University of Basel
Faculty of Business and Economics

Topics in advanced mathematics
Fundamentals and differentiable functions
Dr. Thomas Zehrt

Contents

1 $\mathbb{R}^n$, vectors and matrices

1.1 Vectors

- n-dimensional space $\mathbb{R}^n$
- elements $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are called n-vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T \quad \text{and } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- scalar product and norm:

$$\begin{aligned} \mathbf{x} \bullet \mathbf{y} &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \\ ||\mathbf{x}|| &= \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \end{aligned}$$

$$\boxed{\mathbf{x} \bullet \mathbf{y} = ||\mathbf{x}|| \cdot ||\mathbf{y}|| \cdot \cos \angle(\mathbf{x}, \mathbf{y})}$$

- $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k} \in \mathbb{R}^n$
  - If $a_1, a_2, \ldots, a_k \in \mathbb{R}$, then $\mathbf{z} = a_1\mathbf{x_1} + a_2\mathbf{x_2} + \ldots + a_k\mathbf{x_k}$ is called a linear combination of $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}$.
  - $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}$ are called linearly dependent, if there exist $b_1, b_2, \ldots, b_k \in \mathbb{R}$ such that $b_1\mathbf{x_1} + b_2\mathbf{x_2} + \ldots + b_k\mathbf{x_k} = \mathbf{0}$ and not all $b_j = 0$.
  - $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}$ are called linearly independent, if a linear combination of the zero vector

    $$b_1\mathbf{x_1} + b_2\mathbf{x_2} + \ldots + b_k\mathbf{x_k} = \mathbf{0}$$

    is possible only with $b_1 = b_2 = \ldots = b_k = 0$.

1.2 Matrices

$\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in \mathbb{R}^n$

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix}, \ldots, \mathbf{a}_m = \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} \quad \to \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

is called an $n \times m$ matrix.

- The inverse matrix $\mathbf{A^{-1}}$ of the $n \times n$ matrix $\mathbf{A} = (\mathbf{a_{ij}})$ is defined by

$$\mathbf{A^{-1}} \cdot \mathbf{A} \;=\; \mathbf{A} \cdot \mathbf{A^{-1}} \;=\; \mathbf{I_n} \;=\; \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

- For the $n \times n$ matrix $\mathbf{A}$ let $\mathbf{A_{ij}}$ denote the submatrix of $\mathbf{A}$ generated by cancelling the i-th row and the j-th column of $\mathbf{A}$. Then the determinant $\det \mathbf{A}$ is given (recursively) by

$$\boxed{\det \mathbf{A} = |\mathbf{A}| = \mathbf{a_{11}} \det \mathbf{A_{11}} - \mathbf{a_{12}} \det \mathbf{A_{12}} + \cdots + (\mathbf{-1})^{\mathbf{n+1}} \mathbf{a_{1n}} \det \mathbf{A_{1n}}}$$

Example 1

$$\begin{vmatrix} 1 & 1 & 3 & 3 \\ 1 & 2 & 1 & 2 \\ 1 & -2 & 1 & -2 \\ 0 & 1 & -2 & -1 \end{vmatrix}$$

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

1.3 Matrix calculus

1a.	$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$	1b.	$\mathbf{AB} \neq \mathbf{BA}$
2a.	$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$	2b.	$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
3a.	$\mathbf{A} + \mathbf{0} = \mathbf{A}$	3b.	$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$, ($\mathbf{A}$ square)

4. $\mathbf{AB} = \mathbf{0} \quad \not\Rightarrow \quad \mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$
5. $\mathbf{AB} = \mathbf{AC} \quad \not\Rightarrow \quad \mathbf{B} = \mathbf{C}$

6. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B} \quad \lambda \in \mathbb{R}$
7. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
8. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

9. $(\mathbf{A^{-1}})^{\mathbf{-1}} = \mathbf{A}$
10. $(\mathbf{AB})^{\mathbf{-1}} = \mathbf{B^{-1}A^{-1}}$
11. $(\mathbf{A^T})^{\mathbf{T}} = \mathbf{A}$
12. $(\mathbf{A} + \mathbf{B})^{\mathbf{T}} = \mathbf{A^T} + \mathbf{B^T}$
13. $(\mathbf{AB})^{\mathbf{T}} = \mathbf{B^T A^T}$
14. $(\mathbf{A^{-1}})^{\mathbf{T}} = (\mathbf{A^T})^{\mathbf{-1}}$

For $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc \neq 0$ is $\mathbf{A^{-1}} = \dfrac{\mathbf{1}}{\mathbf{ad - bc}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

1.4 Eigenvalues and eigenvectors

Definition 1.1 *If* $\mathbf{A}$ *is an* $n \times n$ *matrix, then a real number* λ *is an* *eigenvalue* *of* $\mathbf{A}$ *if there is a nonzero vector* $\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$ *such that*

$$\boxed{\mathbf{A}\mathbf{x} = \lambda\mathbf{x}}$$

Then $\mathbf{x}$ *is an* *eigenvector* *of* $\mathbf{A}$ *(associated with* λ*).*

Remark: If $\mathbf{x}$ is an eigenvector associated with the eigenvalue λ, then so is $\alpha\mathbf{x}$ for every real number $\alpha \neq 0$.

$$\mathbf{A}\,(\alpha\,\mathbf{x}) \;=\; \alpha\,\mathbf{A}\,\mathbf{x} = \alpha\,(\lambda\,\mathbf{x}) = \lambda\,(\alpha\,\mathbf{x})$$

