Analysis II

Theresa Sprißler

18. Juli 2021

Inhaltsverzeichnis

1 Grundlagen

Definition 1.1. Sei X eine Menge. Eine Abbildung $d\colon X \times X \to [0, \infty)$ heißt **Metrik** auf X, wenn für alle $x, y, z \in X$ folgendes gilt:

- $d(x, y) = 0 \Leftrightarrow x = y$ (Definitheit)
- $d(x, y) = d(y, x)$ (Symmetrie)
- $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung)

(X, d) nennt man einen **metrischen Raum** und $d(x, y)$ den **Abstand** von zwei Punkten x und y.

Definition 1.2. Sei V ein Vektorraum über $\mathbb{R}$. Dann heißt $\|\cdot\|\colon V \to [0, \infty)$ **Norm**, wenn für alle $v \in V, \lambda \in \mathbb{R}$ gilt:

- $\|v\| = 0 \Leftrightarrow v = 0$ (Definitheit)
- $\|\lambda v\| = |\lambda| \cdot \|v\|$ (Homogenität)
- $\|v + w\| \leq \|v\| + \|w\|$ (Dreiecksungleichung)

Beispiel 1.3. p-Norm: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$.
Für $p = 1$ erhält man die Summennorm : $\|x\|_1 = \sum_{i=1}^n |x_i|$
Für $p = 2$ die euklidische Norm : $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
Maximumsnorm : $\|x\|_\infty = \max\{|x_1|, ..., |x_n|\}$

Definition 1.4. Sei V ein $\mathbb{R}$-Vektorraum. Dann heißt $\langle\cdot,\cdot\rangle\colon V \times V \to \mathbb{R}$ **Skalarprodukt**, wenn für alle $v, w, u \in V, \lambda, \mu \in \mathbb{R}$ gilt:

- $\langle \lambda v + \mu w, u\rangle = \lambda\langle v, w\rangle + \mu\langle w, u\rangle$
- $\langle v, w\rangle = \langle w, v\rangle$
- $\langle v, v\rangle \geq 0$ und $\langle v, v\rangle = 0 \Leftrightarrow v = 0$

Bemerkung 1.5. (a) $\|\cdot\|\colon V \to [0, \infty)$, $\|v\| := \sqrt{\langle v, v\rangle}$ heißt die **induzierte Norm.**
Z.B. induziert das kanonische Skalarprodukt $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$ die euklidische Norm auf dem $\mathbb{R}^n$ $\|x\| := \sqrt{\sum_{i=1}^n x_i^2}$.
(b) $d : V \times V \to [0, \infty)$ mit $d(v, w) := \|v - w\|$ heißt die von $\|\cdot\|$ induzierte Metrik auf V.

Bemerkung 1.6 (Ungleichung von **Cauchy-Schwarz**)**.**
Ist $\|\cdot\|$ die von $\langle\cdot,\cdot\rangle$ induzierte Norm, dann gilt: $|\langle v, w\rangle| \leq \|v\| + \|w\|$

Definition 1.7. Zwei Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ heißen **äquivalent** :$\Leftrightarrow$ Es existieren $c, C > 0$, sodass für alle $v \in V$ gilt: $c\|v\|_1 \leq \|v\|_2 \leq C\|v\|_1$.

2 Folgen und Konvergenz

Definition 2.1. Eine Folge $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^n$ **konvergiert** gegen $a \in \mathbb{R}^n$ $:\Leftrightarrow$
$\forall \varepsilon > 0 \, \exists \, K \in \mathbb{N} : \forall k \geq K : \|x_k - a\| < \varepsilon$.
Man schreibt dann:

$$\lim_{n\to\infty} x_n = a$$

Bemerkung 2.2. Konvergenz in $\mathbb{R}^n$ ist äquivalent zu komponentenweiser Konvergenz, d.h.

$$\lim_{k\to\infty} x_k = a \Leftrightarrow \lim_{k\to\infty} x_{k,j} = a_j$$

für $j = 1, ..., n$

Definition 2.3. Eine Folgen (x_n) in $\mathbb{R}^n$ heißt **Cauchy-Folge** $:\Leftrightarrow$ Für alle $\varepsilon > 0$ existiert ein $n_0 \in \mathbb{N}$, sodass für alle $n, m \geq n_0$ gilt: $d(x_m, x_n) < \varepsilon$.
(mit Quantoren: $\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N} \; \forall n, m \geq n_0 \colon d(x_m, x_n) < \varepsilon$)
Äquivalente Definition: $\forall \epsilon > 0 \, \exists K \geq 1 \, \forall k \geq K, \forall l \geq 1 : \|x_k - x_{k+l}\| < \epsilon$.

