Höhere Mathematik II

Mitschrift der Vorlesung von Prof. Lehn
im Sommersemester 2019 an der Uni Ulm

26. Mai 2019

Inhaltsverzeichnis

Einführung

0.1 Stetigkeit in einer Dimension

$$
\begin{aligned}
&f \text{ ist stetig in } x_0 \\
&\qquad \Leftrightarrow \quad \lim_{x \to x_0} f(x) = f(x_0) \\
&\qquad \Leftrightarrow \quad \forall\, (x_n) \ \text{mit} \ \lim_{n \to \infty} x_n = x_0 \text{ gilt } \lim_{n \to \infty} f(x_n) = f(x_0) \\
&\qquad \Leftrightarrow \quad \forall\, \varepsilon > 0 \quad \exists\, \delta \quad \text{mit} \quad |f(x) - f(x_0)| < \varepsilon \quad \forall\, x \in (x_0 - \delta, x_0 + \delta)
\end{aligned}
$$

Bemerkung: Der Grenzwert von Funktionen ist über den Grenzwert von Folgen definiert und kann auch nur so überprüft werden.

0.2 Zwei Sonderfälle

Skalarfeld

Sei $f : \mathbb{R}^2 \to \mathbb{R}$
Visualisierung durch Höhenlinien: $H_c := \{x \in \mathbb{R}^n : f(x) = c\}$
Beispiel: $f(x, y) = x^2 + y^2$

Vektorfeld

Sei $f : \mathbb{R}^2 \to \mathbb{R}^2$
Beispiel: $f(x, y) = \begin{pmatrix} x \\ y \end{pmatrix}$

Kapitel 1

Differentialrechnung in höheren Dimensionen

1.1 Topologie

Skalarprodukt

Definition: $\langle x, y\rangle := x^\top y = \sum_{k=1}^n x_k y_k$ für $x, y \in \mathbb{R}^n$

Euklidische Norm

Definition: $\|x\|_2 := \sqrt{\langle x, x\rangle} = \sqrt{\sum_{k=1}^n x_k^2}$

1.1.1 Korollar

Sei $x \in \mathbb{R}^n$ mit $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

1.

$$\max_{1\leqslant k\leqslant n} |x_k| \leqslant \|x\| \leqslant \sqrt{n} \max_{1\leqslant k\leqslant n} |x_k|$$

2. Cauchy-Schwarz-Ungleichung:

$$\forall\ x, y \in \mathbb{R}^n \quad : \quad |\langle x, y\rangle| \leqslant \|x\| \cdot \|y\|$$

   Begründung (nicht Beweis!) durch alternative Definition: $\langle x, y\rangle = \|x\| \cdot \|y\| \underbrace{\cos\alpha}_{\leqslant 1}$

   Dabei ist α der Winkel der zwischen x und y eingeschlossen wird.
   Daraus folgt:

$$|\langle x, y\rangle| = \|x\| \cdot \|y\| \Leftrightarrow x, y \text{ sind lin. unabhängig} : x = \lambda y \text{ oder } y = \lambda x \text{ für } \lambda \in \mathbb{R}$$

3. $\|\cdot\|$ ist eine Norm. Eine Norm hat folgende Eigenschaften:

   (i) $\|x\| \geqslant 0$ und $\|x\| = 0 \Leftrightarrow x = 0$

   (ii) $\|\lambda x\| = |\lambda| \cdot \|x\|$

   (iii) $\|x + y\| \leqslant \|x\| + \|y\|$ Dreiecksungleichung

1.1.2 Konvention

Für $A \subset \mathbb{R}^n$ gilt für das Komplement $A^c = \mathbb{R}^n \setminus A$

1.1.3 Definition der ε-Umgebung

Sei $x_0 \in \mathbb{R}^n$ und $\varepsilon > 0$, dann gilt für die ε-Umgebung $U_\varepsilon(x_0)$ von x_0:

$$U_\varepsilon(x_0) := \{x \in \mathbb{R}^n : \|x - x_0\| < \varepsilon\}$$

Bemerkung: Die punktierte ε-Umgebung ist definiert als: $\dot{U}_\varepsilon = U_\varepsilon(a) \setminus \{a\}$

1.1.4 Topologische Grundbegriffe

Sei $A \subset \mathbb{R}^n$, dann heißt ein Punkt $x_0 \in \mathbb{R}^n$

(i) ein **innerer Punkt**, wenn gilt $\exists\ \varepsilon > 0$ mit $U_\varepsilon(x_0) \subset A$
Menge aller inneren Punkte: $\mathring{A} = \{x \in \mathbb{R}^n : \exists\ \varepsilon > 0 \text{ mit } U_\varepsilon(x) \subset A\}$

(ii) ein **Berührungspunkt**, wenn $\forall\ \varepsilon > 0$ gilt $U_\varepsilon(x_0) \cap A \neq \varnothing$
abgeschlossene Hülle: $\overline{A} = \{x \in \mathbb{R}^n : \forall\ \varepsilon > 0 \text{ gilt } U_\varepsilon(x_0) \neq \varnothing\}$

(iii) ein **Häufungspunkt**, wenn $\forall\ \varepsilon > 0$ gilt $(U_\varepsilon(x_0) \setminus \{x_0\}) \cap A \neq \varnothing$
Die Menge aller Häufungspunkte wird mit A' bezeichnet.

(iv) ein **Randpunkt**, wenn $\forall\ \varepsilon > 0$ gilt $U_\varepsilon(x_0) \cap A \neq \varnothing$ und $U_\varepsilon(x_0) \cap A^c \neq \varnothing$
Menge aller Randpunkte oder auch **Rand** von A wird mit ∂A bezeichnet.

