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 Wintersemster 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 rapresentaion of curves and surfces.
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.

1.1 Partial derivatives

Let

$$f(x_1, \ldots, x_n) \text{ a scalar function depending } n \text{ variables}$$

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

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

$$\begin{aligned} p &= p(V, t) = \frac{RT}{V} \\ V &= V(p, T) = \frac{RT}{p} \\ T &= T(p, V) = \frac{pV}{R} \end{aligned}$$

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

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

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

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

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

Thus f is a $\mathcal{C}^1$–function.

- The function

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

at $\mathrm{x}^0 = (0,0)^T$ 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

$$\begin{aligned}
\frac{\partial}{\partial x_i}\Big(\alpha f(\mathrm{x}) + \beta g(\mathrm{x})\Big) &= \alpha \frac{\partial f}{\partial x_i}(\mathrm{x}) + \beta \frac{\partial g}{\partial x_i}(\mathrm{x}) \\
\frac{\partial}{\partial x_i}\Big(f(\mathrm{x}) \cdot g(\mathrm{x})\Big) &= \frac{\partial f}{\partial x_i}(\mathrm{x}) \cdot g(\mathrm{x}) + f(\mathrm{x}) \cdot \frac{\partial g}{\partial x_i}(\mathrm{x}) \\
\frac{\partial}{\partial x_i}\left(\frac{f(\mathrm{x})}{g(\mathrm{x})}\right) &= \frac{\frac{\partial f}{\partial x_i}(\mathrm{x}) \cdot g(\mathrm{x}) - f(\mathrm{x}) \cdot \frac{\partial g}{\partial x_i}(\mathrm{x})}{g(\mathrm{x})^2} \quad \text{for } g(\mathrm{x}) \neq 0
\end{aligned}$$

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

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

Gradient and nabla–operator.

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

- We denote the **row vector**

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

  as gradient of f at x^0.

- We denote the symbolic vector

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

  as nabla–operator.

- Thus we obtain the **column vector**

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

More rules on differentiation.

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

$$\begin{aligned} \operatorname{grad}(\alpha f + \beta g) &= \alpha \cdot \operatorname{grad} f + \beta \cdot \operatorname{grad} g \\ \operatorname{grad}(f \cdot g) &= g \cdot \operatorname{grad} f + f \cdot \operatorname{grad} g \\ \operatorname{grad}\left(\frac{f}{g}\right) &= \frac{1}{g^2}(g \cdot \operatorname{grad} f - f \cdot \operatorname{grad} g), \quad g \neq 0 \end{aligned}$$

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

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

  where $\mathrm{x} = (x_1, \ldots, x_n)$ 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 $f : \mathbb{R}^2 \to \mathbb{R}$ defined as

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

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

$$\begin{aligned} f_x(0,0) &= f_y(0,0) = 0 \\ \frac{\partial f}{\partial x}(x,y) &= \frac{y}{(x^2+y^2)^2} - 4\frac{x^2 y}{(x^2+y^2)^3}, \quad (x,y) \neq (0,0) \\ \frac{\partial f}{\partial y}(x,y) &= \frac{x}{(x^2+y^2)^2} - 4\frac{x y^2}{(x^2+y^2)^3}, \quad (x,y) \neq (0,0) \end{aligned}$$

Example (continuation).

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

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

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

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

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

and thus we have

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

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 $f : D \to \mathbb{R}$ be partial differentiable in a neighborhood of $\mathrm{x}^0 \in D$ and let the partial derivatives $\dfrac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, 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:

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

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

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

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

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

Using the boundedness of the partial derivatives

$$|f(\mathrm{x}) - f(\mathrm{x}^0)| \leq C_1|x_1 - x_1^0| + \cdots + C_n|x_n - x_n^0|$$

for $\|\mathrm{x} - \mathrm{x}^0\|_\infty < \varepsilon$, we obtain the continuity of f at x^0 since

$$f(\mathrm{x}) \to f(\mathrm{x}^0) \qquad \text{für } \|\mathrm{x} - \mathrm{x}^0\|_\infty \to 0$$

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

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

Higher order derivatives.

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

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

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 $\mathcal{C}^k$–function on D. Continuous functions f are called $\mathcal{C}^0$–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 derivaratives are not arbitrarely exchangeable.

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

Example: For the function

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

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 exchangeablity.

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

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

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

Idea of the proof:

Apply the men 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 arbitrarely!

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 exchangealbe 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):

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

- 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 $u(x) = u(x_1, \ldots, x_n)$ we have

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

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

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

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

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

Vector valued functions.

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

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

$$\frac{\partial \mathrm{f}}{\partial x_i}(\mathrm{x}^0) = \lim_{t \to 0} \frac{\mathrm{f}(\mathrm{x}^0 + t\mathrm{e}_i) - \mathrm{f}(\mathrm{x}^0)}{t}$$

exist. The calculation is done componentwise

$$\frac{\partial \mathrm{f}}{\partial x_i}(\mathrm{x}^0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_i} \\ \vdots \\ \dfrac{\partial f_m}{\partial x_i} \end{pmatrix} \qquad \text{for } i = 1, \ldots, n$$

Vectorfields.

