\|f(x_n)\| \ge n$. Let (y_k) be a convergent subsequence with limit $y \in C$. Since f is continuous we have $\lim_{k\to\infty} f(y_k) = f(y)$. But that is impossible if $\|f(x_{i_k})\| \ge i_k$, where $y_k = x_{i_k}$. □

Example. $f : [0,1]^3 \to \mathbb{R}^2$, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x^2 + yz \\ x + y + z \end{pmatrix}$ is bounded.

2.3 Extreme values

Definition 50. Let $S \subseteq \mathbb{R}$. The *infimum* $\inf(S)$ is the largest real number that is less than or equal to every number in S. If no such number exists, then we define $\inf(S) := -\infty$. If $S = \emptyset$, we define $\inf(S) := +\infty$.
The *supremum* $\sup(S)$ is given by $\sup(S) := -\inf(-S)$, where $-S := \{-x \mid x \in S\}$.

Example. $\inf([0,1]) = 0, \sup([0,1)) = 1$.

Definition 51. Let $f : M \to \mathbb{R}$ be a function. f has a *global (or absolute) maximum* at $x^\star \in M$ if $f(x^\star) \geq f(x)$ for all $x \in M$. Similarly, f has a *global (or absolute) minimum* at $x^\star \in M$ if $f(x^\star) \leq f(x)$ for all $x \in M$.

$\sup(\{f(x) | x \in (-1,1)\}) = 1$, but $f(x) < 1 \ \forall x \in (-1,1)$

Example. $f : (-1,1) \to \mathbb{R}, x \mapsto x^2$ has a global minimum (at $x = 0$) but no global maximum.

Theorem 25. *(Extreme value theorem) Let C be a compact subset in $\mathbb{R}^n$ and $f : C \to \mathbb{R}$ a continuous function. If $C \neq \emptyset$ then f has a global maximum and a global minimum.*

2.4 Differentiable functions

Definition 52. Let $f\colon M \to \mathbb{R}$ be a function and $x_0 \in M \subseteq \mathbb{R}$ be an accumulation point. The *derivative of f at x_0* , written $f'(x_0)$ or $\frac{\mathrm{d}f}{\mathrm{d}x}(x_0)$ is given by

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

if this limit exists.
If so we say that f is differentiable at x_0 with derivative $f'(x_0)$.
If f is differentiable at all $x_0 \in M$, we say that f is differentiable with derivative f'.

Observation. Differentiability implies continuity.

Theorem 26 (Linearity of the derivative)**.** *Let $f\colon M \to \mathbb{R}$, $g\colon M \to \mathbb{R}$ be differentiable on $M \subseteq \mathbb{R}$, then we have*

(1) $(r \cdot f)' = r \cdot f' \ \forall r \in \mathbb{R}$, *and*

(2) $(f + g)' = f' + g'$.

Theorem 27 (product and quotient rule)**.** *Let $f\colon M \to \mathbb{R}$, $g\colon M \to \mathbb{R}$ be differentiable on $M \subseteq \mathbb{R}$, then*

(1) $(f \cdot g)' = f' \cdot g + f \cdot g'$, *and*

(2) $(f/g)' = \dfrac{f'g - fg'}{g^2}$ *if* $g(x) \neq 0 \ \forall x \in M$.

Definition 53. Let $f\colon X \to Y$ and $g\colon Y \to Z$ be two functions. The *composition* is given by $g \circ f\colon X \to Z,\ x \mapsto g(f(x))$.

Theorem 28 (Chain rule, one dimension)**.** *Let $g\colon M_1 \to M_2$, $f\colon M_3 \to \mathbb{R}$ with $M_1, M_2, M_3 \subseteq \mathbb{R}$ and $M_2 \subseteq M_3$, then $(f \circ g)' = (f' \circ g) \cdot g'$.*

Example.

$$(x^2 + 4x + 1)^7 = f \circ g \text{ with } f : x \mapsto x^7, g : x \mapsto x^2 + 4x + 1, \text{ so:}$$
$$(f \circ g)' = \underbrace{7 \cdot (\underbrace{x^2 + 4x + 1}_{g(x)})^6}_{f'(g(x))} \cdot \underbrace{2x + 4}_{g'(x)}.$$

Definition 54. A function $f\colon X \to Y$ is called *surjective* if for all $y \in Y$ there exists $x \in X$ such that $f(x) = y$. It is called *injective* if $f(a) = f(b)$ implies $a = b$.
If it is both surjective and injective, then it is called *bijective*.
For all $\tilde{X} \subseteq X$ the *restriction of f to $\tilde{X}$* is given by $f\big|_{\tilde{X}}\colon \tilde{X} \to Y,\ x \mapsto f(x)$.

Definition 55. Let $f\colon X \to Y$ be a function. A function $g\colon Y \to X$ is called *inverse function* of f it $g \circ f = \mathrm{id}_X$, where $\mathrm{id}_X\colon X \to X,\ x \mapsto x$.

Theorem 29. *An inverse function exists for a function $f\colon X \to Y$ if and only if f is bijective. In that case the inverse function is unique and denoted by f^{-1}.*

Definition 56. If a function f is differentiable and its derivative is continuous then we call f *continuously differentiable* and write $f \in C^1$.

Theorem 30 (Inverse function theorem)**.** *Let $f\colon I \to M \subseteq \mathbb{R}$ be a bijective C^1 function, where $I \subseteq \mathbb{R}$ is an interval. If $f'(x) \neq 0\ \forall x \in I$, then*

$$\left(f^{-1}\right)'(\xi) = \frac{1}{f'(f^{-1}(\xi))}$$

for $\xi := f(x)$ and f^{-1} is a C^1 function on the interval $f(I) = \{\, f(x) \mid x \in I \,\}$.

Example. $f\colon [-\pi/2, \pi/2] \to [-1, 1],\ x \mapsto \sin(x)$, $\arcsin := f^{-1}$. Then $f'(x) = \cos(x)$ and

$$\left(f^{-1}\right)'(x) = \arcsin'(x) = \frac{1}{\cos(\arcsin(x))}.$$

Since $\sin^2(x) + \cos^2(x) = 1$ we have $\cos(x) = \pm\sqrt{1 - \sin^2(x)}$. For $a := \arcsin(x) \in [-\pi/2, \pi/2]$ we have $\cos(a) \geq 0$. Hence,

$$\arcsin'(x) = \frac{1}{\sqrt{1 - \sin^2(arcsin(x))}} = \frac{1}{\sqrt{1 - x^2}}.$$

Definition 57. For each $k \in \mathbb{N}_{\geq 0}$ we denote by $\dfrac{\mathrm{d}^k f}{\mathrm{d}x^k}$ the *k-th order derivative* of a funtion f by setting

$$\frac{\mathrm{d}^0 f}{\mathrm{d}x^0} = f \text{ and } \frac{\mathrm{d}^k f}{\mathrm{d}x^k} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}^{k-1} f}{\mathrm{d}x^{k-1}}\right)$$

for all $k \geq 1$. If $\mathrm{d}^k f/\mathrm{d}x^k$ exists and is continuous, we write $f \in C^k$ ($k \geq 0$). If $f \in C^k\ \forall k \in \mathbb{N}_{\geq 0}$, we write $f \in C^\infty$.