Korollar

(i) $\mathring{A} \subset A$

(ii) $\mathring{A} \subset \overline{A}$

(iii) $\partial A \subset \overline{A}$

(iv) $\overline{A} = \mathring{A} \cup \partial A$

(v) $\overline{A} = A \cup \partial A$ (schwächere Aussage als (iv))

1.1.5 Definition von offen und abgeschlossen

Eine Menge $A \subset \mathbb{R}^n$ heißt

(i) **offen**, wenn $A = \mathring{A}$ gilt (A besteht nur aus inneren Punkten)

(ii) **abgeschlossen**, wenn $\partial A \subset A$ gilt (wenn der Rand in der Menge enthalten ist)

1.1.6 Beispiele

1. Jede ε-Umgebung $U_\varepsilon(x_0 \in \mathbb{R}^n)$ ist offen

2. Sei $I \subset \mathbb{R}$, dann gilt

   (i) I ist offen, wenn $I = (a, b)$ mit $-\infty \leqslant a \leqslant b \leqslant \infty$
   für $a = b$ gilt $I = \varnothing$ mit I offen
   und für $a = -\infty, b = \infty$ ist I auch offen

(ii) I ist abgeschlossen, wenn $I = [a, b]$ mit $a, b \in \mathbb{R}$
oder $I = (-\infty, b]$ oder $I = [a, \infty)$ oder $I = (-\infty, \infty) = \mathbb{R}$

(die reellen Zahlen sind offen und abgeschlossen zugleich)

1.1.7 Satz

für $A \subset \mathbb{R}^n$ sind folgenden Aussagen äquivalent:

(i) A ist abgeschlossen $A = \overline{A}$

(ii) A enthält alle Häufungspunkte, $A' \subset A$

(iii) A enthält alle Randpunkte, $\partial A \subset A$

(iv) A^c ist offen

1.1.8 Satz

(i) $\varnothing$ und $\mathbb{R}^n$ sind offen.

(ii) Die Vereinigung beliebig vieler offene Mengen ist offen:

$$\bigcup_{j \in J} (O_j \text{ offen}) = O \text{ offen}$$

(iii) Der Durchschnitt endlich vieler offener Mengen ist offen:

$$\bigcap_{j=1}^{n} (O_j \text{ offen}) = O \text{ offen}$$

Bemerkung: Für unendlich viele offene Mengen gilt dies nicht immer:

$$\bigcap_{k=1}^{\infty} \left(-\frac{1}{k}, \frac{1}{k}\right) = (-1, 1) \cap \left(-\frac{1}{2}, \frac{1}{2}\right) \cap \left(-\frac{1}{3}, \frac{1}{3}\right) \cap ... = \{0\} \text{ abgeschlossen}$$

Beispiel

Seien A_1, A_2 zwei abgeschlossene Mengen, dann gilt

(i) $A_1 \cup A_2$ ist abgeschlossen

Beweisidee: A_1 ist abgeschlossen $\Rightarrow A_1^c$ ist offen

$$(A_1 \cup A_2)^c \overset{\text{De Morgan}}{=} \underbrace{A_1^c}_{\text{offen}} \cap \underbrace{A_2^c}_{\text{offen}} \text{ ist offen wegen Satz 1.1.8}$$

$$((A_1 \cup A_2)^c)^c = A_1 \cup A_2 \text{ ist abgeschlossen}$$

1.1.9 Satz

(i) $\varnothing$ und $\mathbb{R}^n$ sind abgeschlossen.

(ii) Der Durchschnitt beliebig vieler abgeschlossener Mengen ist abgeschlossen:

$$\bigcap_{j\in J}(A_j \text{ abgeschlossen}) = A \text{ abgeschlossen}$$

(iii) Die Vereinigung <u>endlich</u> vieler abgeschlossenen Mengen ist abgeschlossen:

$$\bigcup_{j=1}^{n}(A_j \text{ abgeschlossen}) = A \text{ abgeschlossen}$$

Bemerkung: Für unendlich viele abgeschlossene Mengen gilt dies nicht immer:

$$\bigcup_{k=1}^{\infty}\left[-1+\frac{1}{n}, 1-\frac{1}{n}\right] = \{0\} \cup \left[-\frac{1}{2}, \frac{1}{2}\right] \cup \left[-\frac{2}{3}, \frac{2}{3}\right] \cup ... = (-1,1) \text{ offen}$$

1.1.10 Definition von beschränkt und kompakt

Eine Menge $A \subset \mathbb{R}^n$ heißt:

(i) **beschränkt** wenn $\exists\ c > 0$ mit $\|x\| < c \quad \forall\ x \in A$

(ii) **kompakt**, wenn A abgeschlossen und beschränkt ist.

1.2 Folgen

1.2.1 Definition von Konvergenz und Beschränktheit

Eine Folge $(a_k)_{k=1}^{\infty}$ heißt

(i) **konvergent**, wenn gilt

$$\exists\ a \in \mathbb{R}^n \quad \text{mit} \quad \forall\ \varepsilon > 0 \quad \exists\ N(\varepsilon): \quad \|a_k - a\| \quad \forall\ k \geqslant N(\varepsilon)$$

Dann ist a der Grenzwert der Folge:

$$a = \lim_{k\to\infty} a_k \quad \text{oder} \quad a_k \overset{k\to\infty}{\to} a$$

(ii) **beschränkt**, wenn $\exists\ c > 0$ mit $\|a_k\| < c \quad \forall\ k$

1.2.2 Bemerkung

Wenn eine Folge $(a_k) = \begin{pmatrix} \left(a_1^{(k)}\right) \\ \vdots \\ \left(a_n^{(k)}\right) \end{pmatrix} \in \mathbb{R}^n$ konvergiert, so gilt

(i) $\Leftrightarrow$ jede Komponente $\left(a_1^{(k)}\right), ..., \left(a_n^{(k)}\right)$ konvergiert:

$$\lim_{k\to\infty} a_k = a \quad \Leftrightarrow \quad \lim_{k\to\infty} a_i^{(k)} = a_i \quad \text{für } i = 1, ..., n$$

(ii) $\Leftrightarrow (a_k)$ erfüllt das **Cauchy-Kriterium**:

$$\forall\, \varepsilon > 0 \quad \exists\, N(\varepsilon): \quad \|a_k - a_l\| < \varepsilon \quad \forall\, k, l \geqslant N(\varepsilon)$$

(iii) $\Leftrightarrow$ jede Teilfolge von (a_k) konvergiert gegen a: $a_{l_k} \overset{k\to\infty}{\to} a$ für $l_1 \geqslant 1, l_2 \geqslant 2, \ldots$

(iv) der Grenzwert a ist eindeutig.

1.2.3 Satz von Bolzano Weierstraß

Jede beschränkte Folge im $\mathbb{R}^n$ besitzt einen konvergente Teilfolge.

