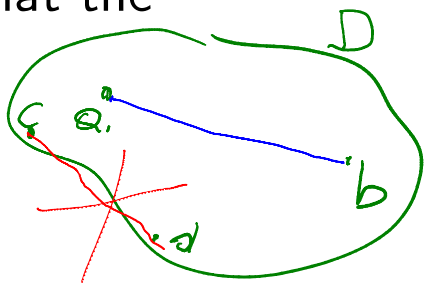
## 1.3 Mean value theorems and Taylor expansion

**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 $\mathbf{a}, \mathbf{b} \in D$ be points in $D$ such that the connecting line segment




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

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

$\theta = 0 \quad a + 0(b-a) = a$
$\theta = 1 \quad a + 1(b-a) = b$

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

1d. MVT $\quad h(t_2) - h(t_1) = h'(t_1 + \theta(t_2 - t_1))(t_2 - t_1)$

**Proof:** We set

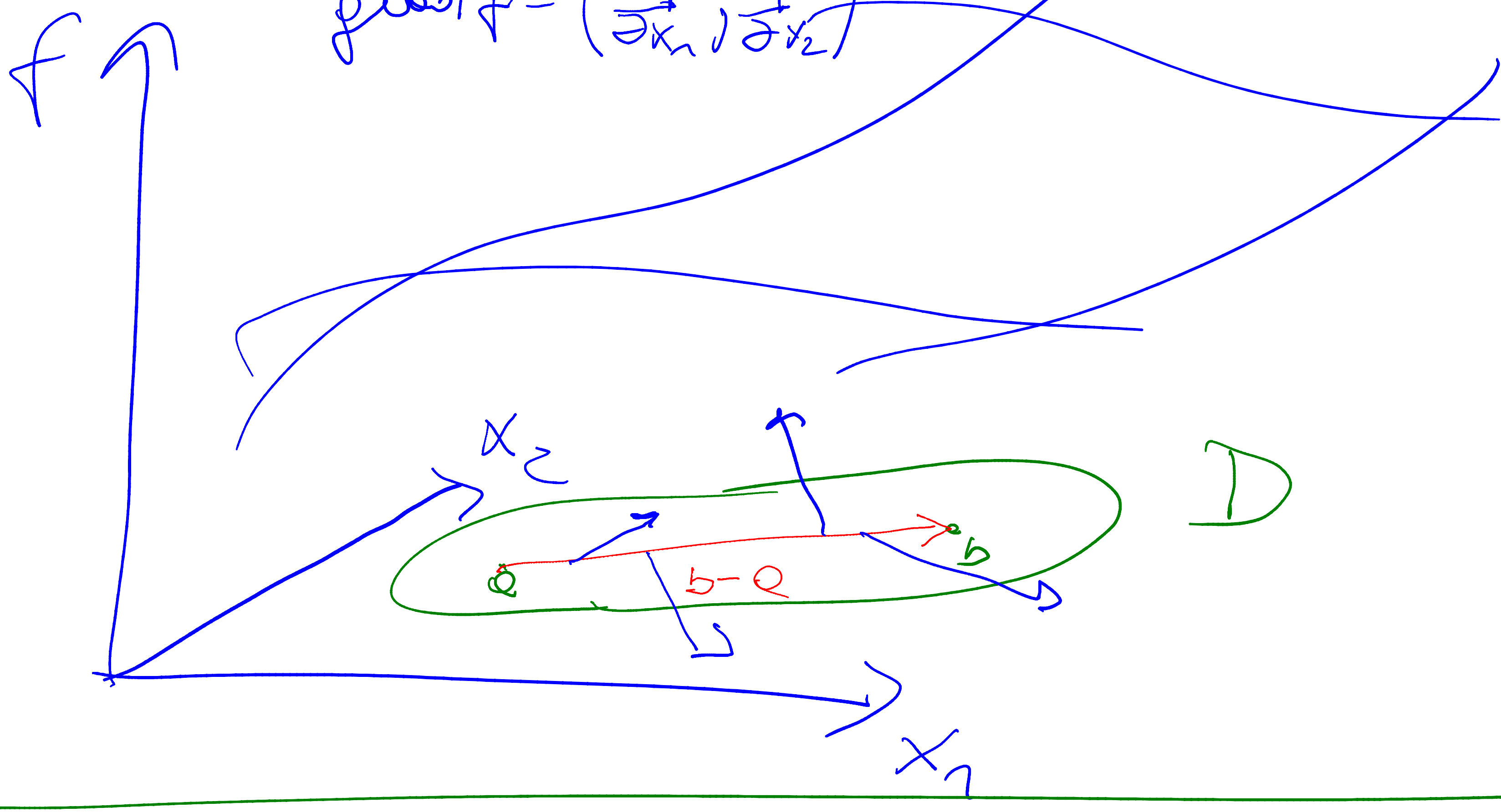
$$h(t) := f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$$