Satz 2.4. Jede konvergente Folge ist eine Cauchy-Folge.
In $\mathbb{R}^n$ konvergiert eine Folge genau dann, wenn sie eine Cauchy-Folge ist.

Bemerkung 2.5. Sind zwei Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ äquivalent, dann konvergiert eine Folge bzgl. $\|\cdot\|_1$ genau dann, wenn sie bzgl. $\|\cdot\|_2$ konvergiert. Auf $\mathbb{R}^n$ sind alle Normen äquivalent.

Definition 2.6. Eine Folge (x_k) in $\mathbb{R}^n$ heißt **beschränkt**, falls ein $M > 0$ existiert, sodass für alle $k \geq 1$ gilt: $\|x_k\| \leq M$.

Satz 2.7. (**Bolzano-Weierstraß in $\mathbb{R}^n$**)
Jede beschränkte Folge in $\mathbb{R}^n$ hat eine konvergente Teilfolge.

3 Umgebungen, offene und abgeschlossene Mengen

Definition 3.1. Sei $a \in \mathbb{R}^n, r \in \mathbb{R}_+$. Dann heißt $B_r(a) := \{x \in \mathbb{R}^n : \|x - a\| < r\}$ die **offene Kugel** um a vom Radius r bzgl. $\|\cdot\|$.

Definition 3.2. Eine Menge $U \subseteq \mathbb{R}^n$ heißt **Umgebung** von $a \in \mathbb{R}^n$, falls ein $\delta > 0$ existiert, sodass $B_\delta(a) \subseteq U$.

Definition 3.3. Eine Menge $U \subseteq \mathbb{R}^n$ heißt **offen**, falls sie Umgebung aller ihrer Punkte ist.

Definition 3.4. Eine Menge $A \subseteq \mathbb{R}^n$ heißt **abgeschlossen**, falls für für jede konvergente Folge in A auch der Grenzwert in A liegt.

Proposition 3.5. A ist genau dann abgeschlossen, wenn A^C offen ist.

Bemerkung 3.6.
1. Endliche Durchschnitte und beliebige Vereinigungen von offenen Mengen sind selbst offen.
2. Endliche Vereinigungen und beliebige Durchschnitte abgeschlossener Mengen sind selbst abgeschlossen.
3. Es gibt Teilmengen, die weder offen noch abgeschlossen sind und Teilmengen, die offen und abgeschlossen sind (z.B. $\varnothing$ und $\mathbb{R}^n$).

4 Kompaktheit

Definition 4.1. Eine Menge $K \subseteq \mathbb{R}^n$ heißt **kompakt**, falls jede Folge in K eine Teilfolge hat, die gegen einen Punkt von K konvergiert.

Theorem 4.2 (Heine-Borel). Eine Teilmenge $K \subseteq \mathbb{R}^n$ ist kompakt $\Leftrightarrow$ K ist beschränkt und abgeschlossen.

Definition 4.3. Eine Menge K ist beschränkt, wenn ein $B > 0$ existiert, sodass für alle $x \in K$ gilt: $\|x\| \leq B$.

Definition 4.4. Eine Familie $(U_i)_{i \in I}$ von Teilmengen U_i heißt eine **offene Überdeckung von** Y $:\Leftrightarrow$
(a) U_i ist offen für alle $i \in I$,
(b) $Y \subseteq \bigcup_{i \in I} U_i$

Satz 4.5. Eine offene Teilmenge $K \subseteq X$ ist **kompakt** $\Leftrightarrow$ Jede offene Überdeckung (U_i) von K enthält eine endliche Teilüberdeckung, d.h. $\exists\, i_1, i_2, ...i_n \in I$:

$$K \subseteq U_{i_1} \cup ... \cup U_{i_n}.$$

Beachte: Das ist nicht gleichbedeutend mit der Forderung, dass eine endliche offene Überdeckung existiert.

Beispiel 4.6.
(a) Sei $K = \{x_1, ..., x_n\}$ endlich. Dann ist K kompakt, da man aus einer offenen Überdeckung (U_i) diejenigen $i_k \in I$ mit $x_k \in U_{i_k}$ wählen kann und dann $K \subseteq U_{i_1} \cup ... \cup U_{i_n}$ gilt.
(b) Sein (x_n) konvergente Folge in X mit dem Grenzwert a. Dann ist $K := \{x_n \in X : n \in \mathbb{N}\} \cup \{a\}$ kompakt.
(c) Das offene Intervall $(0, 1) \subseteq \mathbb{R}$ ist nicht kompakt, da die offene Überdeckung $\bigcup_{n=2}^{\infty} (\frac{1}{n}, 1)$ keine endliche Teilüberdeckung enthält.

