

Analysis III for engineering study programs

Ingenuin Gasser

Departments of Mathematics
Universität Hamburg

Technische Universität Hamburg–Harburg
Wintersemester 2021/22

based on slides of Prof. Jens Struckmeier from Wintersemester 2020/21

Content of the course Analysis III.

- 1 Partial derivatives, differential operators.
- 2 Vector fields, total differential, directional derivative.
- 3 Mean value theorems, Taylor's theorem.
- 4 Extrem values, implicit function theorem.
- 5 Implicit representation of curves and surfaces.
- 6 Extrem values under equality constraints.
- 7 Newton-method, non-linear equations and the least squares method.
- 8 Multiple integrals, Fubini's theorem, transformation theorem.
- 9 Potentials, Green's theorem, Gauß's theorem.
- 10 Green's formulas, Stokes's theorem.

Chapter 1. Multi variable differential calculus

1.1 Partial derivatives

Let

$f(x_1, \dots, x_n)$ a scalar function depending n variables

Example: The constitutive law of an ideal gas $pV = RT$.

Each of the 3 quantities p (pressure), V (volume) and T (temperature) can be expressed as a function of the others (R is the gas constant)

$$p = p(V, T) = \frac{RT}{V}$$

$$V = V(p, T) = \frac{RT}{p}$$

$$T = T(p, V) = \frac{pV}{R}$$

1.1. Partial derivatives

Definition: Let $D \subset \mathbb{R}^n$ be open, $f : D \rightarrow \mathbb{R}$, $x^0 \in D$.

- f is called **partially differentiable** in x^0 with respect to x_i if the limit

$$\begin{aligned}\frac{\partial f}{\partial x_i}(x^0) &:= \lim_{t \rightarrow 0} \frac{f(x^0 + te_i) - f(x^0)}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(x_1^0, \dots, x_i^0 + t, \dots, x_n^0) - f(x_1^0, \dots, x_i^0, \dots, x_n^0)}{t}\end{aligned}$$

exists. e_i denotes the i -th unit vector. The limit is called **partial derivative** of f with respect to x_i at x^0 .

- If at every point x^0 the partial derivatives with respect to every variable $x_i, i = 1, \dots, n$ exist and if the partial derivatives are **continuous functions** then we call f **continuous partial differentiable** or a \mathcal{C}^1 -function.

Examples.

- Consider the function

$$f(x_1, x_2) = x_1^2 + x_2^2$$

At any point $x^0 \in \mathbb{R}^2$ there exist both partial derivatives and both partial derivatives are continuous:

$$\frac{\partial f}{\partial x_1}(x^0) = 2x_1, \quad \frac{\partial f}{\partial x_2}(x^0) = 2x_2$$

Thus f is a \mathcal{C}^1 -function.

- The function

$$f(x_1, x_2) = x_1 + |x_2|$$

at $x^0 = (0, 0)^T$ is partial differentiable with respect to x_1 , but the partial derivative with respect to x_2 does **not** exist!

An engineering example.

The acoustic pressure of a one dimensional acoustic wave is given by

$$p(x, t) = A \sin(\alpha x - \omega t)$$

The partial derivative

$$\frac{\partial p}{\partial x} = \alpha A \cos(\alpha x - \omega t)$$

describes at a given time t the **spacial** rate of change of the pressure.

The partial derivative

$$\frac{\partial p}{\partial t} = -\omega A \cos(\alpha x - \omega t)$$

describes for a fixed position x the **temporal** rate of change of the acoustic pressure.

Rules for differentiation

- Let f, g be differentiable with respect to x_i and $\alpha, \beta \in \mathbb{R}$, then we have the rules

$$\frac{\partial}{\partial x_i} (\alpha f(x) + \beta g(x)) = \alpha \frac{\partial f}{\partial x_i}(x) + \beta \frac{\partial g}{\partial x_i}(x)$$

$$\frac{\partial}{\partial x_i} (f(x) \cdot g(x)) = \frac{\partial f}{\partial x_i}(x) \cdot g(x) + f(x) \cdot \frac{\partial g}{\partial x_i}(x)$$

$$\frac{\partial}{\partial x_i} \left(\frac{f(x)}{g(x)} \right) = \frac{\frac{\partial f}{\partial x_i}(x) \cdot g(x) - f(x) \cdot \frac{\partial g}{\partial x_i}(x)}{g(x)^2} \quad \text{for } g(x) \neq 0$$

- An alternative notation for the partial derivatives of f with respect to x_i at x^0 is given by

$$D_i f(x^0) \quad \text{oder} \quad f_{x_i}(x^0)$$

Gradient and nabla-operator.

Definition: Let $D \subset \mathbb{R}^n$ be an open set and $f : D \rightarrow \mathbb{R}$ partial differentiable.

- We denote the **row vector**

$$\text{grad } f(x^0) := \left(\frac{\partial f}{\partial x_1}(x^0), \dots, \frac{\partial f}{\partial x_n}(x^0) \right)$$

as **gradient** of f at x^0 .

- We denote the symbolic vector

$$\nabla := \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)^T$$

as **nabla-operator**.

- Thus we obtain the **column vector**

$$\nabla f(x^0) := \left(\frac{\partial f}{\partial x_1}(x^0), \dots, \frac{\partial f}{\partial x_n}(x^0) \right)^T$$

More rules on differentiation.

Let f and g be partial differentiable. Then the following **rules on differentiation** hold true:

$$\operatorname{grad}(\alpha f + \beta g) = \alpha \cdot \operatorname{grad} f + \beta \cdot \operatorname{grad} g$$

$$\operatorname{grad}(f \cdot g) = g \cdot \operatorname{grad} f + f \cdot \operatorname{grad} g$$

$$\operatorname{grad}\left(\frac{f}{g}\right) = \frac{1}{g^2}(g \cdot \operatorname{grad} f - f \cdot \operatorname{grad} g), \quad g \neq 0$$

Examples:

- Let $f(x, y) = e^x \cdot \sin y$. Then:

$$\operatorname{grad} f(x, y) = (e^x \cdot \sin y, e^x \cdot \cos y) = e^x(\sin y, \cos y)$$

- For $r(x) := \|x\|_2 = \sqrt{x_1^2 + \dots + x_n^2}$ we have

$$\operatorname{grad} r(x) = \frac{x}{r(x)} = \frac{x}{\|x\|_2} \quad \text{für } x \neq 0,$$

where $x = (x_1, \dots, x_n)$ denotes a row vector.

Partial differentiability does not imply continuity.

Observation: A partial differentiable function (with respect to all coordinates) is not necessarily a **continuous** function.

Example: Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined as

$$f(x, y) := \begin{cases} \frac{x \cdot y}{(x^2 + y^2)^2} & : \text{ for } (x, y) \neq 0 \\ 0 & : \text{ for } (x, y) = 0 \end{cases}$$

The function is partial differentiable on the **entire** \mathbb{R}^2 and we have

$$f_x(0, 0) = f_y(0, 0) = 0$$

$$\frac{\partial f}{\partial x}(x, y) = \frac{y}{(x^2 + y^2)^2} - 4 \frac{x^2 y}{(x^2 + y^2)^3}, \quad (x, y) \neq (0, 0)$$

$$\frac{\partial f}{\partial y}(x, y) = \frac{x}{(x^2 + y^2)^2} - 4 \frac{xy^2}{(x^2 + y^2)^3}, \quad (x, y) \neq (0, 0)$$

Example (continuation).

We calculate the partial derivatives at the origin $(0, 0)$:

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{t \rightarrow 0} \frac{f(t, 0) - f(0, 0)}{t} = \frac{t \cdot 0}{(t^2 + 0^2)^2} = 0$$

$$\frac{\partial f}{\partial y}(0, 0) = \lim_{t \rightarrow 0} \frac{f(0, t) - f(0, 0)}{t} = \frac{0 \cdot t}{(0^2 + t^2)^2} = 0$$

But: At $(0, 0)$ the function is **not** continuous since

$$\lim_{n \rightarrow \infty} f\left(\frac{1}{n}, \frac{1}{n}\right) = \frac{\frac{1}{n} \cdot \frac{1}{n}}{\left(\frac{1}{n} \cdot \frac{1}{n} + \frac{1}{n} \cdot \frac{1}{n}\right)^2} = \frac{\frac{1}{n^2}}{\frac{4}{n^4}} = \frac{n^2}{4} \rightarrow \infty$$

and thus we have

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y) \neq f(0, 0) = 0$$

Boundedness of the derivatives implies continuity.

To guarantee the continuity of a partial differentiable function we need additional conditions on f .

Theorem: Let $D \subset \mathbb{R}^n$ be an open set. Let $f : D \rightarrow \mathbb{R}$ be partial differentiable in a neighborhood of $x^0 \in D$ and let the partial derivatives $\frac{\partial f}{\partial x_i}$, $i = 1, \dots, n$, be **bounded**. Then f is **continuous** in x^0 .

Attention: In the previous example the partial derivatives are **not** bounded in a neighborhood of $(0,0)$ since

$$\frac{\partial f}{\partial x}(x, y) = \frac{y}{(x^2 + y^2)^2} - 4 \frac{x^2 y}{(x^2 + y^2)^3} \quad \text{für } (x, y) \neq (0, 0)$$

Proof of the theorem.

For $\|x - x^0\|_\infty < \varepsilon$, $\varepsilon > 0$ sufficiently small we write:

$$\begin{aligned} f(x) - f(x^0) &= (f(x_1, \dots, x_{n-1}, x_n) - f(x_1, \dots, x_{n-1}, x_n^0)) \\ &+ (f(x_1, \dots, x_{n-1}, x_n^0) - f(x_1, \dots, x_{n-2}, x_{n-1}^0, x_n^0)) \\ &\vdots \\ &+ (f(x_1, x_2^0, \dots, x_n^0) - f(x_1^0, \dots, x_n^0)) \end{aligned}$$

For any difference on the right hand side we consider f as a function in one single variable:

$$g(x_n) - g(x_n^0) := f(x_1, \dots, x_{n-1}, x_n) - f(x_1, \dots, x_{n-1}, x_n^0)$$

Since f is partial differentiable g is differentiable and we can apply the mean value theorem on g :

$$g(x_n) - g(x_n^0) = g'(\xi_n)(x_n - x_n^0)$$

for an appropriate ξ_n between x_n and x_n^0 .

Proof of the theorem (continuation).

Applying the [mean value theorem](#) to every term in the right hand side we obtain

$$\begin{aligned} f(x) - f(x^0) &= \frac{\partial f}{\partial x_n}(x_1, \dots, x_{n-1}, \xi_n) \cdot (x_n - x_n^0) \\ &+ \frac{\partial f}{\partial x_{n-1}}(x_1, \dots, x_{n-2}, \xi_{n-1}, x_n^0) \cdot (x_{n-1} - x_{n-1}^0) \\ &\vdots \\ &+ \frac{\partial f}{\partial x_1}(\xi_1, x_2^0, \dots, x_n^0) \cdot (x_1 - x_1^0) \end{aligned}$$

Using the boundedness of the partial derivatives

$$|f(x) - f(x^0)| \leq C_1|x_1 - x_1^0| + \dots + C_n|x_n - x_n^0|$$

for $\|x - x^0\|_\infty < \varepsilon$, we obtain the [continuity](#) of f at x^0 since

$$f(x) \rightarrow f(x^0) \quad \text{für } \|x - x^0\|_\infty \rightarrow 0$$

Higher order derivatives.

Definition: Let f be a scalar function and partial differentiable on an open set $D \subset \mathbb{R}^n$. If the partial derivatives are differentiable we obtain (by differentiating) the **partial derivatives of second order** of f with

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right)$$

Example: Second order partial derivatives of a function $f(x, y)$:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right), \quad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right), \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y^2}$$

Let $i_1, \dots, i_k \in \{1, \dots, n\}$. Then we define recursively

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}} := \frac{\partial}{\partial x_{i_k}} \left(\frac{\partial^{k-1} f}{\partial x_{i_{k-1}} \partial x_{i_{k-2}} \dots \partial x_{i_1}} \right)$$

Higher order derivatives.

Definition: The function f is called k -times partial differentiable, if all derivatives of order k ,

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}} \quad \text{for all } i_1, \dots, i_k \in \{1, \dots, n\},$$

exist on D .

Alternative notation:

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}} = D_{i_k} D_{i_{k-1}} \dots D_{i_1} f = f_{x_{i_1} \dots x_{i_k}}$$

If all the derivatives of k -th order are continuous the function f is called k -times continuous partial differentiable or called a C^k -function on D . Continuous functions f are called C^0 -functions.

Example: For the function $f(x_1, \dots, x_n) = \prod_{i=1}^n x_i^i$ we have $\frac{\partial^n f}{\partial x_n \dots \partial x_1} = ?$

Partial derivatives are not arbitrarily exchangeable.

ATTENTION: The order how to execute partial derivatives is in general **not** arbitrarily exchangeable!

Example: For the function

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

we calculate

$$f_{xy}(0, 0) = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x}(0, 0) \right) = -1$$

$$f_{yx}(0, 0) = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y}(0, 0) \right) = +1$$

i.e. $f_{xy}(0, 0) \neq f_{yx}(0, 0)$.

Theorem of Schwarz on exchangeability.

Satz: Let $D \subset \mathbb{R}^n$ be open and let $f : D \rightarrow \mathbb{R}$ be a \mathcal{C}^2 -function. Then it holds

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x_1, \dots, x_n) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_n)$$

for all $i, j \in \{1, \dots, n\}$.

