

# Kapitel 9

## Differentialrechnung im $\mathbb{R}^n$

Bisher haben wir uns mit Funktionen beschäftigt, deren Verhalten durch eine einzelne Variable beschrieben wird. In der Praxis reichen solche Funktionen in der Regel nicht aus, um damit reale Problem zu modellieren. Vielmehr hängt das Ergebnis von mehreren relevanten Faktoren ab, womit man zu Funktionen mit mehreren Veränderlichen bzw. Variablen gelangt. Für diese mehrdimensionalen Funktionen lassen sich viele Eigenschaften, die wir für eindimensionale Funktionen kennen gelernt haben, in ähnlicher Form definieren. Einige Aspekte sind im Mehrdimensionalen allerdings komplexer und auch weniger intuitiv. Wir beschränken uns auf eine relativ knappe Einführung der mehrdimensionalen Funktionen. Beispiele beschränken sich im Wesentlichen auf Funktionen mit zwei oder drei Variablen, deren Verhalten sich noch geometrisch vorstellen lässt. Das mathematische Konzept ist aber für beliebige Dimensionen gültig. Bevor wir uns mit mehrdimensionalen Funktionen und deren Eigenschaften beschäftigen, werden im ersten Abschnitt einige Grundlagen des  $\mathbb{R}^n$  definiert.

### 9.1 Grundlagen des $\mathbb{R}^n$

**Definition 9.1** (Kartesisches Produkt). *Seien  $A_1, A_2, \dots, A_n$  beliebige Mengen. Die Menge*

$$A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \dots, a_n \in A_n\}$$

*wird kartesisches Produkt genannt. Ihre Elemente  $(a_1, a_2, \dots, a_n)$  heißen n-Tupel und jeder Eintrag  $a_i$  ( $i = 1, \dots, n$ ) Komponente. Zwei n-Tupel  $a$  und  $b$  sind gleich, wir schreiben  $a = b$ , wenn  $a_i = b_i$  für alle  $i = 1, \dots, n$ .*

**Notation:** Sind alle Mengen  $A_i$  gleich, also etwa  $A_i = A$  für  $i = 1, \dots, n$ , so schreibt man kurz  $A^n$  für das kartesische Produkt.

**Definition 9.2.** *Die Menge  $\mathbb{R}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{R} \text{ für } i = 1, \dots, n\}$  heißt n-dimensionaler Euklidischer Raum. Die n-Tupel  $x \in \mathbb{R}^n$  werden auch als Vektoren bezeichnet.*

Anmerkung: Typischerweise werden Vektoren als Spaltenvektoren aufgefasst. Also

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_1, \dots, x_n)^T, \text{ wobei das hochgestellte } T \text{ kennzeichnet, dass der}$$

Zeilenvektor transponiert und damit ein Spaltenvektor ist. Für Vektoren kann man arithmetische Operatoren und Vergleichsoperatoren definieren.

**Definition 9.3.** Seien  $x, y \in \mathbb{R}^n$  und  $\lambda \in \mathbb{R}$ .

1.  $x + y = (x_1 + y_1, \dots, x_n + y_n)^T$  Addition
2.  $\lambda x = (\lambda x_1, \dots, \lambda x_n)^T$  Skalarmultiplikation
3.  $x^T y = x_1 y_1 + \dots + x_n y_n$  Skalarprodukt
4.  $\|x\| = \sqrt{x^T x}$  Euklidische Norm
5.  $d(x, y) = \|x - y\|$  Abstand von  $x$  und  $y$
6.  $x \leq y \Leftrightarrow \forall i = 1, \dots, n. x_i \leq y_i$  „kleiner gleich“
7.  $x < y \Leftrightarrow \forall i = 1, \dots, n. x_i < y_i$  „kleiner“

**Notation:** Seien  $a, b \in \mathbb{R}^n$ . Dann bezeichnet  $[a, b] = \{x \in \mathbb{R}^n \mid a \leq x \leq b\} = [a_1, b_1] \times \dots \times [a_n, b_n]$  und  $(a, b) = \{x \in \mathbb{R}^n \mid a < x < b\} = (a_1, b_1) \times \dots \times (a_n, b_n)$  sowie entsprechend für  $[a, b)$  und  $(a, b]$ . Damit haben wir abgeschlossene, offene und halboffenen Intervalle für Vektoren definiert.

