Technische Hochschule Wildau [FH]
Technical University of Applied Sciences

WILDAU

Maschinelles Lernen
Mathematische Grundlagen – Multivariate Analysis

Prof. Dr. Rainer Stollhoff

Motivation: Multivariate Regression

Aufgabe: Regression, d.h. Vorhersage $\hat{y} = \hat{y}(x) = f(x)$

Erfahrung: Datensatz $(x_i, y_i)_{i=1}^n$

Qualität: Quadratische Verlustfunktion

$$L(y, \hat{y}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - f(x_i; \theta))^2 = L(\theta)$$

Maschine: vereinfachte lineare Regression mit

$$\mathrm{f}(x_1, x_2; \theta_0, \theta_1, \theta_2) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2$$

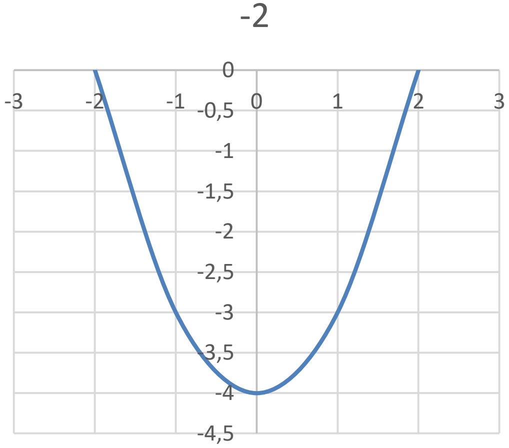
Lernen: Finde Werte für $\theta = (\theta_0, \theta_1, \theta_2)$, die die quadratische Verlustfunktion minimieren

Inhaltsübersicht

Beispiel: Funktion mehrerer Veränderlicher

- $f(x_1) = x_1^2$

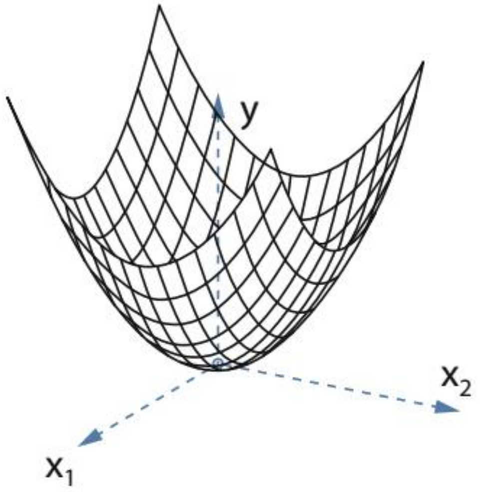
x1	f(x1)
-2	4
-1	1
0	0
1	1
2	4



$$f(x_1) = g(x_1, 0\ )$$

- $g(x_1, x_2) = x_1^2 + x_2^2$

		x2				
	$x_1^2 + x_2^2$	-2	-1	0	1	2
	-2	8	5	4	5	8
	-1	5	2	1	2	5
x1	0	4	1	0	1	4
	1	5	2	1	2	5
	2	8	5	4	5	8

