# Dynamische Prozessoptimierung

## Kapitel 2:

## Mathematische Grundlagen

**Prof. Dr.-Ing. habil. Pu Li**

**Fachgebiet Prozessoptimierung**

Fachgebiet
Prozessoptimierung

# Mathematische Grundlage: Ableitungen

**Grund:** Gradientenbasierte Lösungsverfahren brauchen Ableitungen.

$$\frac{dy}{dx} = ?$$

Eingangsvariable
- Skalar $x$
- Vektor $\mathbf{x}$
- Matrix $\mathbf{X}$

Ausgangsvariable
- Skalar $y$
- Vektor $\mathbf{y}$
- Matrix $\mathbf{Y}$

Es gibt 9 unterschiedliche Möglichkeiten bzw. Kombinationen:

$$\frac{dy}{dx} = ? \qquad \boxed{\frac{d\mathbf{y}}{dx} = ?} \qquad \boxed{\frac{d\mathbf{Y}}{dx} = ?}$$

$$\boxed{\frac{dy}{d\mathbf{x}} = ?} \qquad \boxed{\frac{d\mathbf{y}}{d\mathbf{x}} = ?} \qquad \boxed{\frac{d\mathbf{Y}}{d\mathbf{x}} = ?}$$

$$\frac{dy}{d\mathbf{X}} = ? \qquad \frac{d\mathbf{y}}{d\mathbf{X}} = ? \qquad \frac{d\mathbf{Y}}{d\mathbf{X}} = ?$$

Fachgebiet Prozessoptimierung

# 1. Ableitungen nach einer skalaren Variablen (z.B. Zeit $t$)

**Ein Vektor von Funktionen:** $\mathbf{x}(t) = [x_1(t) \quad x_2(t) \quad \cdots \quad x_n(t)]^T$

$$\frac{d\mathbf{x}}{dt} \equiv \begin{bmatrix} \frac{dx_1}{dt} & \frac{dx_2}{dt} & \ldots & \frac{dx_n}{dt} \end{bmatrix}^T$$

**Inneres Produkt (Skalar-produkt) zweier Vektoren:**

$$\mathbf{x}(t) = [x_1(t) \quad x_2(t) \quad \cdots \quad x_n(t)]^T$$

$$\mathbf{y}(t) = [y_1(t) \quad y_2(t) \quad \cdots \quad y_n(t)]^T$$

Bilineare Systeme:

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i$$

$$\frac{d\left(\mathbf{x}^T\mathbf{y}\right)}{dt} = \frac{d\mathbf{x}^T}{dt}\mathbf{y} + \mathbf{x}^T\frac{d\mathbf{y}}{dt}$$ (Übungsaufgabe!)

**Eine Matrix von Funktionen:**

$$\mathbf{X}(t) = \begin{bmatrix} x_{11}(t) & \cdots & x_{1m}(t) \\ \vdots & \cdots & \vdots \\ x_{n1}(t) & \cdots & x_{nm}(t) \end{bmatrix} = [x_{i,j}]_{n\times m}$$

Die Definition:

$$\frac{d\mathbf{X}}{dt} \equiv \begin{bmatrix} \frac{dx_{11}}{dt} & \cdots & \frac{dx_{1m}}{dt} \\ \vdots & \cdots & \vdots \\ \frac{dx_{n1}}{dt} & \cdots & \frac{dx_{nm}}{dt} \end{bmatrix} \equiv \left[\frac{dx_{i,j}}{dt}\right]_{n\times m}$$

Es gibt:

$$\frac{d(\mathbf{X}_1 \pm \mathbf{X}_2)}{dt} = \frac{d\mathbf{X}_1}{dt} \pm \frac{d\mathbf{X}_2}{dt}$$

$$\frac{d(\lambda\mathbf{X})}{dt} = \frac{d\lambda}{dt}\mathbf{X} + \lambda\frac{d\mathbf{X}}{dt}$$

$$\frac{d(\mathbf{X}_1\mathbf{X}_2)}{dt} = \frac{d\mathbf{X}_1}{dt}\mathbf{X}_2 + \mathbf{X}_1\frac{d\mathbf{X}_2}{dt}$$

(Übungsaufgabe!)