**Definition 9.4.** Sei  $\varepsilon > 0$  und  $x \in \mathbb{R}^n$ . Dann wird

$$\mathcal{U}_\varepsilon(x) = \{\tilde{x} \in \mathbb{R}^n \mid \|\tilde{x} - x\| < \varepsilon\}$$

eine  $\varepsilon$ -Umgebung von  $x$  genannt. Für eine Menge  $M \subseteq \mathbb{R}^n$  heißt ein Element  $x \in \mathbb{R}^n$

- innerer Punkt von  $M$ , wenn ein  $\varepsilon > 0$  mit  $\mathcal{U}_\varepsilon(x) \subset M$  existiert;
- Randpunkt von  $M$ , wenn für jedes  $\varepsilon > 0$  Elemente  $\tilde{x}, \hat{x} \in \mathcal{U}_\varepsilon(x)$  existieren mit  $\tilde{x} \in M$  und  $\hat{x} \notin M$ ;
- isolierter Punkt von  $M$ , wenn ein  $\varepsilon > 0$  mit  $\mathcal{U}_\varepsilon(x) \cap M = \{x\}$  existiert;
- Häufungspunkt von  $M$ , wenn für jedes  $\varepsilon > 0$  ein Element  $\tilde{x} \in \mathcal{U}_\varepsilon(x) \cap M$  mit  $\tilde{x} \neq x$  existiert.

Die Menge aller inneren Punkte von  $M$  wird mit  $\overset{\circ}{M}$  oder  $\text{int}(M)$ , die Menge aller Randpunkte mit  $\partial M$  bezeichnet. Eine Menge  $M \subseteq \mathbb{R}^n$  heißt **offen**, wenn jeder Punkt von  $M$  ein innerer Punkt ist, und sie heißt **abgeschlossen**, wenn sie alle ihre Häufungspunkte enthält.

**Beispiel 9.1.**  $(a, b) \subseteq \mathbb{R}^n$  und  $\mathcal{U}_\varepsilon(x)$  sind offene Mengen, während  $[a, b]$  und  $\{y \in \mathbb{R}^n \mid \|y - x\| \leq \varepsilon\}$  abgeschlossene Mengen sind.

$M \setminus \partial M$  ist offen.

$M \cup \partial M$  ist abgeschlossen.

Der Begriff der Konvergenz lässt sich relativ einfach auf den  $\mathbb{R}^n$  übertragen.

**Definition 9.5** (Normkonvergenz). *Eine Folge  $(x_k)$  mit  $x_k \in \mathbb{R}^n$  für  $k \in \mathbb{N}$  ist konvergent gegen  $a \in \mathbb{R}^n$  genau dann wenn  $\lim_{k \rightarrow \infty} \|x_k - a\| = 0$ .*

**Satz 9.6.** *Im  $\mathbb{R}^n$  ist Normkonvergenz gleichbedeutend mit komponentenweiser Konvergenz, also*

$$\lim_{k \rightarrow \infty} \|x^{(k)} - \tilde{x}\| = 0 \Leftrightarrow \forall i = 1, \dots, n. \lim_{k \rightarrow \infty} x_i^{(k)} = \tilde{x}_i;$$

Für den Beweis dieses Satzes ist folgendes Resultat hilfreich:

**Lemma 9.7.** *Sei  $x \in \mathbb{R}^n$  und  $\|\cdot\|$  die Euklidische Norm. Es gilt*

$$0 \leq \max\{|x_1|, \dots, |x_n|\} \leq \|x\| \leq \sqrt{n} \max\{|x_1|, \dots, |x_n|\}.$$

**Beweis:** Quadrieren der Ungleichungen liefert

$$0 \leq \max\{x_1^2, \dots, x_n^2\} \leq \sum_{i=1}^n x_i^2 \leq n \max\{x_1^2, \dots, x_n^2\}$$

und damit die Behauptung.  $\square$

**Beweis:** (von Satz 9.6)

„ $\Rightarrow$ “ Da wegen Lemma 9.7 gilt, dass

$$\begin{aligned} \|x^{(k)} - \tilde{x}\| &= \sqrt{(x_1^{(k)} - \tilde{x}_1)^2 + \dots + (x_n^{(k)} - \tilde{x}_n)^2} \\ &\geq \max\{|x_i^{(k)} - \tilde{x}_i| : i = 1, \dots, n\} \geq 0 \end{aligned}$$

und  $\lim_{k \rightarrow \infty} \max\{|x_i^{(k)} - \tilde{x}_i| : i = 1, \dots, n\} = 0 \Leftrightarrow \forall i = 1, \dots, n. \lim_{k \rightarrow \infty} x_i^{(k)} = \tilde{x}_i$ , folgt komponentenweise Konvergenz aus Normkonvergenz (s. Satz 3.10, Sandwich-Theorem).