5 Stetigkeit

Definition 5.1. Sei $a \in \mathbb{R}^n, X \subseteq \mathbb{R}^n$. Dann heißt $f\colon X \to \mathbb{R}^m$ **stetig** in a $:\Leftrightarrow$ Für jede Folge (x_n) in X mit $(x_n) \to a$ gilt: $(f(x_n)) \to f(a)$ (Folgenkriterium).
Äquivalent dazu sind folgende Definitionen:
(a) $\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in X$ mit $\|x - a\| < \delta\colon \|f(x) - f(a)\| < \varepsilon$.
(b) $\forall \varepsilon > 0\ \exists \delta > 0\colon f(B_\delta(a)) \subseteq B_\varepsilon(f(a))$.
Man schreibt auch (wie in Analysis 1): f ist stetig, wenn $\lim_{x \to a} f(x) = f(a)$ gilt.

Proposition 5.2. (a) Seien $f\colon X \to Y \subseteq \mathbb{R}^m, g\colon X \to \mathbb{R}^m$, $h\colon Y \to Z \subseteq \mathbb{R}^l$ und $j\colon X \to \mathbb{R}$ stetig.
Dann sind die Funktionen $f + g$ (Addition stetiger Funktionen), $h \circ f$ (Komposition) und $f \cdot j$ ebenfalls stetig.
(b) $f\colon X \to \mathbb{R}^m$ stetig in a $\Leftrightarrow$ $f_j\colon X \to \mathbb{R}$ stetig in a für $j = 1, .., m$.
(c) In einem normierten Raum ist die Norm $\|\cdot\|$ stetig.

Beispiel 5.3. $f : \mathbb{R}^2 \to \mathbb{R}$,

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

ist in $(0,0)$ unstetig. Beweis:
Die Folge $(\frac{1}{n}, \frac{1}{n})_{n\in\mathbb{N}}$ konvergiert für $n \to \infty$ gegen 0, aber es gilt:

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

Also konvergiert $(f(\frac{1}{n}, \frac{1}{n}))_{n\in\mathbb{N}}$ nicht gegen $f(0,0) = 0$.

Satz 5.4. (Hausdorff)
$f\colon X \to Y$ ist stetig.
$\Leftrightarrow$ Das Urbild jeder offenen Menge in Y ist offen in X, d.h. $V \subseteq Y$ offen $\Rightarrow f^{-1}(V) \subseteq X$ offen.
$\Leftrightarrow$ Das Urbild $f^{-1}(A)$ jeder abgeschlossenen Menge $A \subseteq Y$ ist abgeschlossen.

Definition 5.5. Zwei Mengen $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$ heißen **homöomorph** , wenn es eine bijektive Abbildung $f\colon X \to Y$ gibt, sodass f und f^{-1} stetig sind. Die Abbildung f heißt dann **Homöomorphismus** .

Satz 5.6. Sei $K \subseteq \mathbb{R}^n$ kompakt, $f\colon K \to \mathbb{R}^m$ stetig. Dann ist $f(K)$ ebenfalls kompakt.

Theorem 5.7 (Satz von Minimum und Maximum, Weierstraß). Sei $K \subseteq \mathbb{R}^n$ kompakt, $f\colon K \to \mathbb{R}$ stetig. Dann nimmt f ihr Maximum und Minimum an, d.h. es gibt $x_1, x_2 \in K$, sodass für alle $x \in K$ gilt: $f(x_1) \leq f(x) \leq f(x_2)$.

Definition 5.8. Sei $X \subseteq \mathbb{R}^n$. Eine Funktion $f\colon X \to \mathbb{R}^m$ heißt **gleichmäßig stetig**, falls $\forall \varepsilon > 0\, \exists \delta > 0\, \forall x, y \in X$ mit $\|x - y\| < \delta :\ \|f(x) - f(y)\| < \varepsilon$.

Satz 5.9. Sei $K \subseteq \mathbb{R}^n$ kompakt und $f\colon K \to \mathbb{R}^m$ stetig. Dann ist f gleichmäßig stetig.

Definition 5.10. $X \subseteq \mathbb{R}^n$ heißt **wegzusammenhängend**, wenn für alle $a, b \in X$ ein **Weg**, d.h. eine stetige Abbildung $\alpha\colon [0,1] \to X$ mit $\alpha(0) = a$, $\alpha(1) = b$ existiert.

Proposition 5.11. Sei $x \subseteq \mathbb{R}^n$ wegzusammenhängend und $f\colon X \to \mathbb{R}^m$ stetig. Dann ist auch $f(X)$ wegzusammenhängend.