$h(0) = f(a), \quad h(1) = f(b)$

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

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

$$\operatorname{grad} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)$$




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

$$h'(\theta) = \frac{d}{dt} f(a + t(b-a)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\big(a + t(b-a)\big) \cdot \underbrace{(b_i - a_i)}_{\text{inner derivative}} =$$

$$= \operatorname{grad} f\big(a + t(b-a)\big) \cdot (b-a)$$

# Definition and example.

convex ✓

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

not convex

**Example for the mean value theorem:** Given a scalar function

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

$a = (0,0)$
$b = \left(\frac{\pi}{2}, \frac{\pi}{2}\right)$

It is

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

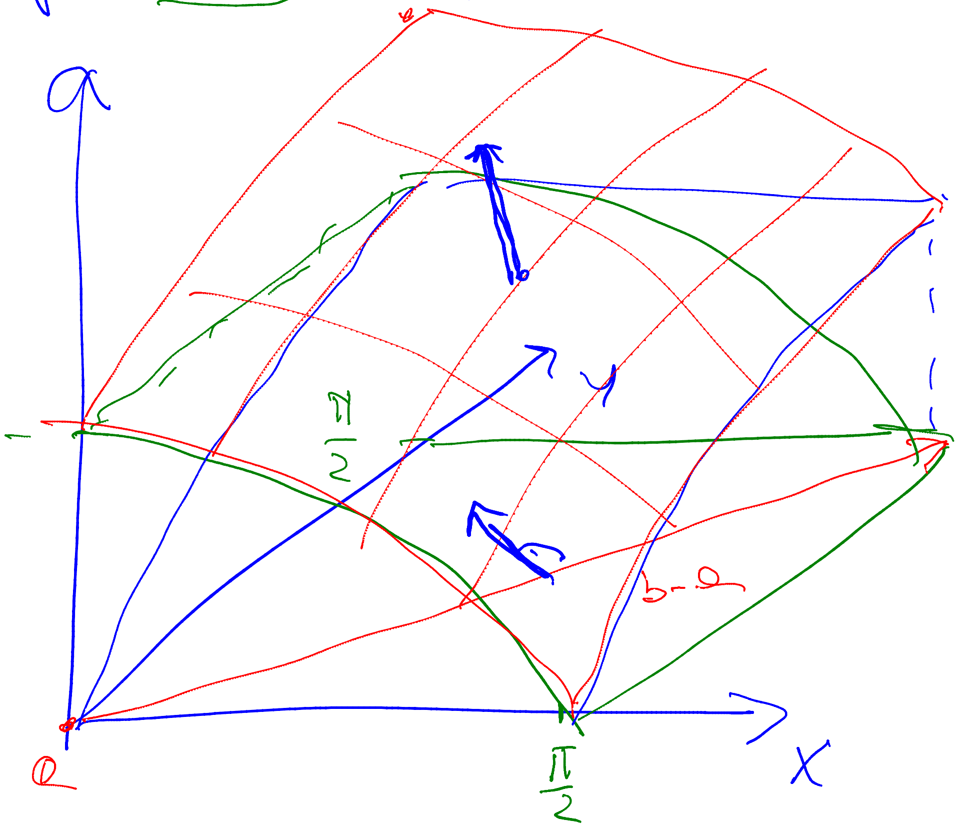
Applying the mean value theorem there exists a $\theta \in (0,1)$ with