„ $\Leftarrow$ “ Da  $\lim_{k \rightarrow \infty} \max\{|x_i^{(k)} - \tilde{x}_i| : i = 1, \dots, n\} = 0 \Leftrightarrow \forall i = 1, \dots, n. \lim_{k \rightarrow \infty} x_i^{(k)} = \tilde{x}_i$  und  $\max\{|x_i^{(k)} - \tilde{x}_i| : i = 1, \dots, n\} \geq \frac{1}{\sqrt{n}} \|x^{(k)} - \tilde{x}\| \geq 0$  wegen Lemma 9.7 folgt Normkonvergenz aus komponentenweiser Konvergenz (s. Satz 3.10, Sandwich-Theorem).  $\square$

$\lim_{k \rightarrow \infty} x^{(k)} = a$  für  $x^{(k)} \in \mathbb{R}^n$  bedeutet also, dass der Abstand  $\lim_{k \rightarrow \infty} d(x^{(k)}, a) = \lim_{k \rightarrow \infty} \|x^{(k)} - a\| = 0$  gegen Null geht. Dies ist gleichbedeutend damit, dass  $x^{(k)}$  komponentenweise gegen  $a$  konvergiert. Zum Beispiel  $\lim_{k \rightarrow \infty} x^{(k)} = \lim_{k \rightarrow \infty} (2 + \frac{1}{k}, \frac{2}{k})^T = (2, 0)$ .

## 9.2 Stetigkeit im $\mathbb{R}^n$

Den Begriff der Stetigkeit von Funktionen erweitern wir nun auf Funktionen mit mehreren Variablen. Dazu definieren wir zuerst den Begriff des Grenzwertes, den wir im vorherigen Abschnitt schon informell benutzt haben.

**Definition 9.8** (Grenzwert im  $\mathbb{R}^n$ ). *Sei  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$  eine Funktion und  $a$  ein Häufungspunkt von  $M$ . Dann heißt  $\lim_{x \rightarrow a} f(x) = b \Leftrightarrow \forall \varepsilon > 0. \exists \delta > 0. |f(x) - b| < \varepsilon$  für alle  $x \in M \setminus \{a\}$  mit  $\|x - a\| < \delta$  der Grenzwert von  $f$  in  $a$ .*

Eine Folge im  $\mathbb{R}^n$  ist analog zu einer Folge in  $\mathbb{R}$  als  $(x_k)_{k \in \mathbb{N}}$  mit  $x_k = (x_{k,1}, \dots, x_{k,n}) \in \mathbb{R}^n$  definiert. Eine Folge konvergiert gegen den Grenzwert  $a \in \mathbb{R}^n$  gdw. es zu jedem  $\varepsilon > 0$  ein  $k_0 \in \mathbb{N}$  gibt, so dass  $|x_{k,i} - a_i| < \varepsilon$  für alle  $k \geq k_0$  und  $i \in \{1, \dots, n\}$  gibt. In analoger Form können wir auch den Begriff der Cauchy-Folge auf den  $\mathbb{R}^n$  übertragen.

**Beispiel 9.2.** Wir betrachten die Folge  $x_k = \left(\frac{1}{k}, \frac{2k}{3k+4}\right)$ . Der Grenzwert dieser Folge ist  $(0, \frac{2}{3})$ , wie man leicht nachprüfen kann. So erhält man für  $\varepsilon = 0.1$   $k_0 = 11$  (nachrechnen).

**Definition 9.9** (Stetigkeit im  $\mathbb{R}^n$ ). Eine Funktion  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$  heißt stetig in  $y \in M \Leftrightarrow$  für jede Folge  $(x_k)$  mit  $x_k \in M$ , die gegen  $y$  strebt, konvergiert  $f(x_k)$  gegen  $f(y)$ .