Idea of the proof:

Apply the mean value theorem twice.

Conclusion:

If f is a C^k -function, then we can exchange the differentiation in order to calculate partial derivatives up to order k **arbitrarily!**

Example for the exchangeability of partial derivatives.

Calculate the partial derivative of third order f_{xyz} for the function

$$f(x, y, z) = y^2 z \sin(x^3) + (\cosh y + 17e^{x^2})z^2$$

The order of execution is exchangeable since $f \in \mathcal{C}^3$.

- Differentiate first with respect to z :

$$\frac{\partial f}{\partial z} = y^2 \sin(x^3) + 2z(\cosh y + 17e^{x^2})$$

- Differentiate then f_z with respect to x (then $\cosh y$ disappears):

$$\begin{aligned} f_{zx} &= \frac{\partial}{\partial x} \left(y^2 \sin(x^3) + 2z(\cosh y + 17e^{x^2}) \right) \\ &= 3x^2 y^2 \cos(x^3) + 68xze^{x^2} \end{aligned}$$

- For the partial derivative of f_{zx} with respect to y we obtain

$$f_{xyz} = 6x^2 y \cos(x^3)$$

The Laplace operator.

The **Laplace-operator** or **Laplacian** is defined as

$$\Delta := \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$$

For a scalar function $u(x) = u(x_1, \dots, x_n)$ we have

$$\Delta u = \sum_{i=1}^n \frac{\partial^2 u}{\partial x_i^2} = u_{x_1 x_1} + \dots + u_{x_n x_n}$$

Examples of important partial differential equations of second order (i.e. equations containing partial derivatives up to order two):

$$\Delta u - \frac{1}{c^2} u_{tt} = 0 \quad (\text{wave equation})$$

$$\Delta u - \frac{1}{k} u_t = 0 \quad (\text{heat equation})$$

$$\Delta u = 0 \quad (\text{Laplace-equation or equation for the potential})$$

Vector valued functions.

Definition: Let $D \subset \mathbb{R}^n$ be open and let $f : D \rightarrow \mathbb{R}^m$ be a vector valued function.

The function f is called **partial differentiable** on $x^0 \in D$, if for all $i = 1, \dots, n$ the limits

$$\frac{\partial f}{\partial x_i}(x^0) = \lim_{t \rightarrow 0} \frac{f(x^0 + te_i) - f(x^0)}{t}$$

exist. The calculation is done componentwise

$$\frac{\partial f}{\partial x_i}(x^0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_i} \\ \frac{\partial f_2}{\partial x_i} \\ \vdots \\ \frac{\partial f_m}{\partial x_i} \end{pmatrix} \quad \text{for } i = 1, \dots, n$$

Vectorfields.

Definition: If $m = n$ the function $f : D \rightarrow \mathbb{R}^n$ is called a **vectorfield** on D . If every (coordinate-) function $f_i(x)$ of $f = (f_1, \dots, f_n)^T$ is a C^k -function, then f is called **C^k -vectorfield**.

Examples of vectorfields:

- velocity fields of liquids or gases;
- elektromagnetic fields;
- temperature gradients in solid states.

Definition: Let $f : D \rightarrow \mathbb{R}^n$ be a partial differentiable vector field. The **divergence** on $x \in D$ is defined as

$$\operatorname{div} f(x^0) := \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x^0)$$

or

$$\operatorname{div} f(x) = \nabla^T f(x) = (\nabla, f(x))$$

Rules of computation and the rotation.

The following rules hold true:

$$\operatorname{div}(\alpha f + \beta g) = \alpha \operatorname{div} f + \beta \operatorname{div} g \quad \text{for } f, g : D \rightarrow \mathbb{R}^n$$

$$\operatorname{div}(\varphi \cdot f) = (\nabla \varphi, f) + \varphi \operatorname{div} f \quad \text{for } \varphi : D \rightarrow \mathbb{R}, f : D \rightarrow \mathbb{R}^n$$

Remark: Let $f : D \rightarrow \mathbb{R}$ be a C^2 -function, then for the Laplacian we have

$$\Delta f = \operatorname{div}(\nabla f)$$

Definition: Let $D \subset \mathbb{R}^3$ open and $f : D \rightarrow \mathbb{R}^3$ a partial differentiable vector field. We define the **rotation** as

$$\operatorname{rot} f(x^0) := \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \right)^T \Big|_{x^0}$$

Alternative notations and additional rules.

$$\operatorname{rot} f(x) = \nabla \times f(x) = \begin{vmatrix} e_1 & e_2 & e_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ f_1 & f_2 & f_3 \end{vmatrix}$$

Remark: The following rules hold true:

$$\operatorname{rot}(\alpha f + \beta g) = \alpha \operatorname{rot} f + \beta \operatorname{rot} g$$

$$\operatorname{rot}(\varphi \cdot f) = (\nabla \varphi) \times f + \varphi \operatorname{rot} f$$

Remark: Let $D \subset \mathbb{R}^3$ and $\varphi : D \rightarrow \mathbb{R}$ be a \mathcal{C}^2 -function. Then

$$\operatorname{rot}(\nabla \varphi) = 0,$$

using the exchangeability theorem of Schwarz. I.e. gradient fields are **rotation-free** everywhere.

1.2 The total differential

Definition: Let $D \subset \mathbb{R}^n$ open, $x^0 \in D$ and $f : D \rightarrow \mathbb{R}^m$. The function $f(x)$ is called **differentiable** in x^0 (or **totally differentiable** in x_0), if there exists a linear map

$$l(x, x^0) := A \cdot (x - x^0)$$

with a matrix $A \in \mathbb{R}^{m \times n}$ which satisfies the following approximation property

$$f(x) = f(x^0) + A \cdot (x - x^0) + o(\|x - x^0\|)$$

i.e.

$$\lim_{x \rightarrow x^0} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|} = 0.$$

The total differential and the Jacobian matrix.

Notation: We call the linear map l the **differential** or the **total differential** of $f(x)$ at the point x^0 . We denote l by $df(x^0)$.

The related matrix A is called **Jacobi-matrix** of $f(x)$ at the point x^0 and is denoted by $Jf(x^0)$ (or $Df(x^0)$ or $f'(x^0)$).

Remark: For $m = n = 1$ we obtain the well known relation

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(|x - x_0|)$$

for the derivative $f'(x_0)$ at the point x_0 .

Remark: In case of a scalar function ($m = 1$) the matrix $A = a$ is a row vector and $a(x - x^0)$ a scalar product $\langle a^T, x - x^0 \rangle$.

Total and partial differentiability.

Theorem: Let $f : D \rightarrow \mathbb{R}^m$, $x^0 \in D \subset \mathbb{R}^n$, D open.

- a) If $f(x)$ is differentiable in x^0 , then $f(x)$ is continuous in x^0 .
- b) If $f(x)$ is differentiable in x^0 , then the (total) differential and thus the Jacobi-matrix are uniquely determined and we have

$$Jf(x^0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x^0) & \dots & \frac{\partial f_1}{\partial x_n}(x^0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x^0) & \dots & \frac{\partial f_m}{\partial x_n}(x^0) \end{pmatrix} = \begin{pmatrix} Df_1(x^0) \\ \vdots \\ Df_m(x^0) \end{pmatrix}$$

- c) If $f(x)$ is a C^1 -function on D , then $f(x)$ is differentiable on D .

Proof of a).

If f is differentiable in x^0 , then by definition

$$\lim_{x \rightarrow x^0} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|} = 0$$

Thus we conclude

$$\lim_{x \rightarrow x^0} \|f(x) - f(x^0) - A \cdot (x - x^0)\| = 0$$

and we obtain

$$\begin{aligned} \|f(x) - f(x^0)\| &\leq \|f(x) - f(x^0) - A \cdot (x - x^0)\| + \|A \cdot (x - x^0)\| \\ &\rightarrow 0 \quad \text{as } x \rightarrow x^0 \end{aligned}$$

Therefore the function f is continuous at x^0 .

Proof of b).

Let $x = x^0 + te_i$, $|t| < \varepsilon$, $i \in \{1, \dots, n\}$. Since f is differentiable at x^0 , we have

$$\lim_{x \rightarrow x^0} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|_\infty} = 0$$

We write

$$\begin{aligned} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|_\infty} &= \frac{f(x^0 + te_i) - f(x^0)}{|t|} - \frac{tAe_i}{|t|} \\ &= \frac{t}{|t|} \cdot \left(\frac{f(x^0 + te_i) - f(x^0)}{t} - Ae_i \right) \\ &\rightarrow 0 \quad \text{as } t \rightarrow 0 \end{aligned}$$

Thus

$$\lim_{t \rightarrow 0} \frac{f(x^0 + te_i) - f(x^0)}{t} = Ae_i \quad i = 1, \dots, n$$

Examples.

- Consider the scalar function $f(x_1, x_2) = x_1 e^{2x_2}$. Then the Jacobian is given by:

$$Jf(x_1, x_2) = Df(x_1, x_2) = e^{2x_2}(1, 2x_1)$$

- Consider the function $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ defined by

$$f(x_1, x_2, x_3) = \begin{pmatrix} x_1 x_2 x_3 \\ \sin(x_1 + 2x_2 + 3x_3) \end{pmatrix}$$

The Jacobian is given by

$$Jf(x_1, x_2, x_3) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{pmatrix} = \begin{pmatrix} x_2 x_3 & x_1 x_3 & x_1 x_2 \\ \cos(s) & 2 \cos(s) & 3 \cos(s) \end{pmatrix}$$

with $s = x_1 + 2x_2 + 3x_3$.

Further examples.

- Let $f(x) = Ax$, $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$. Then

$$Jf(x) = A \quad \text{for all } x \in \mathbb{R}^n$$

- Let $f(x) = x^T Ax = \langle x, Ax \rangle$, $A \in \mathbb{R}^{n \times n}$ and $x \in \mathbb{R}^n$.
Then we have

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \langle e_i, Ax \rangle + \langle x, Ae_i \rangle \\ &= e_i^T Ax + x^T Ae_i \\ &= x^T (A^T + A)e_i \end{aligned}$$

We conclude

$$Jf(x) = \text{grad}f(x) = x^T (A^T + A)$$

Rules for the differentiation.

Theorem:

- a) **Linearität:** LET $f, g : D \rightarrow \mathbb{R}^m$ be differentiable in $x^0 \in D$, D open. Then $\alpha f(x^0) + \beta g(x^0)$, and $\alpha, \beta \in \mathbb{R}$ are differentiable in x^0 and we have

$$d(\alpha f + \beta g)(x^0) = \alpha df(x^0) + \beta dg(x^0)$$

$$J(\alpha f + \beta g)(x^0) = \alpha Jf(x^0) + \beta Jg(x^0)$$

- b) **Chain rule:** Let $f : D \rightarrow \mathbb{R}^m$ be differentiable in $x^0 \in D$, D open. Let $g : E \rightarrow \mathbb{R}^k$ be differentiable in $y^0 = f(x^0) \in E \subset \mathbb{R}^m$, E open. Then $g \circ f$ is differentiable in x^0 .

For the differentials it holds

$$d(g \circ f)(x^0) = dg(y^0) \circ df(x^0)$$

and analogously for the Jacobian matrix

$$J(g \circ f)(x^0) = Jg(y^0) \cdot Jf(x^0)$$

Examples for the chain rule.

Let $I \subset \mathbb{R}$ be an interval. Let $h : I \rightarrow \mathbb{R}^n$ be a curve, differentiable in $t_0 \in I$ with values in $D \subset \mathbb{R}^n$, D open. Let $f : D \rightarrow \mathbb{R}$ be a scalar function, differentiable in $x^0 = h(t_0)$.

Then the composition

$$(f \circ h)(t) = f(h_1(t), \dots, h_n(t))$$

is differentiable in t_0 and we have for the derivative:

$$\begin{aligned}(f \circ h)'(t_0) &= Jf(h(t_0)) \cdot Jh(t_0) \\ &= \operatorname{grad} f(h(t_0)) \cdot h'(t_0) \\ &= \sum_{k=1}^n \frac{\partial f}{\partial x_k}(h(t_0)) \cdot h'_k(t_0)\end{aligned}$$

Directional derivative.

Definition: Let $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$ open, $x^0 \in D$, and $v \in \mathbb{R} \setminus \{0\}$ a vector. Then

$$D_v f(x^0) := \lim_{t \rightarrow 0} \frac{f(x^0 + tv) - f(x^0)}{t}$$

is called the **directional derivative (Gateaux-derivative)** of $f(x)$ in the direction of v .

Example: Let $f(x, y) = x^2 + y^2$ and $v = (1, 1)^T$. Then the directional derivative in the direction of v is given by:

$$\begin{aligned} D_v f(x, y) &= \lim_{t \rightarrow 0} \frac{(x+t)^2 + (y+t)^2 - x^2 - y^2}{t} \\ &= \lim_{t \rightarrow 0} \frac{2xt + t^2 + 2yt + t^2}{t} \\ &= 2(x + y) \end{aligned}$$

Remarks.