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

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

$\begin{pmatrix} \frac{\pi}{2} - 0 \\ \frac{\pi}{2} - 0 \end{pmatrix} \qquad \begin{pmatrix} \frac{\pi}{2} - 0 \\ \frac{\pi}{2} - 0 \end{pmatrix}$

$$f = \cos x + \sin y$$




$$\left(\frac{\pi}{2}, \frac{\pi}{2}\right) = b$$

$$\operatorname{grad} f = \begin{pmatrix} -\sin x \\ \cos y \end{pmatrix}$$

$$\operatorname{grad} f\left(\frac{\pi}{4}, \frac{\pi}{4}\right) = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} = \frac{\sqrt{2}}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\perp \quad (b - a) = \frac{\pi}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\left(\frac{\pi}{4}, \frac{\pi}{4}\right) = \frac{1}{2} \left(\frac{\pi}{2}, \frac{\pi}{2}\right)$$

$$\Theta = \frac{1}{2}$$

# 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 $f: D \to \mathbb{R}^2$

$$\mathbf{f}(t) := \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \qquad t \in [0, \pi/2] = D \subseteq \mathbb{R}^1 \qquad f'(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

It is $a = 0$ $b = \frac{\pi}{2}$

$$f(b) - f(a) = \mathbf{f}(\pi/2) - \mathbf{f}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad \left\| \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\|_2 = \sqrt{2}$$

$\neq$ and $\neq$

$$f'(a + \theta(b-a)) \cdot (b-a) = \underbrace{\mathbf{f}'\left(\theta \frac{\pi}{2}\right)}_{\exists \notin \quad 2\times 1} \cdot \underbrace{\left(\frac{\pi}{2} - 0\right)}_{1} = \frac{\pi}{2} \begin{pmatrix} -\sin(\theta\pi/2) \\ \cos(\theta\pi/2) \end{pmatrix} \qquad \left\| \frac{\pi}{2} \begin{pmatrix} -\sin(\,) \\ \cos(\,) \end{pmatrix} \right\| = \frac{\pi}{2}$$

indep. of $\theta$

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

$$1.41 = \sqrt{2} = \|f(b) - f(a)\|_2 \leq \left\| f'\left(\theta\frac{\pi}{2}\right) \cdot \left(\frac{\pi}{2} - 0\right) \right\|_2 = \frac{\pi}{2} = 1.7\ldots$$

# A mean value estimate for vector–valued functions.

general non scalar

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

$$\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\|_2 \leq \|\underbrace{\mathbf{J}\,\mathbf{f}(\mathbf{a} + \theta(\mathbf{b} - \mathbf{a}))}_{m \times n \text{ Matrix}} \cdot \underbrace{(\mathbf{b} - \mathbf{a})}_{n \text{ Vector}}\|_2$$

**Idea of the proof:** Application of the mean value theorem to the scalar function $g(\mathbf{x})$ definid as

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

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

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

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

$g(x) = (f(b)-f(a))^T f(x)$ $\qquad g: \mathbb{R}^n \to \mathbb{R}$ $\qquad$ Skalar $\qquad$ MWS ✓

$$g(b) - g(a) = (f(b)-f(a))^T (f(b)-f(a)) = \| f(b)-f(a) \|_2^2$$

$(( \exists \theta \in [0,1]$

$$J g(a+\theta(b-a)) \cdot (b-a) = J\left( (f(b)-f(a))^T f(a+\theta(b-a)) \right) \cdot (b-a)$$

$$= \left( \sum_{i=1}^n \frac{\partial}{\partial x_i} \sum_{j=1}^n (f(b)-f(a))_j \, f_j(a+\theta(b-a)) \right) \cdot (b-a)$$