Example. $f(x) = x^3$, $\dfrac{\mathrm{d}f}{\mathrm{d}x} = 3x^2$, $\dfrac{\mathrm{d}^2 f}{\mathrm{d}x^2} = \dfrac{\mathrm{d}f}{\mathrm{d}x}(3x^2) = 6x$, $\dfrac{\mathrm{d}^3 f}{\mathrm{d}x^3} = 6$, $\dfrac{\mathrm{d}^k f}{\mathrm{d}x^k} = 0\ \forall k \geq 4,\ f \in C^\infty$.

Definition 58. Let $f\colon M \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function and $\bar{x} = (x_1, \dots, x_n) \in M$ be an accumulation point. The *partial derivative* $\frac{\partial f}{\partial x_i}$ *at* $\bar{x}$ is given by

$$\frac{\partial f}{\partial x_i}(\bar{x}) = \lim_{h \to 0} \left(f\begin{pmatrix} x_1 \\ \vdots \\ x_i + h \\ \vdots \\ x_n \end{pmatrix} - f\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} \right) / h \quad (1 \leq i \leq n)$$

if this limit exists.

Remark. Treat f as a function with a single variable x_i.

Definition 59. For each $k \in \mathbb{N}_{>0}$ and each $1 \leq i_1, \dots, i_k \leq n$

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}$$

is called a *k-th order partial derivative* of $f\colon M \subseteq \mathbb{R}^n \to \mathbb{R}^m$. For $k \geq 2$ we set

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}} = \frac{\partial}{\partial x_{i_1}} \left(\frac{\partial^{k-1} f}{\partial x_{i_2} \cdots \partial x_{i_k}} \right).$$

Remark. There are n^k k-th order partial derivatives.

Theorem 31 (Schwarz' theorem). *If* $f\colon \mathbb{R}^n \to \mathbb{R}$ *has continuous second order partial derivatives, then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \forall i, j \in \{1, \dots, n\}.$$

Remark. A function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ can be represented by m real-valued functions $f_i\colon \mathbb{R}^n \to \mathbb{R}$ $(1 \leq i \leq m)$, so

$$f\colon \mathbb{R}^n \to \mathbb{R}^m,\ \vec{x} \mapsto \begin{pmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{pmatrix}.$$

Definition 60. Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}^m$ be a function and $\bar{x}$ an accumulation point of U. If there exists a matrix $A_{\bar{x}} \in \mathbb{R}^{n \times m}$ with

$$\lim_{x \to \bar{x}} \frac{\|f(x) - f(\bar{x}) - A_{\bar{x}} \cdot (x - \bar{x})\|}{\|x - \bar{x}\|} = 0$$

then f is *(totally) differentiable* at $\bar{x} \in U$.
The linear map $\mathrm{d}f_{\bar{x}}\colon \Delta x \mapsto A_{\bar{x}} \cdot \Delta x$ is called *(total) derivative* of f at $\bar{x}$. A function is *(totally) differentiable* if its total derivative exists at every point in its domain.

Theorem 32. *Let* $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} : \mathbb{R}^n \supseteq U \to \mathbb{R}^m$, $\bar{x} \in U$, *f (totally) differentiable at* $\bar{x}$ *and*

$$J_f(\bar{x}) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\bar{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\bar{x}) \end{pmatrix}$$

($J_f(\bar{x})$ *is called* Jacobi matrix*). Then we have* $\mathrm{d}f_{\bar{x}}\colon \mathbb{R}^n \to \mathbb{R}^m$, $\Delta x \mapsto J_f(\bar{x})\Delta x$, *i.e. we can compute* $A_{\bar{x}}$ *via partial derivation.*

Theorem 33. *If all partial derivatives of f exist and are continuous, then f is (totally) differentiable.*

Example. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{cases} \frac{xy(x+y)}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$. Note that $f \in C^0$. Then

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} \frac{-y^2(x^2-2xy-y^2)}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} \frac{-x^2(y^2-2xy-x^2)}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}.$$

So $J_f(0,0) = \begin{pmatrix} 0 & 0 \end{pmatrix}$. However, using the sequence $\begin{pmatrix} 1/n \\ 1/n \end{pmatrix}_{n \in \mathbb{N}}$ (with limit $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$) we find that

$$\lim_{x \to \binom{0}{0}} \frac{\|f(x) - f\binom{0}{0} - \begin{pmatrix} 0 & 0 \end{pmatrix} \cdot x\|_1}{\|x\|_1} = \lim_{n \to \infty} \frac{\|\frac{1}{n}\|_1}{\left\|\begin{pmatrix} 1/n \\ 1/n \end{pmatrix}\right\|_1} = \lim_{n \to \infty} \frac{|1/n|}{|1/n| + |1/n|} = \frac{1}{2} \neq 0.$$

Thus, f is not (totally) differentiable at $\binom{0}{0}$.

Theorem 34 (Chain Rule, higher dimensions)**.** *Let* $f\colon D \to W$, $g\colon W \to V$ *be (totally) differentiable, then* $(g \circ f)$ *is (totally) differentiable and* $\mathrm{d}(g \circ f)_{\bar{x}} = \mathrm{d}g_{f(\bar{x})} \circ \mathrm{d}f_{\bar{x}}$*, that is,* $J_{g \circ f}(\bar{x}) = J_g(f(\bar{x})) \cdot J_f(\bar{x})$.