How to find eigenvalues? The equation can be written as

$$\begin{array}{rrcl} & \mathbf{A}\,\mathbf{x} & = & \lambda\,\mathbf{x} \\ \Leftrightarrow & \mathbf{A}\,\mathbf{x} - \lambda\,\mathbf{I}\,\mathbf{x} & = & \mathbf{0} \\ \Leftrightarrow & (\mathbf{A} - \lambda\,\mathbf{I})\,\mathbf{x} & = & \mathbf{0} \end{array}$$

This is a homogeneous linear system of equations. It has a solution $\mathbf{x} \neq \mathbf{0}$ if and only if the coefficient matrix $(\mathbf{A} - \lambda\,\mathbf{I})$ is singular which means that it has determinant equal to 0.

$$\boxed{(\mathbf{A} - \lambda\,\mathbf{I}) \text{ singular} \quad \Leftrightarrow \quad \underbrace{\det(\mathbf{A} - \lambda\,\mathbf{I})}_{p_A(\lambda)} = 0}$$

$p_A(\lambda) = 0$ is called characteristic equation of $\mathbf{A}$. The function $p_A(\lambda)$ is a polynomial of degree n in λ, called the characteristic polynomial of $\mathbf{A}$.

1.5 Diagonalization

Let $\mathbf{A}$ and $\mathbf{P}$ be $n \times n$ matrices with $\mathbf{P}$ invertible. Then $\mathbf{A}$ and $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ have the same eigenvalues (because they have the same characteristic polynomial).

Definition 1.2 *An* $n \times n$ *matrix* $\mathbf{A}$ *is* *diagonalizable* *if there is an invertible matrix* $\mathbf{P}$ *and a diagonal matrix* $\mathbf{D}$ *such that*

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} \;=\; \mathbf{D}.$$

Two natural questions:

1. Which square matrices are diagonalizable?
2. If $\mathbf{A}$ is diagonalizable, how do we find the matrix $\mathbf{P}$?

Theorem 1.1 *An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if it has a set of n linearly independent eigenvectors $\mathbf{x_1}, \ldots, \mathbf{x_n}$. In this case,*

$$\mathbf{P^{-1}AP} \;=\; diag(\lambda_1, \ldots, \lambda_n),$$

where $\mathbf{P}$ is the matrix with $\mathbf{x_1}, \ldots, \mathbf{x_n}$ as its columns, and $\lambda_1, \ldots, \lambda_n$ are the corresponding eigenvalues.

Many of the matrices encountered in economics are symmetric and for these matrices we have the following important result.

Theorem 1.2 (Spectral Theorem for symmetric matrices) *If the $n \times n$ matrix $\mathbf{A}$ is symmetric ($\mathbf{A} = \mathbf{A^T}$), then:*

1. *All n eigenvalues $\lambda_1, \ldots, \lambda_n$ are real.*
2. *Eigenvectors that correspond to different eigenvalues are orthogonal.*
3. *There exists an orthogonal matrix $\mathbf{P}$ ($\mathbf{P^{-1}} = \mathbf{P^T}$) such that*

$$\mathbf{P^{-1}AP} \;=\; diag(\lambda_1, \ldots, \lambda_n).$$

The columns $\mathbf{x_1}, \ldots, \mathbf{x_n}$ of the matrix $\mathbf{P}$ are eigenvectors of unit length corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$.

2 Subsets of $\mathbb{R}^n$

2.1 ϵ-balls and line segments

Definition 2.1 *Let $\mathbf{x} \in \mathbb{R}^n$ and $\epsilon > 0$ a real number. The (open) ϵ-ball with center $\mathbf{x}$ is the set*

$$B_\epsilon(\mathbf{x}) \;=\; \{\, \mathbf{y} \in \mathbb{R}^n \;\mid\; ||\mathbf{x} - \mathbf{y}|| < \epsilon \,\}$$

Definition 2.2 *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The (closed) line segment $\overline{\mathbf{xy}}$ is the set*

$$\overline{\mathbf{xy}} = \{\, t \cdot \mathbf{x} + (1-t) \cdot \mathbf{y} \;\mid\; t \in [0,1] \,\}$$

This is a part of a (straight) line:

$$t \cdot \mathbf{x} + (1-t) \cdot \mathbf{y} \;=\; \mathbf{y} + t \cdot (\mathbf{x} - \mathbf{y}) \;=\; \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} + t \cdot \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix}$$

2.2 Convex subsets

Definition 2.3 *A subset $M \subset \mathbb{R}^n$ is called* <u>*convex*</u>*, if for all* $\mathbf{x}, \mathbf{y} \in M$ *we have* $\overline{\mathbf{xy}} \subset M$ *.*

Theorem 2.1 *Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \qquad \textit{und} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Then the set

$$M = \{ \mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} \leq \mathbf{b} \textit{ and } \mathbf{x} \geq \mathbf{0} \}$$

is convex.

2.3 Bounded subsets, boundary and closed subsets

Let $M \subset \mathbb{R}^n$.