Beispiele

(i) $n = 1$: Sei $A \leqslant (a_k) \leqslant B \quad \forall\, k$. Konstruiert man eine neue Schranke mit $\frac{A+B}{2}$ so liegen wiederum ∞ viele Elemente in der oberen und/oder unteren Hälfte.

(ii) Sei $(a_k) = \begin{pmatrix} (x_k) \\ (y_k) \end{pmatrix}$ eine beschränkte Folge im $\mathbb{R}^2$
$\Rightarrow (x_k), (y_k)$ sind beschränkte Folgen
$\overset{\text{Satz von Bolzano Wierstraß}}{\Rightarrow} \exists\, (x_k), (y_k)$ sind konvergent

1.2.4 Abschließende Bemerkungen

(i) Grenzwert Rechenregeln können aus dem $\mathbb{R}$ für $\mathbb{R}^n$ übernommen werden.
z.b. $a_k \overset{k\to\infty}{\to} a, \quad b_k \overset{k\to\infty}{\to} b \quad \Rightarrow \quad a_k^\top b_k \overset{k\to\infty}{\to} a^\top b$

(ii) Es gibt viele Zusammenhänge zwischen den Eigenschaften von Folgen und den topologischen Eigenschaften von Mengen.
z.b. Sei $A \subset \mathbb{R}^n$ und $a \in \mathbb{R}^n$ ein Häufungspunkt
$\Leftrightarrow \quad \exists (a_k)_{k=1}^\infty$ mit $a_k \in A \setminus \{a\}\, \forall\, k \quad$ und $\quad a_k \overset{k\to\infty}{\to} a$

1.3 Funktionsgrenzwerte und Stetigkeit

1.3.1 Definition

Eine Funktion $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$ nennt man eine Funktion mit n-Veränderlichen.

$$f(x_1, ..., x_n) = f(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}) = \begin{pmatrix} f_1(x_1, ..., x_n) \\ \vdots \\ f_m(x_1, ..., x_n) \end{pmatrix} \quad \text{mit} \quad f_1, ..., f_m : \mathbb{R}^n \to \mathbb{R}$$

1.3.2 Definition Grenzwert/Limes

Sei $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$ und $a \in \overline{A}$. Ein $b \in \mathbb{R}^m$ heißt Grenzwert von f für $x \to a$, wenn gilt:

$$\forall \varepsilon > 0 \quad \exists\, \delta(\varepsilon) > 0: \quad \|f(x) - b\| < \varepsilon \quad \forall\, x \in \dot{U}_{\delta(\varepsilon)}(a) \cap A$$

Bemerkung: Die Funktion f muss in a nicht stetig sein, so kann z.b. gelten:
$\lim_{x\to a} f(x) = b \neq f(a)$

1.3.3 Korollar

Sei $f : A \subset \mathbb{R}^n \to \mathbb{R}^m, a \in \overline{A}, b \in \mathbb{R}^m$ dann sind folgende Aussagen äquivalent:

(i) $f(x) \overset{x\to a}{\to} b$

(ii) $\|f(x) - b\| \overset{x\to a}{\to} 0 \in \mathbb{R}^1$ (Eine Norm bildet immer auf ein Skalar ab)

(iii) $f_1(x) \overset{x\to a}{\to} b_1, ..., f_m(x) \overset{x\to a}{\to} b_m$

Zusätzlich gilt das **Cauchy-Kriterium**:

$$\lim_{x\to a} f(x) = b \quad \Leftrightarrow \quad \forall\, \varepsilon > 0\; \exists\; \delta(\varepsilon) > 0 : \quad \|f(x), f(y)\| < \varepsilon \quad \forall\; x, y \in \dot{U}_{\delta(\varepsilon)}(a) \cap A$$

1.3.4 Beispiel

Sei $f(x,y) = \frac{xy}{x^2+y^2}$

$$a_k = \begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} \frac{1}{k} \\ \frac{1}{k} \end{pmatrix}, \quad f(a_k) = \frac{\frac{1}{k^2}}{\frac{1}{k^2} + \frac{1}{k^2}} = \frac{1}{2} \quad \forall\; k$$

$$b_k = \begin{pmatrix} x_k \\ 0 \end{pmatrix} \text{ mit } x_k \overset{k\to\infty}{\to} 0, \quad f(b_k) = \frac{0}{x_k^2} \quad \forall\; k$$

Da $\lim\limits_{k\to\infty} f(a_k) = \frac{1}{2} \neq 0 = \lim\limits_{k\to\infty} f(b_k)$ kann der Grenzwert nicht existieren.

1.3.5 Lemma Folgenkriterium

Sei $f : A \subset \mathbb{R}^n \to \mathbb{R}^m, a \in \overline{A}$

$$\underbrace{\exists b \in \mathbb{R}^m \text{ mit } \lim_{x\to a} f(x) = b}_{\text{der Grenzwert } b \text{ existiert}} \quad \Leftrightarrow \quad \underbrace{\begin{array}{l}\text{jede Folge } (x_k)_{k=1}^{\infty} \subset A \text{ mit } x_k \neq a\; \forall\; k \text{ und } x_k \overset{k\to\infty}{\to} a \\ \Rightarrow \quad f(x_k) \overset{k\to\infty}{\to} b\end{array}}_{\text{jede beliebige Folge konvergiert gegen } b}$$

1.3.6 Satz zu Grenzwerte verketteter Funktionen

Sei $A \subset \mathbb{R}^n, B \subset \mathbb{R}^m, a \in \overline{A}, f : A \to B, g : \overline{B} \to \mathbb{R}^l$

$$\exists\; b \in \overline{B} \text{ mit } \lim_{x\to a} f(x) = b, \quad \exists\; c \in \mathbb{R}^l \text{ mit } \lim_{y\to b} g(y) = c \quad \Rightarrow \quad \lim_{x\to a} \underbrace{g\left(f(x)\right)}_{(g\circ f)(x)} = \lim_{y\to b} g(y) = c$$

1.3.7 Beispiel

Sei $f(x,y) = e^{-x^2+y^2} = \exp\left(g(x,y)\right)$ mit $g(x,y) = x^2 + y^2$, dabei gilt:

$$\lim_{(x,y)^\top \to (0,0)^\top} g(x,y) = \lim_{(x,y)^\top \to (0,0)^\top} x^2 + y^2 = 0 \quad \Rightarrow \quad \lim_{z\to 0} f(z) = \lim_{z\to 0} e^z = 1$$

1.3.8 Definition der Stetigkeit

Sei $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$

(i) f ist **stetig** in $a \in A$ wenn gilt:

$$\forall\ \varepsilon > 0\ \exists \delta(\varepsilon): \quad \|f(x) - f(a)\| < \varepsilon \quad \forall\ x \in U_{\delta(\varepsilon)}(a) \cap A$$

Bemerkung: Es wird $\lim_{x\to a} f(x) = f(a)$ gefordert.
Diese Definition unterscheidet sich in der nicht punktierten ε-Umgebung und es gilt $f(a)$ anstatt b.