Ist  $f$  in jedem  $x \in M$  stetig, so heißt  $f$  stetig auf  $M$ .

Um zu zeigen, dass eine Funktion in einem Punkt nicht stetig ist, muss man zwei Folgen  $x_k$  und  $y_k$  finden, die den gleichen Grenzwert haben, aber für die

$$\lim_{k \rightarrow \infty} f(x_k) \neq \lim_{k \rightarrow \infty} f(y_k)$$

gilt. Umgekehrt muss man zum Nachweis der Stetigkeit zeigen, dass die Grenzwerte der Funktion für alle konvergenten Folgen identisch sind. Während es bei einer Funktion mit einer Variablen nur zwei Möglichkeiten gibt, sich einem Punkt anzunähern, gibt es im Mehrdimensionalen natürlich unendlich viele Möglichkeiten, was die Analyse in manchen Fällen erschwert.

**Beispiel 9.3.**

1. Seien  $f(x, y) = \frac{xy}{1+x^2+y^2}$  und  $\lim_{k \rightarrow \infty} x_k = a$  sowie  $\lim_{k \rightarrow \infty} y_k = b$ . Dann ist

$$\lim_{k \rightarrow \infty} f(x_k, y_k) = \lim_{k \rightarrow \infty} \frac{x_k y_k}{1+x_k^2+y_k^2} = \frac{ab}{1+a^2+b^2} = f(a, b)$$

für alle  $(a, b)^T \in \mathbb{R}^2$ . Folglich ist  $f(x, y)$  stetig auf  $\mathbb{R}^2$ .

2. Seien

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{für } (x, y)^T \neq (0, 0)^T \\ 0 & \text{sonst} \end{cases}$$

und gelte  $x_k \rightarrow a$  sowie  $y_k \rightarrow b$  für  $k \rightarrow \infty$ . Mit  $(a, b)^T \neq (0, 0)^T$  und  $(x_k, y_k)^T \neq (0, 0)^T$  für  $k \geq 0$  gilt  $f(x_k, y_k) \rightarrow f(a, b)$  für  $x \rightarrow \infty$  und damit Stetigkeit für  $\mathbb{R}^2 \setminus \{(0, 0)^T\}$ .

Seien  $x_k$  und  $y_k$  nun Nullfolgen. Dann gilt

$$\lim_{k \rightarrow \infty} f(x_k, y_k) = \lim_{k \rightarrow \infty} \frac{x_k y_k}{x_k^2 + y_k^2} = \lim_{k \rightarrow \infty} \frac{x_k^2}{2x_k^2} = \frac{1}{2} \neq f(0, 0) = 0.$$

↑  
setze  $x_k = y_k$

Also ist  $f(x, y)$  nicht stetig in  $(0, 0)^T$ . Dies ist nicht das einzige Gegenbeispiel: Wählt man etwa  $x_k = \frac{1}{k}$  und  $y_k = \frac{2}{k}$ , dann ist  $\lim_{k \rightarrow \infty} f(x_k, y_k) = \frac{2}{3}$ .

### 9.3 Partielle Ableitungen

Wir haben die Ableitung bei Funktionen mit einer Variablen durch die Approximation des Funktionswertes in der Umgebung eines Punktes motiviert. Es wird die Tangente in einem Punkt berechnet mit deren Hilfe man Funktionswerte in der Umgebung des Punktes approximieren kann (siehe Satz 5.2). Es stellt sich nun die Frage, wie man diese Idee auf Funktionen mit mehreren Variablen übertragen kann? Für Funktionen mit zwei Variablen lässt sich das Vorgehen gut vorstellen. Sei  $f(x_1, x_2)$  eine solche Funktion. Wenn wir den Funktionswert an einer Stelle  $(y_1, y_2)$  approximieren wollen und das Konzept der Approximation durch eine lineare Funktion übertragen, so müssen die Änderungen in beiden Variablen berücksichtigt werden. Somit ergibt sich für jede Variable eine Tangente, im zweidimensionalen Fall also eine Ebene. Der Funktionswert an der Stelle  $(y_1, y_2)$  wird dann durch

$$f(y_1, y_2) \approx f(x_1, x_2) + a_1(y_1 - x_1) + a_2(y_2 - x_2)$$

approximiert. Die Werte für  $a_1$  und  $a_2$  müssen natürlich entsprechend bestimmt werden. Wie dies geschehen muss zeigt die partielle Ableitung.