Definition 2.4 *M is called* <u>*bounded*</u>*, if there exists a real positive number ϵ such that $M \subset B_\epsilon(0)$.*

A point $\mathbf{x}$ is called <u>*boundary point*</u> *of M, if* **each** *ϵ-ball with center $\mathbf{x}$ contains both points of M and points of the complement $M^c = \mathbb{R}^n - M$.*

A point in M which is not a boundary point of M is called an <u>*inner point*</u> *of M.*

The <u>*boundary*</u> *∂M of M is*

$$\partial M = \{ \mathbf{x} \in \mathbb{R}^n \mid \textit{for all } \epsilon > 0 \textit{ is } B_\epsilon(\mathbf{x}) \cap M \neq \emptyset \textit{ and } B_\epsilon(\mathbf{x}) \cap M^c \neq \emptyset \}$$

M is called <u>*closed*</u> *if $\partial M \subset M$.*

3 Functions

3.1 Definition

Definition 3.1 *Let $D \subset \mathbb{R}^n$. A function from D to $\mathbb{R}$ is an assignment of exatly one element of $\mathbb{R}$ to each element of D. We write:*

$$\begin{array}{rcccl} & f : D & \longrightarrow & \mathbb{R} & \\ \mathbf{x} = & (x_1, x_2, \ldots, x_n)^T & \longmapsto & f(x_1, x_2, \ldots, x_n) = f(\mathbf{x}) = y \end{array}$$

D is called domain and $W = \{y \in \mathbb{R} \mid f(\mathbf{x}) = y \text{ for an } \mathbf{x} \in D\}$ the range of f.

The assignment of $f(\mathbf{x})$ to $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ can be given by

1. an explicit calculation $f(x_1, x_2) = 3x_1x_2^2 + x_1$ with $D = \mathbb{R}^2$,
2. an implicit equation, for example let $f(x_1, x_2)$ be the solution x_3 of the equation $x_3^3 - x_1x_3 - x_2 = 0$ with $x_1 \in \mathbb{R}$ and $x_2 > 0$,
3. by a differential equation.

3.2 Graph and level sets

Definition 3.2 *Let $D \subset \mathbb{R}^n$ and $f : D \to \mathbb{R}$ be a function. The set*

$$\boxed{G_f = \{ (\mathbf{x}, f(\mathbf{x}))^T \mid \mathbf{x} \in D \} \subset \mathbb{R}^{n+1}}$$

is called graph of f.

Example 2 $f : \mathbb{R}^2 \to \mathbb{R}$, $f(\mathbf{x}) = f(x_1, x_2) = x_1^2 - x_2^2$.

$$G_f = \{ (x_1, x_2, x_1^2 - x_2^2)^T \mid (x_1, x_2)^T \in \mathbb{R}^2 \} \subset \mathbb{R}^3$$

Definition 3.3 *Let $D \subset \mathbb{R}^n$, $f : D \to \mathbb{R}$ a function and $c \in \mathbb{R}$. The set*

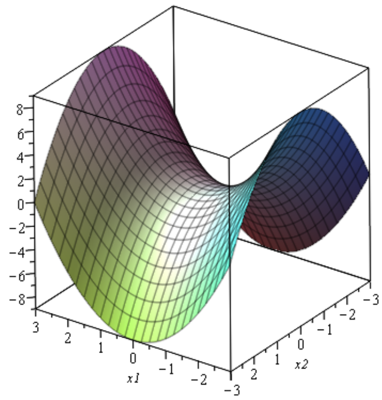
$$\boxed{N_c = \{ \mathbf{x} \in D \mid f(\mathbf{x}) = c \} \subset D \subset \mathbb{R}^n}$$

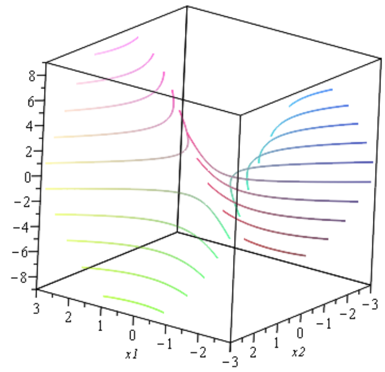
is called level set.

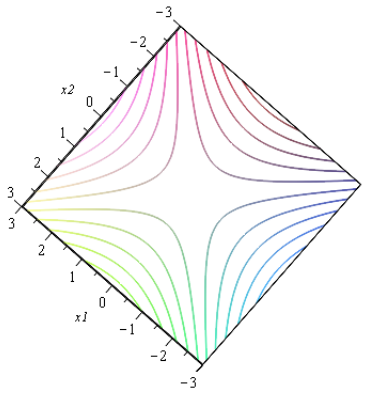
Example 3 $f(\mathbf{x}) = f(x_1, x_2) = x_1^2 - x_2^2$

$$N_c = \{ x_1^2 - x_2^2 = c \mid (x_1, x_2)^T \in \mathbb{R}^2 \}$$

are hyperbolas.







3.3 Continous functions

Definition 3.4 *A function $f : D \to \mathbb{R}$ is called continuous in $\mathbf{a} \in D$, if for each sequence $(\mathbf{x}_k)$ in D with $\lim\limits_{k\to\infty} \mathbf{x}_k = \mathbf{a}$ we have $\lim\limits_{k\to\infty} f(\mathbf{x}_k) = f(\mathbf{a})$.*

Theorem 3.1 (Theorem of Weierstraß) *Let $D \subset \mathbb{R}^n$ be a closed and bounded set and $f : D \to \mathbb{R}$ continuous. Then there exist $\mathbf{x}_{min}, \mathbf{x}_{max} \in D$ such that*

$$f(\mathbf{x}_{min}) \ \leq \ f(\mathbf{x}) \ \leq \ f(\mathbf{x}_{max})$$

for all $\mathbf{x} \in D$.

3.4 Important functions

Definition 3.5 *A linear function is given by*

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2, \ldots, x_n)^T &\longmapsto a_1x_1 + a_2x_2 + \ldots + a_nx_n \ = \ \mathbf{a} \bullet \mathbf{x} \ = \ f(\mathbf{x}) \end{aligned}$$

for real numbers $a_1, a_2, \ldots, a_n$ or $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$.