(ii) f ist stetig auf A, wenn f in jedem Punkt $a \in A$ stetig ist.

1.3.9 Bemerkung

(i) Kompositionen stetiger Funktionen sind wieder stetig: f, g stetig $\Rightarrow f + g, f - g, ...$ stetig

(ii) Das Folgenkriterium überträgt sich:
Sei $(a_k)_{k=1}^{\infty}$ eine Folge in A mit $\lim_{k\to\infty} a_k = a \quad \Leftrightarrow \quad \lim_{k\to\infty} f(a_k) = f(a)$

(iii) Ist A kompakt, dann nimmt eine stetige Funktion $f : A \to \mathbb{R}$ immer ein Maximum und Minimum an:

$$\exists\ x_m, x_M \in A \text{ mit } f(x_m) = \min_{x\in A} f(x), f(x_M) = \max_{x\in A} f(x)$$

1.4 Partielle Ableitungen, Richtungsableitungen

1.4.1 Definition der partiellen Ableitung

Die Funktion $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ heißt **partielle differenzierbar** in $a \in A$ nach der k-ten Variable x_k mit $k \in \{1, ..., n\}$ wenn der folgender Grenzwert existiert:

$$\frac{\partial}{\partial x_k} f(a) = f_{x_k}(a) = \lim_{h\to 0} \frac{f(a + h \cdot e_k) - f(a)}{h}$$

Existieren alle partielle Ableitungen $f_{x_1}(a), ..., f_{x_n}(a)$, dann ist der **Gradient** von f wie folgt definiert:

$$\nabla f(a) = \begin{pmatrix} f_{x_1}(a) \\ \vdots \\ f_{x_n}(a) \end{pmatrix}$$

und die Funktion f heißt mindestens einmal partielle differenzierbar. Sind die partiellen Ableitungen $f_{x_1}(a), ..., f_{x_n}(a)$ zudem stetig, so heißt f einmal stetig differenzierbar: $f \in C^1(A, \mathbb{R}^m)$ oder kurz $f \in C^1(A)$.

1.4.2 Beispiel

Sei $f(x,y,z) = x^2 - xy + 3z$

$$\begin{aligned}
\frac{\partial}{\partial x} f(x,y,z) &= \lim_{h\to 0} \frac{f(x+h,y,z) - f(x,y,z)}{h} \\
&= \lim_{h\to 0} \frac{(x+h)^2 - (x+h)y + 3z - (x^2 - xy + 3z)}{h} \\
&= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} - \frac{(x+h)y - xy}{h} + \frac{3z - 3z}{h} \\
&= \left(\frac{d}{dx} x^2\right) - \left(\frac{d}{dx} x\right) y + \left(\frac{d}{dx} 0\right) z \\
&= 2x - y + 0 \\
\Rightarrow \nabla f(x,y,z) &= \begin{pmatrix} 2x - y \\ -x \\ 3 \end{pmatrix}
\end{aligned}$$

1.4.3 Definition der Richtungsableitung

Sei $a, r \in \mathbb{R}^n$ mit $\|r\| = 1$ (normiert), $f : \mathbb{R}^n \to \mathbb{R}^m$, dann heißt der folgende Grenzwert die Richtungsableitung von f bei a in Richtung r:

$$\frac{\partial}{\partial r} f(a) = f_r(a) = \lim_{h\to 0} \frac{f(a + h\cdot r) - f(a)}{h}$$

Bemerkung

(i) Ist $r = e_k$, dann erhalten wir gerade eine partielle Ableitung.

(ii) Es gibt Funktionen die in a in <u>jede Richtung differenzierbar</u> sind, aber in a <u>nicht stetig</u> sind!

1.5 Total Differenzierbarkeit

Idee: Differenzierbare Funktionen sind lokal im Punkt x_0 linear approximierbar:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \underbrace{r(x)\|x - x_0\|}_{\tilde{r}(x)}$$

Dabei muss der Fehler $\tilde{r}(x) = r(x)\|x - x_0\|$ *schneller gegen Null gehen als x gegen x_0* also muss $\tilde{r}(x) = \mathrm{o}(x - x_0)$ gelten (Landau-Notation: klein-oh).

1.5.1 Definition der totalen Differenzierbarkeit

Sei $f : A \subset \mathbb{R}^n \to \mathbb{R}^m, A$ offen, $x_0 \in A$

(i) Die Funktion f nennt man **total differenzierbar** bei x_0, wenn eine Matrix $A \in \mathbb{R}^{m\times n}$ existiert, mit der sich die Funktion f in einer ε-Umgebung um x_0 mittels einer Hyperebene approximieren lässt:

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

Dann nennt man die Matrix $A = f'(x_0) = \frac{\partial}{\partial x} f(x_0)$ die total Ableitung von f in x_0.

(ii) Ist $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$ partiell diff'bar, so nennt man die Ableitung **Jacobi-Matrix**:

$$f'(x_0) = \frac{\partial}{\partial x} f(x_0) = J_f(x_0) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x_0) & \dots & \frac{\partial}{\partial x_n} f_1(x_0) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} f_m(x_0) & \dots & \frac{\partial}{\partial x_n} f_m(x_0) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Bemerkung: Es gilt: $\exists f'(x_0) \quad \Rightarrow \quad f'(x_0) = J_f(x_0)$, nicht aber die Gegenrichtung! Es kann also sein, dass die Jacobi-Matrix J_f existiert die Funktion aber nicht total diff'bar ist.