- For $v = e_i$ the directional derivative in the direction of v is given by the partial derivative with respect to x_i :

$$D_v f(x^0) = \frac{\partial f}{\partial x_i}(x^0)$$

- If v is a unit vector, i.e. $\|v\| = 1$, then the directional derivative $D_v f(x^0)$ describes the **slope** of $f(x)$ in the direction of v .
- If $f(x)$ is differentiable in x^0 , then all directional derivatives of $f(x)$ in x^0 exist. With $h(t) = x^0 + tv$ we have

$$D_v f(x^0) = \frac{d}{dt}(f \circ h)|_{t=0} = \text{grad } f(x^0) \cdot v$$

This follows directly applying the chain rule.

Properties of the gradient.

Theorem: Let $D \subset \mathbb{R}^n$ open, $f : D \rightarrow \mathbb{R}$ differentiable in $x^0 \in D$. Then we have

- a) The gradient vector $\text{grad } f(x^0) \in \mathbb{R}^n$ is orthogonal in the **level set**

$$N_{x^0} := \{x \in D \mid f(x) = f(x^0)\}$$

In the case of $n = 2$ we call the level sets **contour lines**, in $n = 3$ we call the level sets **equipotential surfaces**.

- 2) The gradient $\text{grad } f(x^0)$ gives the direction of the steepest slope of $f(x)$ in x^0 .

Idea of the proof:

- a) application of the chain rule.
b) for an arbitrary direction v we conclude with the Cauchy–Schwarz inequality

$$|D_v f(x^0)| = |(\text{grad } f(x^0), v)| \leq \|\text{grad } f(x^0)\|_2$$

Equality is obtained for $v = \text{grad } f(x^0) / \|\text{grad } f(x^0)\|_2$.

Curvilinear coordinates.

Definition: Let $U, V \subset \mathbb{R}^n$ be open and $\Phi : U \rightarrow V$ be a \mathcal{C}^1 -map, for which the Jacobimatrix $J\Phi(u^0)$ is regular (invertible) at every $u^0 \in U$. In addition there exists the inverse map $\Phi^{-1} : V \rightarrow U$ and the inverse map is also a \mathcal{C}^1 -map.

Then $x = \Phi(u)$ defines a **coordinate transformation** from the coordinates u to x .

Example: Consider for $n = 2$ the **polar coordinates** $u = (r, \varphi)$ with $r > 0$ and $-\pi < \varphi < \pi$ and set

$$x = r \cos \varphi$$

$$y = r \sin \varphi$$

with the **cartesian coordinates** $x = (x, y)$.

Calculation of the partial derivatives.

For all $u \in U$ with $x = \Phi(u)$ the following relations hold

$$\Phi^{-1}(\Phi(u)) = u$$

$$J\Phi^{-1}(x) \cdot J\Phi(u) = I_n \quad (\text{chain rule})$$

$$J\Phi^{-1}(x) = (J\Phi(u))^{-1}$$

Let $\tilde{f} : V \rightarrow \mathbb{R}$ be a given function. Set

$$f(u) := \tilde{f}(\Phi(u))$$

the by using the chain rule we obtain

$$\frac{\partial f}{\partial u_i} = \sum_{j=1}^n \frac{\partial \tilde{f}}{\partial x_j} \frac{\partial \Phi_j}{\partial u_i} =: \sum_{j=1}^n g^{ij} \frac{\partial \tilde{f}}{\partial x_j}$$

with

$$g^{ij} := \frac{\partial \Phi_j}{\partial u_i}, \quad G(u) := (g^{ij}) = (J\Phi(u))^T$$

Notations.

We use the short notation

$$\frac{\partial}{\partial u_i} = \sum_{j=1}^n g^{ij} \frac{\partial}{\partial x_j}$$

Analogously we can express the partial derivatives with respect to x_i by the partial derivatives with respect to u_j

$$\frac{\partial}{\partial x_i} = \sum_{j=1}^n g_{ij} \frac{\partial}{\partial u_j}$$

where

$$(g_{ij}) := (g^{ij})^{-1} = (J\Phi)^{-T} = (J\Phi^{-1})^T$$

We obtain these relations by applying the chain rule on Φ^{-1} .

Example: polar coordinates.

We consider polar coordinates

$$x = \Phi(u) = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix}$$

We calculate

$$J\Phi(u) = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix}$$

and thus

$$(g^{ij}) = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -r \sin \varphi & r \cos \varphi \end{pmatrix} \quad (g_{ij}) = \begin{pmatrix} \cos \varphi & -\frac{1}{r} \sin \varphi \\ \sin \varphi & \frac{1}{r} \cos \varphi \end{pmatrix}$$

Partial derivatives for polar coordinates.

The calculation of the partial derivatives gives

$$\frac{\partial}{\partial x} = \cos \varphi \frac{\partial}{\partial r} - \frac{1}{r} \sin \varphi \frac{\partial}{\partial \varphi}$$

$$\frac{\partial}{\partial y} = \sin \varphi \frac{\partial}{\partial r} + \frac{1}{r} \cos \varphi \frac{\partial}{\partial \varphi}$$

Example: Calculation of the **Laplacian-operator** in polar coordinates

$$\frac{\partial^2}{\partial x^2} = \cos^2 \varphi \frac{\partial^2}{\partial r^2} - \frac{\sin(2\varphi)}{r} \frac{\partial^2}{\partial r \partial \varphi} + \frac{\sin^2 \varphi}{r^2} \frac{\partial^2}{\partial \varphi^2} + \frac{\sin(2\varphi)}{r^2} \frac{\partial}{\partial \varphi} + \frac{\sin^2 \varphi}{r} \frac{\partial}{\partial r}$$

$$\frac{\partial^2}{\partial y^2} = \sin^2 \varphi \frac{\partial^2}{\partial r^2} + \frac{\sin(2\varphi)}{r} \frac{\partial^2}{\partial r \partial \varphi} + \frac{\cos^2 \varphi}{r^2} \frac{\partial^2}{\partial \varphi^2} - \frac{\sin(2\varphi)}{r^2} \frac{\partial}{\partial \varphi} + \frac{\cos^2 \varphi}{r} \frac{\partial}{\partial r}$$

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2} + \frac{1}{r} \frac{\partial}{\partial r}$$

Example: spherical coordinates.

We consider spherical coordinates

$$\mathbf{x} = \Phi(\mathbf{u}) = \begin{pmatrix} r \cos \varphi \cos \theta \\ r \sin \varphi \cos \theta \\ r \sin \theta \end{pmatrix}$$

The Jacobian–matrix is given by:

$$J\Phi(\mathbf{u}) = \begin{pmatrix} \cos \varphi \cos \theta & -r \sin \varphi \cos \theta & -r \cos \varphi \sin \theta \\ \sin \varphi \cos \theta & r \cos \varphi \cos \theta & -r \sin \varphi \sin \theta \\ \sin \theta & 0 & r \cos \theta \end{pmatrix}$$

Partial derivatives for spherical coordinates.

Calculating the partial derivatives gives

$$\frac{\partial}{\partial x} = \cos \varphi \cos \theta \frac{\partial}{\partial r} - \frac{\sin \varphi}{r \cos \theta} \frac{\partial}{\partial \varphi} - \frac{1}{r} \cos \varphi \sin \theta \frac{\partial}{\partial \theta}$$

$$\frac{\partial}{\partial y} = \sin \varphi \cos \theta \frac{\partial}{\partial r} + \frac{\cos \varphi}{r \cos \theta} \frac{\partial}{\partial \varphi} - \frac{1}{r} \sin \varphi \sin \theta \frac{\partial}{\partial \theta}$$

$$\frac{\partial}{\partial z} = \sin \theta \frac{\partial}{\partial r} + \frac{1}{r} \cos \theta \frac{\partial}{\partial \theta}$$

Example: calculation of the [Laplace-operator](#) in spherical coordinates

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2 \cos^2 \theta} \frac{\partial^2}{\partial \varphi^2} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{\tan \theta}{r^2} \frac{\partial}{\partial \theta}$$

Chapter 1. Multivariate differential calculus

1.3 Mean value theorems and Taylor expansion

Theorem (Mean value theorem): Let $f : D \rightarrow \mathbb{R}$ be a scalar differentiable function on an open set $D \subset \mathbb{R}^n$. Let $a, b \in D$ be points in D such that the connecting line segment

$$[a, b] := \{a + t(b - a) \mid t \in [0, 1]\}$$

lies entirely in D . Then there exists a number $\theta \in (0, 1)$ with

$$f(b) - f(a) = \text{grad } f(a + \theta(b - a)) \cdot (b - a)$$

Proof: We set

$$h(t) := f(a + t(b - a))$$

with the mean value theorem for a **single** variable and the chain rules we conclude

$$\begin{aligned} f(b) - f(a) &= h(1) - h(0) = h'(\theta) \cdot (1 - 0) \\ &= \text{grad } f(a + \theta(b - a)) \cdot (b - a) \end{aligned}$$

Definition and example.

Definition: If the condition $[a, b] \subset D$ holds true for **all** points $a, b \in D$, then the set D is called **convex**.

Example for the mean value theorem: Given a scalar function

$$f(x, y) := \cos x + \sin y$$

It is

$$f(0, 0) = f(\pi/2, \pi/2) = 1 \quad \Rightarrow \quad f(\pi/2, \pi/2) - f(0, 0) = 0$$

Applying the mean value theorem there exists a $\theta \in (0, 1)$ with

$$\text{grad } f \left(\theta \begin{pmatrix} \pi/2 \\ \pi/2 \end{pmatrix} \right) \cdot \begin{pmatrix} \pi/2 \\ \pi/2 \end{pmatrix} = 0$$

Indeed this is true for $\theta = \frac{1}{2}$.

Mean value theorem is only true for **scalar** functions.

Attention: The mean value theorem for multivariate functions is only true for **scalar** functions but in general not for **vector-valued** functions!

Examples: Consider the **vector-valued** Function

$$f(t) := \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad t \in [0, \pi/2]$$

It is

$$f(\pi/2) - f(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

and

$$f' \left(\theta \frac{\pi}{2} \right) \cdot \left(\frac{\pi}{2} - 0 \right) = \frac{\pi}{2} \begin{pmatrix} -\sin(\theta\pi/2) \\ \cos(\theta\pi/2) \end{pmatrix}$$

BUT: the vectors on the right hand side have length $\sqrt{2}$ and $\pi/2$!

A mean value estimate for **vector-valued** functions.

Theorem: Let $f : D \rightarrow \mathbb{R}^m$ be differentiable on an open set $D \subset \mathbb{R}^n$. Let a, b be points in D with $[a, b] \subset D$. Then there exists a $\theta \in (0, 1)$ with

$$\|f(b) - f(a)\|_2 \leq \|Jf(a + \theta(b - a)) \cdot (b - a)\|_2$$

Idea of the proof: Application of the mean value theorem to the **scalar** function $g(x)$ defined as

$$g(x) := (f(b) - f(a))^T f(x) \quad (\text{scalar product!})$$

Remark: Another (weaker) form of the mean value estimate is

$$\|f(b) - f(a)\| \leq \sup_{\xi \in [a, b]} \|Jf(\xi)\| \cdot \|(b - a)\|$$

where $\|\cdot\|$ denotes an arbitrary vector norm with related matrix norm.

Taylor series: notations.

We define the **multi-index** $\alpha \in \mathbb{N}_0^n$ as

$$\alpha := (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$$

Let

$$|\alpha| := \alpha_1 + \dots + \alpha_n \quad \alpha! := \alpha_1! \cdots \alpha_n!$$

Let $f : D \rightarrow \mathbb{R}$ be $|\alpha|$ times continuous differentiable. Then we set

$$D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \dots D_n^{\alpha_n} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}},$$

where $D_i^{\alpha_i} = \underbrace{D_i \dots D_i}_{\alpha_i\text{-mal}}$. We write

$$x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n} \quad \text{for } x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

The Taylor theorem.

Theorem: (Taylor)

Let $D \subset \mathbb{R}^n$ be open and convex. Let $f : D \rightarrow \mathbb{R}$ be a \mathcal{C}^{m+1} -function and $x_0 \in D$. Then the Taylor-expansion holds true in $x \in D$

$$f(x) = T_m(x; x_0) + R_m(x; x_0)$$

$$T_m(x; x_0) = \sum_{|\alpha| \leq m} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha$$

$$R_m(x; x_0) = \sum_{|\alpha|=m+1} \frac{D^\alpha f(x_0 + \theta(x - x_0))}{\alpha!} (x - x_0)^\alpha$$

for an appropriate $\theta \in (0, 1)$.

Notation: In the Taylor-expansion we denote $T_m(x; x_0)$ Taylor-polynom of degree m and $R_m(x; x_0)$ Lagrange-remainder.

Derivation of the Taylor expansion.

We define a scalar function in **one single** variable $t \in [0, 1]$ as

$$g(t) := f(x_0 + t(x - x_0))$$

and calculate the (univariate) Taylor–expansion **at $t = 0$** . It is

$$g(1) = g(0) + g'(0) \cdot (1 - 0) + \frac{1}{2}g''(\xi) \cdot (1 - 0)^2 \quad \text{for a } \xi \in (0, 1).$$

The calculation of $g'(0)$ is given by the chain rule

$$\begin{aligned} g'(0) &= \left. \frac{d}{dt} f(x_1^0 + t(x_1 - x_1^0), x_2^0 + t(x_2 - x_2^0), \dots, x_n^0 + t(x_n - x_n^0)) \right|_{t=0} \\ &= D_1 f(x_0) \cdot (x_1 - x_1^0) + \dots + D_n f(x_0) \cdot (x_n - x_n^0) \\ &= \sum_{|\alpha|=1} \frac{D^\alpha f(x_0)}{\alpha!} \cdot (x - x_0)^\alpha \end{aligned}$$

Continuation of the derivation.