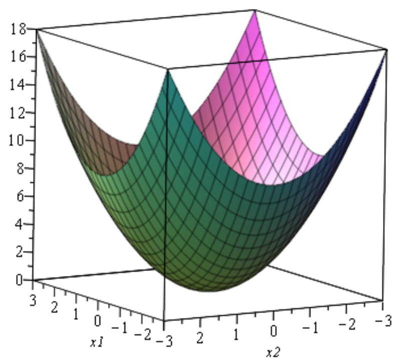
Definition 3.6 *Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. The function*

$$\begin{aligned} Q_A : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2, \ldots, x_n)^T &\longmapsto \mathbf{x} \bullet A\mathbf{x} \ = \ \mathbf{x}^T A\mathbf{x} \ = \ \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\, x_i\, x_j \end{aligned}$$

is called quadratic form corresponding to $\mathbf{A}$.

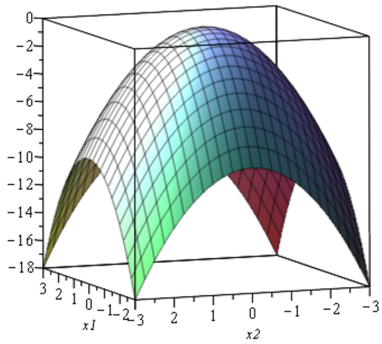
Example 4

$$Q_A(\mathbf{x}) \ = \ (x_1, x_2) \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{=\mathbf{A}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ = \ x_1^2 + x_2^2.$$



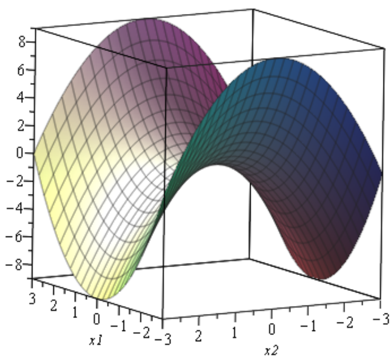
Example 5

$$Q_B(\mathbf{x}) \quad = \quad (x_1, x_2)\underbrace{\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}}_{=\mathbf{B}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \; = \; -x_1^2 - x_2^2.$$



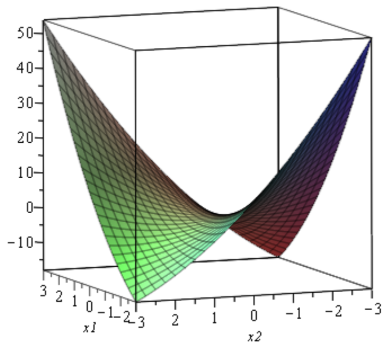
Example 6

$$Q_C(\mathbf{x}) \quad = \quad (x_1, x_2)\underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{=\mathbf{C}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \; = \; x_1^2 - x_2^2.$$



Example 7

$$Q_D(\mathbf{x}) \quad = \quad (x_1, x_2)\underbrace{\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}}_{=\mathbf{D}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad = \quad x_1^2 + 4x_1x_2 + x_2^2.$$



Definition 3.7 *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then* $\mathbf{A}$ *is called*

- *positive definite, if* $Q_A(\mathbf{x}) > 0$*;*
- *positive semidefinite, if* $Q_A(\mathbf{x}) \geq 0$*;*
- *negative definite, if* $Q_A(\mathbf{x}) < 0$*;*
- *negative semidefinite, if* $Q_A(\mathbf{x}) \leq 0$*;*

for all $\mathbf{x} \neq \mathbf{0}$.

$\mathbf{A}$ *is called indefinite, if there exist vectors* $\mathbf{x}$ *with* $Q_A(\mathbf{x}) > 0$ *as well as vectors* $\mathbf{y}$ *with* $Q_A(\mathbf{y}) < 0$.

Definition 3.8 *Let* $D \subset \mathbb{R}^n$ *be a convex set. A function* $f : D \to \mathbb{R}$ *is called*

- *(strongly) concave on* D*, if*

  $$f((1-t)\mathbf{a} + t\mathbf{b}) \quad (>) \;\geq \quad (1-t)f(\mathbf{a}) + tf(\mathbf{b})$$

  for all $\mathbf{a}, \mathbf{b} \in D$ *and all* $t \in (0,1)$*;*
- *(strongly) convex on* D*, if*

  $$f((1-t)\mathbf{a} + t\mathbf{b}) \quad (<) \;\leq \quad (1-t)f(\mathbf{a}) + tf(\mathbf{b})$$

  for all $\mathbf{a}, \mathbf{b} \in D$ *and all* $t \in (0,1)$.

4 Differentiable functions

4.1 Partial derivations

Definition 4.1 *Let $y = f(\mathbf{x}) = f(x_1, \ldots, x_i, \ldots, x_n)$ be a function. For $i = 1, 2, \ldots, n$ the i-th partial derivation of f is defined by*

$$\boxed{\frac{\partial f}{\partial x_i}(\mathbf{x}) = f_{x_i}(\mathbf{x}) \quad = \quad \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e_i}) - f(\mathbf{x})}{t}}$$

Example 8

$$\begin{aligned}
f(x_1, x_2) &= x_1^2 + x_1x_2 + 2x_2^2 + 3 \\
\frac{\partial f}{\partial x_1} &= \lim_{t \to 0} \frac{(x_1 + t)^2 + (x_1 + t)x_2 + 2x_2^2 + 3 - (x_1^2 + x_1x_2 + 2x_2^2 + 3)}{t} \\
&= \lim_{t \to 0} \frac{2x_1t + t^2 + tx_2}{t} \;=\; \lim_{t \to 0} (2x_1 + t + x_2) \;=\; 2x_1 + x_2
\end{aligned}$$

Definition 4.2 *The function f is called 2-times (k)-times partially differentiable, if all partial derivations of second order*

$$f_{x_ix_j} \;=\; (f_{x_i})_{x_j} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \qquad (1 \leq i, j \leq n)$$

exist.