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

Examples of vectorfields:

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

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

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

or

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

Rules of computation and the rotation.

The following rules hold true:

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

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

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

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

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

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

Alternative notations and additional rules.

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

Remark: The following rules hold true:

$$\begin{aligned} \text{rot}\,(\alpha\, f + \beta\, g) &= \alpha\, \text{rot}\, f + \beta\, \text{rot}\, g \\ \text{rot}\,(\varphi \cdot f) &= (\nabla\varphi) \times f + \varphi\, \text{rot}\, f \end{aligned}$$

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

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

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

1.2 The total differential

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

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

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 $J f(x^0)$ (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 vextor and $a(x - x^0)$ a scalar product $\langle a^T, x - x^0 \rangle$.

Total and partial differentiability.

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

a) If $\mathrm{f}(\mathrm{x})$ is differentiable in x^0, then $\mathrm{f}(\mathrm{x})$ is continuous in x^0.

b) If $\mathrm{f}(\mathrm{x})$ is differentiable in x^0, then the (total) differential and thus the Jacobi–matrix are uniquely determined and we have

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

c) If $\mathrm{f}(\mathrm{x})$ is a $\mathcal{C}^1$–function on D, then $\mathrm{f}(\mathrm{x})$ is differentiable on D.

Proof of a).

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

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

Thus we conclude

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

and we obtain

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

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 in differentiable at x^0, we have

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

We write

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

Thus

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

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 $\mathrm{f} : \mathbb{R}^3 \to \mathbb{R}^2$ defined by
$$\mathrm{f}(x_1, x_2, x_3) = \begin{pmatrix} x_1 x_2 x_3 \\ \sin(x_1 + 2x_2 + 3x_3) \end{pmatrix}$$

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

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

Further examples.

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

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

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

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

  We conclude

$$\mathrm{J}f(\mathrm{x}) = \mathrm{grad} f(\mathrm{x}) = \mathrm{x}^T (\mathrm{A}^T + \mathrm{A})$$

Rules for the differentiation.

Theorem:

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

$$\begin{aligned} \mathrm{d}(\alpha\mathrm{f} + \beta\mathrm{g})(\mathrm{x}^0) &= \alpha\,\mathrm{df}(\mathrm{x}^0) + \beta\,\mathrm{dg}(\mathrm{x}^0) \\ \mathrm{J}(\alpha\mathrm{f} + \beta\mathrm{g})(\mathrm{x}^0) &= \alpha\,\mathrm{Jf}(\mathrm{x}^0) + \beta\,\mathrm{Jg}(\mathrm{x}^0) \end{aligned}$$

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

For the differentials it holds

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

and analogously for the Jacobian matrix

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

Examples for the chain rule.

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

Then the composition

$$(f \circ \mathrm{h})(t) = f(h_1(t), \ldots, h_n(t))$$

is differentiable in t_0 and we have for the derivative:

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

Directional derivative.

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

$$D_{\mathrm{v}} f(\mathrm{x}^0) := \lim_{t \to 0} \frac{f(\mathrm{x}^0 + t\mathrm{v}) - f(\mathrm{x}^0)}{t}$$

is called the directional derivative (Gateaux–derivative) of $f(\mathrm{x})$ in the direction of v.

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

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

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

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

This follows directely applying the chain rule.

Properties of the gradient.

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

a) The gradient vector $\operatorname{grad} f(x^0) \in \mathbb{R}^n$ 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 $\operatorname{grad} f(x^0)$ 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

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

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

Curvilinear coordinates.

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

In addition there exists the inverse map $\Phi^{-1} : V \to U$ and the inverse map is also a $\mathcal{C}^1$–map.

Then $\mathrm{x} = \Phi(\mathrm{u})$ defines a coodinate transformation from the coordinates u to x.

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

$$x = r \cos \varphi$$

$$y = r \sin \varphi$$

with the cartesian coordinates $\mathrm{x} = (x, y)$.

Calculation of the partial derivatives.

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

$$\begin{aligned} \Phi^{-1}(\Phi(u)) &= u \\ J\,\Phi^{-1}(x) \cdot J\,\Phi(u) &= I_n \qquad \text{(chain rule)} \\ J\,\Phi^{-1}(x) &= (J\,\Phi(u))^{-1} \end{aligned}$$

Let $\tilde{f} : V \to \mathbb{R}$ 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

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

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

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

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

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

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

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

The Jacobian–matrix is given by:

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

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

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

1.3 Mean value theorems and Taylor expansion

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

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

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

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

Proof: We set

$$h(t) := f(\mathrm{a} + t(\mathrm{b} - \mathrm{a}))$$

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

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

Definition and example.

Definition: If the condition $[\mathrm{a}, \mathrm{b}] \subset D$ holds true for **all** points $\mathrm{a}, \mathrm{b} \in D$, 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

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

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

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

It is

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

and

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

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

A mean value estimate for vector–valued functions.