**Definition 9.10** (Partielle Ableitung). *Seien  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$  und  $e^{(k)}$  der  $k$ -te Einheitsvektor für  $k = 1, \dots, n$  im  $\mathbb{R}^n$  mit  $e_k^{(k)} = 1$  und  $e_i^{(k)} = 0$  für  $i \neq k$ . Der Grenzwert*

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{f(x + h \cdot e^{(i)}) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)}{h} \\ &=: \frac{\partial f(x)}{\partial x_i} =: f_{x_i}(x) =: \mathcal{D}_i f(x) \end{aligned}$$

heißt partielle Ableitung (1. Ordnung) von  $f$  nach  $x_i$  an der Stelle  $x \in M$ , sofern er existiert. Der aus den partiellen Ableitungen 1. Ordnung gebildete Vektor  $\nabla f(x) := \text{grad } f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T$  wird Gradient genannt. Sind diese partiellen Ableitungen stetig, dann ist  $f$  in  $x$  stetig partiell differenzierbar und wir schreiben  $f \in C^1$ .

Handwerkliches: Man erhält die partielle Ableitung von  $f$  nach  $x_i$ , indem man alle Variablen außer  $x_i$  als Konstanten auffasst und die dann nur noch von einer Variablen, nämlich  $x_i$ , abhängige Funktion in gewohnter Weise ableitet. Dadurch kann man die bereits bekannten Ableitungsregeln auch zur Bestimmung der partiellen Ableitungen nutzen.

#### Beispiel 9.4.

a) Sei  $f(x, y, z) = 2x^2 + 3xy + z$ . Die ersten partiellen Ableitungen lauten

$$\begin{aligned} f_x(x, y, z) &= \frac{\partial f(x, y, z)}{\partial x} = 4x + 3y \\ f_y(x, y, z) &= \frac{\partial f(x, y, z)}{\partial y} = 3x \\ f_z(x, y, z) &= \frac{\partial f(x, y, z)}{\partial z} = 1 \end{aligned}$$

mit Gradienten  $\nabla f(x, y, z) = (4x + 3y, 3x, 1)^T$

b) Sei  $f(x, y) = xy \cdot \sin(xy)$ .

Die Produktregel führt zu:

$$\begin{aligned} u &= xy & v &= \sin(xy) \\ u_x &= y & v_x &= y \cos(xy) \\ u_y &= x & v_y &= x \cos(xy) \\ f_x &= u_x \cdot v + v_x \cdot u = y \sin(xy) + xy^2 \cos(xy) \\ f_y &= u_y \cdot v + v_y \cdot u = x \sin(xy) + x^2 y \cos(xy) \\ \nabla f(x, y) &= (f_x, f_y)^T \end{aligned}$$

Es ist nun naheliegend, das Konzept der Ableitung, wie im eindimensionalen Fall, mehrfach anzuwenden und so zu Ableitungen höherer Ordnung zu gelangen. Im Prinzip ist dies auch möglich. Man muss sich allerdings vor Augen führen, dass man die partiellen Ableitungen nun aber mischen kann, so dass man erst nach  $x_i$  und anschließend nach  $x_j$  partiell ableitet. Dies führt zu  $n^2$  partiellen zweiten Ableitungen. Allgemein gibt es  $n^k$  partielle  $k$ -te Ableitungen. Wir werden uns deshalb auf die zweiten partiellen Ableitungen beschränken. Diese kann man auch noch in Form einer Matrix interpretieren.

**Definition 9.11.** Sei  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$ . Wenn die  $n$  partiellen Ableitungen 1. Ordnung existieren und stetig sind, dann werden für  $i, j = 1, \dots, n$  die Ausdrücke

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} := \frac{\partial f_{x_i}(x)}{\partial x_j} := f_{x_i x_j}(x)$$

partielle Ableitungen 2. Ordnung von  $f$  nach  $x_i$  und  $x_j$  an der Stelle  $x \in M$  genannt. Die aus den partiellen Ableitungen 2. Ordnung gebildete quadratische Matrix

$$\nabla^2 f(x) := H(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

heißt Hesse-Matrix. Sind die partiellen Ableitungen stetig, so schreiben wir  $f \in C^2$ .