The following fact is sometimes important:

Theorem 4.1 *If all partial derivations of second order exist and are continuous functions, then $f_{x_ix_j} = f_{x_jx_i}$.*

Definition 4.3 *Let $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in D \subset \mathbb{R}^n$ be a point in the domain of f. The vector*

$$\mathbf{grad}\; f(\mathbf{a}) \;=\; \begin{pmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{pmatrix}$$

is called gradient of f in $\mathbf{a}$. The matrix

$$H_f(\mathbf{a}) \;=\; \begin{pmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & \ldots & f_{x_1x_n}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & \ldots & f_{x_2x_n}(\mathbf{a}) \\ \vdots & \vdots & \vdots & \vdots \\ f_{x_nx_1}(\mathbf{a}) & f_{x_nx_2}(\mathbf{a}) & \ldots & f_{x_nx_n}(\mathbf{a}) \end{pmatrix}$$

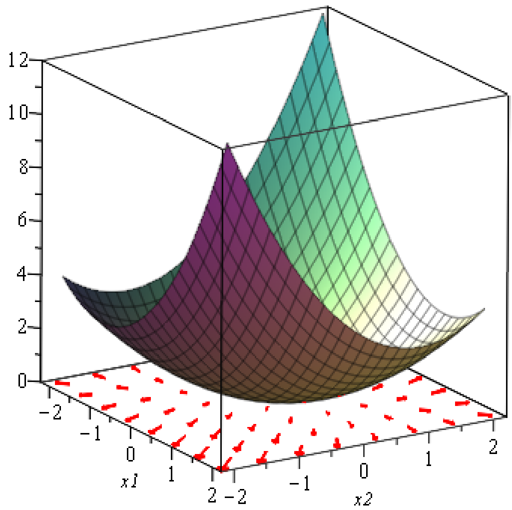
is called Hesse matrix of f in $\mathbf{a}$.

Theorem 4.2 (Properties of the gradient)

- *The gradient of f in $\mathbf{a}$ is orthogonal to the level set $f(\mathbf{x}) = f(\mathbf{a})$.*
- *The gradient of f in $\mathbf{a}$ points in the direction of the greatest rate of increase of the function f in $\mathbf{a}$.*

Example 9 *For $f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2$ we have*

$$\mathbf{grad}\, f(x_1, x_2) \quad = \quad \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 \end{pmatrix}.$$



4.2 Differential and differentiable functions

Definition 4.4 *The (total) differential df of f is defined by*

$$\boxed{df \;=\; df(\mathbf{x}, d\mathbf{x}) \;=\; f_{x_1}(\mathbf{x}) \cdot dx_1 + \ldots + f_{x_n}(\mathbf{x}) \cdot dx_n}$$

Example 10 *Let* $f(x_1, x_2, x_3) = \sin(x_1 x_2) + x_3^2$. *Then*

$$\begin{aligned} df(x_1, x_2, x_3, dx_1, dx_2, dx_3) &= x_2 \cos(x_1 x_2)\, dx_1 + x_1 \cos(x_1 x_2)\, dx_2 + 2x_3\, dx_3 \\ df(x_1, x_2, x_3, dx_1, dx_2, dx_3) &= x_2 \cos(x_1 x_2)\, dx_1 + x_1 \cos(x_1 x_2)\, dx_2 + 2x_3\, dx_3 \end{aligned}$$

Definition 4.5 *Let $D \subset \mathbb{R}^n$ be an open set. A function $f : D \to \mathbb{R}$ is called differentiable in $\mathbf{a} \in D$, if*

$$\underbrace{f(\mathbf{x}) \;=\; f(\mathbf{a}) + \mathbf{grad}\ f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) + R(\mathbf{x}, \mathbf{a})}_{*} \quad \textit{and} \quad \underbrace{\lim_{\mathbf{x} \to \mathbf{a}} \frac{R(\mathbf{x}, \mathbf{a})}{||\mathbf{x} - \mathbf{a}||} \;=\; 0}_{\star}$$

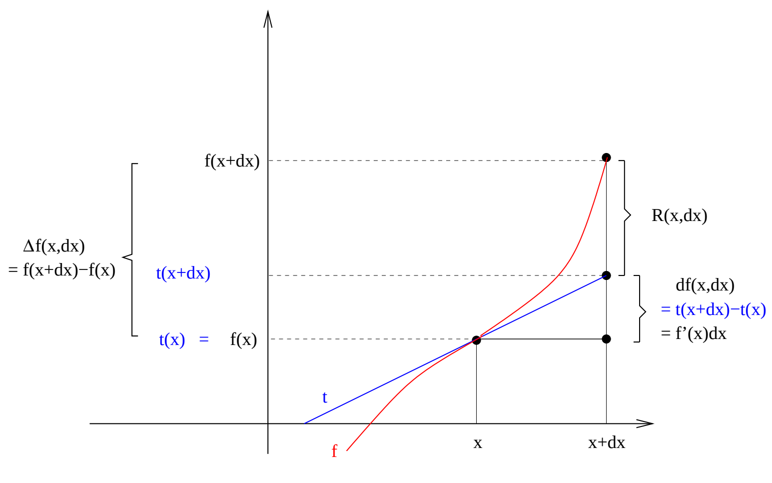
- The function $t(\mathbf{x}) = f(\mathbf{a}) + \mathbf{grad}\ f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a})$ is called tangent hyperplane of f in $\mathbf{a}$:

$$t(\mathbf{x}) \;=\; f(\mathbf{a}) + \mathbf{grad}\ f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) \;=\; f(\mathbf{a}) + df(\mathbf{a}, d\mathbf{x})$$

- A differentiable function can be approximated (very well) by a linear function and the claim $\star$ is essential.