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

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

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

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

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

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

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

$$|\alpha| := \alpha_1 + \dots + \alpha_n \qquad \alpha! := \alpha_1! \cdot \dots \cdot \alpha_n!$$

Let $f : D \to \mathbb{R}$ 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

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

The Taylor theorem.

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

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

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

$$g(t) := f(\mathrm{x}_0 + t(\mathrm{x} - \mathrm{x}_0))$$

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

$$\begin{aligned} g'(0) &= \frac{d}{dt} f(x_1^0 + t(x_1 - x_1^0), x_2^0 + t(x_2 - x_2^0), \ldots, x_n^0 + t(x_n - x_n^0))\Big|_{t=0} \\ &= D_1 f(\mathrm{x}_0) \cdot (x_1 - x_1^0) + \ldots + D_n f(\mathrm{x}_0) \cdot (x_n - x_n^0) \\ &= \sum_{|\alpha|=1} \frac{D^\alpha f(\mathrm{x}_0)}{\alpha!} \cdot (\mathrm{x} - \mathrm{x}_0)^\alpha \end{aligned}$$

Continuation of the derivation.

Calculation of $g''(0)$ gives

$$\begin{aligned} g''(0) &= \frac{d^2}{dt^2} f(\mathrm{x}_0 + t(\mathrm{x}-\mathrm{x}_0))\Big|_{t=0} = \frac{d}{dt}\sum_{k=1}^{n} D_k f(\mathrm{x}^0 + t(\mathrm{x}-\mathrm{x}^0))(x_k - x_k^0)\Big|_{t=0} \\ &= D_{11} f(\mathrm{x}_0)(x_1 - x_1^0)^2 + D_{21} f(\mathrm{x}_0)(x_1 - x_1^0)(x_2 - x_2^0) \\ &\quad + \ldots + D_{ij} f(\mathrm{x}_0)(x_i - x_i^0)(x_j - x_j^0) + \ldots + \\ &\quad + D_{n-1,n} f(\mathrm{x}_0)(x_{n-1} - x_{n-1}^0)(x_n - x_n^0) + D_{nn} f(\mathrm{x}_0)(x_n - x_n^0)^2) \\ &= \sum_{|\alpha|=2} \frac{D^\alpha f(\mathrm{x}_0)}{\alpha!}(\mathrm{x}-\mathrm{x}_0)^\alpha \qquad \text{(exchange theorem of Schwarz!)} \end{aligned}$$

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

Proof of the Taylor theorem.

The function

$$g(t) := f(\mathrm{x}^0 + t(\mathrm{x} - \mathrm{x}^0))$$

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)

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

and

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

Examples for the Taylor–expansion.

1. Calculate the Taylor–polynom $T_2(\mathrm{x};\mathrm{x}_0)$ 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 $T_2(\mathrm{x};\mathrm{x}_0)$ 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 $T_2(\mathrm{x};\mathrm{x}_0)$ in the form

$$T_2(\mathrm{x};\mathrm{x}_0) = 4z(x+y-2)$$

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 $(m+1)$:

$$R_m(\mathrm{x};\mathrm{x}_0) = \sum_{|\alpha|=m+1} \frac{D^\alpha f(\mathrm{x}_0 + \theta(\mathrm{x}-\mathrm{x}_0))}{\alpha!}(\mathrm{x}-\mathrm{x}_0)^\alpha$$

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

$$|R_m(\mathrm{x};\mathrm{x}_0)| \le \frac{n^{m+1}}{(m+1)!}\, C\, \|\mathrm{x}-\mathrm{x}_0\|_\infty^{m+1}$$

We conlude for the quality of the approximation of a $\mathcal{C}^{m+1}$–function by the Taylor–polynom

$$f(\mathrm{x}) = T_m(\mathrm{x};\mathrm{x}_0) + O\left(\|\mathrm{x}-\mathrm{x}_0\|^{m+1}\right)$$

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

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

The Hesse–matrix.

The matrix

$$\mathrm{H}f(\mathrm{x}_0) := \begin{pmatrix} f_{x_1x_1}(\mathrm{x}_0) & \dots & f_{x_1x_n}(\mathrm{x}_0) \\ \vdots & & \vdots \\ f_{x_nx_1}(\mathrm{x}_0) & \dots & f_{x_nx_n}(\mathrm{x}_0) \end{pmatrix}$$

is called Hesse–matrix of f at x_0.

$$\text{Hesse–matrix} = \text{Jacobi–matrix of the gradient } \nabla f$$

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

$$f(\mathrm{x}) = f(\mathrm{x}_0) + \operatorname{grad} f(\mathrm{x}_0)(\mathrm{x}-\mathrm{x}_0) + \frac{1}{2}(\mathrm{x}-\mathrm{x}_0)^T \mathrm{H}f(\mathrm{x}_0)(\mathrm{x}-\mathrm{x}_0) + O(\|\mathrm{x}-\mathrm{x}_0\|^3)$$

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

2.1 Extrem values of multivariate functions

Definition: Let $D \subset \mathbb{R}^n$, $f : D \to \mathbb{R}$ 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

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

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

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

Analogously we define the different forms of minima.