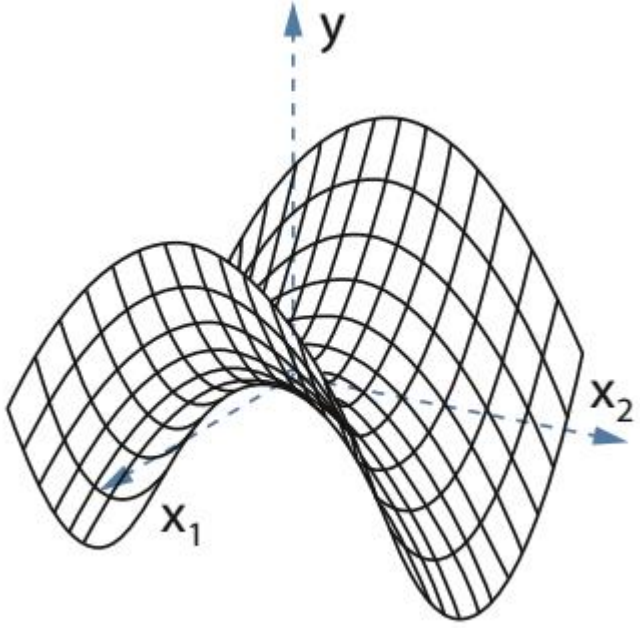


$y = x_1^2 + x_2^2$

Funktion mehrerer Veränderlicher

- $f(x_1, x_2) = x_1^2$
- $g(x_1, x_2) = x_1^2 + x_2^2$
- $h(x_1, x_2) = x_1^2 - x_2^2$

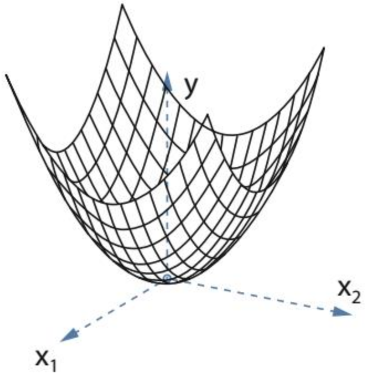
		x2				
	$x_1^2 - x_2^2$	-2	-1	0	1	2
x1	-2	0	3	4	3	0
	-1	-3	0	1	0	-3
	0	-4	-1	0	-1	-4
	1	-3	0	1	0	-3
	2	0	3	4	3	0



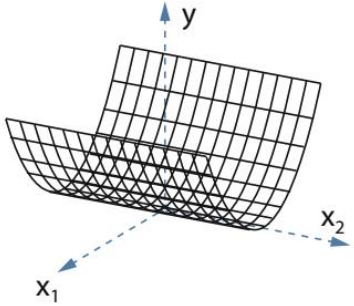
$y = x_1^2 - x_2^2$

Funktion mehrerer Veränderlicher

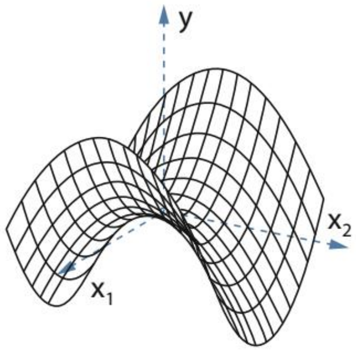
- $f(x_1, x_2) = x_1^2$
- $g(x_1, x_2) = x_1^2 + x_2^2$
- $h(x_1, x_2) = x_1^2 - x_2^2$



$y = x_1^2 + x_2^2$



$y = x_1^2$



$y = x_1^2 - x_2^2$

Wir gehen davon aus, dass $x_1, x_2, \ldots, x_n$ reellwertige, voneinander unabhängige Veränderliche sind