Calculation of $g''(0)$ gives

$$\begin{aligned}g''(0) &= \left. \frac{d^2}{dt^2} f(x_0 + t(x - x_0)) \right|_{t=0} = \left. \frac{d}{dt} \sum_{k=1}^n D_k f(x_0 + t(x - x_0)) (x_k - x_k^0) \right|_{t=0} \\&= D_{11} f(x_0) (x_1 - x_1^0)^2 + D_{21} f(x_0) (x_1 - x_1^0) (x_2 - x_2^0) \\&\quad + \dots + D_{ij} f(x_0) (x_i - x_i^0) (x_j - x_j^0) + \dots + \\&\quad + D_{n-1,n} f(x_0) (x_{n-1} - x_{n-1}^0) (x_n - x_n^0) + D_{nn} f(x_0) (x_n - x_n^0)^2 \\&= \sum_{|\alpha|=2} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha \quad (\text{exchange theorem of Schwarz!})\end{aligned}$$

Continuation: Proof of the Taylor-formula by (mathematical) induction!

Proof of the Taylor theorem.

The function

$$g(t) := f(x^0 + t(x - x^0))$$

is $(m + 1)$ -times continuous differentiable and we have

$$g(1) = \sum_{k=0}^m \frac{g^{(k)}(0)}{k!} + \frac{g^{(m+1)}(\theta)}{(m+1)!} \quad \text{for a } \theta \in [0, 1].$$

In addition we have (by induction over k)

$$\frac{g^{(k)}(0)}{k!} = \sum_{|\alpha|=k} \frac{D^\alpha f(x^0)}{\alpha!} (x - x^0)^\alpha$$

and

$$\frac{g^{(m+1)}(\theta)}{(m+1)!} = \sum_{|\alpha|=m+1} \frac{D^\alpha f(x^0 + \theta(x - x^0))}{\alpha!} (x - x^0)^\alpha$$

Examples for the Taylor–expansion.

- 1 Calculate the Taylor–polynom $T_2(x; x_0)$ of degree 2 of the function

$$f(x, y, z) = x y^2 \sin z$$

at $(x, y, z) = (1, 2, 0)^T$.

- 2 The calculation of $T_2(x; x_0)$ requires the partial derivatives up to order 2.
- 3 These derivatives have to be evaluated at $(x, y, z) = (1, 2, 0)^T$.
- 4 The result is $T_2(x; x_0)$ in the form

$$T_2(x; x_0) = 4z(x + y - 2)$$

- 5 Details on extra slide.

Remarks to the remainder of a Taylor–expansion.

Remark: The remainder of a Taylor–expansion contains **all** partial derivatives of order $(m + 1)$:

$$R_m(x; x_0) = \sum_{|\alpha|=m+1} \frac{D^\alpha f(x_0 + \theta(x - x_0))}{\alpha!} (x - x_0)^\alpha$$

If all these derivative are bounded by a constant C in a neighborhood of x_0 then the **estimate for the remainder** hold true

$$|R_m(x; x_0)| \leq \frac{n^{m+1}}{(m+1)!} C \|x - x_0\|_\infty^{m+1}$$

We conclude for the quality of the approximation of a C^{m+1} –function by the Taylor–polynom

$$f(x) = T_m(x; x_0) + O(\|x - x_0\|^{m+1})$$

Special case $m = 1$: For a C^2 –function $f(x)$ we obtain

$$f(x) = f(x^0) + \text{grad } f(x^0) \cdot (x - x^0) + O(\|x - x^0\|^2).$$

The Hesse-matrix.

The matrix

$$Hf(x_0) := \begin{pmatrix} f_{x_1 x_1}(x_0) & \cdots & f_{x_1 x_n}(x_0) \\ \vdots & & \vdots \\ f_{x_n x_1}(x_0) & \cdots & f_{x_n x_n}(x_0) \end{pmatrix}$$

is called **Hesse-matrix** of f at x_0 .

Hesse-matrix = Jacobi-matrix of the gradient ∇f

The Taylor-expansion of a \mathcal{C}^3 -function can be written as

$$f(x) = f(x_0) + \text{grad } f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T Hf(x_0)(x - x_0) + O(\|x - x_0\|^3)$$

The Hesse-matrix of a \mathcal{C}^2 -function is symmetric.

2.1 Extrem values of multivariate functions

Definition: Let $D \subset \mathbb{R}^n$, $f : D \rightarrow \mathbb{R}$ and $x^0 \in D$. Then at x^0 the function f has

- a **global maximum** if $f(x) \leq f(x^0)$ for all $x \in D$.
- a **strict global maximum** if $f(x) < f(x^0)$ for all $x \in D$.
- a **local maximum** if there exists an $\varepsilon > 0$ such that

$$f(x) \leq f(x^0) \quad \text{for all } x \in D \text{ with } \|x - x^0\| < \varepsilon.$$

- a **strict local maximum** if there exists an $\varepsilon > 0$ such that

$$f(x) < f(x^0) \quad \text{for all } x \in D \text{ with } \|x - x^0\| < \varepsilon.$$

Analogously we define the different forms of minima.

Necessary conditions for local extrem values.

Theorem: If a \mathcal{C}^1 -function $f(x)$ has a local extrem value (minimum or maximum) at $x^0 \in D^0$, then

$$\text{grad } f(x^0) = 0 \in \mathbb{R}^n$$

Proof: For an arbitrary $v \in \mathbb{R}^n$, $v \neq 0$ the function

$$\varphi(t) := f(x^0 + tv)$$

is differentiable in a neighborhood of $t^0 = 0$.

$\varphi(t)$ has a local extrem value at $t^0 = 0$. We conclude:

$$\varphi'(0) = \text{grad } f(x^0) v = 0$$

Since this holds true for all $v \neq 0$ we obtain

$$\text{grad } f(x^0) = (0, \dots, 0)^T$$

Remarks to local extrem values.

Bemerkungen:

- Typically the condition $\text{grad } f(x^0) = 0$ gives a **non-linear** system of n equations for n unknowns for the calculation of $x = x^0$.
- The points $x^0 \in D^0$ with $\text{grad } f(x^0) = 0$ are called **stationary points** of f . Stationary points are **not** necessarily local extrem values. As an example take

$$f(x, y) := x^2 - y^2$$

with the gradient

$$\text{grad } f(x, y) = 2(x, -y)$$

and therefore with the only stationary point $x^0 = (0, 0)^T$. However, the point x^0 is a **saddel point** of f , i.e. in every neighborhood of x^0 there exist two points x^1 and x^2 with

$$f(x^1) < f(x^0) < f(x^2).$$

Classification of stationary points.

Theorem: Let $f(x)$ be a \mathcal{C}^2 -function on D^0 and let $x^0 \in D^0$ be a stationary point of $f(x)$, i.e. $\text{grad } f(x^0) = 0$.

a) **necessary condition**

If x^0 is a local extrem value of f , then:

x^0 local minimum $\Rightarrow H f(x^0)$ positiv semidefinit

x^0 local maximum $\Rightarrow H f(x^0)$ negativ semidefinit

b) **sufficient condition**

If $H f(x^0)$ is positiv definit (negativ definit) then x^0 is a strict local minimum (maximum) of f .

If $H f(x^0)$ is indefinit then x^0 is a saddle point, i.e. in every neighborhood of x^0 there exist points x^1 and x^2 with $f(x^1) < f(x^0) < f(x^2)$.

Proof of the theorem, part a).

Let x^0 be a local minimum. For $v \neq 0$ and $\varepsilon > 0$ sufficiently small we conclude from the Taylor–expansion

$$f(x^0 + \varepsilon v) - f(x^0) = \frac{1}{2}(\varepsilon v)^T H f(x^0 + \theta \varepsilon v)(\varepsilon v) \geq 0 \quad (1)$$

with $\theta = \theta(\varepsilon, v) \in (0, 1)$.

The gradient in the Taylor expansion $\text{grad } f(x^0) = 0$ vanishes since x^0 is stationary.

From (1) it follows

$$v^T H f(x^0 + \theta \varepsilon v) v \geq 0 \quad (2)$$

Since f is a \mathcal{C}^2 –function, the Hesse–matrix is a **continuous** map. In the limit $\varepsilon \rightarrow 0$ we conclude from (2),

$$v^T H f(x^0) v \geq 0$$

i.e. $H f(x^0)$ is positiv semidefinit.

Proof of the theorem, part b).

If $Hf(x^0)$ is positiv definit, then $Hf(x)$ is positiv definit in a sufficiently small neighborhood $x \in K_\varepsilon(x^0) \subset D$ around x^0 . This follows from the continuity of the second partial derivatives.

For $x \in K_\varepsilon(x^0)$, $x \neq x^0$ we have

$$\begin{aligned} f(x) - f(x^0) &= \frac{1}{2}(x - x^0)^T Hf(x^0 + \theta(x - x^0))(x - x^0) \\ &> 0 \end{aligned}$$

with $\theta \in (0, 1)$, i.e. f has a strict local minimum at x^0 .

If $Hf(x^0)$ is indefinit, then there exist Eigenvectors v, w for Eigenvalues of $Hf(x^0)$ with opposite sign with

$$v^T Hf(x^0)v > 0 \quad w^T Hf(x^0)w < 0$$

and thus x^0 is a saddle point.

Remarks.

- A stationary point x^0 with $\det Hf(x^0) = 0$ is called **degenerate**. The Hesse–matrix has an Eigenvalue $\lambda = 0$.
- If x^0 is **not** degenerate, then there exist 3 cases for the Eigenvalues of $Hf(x^0)$:

all Eigenvalues are strictly positive $\Rightarrow x^0$ is a strict local minimum

all Eigenvalues are strictly negative $\Rightarrow x^0$ is a strict local maximum

there are strictly positive and negative Eigenvalues $\Rightarrow x^0$ saddle point

- The following implications are true (**but not the inverse**)

x^0 local minimum $\Leftarrow x^0$ strict local minimum

\Downarrow

\Uparrow

$Hf(x^0)$ positiv semidefinit $\Leftarrow Hf(x^0)$ positiv definit

Further remarks.

- If f is a C^3 -function, x^0 a stationary point of f and $Hf(x^0)$ positiv definit. Then the following estimate is true:

$$(x - x^0)^T Hf(x^0) (x - x^0) \geq \lambda_{\min} \cdot \|x - x^0\|^2$$

where λ_{\min} denoted the **smallest** Eigenvalue of the Hesse-matrix.

Using the Taylor theorem we obtain:

$$\begin{aligned} f(x) - f(x^0) &\geq \frac{1}{2} \lambda_{\min} \|x - x^0\|^2 + R_3(x; x^0) \\ &\geq \|x - x^0\|^2 \left(\frac{\lambda_{\min}}{2} - C \|x - x^0\| \right) \end{aligned}$$

with an appropriate constant $C > 0$.

The function f grows at least quadratically around x^0 .

Example .

We consider the function

$$f(x, y) := y^2(x - 1) + x^2(x + 1)$$

and look for stationary points :

$$\text{grad } f(x, y) = (y^2 + x(3x + 2), 2y(x - 1))^T$$

The condition $\text{grad } f(x, y) = 0$ gives two stationary points

$$x^0 = (0, 0)^T \quad \text{und} \quad x^1 = (-2/3, 0)^T.$$

The related Hesse-matrices of f at x^0 and x^1 are

$$Hf(x^0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \quad \text{and} \quad Hf(x^1) = \begin{pmatrix} -2 & 0 \\ 0 & -10/3 \end{pmatrix}$$

The matrix $Hf(x^0)$ is indefinit, therefore x^0 is a saddle point. $Hf(x^1)$ is negativ definit and thus x^1 is a strict local ein strenges maximum of f .

2.2 Implicitly defined functions

Aim: study the set of solutions of the system of *non-linear* equations of the form

$$g(x) = 0$$

with $g : D \rightarrow \mathbb{R}^m$, $D \subset \mathbb{R}^n$. I.e. we consider m equations for n unknowns with

$$m < n.$$

Thus: there are **less** equations than unknowns.

We call such a system of equations **underdetermined** and the set of solutions $G \subset \mathbb{R}^n$ contains typically *infinitely* many points.

Solvability of (non-linear) equations.

Question: can we **solve** the system $g(x) = 0$ with respect to certain unknowns, i.e. with respect to the last m variables x_{n-m+1}, \dots, x_n ?

In other words: is there a function $f(x_1, \dots, x_{n-m})$ with

$$g(x) = 0 \iff (x_{n-m+1}, \dots, x_n)^T = f(x_1, \dots, x_{n-m})$$

Terminology: "solve" means express the last m variables by the first $n - m$ variables?

Other question: with respect to which m variables can we solve the system? Is the solution possible *globally* on the domain of definition D ? Or only *locally* on a subdomain $\tilde{D} \subset D$?

Geometrical interpretation: The set of solution G of $g(x) = 0$ can be expressed (at least locally) as graph of a function $f : \mathbb{R}^{n-m} \rightarrow \mathbb{R}^m$.

Example.

The equation for a circle

$$g(x, y) = x^2 + y^2 - r^2 = 0 \quad \text{mit } r > 0$$

defines an **underdetermined** non-linear system of equations since we have **two** unknowns (x, y) , but only **one** scalar equation.