Definition 61. The *directional derivative of* $f\colon \mathbb{R}^n \to \mathbb{R}$ *along the vector* $\vec{a}$, denoted by $f'_{\vec{a}}(\vec{x})$, is given by

$$f'_{\vec{a}}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{a}) - f(\vec{x})}{h}.$$

Lemma 35. *Let* $f\colon \mathbb{R}^n \to \mathbb{R}$ *be (totally) differentiable function and* $\vec{a} \in \mathbb{R}^n$ *be a direction. Then*

$$f'_{\vec{a}}(\vec{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{x}) \cdot a_i.$$

Proof. For some $\vec{x}, \vec{a} \in \mathbb{R}^n$ set $g(t) := f(\vec{x} + t \cdot \vec{a})$. Note that g is a univariate function. It holds

$$g'(t) = \lim_{h \to 0} \frac{g(t+h) - g(t)}{h} \text{ and therefore } g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h}.$$

Plugging in we have

$$g'(0) = \lim_{h \to 0} \frac{f(\vec{x} + h \cdot \vec{a}) - f(\vec{x})}{h} = f'_{\vec{a}}(\vec{x}).$$

Setting $l(t) := \vec{x} + t \cdot \vec{a}$ we see that we may apply the chain rule (for higher dimensions) in order to differentiate $g(t)$ since $g = f \circ l$. We have $\mathrm{d}g_t = \mathrm{d}f_{l(t)} \circ \mathrm{d}l_t$. The corresponding Jacobians are

$$J_f(l(t)) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x} + t \cdot \vec{a}) & \cdots & \frac{\partial f}{\partial x_n}(\vec{x} + t \cdot \vec{a}) \end{pmatrix} \text{ and } J_l(t) = \vec{a}.$$

Multiplying the two matrices and plugging in $t = 0$ proves the lemma. □

Remark. The directional derivative along the i-th unit vector e_i coincides with the corresponding partial derivative:

$$f'_{e_i}(\vec{x}) = \frac{\partial f}{\partial x_i}(\vec{x}).$$

Example. The function of the previous example was shown to be not (totally) differentiable at $\binom{0}{0}$. We now show that Lemma 35 is indeed not applicable. Let $\vec{a} = \binom{1}{1}$. Using the definition of the directional derivative we have:

$$f'_{\vec{a}}(0,0) = \lim_{h \to 0} \frac{f(h,h)}{h} = \lim_{h \to 0} \frac{2h^3}{2h^2 \cdot h} = 1.$$

However, Lemma 35 would yield:

$$f'_{\vec{a}}(0,0) = \frac{\partial f}{\partial x}(0,0) \cdot 1 + \frac{\partial f}{\partial y}(0,0) \cdot 1 = 0.$$

Definition 62. For each $k \in \mathbb{N}_{\geq 0}$ the *k-th order Taylor polynomial* of a function $f \colon \mathbb{R} \supseteq U \to \mathbb{R}$ at a is given by

$$T_k^a(x) = \sum_{i=0}^{k} \frac{\mathrm{d}^i f}{\mathrm{d}x^i}(a) \cdot \frac{(x-a)^i}{i!}.$$

Example. Let $f(x) = \sqrt{x}$, $k = 2$ and $a = 4$. Then $f'(x) = 0.5 \cdot x^{-0.5}$, $f''(x) = -0.25 \cdot x^{-1.5}$, $f(4) = 2$, $f'(4) = 0.25$, $f''(4) = -0.03125$, so

$$T_2^4(x) = 2 + \frac{x-4}{4} - \frac{(x-4)^2}{64}.$$

With this $T_2^4(5) = 143/64 = 2.234375 \approx 2.236067977 \approx \sqrt{5}$.

Definition 63. For a function $f \colon \mathbb{R}^n \supseteq U \to \mathbb{R}^m$ we write $f \in C^k$ if all k-th order partial derivatives exist and are continuous. $f \in C^\infty$ if $f \in C^k$ $\forall k \in \mathbb{N}_{>0}$.

Theorem 36 (Taylor expansion)**.** *Let $f\colon \mathbb{R} \supseteq U \to \mathbb{R}$ be a C^{k+1} function. For any points $x, a \in U$, there exists a point $\tilde{x}$ between x and a in U such that*

$$R_k^a(x) = \frac{(x-a)^{k+1}}{(k+1)!} \cdot \frac{\mathrm{d}^{k+1} f}{\mathrm{d}x^{k+1}}(\tilde{x}),$$

where $f(x) = T_k^a(x) + R_k^a(x)$ ($R_k^a(x)$ is called remainder).

Remark. $\lim_{x\to a} \frac{R_k^a(x)}{(x-a)^k} = \lim_{x\to a} \frac{1}{(k+1)!} \cdot \underbrace{\frac{\mathrm{d}^{k+1} f}{\mathrm{d}x^{k+1}}(\tilde{x})}_{\text{This term is bounded.}} \cdot (x-a) = 0.$

Definition 64. For $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$ the *gradient* of f is given by $\nabla f(x) = J_f(x)^T$.

Definition 65. For each $k \in \mathbb{N}_{>0}$ the *k-th order Taylor polynomial* of a function $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}^m$ at $\vec{a} \in U$ is given by

$$\begin{aligned} T_k^{\vec{a}}(\vec{x}) = f(\vec{a}) &+ \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a}) \cdot (x_i - a_i) \\ &+ \frac{1}{2} \sum_{i,j}^{n} \frac{\partial f}{\partial x_i \partial x_j}(\vec{a}) \cdot (x_i - a_i)(x_j - a_j) \\ &+ \dots \\ &+ \frac{1}{k!} \sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k}{\partial x_1 \cdots \partial x_k} f(\vec{a}) \cdot \prod_{j=1}^{k} (x_{i_j} - a_{i_j}). \end{aligned}$$

Definition 66. Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$ be a C^2 function, then the *Hessian matrix at $\vec{a}$* is given by

$$H_f(\vec{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\vec{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\vec{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\vec{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\vec{a}) \end{pmatrix}$$

Observation. The second order Taylor polynomial for a function $f\colon \mathbb{R}^n \supseteq\to \mathbb{R}$ is

$$\begin{aligned} T_2^{\vec{a}}(\vec{x}) &= f(\vec{a}) + J_f(\vec{a}) \cdot (\vec{x} - \vec{a}) + \frac{1}{2}(\vec{x} - \vec{a})^T H_f(\vec{a})(\vec{x} - \vec{a}) \\ &= f(\vec{a}) + (\vec{x} - \vec{a})^T \nabla f(\vec{a}) + \frac{1}{2}(\vec{x} - \vec{a})^T H_f(\vec{a})(\vec{x} - \vec{a}). \end{aligned}$$

Theorem 37. *Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$, $\vec{x}, \vec{a} \in U$ such that the line segment from $\vec{x}$ to $\vec{a}$ lies in U and $f \in C^2$. For $R_2^{\vec{a}}(\vec{x}) = f(\vec{x}) - T_2^{\vec{a}}(\vec{x})$ we have*

$$\lim_{\vec{x}\to\vec{a}} \frac{R_2^{\vec{a}}(\vec{x})}{\|\vec{x} - \vec{a}\|^2} = 0.$$

Chapter 3

Unconstrained optimisation

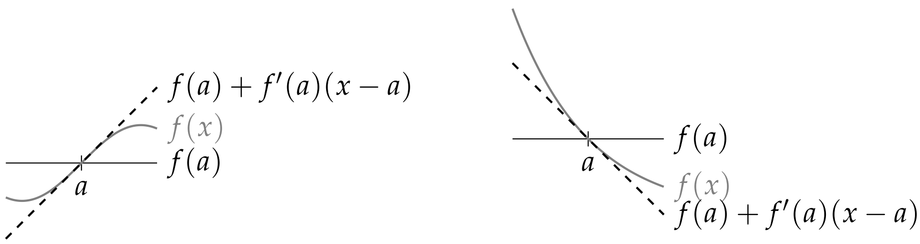
Definition 67. Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$.