$$x_i \in \mathbb{R} \quad (i = 1, 2, \ldots, n)$$

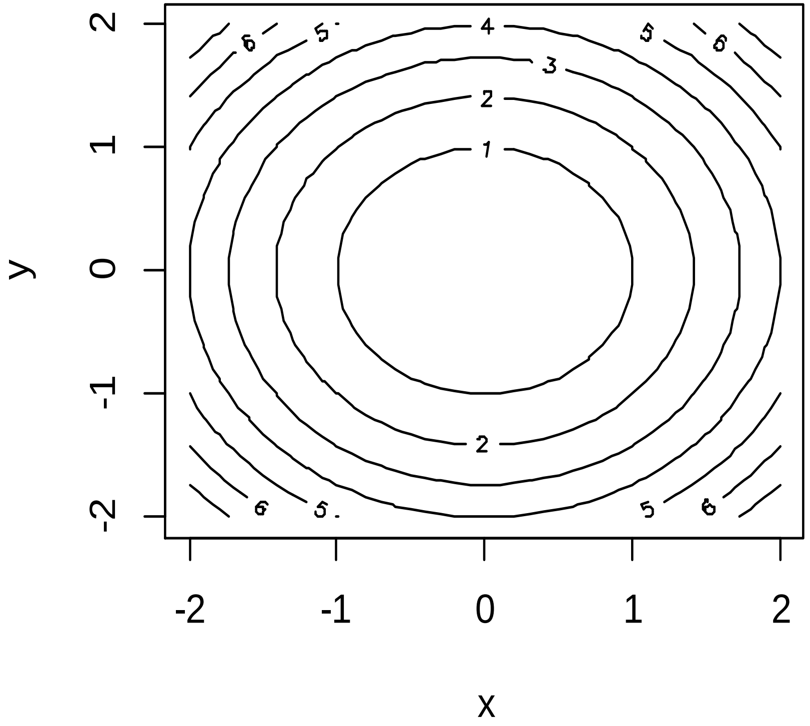
und nennen eine Vorschrift f, mit der jeder zum Definitionsbereich $\mathbb{D}(f)$ von f gehörenden Kombination $(x_1, x_2, \ldots, x_n)$ der Veränderlichen genau ein Wert $f(x_1, x_2, \ldots, x_n) \in \mathbb{R}$ zugeordnet wird, eine reelle **Funktion f der n unabhängigen Variablen $x_1, x_2, \ldots, x_n$**. Als Zuordnungsvorschrift benutzen wir die **Funktionsgleichung**

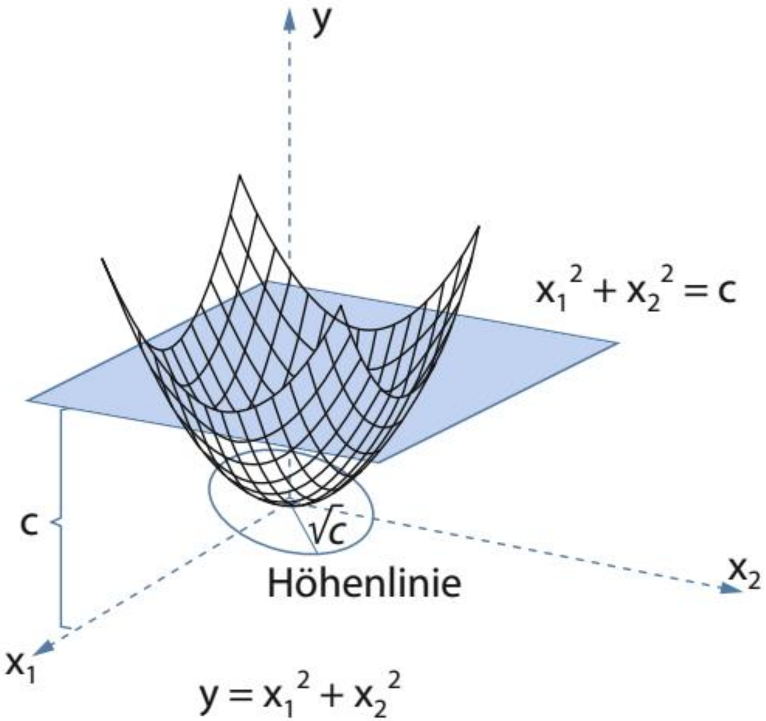
$$y = f(x_1, x_2, \ldots, x_n)\,.$$

Höhenlinien

- $f(x_1, x_2) = x_1^2$
- $g(x_1, x_2) = x_1^2 + x_2^2$
- $g(x_1, x_2) = c$?

		x2				
	$x_1^2 + x_2^2$	-2	-1	0	1	2
x1	-2	8	5	4	5	8
	-1	5	2	1	2	5
	0	4	1	0	1	4
	1	5	2	1	2	5
	2	8	5	4	5	8