The equation for the circle can be solved **locally** and defines the four functions :

$$y = \sqrt{r^2 - x^2}, \quad -r \leq x \leq r$$

$$y = -\sqrt{r^2 - x^2}, \quad -r \leq x \leq r$$

$$x = \sqrt{r^2 - y^2}, \quad -r \leq y \leq r$$

$$x = -\sqrt{r^2 - y^2}, \quad -r \leq y \leq r$$

Example.

Let g be an affin-linear function, i.e. g has the form

$$g(x) = Cx + b \quad \text{for } C \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

We split the variables x into two vectors

$$x^{(1)} = (x_1, \dots, x_{n-m})^T \in \mathbb{R}^{n-m} \quad \text{and} \quad x^{(2)} = (x_{n-m+1}, \dots, x_n)^T \in \mathbb{R}^m$$

Splitting of the matrix $C = [B, A]$ gives the form

$$g(x) = Bx^{(1)} + Ax^{(2)} + b$$

with $B \in \mathbb{R}^{m \times (n-m)}$, $A \in \mathbb{R}^{m \times m}$.

The system of equations $g(x) = 0$ can be solved (uniquely) with respect to the variables $x^{(2)}$, if A is regular. Then

$$g(x) = 0 \quad \iff \quad x^{(2)} = -A^{-1}(Bx^{(1)} + b) = f(x^{(1)})$$

Continuation of the example.

Question: How can we write the matrix A as dependent of g ?

From the equation

$$g(x) = Bx^{(1)} + Ax^{(2)} + b$$

we see that

$$A = \frac{\partial g}{\partial x^{(2)}}(x^{(1)}, x^{(2)})$$

holds, i.e. A is the Jacobian of the map

$$x^{(2)} \rightarrow g(x^{(1)}, x^{(2)})$$

for fixed $x^{(1)}$!

We conclude: Solvability is given if the Jacobian is regular (invertible).

Implicit function theorem.

Theorem: Let $g : D \rightarrow \mathbb{R}^m$ be a C^1 -function, $D \subset \mathbb{R}^n$ open. We denote the variables in D by (x, y) with $x \in \mathbb{R}^{n-m}$ und $y \in \mathbb{R}^m$. Let $\text{Der } (x^0, y^0) \in D$ be a solution of $g(x^0, y^0) = 0$.

If the Jacobi-matrix

$$\frac{\partial g}{\partial y}(x^0, y^0) := \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(x^0, y^0) & \dots & \frac{\partial g_1}{\partial y_m}(x^0, y^0) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial y_1}(x^0, y^0) & \dots & \frac{\partial g_m}{\partial y_m}(x^0, y^0) \end{pmatrix}$$

is **regular**, then there exist neighborhoods U of x^0 and V of y^0 , $U \times V \subset D$ and a uniquely determined continuous differentiable function $f : U \rightarrow V$ with

$$f(x^0) = y^0 \quad \text{und} \quad g(x, f(x)) = 0 \quad \text{für alle } x \in U$$

and

$$Jf(x) = - \left(\frac{\partial g}{\partial y}(x, f(x)) \right)^{-1} \left(\frac{\partial g}{\partial x}(x, f(x)) \right)$$

Example.

For the equation of a circle $g(x, y) = x^2 + y^2 - r^2 = 0, r > 0$ we have at $(x^0, y^0) = (0, r)$

$$\frac{\partial g}{\partial x}(0, r) = 0, \quad \frac{\partial g}{\partial y}(0, r) = 2r \neq 0$$

Thus we can solve the equation of a circle in a neighborhood of $(0, r)$ with respect to y :

$$f(x) = \sqrt{r^2 - x^2}$$

The derivative $f'(x)$ can be calculated by **implicit differentiation**:

$$g(x, y(x)) = 0 \quad \implies \quad g_x(x, y(x)) + g_y(x, y(x))y'(x) = 0$$

and therefore

$$2x + 2y(x)y'(x) = 0 \quad \implies \quad y'(x) = f'(x) = -\frac{x}{y(x)}$$

Another example.

Consider the equation $g(x, y) = e^{y-x} + 3y + x^2 - 1 = 0$.

It is

$$\frac{\partial g}{\partial y}(x, y) = e^{y-x} + 3 > 0 \quad \text{for all } x \in \mathbb{R}.$$

Therefore the equation can be solved for every $x \in \mathbb{R}$ with respect to $y =: f(x)$ and $f(x)$ is a continuous differentiable function. Implicit differentiation gives

$$e^{y-x}(y' - 1) + 3y' + 2x = 0 \quad \implies \quad y' = \frac{e^{y-x} - 2x}{e^{y-x} + 3}$$

Differentiating again gives

$$e^{y-x}y'' + e^{y-x}(y' - 1)^2 + 3y'' + 2 = 0 \quad \implies \quad y' = -\frac{2 + e^{y-x}(y' - 1)^2}{e^{y-x} + 3}$$

But: Solving the equation with respect to y (in terms of elementary functions) is not possible in this case!

general remark.

Implicit differentiation of a implicitly defined function

$$g(x, y) = 0, \quad \frac{\partial g}{\partial y} \neq 0$$

$y = f(x)$, with $x, y \in \mathbb{R}$, gives

$$f'(x) = -\frac{g_x}{g_y}$$

$$f''(x) = -\frac{g_{xx}g_y^2 - 2g_{xy}g_xg_y + g_{yy}g_x^2}{g_y^3}$$

Therefore the point x^0 is a **stationary** point of $f(x)$ if

$$g(x^0, y^0) = g_x(x^0, y^0) = 0 \quad \text{and} \quad g_y(x^0, y^0) \neq 0$$

And x^0 is a **local maximum** (**minimum**) if

$$\frac{g_{xx}(x^0, y^0)}{g_y(x^0, y^0)} > 0 \quad \left(\text{bzw. } \frac{g_{xx}(x^0, y^0)}{g_y(x^0, y^0)} < 0 \right)$$

Implicit representation of curves.

Consider the set of solutions of a scalar equation

$$g(x, y) = 0$$

If

$$\text{grad } g = (g_x, g_y) \neq 0$$

then $g(x, y)$ defines locally a function $y = f(x)$ or $x = \bar{f}(y)$.

Definition: A solution point (x^0, y^0) of the equation $g(x, y) = 0$ with

- $\text{grad } g(x^0, y^0) \neq 0$ is called **regular** point,
- $\text{grad } g(x^0, y^0) = 0$ is called **singular** point.

Example: Consider (again) the equation for a circle

$$g(x, y) = x^2 + y^2 - r = 0 \quad \text{mit } r > 0.$$

on the circle there are **no** singular points!

Horizontal and vertical tangents.

Remarks:

a) If for a regular point (x^0, y^0) we have

$$g_x(x^0) = 0 \quad \text{und} \quad g_y(x^0) \neq 0$$

then the set of solutions contains a **horizontal tangent** in x^0 .

b) If for a regular point (x^0, y^0) we have

$$g_x(x^0) \neq 0 \quad \text{und} \quad g_y(x^0) = 0$$

then the set of solutions contains a **vertical tangent** in x^0 .

c) If x^0 is a **singular point**, then the set of solutions is approximated at x^0 “in second order” by the following **quadratic equation**

$$g_{xx}(x^0)(x - x^0)^2 + 2g_{xy}(x^0)(x - x^0)(y - y^0) + g_{yy}(x^0)(y - y^0)^2 = 0$$

Remarks.

Due to c) for $g_{xx}, g_{xy}, g_{yy} \neq 0$ we obtain:

$\det Hg(x^0) > 0$: x^0 is an **isolated point** of the set of solutions

$\det Hg(x^0) < 0$: x^0 is a **double point**

$\det Hg(x^0) = 0$: x^0 is a **return point** or a **cusps**

Geometric interpretation:

- If $\det Hg(x^0) > 0$, then both Eigenvalues of $Hg(x^0)$ are or strictly positive or strictly negative, i.e. x^0 is a strict local **minimum** or **maximum** of $g(x)$.
- If $\det Hg(x^0) < 0$, then both Eigenvalues of $Hg(x^0)$ have opposite sign, i.e. x^0 is a **saddle point** of $g(x)$.
- If $\det Hg(x^0) = 0$, then the stationary point x^0 of $g(x)$ is **degenerate**.

Example 1.

Consider the singular point $x^0 = 0$ of the implicit equation

$$g(x, y) = y^2(x - 1) + x^2(x - 2) = 0$$

Calculate the partial derivatives up to order 2:

$$g_x = y^2 + 3x^2 - 4x$$

$$g_y = 2y(x - 1)$$

$$g_{xx} = 6x - 4$$

$$g_{xy} = 2y$$

$$g_{yy} = 2(x - 1)$$

$$Hg(0) = \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix}$$

Therefore $x^0 = 0$ is an **isolated point**.

Example 2.

Consider the singular point $x^0 = 0$ of the implicit equation

$$g(x, y) = y^2(x - 1) + x^2(x + q^2) = 0$$

Calculate the partial derivatives up to order 2:

$$g_x = y^2 + 3x^2 + 2xq^2$$

$$g_y = 2y(x - 1)$$

$$g_{xx} = 6x + 2q^2$$

$$g_{xy} = 2y$$

$$g_{yy} = 2(x - 1)$$

$$Hg(0) = \begin{pmatrix} 2q^2 & 0 \\ 0 & -2 \end{pmatrix}$$

Therefore $x^0 = 0$ is an **double point**.

Example 3.

Consider the singular point $x^0 = 0$ of the implicit equation

$$g(x, y) = y^2(x - 1) + x^3 = 0$$

Calculate the partial derivatives up to order 2:

$$g_x = y^2 + 3x^2$$

$$g_y = 2y(x - 1)$$

$$g_{xx} = 6x$$

$$g_{xy} = 2y$$

$$g_{yy} = 2(x - 1)$$

$$Hg(0) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}$$

Therefore $x^0 = 0$ is a **cuspid** (or a **return point**).

Implicit representation of surfaces.

- The set of solutions of a scalar equation $g(x, y, z) = 0$ for $\text{grad } g \neq 0$ is *locally* a **surface** in \mathbb{R}^3 .
- For the **tangential** in $x^0 = (x^0, y^0, z^0)^T$ with $g(x^0) = 0$ and $\text{grad } g(x^0) \neq 0^T$ we obtain by Taylor expanding (denoting $\Delta x^0 = x - x^0$)

$$\text{grad } g \cdot \Delta x^0 = g_x(x^0)(x - x^0) + g_y(x^0)(y - y^0) + g_z(x^0)(z - z_0) = 0$$

i.e. the gradient is vertical to the surface $g(x, y, z) = 0$.

- If for example $g_z(x^0) \neq 0$, then locally there exists a a representation at x^0 of the form

$$z = f(x, y)$$

and for the **partial derivatives** of $f(x, y)$ we obtain

$$\text{grad } f(x, y) = (f_x, f_y) = -\frac{1}{g_z}(g_x, g_y) = \left(-\frac{g_x}{g_z}, \frac{g_y}{g_z} \right)$$

using the implicit function theorem.

The inverted Problem.

Question: Given the set of equations

$$y = f(x)$$

with $f : D \rightarrow \mathbb{R}^n$, $D \subset \mathbb{R}^n$ open. Can we solve it with respect to x , i.e. can we **invert** the problem?

Theorem: ([Inversion theorem](#))

Let $D \subset \mathbb{R}^n$ be open and $f : D \rightarrow \mathbb{R}^n$ a \mathcal{C}^1 -function. If the Jacobian-matrix $Jf(x^0)$ is regular for an $x^0 \in D$, then there exist neighborhoods U and V of x^0 and $y^0 = f(x^0)$ such that f maps U on V **bijectively**.

The inverse function $f^{-1} : V \rightarrow U$ is also \mathcal{C}^1 and for all $x \in U$ we have:

$$Jf^{-1}(y) = (Jf(x))^{-1}, \quad y = f(x)$$

Remark: We call f locally a \mathcal{C}^1 -diffeomorphism.

Chapter 2. Applications of multivariate differential calculus

2.3 Extrem value problems under constraints

Question: What is the size of a metallic cylindrical can in order to minimize the material amount by given volume?

Ansatz for solution: Let $r > 0$ be the radius and $h > 0$ the height of the can. Then

$$V = \pi r^2 h$$

$$O = 2\pi r^2 + 2\pi rh$$

Let $c \in \mathbb{R}_+$ be the given volume (with $x := r, y := h$),

$$f(x, y) = 2\pi x^2 + 2\pi xy$$

$$g(x, y) = \pi x^2 y - c = 0$$

Determine the minimum of the function $f(x, y)$ on the set

$$G := \{(x, y) \in \mathbb{R}_+^2 \mid g(x, y) = 0\}$$

Solution of the constraint minimisation problem.

From $g(x, y) = \pi x^2 y - c = 0$ follows

$$y = \frac{c}{\pi x^2}$$

We plug this into $f(x, y)$ and obtain

$$h(x) := 2\pi x^2 + 2\pi x \frac{c}{\pi x^2} = 2\pi x^2 + \frac{2c}{x}$$

Determine the minimum of the function $h(x)$:

$$h'(x) = 4\pi x - \frac{2c}{x^2} = 0 \quad \Rightarrow \quad 4\pi x = \frac{2c}{x^2} \quad \Rightarrow \quad x = \left(\frac{c}{2\pi}\right)^{1/3}$$

Sufficient condition

$$h''(x) = 4\pi + \frac{4c}{x^3} \quad \Rightarrow \quad h''\left(\left(\frac{c}{\pi}\right)^{1/3}\right) = 12\pi > 0$$

General formulation of the problem.