Necessary conditions for local extrem values.

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

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

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

$$\varphi(t) := f(\mathrm{x}^0 + t\mathrm{v})$$

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

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

$$\varphi'(0) = \operatorname{grad} f(\mathrm{x}^0)\,\mathrm{v} = 0$$

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

$$\operatorname{grad} f(\mathrm{x}^0) = (0, \ldots, 0)^T$$

Remarks to local extrem values.

Bemerkungen:

- Typically the condition grad $f(\mathrm{x}^0) = 0$ gives a non-linear system of n equations for n unknwons for the calculation of $\mathrm{x} = \mathrm{x}^0$.
- The points $\mathrm{x}^0 \in D^0$ with grad $f(\mathrm{x}^0) = 0$ are called stationary points of f. Stationary points are **not** necessarily local extram 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 $\mathrm{x}^0 = (0, 0)^T$. 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

$$f(\mathrm{x}^1) < f(\mathrm{x}^0) < f(\mathrm{x}^2).$$

Classification of stationary points.

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

a) necessary condition

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

$$\begin{array}{lcl} \mathrm{x}^0 \text{ local minimum} & \Rightarrow & \mathrm{H}\, f(\mathrm{x}^0) \text{ positiv semidefinit} \\ \mathrm{x}^0 \text{ local maximum} & \Rightarrow & \mathrm{H}\, f(\mathrm{x}^0) \text{ negativ semidefinit} \end{array}$$

b) sufficient condition

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

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

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

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

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

$$v^T H\, f(x^0 + \theta\varepsilon v)v \geq 0 \tag{2}$$

Since f is a $\mathcal{C}^2$–function, the Hesse–matrix is a continuous map. In the limit $\varepsilon \to 0$ 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 $\mathrm{H}\,f(\mathrm{x}^0)$ is positiv definit, then $\mathrm{H}\,f(\mathrm{x})$ is positiv definit in a sufficiently small neighborhood $\mathrm{x} \in K_\varepsilon(\mathrm{x}^0) \subset D$ around x^0. This follows from the continuity of the second partial derivatives.

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

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

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

If $\mathrm{H}\,f(\mathrm{x}^0)$ is indefinit, then there exist Eigenvectors v, w for Eigenvalues of $\mathrm{H}\,f(\mathrm{x}^0)$ with opposite sign with

$$\mathrm{v}^T \mathrm{H} f(\mathrm{x}^0)\mathrm{v} > 0 \qquad \mathrm{w}^T \mathrm{H} f(\mathrm{x}^0)\mathrm{w} < 0$$

and thus x^0 is a saddel 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 $Hf(x^0)$:

$$
\begin{aligned}
\text{all Eigenvalues are strictly positive} \quad &\Rightarrow \quad x^0 \text{ is a strict local mi}\\
\text{all Eigenvalues are strictly negative} \quad &\Rightarrow \quad x^0 \text{ is a strict local ma}\\
\text{there are strictly positive and negative Eigenvalues} \quad &\Rightarrow \quad x^0 \text{ saddel point}
\end{aligned}
$$

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

$$
\begin{array}{ccc}
x^0 \text{ local minimum} & \Leftarrow & x^0 \text{ strict local minimum} \\
\Downarrow & & \Uparrow \\
Hf(x^0) \text{ positiv semidefinit} & \Leftarrow & Hf(x^0) \text{ positiv definit}
\end{array}
$$

Further remarks.

- If f is a $\mathcal{C}^3$–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 ot 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 :

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

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

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

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

$$\mathrm{H}f(\mathrm{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \qquad \text{and} \qquad \mathrm{H}f(\mathrm{x}^1) = \begin{pmatrix} -2 & 0 \\ 0 & -10/3 \end{pmatrix}$$

The matrix $\mathrm{H}f(\mathrm{x}^0)$ is indefinit, therefore x^0 is a saddel point. $\mathrm{H}f(\mathrm{x}^1)$ is negativ definit and thus x^1 is a strict local ein strenges maximum of f.

2.2 Implicitely defined functions

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

$$\mathrm{g}(\mathrm{x}) = 0$$

with $\mathrm{g} : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$. 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 $\mathrm{g}(\mathrm{x}) = 0$ with respect to certain unknowns, i.e. with respect to the last m variables $x_{n-m+1}, \ldots, x_n$?

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

$$\mathrm{g}(\mathrm{x}) = 0 \quad \Longleftrightarrow \quad (x_{n-m+1}, \ldots, x_n)^T = \mathrm{f}(x_1, ..., x_{n-m})$$

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 defintion D? Or only *locally* on a subdomain $\tilde{D} \subset D$?

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

Example.