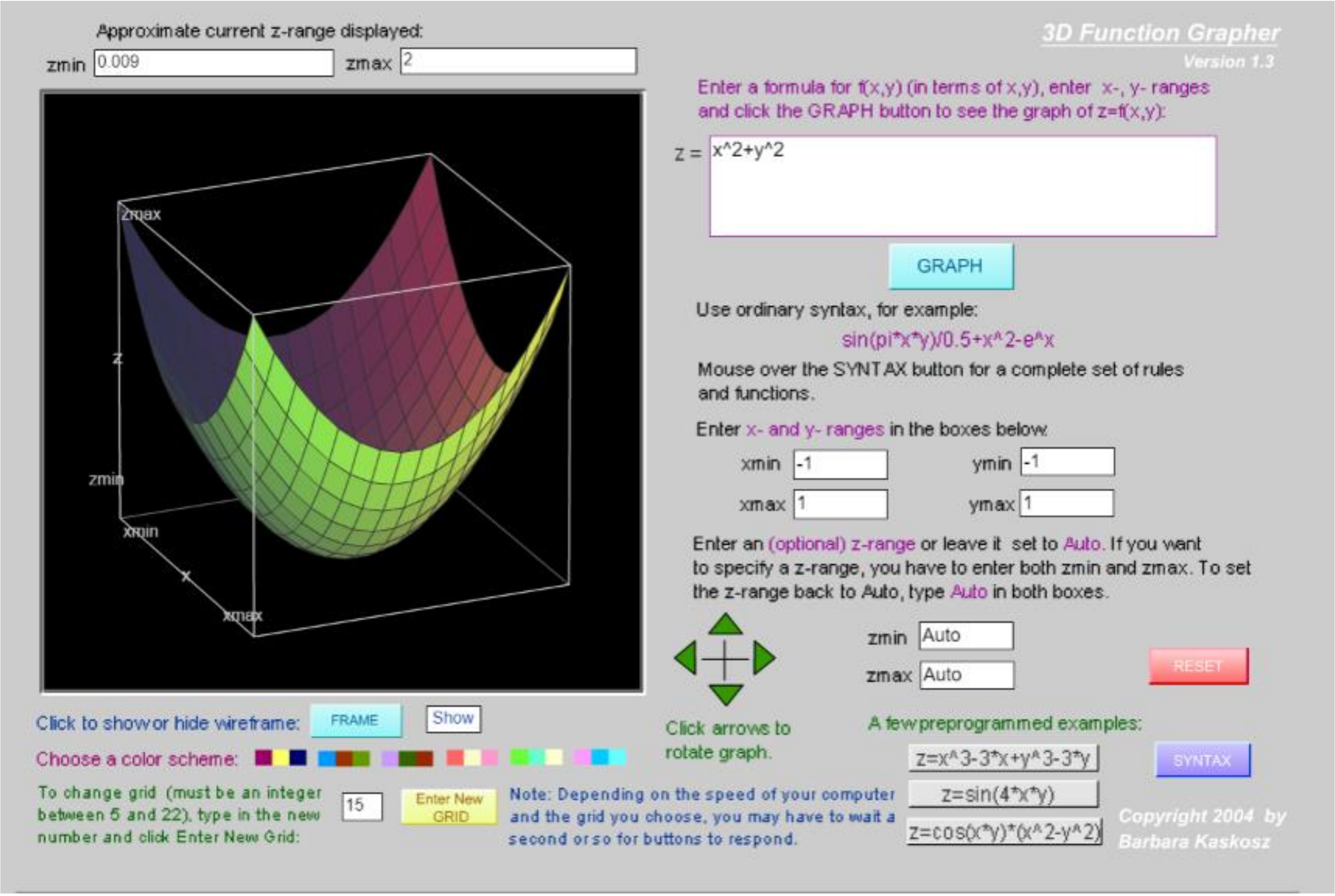




3D Function Grapher

Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU

http://www.math.uri.edu/~bkaskosz/flashmo/graph3d2/



Inhaltsübersicht

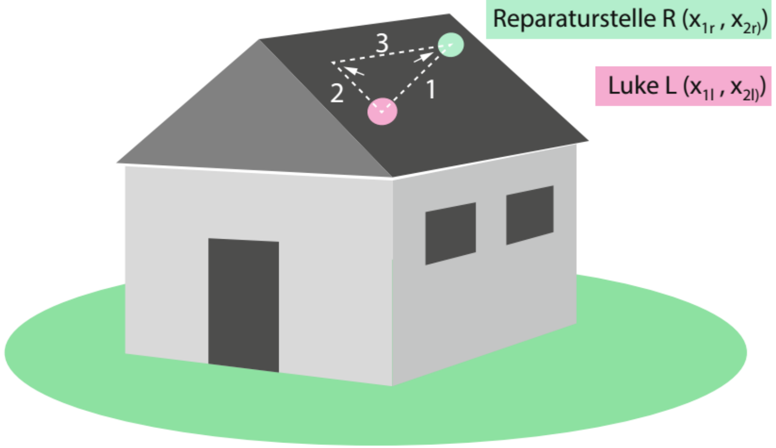
- Motivation / Einleitung
- Funktionen mehrerer Veränderlicher
- **Partielle Ableitungen**
- Lineare Algebra

Partielle Ableitungen

- Erster Ordnung = Ableitung in Richtung einer der Veränderlichen

- Zweiter Ordnung

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i} f = f'_{x_i} \,.$$



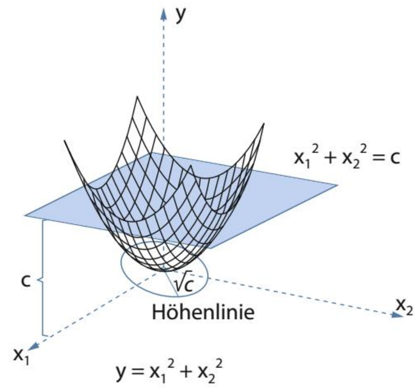


Abb. 7.4 Höhenlinie von $y = x_1^2 + x_2^2$

$$\begin{aligned}\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x_1, x_2, \ldots, x_n) &= \frac{\partial^2}{\partial x_j \partial x_i} f(x_1, x_2, \ldots, x_n) \\ &= \frac{\partial^2 f(x_1, x_2, \ldots, x_n)}{\partial x_j \partial x_i} \\ &= f''_{x_i x_j}(x_1, x_2, \ldots, x_n)\,.\end{aligned}$$

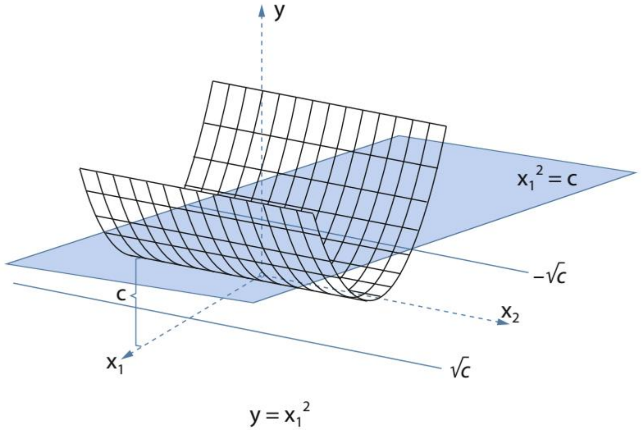


Abb. 7.5 Höhenlinien von $y = x_1^2$

Partielle Ableitungen

Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU

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

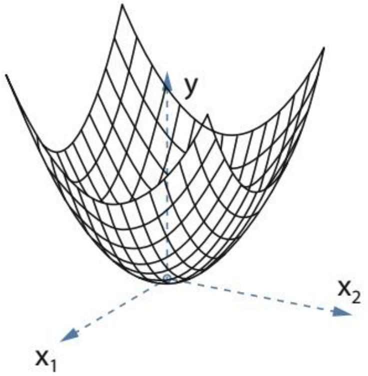
- $\frac{\partial f(x_1,x_2)}{\partial x_1} =$
- $\frac{\partial f(x_1,x_2)}{\partial x_2} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial {x_1}^2} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial {x_2}^2} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial x_2 \quad \partial x_1} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial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

$$f(x_1, x_2) = 5\sqrt{x_1 x_2}$$

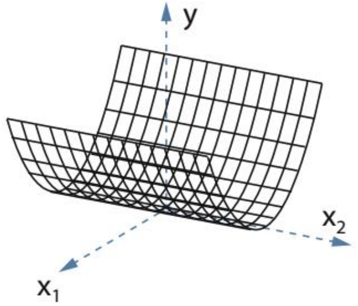
- $\frac{\partial f(x_1,x_2)}{\partial x_1} =$
- $\frac{\partial f(x_1,x_2)}{\partial x_2} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial {x_1}^2} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial {x_2}^2} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial x_2 \quad \partial x_1} =$
- $\frac{\partial^2 f(x_1,x_2)}{\partial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