- If we use the notation $\Delta f(\mathbf{a}, d\mathbf{x}) = f(\mathbf{a} + d\mathbf{x}) - f(\mathbf{a})$ for the real change of f and $\mathbf{x} = \mathbf{a} + d\mathbf{x}$ we get

$$\Delta f(\mathbf{a}, d\mathbf{x}) \;=\; df(\mathbf{a}, d\mathbf{x}) + R(\mathbf{x}, \mathbf{a})$$



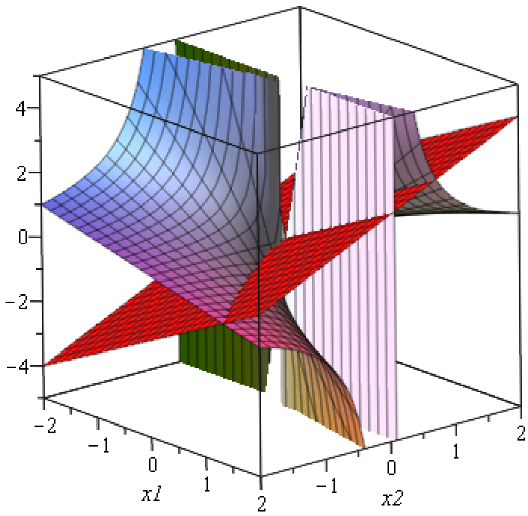
$\Delta x = dx$	change of x.
$\Delta f = \Delta f(x, dx)$	change of f, if we change x by dx.
$df = df(x, dx)$	change of the linear approximation, if we change x by dx.

Example 11 $z = f(x_1, x_2) = \dfrac{x_1}{x_2}$, $\mathbf{a} = (1, 1)$

$$f_{x_1} = \frac{1}{x_2} \quad f_{x_1}(1,1) = 1 \qquad f_{x_2} = -\frac{x_1}{x_2^2} \quad f_{x_2}(1,1) = -1$$

$$\begin{aligned} f(x_1, x_2) = \frac{x_1}{x_2} &= f(1,1) + \begin{pmatrix} f_{x_1}(1,1) \\ f_{x_2}(1,1) \end{pmatrix} \bullet \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} + R(x_1, x_2, 1, 1) \\ &= 1 + (x_1 - 1) - (x_2 - 1) + R(x_1, x_2, 1, 1) \\ &= 1 + x_1 - x_2 + R(x_1, x_2, 1, 1) \end{aligned}$$

Graph of f and of the tangent plane (red)



We should show:

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R(\mathbf{x}, \mathbf{a})}{||\mathbf{x} - \mathbf{a}||} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\dfrac{x_1}{x_2} - 1 - x_1 + x_2}{\sqrt{(x_1 - 1)^2 + (x_2 - 1)^2}} = 0$$

The tangent plane is

$$\begin{aligned} t(x_1, x_2) &= f(\mathbf{a}) + f_{x_1}(\mathbf{a}) \cdot (x_1 - a_1) + f_{x_2}(\mathbf{a}) \cdot (x_2 - a_2) \\ &= 1 + 1 \cdot (x_1 - 1) + (-1) \cdot (x_2 - 1) \\ &= 1 + x_1 - x_2 \end{aligned}$$

4.3 The directional derivation

Definition 4.6 *Let* $\mathbf{v} \in \mathbb{R}^n$ *be a vector of length* 1 *(unit vector). The limit (if it exists)*

$$\boxed{\partial_{\mathbf{v}}\, f(\mathbf{a}) \;=\; \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t}}$$

is called the directional derivation of f *in* $\mathbf{a}$ *in direction* $\mathbf{v}$*.*

Theorem 4.3 *Let* D *be open,* f *differentiable on* D *and* $\mathbf{v} \in \mathbb{R}^n$ *with* $||\mathbf{v}|| = 1$*. Then*

$$\boxed{\partial_{\mathbf{v}}\, f(\mathbf{a}) \;=\; \mathbf{grad}\, f(\mathbf{a}) \bullet \mathbf{v} \;=\; \sum_{i=1}^{n} f_{x_i}(\mathbf{a})\, v_i}$$

Proof: Let f be totally differentiable in $\mathbf{a}$, then

$$f(\mathbf{x}) \;=\; f(\mathbf{a}) + \mathbf{grad}\, f(\mathbf{a}) \bullet (\mathbf{x}-\mathbf{a}) + R(\mathbf{x},\mathbf{a}) \quad \text{und} \quad \lim_{\mathbf{x}\to\mathbf{a}} \frac{R(\mathbf{x},\mathbf{a})}{||\mathbf{x}-\mathbf{a}||} \;=\; 0$$