$$= \sum_{i=1}^n \sum_{j=1}^n (f(b)-f(a))_j \underbrace{\frac{\partial}{\partial x_i} f_j(a+\theta(b-a))}_{Jf} (b-a)$$

$$= \underbrace{(f(b)-f(a))^T}_{m \text{ Vektor}} \underbrace{Jf(a+\theta(b-a))}_{m \times n} \underbrace{(b-a)}_{n \text{ Vektor}}$$

$$\| Jg(a+\theta(b-a)) \cdot b-a \|_2 = \|_2 (f(b)-f(a))^T Jf(a+\theta(b-a))(b-a) \|$$

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

# Taylor series: notations.

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

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

$n = 2$

$|\alpha| = 0 \quad (\alpha_1, \alpha_2) = (0,0)$

$|\alpha| = 1 \quad (\alpha_1, \alpha_2) = \begin{cases} (1,0) \\ (0,1) \end{cases}$

$\frac{\partial}{\partial x_1} = D^{(1,0)}$

$\frac{\partial}{\partial x_2} = D^{(0,1)}$

Let

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

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

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

$|\alpha| = 2 \quad (\alpha_1, \alpha_2) = \begin{cases} (2,0) \\ (1,1) \\ (0,2) \end{cases}$

$D^{(2,0)} = \frac{\partial^2}{\partial x_1^2}$

$D^{(1,1)} = \frac{\partial^2}{\partial x_1 \partial x_2}$

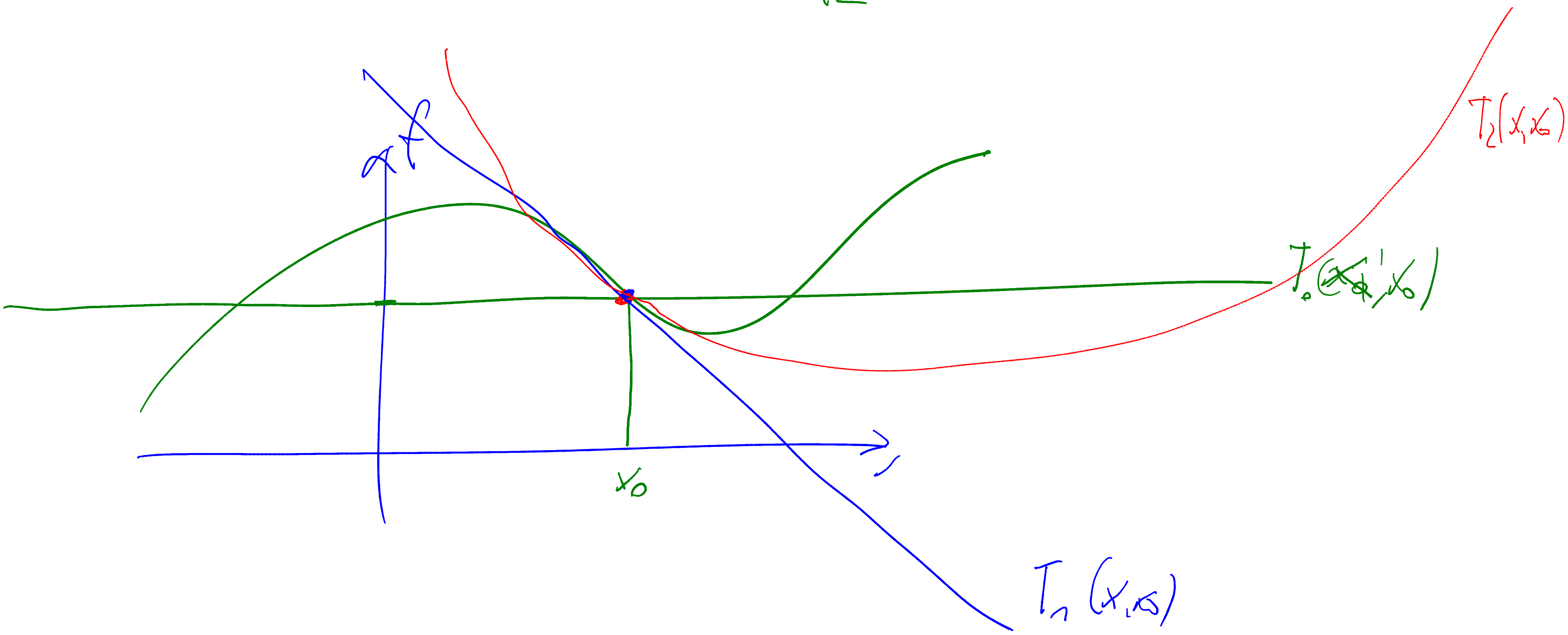
$D^{(0,2)} = \frac{\partial^2}{\partial x_2^2}$