- $x^* \in \operatorname{int}(U)$ is a *a local (or relative) maximum or minimum of f*, it there exists a ball $B_\varepsilon(x^*)$ such that $f(x^*) \geq f(x)$ or $f(x^*) \leq f(x)$ for all $x \in B_\varepsilon(x^*) \cap U$.
- $x^* \in \operatorname{int}(U)$ is a strict local (or relative) maximum or minimum of f, if there exists a ball $B_\varepsilon(x^*)$ such that $f(x^*) > f(x)$ or $f(x^*) < f(x)$ for all $x^* \neq x \in B_\varepsilon(x^*) \cap U$.

Figure 3.1 illustrates above definition.
Consider the one-dimensional case to get a notion of how to find extrema: From Theorem 36 (page 30) we know that a function $f\colon \mathbb{R} \to \mathbb{R}$ may be represented at a point a as

$$f(x) = f(a) + f'(a)(x-a) + R_1^a(x).$$

Assume that the error term $R_1^a(x)$ is negligibly small and that $f'(a) \neq 0$. Then the situation is similar to either the left ($f(a) > 0$) or the right ($f(a) < 0$) picture:



Clearly neither situation shows an extremum: In the left picture $f(a+\varepsilon) > f(a)$ and $f(a-\varepsilon) < f(a)$ for $\varepsilon \in \mathbb{R}_{>0}$ (the situation on the right is similar). We conclude that $f(a) = 0$ is a *necessary* condition for an extremum, hence the following defintion:

Definition 68. Let $f\colon \mathbb{R}^n \supseteq\to \mathbb{R}$. Any point $x^* \in U$ that satisfies $\nabla f(x^*) = \vec{0}$ is called *critical or stationary point*.

Observation. Global maxima/minima are either

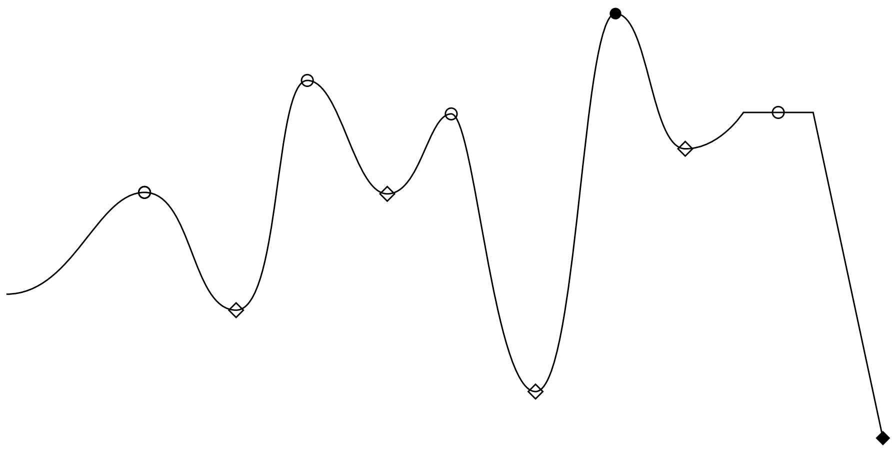


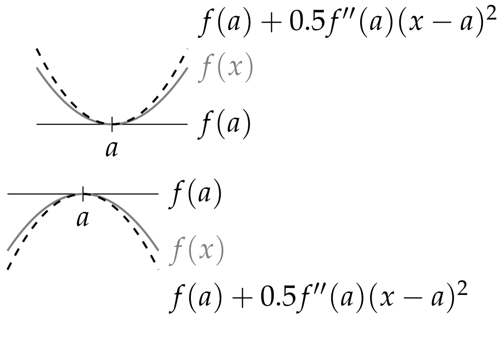
Figure 3.1: Graph with extrema highlighted: outlined circles/diamonds indicate local maxima/minima (all but the last are strict); filled circles/diamonds indicate global maxima/minima.

- critical points,
- points where the function is not differentiable, or
- boundary points.

Looking at the one-dimensional case proved useful before so let us do that again. The 2nd order Taylor expansion of f at a is

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + R_2^a(x).$$

Suppose that $R_2^a(x)$ is negligibly small and that $f'(a) = 0$. That leaves us with $f(a) + 0.5 \cdot f''(a)(x-a)^2$. If $f''(a) > 0$ or $f''(a) < 0$ we have a strict local minimum (left picture) or a strict local maximum (right picture):



If $f''(a) = 0$ we are unable to make a clear statement (it all depends on $R_2^a(x)$). For higher dimensions we resort to quadratic forms:

Theorem 38. *Let $f \colon \mathbb{R}^n \supseteq U \to \mathbb{R}$ be a C^2 function and $x^* \in \operatorname{int}(U)$ a critical point.*

(1) If $H_f(x^*)$ is negative definite, then x^* is a strict local maximum of f.

(2) *If $H_f(x^*)$ is positive definite, then x^* is a strict local minimum of f.*

(3) *If $H_f(x^*)$ is indefinite, then x^* is neither a local maximum nor a local minimum. (saddle point)*

Recall that $T_2^a(x) = f(a) + J_f(a)(x-a) + \frac{1}{2}(x-a)^T H_f(a)(x-a)$.

Remark. For $R_2^a(x)$ different limit results are available.

Theorem 39. *Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$ be a C^2 function and $x^* \in \operatorname{int}(U)$. If x^* is a local maximum (or local minimum) of f, then $J_f(x^*) = 0$ and $H_f(x^*)$ is negative (positive) semi-definite.*

Example. Let $f\colon U \to \mathbb{R}$, $x \mapsto 2 - |x| + 2|x-2| + 2|x+2|$ with $U = [-3,-1] \cup [1,3]$. Since f is continuous and U is compact (closed and bounded) a global maximum and a global minimum exists.
$\operatorname{int}(U) = (-3,-1) \cup (1,3)$ and f can be written as

$$f(x) = \begin{cases} 2-3x & \text{if } x \le -2 \\ 10+x & \text{if } -2 < x \le -1 \\ 10-x & \text{if } 1 \le x \le 2 \\ 2+3x & \text{if } 2 < x \end{cases}.$$