With $\mathbf{x} = \mathbf{a} + t\mathbf{v}$ we get:

$$f(\mathbf{x}) - f(\mathbf{a}) \;=\; f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a}) \;=\; \mathbf{grad}\, f(\mathbf{a}) \bullet t\mathbf{v} + R(\mathbf{x},\mathbf{a}).$$

Hence:

$$\begin{aligned}
\partial_{\mathbf{v}} f(\mathbf{a}) &= \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t} \\
&= \lim_{t\to 0} \frac{\mathbf{grad}\, f(\mathbf{a}) \bullet t\mathbf{v} + R(\mathbf{x},\mathbf{a})}{t} \\
&= \mathbf{grad}\, f(\mathbf{a}) \bullet \mathbf{v} + \lim_{t\to 0} \frac{R(\mathbf{x},\mathbf{a})}{t} \\
&= \mathbf{grad}\, f(\mathbf{a}) \bullet \mathbf{v}.
\end{aligned}$$

$\square$

4.4 The chain rule

Theorem 4.4 *Let* $D \subset \mathbb{R}^n$ *be open and* $f : D \to \mathbb{R}$ *continously partially differentiable.* $I \subset \mathbb{R}$ *and*

$$\mathbf{x} : I \to D \subset \mathbb{R}^n \quad \textit{with} \quad \mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

with differentiable coordinat functions $x_i(t)$ *für* $1 \leq i \leq n$*. Then the composition* $f \circ \mathbf{x} : I \to \mathbb{R}$ *mit* $f \circ \mathbf{x}(t) = f(\mathbf{x}(t))$ *is differentiable with*

$$\boxed{\frac{d}{dt}\, f(\mathbf{x}(t)) \quad = \mathbf{grad}\, f(\mathbf{x}(t)) \bullet \frac{d}{dt}\, \mathbf{x}(t)}$$

Expansion:

$$\begin{aligned}
& \frac{d}{dt}\, f(\mathbf{x}(t)) \\
= \; & \mathbf{grad}\, f(\mathbf{x}(t)) \bullet \frac{d}{dt}\, \mathbf{x}(t) \\
= \; & \frac{d}{dt}\, f(x_1(t), x_2(t), \ldots, x_n(t)) \\
= \; & f_{x_1}(\mathbf{x}(t))\, \frac{d}{dt} x_1(t) + f_{x_2}(\mathbf{x}(t))\, \frac{d}{dt} x_2(t) + \ldots + f_{x_n}(\mathbf{x}(t))\, \frac{d}{dt} x_n(t) \\
= \; & f_{x_1}(\mathbf{x}(t))\, \dot{x}_1(t) + f_{x_2}(\mathbf{x}(t))\, \dot{x}_2(t) + \ldots + f_{x_n}(\mathbf{x}(t))\, \dot{x}_n(t)
\end{aligned}$$

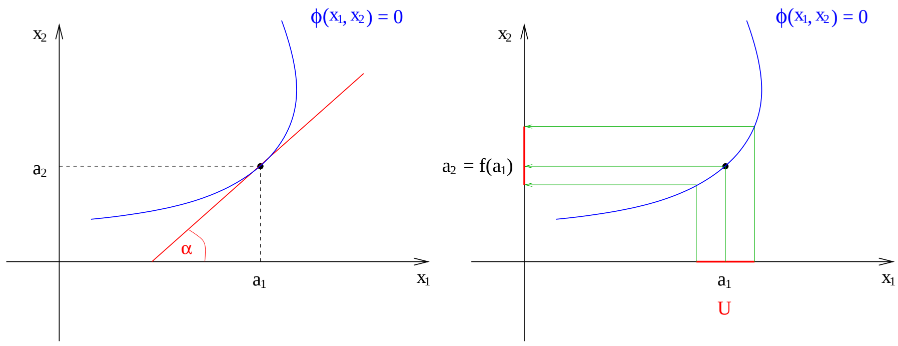
4.5 Implicite Derivation

Notation: $(\mathbf{x}, y) = (x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1}$

Theorem 4.5 *Let $M \subset \mathbb{R}^{n+1}$ be open, $\phi : M \to \mathbb{R}$ continuously partially differentiable and $\mathbf{a} = (a_1, \ldots, a_n, a_{n+1}) \in M$ with $\phi(\mathbf{a}) = 0$ and $\phi_y(\mathbf{a}) \neq 0$. Then there is a neighbourhood U of $(a_1, \ldots, a_n)$ and an open interval $I \subset \mathbb{R}$ with $a_{n+1} \in I$ such that:*

1. *$R := \{\ (\mathbf{x}, y) \subset \mathbb{R}^{n+1} \mid \mathbf{x} \subset U$ and $y \in I\ \} \ \subset\ M$ and $\phi_y(\mathbf{x}) \neq 0$ for all $(\mathbf{x}, y) \in R$.*
2. *For each $\mathbf{x} \in U$ there exists exactly one $y \in I$ with $\phi(\mathbf{x}, y) = 0$. The function $y := f(\mathbf{x})$ is partially differentiable ($f : U \to I$) and*

$$\boxed{\phi(\mathbf{x}, y) \;=\; \phi(\mathbf{x}, f(\mathbf{x})) \;=\; 0 \quad \longrightarrow \quad \frac{\partial}{\partial x_i}\, f(\mathbf{x}) \;=\; -\frac{\frac{\partial}{\partial x_i}\,\phi(\mathbf{x}, y)}{\frac{\partial}{\partial y}\,\phi(\mathbf{x}, y)}}$$