Handwerkliches: Man erhält die partielle Ableitung 2. Ordnung von  $f$  nach  $x_i$  und  $x_j$ , wenn man zunächst partiell nach  $x_i$  ableitet und anschließend das Resultat partiell nach  $x_j$  ableitet.

**Beispiel 9.5.** Sei  $f(x, y) = 2xy + 3x + 4x^2y$ . Die ersten partiellen Ableitungen lauten

$$f_x = 2y + 3 + 8xy \quad \text{und} \quad f_y = 2x + 4x^2.$$

Erneute partielle Ableitungen jeweils nach  $x$  und  $y$  liefern die partiellen Ableitungen zweiter Ordnung

$$f_{xx} = 8y, \quad f_{xy} = 2 + 8x, \quad f_{yx} = 2 + 8x \quad \text{und} \quad f_{yy} = 0,$$

die sich in der Hessematrix  $\nabla^2 f(x, y) = \begin{pmatrix} 8y & 2 + 8x \\ 2 + 8x & 0 \end{pmatrix}$  zusammenfassen lassen.

**Satz 9.12** (Satz von Schwarz). *(Ohne Beweis)*

Ist  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$  in  $C^2$ , dann  $f_{x_i x_j} = f_{x_j x_i} \forall i, j = 1, \dots, n$ .

Aus dem Satz von Schwarz folgt, dass die Reihenfolge, in der die partiellen Ableitungen vorgenommen werden, unter den genannten Voraussetzungen irrelevant ist. In diesem Fall gilt, dass die Hesse-Matrix symmetrisch ist.

## 9.4 Minima und Maxima

Wir schon bei Funktionen mit einer Variablen ist auch im mehrdimensionalen Fall die Bestimmung von Minima und Maxima ein wichtiges Problem, bei dessen Lösung uns nun die partiellen Ableitungen helfen werden.

**Definition 9.13.** Sei  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$

- $x^* \in M$  heißt **globale Minimalstelle** von  $f$ , falls  $\forall x \in M. f(x^*) \leq f(x)$ .  
Der Wert  $f(x^*)$  wird dann **globales Minimum** genannt.
- $x^* \in M$  heißt **lokale Minimalstelle** von  $f$ , falls  $\exists \varepsilon > 0. \forall x \in \mathcal{U}_\varepsilon(x^*) \cap M. f(x^*) \leq f(x)$ .  
Der Wert  $f(x^*)$  wird dann **lokales Minimum** genannt.
- Entsprechend spricht man von **Maximalstellen und Maxima**, wenn die Ungleichungen umgekehrt werden.

Offensichtlich ist jede globale Minimal- bzw. Maximalstelle auch eine lokale Minimal- bzw. Maximalstelle. Die Umkehrung ist i.A. falsch.

**Satz 9.14** (Notwendiges Kriterium). Sei  $f : M \rightarrow \mathbb{R}$  für offenes  $M \subseteq \mathbb{R}^n$ . Ist  $x^* \in M$  lokale Extremalstelle von  $f$  und ist  $f$  in  $x^*$  partiell differenzierbar, so ist  $\nabla f(x^*) = (0, 0, \dots, 0)^T$ .

**Beweis:** Ist  $x^*$  lokale Extremalstelle von  $f$ , so auch für alle  $f_i(x_i) := f(x_1^*, \dots, x_{i-1}^*, x_i, x_{i+1}^*, \dots, x_n^*)$  mit  $x_i = x_i^*$ . Folglich gilt laut Satz 5.14 notwendig  $f'_i(x_i^*) = 0$ . Da  $f'_i(x_i^*) = \frac{\partial f(x^*)}{\partial x_i}$ , folgt die Behauptung.  $\square$

Handwerkliches: Nach Nullsetzen der  $n$  partiellen Ableitungen 1. Ordnung muss man im Allgemeinen ein nichtlineares Gleichungssystem lösen. Dies ist u.U. ein schwieriges mathematisches Problem, das nur mit Hilfe von numerischen Algorithmen approximativ gelöst werden kann. Alle Lösungen dieses Gleichungssystems sind kritische Punkte oder stationäre Lösungen oder Lösungskandidaten für die Extremalaufgabe. Ob Extrema an diesen Stellen vorliegen und wenn ja, ob Minima oder Maxima, bedarf weiterer Untersuchung. Der folgende Satz zeigt, wie uns die zweiten Ableitungen bei der Bewertung der kritischen Punkte helfen können.