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

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

Taylor 1d

$$\underbrace{\underbrace{\underbrace{f(x) = f(x_0)}_{T_0} + f'(x_0)(x-x_0)}_{T_1} + \frac{f''(x_0)}{2!}(x-x_0)^2}_{T_2} + \dots$$



# The Taylor theorem.

**Theorem:** (Taylor) Skalar

Let $D \subset \mathbb{R}^n$ be open and convex. Let $f : D \to \mathbb{R}$ be a $\mathcal{C}^{m+1}$–function and $\mathbf{x}_0 \in D$. Then the Taylor–expansion holds true in $\mathbf{x} \in D$

$$
\begin{aligned}
f(\mathbf{x}) &= T_m(\mathbf{x}; \mathbf{x}_0) + R_m(\mathbf{x}; \mathbf{x}_0) \\
T_m(\mathbf{x}; \mathbf{x}_0) &= \sum_{|\alpha| \le m} \frac{D^\alpha f(\mathbf{x}_0)}{\alpha!} (\mathbf{x} - \mathbf{x}_0)^\alpha \\
R_m(\mathbf{x}; \mathbf{x}_0) &= \sum_{|\alpha| = m+1} \frac{D^\alpha f(\mathbf{x}_0 + \theta(\mathbf{x} - \mathbf{x}_0))}{\alpha!} (\mathbf{x} - \mathbf{x}_0)^\alpha
\end{aligned}
$$

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

**Notation:** In the Taylor–expansion we denote $T_m(\mathbf{x}; \mathbf{x}_0)$ Taylor–polynom of degree $m$ and $R_m(\mathbf{x}; \mathbf{x}_0)$ Lagrange–remainder.

$n=2$ scalar

$$f(x) = f(x_0) + D^{(1,0)} f(x_0)(x-x_0)^{(1,0)} + D^{(0,1)} f(x_0)\cdot(x-x_0)^{(0,1)}$$

$$+ \frac{1}{2}\left[D^{(2,0)} f(x_0)(x-x_0)^{(2,0)} + D^{(1,1)} f(x_0)(x-x_0)^{(1,1)} + D^{(0,2)} f(x_0)(x-x_0)^{(0,2)}\right] + \dots$$

$$= \underbrace{f(x_0)}_{T_0(x,x_0)} + \underbrace{\frac{\partial}{\partial x_1} f(x_0)(x_1-x_{01}) + \frac{\partial}{\partial x_2} f(x_0)\cdot(x_2-x_{02})}_{\operatorname{grad} f(x_0)\cdot(x-x_0)}$$

$$+ \frac{1}{2}\underbrace{\left(\frac{\partial^2}{\partial x_1^2} f(x_0)(x_1-x_{01})^2 + 2\frac{\partial^2}{\partial x_1 \partial x_2} f(x_0)(x_1-x_{01})(x_2-x_{02}) + \frac{\partial^2}{\partial x_2^2} f(x_0)(x_2-x_{02})^2\right)}_{\frac{1}{2}(x-x_0)^T Hf(x_0)(x-x_0)}$$

$$Hf(x_0) = \underbrace{\begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f & \frac{\partial^2}{\partial x_2 \partial x_1} f \\ \frac{\partial^2}{\partial x_1 \partial x_2} f & \frac{\partial^2}{\partial x_2^2} f \end{pmatrix}}_{2\times 2}(x_0)$$