Satz 5.12. (**Zwischenwertsatz**)
Sei $X \subseteq \mathbb{R}^n$ wegzusammenhängend, $f\colon X \to \mathbb{R}$ stetig und $x_0, x_1 \in X$. Dann existiert für alle $c \in \mathbb{R}$ zwischen $f(x_0)$ und $f(x_1)$ ein $\xi \in X$ mit $f(\xi) = c$.

5.1 Gleichmäßige Konvergenz

Definition 5.13. Seien $f_n, f\colon X \to Y$ $(n \in N)$. Dann **konvergiert** f_n **gleichmäßig** gegen f :$\Leftrightarrow$ Für alle $\varepsilon > 0$ existiert ein $n_0 \in \mathbb{N}$, sodass für alle $x \in X$ und $n \geq n_0$ gilt: $\|f_n(x) - f(x)\| < \varepsilon$. (Beachte: n_0 hängt nur von ε, aber nicht von x ab.)
Zum Vergleich: (f_n) **konvergiert punktweise** gegen f :$\Leftrightarrow$ Für alle $x \in X$ und für alle $\varepsilon > 0$ existiert ein $n_0 \in \mathbb{N}$, sodass für alle $n \geq n_0$ gilt: $\|f_n(x) - f(x)\| < \varepsilon$ (n_0 hängt auch von x ab).

Satz 5.14. Seien $f_n\colon X \to Y$ stetig und gleichmäßig konvergent gegen $f\colon X \to Y$. Dann ist auch f stetig.

6 Differenzierbarkeit

Im Folgenden sei $G \subseteq \mathbb{R}^n$ offen.

Definition 6.1. Sei $x_0 \in G$ und $f\colon G \to \mathbb{R}^m$. f heißt in x_0 **differenzierbar**, wenn es eine lineare Abbildung $f'(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ und eine in x_0 stetige Funktion $r\colon G \to \mathbb{R}^m$ mit $r(x_0) = 0$ gibt, sodass für alle $x \in G$ gilt:

$$f(x) = f(x_0) + f'(x_0(x - x_0) + r(x)\|x - x_0\|$$

Die lineare Abbildung $f'(x_0)$ heißt **Ableitung** von f in x_0 und wird mit $Df(x_0)$ bezeichnet.

Proposition 6.2. $f\colon G \to \mathbb{R}^m$ ist genau dann differenzierbar in x_0, wenn $f_j\colon G \to \mathbb{R}$ $(j = 1, ..., m)$ differenzierbar sind.

Satz 6.3. Falls f in x_0 differenzierbar ist, gilt für alle $v \in \mathbb{R}^n$

$$f'(x_0)v = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

Der Grenzwert wird mit $\partial_v f(x_0)$ bezeichnet und heißt **Richtungsableitung** von f in Richtung v.

Definition 6.4. Sei $a \in \mathbb{R}^n$, $f\colon \mathbb{R}^n \to \mathbb{R}^m$. Der Grenzwert

$$\lim_{h \to 0} \frac{f(a + h \cdot e_j) - f(a)}{h} = \lim_{h \to 0} \frac{f(a_1, ..., a_j + h, ..., a_n) - f(a_1, ..., a_n)}{h}$$

(wobei e_j der j-te Einheitsvektor ist) heißt **partielle Ableitung** von f nach x_j an der stelle a und wird mit $\dfrac{\partial f}{\partial x_j}(a)$ oder $\partial_j f(a)$ bezeichnet.

Bemerkung 6.5. Man berechnet die j-te partielle Ableitung $\partial_j f(x)$, indem man nur nach der Variable x_j differenziert und alle anderen x_i mit $i \neq j$ wie Konstanten behandelt.

Beispiel 6.6. Die Radiusfunktion

$$r\colon \mathbb{R}^n \to \mathbb{R}, r(x) = \|x\|_2 = \sqrt{x_1^2 + ... + x_n^2}$$

ist für alle $x \in \mathbb{R}^n \backslash \{0\}$ und für alle $j \in \{1, ..., n\}$ in die j-te Richtung partiell differenzierbar mit (verwende Kettenregel)

$$\frac{\partial f}{\partial x_j}(x) = \frac{1}{2} \frac{2x_j}{\sqrt{x_1^2 + ... + x_n^2}} = \frac{x_j}{\|x\|_2}$$

Satz 6.7. Die **Jacobi-Matrix** (oder **Funktionalmatrix**) von $f = (f_1, ..., f_m)$, $f\colon \mathbb{R}^n \to \mathbb{R}^m$ in x_0, $Df(x_0)$, ist die Matrix, durch die die lineare Abbildung $f'(x_0)$ beschrieben wird. Für ihre Einträge a_{ij} gilt $(i = 1, ..., m; j = 1, ..., n)$ $a_{ij} = \dfrac{\partial f_i}{\partial x_j}(x_0)$, also:

$$Df(x_0) = \begin{pmatrix} \partial_1 f_1(x_0) & \dots & \partial_n f_1(x_0) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x_0) & \dots & \partial_n f_m(x_0) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Satz 6.8. (**Kriterium für Differenzierbarkeit**) Existieren in einer Umgebung von x_0 alle partiallen Ableitungen $\partial_1 f, ..., \partial_n f$ und sind in x_0 stetig, dann ist f in x_0 differenzierbar.