The equation for a circle

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

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 :

$$\begin{aligned}
y &= \sqrt{r^2 - x^2}, \quad -r \le x \le r \\
y &= -\sqrt{r^2 - x^2}, \quad -r \le x \le r \\
x &= \sqrt{r^2 - y^2}, \quad -r \le y \le r \\
x &= -\sqrt{r^2 - y^2}, \quad -r \le y \le r
\end{aligned}$$

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

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

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

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

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

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

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

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

Implicit function theorem.

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

If the Jacobi–matrix

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

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 $\mathrm{f} : U \to V$ with

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

and

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

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 neighborhod of $(0, r)$ with respect to y:

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

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

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

and therefore

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

Another example.

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

It is

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

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

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

Differentiating again gives

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

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 implicitely 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 opint 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

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

Implicit representation of curves.

Consider the set of solutions of a scalar equation

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

If

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

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

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

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

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

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

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

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

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

Remarks.

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

$$\det \mathrm{H}g(\mathrm{x}^0) > 0 \quad : \quad \mathrm{x}^0 \text{ is an isolated point of the set of solutions}$$

$$\det \mathrm{H}g(\mathrm{x}^0) < 0 \quad : \quad \mathrm{x}^0 \text{ is a double point}$$

$$\det \mathrm{H}g(\mathrm{x}^0) = 0 \quad : \quad \mathrm{x}^0 \text{ is a return point or a cusp}$$

Geometric interpretation:

a) If $\det \mathrm{H}g(\mathrm{x}^0) > 0$, then both Eigenvalues of $\mathrm{H}g(\mathrm{x}^0)$ are or strictly positiv or strictly negativ, i.e. x^0 is a strict local minimum or maximum of $g(\mathrm{x})$.

b) If $\det \mathrm{H}g(\mathrm{x}^0) < 0$, then both Eigenvalues of $\mathrm{H}g(\mathrm{x}^0)$ have opposite sign, i.e. x^0 is a saddel point of $g(\mathrm{x})$.

c) If $\det \mathrm{H}g(\mathrm{x}^0) = 0$, then the stationary point x^0 of $g(\mathrm{x})$ 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:

$$\begin{aligned} g_x &= y^2 + 3x^2 - 4x \\ g_y &= 2y(x-1) \\ g_{xx} &= 6x - 4 \\ g_{xy} &= 2y \\ g_{yy} &= 2(x-1) \\ \mathrm{H}g(0) &= \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix} \end{aligned}$$

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:

$$\begin{aligned} g_x &= y^2 + 3x^2 + 2xq^2 \\ g_y &= 2y(x-1) \\ g_{xx} &= 6x + 2q^2 \\ g_{xy} &= 2y \\ g_{yy} &= 2(x-1) \\ \mathrm{H}g(0) &= \begin{pmatrix} 2q^2 & 0 \\ 0 & -2 \end{pmatrix} \end{aligned}$$

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:

$$\begin{aligned} g_x &= y^2 + 3x^2 \\ g_y &= 2y(x-1) \\ g_{xx} &= 6x \\ g_{xy} &= 2y \\ g_{yy} &= 2(x-1) \\ \mathrm{H}g(0) &= \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \end{aligned}$$

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

Implicit representation of surfaces.

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

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

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

- If for example $g_z(\mathrm{x}^0) \neq 0$, 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

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

using the implicit function theorem.

The inverted Problem.

Question: Given the set of equations

$$y = f(x)$$

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

Theorem: (Inversion theorem)

Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^n$ a $\mathcal{C}^1$–function. If the Jacobian–matrix $J f(x^0)$ 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 $f^{-1} : V \to U$ is also $\mathcal{C}^1$ and for all $x \in U$ we have:

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

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

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

$$\begin{aligned} V &= \pi r^2 h \\ O &= 2\pi r^2 + 2\pi r h \end{aligned}$$

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

$$\begin{aligned} f(x,y) &= 2\pi x^2 + 2\pi xy \\ g(x,y) &= \pi x^2 y - c = 0 \end{aligned}$$

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 $f : \mathbb{R}^n \to \mathbb{R}$ under the constraint

$$\mathrm{g}(\mathrm{x}) = 0$$

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

The constraints are

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, \ldots, x_n) &= 0 \end{aligned}$$

Alternatively: Determine the extrem values of the function $f(\mathrm{x})$ on the set

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

The Lagrange–function and the Lagrange–Lemma.

We define the Lagrange–function

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

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

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

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

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

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

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

A necessary condition for local extrema.

Let f and g_i, $i = 1, \ldots, m$, $\mathcal{C}^1$–functions, then a necessary condition for an extrem value x^0 of $F(\mathrm{x})$ is given by

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

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

The solutions $(\mathrm{x}^0, \lambda^0)$ are the candidates for the extrem values, since these solutions satisfy the above necessary condition.

Alternatively: Define a Langrange–function

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

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

Some remarks on suffiecient conditions.