Determine the extrem values of the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ under the constraint

$$g(x) = 0$$

where $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

The constraints are

$$g_1(x_1, \dots, x_n) = 0$$

$$\vdots$$

$$g_m(x_1, \dots, x_n) = 0$$

Alternatively: Determine the extrem values of the function $f(x)$ on the set

$$G := \{x \in \mathbb{R}^n \mid g(x) = 0\}$$

The Lagrange–function and the Lagrange–Lemma.

We define the **Lagrange–function**

$$F(x) := f(x) + \sum_{i=1}^m \lambda_i g_i(x)$$

and look for the extrem values of $F(x)$ for fixed $\lambda = (\lambda_1, \dots, \lambda_m)^T$.

The numbers λ_i , $i = 1, \dots, m$ are called **Lagrange–multiplier**.

Theorem: (**Lagrange–Lemma**) If x^0 minimizes (or maximizes) the Lagrange–function $F(x)$ (for a fixed λ) on D and if $g(x^0) = 0$ holds, then x^0 is the minimum (or maximum) of $f(x)$ on $G := \{x \in D \mid g(x) = 0\}$.

Proof: For an arbitrary $x \in D$ we have

$$f(x^0) + \lambda^T g(x^0) \leq f(x) + \lambda^T g(x)$$

If we choose $x \in G$, then $g(x) = g(x^0) = 0$, thus $f(x^0) \leq f(x)$.

A necessary condition for local extrema.

Let f and g_i , $i = 1, \dots, m$, C^1 -functions, then a necessary condition for an extrem value x^0 of $F(x)$ is given by

$$\text{grad } F(x) = \text{grad } f(x) + \sum_{i=1}^m \lambda_i \text{grad } g_i(x) = 0$$

Together with the constraints $g(x) = 0$ we obtain a set of (non-linear) equations with $(n + m)$ equations and $(n + m)$ unknowns x and λ .

The solutions (x^0, λ^0) are the candidates for the extrem values, since these solutions satisfy the above necessary condition.

Alternatively: Define a Lagrange-function

$$G(x, \lambda) := f(x) + \sum_{i=1}^m \lambda_i g_i(x)$$

and look for the extrem values of $G(x, \lambda)$ with respect to x **and** λ .

Some remarks on sufficient conditions.

- 1 We can formulate a **sufficient** condition:
If the functions f and g are \mathcal{C}^2 -functions and if the Hesse-matrix $HF(x^0)$ of the Lagrange-function is positiv (negativ) definit, then x^0 is a strict local minimum (maximum) of $f(x)$ on G .
- 2 In most of the applications the necessary condition are **not** satisfied, although x^0 is a strict local extremum.
- 3 And from the indefinitness of the Hesse-matrix $HF(x^0)$ we **cannot** conclude, that x^0 is not an extremum.
- 4 We have a similar problem with the necessary condition which is obtained from the Hesse-matrix of the Lagrange-function $G(x, \lambda)$ with respect to x **and** λ .

An example of a minimisation problem with constraints.

We look for extrem values of $f(x, y) := xy$ on the disc

$$K := \{(x, y)^T \mid x^2 + y^2 \leq 1\}$$

Since the function f is continuous and $K \subset \mathbb{R}^2$ compact we conclude from the min–max–property the existence of global maxima and minima on K .

We consider first the interior K^0 of K , i.e. the **open** set

$$K^0 := \{(x, y)^T \mid x^2 + y^2 < 1\}$$

The necessary condition for an extrem value is given by

$$\text{grad } f = (y, x) = 0$$

Thus the origin $x^0 = 0$ is a candidate for a (local) extrem value.

continuation of the example.

The Hesse-matrix at the origin is given by

$$Hf(0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and is **indefinit**. Thus x^0 is a **saddel point**.

Therefore the extrem values have to be on the boundary which is represented by a **constraint equation**:

$$g(x, y) = x^2 + y^2 - 1 = 0$$

Therefore we look for the extrem values of $f(x, y) = xy$ under the constraint $g(x, y) = 0$.

The Lagrange-function is given by

$$F(x, y) = xy + \lambda(x^2 + y^2 - 1)$$

Completion of the example.

We obtain the non-linear system of equations

$$y + 2\lambda x = 0$$

$$x + 2\lambda y = 0$$

$$x^2 + y^2 = 1$$

with the four solutions

$$\lambda = \frac{1}{2} \quad : \quad x^{(1)} = (\sqrt{1/2}, -\sqrt{1/2})^T \quad x^{(2)} = (-\sqrt{1/2}, \sqrt{1/2})^T$$

$$\lambda = -\frac{1}{2} \quad : \quad x^{(3)} = (\sqrt{1/2}, \sqrt{1/2})^T \quad x^{(4)} = (-\sqrt{1/2}, -\sqrt{1/2})^T$$

Minima and **Maxima** can be concluded from the **values of the function**

$$f(x^{(1)}) = f(x^{(2)}) = -1/2 \quad f(x^{(3)}) = f(x^{(4)}) = 1/2$$

i.e. minima are $x^{(1)}$ and $x^{(2)}$, maxima are $x^{(3)}$ and $x^{(4)}$.

Lagrange–multiplier–rule.

Satz: Let $f, g_1, \dots, g_m : D \rightarrow \mathbb{R}$ be \mathcal{C}^1 -functions, und let $x^0 \in D$ a local extrem value of $f(x)$ under the constraint $g(x) = 0$. In addition let the **regularity condition**

$$\text{rang} \left(Jg(x^0) \right) = m$$

hold true. Then there exist **Lagrange–multiplier** $\lambda_1, \dots, \lambda_m$, such that for the **Lagrange function**

$$F(x) := f(x) + \sum_{i=1}^m \lambda_i g_i(x)$$

the following **first order necessary condition** holds true:

$$\text{grad} F(x^0) = 0$$

Necessary condition of second order and sufficient condition.

Theorem: 1) Let $x^0 \in D$ a **local minimum** of $f(x)$ under the constraint $g(x) = 0$, let the regularity condition be satisfied and let $\lambda_1, \dots, \lambda_m$ be the related Lagrange–multiplier. Then the Hesse–matrix $HF(x^0)$ of the Lagrange–function is **positiv semi-definit** on the tangential space

$$TG(x^0) := \{y \in \mathbb{R}^n \mid \text{grad } g_i(x^0) \cdot y = 0 \text{ for } i = 1, \dots, m\}$$

i.e. it is $y^T HF(x^0) y \geq 0$ for all $y \in TG(x^0)$.

2) Let the regularity condition for a point $x^0 \in G$ be satisfied. If there exist Lagrange–multiplier $\lambda_1, \dots, \lambda_m$, such that x^0 is a stationary point of the related Lagrange–function. Let the Hesse–matrix $HF(x^0)$ be **positiv definit** on the tangential space $TG(x^0)$, i.e. it holds

$$y^T HF(x^0) y > 0 \quad \forall y \in TG(x^0) \setminus \{0\},$$

then x^0 is a **strict local minimum** of $f(x)$ under the constraint $g(x) = 0$.

Example.

Determine the global maximum of the function

$$f(x, y) = -x^2 + 8x - y^2 + 9$$

under the constraint

$$g(x, y) = x^2 + y^2 - 1 = 0$$

The Lagrange–function is given by

$$F(x, y) = -x^2 + 8x - y^2 + 9 + \lambda(x^2 + y^2 - 1)$$

From the necessary condition we obtain the non-linear system

$$-2x + 8 = -2\lambda x$$

$$-2y = -2\lambda y$$

$$x^2 + y^2 = 1$$

Continuation of the example.

From the necessary condition we obtain the non-linear system

$$-2x + 8 = -2\lambda x$$

$$-2y = -2\lambda y$$

$$x^2 + y^2 = 1$$

The first equation gives $\lambda \neq 1$. Using this in the second equation we get $y = 0$. From the third equation we obtain $x = \pm 1$.

Therefore the two points $(x, y) = (1, 0)$ and $(x, y) = (-1, 0)$ are candidates for a global maximum. Since

$$f(1, 0) = 16 \quad f(-1, 0) = 0$$

the global maximum of $f(x, y)$ under the constraint $g(x, y) = 0$ is given at the point $(x, y) = (1, 0)$.

Another example.

Determine the local extrem values of

$$f(x, y, z) = 2x + 3y + 2z$$

on the intersection of the cylinder surface

$$M_Z := \{(x, y, z)^T \in \mathbb{R}^3 \mid x^2 + y^2 = 2\}$$

with the plane

$$E := \{(x, y, z)^T \in \mathbb{R}^3 \mid x + z = 1\}$$

Reformulation: Determine the extrem values of the function $f(x, y, z)$ under the constraint

$$g_1(x, y, z) := x^2 + y^2 - 2 = 0$$

$$g_2(x, y, z) := x + z - 1 = 0$$

Continuation of the example.

The Jacobi-matrix

$$Jg(x) = \begin{pmatrix} 2x & 2y & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

has rank 2, i.e. we can determine extrem values using the Lagrange-function:

$$F(x, y, z) = 2x + 3y + 2z + \lambda_1(x^2 + y^2 - 2) + \lambda_2(x + z - 1)$$

The necessary condition gives the non-linear system

$$2 + 2\lambda_1 x + \lambda_2 = 0$$

$$3 + 2\lambda_1 y = 0$$

$$2 + \lambda_2 = 0$$

$$x^2 + y^2 = 2$$

$$x + z = 1$$

Continuation of the example.

The necessary condition gives the non-linear system

$$2 + 2\lambda_1 x + \lambda_2 = 0$$

$$3 + 2\lambda_1 y = 0$$

$$2 + \lambda_2 = 0$$

$$x^2 + y^2 = 2$$

$$x + z = 1$$

From the first and the third equation it follows

$$2\lambda_1 x = 0$$

From the second equation it follows $\lambda_1 \neq 0$, i.e. $x = 0$.

Thus we have possible extrem values

$$(x, y, z) = (0, \sqrt{2}, 1) \quad (x, y, z) = (0, -\sqrt{2}, 1)$$

Completion if the example.

The possible extrem values are

$$(x, y, z) = (0, \sqrt{2}, 1) \quad (x, y, z) = (0, -\sqrt{2}, 1)$$

and lie on the cylinder surface M_Z of the cylinder Z with

$$Z = \{(x, y, z)^T \in \mathbb{R}^3 \mid x^2 + y^2 \leq 2\}$$

$$M_Z = \{(x, y, z)^T \in \mathbb{R}^3 \mid x^2 + y^2 = 2\}$$

We calculate the related function values

$$f(0, \sqrt{2}, 1) = 3\sqrt{2} + 2$$

$$f(0, -\sqrt{2}, 1) = -3\sqrt{2} + 2$$

Thus the point $(x, y, z) = (0, \sqrt{2}, 1)$ is a maximum and the point $(x, y, z) = (0, -\sqrt{2}, 1)$ a minimum.

2.4 the Newton–method

Aim: We look for the zero's of a function $f : D \rightarrow \mathbb{R}^n$, $D \subset \mathbb{R}^n$:

$$f(x) = 0$$

- We already know the **fixed-point iteration**

$$x^{k+1} := \Phi(x^k)$$

with starting point x^0 and iteration map $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

- Convergence results are given by the **Banach Fixed Point Theorem**.

Advantage: this method is **derivative-free**.

Disadvantages:

- the numerical scheme converges to slow (only linear),
- there is no unique iteratin map.

The construction of the Newton method.

Starting point: Let \mathcal{C}^1 -function $f : D \rightarrow \mathbb{R}^n$, $D \subset \mathbb{R}^n$ open.

We look for a zero of f , i.e. a $x^* \in D$ with

$$f(x^*) = 0$$

Construction of the Newton-method:

The Taylor-expansion of $f(x)$ at x^0 is given by

$$f(x) = f(x^0) + Jf(x^0)(x - x^0) + o(\|x - x^0\|)$$

Setting $x = x^*$ we obtain

$$Jf(x^0)(x^* - x^0) \approx -f(x^0)$$

An approximative solution for x^* is given by x^1 , $x^1 \approx x^*$, the solution of the linear system of equations

$$Jf(x^0)(x^1 - x^0) = -f(x^0)$$

The Newton–method as algorithm.

The **Newton–method** can be formulated as algorithm.

Algorithm (**Newton–method**):

(1) FOR $k = 0, 1, 2, \dots$

(2a) Solve $Jf(x^k) \cdot \Delta x^k = -f(x^k)$;

(2b) Set $x^{k+1} = x^k + \Delta x^k$;

- In every Newton–step we solve a set of linear equations.
- The solution Δx^k is called **Newton–correction**.
- The Newton–method is **scaling-invariant**.

Scaling-invariance of the Newton–method.

Theorem: the Newton–method is invariant under linear transformations of the form

$$f(x) \rightarrow g(x) = Af(x) \quad \text{for } A \in \mathbb{R}^{n \times n} \text{ regular,}$$

i.e. the iterates for f and g are identical.