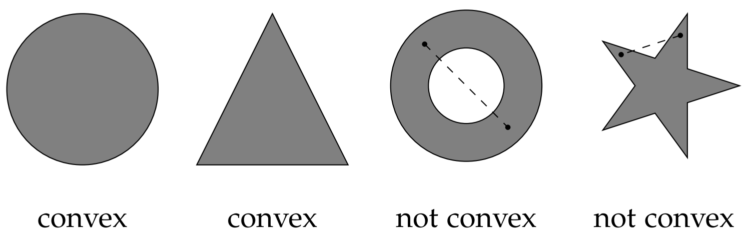
$f|_{U\setminus\{-2,2\}}$ is differentiable with

$$\frac{\mathrm{d}f|_{U\setminus\{-2,2\}}}{\mathrm{d}x} = \begin{cases} -3 & \text{if } x \le -2 \\ +1 & \text{if } -2 < x \le 1 \\ -1 & \text{if } 1 < x \le 2 \\ +3 & \text{if } 2 < x \end{cases}.$$

Note that $\mathrm{d}f|_{U\setminus\{-2,2\}}/\mathrm{d}x \neq 0$ for all $x \in U$. Looking at the boundary we find that $f(-3) = 11 = f(3)$ and $f(-1) = 9 = f(1)$. And at the non-differentiable points we have $f(-2) = 8 = f(2)$. So we have global maxima at ± 3 and global minima at ± 2.

3.1 Global optima – convex and concave optimisation

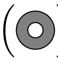
Definition 69. A set $S \subseteq \mathbb{R}^n$ is called *convex* if $\lambda x + (1-\lambda)y \in S$ for all $x, y \in S$ and all $\lambda \in [0,1]$.



Definition 70. Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$ be a function where U is convex.

- f is *convex* if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for all $x, y \in U$ and all $\lambda \in [0,1]$.
- f is *concave* if $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$ for all $x, y \in U$ and all $\lambda \in [0,1]$.

Definition 71. The *convex hull* $\operatorname{conv}(U)$ is the smallest convex set containing U.

Remark. $\operatorname{conv}(\{u_1, \dots, u_n\}) = \{\sum_{i=1}^n \lambda_i u_i \mid \lambda_i \geq 0,\ \sum_{i=1}^n \lambda_i = 1\}$.
conv () = 

Example. $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\binom{x_1}{x_2} \mapsto x_1^2 + x_2^2$ is convex.

Theorem 40. *Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$ be a C^1 function, where U is convex:*

- *f is concave on U iff $f(y) - f(x) \leq J_f(x)(y-x)$ for all $x, y \in U$, i.e. $f(y) - f(x) \leq \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)(y_i - x_i)$.*
- *f is convex on U iff $f(y) - f(x) \geq J_f(x)(y-x)$.*

Theorem 41. *Let f be a concave (convex) function on an open, convex subset $U \subseteq \mathbb{R}^n$. If x^* is a critical point of f, then x^* ($\in \operatorname{int}(U)$) is a global maximizer (global minimizer) of f on U.*

Theorem 42. *Let $f\colon \mathbb{R}^n \supseteq U \to \mathbb{R}$, where U is convex. If $f \in C^1$ is concave (convex) and $x^* \in U$ satisfies $J_f(x^*)(y - x^*) \leq 0$ (≥ 0) for all $y \in U$, then x^* is a global maximizer (minimizer) of f on U.*

Example. $f\colon T \to \mathbb{R}$, $\binom{x}{y} \mapsto x^{1/4}y^{1/4}$ where $T = \{(x,y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0, x + y \leq 2\} = \operatorname{conv}(\{\binom{0}{0}, \binom{2}{0}, \binom{0}{2}\})$ (T is convex). Then

$$\frac{\partial f}{\partial x}(1,1)\cdot(x-1) + \frac{\partial f}{\partial y}(1,1)\cdot(y-1) = \frac{x-1}{4} + \frac{y-1}{4} = \frac{x+y-2}{4} \leq 0,$$

so $(1,1)$ is a global maximum.

Chapter 4

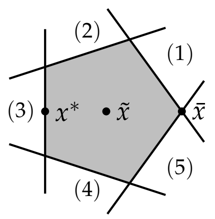
Constrained optimisation

Definition 72. A *non-linear optimisation problem* (NLP) is given by

$$\begin{aligned} \min_{\vec{x}\in\mathbb{R}^n} & f(x) \\ \text{subject to } & h_i(\vec{x}) \geq 0 \quad \forall i \in I \\ & h_i(\vec{x}) = 0 \quad \forall i \in E \end{aligned}$$

where $I = \{1, \dots, m\}$ is the index set of inequality constraints, $E = \{m + 1, \dots, m + p\}$ the index set of equality constraints (note that both are finite), $f \colon \mathbb{R}^n \to \mathbb{R} \in C^1$ is called cost function and $h_i \colon \mathbb{R}^n \to \mathbb{R} \in C^1$ for $i \in I \cup E$ are called constraint functions.

Notation. The *admissible set* is defined as $Z := \{\vec{x} \in \mathbb{R}^n \mid h_i(\vec{x}) \geq 0 \ \forall i \in I,\ h_i(\vec{x}) = 0 \ \forall i \in R\}$.
$\bar{x} \in Z$ is a *global minimum* if $f(\bar{x}) \leq f(\vec{x})$ for all $\vec{x} \in Z$.
$\bar{x} \in Z$ is a *local minimum* if $f(\bar{x}) \leq f(\vec{x})$ for all $\vec{x} \in Z \cup B_\varepsilon(\bar{x})$ for some $\varepsilon \in \mathbb{R}_{>0}$.
An inequality $h_i(\vec{x}) \geq 0$ ($i \in I$) is called *active* (*inactive*) at the point $\vec{x}$ if $h_i(\vec{x}) = 0$ ($h_i(\vec{x}) \neq 0$).
The *active set* at the point $\vec{x}$ is defined as $A(\vec{x}) := \{i \in I \cup E \mid h_i(\vec{x}) = 0\}$.



Example. The lines in the picture on the right depict inequality constraints. The shaded area is the admissible set. The actice set for the points $\tilde{x}$, x^* and $\bar{x}$ are $A(\tilde{x}) = \{\}$, $A(x^*) = \{3\}$ and $A(\bar{x}) = \{1, 5\}$, respectively.

For *affine-linear constraints*, i.e. $h_i(\vec{x}) = \vec{a}_i^T + b_i \geq 0$, the admissible set is a *polyhedron* (*polytope* if bounded, like on the right).

4.1 Necessary conditions

Definition 73. For (NLP) the *Lagrange function* is defined as

$$L\left(\vec{x}, \vec{\lambda}\right) = f(\vec{x}) - \sum_{i \in I \cup E} \lambda_i h_i(x)$$

(the λ_i are called *Lagrange multipliers*).