Fachgebiet Prozessoptimierung

# Eine quadratische Funktion: $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$



mit $\mathbf{x}(t) = [x_1(t) \quad x_2(t) \quad \cdots \quad x_n(t)]^T$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{bmatrix} = [a_{i,j}]_{n\times n}$$

$\mathbf{A}$ ist eine symmetrische, konstante Matrix.

$$\frac{d\left(\mathbf{x}^T \mathbf{A}\mathbf{x}\right)}{dt} = \frac{d\mathbf{x}^T}{dt}(\mathbf{A}\mathbf{x}) + \mathbf{x}^T \frac{d(\mathbf{A}\mathbf{x})}{dt}$$

$$= \frac{d\mathbf{x}^T}{dt}(\mathbf{A}\mathbf{x}) + \mathbf{x}^T \mathbf{A}\frac{d\mathbf{x}}{dt} = 2\mathbf{x}^T \mathbf{A}\frac{d\mathbf{x}}{dt} = 2\mathbf{x}^T \mathbf{A}\dot{\mathbf{x}}$$

## Beispiel:

| | Istwert | Sollwert |
|---|---|---|
| Position: | $u$ | $u_s$ |
| Geschwindigkeit: | $v$ | $v_s$ |
| Beschleunigung: | $w$ | $w_s$ |

Minimierung des Fehlerquadrates:

$$\min_{u,v,w} \int_{t_0}^{t_f} \left[\alpha_1 (u-u_s)^2 + \alpha_2 (v-v_s)^2 + \alpha_3 (w-w_s)^2\right] dt$$

Wie kann man die Funktion mit der Form $f(x) = \mathbf{x}^T \mathbf{A}\mathbf{x}$ schreiben?

**Inversion einer Matrix:** $\mathbf{X}^{-1}(t)$ mit $\mathbf{X}(t)=\begin{bmatrix} x_{11}(t) & \cdots & x_{1n}(t) \\ \vdots & \cdots & \vdots \\ x_{n1}(t) & \cdots & x_{nn}(t) \end{bmatrix}=[x_{i,j}]_{n\times n}$

Da $\mathbf{X}\,\mathbf{X}^{-1}=\mathbf{E}$, $\dfrac{d\mathbf{E}}{dt}=\mathbf{0}$ , gibt es

$$\frac{d\left(\mathbf{X}\,\mathbf{X}^{-1}\right)}{dt}=\frac{d\mathbf{X}}{dt}\mathbf{X}^{-1}+\mathbf{X}\frac{d\mathbf{X}^{-1}}{dt}=\mathbf{0}$$

Dann $\mathbf{X}\dfrac{d\mathbf{X}^{-1}}{dt}=-\dfrac{d\mathbf{X}}{dt}\mathbf{X}^{-1}$

$$\frac{d\mathbf{X}^{-1}}{dt}=-\mathbf{X}^{-1}\frac{d\mathbf{X}}{dt}\mathbf{X}^{-1}$$

Achtung: $\mathbf{X}$ muss invertierbar sein.
d.h. $\det\left[\mathbf{X}(t)\right]\neq 0$

Eine Ableitung nach der Zeit wird die Dimension nicht verändern!
Sie ist eine Totalableitung / ein totales Differential!

Fachgebiet Prozessoptimierung

# 2. Ableitungen nach einem Vektor von Variablen

**Eine Funktion:** $f(\mathbf{x})$ mit $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$

$$\frac{df(\mathbf{x})}{d\mathbf{x}} \equiv \left[ \frac{\partial f(\mathbf{x})}{\partial x_1} \quad \frac{\partial f(\mathbf{x})}{\partial x_1} \quad \cdots \quad \frac{\partial f(\mathbf{x})}{\partial x_n} \right]^T \equiv \nabla f(\mathbf{x})$$

$$\frac{df(\mathbf{x})}{d\mathbf{x}^T} \equiv \left[ \frac{\partial f(\mathbf{x})}{\partial x_1} \quad \frac{\partial f(\mathbf{x})}{\partial x_1} \quad \cdots \quad \frac{\partial f(\mathbf{x})}{\partial x_n} \right] \equiv \left(\nabla f(\mathbf{x})\right)^T$$