Proof: Constructing the Newton–method for $g(x)$, then the Newton–correction is given by

$$\begin{aligned} \Delta x^k &= -(Jg(x^k))^{-1} \cdot g(x^k) \\ &= -(AJf(x^k))^{-1} \cdot Af(x^k) \\ &= -(Jf(x^k))^{-1} \cdot A^{-1}A \cdot f(x^k) \\ &= -(Jf(x^k))^{-1} \cdot f(x^k) \end{aligned}$$

and thus the Newton–correction of f and g coincide.

Using the same starting point x^0 we obtain the same iterates x^k .

Local convergence of the Newton–method.

Theorem: Let $f : D \rightarrow \mathbb{R}^n$ be a \mathcal{C}^1 -function, $D \subset \mathbb{R}^n$ open and convex. Let $x^* \in D$ a zero of f , i.e. $f(x^*) = 0$.

Let the Jacobi–matrix $Jf(x)$ be regular for $x \in D$, and suppose the **Lipschitz–condition**

$$\|(Jf(x))^{-1}(Jf(y) - Jf(x))\| \leq L\|y - x\| \quad \text{for all } x, y \in D,$$

holds true with $L > 0$. Then the Newton–method is well defined for all starting points $x^0 \in D$ with

$$\|x^0 - x^*\| < \frac{2}{L} =: r \quad \text{and} \quad K_r(x^*) \subset D$$

with $x^k \in K_r(x^*)$, $k = 0, 1, 2, \dots$, and the Newton–iterates x^k converge **quadratically** to x^* , i.e.

$$\|x^{k+1} - x^*\| \leq \frac{L}{2} \|x^k - x^*\|^2$$

x^* is the unique zero of $f(x)$ within the ball $K_r(x^*)$.

The damped Newton–method.

Additional observations:

- The Newton–method converges quadratically, but only **locally**.
- **Global** convergence can be obtained - if applicable - by a damping term:

Algorithm (Damped Newton–method):

(1) FOR $k = 0, 1, 2, \dots$

(2a) Solve $Jf(x^k) \cdot \Delta x^k = -f(x^k)$;

(2b) Set $x^{k+1} = x^k + \lambda_k \Delta x^k$;

Frage: How should we choose the **damping parameters** λ_k ?

Choice of the damping parameter.

Strategy: Use a **testfunction** $T(x) = \|f(x)\|$ such that

$$T(x) \geq 0, \quad \forall x \in D$$

$$T(x) = 0 \Leftrightarrow f(x) = 0$$

Choose $\lambda_k \in (0, 1)$ such that the sequence $T(x^k)$ decreases strictly monotonically, i.e.

$$\|f(x^{k+1})\| < \|f(x^k)\| \quad \text{für } k \geq 0.$$

Close to the solution x^* we should choose $\lambda_k = 1$ to guarantee (local) quadratic convergence.

The following Theorem guarantees the existence of damping parameters.

Theorem: Let f a C^1 -function on the open and convex set $D \subset \mathbb{R}^n$. For $x^k \in D$ with $f(x^k) \neq 0$ there exists a $\mu_k > 0$ such that

$$\|f(x^k + \lambda \Delta x^k)\|_2^2 < \|f(x^k)\|_2^2 \quad \text{for all } \lambda \in (0, \mu_k).$$

Damping strategy.

For the **initial iteration** $k = 0$: Choose $\lambda_0 \in \{1, \frac{1}{2}, \frac{1}{4}, \dots, \lambda_{min}\}$ as big as possible such that

$$\|f(x^0)\|_2 > \|f(x^0 + \lambda_0 \Delta x^0)\|_2$$

holds. For **subsequent iterations** $k > 0$: Set $\lambda_k = \lambda_{k-1}$.

IF $\|f(x^k)\|_2 > \|f(x^k + \lambda_k \Delta x^k)\|_2$ **THEN**

- $x^{k+1} := x^k + \lambda_k \Delta x^k$
- $\lambda_k := 2\lambda_k$, falls $\lambda_k < 1$.

ELSE

- Determine $\mu = \max\{\lambda_k/2, \lambda_k/4, \dots, \lambda_{min}\}$ with

$$\|f(x^k)\|_2 > \|f(x^k + \lambda_k \Delta x^k)\|_2$$

- $\lambda_k := \mu$

END

Chapter 3. Integration in higher dimensions

3.1 Area integrals

Given a function $f : D \rightarrow \mathbb{R}$ with domain of definition $D \subset \mathbb{R}^n$.

Aim: Calculate the volume under the graph of $f(x)$:

$$V = \int_D f(x) dx$$

Remember (Analysis II): Riemann-Integral of a function f on the interval $[a, b]$:

$$I = \int_a^b f(x) dx$$

The integral I is defined as limit of Riemann upper- and lower-sums, if the limits exist and coincide.

Construction of area integrals.

Procedure: Same as in the one dimensional case.

But: the domain of definition D is more complex.

Starting point: consider the case of two variables $n = 2$ and a domain of definition $D \subset \mathbb{R}^2$ of the form

$$D = [a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$$

i.e. D is compact cuboid (rectangle).

Let $f : D \rightarrow \mathbb{R}$ be a bounded function.

Definition: We call $Z = \{(x_0, x_1, \dots, x_n), (y_0, y_1, \dots, y_m)\}$ a **partition** of the cuboid $D = [a_1, b_1] \times [a_2, b_2]$ if it holds

$$a_1 = x_0 < x_1 < \dots < x_n = b_1$$

$$a_2 = y_0 < y_1 < \dots < y_m = b_2$$

$Z(D)$ denotes the **set of partitions** of D .

Partitions and Riemann sums.

Definition:

- The **fineness** of a partition $Z \in Z(D)$ is given by

$$\|Z\| := \max_{i,j} \{|x_{i+1} - x_i|, |y_{j+1} - y_j|\}$$

- For a given partition Z the sets

$$Q_{ij} := [x_i, x_{i+1}] \times [y_j, y_{j+1}]$$

are called the **subcuboid** of the partition Z . The **volume** of the subcuboid Q_{ij} is given by

$$\text{vol}(Q_{ij}) := (x_{i+1} - x_i) \cdot (y_{j+1} - y_j)$$

- For arbitrary points $x_{ij} \in Q_{ij}$ of the subcuboids we call

$$R_f(Z) := \sum_{i,j} f(x_{ij}) \cdot \text{vol}(Q_{ij})$$

a **Riemann sum** of the partition Z .

Riemann upper and lower sums.

Definition:

In analogy to the integral for the univariate case we call for a partition Z

$$U_f(Z) := \sum_{i,j} \inf_{x \in Q_{ij}} f(x) \cdot \text{vol}(Q_{ij})$$

$$O_f(Z) := \sum_{i,j} \sup_{x \in Q_{ij}} f(x) \cdot \text{vol}(Q_{ij})$$

the **Riemann lower sum** and the **Riemann upper sum** of $f(x)$, respectively.

Remark:

A Riemann sum for the partition Z lies always between the lower and the upper sum of that partition i.e.

$$U_f(Z) \leq R_f(Z) \leq O_f(Z)$$

Remark.

If a partition Z_2 is obtained from a partition Z_1 by adding additional intermediate points x_i and/or y_j , then

$$U_f(Z_2) \geq U_f(Z_1) \quad \text{and} \quad O_f(Z_2) \leq O_f(Z_1)$$

For arbitrary two partitions Z_1 and Z_2 we always have:

$$U_f(Z_1) \leq O_f(Z_2)$$

Question: what happens to the lower and upper sums in the limit $\|Z\| \rightarrow 0$:

$$U_f := \sup\{U_f(Z) : Z \in \mathcal{Z}(D)\}$$

$$O_f := \inf\{O_f(Z) : Z \in \mathcal{Z}(D)\}$$

Observation: Both values U_f and O_f exist since lower and upper sum are monoton and bounded.

Riemann upper and lower integrals.

Definition:

- ① The **Riemann lower and upper integral** of a function $f(x)$ on D is given by

$$\int_{\underline{D}} f(x) dx := \sup\{U_f(Z) : Z \in \mathcal{Z}(D)\}$$

$$\int_{\overline{D}} f(x) dx := \inf\{O_f(Z) : Z \in \mathcal{Z}(D)\}$$

- ② The function $f(x)$ is called **Riemann-integrable** on D , if lower and upper intergral conincide. The **Riemann-integral** of $f(x)$ on D is then given by

$$\int_D f(x) dx := \int_{\underline{D}} f(x) dx = \int_{\overline{D}} f(x) dx$$

Remark.

Up to now we have "only" considered the case of **two** variables:

$$f : D \rightarrow \mathbb{R}, \quad D \in \mathbb{R}^2$$

In higher dimensions, $n > 2$, the procedure is the same.

Notation: for $n = 2$ and $n = 3$

$$\int_D f(x, y) dx dy \quad \text{bzw.} \quad \int_D f(x, y, z) dx dy dz$$

or

$$\iint_D f(x, y) dx dy \quad \text{bzw.} \quad \iiint_D f(x, y, z) dx dy dz$$

respectively.

Elementary properties of the integral.

Theorem:

a) Linearity

$$\int_D (\alpha f(x) + \beta g(x)) dx = \alpha \int_D f(x) dx + \beta \int_D g(x) dx$$

b) Monotonicity

If $f(x) \leq g(x)$ for all $x \in D$, then:

$$\int_D f(x) dx \leq \int_D g(x) dx$$

c) Positivity

If for all $x \in D$ the relation $f(x) \geq 0$ holds, i.e. $f(x)$ is **non-negative**, then

$$\int_D f(x) dx \geq 0$$

Additional properties of the integral.

Theorem:

- a) Let D_1 , D_2 and D be cuboids, $D = D_1 \cup D_2$ and $\text{vol}(D_1 \cap D_2) = 0$, then $f(x)$ is on D integrable if and only if $f(x)$ is integrable on D_1 and D_2 . And we have

$$\int_D f(x) dx = \int_{D_1} f(x) dx + \int_{D_2} f(x) dx$$

- b) The following **estimate** holds for the integral

$$\left| \int_D f(x) dx \right| \leq \sup_{x \in D} |f(x)| \cdot \text{vol}(D)$$

- c) **Riemann criterion**

$f(x)$ is integrable on D if and only if :

$$\forall \varepsilon > 0 \quad \exists Z \in \mathcal{Z}(D) \quad : \quad O_f(Z) - U_f(Z) < \varepsilon$$

Fubini's theorem.

Theorem: (Fubini's theorem) Let $f : D \rightarrow \mathbb{R}$ be integrable, $D = [a_1, b_1] \times [a_2, b_2]$ be a cuboid. If the integrals

$$F(x) = \int_{a_2}^{b_2} f(x, y) dy \quad \text{und} \quad G(y) = \int_{a_1}^{b_1} f(x, y) dx$$

exist for all $x \in [a_1, b_1]$ and $y \in [a_2, b_2]$, respectively, then

$$\int_D f(x) dx = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y) dy dx$$

$$\int_D f(x) dx = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y) dx dy$$

holds true.

Importance:

Fubini's theorem allows to reduce higher-dimensional integrals to one-dimensional integrals.

Example.

Given the cuboid $D = [0, 1] \times [0, 2]$ and the function

$$f(x, y) = 2 - xy$$

We will show that continuous functions are integrable on cuboids. Thus we can apply Fubini's theorem:

$$\begin{aligned} \int_D f(x, y) dx dy &= \int_0^2 \int_0^1 f(x, y) dx dy = \int_0^2 \left[2x - \frac{x^2 y}{2} \right]_{x=0}^{x=1} dy \\ &= \int_0^2 \left(2 - \frac{y}{2} \right) dy = \left[2y - \frac{y^2}{4} \right]_{y=0}^{y=2} = 3 \end{aligned}$$

Remark: Fubini's theorem requires the integrability of $f(x, y)$. The existence of the two integrals $F(x)$ and $G(y)$ does **not** guarantee the integrability of $f(x, y)$!

The characteristic function.

Definition: Let $D \subset \mathbb{R}^n$ compact and $f : D \rightarrow \mathbb{R}$ bounded. We set

$$f^*(x) := \begin{cases} f(x) & : \text{if } x \in D \\ 0 & : \text{if } x \in \mathbb{R}^n \setminus D \end{cases}$$

In particular for $f(x) = 1$ we call $f^*(x)$ the **characteristic function** of D . The characteristic function of D is called $\mathcal{X}_D(x)$.

Let Q be the smallest cuboid with $D \subset Q$. The function $f(x)$ is called **integrable** on D , if $f^*(x)$ is integrable on Q . We set

$$\int_D f(x) dx := \int_Q f^*(x) dx$$

Measurability and null sets.

Definition: The compact set $D \subset \mathbb{R}^n$ is called **measurable**, if the integral

$$\text{vol}(D) := \int_D 1 dx = \int_Q \chi_D(x) dx$$

exists. We call $\text{vol}(D)$ the **volume** of D in \mathbb{R}^n .

The compact set D is called **null set**, if D is measurable and if $\text{vol}(D) = 0$ holds.

Remark:

- If D a cuboid, then $Q = D$ and thus

$$\int_D f(x) dx = \int_Q f^*(x) dx = \int_Q f(x) dx$$

i.e. the introduced concepts of integrability coincide.

- Cuboids are measurable sets.
- $\text{vol}(D)$ is the volume of the cuboid on \mathbb{R}^n .

Three more properties of integration.

We have the following theorems for integrals in higher dimensions.