1.5.2 Beispiele

(i)

$$f(r, \varphi) = r \cdot \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} \quad \Rightarrow \quad J_f = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

(ii) $f(x) = a + b^\top(x - x_0), \quad f : \mathbb{R}^n \to \mathbb{R}, \quad a \in \mathbb{R}, \quad b, x_0 \in \mathbb{R}^n$
$\Rightarrow \quad f(x_0) = a, \quad f'(x_0) = b^\top$

(iii) $f(x) = a + A(x - x_0), \quad f : \mathbb{R}^n \to \mathbb{R}^m, \quad a \in \mathbb{R}^m, \quad A \in \mathbb{R}^{m \times n}, \quad x_0 \in \mathbb{R}^n$
$\Rightarrow \quad f(x_0) = a, \quad f'(x_0) = A$

Bemerkung: Beispiel (ii) und (iii) sind lineare Funktionen.

1.5.3 Satz

Ist $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ in jedem Punkt $x_0 \in A$ total differenzierbar, so ist f stetig in A.

Beweis:

$$f(x) = \underbrace{f(x_0)}_{\overset{x \to x_0}{\to} f(x_0)} + \underbrace{A\,\underbrace{(x - x_0)}_{\overset{x \to x_0}{\to} 0 \in \mathbb{R}^n}}_{\overset{x \to x_0}{\to} 0 \in \mathbb{R}^n} + \underbrace{r(x)}_{\overset{x \to x_0}{\to} 0 \in \mathbb{R}^m} \underbrace{\|x - x_0\|}_{\overset{x \to x_0}{\to} 0 \in \mathbb{R}} \qquad \text{mit } r(x) \overset{x \to x_0}{\to} 0$$

$\lim\limits_{x \to a} f(x) = f(x_0) \quad \square$

1.5.4 Satz

Sei $f : A \subset \mathbb{R}^n \to \mathbb{R}^m, x_0 \in A$

a. Ist f total differenzierbar in x_0, so gilt

(i) $f'(x_0) = J_f(x_0)$

(ii) f ist in jede Richtung r differenzierbar mit: $\frac{\partial}{\partial r} f(x_0) = J_f(x_0) \cdot r$

Beweis: Es ist zu zeigen, dass wenn f differenzierbar in x_0 die Ableitung gerade die Form $\frac{\partial}{\partial r}f(x_0) = J_f(x_0)\cdot r$ besitzt. Für diese Ableitung muss folgendes gelten:

$$f(x) = f(x_0) + A(x-x_0) + \tilde{r}(x) \text{ mit } A = f'(x_0) \text{ und } \tilde{r} \in \mathrm{o}(\|x-x_0\|) \Rightarrow \frac{\tilde{r}(x)}{\|x-x_0\|} \overset{x\to x_0}{\to} 0$$

$$\overset{f(x)=f(x_0+r\cdot h}{\Leftrightarrow} \qquad f(x_0+r\cdot h) - f(x_0) = A\cdot r\cdot h + \tilde{r}(x) = f'(x_0)rh + \tilde{r}(x)$$

also muss folgendes gezeigt werden:

$$\left\|\underbrace{\frac{f(x_0+r\cdot h)-f(x_0)}{h}}_{\text{Diff'Quotient für } \frac{\partial f}{\partial r}} - \underbrace{f'(x_0)\cdot r}_{\substack{\text{Grenzwert-}\\ \text{kandidat}}}\right\| \overset{h\to 0}{\to} 0$$

$$\left\|\frac{f(x_0+r\cdot h)-f(x_0)}{h} - f'(x_0)\cdot r\right\| = \left\|\frac{f'(x_0)rh+\tilde{r}(x)}{h} - f'(x_0)r\right\| =$$

$$\left\|\frac{f'(x_0)r\cancel{h}}{\cancel{h}} + \frac{\tilde{r}(x)}{h} - f'(x_0)r\right\| = \left\|\frac{\tilde{r}(x)}{h}\right\| = \left\|\frac{\tilde{r}(x)}{x-x_0}\right\| \overset{x\to x_0}{\to} 0$$

Ist $r = e_k$ so erhält man gerade eine Spalte der Jacobi-Matrix.

b. Existieren in x_0 alle partiellen Ableitungen (also alle Komponenten der Jacobi-Matrix) und diese stetig sind $\quad\Rightarrow\quad f$ ist in x_0 total differenzierbar.

Beweis: Für den Fall $n = 2, m = 1$ muss folgendes gezeigt werden:

$$\exists\ \nabla f(x_0) \text{ und } \tilde{r}(x) \text{ mit } f(x) = f(x_0) + \nabla f(x_0)^\top(x-x_0) + \tilde{r}(x)$$

$$\text{oder } \left\|\frac{f(x)-f(x_0)}{x-x_0} - \nabla f(x_0)\right\| = \frac{\|f(x)-f(x_0)-\nabla f(x_0)(x-x_0)\|}{\|x-x_0\|} \overset{x\to x_0}{\to} 0$$

Sei $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ und sei $x_0 = a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.
Nebenrechnung: Definition zweier Hilfsfunktionen g_1, g_2:

$$\begin{aligned}
& \text{Sei } g_1(t) = f(t, x_2) \quad g_2 : \mathbb{R} \to \mathbb{R} \\
\overset{\text{MWS}}{\Rightarrow} \quad & \exists\ \xi_1 \in (a_1, x_1) \text{ mit } g_1'(\xi_1) = \frac{g_1(x_1)-g_1(a_1)}{x_1-a_1} \\
& = \frac{\partial}{\partial x_1} f(\xi_1, x_2) = \frac{f(x_1,x_2)-f(a_1,x_2)}{x_1-a_1} \\
\Leftrightarrow \quad & f(x_1,x_2) - f(a_1,x_2) = \frac{\partial}{\partial x_1} f(\xi_1,x_2)(x_1-a_1) \\
\text{analog gilt für} \quad & g_2(t) = f(a_1, t) \quad g_2 : \mathbb{R} \to \mathbb{R} \\
& f(a_1,a_2) - f(a_1,x_2) = \frac{\partial}{\partial x_2} f(a_1,\xi_2)(x_2-a_2)
\end{aligned}$$