Extremwerte Beispiele

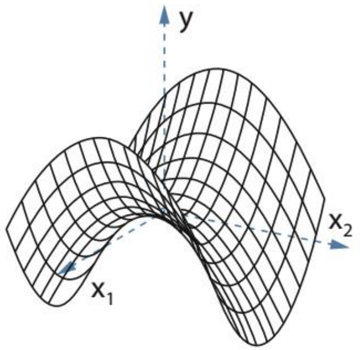
- $f(x_1, x_2) = x_1^2$
- $g(x_1, x_2) = x_1^2 + x_2^2$
- $h(x_1, x_2) = x_1^2 - x_2^2$



$y = x_1^2 + x_2^2$



$y = x_1^2$



$y = x_1^2 - x_2^2$

Merksatz

Sei $f = f(x_1, x_2)$ eine Funktion mit zwei voneinander unabhängigen Variablen x_1 und x_2. Die Funktion f sei zweimal partiell differenzierbar, d. h., es mögen die partiellen Ableitungen erster Ordnung f'_{x_1} und f'_{x_2} sowie die partiellen Ableitungen zweiter Ordnung $f'_{x_1x_1}$ und $f'_{x_1x_2}$ ebenso wie $f'_{x_2x_1}$ und $f'_{x_2x_2}$ existieren. Dann gilt: Die Funktion f besitzt an der Stelle $P_0 = P_0(x_{1_0}, x_{2_0})$ einen **lokalen Extremwert**, wenn die Bedingungen

$$f'_{x_1}(x_{1_0}, x_{2_0}) = f'_{x_2}(x_{1_0}, x_{2_0}) = 0$$

und

$$f''_{x_1x_1}(x_{1_0}, x_{2_0}) \cdot f''_{x_2x_2}(x_{1_0}, x_{2_0}) - (f''_{x_1x_2}(x_{1_0}, x_{2_0}))^2 > 0$$

erfüllt sind.

Sofern $f''_{x_1x_1}(x_{1_0}, x_{2_0}) < 0$ ist, handelt es sich um ein **lokales Maximum**. Ist $f''_{x_1x_1}(x_{1_0}, x_{2_0}) > 0$, liegt ein **lokales Minimum** vor.

Beispiel:

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

- $\frac{\partial f(x_1, x_2)}{\partial x_1} = 0$
- $\frac{\partial f(x_1, x_2)}{\partial x_2} = 0$
- $\frac{\partial^2 f(x_1, x_2)}{\partial {x_1}^2} \cdot \frac{\partial^2 f(x_1, x_2)}{\partial {x_2}^2} - \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \ \partial x_2} > 0$

Gradientenvektor und Hesse-Matrix

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i} f = f'_{x_i} \,.$$

$$\begin{aligned}\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x_1, x_2, \ldots, x_n) &= \frac{\partial^2}{\partial x_j \partial x_i} f(x_1, x_2, \ldots, x_n) \\ &= \frac{\partial^2 f(x_1, x_2, \ldots, x_n)}{\partial x_j \partial x_i} \\ &= f''_{x_i x_j}(x_1, x_2, \ldots, x_n)\,.\end{aligned}$$

Gradientenvektor
Partielle Ableitungen 1. Ordnung

$$\nabla f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix}$$

Hesse-Matrix
Partielle Ableitungen 2. Ordnung

$$\nabla(\nabla f(x)) = \mathrm{H}_f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1} f(x) & \cdots & \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_n} f(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n}\frac{\partial}{\partial x_1} f(x) & \cdots & \frac{\partial}{\partial x_n}\frac{\partial}{\partial x_n} f(x) \end{pmatrix}$$