Bemerkung 6.9. Zusammenfassung:

$$\begin{array}{ccc} & \text{stetig partiell differenzierbar} & \\ & \Downarrow \not\Uparrow & \\ & \text{differenzierbar} & \\ \Downarrow \not\Uparrow & & \Downarrow \not\Uparrow \\ \text{partiell diff.bar} & \not\Leftrightarrow & \text{stetig} \end{array}$$

Beispiel 6.10. (Aus partieller Differenzierbarkeit folgt nicht Stetigkeit.)
$f : \mathbb{R}^2 \to \mathbb{R}$,

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

ist auf ganz $\mathbb{R}^2$ partiell differenzierbar (aber in $(0,0)$ nicht stetig, siehe Bsp. 3.3). Beweis: Auf $\mathbb{R}^2 \backslash \{0\}$ ist f eine Zusammensetzung von partiell differenzierbaren Funktionen und daher selbst partiell differenzierbar. Betrachte den Fall $(x,y) = (0,0)$:

$$\frac{1}{h}(f(0+h,0) - f(0,0)) = \frac{1}{h}\left(\frac{0+h\cdot 0}{h^2+0^2} - 0\right) = \frac{1}{h}\cdot 0 \xrightarrow{h\to 0} 0$$
$$\frac{1}{h}(f(0,0+h) - f(0,0)) = \frac{1}{h}\left(\frac{0\cdot h}{0^2+h^2} - 0\right) = \frac{1}{h}\cdot 0 \xrightarrow{h\to 0} 0$$

Also ist f auch in $(0,0)$ in x- und y-Richtung partiell differenzierbar.

Satz 6.11. (**Kettenregel**)
Sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$, $g\colon \mathbb{R}^m \to \mathbb{R}^l$, $x_0 \in \mathbb{R}^n$, $y_0 := f(x_0)$, f differenzierbar in x_0 und g differenzierbar in y_0. Dann ist $g \circ f$ differenzierbar in x_0 und es gilt

$$(g\circ f)'(x_0) = g'(y_0)f'(x_0) \quad \text{bzw.} \quad D(g\circ f)(x_0) = Dg(y_0)Df(x_0).$$

Definition 6.12. Für $f\colon \mathbb{R}^n \to \mathbb{R}$ heißt $\nabla f(x) = \operatorname{grad} f(x) := (\partial_1 f(x), \partial_2 f(x), ..., \partial_n f(x))^T$ der **Gradient** von f in x.

Theorem 6.13. (**Mittelwertsatz**)
Sei $f\colon G \to \mathbb{R}$ differenzierbar, $a, b \in G$, deren Verbindungsstrecke in G liegt (d.h. $[a,b] := \{a + t(b-a) | t \in [0,1]\} \subseteq G$. Dann existiert ein $\xi \in [a,b]$, sodass gilt:

$$f(b) - f(a) = f'(\xi)(b-a)$$

Satz 6.14. (**Schrankensatz**)
Sei $f\colon G \to \mathbb{R}^m$ stetig differenzierbar, $a, b \in G$, sodass $[a,b] \subseteq G$. Ist

$$\|f'(x)(b-a)\| \leq M\|b-a\| \forall x \in [a,b]$$

dann folgt

$$\|f(b) - f(a)\| \leq M\|b-a\|.$$

6.1 Höhere partielle Ableitungen

Satz 6.15. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}$ so, dass in einer Umgebung von (x_0, y_0) die partiellen Ableitungen $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial}{\partial y}\frac{\partial f}{\partial x}$ existieren und stetig sind. Dann existiert auch $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}$ und es gilt:

$$\frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x_0, y_0)$$

Definition 6.16. $f\colon G \to \mathbb{R}^m$ heißt **k-mal stetig differenzierbar**, falls in jedem $x \in G$ alle partiellen Ableitungen bis zur k-ten Ordnung existieren und stetig sind.

Definition 6.17.

$$\mathrm{H}(x) = \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix} \in \mathrm{Mat_n}(\mathbb{R})$$

ist die **Hesse-Matrix von f in** x .
Wenn f zweimal stetig differenzierbar ist, ist die Hesse-Matrix symmetrisch.