Es existieren:

$$\frac{d\left(f(\mathbf{x}) \pm g(\mathbf{x})\right)}{d\mathbf{x}} = \frac{df(\mathbf{x})}{d\mathbf{x}} \pm \frac{dg(\mathbf{x})}{d\mathbf{x}}$$

$$\frac{d\left(f(\mathbf{x}) \cdot g(\mathbf{x})\right)}{d\mathbf{x}} = g(\mathbf{x})\frac{df(\mathbf{x})}{d\mathbf{x}} + f(\mathbf{x})\frac{dg(\mathbf{x})}{d\mathbf{x}}$$

Die Ableitung einer Funktion nach einem Vektor führt zu einem Vektor!

Die Elemente sind Partialableitungen!

Fachgebiet Prozessoptimierung

**Ein Vektor:** $\mathbf{f}(\mathbf{x}) = [\,f_1(\mathbf{x}) \quad f_2(\mathbf{x}) \quad \cdots \quad f_m(\mathbf{x})]^T$

mit $\mathbf{x} = [\,x_1 \quad x_2 \quad \cdots \quad x_n\,]^T$

Die Definition:

$$\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}} \equiv \frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}^T} \equiv \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right]_{m\times n} \quad \text{(Jacobi-Matrix)}$$

und

$$\frac{d\mathbf{f}(\mathbf{x})^T}{d\mathbf{x}} \equiv \left(\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right)^T = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_1} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f_j(\mathbf{x})}{\partial x_i}\right]_{n\times m}$$

Daher $\frac{d\mathbf{x}}{d\mathbf{x}^T} = \frac{d\mathbf{x}^T}{d\mathbf{x}} = \mathbf{E}$

Achtung: $d\mathbf{x} = d\mathbf{x}^T \quad \Rightarrow \quad \frac{d\mathbf{x}}{d\mathbf{x}^T} = \mathbf{E}$

Die Ableitung eines Vektors nach einem Vektor führt zu einer Matrix!

Fachgebiet Prozessoptimierung

Es existieren: $$\frac{d(\mathbf{f}(\mathbf{x}) \pm \mathbf{g}(\mathbf{x}))}{d\mathbf{x}} = \frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}} \pm \frac{d\mathbf{g}(\mathbf{x})}{d\mathbf{x}}$$

$$\frac{d(\mathbf{f}(\mathbf{x})^T \mathbf{g}(\mathbf{x}))}{d\mathbf{x}} = \frac{d\mathbf{f}(\mathbf{x})^T}{d\mathbf{x}} \mathbf{g}(\mathbf{x}) + \frac{d\mathbf{g}(\mathbf{x})^T}{d\mathbf{x}} \mathbf{f}(\mathbf{x})$$ (Übungsaufgabe!)

daher $$\frac{d(\mathbf{a}^T \mathbf{f}(\mathbf{x}))}{d\mathbf{x}} = \frac{d\mathbf{a}^T}{d\mathbf{x}} \mathbf{f}(\mathbf{x}) + \frac{d\mathbf{f}(\mathbf{x})^T}{d\mathbf{x}} \mathbf{a} = \frac{d\mathbf{f}(\mathbf{x})^T}{d\mathbf{x}} \mathbf{a}$$

**Lineares System:**

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{m1} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{A}\mathbf{x}$$

$$\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}} \equiv \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{m1} \end{bmatrix} = \mathbf{A}$$

**Eine quadratische Funktion:** $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$

mit $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$ $\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{bmatrix} = [a_{i,j}]_{n\times n}$

$\mathbf{A}$ ist eine symmetrische, konstante Matrix.