Inhaltsübersicht

Technische Hochschule Wildau
Technical University of Applied Sciences

- Motivation / Einleitung
- Funktionen mehrerer Veränderlicher
- Partielle Ableitungen
- **Lineare Algebra**

Vektoren und Matrizen

Tabellen

	Produkt 1	Produkt 2	Produkt 3	Produkt 4
Kunde 1	20	300	10	100
Kunde 2	55	80	0	210
Kunde 3	0	1000	70	350

Matrizen

mxn

$$\begin{pmatrix} 20 & 300 & 10 & 100 \\ 55 & 80 & 0 & 210 \\ 0 & 1000 & 70 & 350 \end{pmatrix}$$

Zeilenvektor 1xn Spaltenvektor mx1

$$(10; 4; 6; 2) \quad \text{bzw.} \quad \begin{pmatrix} 10 \\ 4 \\ 6 \\ 2 \end{pmatrix}$$

Quadratische Matrix

$$\begin{pmatrix} 1 & 3 & -4 \\ 3 & 12 & -2 \\ -4 & -2 & 5 \end{pmatrix}.$$

Obere (untere) Dreiecksmatrix

$$\begin{pmatrix} -2 & 1 & 10 & -5 \\ 0 & 4 & -3 & 9 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 11 \end{pmatrix}$$

Diagonalmatrix

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Einheitsmatrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Rechnen mit Vektoren und Matrizen

Technische Hochschule Wildau
Technical University of Applied Sciences

- Transponieren

$$\begin{pmatrix} 20 & 300 & 10 & 100 \\ 55 & 80 & 0 & 210 \\ 0 & 1000 & 70 & 350 \end{pmatrix}^T = \begin{pmatrix} 20 & 55 & 0 \\ 300 & 80 & 1000 \\ 10 & 0 & 70 \\ 100 & 210 & 350 \end{pmatrix}$$

- Addieren

$$\begin{pmatrix} 20 & 300 & 10 & 100 \\ 55 & 80 & 0 & 210 \\ 0 & 1000 & 70 & 350 \end{pmatrix} + \begin{pmatrix} 16 & 240 & 8 & 80 \\ 44 & 64 & 0 & 168 \\ 0 & 800 & 56 & 280 \end{pmatrix} = \begin{pmatrix} 36 & 540 & 18 & 180 \\ 99 & 144 & 0 & 378 \\ 0 & 1800 & 126 & 630 \end{pmatrix}.$$

- Multiplizieren mit Ska

$$0{,}2 \cdot \begin{pmatrix} 20 & 300 & 10 & 100 \\ 55 & 80 & 0 & 210 \\ 0 & 1000 & 70 & 350 \end{pmatrix} = \begin{pmatrix} 4 & 60 & 2 & 20 \\ 11 & 16 & 0 & 42 \\ 0 & 200 & 14 & 70 \end{pmatrix}$$

- Matrix mal Vektor

$$\begin{array}{c} \text{Richtung} \rightarrow \\ \begin{pmatrix} \boxed{20 \quad 300 \quad 10 \quad 100} \\ 55 \quad 80 \quad 0 \quad 210 \\ 0 \quad 1000 \quad 70 \quad 350 \end{pmatrix} \end{array} \rightarrow \begin{pmatrix} \boxed{\begin{matrix} 10 \\ 4 \\ 6 \\ 2 \end{matrix}} \end{pmatrix} \downarrow \text{Richtung}$$

$$= \begin{pmatrix} 1660 \\ 1290 \\ 5120 \end{pmatrix}.$$

(3x4)·(4x1)=(3x1)

$A \cdot B = C$					0	4	6
					-1	2	-2
					3	4	1
					-4	7	0
					1	-2	6
1	-0,5	0	1,5	-2	-7,5	$c_{1,2}$	-5
6	4	-2	1	0	-14	31	26
0,5	3	-1	-3	10	$c_{3,1}$	-37	56
2	2	-6	1	4	-20	-13	$c_{3,3}$