7 Wichtige Sätze

7.1 Banachscher Fixpunktsatz

Definition 7.1. $f\colon A \to A$ heißt **Kontraktion** $:\Leftrightarrow \exists\, \vartheta \in (0,1)\colon \forall x, y \in A\colon$
$\|f(x) - f(y)\| \leq \vartheta \cdot \|x - y\|$.

Theorem 7.2 (Banachscher Fixpunktsatz).
Sei $A \subseteq \mathbb{R}^n$ abgeschlossen, $f\colon A \to A$ eine Kontraktion. Dann gilt: f hat genau einen Fixpunkt, d.h. ein $a \in A$ mit $f(a) = a$.

7.2 Satz von der lokalen Umkehrbarkeit

Definition 7.3. Seien $G, D \subseteq \mathbb{R}^n$ Gebiete. Eine stetig differenzierbare Abbildung $f\colon G \to D$ heißt **Diffeomorphismus**, wenn gilt:
(a) f ist bijektiv
(b) $f^{-1}\colon D \to G$ ist stetig differenzierbar.

Theorem 7.4 (Satz von der lokalen Umkehrbarkeit).
Sei $G \subseteq \mathbb{R}^n$ offen, $x_0 \in G$, $f\colon G \to \mathbb{R}^n$ stetig differenzierbar und $\mathrm{D}\, f(x_0)$ invertierbar. Dann existieren offene Umgebungen $U \subseteq G$ von x_0 und $V \subseteq \mathbb{R}^n$ von $y_0 := f(x_0)$, sodass $f\colon U \to V$ ein Diffeomorphismus ist.
Für $g := f^{-1}\colon V \to U$ gilt dann: $Dg(y_0) = Df(x_0)^{-1}$.

7.3 Satz über implizite Funktionen

Motivation: Eine Funktion g ist nicht immer explizit durch $y = g(x)$ gegeben, sondern häufig nur implizit durch eine Gleichung der Form $F(x, g(x)) = 0$, wobei F gegeben ist. Die implizite Gleichung $F(x, y) = 0$ möchte man dann „nach y auflösen“.

Beispiel 7.5. Betrachte $F\colon \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto x^2+y^2-1$. $C = \{(x,y) \in \mathbb{R}^2 : F(x,y) = 0\}$ ist die Einheitskreislinie, aber nicht der Graph einer Funktion $g\colon \mathbb{R} \to \mathbb{R}$, denn für $x \in \mathbb{R}$ mit $x > 1$ existiert kein $y \in \mathbb{R}$ mit $F(x,y) = 0$ und für $x \in \mathbb{R}$ mit $|x| < 1$ existieren zwei $y \in \mathbb{R}$ mit $F(x,y) = 0$, nämlich $y = \pm\sqrt{1-x^2}$.
Ist allerdings $(a,b) \in C$ mit $b \neq 0$, so kann man $F(x,y)$ lokal um (a,b) nach y oder x auflösen, d.h. es existiert eine Umgebung $U_1 \subseteq \mathbb{R}^2$ von a und eine Umgebung $U_2 \subseteq \mathbb{R}^2$ von b und eine Funktion $g\colon U_1 \to U_2$, sodass für alle $(x,y) \in U_1 \times U_2$ gilt:

$$F(x,y) = 0 \Leftrightarrow g(x) = y$$

Theorem 7.6 (Satz über implizite Funktionen). Sei $W \subseteq \mathbb{R}^n \times \mathbb{R}^m$ offen und $F\colon W \to \mathbb{R}^m$ stetig differenzierbar, sowie $(a,b) \in W$ mit $F(a,b) = 0$. Außerdem sei

$$\mathrm{D}_y F(a,b) = \frac{\partial F}{\partial y}(a,b) = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}(a,b)$$

invertierbar. Dann gibt es offene Umgebungen $U_1 \subseteq \mathbb{R}^n$ von a und $U_2 \subseteq \mathbb{R}^m$ von b mit $U_1 \times U_2 \subseteq W$ und eine stetig differenzierbare Funktion $g\colon U_1 \to U_2$, sodass für alle $(x,y) \in U_1 \times U_2$ gilt:

$$F(x,y) = 0 \;\Leftrightarrow\; y = g(x)$$

Die Ableitung von g ist

$$\mathrm{D}\, g(a) = -\underbrace{(\frac{\partial F}{\partial y}(a,b))^{-1}}_{m\times m}\underbrace{\frac{\partial F}{\partial x}(a,b)}_{m\times n}.$$

Bemerkung 7.7. Der Satz besagt, dass es unter den genannten Voraussetzungen eine stetig differenzierbare Funktion g mit $g(x) = y$ gibt, sodass die implizit gegebene Gleichung $F(x,y) = 0$ für alle (x,y) in einer Umgebung von (a,b) erfüllt ist.