**Satz 9.15.** *(Ohne Beweis)*

Sei  $f : M \rightarrow \mathbb{R}$  mit offenem  $M \subseteq \mathbb{R}^n$ . Wenn die partiellen Ableitungen zweiter Ordnung existieren und stetig sind und außerdem  $\nabla f(x^*) = (0, 0, \dots, 0)^T$  für ein  $x^* \in M$  gilt, dann ist  $x^*$  eine

- lokale Minimalstelle, wenn  $x^T \cdot \nabla^2 f(x^*) x > 0$  für alle  $x \in \mathbb{R}^n \setminus \{0\}$ <sup>1</sup>

<sup>1</sup>Die Hesse-Matrix ist in diesem Fall positiv definit.

b) lokale Maximalstelle, wenn  $x^T \cdot \nabla^2 f(x^*)x < 0$  für alle  $x \in \mathbb{R}^n \setminus \{0\}^2$

Ist  $x^T \nabla^2 f(x^*)x$  für mindestens ein  $x_1$  negativ und ein  $x_2$  positiv, so liegt kein Extremum vor.

Spezialfall  $n = 2$

$$\nabla^2 f(x) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

$$\Delta_1 = f_{xx}(x^*) \quad \Delta_2 = f_{xx}(x^*) \cdot f_{yy}(x^*) - f_{xy}^2(x^*)$$

$$x^T \nabla^2 f(x^*)x > 0 \Leftrightarrow \overbrace{f_{xx}(x^*)}^{\Delta_1} > 0 \text{ und } \overbrace{f_{xx}(x^*) \cdot f_{yy}(x^*) - f_{xy}^2(x^*)}^{\Delta_2} > 0$$

$$x^T \nabla^2 f(x^*)x < 0 \Leftrightarrow f_{xx}(x^*) < 0 \text{ und } f_{xx}(x^*) \cdot f_{yy}(x^*) - f_{xy}^2(x^*) > 0$$

Ist  $f_{xx}(x^*) \cdot f_{yy}(x^*) - f_{xy}^2(x^*) < 0 \Rightarrow$  kein Extremum in  $x^*$

Ist  $f_{xx}(x^*) \cdot f_{yy}(x^*) - f_{xy}^2(x^*) = 0 \Rightarrow$  keine Aussage möglich.

**Beispiel 9.6.**

$$a) \left. \begin{aligned} f(x, y) &= x^3 + y^3 - 3xy \\ f_x &= 3x^2 - 3y \stackrel{!}{=} 0 \Leftrightarrow x^2 \stackrel{!}{=} y \\ f_y &= 3y^2 - 3x \stackrel{!}{=} 0 \Leftrightarrow x \stackrel{!}{=} y^2 \end{aligned} \right\} \Rightarrow \text{kritische Punkte } (0,0)^T \text{ und } (1,1)^T$$

$$\nabla^2 f(x, y) = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$$

$$\nabla^2 f(0,0) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix} \text{ ist indefinit } (\Delta_2 < 0) \Rightarrow (0,0)^T \text{ keine Extremalstelle}$$

$$\nabla^2 f(1,1) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}$$

$$\left. \begin{aligned} \Delta_1 &= 6 > 0 \\ \Delta_2 &= 6 \cdot 6 - (-3)(-3) = 27 > 0 \end{aligned} \right\} \Rightarrow \text{lokales Minimum: } f(1,1) = -1$$

Globales Minimum?