Theorem 43 (Karush-Kuhn-Tucker (KKT) conditions)**.** *Let $\bar{x}$ be a local minimum of (NLP) and suppose a* constraint qualification (CQ) *is satisfied (see below). Then it holds:*

(1) $\nabla f(\bar{x}) = \sum_{i \in I \cap E} \bar{\lambda}_i \nabla h_i(\bar{x})$ ($\iff$ $\nabla L\left(\bar{x}, \bar{\lambda}\right) = \vec{0}$); the gradient of the cost function is a linear combination of the gradients of the constraint functions.

(2) $h_i(\bar{x}) \geq 0 \ \forall i \in I$.

(3) $h_i(\bar{x}) = 0 \ \forall i \in E$; *(2) & (3):* primal admissibility.

(4) $\bar{\lambda}_i \geq 0 \ \forall i \in I$; sign condition for Lagrange multipliers of inequality constraints.

(5) $\bar{\lambda}_i \cdot h_i(\bar{x}) = 0 \ \forall i \in I$; complementarity condition.

Definition 74. A point $\bar{x}$ satisfying (1)–(5) (together with a Lagrange multiplier $\vec{\lambda} \in \mathbb{R}^{|I|+|E|}$ is called a *KKT point*.

Above definition is a generalisation of a critical/stationary point. At least eight different versions of the KKT-Theorem may be found in books. Merging results from different books should be avoided.
The interest in local minima stems from engineering problems. In economics we are usually interested in local maxima, however, $\min f(x) = -\max -f(x)$, so above theorem is applicable. Depending on the context the Lagrange function might also be given as $L(x, \lambda) = f(x) + \sum_i \lambda_i h_i(x)$. Finally, the inequality constraints are sometimes formulated as $h_i(x) \leq 0$, but since $h_i(x) \geq 0 \iff -h_i(x) \leq 0$ this is inconsequential as well.

Definition 75 (Constraint qualifications (CQ))**.**

v1: All constraint functions affine-linear, i.e. $h_i(\vec{x}) = \vec{a}_i^T \vec{x} + b_i \ \forall i \in I \cup E$. (Linear programming)

v2: All active constraint functions at the point $\bar{x}$ are affine linear, i.e., $h_i(\vec{x}) = \vec{a}_i^T \vec{x} + b_i \ \forall i \in A(\bar{x}) \subseteq I \cup E$.

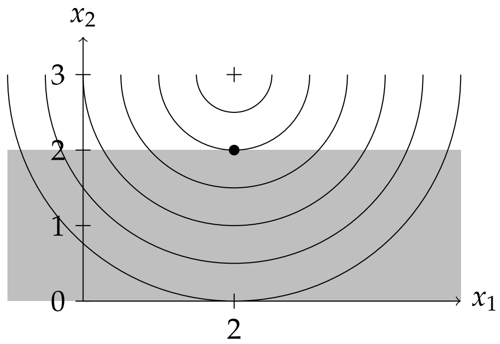
v3: The set $\{ \nabla h_i(\bar{x}) \mid i \in A(\bar{x}) \}$ is linear independent. (Linear independence constraint qualification, LICQ)

Refer to books about optimisation for more CQs.

Example. Consider the (NLP)

$$\begin{aligned} \min f(x_1, x_2) &= (x_1 - 2)^2 + (x_2 - 3)^2 \\ \text{s.t. } h_1(x_1, x_2) &= 2 - x_2 \geq 0 \\ h_2(x_1, x_2) &= x_2 \geq 0 \end{aligned}$$

and observe the the requirements of Theorem 43 are satisfied (all constraint functions are linear). Let us plot the situation:



The admissible set is shaded. The circular lines around $(2, 3)$ are *contour lines* of the cost function f, i.e. $\{ (x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = c \}$ where $c \in \mathbb{R}$ is constant. (For instance, the circle with the smallest radius is the contour line with $c = 0.25$, meaning that all points on it have the same function value.) Circles with bigger radius correspond to greater function values. Therefore the black dot is the optimal solution.
Now let us compute what we just derived graphically. The Lagrange function is

$$L(x_1, x_2, \lambda_1, \lambda_2) = (x_1 - 2)^2 + (x_2 - 3)^2 - \lambda_1(2 - x_1) - \lambda_2 x_2.$$

Theorem 43 yields:

(1) $\nabla_x L(x_1, x_2, \lambda_1, \lambda_2) = \begin{pmatrix} 2(x_1 - 2) \\ 2(x_2 - 3) + \lambda_1 - \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$

(2) $2 - x_2 \geq 0$

(3) not applicable

(4) $\lambda_1, \lambda_2 \geq 0$

(5) $\lambda_1(2 - x_2) = 0$, $\lambda_2 x_2 = 0$

The first equation in (1) yields $2(x_1 - 2) = 0 \iff x_1 = 2$. From the second equation together with (5) we have three cases:

- (5) is satisfied by $\lambda_1 = \lambda_2 = 0$, so $x_2 = 3$, which is infeasible due to (2).
- (5) is satisfied by $x_2 = 2$ and $\lambda_2 = 0$, so $\lambda_1 = 2$, which is feasible.
- (5) is satisfied by $x_2 = 0$ and $\lambda_1 = 0$, so $\lambda_2 = -6$, which is infeasible due to (4).

So we have one KKT point at $(2, 2)$ with $\lambda_1 = 2$ and $\lambda_2 = 0$.

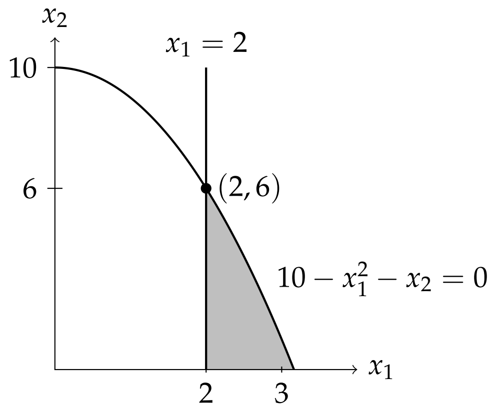
Example. Consider the following (NLP):

$$\begin{aligned} \max\ & x_2 - x_1^2 \\ \text{s.t. } & -(10 - x_1^2 - x_2)^3 \leq 0 \\ & -x_1 \leq -2 \\ & x_1, x_2 \geq 0 \end{aligned} \quad \iff \quad \begin{aligned} -\min\ & x_1^2 - x_2 \\ \text{s.t. } & (10 - x_1^2 - x_2)^3 \geq 0 \\ & x_1 - 2 \geq 0 \\ & x_1 \geq 0 \end{aligned}$$

The Lagrange function is

$$L(x_1, x_1, \lambda_1, \lambda_2, \lambda_3) = x_1^2 - x_2 - \lambda_1(10 - x_1^2 - x_2)^3 - \lambda_2(x_1 - 2) - \lambda_3 x_2.$$

Since $(10 - x_1^2 - x_2)^3 \geq 0$ implies $10 - x_1^2 - x_2 \geq 0$ we know that $10 - x_1^2 \geq 0$ (for x_2 is non-negative) and so $x_1 \leq 4$. Similarly we arrive at $x_2 \leq 10$. Therefore the admissible set is closed and contained in $[0, 4] \times [0, 10]$. Thus, it is compact and global maxima and minima exist. The shaded area in the following picture show the actual admissible set.