Beispiel 7.8. Sei $f\colon G \subseteq \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Eine **Niveaufläche** von f ist gegeben durch $N_c := \{x \in G : f(x) = c\}$ („Höhenlinien“).
Wenn $a = (a', a_n) \in N_c$, $a' = (a_1, ..., a_{n-1})$ und $\frac{\partial f}{\partial x_n}(a', a_n) \neq 0$ ist, dann gibt es Umgebungen $U' \subseteq \mathbb{R}^{n-1}$ von a' und $I \subseteq \mathbb{R}$ von a_n mit $U' \times I \subseteq G$ und eine $\mathcal{C}^1$-Funktion $g\colon U' \to I$, sodass für alle $x \in U' \times I$ gilt:

$$f(x', x_n) - c = 0 \;\Leftrightarrow g(x') = x_n,$$

also $N_c = \mathrm{Graph}(g)$.

8 Lokale Extrema

Definition 8.1. Sei $f\colon G \to \mathbb{R}$. $x_0 \in G$ heißt
(a) **lokales Minimum** von f, wenn es ein $\delta > 0$ gibt mit $f(x) \geq f(x_0)$ für alle $x \in B_\delta(x_0) \cap G$.
(b) **lokales Maximum** von f, wenn es ein $\delta > 0$ gibt mit $f(x) \leq f(x_0)$ für alle $x \in B_\delta(x_0) \cap G$.
(c) Gilt sogar $f(x) > f(x_0)$ bzw. $f(x) < f(x_0)$, dann heißt das lokale Minimum bzw. Maximum **strikt**.

Satz 8.2 (Notwendige Bedingung für lokale Extrema). Sei $f\colon G \to \mathbb{R}$ stetig differenzierbar. Wenn $x_0 \in G$ ein lokales Extremum von f ist, dann gilt $\nabla f(x_0) = 0$.
Ähnlich wie im eindimensionalen Fall ist dieses Kriterium nur notwendig, aber nicht hinreichend.

Definition 8.3. Sei V ein $\mathbb{R}$-Vektorraum. $q\colon V \to V$ ist eine **quadratische Form**, wenn eine symmetrische Bilinearform $s\colon V \times V \to \mathbb{R}$ existiert, sodass $q(v) = s(v,v)$.
Beispiel: $A \in \mathrm{Mat_n}(\mathbb{R})$ symmetrisch, d.h $A^t = A$. Dann ist $s_a\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, s_A(x,y) = x^t A y = \langle x, Ay\rangle$ eine symmetrische Bilinearform.

Definition 8.4. Eine quadratische Form $q\colon V \to V$ heißt

- **positiv definit**, falls $q(v) > 0$ für alle $v \in V \backslash \{0\}$
- **negativ definit**, falls $q(v) < 0$ für alle $v \in V \backslash \{0\}$
- **indefinit**, falls $v, w \in V$ existieren mit $q(v) < 0$ und $q(w) > 0$.
- **positiv semidefinit**, falls $q(v) \geq 0$ für alle $v \in V \backslash \{0\}$
- **negativ semidefinit**, falls $q(v) \leq 0$ für alle $v \in V \backslash \{0\}$

Eine symmetrische Bilinearform heißt positiv definit, wenn ihre zugehörige quadratische Form positiv definit ist.
Eine symmetrische Matrix $A \in \mathrm{Sym_n}(\mathbb{R})$ heißt **positiv definit** (negativ definit und indefinit analog), falls $\langle x, Ax\rangle > 0$ für alle $x \in \mathbb{R}\backslash\{0\}$.

Bemerkung 8.5. $A \in \mathrm{Sym_n}(\mathbb{R})$ ist genau dann positiv bzw. negativ definit, wenn alle Eigenwerte > 0 bzw. < 0 sind. Sie ist genau dann positiv bwz. negativ semidefinit, wenn alle Eigenwerte ≥ 0 bzw. ≤ 0 sind und genau dann indefinit, wenn es einen Eigenwert > 0 und einen Eigenwert < 0 gibt.

Satz 8.6 (Hinreichendes Kriterium für lokale Extrema). Sei $f\colon G \to \mathbb{R}$ zweimal stetig differenzierbar, $x_0 \in G$.
Wenn $\nabla f(x_0) = 0$ und ...

... $\nabla^2 f(x_0)$ positiv definit ist, dann besitzt f in x_0 ein (striktes) **lokales Minimum**.

... $\nabla^2 f(x_0)$ negativ definit ist, dann besitzt f in x_0 ein (striktes) **lokales Maximum**.

... $\nabla^2 f(x_0)$ indefinit ist, dann besitzt f in x_0 **kein lokales Extremum**.