1. We can formulate a **necessary** condition:
   If the functions f and g are $\mathcal{C}^2$–functions and if the Hesse–matrix $\mathrm{H}F(\mathrm{x}^0)$ of the Lagrange–function is positiv (negativ) definit, then x^0 is a strict local minimum (maximum) of $f(\mathrm{x})$ on G.

2. In most of the applications the necessary condition are **not** satisfied, allthough x^0 is a strict local extremum.

3. And from the indefinitness of the Hesse–matrix $\mathrm{H}F(\mathrm{x}^0)$ 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 $G(\mathrm{x}, \lambda)$ 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

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

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

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

The necessary condition for an extrem value is given by

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

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

continuation of the example.

The Hesse–matrix at the origin is given by

$$\mathrm{H}f(0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

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

$$\begin{aligned} y + 2\lambda x &= 0 \\ x + 2\lambda y &= 0 \\ x^2 + y^2 &= 1 \end{aligned}$$

with the four solution

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

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

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

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

Lagrange–multiplier–rule.

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

$$\operatorname{rang}\left(\mathrm{J}\,\mathrm{g}(\mathrm{x}^0)\right) = m$$

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

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

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

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

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 $\lambda_1, \ldots, \lambda_m$ be the related Lagrange–multiplier. Then the Hesse–matrix $HF(x^0)$ of the Lagrange–function is positiv semi-definit on the tangential space

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

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 staisfied. If there exist Lagrange–multiplier $\lambda_1, \ldots, \lambda_m$, 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

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

From the necessary condition we obtain the non-linear system

$$\begin{aligned} -2x + 8 &= -2\lambda x \\ -2y &= -2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

Continuation of the example.

From the necessary condition we obtain the non-linear system

$$\begin{aligned} -2x + 8 &= -2\lambda x \\ -2y &= -2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

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

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

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

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

with the plane

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

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

$$\begin{aligned} g_1(x,y,z) &:= x^2 + y^2 - 2 = 0 \\ g_2(x,y,z) &:= x + z - 1 = 0 \end{aligned}$$

Continuation of the example.

The Jacobi–matrix

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

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

$$\begin{aligned} 2 + 2\lambda_1 x + \lambda_2 &= 0 \\ 3 + 2\lambda_1 y &= 0 \\ 2 + \lambda_2 &= 0 \\ x^2 + y^2 &= 2 \\ x + z &= 1 \end{aligned}$$

Continuation of the example.

The necessary condition gives the non-linear system

$$\begin{aligned} 2 + 2\lambda_1 x + \lambda_2 &= 0 \\ 3 + 2\lambda_1 y &= 0 \\ 2 + \lambda_2 &= 0 \\ x^2 + y^2 &= 2 \\ x + z &= 1 \end{aligned}$$

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

$$\begin{aligned} Z &= \{(x, y, z)^T \in \mathbb{R}^3 \,|\, x^2 + y^2 \leq 2\} \\ M_Z &= \{(x, y, z)^T \in \mathbb{R}^3 \,|\, x^2 + y^2 = 2\} \end{aligned}$$

We calculate the related functiuon values

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

Thus the point $(x, y, z) = (0, \sqrt{2}, 1)$ is a maximum an 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 $\mathrm{f} : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$:

$$\mathrm{f}(\mathrm{x}) = 0$$

- We already know the fixed-point iteration

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

  with starting point x^0 and iteration map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$.
- Convergence results are given by the Banach Fixed Point Theorem.

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

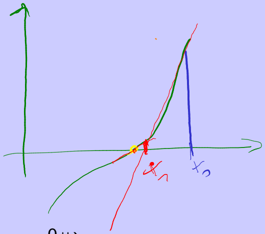
Disadvantages:

- 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 $f : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$ 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

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

$f : \mathbb{R}^n \to \mathbb{R}^n$

$Jf(x_0)$ $n \times n$

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 $k = 0, 1, 2, \ldots$

(2a) Solve $\mathrm{J}f(\mathrm{x}^k) \cdot \Delta\mathrm{x}^k = -f(\mathrm{x}^k)$**;**

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

- In *every* Newton–step we solve a set of linear equations.
- The solution $\Delta\mathrm{x}^k$ 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

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

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 conincide.

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

Local convergence of the Newton–method.

Theorem: Let $f : D \to \mathbb{R}^n$ 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

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

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 $x^k \in K_r(x^*)$, $k = 0, 1, 2, \ldots$, 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 $K_r(x^*)$.

Local convergence of the Newton–method.

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

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

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

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

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

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

$$\|\mathrm{x}^{k+1} - \mathrm{x}^*\| \;\le\; \frac{L}{2}\,\|\mathrm{x}^k - \mathrm{x}^*\|^2$$

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

Local convergence of the Newton–method.

Theorem: Let $f : D \to \mathbb{R}^n$ 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

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

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

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

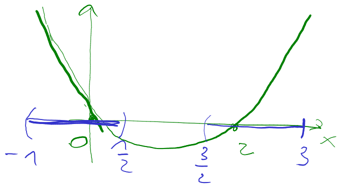
with $x^k \in K_r(x^*)$, $k = 0, 1, 2, \ldots$, and the Newton–iterates x^k converge quadratically to x^*, i.e.