Daher
$$\frac{d\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)}{d\mathbf{x}} = \frac{d\mathbf{x}^T}{d\mathbf{x}}\left(\mathbf{A}\mathbf{x}\right) + \frac{d\left(\mathbf{A}\mathbf{x}\right)^T}{d\mathbf{x}}\mathbf{x}$$
$$= \mathbf{E}\left(\mathbf{A}\mathbf{x}\right) + \left(\frac{d\left(\mathbf{A}\mathbf{x}\right)}{d\mathbf{x}}\right)^T \mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} = 2\mathbf{A}\mathbf{x}$$

Dann
$$\frac{d\left(\boldsymbol{\lambda}^T\mathbf{B}\mathbf{x}\right)}{d\mathbf{x}} = \frac{d\boldsymbol{\lambda}^T}{d\mathbf{x}}\left(\mathbf{B}\mathbf{x}\right) + \frac{d\left(\mathbf{B}\mathbf{x}\right)^T}{d\mathbf{x}}\boldsymbol{\lambda}$$
$$= \left(\frac{d\left(\mathbf{B}\mathbf{x}\right)}{d\mathbf{x}}\right)^T \boldsymbol{\lambda} = \mathbf{B}^T\boldsymbol{\lambda}$$

Fachgebiet Prozessoptimierung

# Beispiel: $f(\mathbf{x}) = x_1^2 + x_1x_2 + x_2^2 - 4x_2x_3 + x_3^2 + 3x_1 - 2x_2 - 6x_3$

Die Matrix/Vektor-Form:

$$f(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & -2 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 3 & -2 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x}$$

Die ersten Ableitungen:

$$\nabla f(\mathbf{x}) = \frac{df(\mathbf{x})}{d\mathbf{x}} = \frac{d(\mathbf{x}^T \mathbf{Q} \mathbf{x})}{d\mathbf{x}} + \frac{d(\mathbf{c}^T \mathbf{x})}{d\mathbf{x}} = 2\mathbf{Q}\mathbf{x} + \mathbf{c}$$

$$= 2 \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & -2 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \\ -6 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -4 \\ 0 & -4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \\ -6 \end{bmatrix}$$

Die zweiten Ableitungen (die Hesse-Matrix):

$$\nabla^2 f(\mathbf{x}) = \frac{d(\nabla f(\mathbf{x}))}{d\mathbf{x}} = \frac{d(2\mathbf{Q}\mathbf{x} + \mathbf{c})}{d\mathbf{x}} = 2\mathbf{Q} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -4 \\ 0 & -4 & 2 \end{bmatrix}$$

**Eine Matrix:**

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_{11}(\mathbf{x}) & \cdots & f_{1l}(\mathbf{x}) \\ \vdots & \cdots & \vdots \\ f_{m1}(\mathbf{x}) & \cdots & x_{ml}(\mathbf{x}) \end{bmatrix} = [\, f_{i,j}(\mathbf{x})]_{m\times l}$$

mit

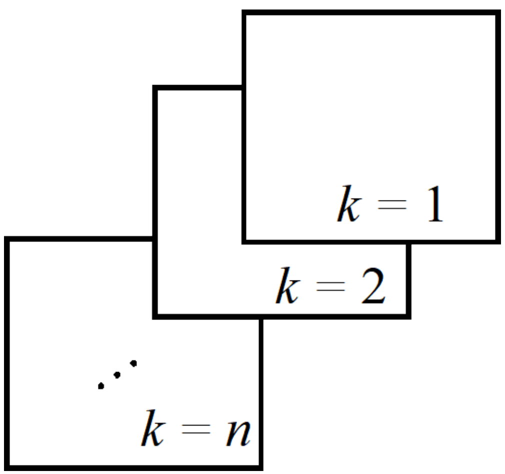
$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$$

Die Definition:

$$\frac{d\mathbf{F}(\mathbf{x})}{d\mathbf{x}} \equiv \begin{bmatrix} \frac{\partial \mathbf{F}(\mathbf{x})}{\partial x_1} & \frac{\partial \mathbf{F}(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial \mathbf{F}(\mathbf{x})}{\partial x_n} \end{bmatrix}^T$$

wobei

$$\frac{\partial \mathbf{F}(\mathbf{x})}{\partial x_k} = \begin{bmatrix} \frac{\partial f_{11}(\mathbf{x})}{\partial x_k} & \cdots & \frac{\partial f_{1l}(\mathbf{x})}{\partial x_k} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_{m1}(\mathbf{x})}{\partial x_k} & \cdots & \frac{\partial f_{m1}(\mathbf{x})}{\partial x_k} \end{bmatrix} = \left[ \frac{\partial f_{i,j}(\mathbf{x})}{\partial x_k} \right]_{m\times l} \quad k = 1,\cdots,n$$