Beispiel 8.7. $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2+y^2$. Der Gradient $\nabla f(x,y) = (2x, 2y)$ verschwindet in $(0,0)$ und die Hesse-Matrix

$$\nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

ist positiv definit. Also hat f in $(0,0)$ ein striktes lokales Minimum.
$g\colon \mathbb{R}^2 \to \mathbb{R}$, $g(x,y) = x^2 - y^2$. Der Gradient $\nabla(g)(x,y) = (2x, -2y)$ verschwindet zwar in $(0,0)$, aber

$$\nabla^2 g = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

ist indefinit, also besitzt g in $(0,0)$ kein lokales Extremum.

Bemerkung 8.8. Ist $\nabla^2 f(x)$ nur positiv semidefinit, kann man i.A. keine Aussage über ein lokales Extremum in x treffen.

8.1 Extrema unter Nebenbedingungen

Definition 8.9. Seien $f, h\colon G \to \mathbb{R}$ stetig differenzierbar. $a \in G$ heißt ein **lokales Maximum bzw. Minimum von f unter der Nebenbedingung** $h = 0$, wenn $h(a) = 0$ und es eine offene Umgebung $U \subseteq G$ von a gibt, sodass für alle $x \in U$ mit $h(x) = 0$ gilt: $f(x) \le f(a)$ bzw. $f(x) \ge f(a)$.

Theorem 8.10 (Satz von Lagrange; notwendige Bedingung).
Seien $f\colon G \to \mathbb{R}, h\colon G \to \mathbb{R}^m$ stetig differenzierbar und $\nabla h_1(a), ..., \nabla h_m(a)$ linear unabhängig. Sei $a \in G$ ein lokales Extremum von f unter der Nebenbedingung $h(x) = 0$. Dann gibt es $\lambda_1, ..., \lambda_m \in \mathbb{R}$ (**„Lagrange-Multiplikatoren“**), sodass gilt:

$$\nabla f(a) - \sum_{i=1}^{m} \lambda_i \nabla h_i(a) = 0$$

Bemerkung 8.11. Der Satz besagt, dass für ein lokales Extremum unter einer Nebenbedingung der Gradient senkrecht auf den Tangentialvektoren von $M = \{x \in G : h(x) = 0\}$ stehen muss.

Beispiel 8.12. Wir betrachten auf dem Kompaktum $K = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ (Kreisscheibe um den Ursprung mit Radius 1) die stetige Funktion $f\colon K \to \mathbb{R}$, $f(x,y) = y^2 - x^2$. Nach dem Satz von Weierstraß nimmt f ihr Maximum an. Wo liegen die globalen Maxima von f?
Angenommen, $(x_0, y_0) \in \mathring{K} = \{(x,y) \in K : x^2 + y^2 < 1\}$ wäre ein Maximum von f. Dann muss gelten: $\operatorname{grad}(f)(x_0, y_0) = (-2x_0, 2y_0) = 0$. Also ist $x_0 = y_0 = 0$. Aber:

$$\mathrm{Hess(f)}(0,0) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$

ist indefinit. Das globale Maximum kann also nicht in $\{(x,y) \in K : x^2 + y^2 < 1\}$ liegen. Es muss also auf dem Rand $M := \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ liegen. Für $(x,y) \in M$ verschwindet der Gradient nicht, wir suchen also ein Maximum unter einer Nebenbedingung.

Setze dazu $h\colon \mathbb{R}^2 \to \mathbb{R},\ h(x,y) = x^2 + y^2 - 1$. Wenn f in (x_0, y_0) ein Extremum unter der Nebenbedingung $h = 0$ hat, muss es ein $\lambda \in \mathbb{R}$ geben mit $\operatorname{grad}(f)(x_0, y_0) = \lambda \cdot \operatorname{grad}(h)(x_0, y_0)$. Wir müssen also das folgende LGS lösen:

$$-2x = \lambda 2x \tag{1}$$
$$2y = \lambda 2y \tag{2}$$
$$x^2 + y^2 = 1 \tag{3}$$

Fall 1: $x \neq 0$. Dann muss nach (1) $\lambda = -1$ sein. Aus (2) erhält man damit $2y = -2y$ und damit ist $y = 0$. Mit (3) folgt dann $x = \pm 1$.
Fall 2: $x = 0$. Aus (3) folgt damit $y = \pm 1$.
Betrachte nun die Funktionswerte an den Punkten $(\pm 1, 0)$ und $(0, \pm 1)$:
$f(\pm 1, 0) = -1$ und $f(0, \pm 1) = 1$. Das Maximum von f wird also in $(0, \pm 1)$ angenommen.

Wortverzeichnis