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

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

$f(x) = (x-1)^2 - 1$

$f'(x) = 2(x-1) = Df(x)$



$\left| \frac{1}{2(x-1)} \left(2(y-1) - 2(x-1) \right) \right| \leq L |y-x|$

$Df(x)^{-1}$ $\quad Df(y)$ $\quad Df(x)$

$\left| \frac{1}{x-1} \right| |y-x| \leq L |y-x|$

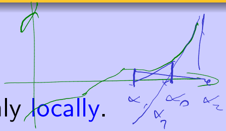
$\frac{1}{|x-1|} \leq L \quad$ e.g. $L = 2 \quad |x-1| \geq \frac{1}{2} \quad x-1 \geq \frac{1}{2} \quad \boxed{x \geq \frac{3}{2}}$

$1-x \geq \frac{1}{2} \quad x \leq \frac{1}{2}$

$x_0 \in \left(\frac{3}{2}, 3\right) \Rightarrow x_k \xrightarrow[k\to\infty]{} 2$

$x_0 \in \left(-1, \frac{1}{2}\right) \Rightarrow x_k \xrightarrow[k\to\infty]{} 0$

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** $\mathrm{Jf}(\mathrm{x}^k) \cdot \Delta \mathrm{x}^k = -f(\mathrm{x}^k)$**;**
>
> **(2b) Set** $\mathrm{x}^{k+1} = \mathrm{x}^k + \lambda_k \Delta \mathrm{x}^k$**;**
>
> instead of 1

Frage: How should we choose the damping parameters λ_k?

Choice of the damping paramter.

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

$$\begin{aligned} T(\mathrm{x}) &\geq 0, \quad \forall \mathrm{x} \in D \\ T(\mathrm{x}) &= 0 \quad \Leftrightarrow \quad f(\mathrm{x}) = 0 \end{aligned}$$

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

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

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

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

Damping strategy.

For the **initial iteration** $k = 0$: Choose $\lambda_0 \in \{1, \frac{1}{2}, \frac{1}{4}, \ldots, \lambda_{min}\}$ 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 $\lambda_k < 1$.

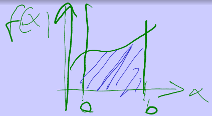
ELSE

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

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

- $\lambda_k := \mu$

END



3.1 Area integrals

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

Aim: Calculate the volume under the graph of $f(\mathrm{x})$:

$f(x)$, $F(x)$, $F'(x) = f(x)$

$\int_a^b f(x)dx = F(b) - F(a)$

$$V = \int_D f(\mathrm{x})d\mathrm{x}$$

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

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

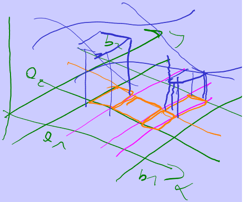
i.e. D is compact cuboid (rectangle).

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

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

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

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



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

Definition:

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

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

maximal distance of x_i's or y_j's

- 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.

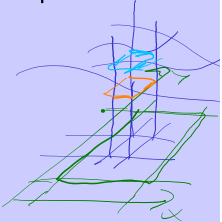
Riemann upper and lower sums.

Definition:

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

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

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



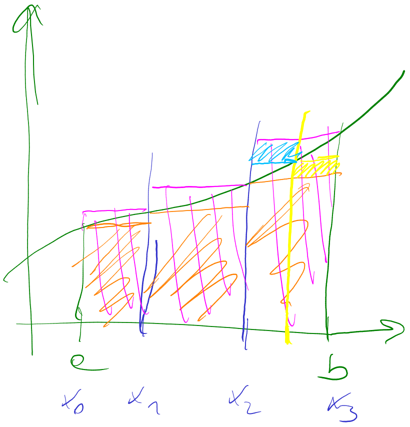
the Riemann lower sum and the Riemann upper sum of $f(x)$, respectively.

Remark:

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

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

based on inf arbitrary sup



$O_f(Z) \geq O_f(\tilde{Z})$

$U_f(Z) \leq U_f(\tilde{Z})$

Z_1, Z_2 arbitrary

$Z_3 = Z_1 \cup Z_2$

Z_3 finer than Z_1, Z_2

$U_f(Z_1) \leq U_f(Z_3) \leq O_f(Z_3) \leq O_f(Z_2)$

finer

$\Rightarrow$ fix Z_2 $\quad U_f(Z_1) \leq O_f(Z_2) = c_2 \quad \forall Z_1 \in Z(D)$

$\Rightarrow$ fix Z_1 $\quad c_1 = U_f(Z_1) \leq O_f(Z_2) \quad \forall Z_2 \in Z(D)$

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 $\|Z\| \to 0$:

$$\begin{aligned} U_f &:= \sup\{U_f(Z) \, : \, Z \in \mathrm{Z}(D)\} \\ O_f &:= \inf\{O_f(Z) \, : \, Z \in \mathrm{Z}(D)\} \end{aligned}$$