Let $y := f(\mathbf{x})$ for all $\mathbf{x} \in U$ the function above. Then

$$\phi(\mathbf{x}, y) \;=\; \phi(\mathbf{x}, f(\mathbf{x})) \;=\; 0$$

By the chain rule we get:

$$\begin{aligned}
0 \;=\; \frac{\partial}{\partial x_i}\, 0 \;&=\; \frac{\partial}{\partial x_i}\, \phi(\overbrace{x_1, \ldots, x_n}^{\mathbf{x}}, \overbrace{f(x_1, \ldots, x_n)}^{y}) \\
&=\; \sum_{j=1}^{n} \frac{\partial}{\partial x_j}\, \phi\,(\mathbf{x}, y) \cdot \frac{\partial x_j}{\partial x_i} + \frac{\partial}{\partial y}\, \phi(\mathbf{x}, y) \cdot \frac{\partial y}{\partial x_i} \\
&=\; \frac{\partial}{\partial x_i}\, \phi(\mathbf{x}, y) \cdot \frac{\partial x_i}{\partial x_i} + \frac{\partial}{\partial y}\, \phi(\mathbf{x}, y) \cdot \frac{\partial y}{\partial x_i} \\
&=\; \frac{\partial}{\partial x_i}\, \phi(\mathbf{x}, y) + \frac{\partial}{\partial y}\, \phi(\mathbf{x}, y) \cdot \frac{\partial}{\partial x_i}\, f(\mathbf{x})
\end{aligned}$$

Solving this equation for $\frac{\partial}{\partial x_i}\, f(\mathbf{x})$ proves the Theorem.

4.6 The Taylor Formula

Let $D \subset \mathbb{R}^n$ be convex and open, $\mathbf{a}, \mathbf{x} \in D$ and $f : D \to \mathbb{R}$ a 3-times continuously partially differentiable function.

Definition 4.7 *The 2-nd Taylor polynom of f in $\mathbf{a}$ is defined by:*

$$\boxed{t_2(\mathbf{x}) \;=\; f(\mathbf{a}) + \mathbf{grad}\, f(\mathbf{a}) \bullet (\mathbf{x}-\mathbf{a}) + \frac{1}{2}\,(\mathbf{x}-\mathbf{a})^T\, H_f(\mathbf{a})\;(\mathbf{x}-\mathbf{a})}$$

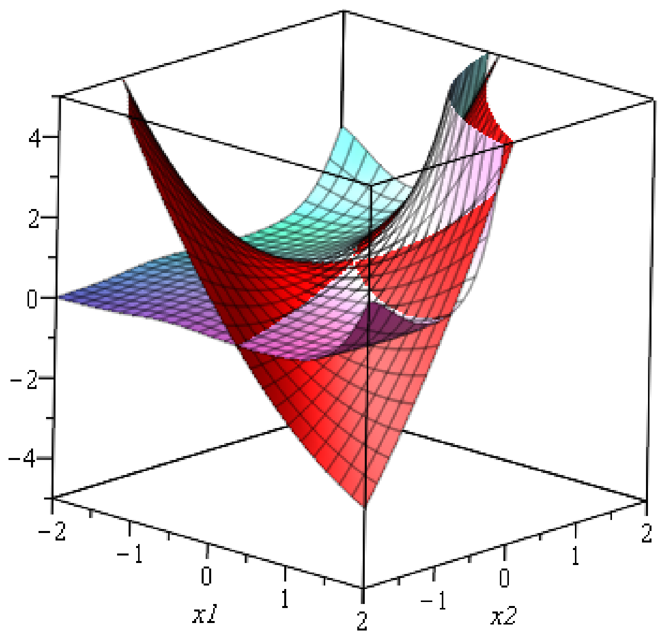
Theorem 4.6

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - t_2(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^2} \;=\; 0$$

Example 12 $f(x_1, x_2) = e^{x_1+x_2} + \sin(x_1 x_2)$, $\mathbf{a} = (0,0)$

$$\begin{aligned} t_2(\mathbf{x}) &= 1 + (1,1)\,\mathbf{x} + \mathbf{x}^T\,\frac{1}{2}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\mathbf{x} \\ &= 1 + x_1 + x_2 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 \end{aligned}$$

Graph of f and t_2 (red)



4.7 Concave and convex functions

Theorem 4.7 *Let $D \subset \mathbb{R}^n$ be a convex set and $f : D \to \mathbb{R}$ a 2-times continuously partially differentiable function. Furthermore, let H_f be the Hesse matrix of f. Then we have*

$$\boxed{\begin{array}{lcl} H_f(\mathbf{x}) \textit{ for all } \mathbf{x} \in D \textit{ negative semidefinite} & \Longleftrightarrow & f \textit{ concave} \\[1ex] H_f(\mathbf{x}) \textit{ for all } \mathbf{x} \in D \textit{ negative definite} & \Longrightarrow & f \textit{ is striktly concave} \\[1ex] H_f(\mathbf{x}) \textit{ for all } \mathbf{x} \in D \textit{ positive semidefinite} & \Longleftrightarrow & f \textit{ is convex} \\[1ex] H_f(\mathbf{x}) \textit{ for all } \mathbf{x} \in D \textit{ positive definite} & \Longrightarrow & f \textit{ is striktly convex} \end{array}}$$

Example 13 *The function $f(x_1, x_2) = 2x_1 - x_2 - x_1^2 + 2x_1x_2 - x_2^2$ is defined on $D = \mathbb{R}^2$ and*

$$H_f(\mathbf{x}) \quad = \quad \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$$

is (always) negative semidefinite. Hence f is concave.