**Der Rechenaufwand ist sehr hoch!**

# 4. Ableitungen nach Variablen in Funktionen



**Eine Funktion:** $f(\mathbf{x})$ mit $\mathbf{x}(t) = [x_1(t) \quad x_2(t) \quad \cdots \quad x_n(t)]^T$

$$\frac{df(\mathbf{x}(t))}{dt} = \frac{df(\mathbf{x}(t))}{d\mathbf{x}^T}\frac{d\mathbf{x}(t)}{dt}$$ (Übungsaufgabe!)

**Beispiel:** $f(\mathbf{x}) = \sin\left(\mathbf{c}^T\mathbf{x}\right)$

mit $\mathbf{c}^T\mathbf{x} = c_1x_1 + c_2x_2 + \cdots + c_nx_n = \sum_{i=1}^{n} c_ix_i$

$$\frac{df}{d\mathbf{x}^T} = \left(\frac{df}{d\mathbf{x}}\right)^T = \left(\frac{d}{d\mathbf{x}}\sin\left(\mathbf{c}^T\mathbf{x}\right)\right)^T = \left[\mathbf{c}\cos\left(\mathbf{c}^T\mathbf{x}\right)\right]^T = \mathbf{c}^T\cos\left(\mathbf{c}^T\mathbf{x}\right)$$

daher $$\frac{df(\mathbf{x}(t))}{dt} = \frac{df(\mathbf{x}(t))}{d\mathbf{x}^T}\frac{d\mathbf{x}(t)}{dt} = \left[\cos\left(\mathbf{c}^T\mathbf{x}(t)\right)\right]\mathbf{c}^T\dot{\mathbf{x}}(t)$$

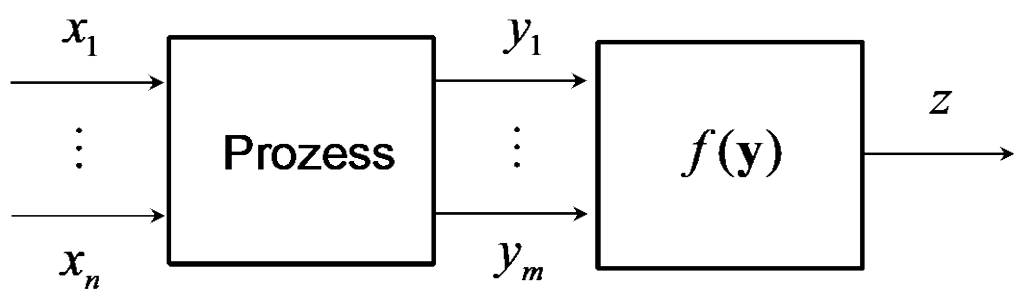
Fachgebiet Prozessoptimierung

**Eine Funktion:** $z = f(\mathbf{y})$

mit $\mathbf{y}(\mathbf{x}) = [y_1(\mathbf{x}) \quad y_2(\mathbf{x}) \quad \cdots \quad y_m(\mathbf{x})]^T$

$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$

Physikalische Bedeutung:




Weil $df = \dfrac{df}{d\mathbf{y}^T} d\mathbf{y}, \quad d\mathbf{y} = \dfrac{d\mathbf{y}}{d\mathbf{x}^T} d\mathbf{x}$ dann $df = \dfrac{df}{d\mathbf{y}^T} \dfrac{d\mathbf{y}}{d\mathbf{x}^T} d\mathbf{x}$

daher

$$\frac{df}{d\mathbf{x}} = \left( \frac{df}{d\mathbf{y}^T} \frac{d\mathbf{y}}{d\mathbf{x}^T} \right)^T = \left( \frac{d\mathbf{y}}{d\mathbf{x}^T} \right)^T \left( \frac{df}{d\mathbf{y}^T} \right)^T = \frac{d\mathbf{y}^T}{d\mathbf{x}} \frac{df}{d\mathbf{y}}$$