(4x5)·(5x3)=(4x3)

Inverse einer Matrix

Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU

- Inverse

$$(X^{-1})X = E = X(X^{-1})$$

  - Lineares Gleichungssystem
    Gauß-Verfahren

$$\left(\begin{matrix} 3000 & 2500 \\ 7500 & 1300 \end{matrix}\;\middle|\;\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}\right)$$

  - Ziel & Ergebnis

$$\left(\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}\;\middle|\;\begin{matrix} X_{11}^{-1} & X_{12}^{-1} \\ X_{21}^{-1} & X_{22}^{-1} \end{matrix}\right)$$

  - Lösung

$$X^{-1} = \begin{pmatrix} X_{11}^{-1} & X_{12}^{-1} \\ X_{21}^{-1} & X_{22}^{-1} \end{pmatrix}$$

  - Probe

$$(X^{-1})X =$$

Lineare Gleichungssysteme und Inverse

Technische Hochschule Wildau
Technical University of Applied Sciences

- Lineares Gleichungssystem
  Gauß-Verfahren
  - Kurzschreibweise $\begin{pmatrix} 2 & 2 & 4 \\ 4 & 6 & 5 \\ 2 & 3 & 3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$
  - Rechenoperationen
  - Ziel und Ergebnis
  - Lösung:

$$\left(\begin{array}{ccc|c} 2 & 2 & 4 & 2 \\ 4 & 6 & 5 & 2 \\ 2 & 3 & 3 & 4 \end{array}\right)$$

- Vertauschen zweier Zeilen,
- Multiplizieren einer Zeile mit einer Zahl ungleich null,
- Addition einer Zeile zu einer anderen Zeile.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -19 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 6 \end{array}\right)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -19 \\ 8 \\ 6 \end{pmatrix}$$

- Inverse

$$(A^{-1})A = E = A(A^{-1})$$

  - Lineares Gleichungssytem
    Gauß-Verfahren
  - Ziel & Ergebnis

$$\left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ -4 & 3 & 0 & 1 \end{array}\right)$$

  - Lösung und Probe

$$\left(\begin{array}{cc|cc} 1 & 0 & \frac{3}{10} & -\frac{1}{10} \\ 0 & 1 & \frac{2}{5} & \frac{1}{5} \end{array}\right)$$

- LGS mittels Inverser

$$A^{-1} \cdot A = \begin{pmatrix} \frac{3}{10} & -\frac{1}{10} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ -4 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A \cdot x = b \quad \rightarrow \quad A^{-1} \cdot A \cdot x = x = A^{-1} \cdot b$$

Determinanten und lineare Abhängigkeit

Matrix mit linear unabhängigen Zeilen

$$X = \begin{pmatrix} 3 & 4 \\ 6 & 1 \end{pmatrix}$$

$\Longleftrightarrow$

Determinante ungleich Null

$$\det(X) = |X| = 3 \cdot 1 - 6 \cdot 4 = -21$$

Matrix mit linear abhängigen Zeilen

$$Y = \begin{pmatrix} 3 & 4 \\ 6 & 8 \end{pmatrix}$$

$\Longleftrightarrow$

Determinante gleich Null

$$\det(X) = |X| = 3 \cdot 8 - 6 \cdot 4 = 0$$

Matrix mit linear abhängigen oder unabhängigen Zeilen?

$$Z = \begin{pmatrix} 3 & 4 & 1 \\ 7 & 8 & 3 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\begin{aligned} &\det(X) = |X| = \\ &+3 \cdot (8 \cdot 1 - 0 \cdot 3) \\ &-7 \cdot (4 \cdot 1 - 0 \cdot 1) \\ &+1 \cdot (4 \cdot 3 - 8 \cdot 1) \\ &= 3 \cdot 8 - 7 \cdot 4 + 1 \cdot 4 = 0 \end{aligned}$$

$\Rightarrow$

Matrix mit linear abhängigen Zeilen