(because of $n=2$)

# Derivation of the Taylor expansion.

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

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

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

$$f(x) = g(1) = \underbrace{\underbrace{\overbrace{g(0)}^{f(x_0)}}_{T_0(x,x_0)} + g'(0)\cdot(1-0)}_{T_1(x,x_0)} + \frac{1}{2} g''(\xi)\cdot(1-0)^2 \quad \text{for a } \xi \in (0,1).$$

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

$$\begin{aligned}
g'(0) &= \frac{d}{dt} f(x_1^0 + t(x_1 - x_1^0), x_2^0 + t(x_2 - x_2^0), \ldots, x_n^0 + t(x_n - x_n^0))\Big|_{t=0} \\
\text{chain rule} \quad &= D_1 f(\mathbf{x}_0)\cdot(x_1 - x_1^0) + \ldots + D_n f(\mathbf{x}_0)\cdot(x_n - x_n^0) = \operatorname{grad} f(x_0)\cdot(x - x_0) \\
&= \sum_{|\alpha|=1} \frac{D^\alpha f(\mathbf{x}_0)}{\alpha!}\cdot(\mathbf{x} - \mathbf{x}_0)^\alpha
\end{aligned}$$

# Continuation of the derivation.

Calculation of $g''(0)$ gives

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

$\frac{1}{2}(x-x_0)^T Hf(x_0)(x-x_0)$

$$
= \sum_{|\alpha|=2} \frac{D^\alpha f(\mathbf{x}_0)}{\alpha!}(\mathbf{x}-\mathbf{x}_0)^\alpha \qquad \text{(exchange theorem of Schwarz!)}
$$

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

# Proof of the Taylor theorem.

The function

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

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 $k$)

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

and

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

# Examples for the Taylor–expansion.

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

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

5. Details on extra slide.

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

$$(x_0,y_0,z_0) = (1,2,0)$$

$$f(x_0,y_0,z_0) = 0$$

$$\operatorname{grad} f(x_0,y_0,z_0) = \left(y^2 \sin z,\; x2y \sin z,\; xy^2 \cos z\right)\Big|_{(1,2,0)} = (0,0,4)$$

$$Hf(x_0,y_0,z_0) = \begin{pmatrix} 0 & 2y \sin z & \\ 2y \sin z & 2x \sin z & \\ y^2 \cos z & 2xy \cos z & -xy^2 \sin z \end{pmatrix}\Bigg|_{(1,2,0)} = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 4 \\ 4 & 4 & 0 \end{pmatrix}$$

$$f(x,y,z) = 0 + 4(z-0) + \frac{1}{2}(x-1, y-2, z)^T \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 4 \\ 4 & 4 & 0 \end{pmatrix} \begin{pmatrix} x-1 \\ y-2 \\ z \end{pmatrix}$$

$$= 4z + \underbrace{\frac{1}{2}\left( \qquad \right) \begin{pmatrix} 4z \\ 4z \\ 4(x-1)+4(y-2) \end{pmatrix}}_{\frac{1}{2}\left\{4z(x-1) + 4z(y-2) + [4(x-1)+4(y-2)]z\right\} = 4z(x-1+y-2) = (*)} =$$

$$\boxed{(*) = 4z(x+y-2)}$$

# Remarks to the remainder of a Taylor–expansion.

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

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

If all these derivative are bounded by aconstant $C$ in a neighborhood of $\mathbf{x}_0$ then the estimate for the remainder hold true

number of possible terms.

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

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

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

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

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

# The Hesse–matrix.

The matrix

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

is called Hesse–matrix of $f$ at $\mathbf{x}_0$.

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

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

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

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