Damit gilt:

$$f(x) - f(a) = f(x_1, x_2) - f(a_1, a_2) = f(x_1, x_2) \underbrace{- f(a_1, x_2) + f(a_1, x_2)}_{=0} - f(a_1, a_2)$$

$$\overset{\substack{\text{mit Resultat}\\\text{aus Neben-}\\\text{rechnung}}}{=} \frac{\partial}{\partial x_1} f(\xi_1, x_2)(x_1 - a_1) + \frac{\partial}{\partial x_2} f(a_1, \xi_2)(x_2 - a_2)$$

$$= \begin{pmatrix} f_{x_1}(\xi_1, x_2) \\ f_{x_2}(a_1, \xi_2) \end{pmatrix}^\top \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix} = \begin{pmatrix} f_{x_1}(\xi_1, x_2) \\ f_{x_2}(a_1, \xi_2) \end{pmatrix}^\top (x - a)$$

Für $x \to a$ gilt:

$$\begin{aligned} x_1 &\to a_1 \quad x_2 \to a_2 \\ \xi_1 &\to a_1 \quad \xi_2 \to a_2 \end{aligned}$$

da f_{x_1}, f_{x_2} stetig, folgt:

$$\begin{aligned} f_{x_1}(\xi_1, x_2) &\to f_{x_1}(a_1, a_2) \\ f_{x_2}(a_1, \xi_2) &\to f_{x_2}(a_1, a_2) \\ \Rightarrow \begin{pmatrix} f_{x_1}(\xi_1, x_2) \\ f_{x_2}(a_1, \xi_2) \end{pmatrix}^\top &\to \nabla f(a_1, a_2) = \nabla f(x_0) \end{aligned}$$

und es gilt:

$$\frac{\left\| f(x) - f(x_0) - \overbrace{\begin{pmatrix} f_{x_1}(\xi_1, x_2) \\ f_{x_2}(a_1, \xi_2) \end{pmatrix}^\top}^{\overset{x \to x_0}{\to} \nabla f(x_0)} (x - x_0) \right\|}{\|x - x_0\|} \overset{x \to x_0}{\to} 0$$

1.5.5 Bemerkung

Sei r eine Richtung mit $\|r\| = 1$ und $x = x_0 + r$, dann gilt:

$$\begin{aligned} & f(x) \approx f(x_0) + \nabla f(x_0)^\top \cdot r \\ & \quad \Rightarrow 1.\ \text{Fall}: \quad r, \nabla f(x_0) \text{ zeigen in dieselbe Richtung}: \\ & \qquad f(x) - f(x_0) \approx \|\nabla f(x_0)\| \|r\| = \|\nabla f(x_0)\| > 0 \\ & \quad \Rightarrow 2.\ \text{Fall}: \quad r, \nabla f(x_0) \text{ zeigen in entgegengesetzte Richtungen}: \\ & \qquad f(x) - f(x_0) \approx -\|\nabla f(x_0)\| < 0 \end{aligned}$$

In allen Fällen gilt Näherungsweise:

$$-\|\nabla f(x_0)\| < \nabla f(x_0)^\top r \leqslant \|\nabla f(x_0)\|$$

Fazit: Beim Reinzoomen sind die Höhenlinien parallel. Der Gradient zeigt in Richtung des steilsten Anstieges.

1.5.6 Satz zur Kettenregel

Ist $f : A \subset \mathbb{R}^n \to B \subset \mathbb{R}^m$ differenzierbar in $a \in A$ und $g : B \subset \mathbb{R}^m \to \mathbb{R}^l$ differenzierbar in $b \in B$, so gilt:

$$(g \circ f)'(a) = g'\left(\underbrace{f(a)}_{=b}\right) f'(a) = \underbrace{J_g(b)}_{\in \mathbb{R}^{l \times m}} \underbrace{J_f(a)}_{\in \mathbb{R}^{m \times n}}$$

Beispiel aus der Strömungsmechanik: Die Funktion $f : \mathbb{R}^4 \to \mathbb{R}, f = f(x, y, z, t)$ beschreibe die Eigenschaften eines Teilchens in einer Strömung. Dabei kann die Bewegung der Position im Raum x, y, z als Abhängigkeit von der Zeit beschrieben werden. Dazu definieren wir den Weg $\gamma(t)$:

$$\gamma : \mathbb{R} \to \mathbb{R}^4, \quad \gamma(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \\ t \end{pmatrix} \quad \text{mit} \quad \frac{d\gamma}{dt} = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \\ 1 \end{pmatrix}$$

Nun leiten wir die verkettete Funktion $\hat{f}(t) = (f \circ \gamma)(t) = f(\gamma(t))$ nach der Zeit t ab:

$$\begin{aligned} \frac{d\hat{f}}{dt} &= \left(\hat{f}(\gamma(t))\right)' = f'(h(t))\gamma'(t) = \nabla f \cdot \frac{d\gamma}{dt} \\ &= \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}^\top \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \\ 1 \end{pmatrix} \\ &= \frac{\partial f}{\partial x} \underbrace{\frac{dx}{dt}}_{u} + \frac{\partial f}{\partial y} \underbrace{\frac{dy}{dt}}_{v} + \frac{\partial f}{\partial z} \underbrace{\frac{dz}{dt}}_{w} + \frac{\partial f}{\partial t} \end{aligned}$$

Dabei beschreibt der Vektor $\begin{pmatrix} u \\ v \\ w \end{pmatrix}$ die Geschwindigkeit im Raum.

1.6 Lokale Extremstellen und Mittelwertsätze

In einer Dimension gilt:

1. Mittelwertsatz

Ist f differenzierbar auf (a, b) und stetig auf $[a, b]$, so gilt:

$$\exists\, \xi \in (a, b) \text{ mit } f'(\xi) = \frac{f(b) - f(a)}{b - a}$$

Satz von Rolle

Ist f differenzierbar auf (a, b) und stetig auf $[a, b]$ und gilt $f(a) = f(b)$, so gilt:

$$\exists\, \xi \in (a, b) \text{ mit } f'(\xi) = 0$$

1.6.1 Definition lokale/globale Extremstellen

(i) Eine Funktion $f : A \subset \mathbb{R}^n \to \mathbb{R}$ (Skalarfeld) hat bei $x_0 \in A$ ein lokales Minimum (Maximum) wenn in einer Umgebung $U = U_\varepsilon(x_0) \cap A$ für $\varepsilon > 0$ (offen bezüglich A) von x_0 gilt:

$$f(x_0) \overset{(\geqslant)}{\leqslant} f(x) \quad \forall x \in U$$

Ist bei x_0 ein lokales Minimum (Maximum) dann nennt man x_0 eine lokale Extremstelle.