**Beispiel:** $$f(\mathbf{x}) = (\mathbf{Ax}-\mathbf{b})^T \mathbf{R}(\mathbf{Ax}-\mathbf{b})$$

mit $\mathbf{A} = [a_{ij}]_{m\times n}$, $\mathbf{R} = [r_{ij}]_{m\times m}$, $\mathbf{b} = [b_1 \quad \cdots \quad b_m]$

Definiert man $\mathbf{y} = \mathbf{Ax}-\mathbf{b}$ dann $f(\mathbf{y}) = \mathbf{y}^T\mathbf{R}\mathbf{y}$

Weil

$$\frac{df}{d\mathbf{x}} = \frac{d\mathbf{y}^T}{d\mathbf{x}}\frac{df}{d\mathbf{y}}$$

mit

$$\frac{d\mathbf{y}^T}{d\mathbf{x}} = \frac{d(\mathbf{Ax}-\mathbf{b})^T}{d\mathbf{x}} = \frac{d(\mathbf{Ax})^T}{d\mathbf{x}} = \left(\frac{d(\mathbf{Ax})}{d\mathbf{x}}\right)^T = \mathbf{A}^T$$

$$\frac{df}{d\mathbf{y}} = \frac{d(\mathbf{y}^T\mathbf{R}\mathbf{y})}{d\mathbf{y}} = (\mathbf{R}^T+\mathbf{R})\mathbf{y}$$

Daher

$$\frac{df}{d\mathbf{x}} = \frac{d\mathbf{y}^T}{d\mathbf{x}}\frac{df}{d\mathbf{y}} = \mathbf{A}^T(\mathbf{R}^T+\mathbf{R})(\mathbf{Ax}-\mathbf{b})$$

Fachgebiet Prozessoptimierung

# 5. Taylor-Entwicklung und Extrempunkte einer Funktion 

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \left(\frac{df}{d\mathbf{x}}\right)_0^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \left(\frac{d^2 f}{d\mathbf{x}^2}\right)_0 \Delta\mathbf{x} + \left(O^3\right)$$

$$= f(\mathbf{x}_0) + \nabla f_0^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \left(\nabla^2 f\right)_0 \Delta\mathbf{x} + \left(O^3\right)$$

$$f(\mathbf{x},\mathbf{u}) = f(\mathbf{x}_0,\mathbf{u}_0) + \begin{bmatrix} \dfrac{\partial f}{\partial \mathbf{x}^T} & \dfrac{\partial f}{\partial \mathbf{u}^T} \end{bmatrix}_0 \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{u} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \Delta\mathbf{x} & \Delta\mathbf{u} \end{bmatrix} \begin{pmatrix} \dfrac{\partial^2 f}{\partial \mathbf{x}^2} & \dfrac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{u}} \\ \dfrac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{u}} & \dfrac{\partial^2 f}{\partial \mathbf{u}^2} \end{pmatrix}_0 \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{u} \end{bmatrix} + \left(O^3\right)$$

Ist $(\mathbf{x}_0, \mathbf{u}_0)$ ein Extrempunkt, dann $\begin{bmatrix} \dfrac{\partial f}{\partial \mathbf{x}^T} & \dfrac{\partial f}{\partial \mathbf{u}^T} \end{bmatrix}_0 = \mathbf{0}$

Ist die Hesse-Matrix positiv definit, dann ist der Punkt ein Minimumpunkt.

Ist die Hesse-Matrix negativ definit, dann ist der Punkt ein Maximumpunkt.

Achtung: die Funktion $f(\mathbf{x})$ muss zweimal differenzierbar sein!

**Beispiel:** $f(\mathbf{x}) = x_1^2 + x_1 x_2 + x_2^2 - 4x_2 x_3 + x_3^2 + 3x_1 - 2x_2 - 6x_3$

(Übungsaufgabe!)

Fachgebiet Prozessoptimierung

# 6. Das Lagrange-Verfahren



**Problemdefinition eines Problems mit Gleichungsbeschränkungen:**

$$\min_{x} \quad f(\mathbf{x})$$

$$\text{mit} \quad \mathbf{g}(\mathbf{x}) = \mathbf{0}$$

wobei $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$

$\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}) \quad g_2(\mathbf{x}) \quad \cdots \quad g_m(\mathbf{x})]^T$

Welche Beziehung besteht zwischen $n$ und $m$ ?