Theorem 43 yields:

$$(1)\quad \begin{aligned} 2x_1 + 6\lambda_1 x_1(10 - x_1^2 - x_2)^2 - \lambda_2 &= 0 \qquad (1.1)\\ -1 + 3\lambda_1(10 - x_1^2 - x_2)^2 - \lambda_3 &= 0 \qquad (1.2) \end{aligned}$$

$$(2)\quad \begin{aligned} (10 - x_1^2 - x_2)^3 &\geq 0 \qquad (2.1)\\ x_1 - 2 &\geq 0 \qquad (2.2)\\ x_1, x_2 &\geq 0 \qquad (2.3) \end{aligned}$$

(3) not applicable

(4) $\lambda_1, \lambda_2, \lambda_3 \geq 0$

$$(5)\quad \begin{aligned} \lambda_1(10 - x_1^2 - x_2)^3 &= 0\\ \lambda_2(x_1 - 2) &= 0\\ \lambda_3 x_2 &= 0 \end{aligned}$$

Consider $\bar{x} = \binom{2}{6}$: The third equation of (5) implies $\lambda_3 = 0$, however, plugging into the second equation of (1) leads to the contradiction $-1 = 0$.
Let us go through several cases:

(a) $A(\bar{x}) = \{\,\}$: $\lambda_1 = \lambda_2 = \lambda_3 = 0$, contradicts (1.2).

(b) $A(\bar{x}) = \{3\}$: $\lambda_1 = \lambda_2 = 0$, $x_2 = 0$, using (1) we have $x_1 = 0$, $\lambda_3 = -1$, a contradiction.

(c) $A(\bar{x}) = \{2\}$: $\lambda_1 = \lambda_3 = 0$, $x_1 = 2$, which contradicts (1).

(d) $A(\bar{x}) = \{2, 3\}$: $x_1 = 2$, $x_2 = 0$, $\lambda_1 = 0$, using (1) we have $\lambda_2 = 4$ and $\lambda_3 = -1$, a contradiction.

(e) $A(\bar{x}) = \{1\}$: $\lambda_2 = \lambda_3 = 0$, plugging into the first (1.1) we have $3\lambda_1(10 - x_1^2 - x_2)^2 = 1$. Multiplying this by $2x_1$ and subtracting it from (1.2) of (1) yields $2x_1 + 2x_1 = 0$. But $x_1 = 0$ is infeasible.

(f) $A(\bar{x}) = \{1, 2\}$: $x_1 = 2$, $x_2 = 6$ was already shown to be infeasible.

(g) $A(\bar{x}) = \{1, 3\}$: $x_2 = 0$, $x_1 = \sqrt{10}$, $\lambda_2 = 0$ leads to a contradiction in (1.1).

(h) $A(\bar{x}) = \{1, 2, 3\}$: no solution.

We still have not made any progress. Let us try a direct approach. Suppose (2.1) is active. Then $(10 - x_1^2 - x_2)^3 = 0 \iff x_2 = 10 - x_1^2$. Substituting this into our original problem yields:

$$\begin{aligned} \max\ & 10 - 3x_1^2 =: g(x_1) \\ \text{s.t. } & 10 - x_1^2 \geq 0 \\ & x_1 \geq 2 \end{aligned}$$

The constraints imply $2 \leq x_1 \leq \sqrt{10}$. Differentiating g we have

$$\frac{\mathrm{d}g}{\mathrm{d}x_1}(x_1) = -4x_1 \iff x_1 = 0,$$

which contradicts the constraints $x_1 \geq 2$ and is therefore not feasible. It remains to check the border: $g(2) = 2$, $g(\sqrt{10}) = -10$. Thus there is a global minimum at $(-10, 0)$ and a global maximum at $(2, 6)$.

Chapter 5

Ordinary differential equations

Example. Assume that the rate of change of the gross domestic product (GDP) is proportional to the current GDP. This can be modeled by

$x(t)$	GDP at time t
$\frac{\mathrm{d}}{\mathrm{d}t}x(t) =: \dot{x}(t)$	rate of change at time t
$g \in \mathbb{R}$	proportionality constant

and the equation

$$\dot{x}(t) = g \cdot x(t).$$

The solution is given by $x(t) = c \cdot \mathrm{e}^{gt}$ for each $c \in \mathbb{R}$. (Check: $\dot{x}(t) = cg \cdot \mathrm{e}^{gt} = g \cdot (c \cdot \mathrm{e}^{gt})$.) Additionally there might be some sort of boundary condition, for example $c = x(0)$. Since it pertains the initial time a *boundary condition* at $t = 0$ is called *initial condition*.

In the following we often omit the dependent variable, e.g. $\dot{x} = \dot{x}(t)$.

Example. For the equation $\dot{x} = x + t$

- both $x = -t - 1$ and $x = \mathrm{e}^t - t - 1$ are *particular solutions* on $\mathbb{R}$,
- the *general solution* is given by $x = c \cdot \mathrm{e}^t - t - 1$ for each $c \in \mathbb{R}$,
- $x = \mathrm{e}^t - 1$ is not a solution.

Definition 76. An *ordinary differential equation* (ODE) is an equation $\dot{y} = F(y, t)$ that connects the derivative of an unknown function $y(t)$ and an expression $F(y, t)$. If $F(y, t)$ does not explicitly involve t, i.e. $F(y, t) = \tilde{F}(y)$, it is called *autonomous* or *time-independent* differential equation and *non-autonomous* or *time-dependent* otherwise. An equation which involves derivatives up to and including the i-th derivative is called an i-th order differential equation.