(ii) f besitzt in x_0 ein globales Minimum (Maximum), wenn gilt:

$$f(x_0) \overset{(\geqslant)}{\leqslant} f(x) \quad \forall x \in A$$

1.6.2 Satz zur notwendigen Bedingung für eine lokale Extremstelle

Besitzt $f : \mathring{A} \subset \mathbb{R}^n \to \mathbb{R}$ bei $x_0 \in A$ eine lokale Extremstelle und f ist partiell differenzierbar, dann ist

$$\nabla f(x_0) = 0$$

Bemerkung: Der Rand ist ausgeschlossen da $\mathring{A}$ (alle inneren Punkte) in der Definition verwendet wurde.
Auch gilt:

$$x_0 \text{ ist eine lokale Extremstelle} \quad \overset{\not\Leftarrow}{\Rightarrow} \quad f'(x_0) = 0$$

Aus $f'(x_0)$ folgt nicht direkt die Extremstelle, denn Sattelpunkte sind keine Extremstellen.

Beweis: ...

1.6.3 Definition des kritischen Punktes

Ein $x_0 \in \mathbb{R}^n$ mit $\nabla f(x_0) = 0$ heißt **kritischer** oder stationärer Punkt.

1.6.4 Mittelwertsatz

Sei $f : G \subset \mathbb{R}^n \to \mathbb{R}$ differenzierbar und sei G offen und enthalte die Menge $\overline{a, b} = \{a, b \in G \text{ mit } a + t(b-a) : t \in [0,1]\}$ (a, b können durch eine Gerade verbunden werden). Dann:

$$\exists\, \xi \in (0,1) \quad \text{mit} \quad f(b) = f(a) + \nabla f(a + \xi(b-a))^\top (b-a)$$

Bemerkung:

$$h(t) = a + t(b-a) \quad g(t) = f(h(t)) \quad \text{(differenzierbar)}$$
$$\Rightarrow \quad \exists\, \xi \in (0,1) \quad \text{mit} \quad g'(\xi) = \frac{g(1) - g(0)}{1 - 0}$$

Beweis: Definiere $h(t) = a + t(b-a)$ und $g : [0,1] \to \mathbb{R}, g(t) = f(h(t))$ differenzierbar, damit gilt:

$$
\begin{aligned}
\exists\ \xi \in (0,1) \quad \text{mit} \quad g'(\xi) &= \frac{g(1)-g(0)}{1-0} = g(1) - g(0) = f(a) - f(b) \\
g'(\xi) &= \frac{d}{dt} g(t)|_{t=\xi} = \frac{d}{dt} f(h(t))|_{t=\xi} \\
&\overset{\substack{\text{Ketten-}\\\text{regel}}}{=} f'(h(t))h'(t)|_{t=\xi} \\
&= \nabla f(a + \xi(b-a))^\top (b-a) = f(a) - f(b) \\
\Leftrightarrow \quad & f(b) = f(a) + \nabla f(a + \xi(b-a))^\top (b-a)
\end{aligned}
$$

1.6.5 Definition eines Gebiets

(i) Ein Menge, die wie folgt konstruiert werden kann, heißt **Polygonzug**:

$$\overline{a_0, ..., a_k} = \bigcup_{j=1}^{k} \overline{a_{j-1}, a_j} \qquad \text{mit} \qquad a_0, ..., a_k \in \mathbb{R}^n$$

(ii) Eine Menge $M \subset \mathbb{R}^n$ heißt **kurvenweise zusammenhängend** wenn zu beliebigen $a, b \in M$ eine stetige Funktion $\gamma : [0,1] \to M$ mit $\gamma(0) = a, \gamma(1) = b$ existiert.

(iii) Eine Menge $G \subset \mathbb{R}^n$ heißt **Gebiet**, wenn G offen und kurvenweise zusammenhängend ist.

1.6.6 Bemerkungen zu Gebieten

(i) Ein Gebiet G entspricht einem offenen Intervall $(a, b) \subset \mathbb{R}$ im Eindimensionalen: Der Rand ist nicht dabei, es hat keine Inseln.

(ii) Man kann zeigen, dass es reicht, wenn $a, b \in G$ mit einem Polygonzug verbunden werden kann.

1.6.7 Satz

Sei $G \subset \mathbb{R}^n$ ein Gebiet, $G \neq \varnothing$, und $f : G \to \mathbb{R}$ differenzierbar, dann gilt:

$$f(x) = \text{konst.} \quad \Leftrightarrow \quad \nabla f(x) = 0 \quad \forall\ x \in G$$

Beweis: ...

1.6.8 Definition partieller Ableitungen r'ter Ordnung

Für $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ definiert man (wenn diese auch existieren) induktiv für $x_0 \in A$ und $k_1, ..., k_r \in \{1, ..., n\}$ die partiellen Ableitungen r'ter Ordnung als:

$$\frac{\partial^n}{\partial x_{k_1} \dots \partial x_{k_r}} f(x_0) = f_{x_{k_1} \dots x_{k_r}} = \begin{cases} f(x_0) & r = 0 \\ \frac{\partial}{\partial x_{k_1}} f(x_0) & r = 1 \\ \frac{\partial}{\partial x_{k_1}} \left(\frac{\partial^{r-1}}{\partial x_{k_2} \dots \partial x_{k_r}} f(x_0) \right) & r > 1 \end{cases}$$

Existieren alle Ableitungen r'ter Ordnung und sind diese zudem stetig, so nennt man die Funktion f r-mal stetig differenzierbar: $f \in C^r(A; \mathbb{R}^m)$.

1.6.9 Definition der Hessematrix

Ist $f : A \subset \mathbb{R}^n \to \mathbb{R}$ 2-mal stetig differenzierbar bei $x_0 \in A$, dann ist die **Hessematrix** wie folgt definiert:

$$H_f(x_0) = \begin{pmatrix} f_{x_1,x_1}(x_0) & \dots & f_{x_1,x_n}(x_0) \\ \vdots & & \vdots \\ f_{x_m,x_1}(x_0) & \dots & f_{x_m,x_n}(x_0) \end{pmatrix}$$

1.6.10 Beispiele

(i)

$$f(x,y) = 2xy^3 + y\log x$$

(ii)

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

$\Rightarrow$ Hessematrix ist nicht symmetrisch.