Theorem: Let $D \subset \mathbb{R}^n$ be compact. D is measurable if and only if the boundary ∂D of D is a null set.

Theorem: Let $D \subset \mathbb{R}^n$ be compact and measurable. Let $f : D \rightarrow \mathbb{R}$ be continuous. Then $f(x)$ is integrable on D .

Theorem: (Mean value theorem) Let $D \subset \mathbb{R}^n$ be compact, connected and measurable, and let $f : D \rightarrow \mathbb{R}$ be continuous, then there exist a point $\xi \in D$ with

$$\int_D f(x) dx = f(\xi) \cdot \text{vol}(D)$$

"Normal" areas.

Definition:

- A subset $D \subset \mathbb{R}^2$ is called "normal" area, there exist continuous functions g, h and \tilde{g}, \tilde{h} with

$$D = \{(x, y) \mid a \leq x \leq b \text{ und } g(x) \leq y \leq h(x)\}$$

and

$$D = \{(x, y) \mid \tilde{a} \leq y \leq \tilde{b} \text{ und } \tilde{g}(y) \leq x \leq \tilde{h}(y)\}$$

respectively.

- A subset $D \subset \mathbb{R}^3$ is called "normal" area, if there is a representation

$$D = \{(x_1, x_2, x_3) \mid a \leq x_i \leq b, g(x_i) \leq x_j \leq h(x_i) \\ \text{and } \varphi(x_i, x_j) \leq x_k \leq \psi(x_i, x_j)\}$$

with a permutation (i, j, k) of $(1, 2, 3)$ and continuous functions g, h, φ and ψ .

Projectable sets.

Definition: A subset $D \subset \mathbb{R}^n$ is called **projectable** in the direction x_i , $i \in \{1, \dots, n\}$, if there exist a measurable set $B \subset \mathbb{R}^{n-1}$ and continuous functions φ, ψ such that

$$D = \{ x \in \mathbb{R}^n \mid \tilde{x} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)^T \in B \\ \text{und } \varphi(\tilde{x}) \leq x_i \leq \psi(\tilde{x}) \}$$

Remark:

- Projectable sets are measurable sets. Since "normal" areas are projectable, "normal" areas are measurable.
- Often the area of integration D can be represented by a union of finite many "normal" areas. Such areas are then also measurable.

Integration on "normal" areas and projectable sets.

Theorem: If $f(x)$ is a **continuous** function on a "normal" area

$$D = \{ (x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ and } g(x) \leq y \leq h(x) \}$$

then we have

$$\int_D f(x) dx = \int_a^b \int_{g(x)}^{h(x)} f(x, y) dy dx$$

Analogous relations hold in higher dimensions: If $D \subset \mathbb{R}^n$ is a **projectable set** in the direction x_i , i.e. D has a representation of the form

$$D = \{ x \in \mathbb{R}^n \mid \tilde{x} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)^T \in B$$

$$\text{and } \varphi(\tilde{x}) \leq x_i \leq \psi(\tilde{x}) \}$$

then it holds

$$\int_D f(x) dx = \int_B \left(\int_{\varphi(\tilde{x})}^{\psi(\tilde{x})} f(x) dx_i \right) d\tilde{x}$$

Example.

Given a function

$$f(x, y) := x + 2y$$

Calculate the integral on the area bounded by two parabolas

$$D := \{(x, y) \mid -1 \leq x \leq 1 \text{ und } x^2 \leq y \leq 2 - x^2\}$$

The set D is a "normal" area and $f(x, y)$ is continuous. Thus

$$\begin{aligned} \int_D f(x, y) dx &= \int_{-1}^1 \left(\int_{x^2}^{2-x^2} (x + 2y) dy \right) dx = \int_{-1}^1 [xy + y^2]_{x^2}^{2-x^2} dx \\ &= \int_{-1}^1 (x(2 - x^2) + (2 - x^2)^2 - x^3 - x^4) dx \\ &= \int_{-1}^1 (-2x^3 - 4x^2 + 2x + 4) dx = \frac{16}{3} \end{aligned}$$

Example.

Calculate the volume of the **rotational paraboloid**

$$V := \{(x, y, z)^T \mid x^2 + y^2 \leq 1 \text{ and } x^2 + y^2 \leq z \leq 1\}$$

Representation of V as "normal" area

$$V = \{(x, y, z)^T \mid -1 \leq x \leq 1, -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2} \text{ and } x^2 + y^2 \leq z \leq 1\}$$

Then we have

$$\begin{aligned} \text{vol}(V) &= \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{x^2+y^2}^1 dz dy dx = \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} (1 - x^2 - y^2) dy dx \\ &= \int_{-1}^1 \left[(1 - x^2)y - \frac{y^3}{3} \right]_{y=-\sqrt{1-x^2}}^{y=\sqrt{1-x^2}} dx = \frac{4}{3} \int_{-1}^1 (1 - x^2)^{3/2} dx \\ &= \frac{1}{3} \left[x(\sqrt{1-x^2})^3 + \frac{3}{2}x\sqrt{1-x^2} + \frac{3}{2}\arcsin(x) \right]_{-1}^1 = \frac{\pi}{2} \end{aligned}$$

Integration over arbitrary domains.

Definition: Let $D \subset \mathbb{R}^n$ be a compact and measurable set. We call $Z = \{D_1, \dots, D_m\}$ an **universal partition** of D , if the sets D_k are compact, measurable and connected and if

$$\bigcup_{j=1}^m D_j = D \quad \text{and} \quad \forall i \neq j : D_i^0 \cap D_j^0 = \emptyset.$$

We call

$$\text{diam}(D_j) := \sup \{ \|x - y\| \mid x, y \in D_j \}$$

the **diameter** of the set D_j and

$$\|Z\| := \max \{ \text{diam}(D_j) \mid j = 1, \dots, m \}$$

the **fineness** of the universal partition Z .

Riemann sums for universal partitions.

For a continuous function $f : D \rightarrow \mathbb{R}$ we define the **Riemann sums**

$$R_f(Z) = \sum_{j=1}^m f(x^j) \operatorname{vol}(D_j)$$

with arbitrary $x^j \in D_j$, $j = 1, \dots, m$.

Theorem: For any sequence $(Z_k)_{k \in \mathbb{N}}$ of universal partitions of D with $\|Z_k\| \rightarrow 0$ (as $k \rightarrow \infty$) and for any sequence of related Riemann sums $R_f(Z_k)$ we have

$$\lim_{k \rightarrow \infty} R_f(Z_k) = \int_D f(x) dx$$

Center (of mass) of areas and solids.

An important **application** of the area integrals is the calculation of the **centers (of mass)** of areas and solids.

Definition: Let $D \subset \mathbb{R}^2$ (or \mathbb{R}^3) be a measurable set and $\rho(x)$, $x \in D$, a given mass density. Then the **center (of mass)** of the area (or the solid) D is given by

$$x_s := \frac{\int_D \rho(x)x dx}{\int_D \rho(x) dx}$$

The numerator integral (over a vector valued function) is intended componentwise (and gives as result a vector).

Example.

Calculate the center of mass of the pyramid P

$$P := \left\{ (x, y, z)^T \mid \max(|y|, |z|) \leq \frac{ax}{2h}, \quad 0 \leq x \leq h \right\}$$

Calculate the volume of P under assumption of constant mass density

$$\begin{aligned} \text{vol}(P) &= \int_0^h \int_{-\frac{ax}{2h}}^{\frac{ax}{2h}} \int_{-\frac{ax}{2h}}^{\frac{ax}{2h}} dz dy dx \\ &= \int_0^h \int_{-\frac{ax}{2h}}^{\frac{ax}{2h}} \frac{ax}{h} dy dx \\ &= \int_0^h \left(\frac{ax}{h}\right)^2 dx = \frac{1}{3} a^2 h \end{aligned}$$

Continuation of the example.

and

$$\begin{aligned} \int_0^h \int_{-\frac{ax}{2h}}^{\frac{ax}{2h}} \int_{-\frac{ax}{2h}}^{\frac{ax}{2h}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} dz dy dx &= \int_0^h \int_{-\frac{ax}{2h}}^{\frac{ax}{2h}} \begin{pmatrix} \frac{ax^2}{h} \\ \frac{axy}{h} \\ 0 \end{pmatrix} dy dx \\ &= \int_0^h \begin{pmatrix} \frac{a^2x^3}{h^2} \\ 0 \\ 0 \end{pmatrix} dx \\ &= \begin{pmatrix} \frac{1}{4}a^2h^2 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

The center of mass of P lies in the point $x_S = \left(\frac{3}{4}h, 0, 0\right)^T$.

Moments of inertia of areas and solids.

Another important **application** of area integrals is the calculation of **moments of inertia** of areas and solids.

Definition: (moments of inertia with respect to an axis)

Let $D \subset \mathbb{R}^2$ (or \mathbb{R}^3) be a measurable set, $\rho(x)$ denotes for $x \in D$ a mass density and $r(x)$ the distance of the point $x \in D$ from the given axis of rotation.

Then the moment of inertia of D with respect to this axis is given by

$$\Theta := \int_D \rho(x)r^2(x)dx$$

Example.

We calculate the moment of inertia of a **homogeneous cylinder**

$$Z := \{ (x, y, z)^T : x^2 + y^2 \leq r^2, -l/2 \leq z \leq l/2 \}$$

with respect to the x -axis assuming a constant density ρ .

$$\begin{aligned} \Theta &= \int_Z \rho(y^2 + z^2) d(x, y, z) = \rho \int_Z (y^2 + z^2) d(x, y, z) \\ &= \rho \int_{-r}^r \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \int_{-l/2}^{l/2} (y^2 + z^2) dz dy dx \\ &= \rho \int_{-r}^r \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \left(ly^2 + \frac{l^3}{12} \right) dy dx \\ &= \rho \frac{\pi lr^2}{12} (3r^2 + l^2) \end{aligned}$$

The theorem of transformation.

Aim: A generalisation of the (one dimensional) [rule of substitution](#)

$$\int_{\varphi(a)}^{\varphi(b)} f(x) dx = \int_a^b f(\varphi(t))\varphi'(t) dt$$

Theorem: ([Theorem of transformation](#)) Let $\Phi : U \rightarrow \mathbb{R}^n$, $U \subset \mathbb{R}^n$ be open and a \mathcal{C}^1 -map. Let $D \subset U$ be a compact, measurable set such that Φ is a \mathcal{C}^1 -diffeomorphism on D° . Then $\Phi(D)$ is compact and measurable and for any continuous function $f : \Phi(D) \rightarrow \mathbb{R}$ the [rule of transformation](#)

$$\int_{\Phi(D)} f(x) dx = \int_D f(\Phi(u)) |\det J\Phi(u)| du$$

holds.

Remark: Note that the rule of transformation requires the bijectivity of Φ only on the interior D° of D – not on the boundary ∂D !

Example.

Calculate the center of mass of a homogeneous **spherical octant**

$$V = \{(x, y, z)^T \mid x^2 + y^2 + z^2 \leq 1 \text{ und } x, y, z \geq 0\}$$

It is easier to calculate the center of mass using **spherical coordinates**
Kugelkoordinaten:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cos \varphi \cos \psi \\ r \sin \varphi \cos \psi \\ r \sin \psi \end{pmatrix} = \Phi(r, \varphi, \psi)$$

The transformation is defined on \mathbb{R}^3 and with

$$D = [0, 1] \times \left[0, \frac{\pi}{2}\right] \times \left[0, \frac{\pi}{2}\right]$$

we have $\Phi(D) = V$. It is Φ on D^0 a \mathcal{C}^1 -diffeomorphism with

$$\det J\Phi(r, \varphi, \psi) = r^2 \cos \psi$$

Continuation of the example.

According to the theorem of transformation it follows

$$\text{vol}(V) = \int_V dx = \int_0^1 \int_0^{\pi/2} \int_0^{\pi/2} r^2 \cos \psi \, d\psi \, d\varphi \, dr = \frac{\pi}{6}$$

and

$$\begin{aligned} \text{vol}(V) \cdot x_s &= \int_V x \, dx = \int_0^1 \int_0^{\pi/2} \int_0^{\pi/2} (r \cos \varphi \cos \psi) r^2 \cos \psi \, d\psi \, d\varphi \, dr \\ &= \int_0^1 r^3 \, dr \cdot \int_0^{\pi/2} \cos \varphi \, d\varphi \cdot \int_0^{\pi/2} \cos^2 \psi \, d\psi = \frac{\pi}{16} \end{aligned}$$

The it follows $x_s = \frac{3}{8}$.

In Analogy we calculate $y_s = z_s = \frac{3}{8}$.

The Theorem of Steiner.

Theorem: (Theorem of Steiner) For the moment of inertia of a homogeneous solid K with total mass m with respect to a given axis of rotation A we have

$$\Theta_A = md^2 + \Theta_S$$

S is the axis through to center of mass of the solid K parallel to the axis A and d the distance of the center of mass x_S from the axis A .

Idea of the proof: Set $x := \Phi(u) = x_S + u$. Then with the unit vector a in direction of the axis A

$$\begin{aligned}\Theta_A &= \rho \int_K (\langle x, x \rangle - \langle x, a \rangle^2) dx \\ &= \rho \int_D (\langle x_S + u, x_S + u \rangle - \langle x_S + u, a \rangle^2) dx\end{aligned}$$

where

$$D := \{x - x_S \mid x \in K\}$$