Example. By Hooke's law the force $F(y)$ on a frictionless spring is proportional to the displacement y of the spring from its equilibrium position. Since force equals mass times acceleration we have, for mass 1

$$F(y) = -ay \text{ and } \ddot{y} = -ay,$$

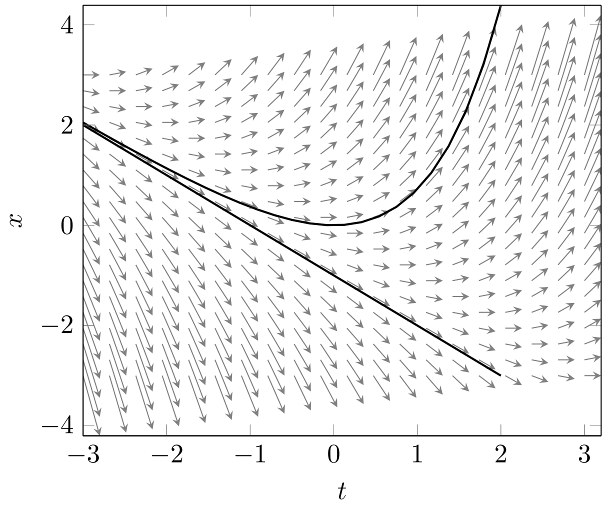


Figure 5.1: Slope field of $\dot{x} = x + t$ in gray and two particular solutions in black.

a second order ODE with general solution $y\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, $t \mapsto k_1 \cos(\sqrt{a} \cdot t) + k_2 \sin(\sqrt{a} \cdot t)$, with free parameters k_1 and k_2. Check:

$$\begin{aligned}\dot{y} &= -k_1 \sin(\sqrt{a} \cdot t)\sqrt{a} + k_2 \cos(\sqrt{a} \cdot t)\sqrt{a} \\ \ddot{y} &= -k_1 \cos(\sqrt{a} \cdot t)a - k_2 \sin(\sqrt{a} \cdot t)a \\ &= -a(k_1 \cos(\sqrt{a} \cdot t) - k_2 \sin(\sqrt{a} \cdot t)) = -ay.\end{aligned}$$

Initial conditions allow us to determine k_1 and k_2:

$$y(0) = c_1 = k_1 \text{ and } \dot{y}(0) = c_2 = k_2\sqrt{a}$$

Definition 77. A solution of $\dot{y} = F(y, t)$ on an interval $I \subseteq \mathbb{R}$ is every differentiable function φ defined on I such that $\dot{\varphi}(t) = F(t, \varphi(t))$ for all $t \in I$.

Remark. The direction in which the graph is going is known, we need to find the graph itself: If $x = x(t)$ is a solution of $\dot{x} = x + t$, then the slope of the tangent to the graph at the point (t, x) is equal to $x + t$, e.g. slope 0 at $(0, 0)$ and slope 3 at $(1, 2)$. We can use this to plot the *slope field* (*direction field*) of the differential equation, see Figure 5.1: for every point (t, x) an arrow indicates the direction in which the graph of the solution is going.

Theorem 44 (Fundamental theorem of differential equations)**.** *Consider the initial value problem*

$$\dot{y} = f(t, y),\ y(t_0) = y_0. \qquad (\star)$$

Suppose that f is a continuous function at (t_0, y_0). There exists a C^1 function $y\colon I \to \mathbb{R}$ defined on an open interval $I = (\,t_0 - a, t_0 + a\,)$ about t_0 such that $y(t_0) = y_0$ and $\dot{y}(t) = f(t, y(t))$ for all $t \in I$. If f is C^1 at (t_0, y_0), then the solution $y(t)$ is unique, i.e. any two solution of $(\star)$ must be equal to each other on the intersection of their domains.

Definition 78. A *separable equation* is a first order ODE in which the expression for $\dot{y} = \frac{\mathrm{d}y}{\mathrm{d}t}$ can be factorised into a function of t times a function of y:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t) \cdot g(y).$$

To solve a separable equation we rewrite it as

$$\int \frac{1}{g(y)}\,\mathrm{d}y = \int f(t)\,\mathrm{d}t$$

which is solved by $G(y) = F(t) + C$, where C is a constant.
If initial values $y(t_0) = y_0$ are given, we have:

$$\int_{y_0}^{y} \frac{1}{g(\tilde{y})}\,\mathrm{d}\tilde{y} = \int_{t_0}^{t} f(\tilde{t})\,\mathrm{d}\tilde{t}$$

Example. Let $\frac{\mathrm{d}y}{\mathrm{d}x} = 6y^2x$ and $y(1) = 1/25$. So

$$\int_{1/25}^{y} \tilde{y}^{-2}\,\mathrm{d}\tilde{y} = \int_{1}^{x} 6\tilde{x}\,\mathrm{d}\tilde{x} \iff \left[-\frac{1}{\tilde{y}}\right]_{1/25}^{y} = \left[3\tilde{x}^2\right]_1^x$$
$$\iff 25 - \frac{1}{y} = 3x^2 - 3 \iff y(x) = \frac{1}{28 - 3x^2}$$

Since $28 - 3x^2 = 0$ yields $x = \pm\sqrt{28/3}$ we have $I \subseteq (-\sqrt{28/3}, \sqrt{28/3})$ with $1 \in I$. ($y \equiv 0$ violates the initial condition.)

Remark. An autonomous linear first order ODE

$$\dot{y}(t) = a \cdot y(t) + b$$

with $b \in \mathbb{R}$, $a \in \mathbb{R} \setminus \{0\}$ has the general solution

$$y(t) = -b/a + c \cdot \mathrm{e}^{at}$$

($c \in \mathbb{R}$, constant).
A non-autonomous homogeneous linear first order ODE

$$\dot{y}(t) = a(t) \cdot y(t) + b$$

is separable (choose $F(t) = \int_{t_0}^{t} a(\tilde{t})\,\mathrm{d}\tilde{t}$) has the general solution

$$y(t) = c \cdot \mathrm{e}^{F(t)}$$

($c \in \mathbb{R}$, constant).
A non-autonomous inhomogeneous linear first order ODE

$$\dot{y}(t) = a(t) \cdot y(t) + b(t) \tag{5.1}$$

has the general solution

$$y(t) = \mathrm{e}^{F(t)} \left(\int_{t_0}^{t} b(\tilde{t}) \cdot \mathrm{e}^{-F(\tilde{t})}\mathrm{d}\tilde{t} + c \right) \tag{5.2}$$

($c \in \mathbb{R}$, constant).

Example. Let $\dot{y} + y/t = t^3$, $t_0 = 1$. So $a(t) = -1/t$ and $F(t) = -\ln(t) - \ln(1) = -\ln(t)$ (assuming $t > 0$). $b(t) = t^3$, $e^{F(t)} = 1/t$, $e^{-F(t)} = t$. Therefore

$$\int_{t_0}^{t} \tilde{t}^3 \cdot \tilde{t} \, d\tilde{t} = \int_{1}^{t} \tilde{t}^4 \, d\tilde{t} = \left[\frac{\tilde{t}^5}{5} \right|_1^t = \frac{t^5}{5} - \frac{1}{5}.$$

Finally $y(t) = (t^5/5 - 1/5 + c)/t = k/t + t^4/5$ with $k := -1/5 + c$, $k \in \mathbb{R}$, constant.

Observation. The general solution of (5.1) in addition to a particular solution of (5.2) gives the general solution of (5.2).