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

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

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

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 habe "only" considered the case of **two** variables:

$$f : D \to \mathbb{R}, \qquad D \in \mathbb{R}^2$$

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

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

$$\int_D f(x,y)dxdy \quad \text{bzw.} \quad \int_D f(x,y,z)dxdydz$$

or

$$\iint_D f(x,y)dxdy \quad \text{bzw.} \quad \iiint_D f(x,y,z)dxdydz$$

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

$U_f(Z) \le U_g(Z)$, $O_f(Z) \le O_g(Z)$

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

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

$0 \le \int_D g(x)dx$

$f(x) = 0 \le g(x)$ $\quad \int_D 0\,dx = 0 \quad U_{f=0}(Z) = 0 \quad O_{f=0}(Z) = 0$

c) Positivity

If for all $x \in D$ the relation $f(x) \ge 0$ holds, i.e. $f(x)$ is non–negativ, then

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

$D = [0,1]$ $f(x) = \begin{cases} 0 & x \in \mathbb{Q} \cap [0,1] \\ 1 & x \notin \mathbb{Q} \cap [0,1] \end{cases}$

$\nexists \varepsilon > 0$

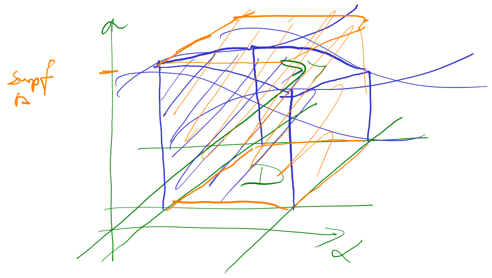
$-U_f(Z) + O_f(Z) \leq \varepsilon$

$\int_0^1 f(x)\,dx$ $U_f(Z) = 0$ $\int_{[0,1]} f(x)\,dx = 0$

$O_f(Z) = 1$ $\int_{[0,1]} f(x)\,dx = 1$ #

not Riemann integrable

(*)



$\int_D f(x)\,dx$

Additional properties of the integral.

Theorem
~~**Satz:**~~

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

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

vol$(D_1 \cap D_2) \stackrel{!}{=} 0$

b) The following estimate holds for the integral

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

(*)

c) Riemann criterion

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

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

Fubini's theorem.

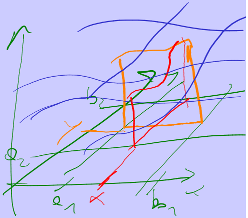
Theorem: (Fubini's theorem) Let $f : D \to \mathbb{R}$ be integrable, $D = [a_1, b_1] \times [a_2, b_2]$ be a cuboid. If the integrals

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

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

$$\int_D f(\mathrm{x})d\mathrm{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x,y)dydx$$

$$\int_D f(\mathrm{x})d\mathrm{x} = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x,y)dxdy$$



holds true.

Importance:

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

Example.

$F(x) = \int_0^2 f(x,y)\,dy = \int_0^2 (2-xy)\,dy = 4-2x \qquad G(y) = \int_0^1 f(x,y)\,dx = 2-\frac{y}{2}$

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:

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

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

$= \int_0^1 \int_0^2 f(x,y)\,dy\,dx = \int_0^1 (4-2x)\,dx = 4 - \frac{2}{2} = 3$

The characteristic function.

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

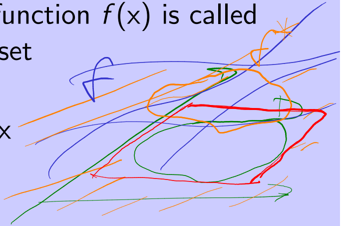
restriction of f on D

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

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

$$\mathrm{vol}(D) := \int_D 1 dx = \int_Q \mathcal{X}_D(x) dx$$

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

The compact set D is called null set, if D is measurable and if $\mathrm{vol}(D) = 0$ 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.
- $\mathrm{vol}(D)$ 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 $f : D \to \mathbb{R}$ be continuous. Then $f(\mathrm{x})$ is integrable on D.

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

$$\int_D f(\mathrm{x}) d\mathrm{x} = f(\xi) \cdot \mathrm{vol}(D)$$

"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 continuos functions g, h, φ and ψ.

Projectable sets.

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

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

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 $f(\mathrm{x})$ is a continuous function on a "normal" area

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

then we have

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

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

$$\begin{aligned} D = \{\ \mathrm{x} \in \mathbb{R}^n \mid \tilde{\mathrm{x}} &= (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)^T \in B \\ &\text{and } \varphi(\tilde{\mathrm{x}}) \le x_i \le \psi(\tilde{\mathrm{x}})\ \} \end{aligned}$$

then it holds

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

Example.

Given a function

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

Calculate the integral on the area bounded by two parabolas

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

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

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

Example.

Calculate the volume of the rotational paraboloid

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

Representation of V as "normal" area

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

Then we have

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