$f(x, x) \rightarrow +\infty$  für  $x \rightarrow \infty$  und  $f(x, x) \rightarrow -\infty$  für  $x \rightarrow -\infty \Rightarrow f$  unbeschränkt  $\Rightarrow$  kein globales Minimum.

$$b) \left. \begin{aligned} f(x, y) &= x^2 + y^2 - 2xy + 1 \\ f_x &= 2x - 2y \stackrel{!}{=} 0 \Leftrightarrow x \stackrel{!}{=} y \\ f_y &= 2y - 2x \stackrel{!}{=} 0 \Leftrightarrow x \stackrel{!}{=} y \end{aligned} \right\} \Rightarrow \text{kritische Punkte } (x, y)^T \text{ mit } x = y$$

$$\nabla^2 f(x, y) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

$\Delta_1 > 0$  und  $\Delta_2 = 0 \rightarrow$  so keine Entscheidung möglich

Analyse durch Nachdenken und Abschätzung der Funktion

$$f(x, y) = x^2 + y^2 - 2xy + 1 = \underbrace{(x - y)^2}_{\geq 0} + 1 \geq 1 \text{ wird angenommen, wenn } x = y; \text{ also}$$

für alle kritischen Punkte  $\Rightarrow$  alle Elemente in  $\{(x, y)^T \in \mathbb{R}^2 \mid x = y\}$  sind lokale und globale Minimalstellen.

$$c) f(x, y) = \sqrt{x^2 + y^2}$$

Offensichtlich: globales Minimum in  $(0,0)^T$

<sup>2</sup>Entsprechend ist die Hesse-Matrix negativ definit.

Analyse durch partielle Ableitung

$$f_x = 2x \cdot \frac{1}{2\sqrt{x^2 y^2}} \quad f_y = 2y \cdot \frac{1}{2\sqrt{x^2 y^2}}$$

$$\left. \begin{array}{l} f_x(0, y) = 0 \text{ für } y \neq 0 \\ f_y(x, 0) = 0 \text{ für } x \neq 0 \end{array} \right\} \text{ nicht simultan erfüllbar}$$

$f_x(0, 0)$  und  $f_y(0, 0)$  nicht definiert  $\Rightarrow$   $\nexists$  partiellen Ableitungen in  $(0, 0)^T$

d) Gegeben seien  $N$  Punkte  $x^{(1)}, \dots, x^{(N)} \in \mathbb{R}^n$ . Für welches  $x \in \mathbb{R}^n$  wird die Summe der quadratischen Abstände zwischen  $x^{(i)}$  und  $x$  minimal?

Formal:  $\sum_{i=1}^N \|x^{(i)} - x\|^2 \rightarrow \min!$

$$f(x) = \sum_{i=1}^N \sum_{j=1}^n \underbrace{(x_j^{(i)} - x_j)}_{\text{bekannt Variable}}^2$$

alle Variablen mit Index  $j \neq k$  sind Konstanten!

$$\begin{aligned} \frac{\partial f(x)}{\partial x_k} &= \sum_{i=1}^N \sum_{j=1}^n \underbrace{\frac{\partial}{\partial x_k} (x_j^{(i)} - x_j)^2}_{=0 \text{ für } j \neq k} \\ &= \sum_{i=1}^N \frac{\partial}{\partial x_k} \left( \underbrace{x_k^{(i)}}_{\text{konstant}} - \underbrace{x_k}_{\text{Variable}} \right)^2 \\ &= \sum_{i=1}^N 2(x_k^{(i)} - x_k)(-1) \\ &= 2 \sum_{i=1}^N (x_k - x_k^{(i)}) \\ &= 2 \left( \sum_{i=1}^N x_k - \sum_{i=1}^N x_k^{(i)} \right) \\ &= 2Nx_k - 2 \sum_{i=1}^N x_k^{(i)} \quad \stackrel{!}{=} 0 \end{aligned}$$

für  $k = 1, \dots, n$

$$\Leftrightarrow Nx_k \stackrel{!}{=} \sum_{i=1}^N x_k^{(i)} \quad \Leftrightarrow \quad x_k \stackrel{!}{=} \frac{1}{N} \sum_{i=1}^N x_k^{(i)}$$

also:  $x^* = \frac{1}{N} \sum_{i=1}^N x^{(i)}$  ist Lösungskandidat

$$\frac{\partial^2 f(x)}{\partial x_k^2} = 2N \quad \text{und} \quad \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = 0 \text{ für } k \neq l \Rightarrow \nabla^2 f(x) = \begin{pmatrix} 2N & & 0 \\ & \ddots & \\ 0 & & 2N \end{pmatrix}$$

Offensichtlich ist  $x^T \nabla^2 f(x) x = 2N \sum_{j=1}^n x_j^2 = 2N \|x\|^2 > 0$  für  $x \neq 0$ . Lokales Minimum!