Lagrange-Funktion: $$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x})$$

Die notwendige Bedingung: $$\frac{\partial L}{\partial \mathbf{x}} = \frac{\partial f}{\partial \mathbf{x}} + \frac{\partial \mathbf{g}^T}{\partial \mathbf{x}} \boldsymbol{\lambda} = \mathbf{0}$$

also

$$\boxed{\begin{aligned} &\nabla L = \nabla f + \nabla \mathbf{g}^T \boldsymbol{\lambda} = \mathbf{0} \\ &\mathbf{g}(\mathbf{x}) = \mathbf{0} \end{aligned}}$$

Die Modellgleichungen:

Die hinreichende Bedingung: positiv definite Hesse-Matrix von $L(\mathbf{x}, \boldsymbol{\lambda})$

# Beispiel:

$$\min_{x} \quad x_1^2 + x_2^2$$

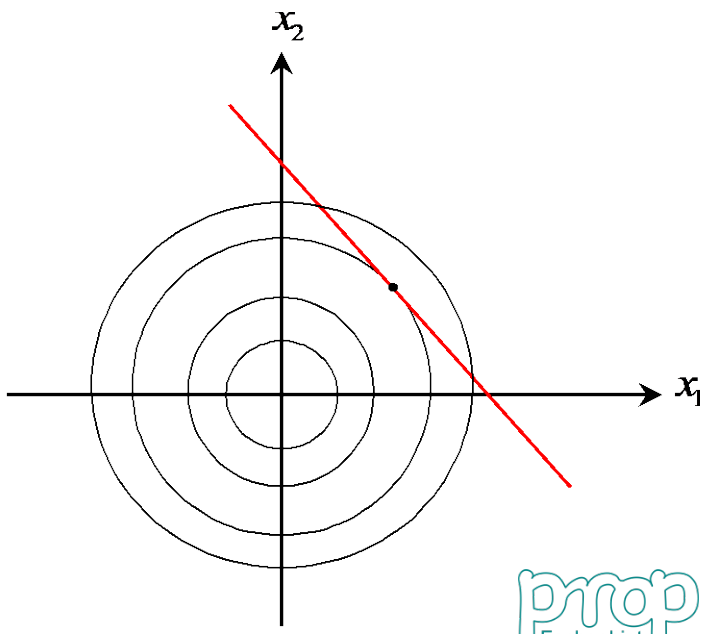
$$\text{mit} \quad x_1 + x_2 = 1$$

Lagrange-Funktion: $L(\mathbf{x}, \boldsymbol{\lambda}) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1)$

dann

$$\frac{\partial L}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial L}{\partial x_1} \\ \dfrac{\partial L}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 - 1 = 0$$




Der Lösungspunkt: $x_1^* = x_2^* = \dfrac{1}{2}$

Die Hesse-Matrix: $\mathbf{H}_L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

# 7. Die Karush-Kuhn-Tucker-Bedingungen (KKT)



**Problemdefinition eines Problems mit Gleichungs- und Ungleichungsbeschränkungen:**

$$\min_x \; f(\mathbf{x}) \qquad \text{wobei} \qquad \mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$$

$$\text{mit} \quad \mathbf{g}(\mathbf{x}) = \mathbf{0} \qquad \mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}) \quad g_2(\mathbf{x}) \quad \cdots \quad g_m(\mathbf{x})]^T$$

$$\mathbf{h}(\mathbf{x}) \geq \mathbf{0} \qquad \mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}) \quad h_2(\mathbf{x}) \quad \cdots \quad h_r(\mathbf{x})]^T$$

Lagrange-Funktion: $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) - \boldsymbol{\mu}^T \mathbf{h}(\mathbf{x})$

Beim Lösungspunkt:

- $\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$
- $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$
- $\mathbf{h}(\mathbf{x}^*) \geq \mathbf{0}$
- $\mu_i \, h_i(\mathbf{x}^*) = 0, \quad \mu_i \geq 0, \quad i = 1, \cdots, r$
- $\nabla^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ positiv definit