(iii) $A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n, c \in \mathbb{R}, Q : \mathbb{R}^n \to \mathbb{R}$

$$\begin{aligned} Q(x) &= x^\top Ax + b^\top x + c \\ \nabla Q(x) &= \left(A + A^\top\right) x + b \quad \overbrace{(= 2Ax + b)}^{\text{wenn } A \text{ sym.}} \\ H_Q(x) &= A + A^\top \quad \underbrace{(= 2A)}_{\text{wenn } A \text{ sym.}} \end{aligned}$$

Q wird eine quadratische Funktion genannt.

1.6.11 Satz von Schwarz

Ist $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$ in x_0 2-mal stetig partielle differenzierbar ($f \in C^2(A)$), dann ist $H_f(x_0) \quad \forall\, x_0 \in A$ symmetrisch und es gilt:

$$f_{x_l,x_k}(x_0) = \frac{\partial}{\partial x_l \partial x_k} f(x_0) = \frac{\partial}{\partial x_k \partial x_l} f(x_0) = f_{x_k,x_l}(x_0) \quad \forall\, l,k \in \{1,...,k\}$$

1.6.12 Satz von Taylor mit quadratischem Restglied

Seien $a, b \in G \subset \mathbb{R}^n, f \in C^2(G), G$ ein Gebiet, dann:

$$\exists\, \xi \in \overline{a,b} \quad \text{mit} \quad f(b) = f(a) + \nabla f(a)^\top (b-a) + \frac{1}{2}(b-a)^\top H_f(\xi)(b-a)$$

Beweis: Definiere $g(t) = f(h(t))$ mit $h(t) = a + t(b-a) \quad \Rightarrow \quad g(0) = f(a), g(1) = f(b)$. Bei Funktionen mit einem Skalar, gilt der eindimensionale Taylor mit einer Zwischenstelle $z \in [0,1]$:

$$\Rightarrow \underbrace{g(1)}_{f(b)} = \underbrace{g(0)}_{f(a)} + \underbrace{g'(0)(1-0)}_{\substack{g'(t)=\frac{d}{dt}f(h(t)) \\ =f'(h(t))h'(t) \\ =\nabla f(a+t(b-a))^\top(b-a) \\ =(b-a)^\top \nabla f(a+t(b-a))}} + \underbrace{\frac{1}{2}g''(z)(1-0)^2}_{g''(t)=\ldots=(b-a)^\top H_f(a+t(b-a))(b-a)}$$

$$= f(a) + \nabla f(a)^\top (b-a) + \frac{1}{2}(b-a)^\top H_f(\underbrace{\xi}_{\xi=a+t(b-a)})(b-a)$$

1.6.13 Definition von Definitheit

Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch ($A = A^\top$):

a. Die durch $Q_A(x) = x^\top Ax$ definierte Funktion $Q : \mathbb{R}^n \to \mathbb{R}$ heißt quadratische Form von A.

b. Die Matrix A und ihre quadratische Form Q_A heißen:

(i) **positiv definit**, wenn

$$Q_A(x) = x^\top Ax > 0 \quad \forall\, x \in \mathbb{R}^n \text{ mit } x \neq 0$$

negativ definit, wenn

$$Q_A(x) = x^\top Ax < 0 \quad \forall\, x \in \mathbb{R}^n \text{ mit } x \neq 0$$

oder kurz **definit** falls die Matrix A positiv oder negativ definit ist.

(ii) **semi definit** falls die Matrix A positiv semi definit oder (negativ semi definit) ist:

$$Q_A(x) = x^\top Ax \overset{(\leqslant)}{\geqslant} 0 \quad \forall\, x \in \mathbb{R}^n \text{ mit } x \neq 0$$

(iii) **indefinit**, wenn ein $x_1, x_2 \in \mathbb{R}^n$ existieren mit:

$$\underbrace{x_1^\top Ax_1}_{Q_A(x_1)} < 0 \quad \text{und} \quad \underbrace{x_2^\top Ax_2}_{Q_A(x_2)} > 0$$

1.6.14 Beispiele

(i) $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

(ii) $A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$

(iii) $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

(iv) $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$

(v) $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

(vi) $A = \begin{pmatrix} a & b & 0 \\ b & c & 0 \\ 0 & 0 & d \end{pmatrix}$

(vii) $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

1.6.15 Satz zum Hauptminorenkriterium

Sei $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \in \mathbb{R}^{n\times n}$ symmetrisch,

dann ist der k'te Hauptminor definiert als: $D_k = \det \underbrace{\begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix}}_{k\times k \text{ Teilmatrix oben links}}$

Für die Hauptminoren $D_1, ..., D_n$ gilt:

a. A ist positiv definit $\Leftrightarrow$ $D_1 > 0, D_2 > 0, ..., D_n > 0$ alle Hauptminoren sind positiv

b. A ist negativ definit $\Leftrightarrow$ $D_1 < 0, D_2 > 0, D_3 < 0, ...$ oder $D_k = \begin{cases} < 0 & k \text{ ungerade} \\ > 0 & k \text{ gerade} \end{cases}$

c. (i) $D_k < 0$ mit k gerade $\Rightarrow$ A ist indefinit

(ii) $D_k < 0 < D_l$ mit k, l ungerade $\Rightarrow$ A ist indefinit

Beispiele

(i) $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

(ii) $A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$

(iii) $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

1.6.16 Satz über die hinreichenden Bedingungen für lokale Extremstellen

Sei $f \in C^2(U)$ in einer Umgebung U um x_0 und gilt $\nabla f(x_0) = 0$ sowie:

(i) $H_f(x_0)$ ist positiv definit $\Rightarrow$ x_0 ist eine lokale Minimalstelle.

(ii) $H_f(x_0)$ ist negativ definit $\Rightarrow$ x_0 ist eine lokale Maximalstelle.

Beweis: o.B.d.A nur für (i), denn (ii) folgt analog mit gedrehtem Ungleichungszeichen.

Sei $U = U_\varepsilon(x_0)$ eine ε-Umgebung um den Punkt $x_0 \in \mathbb{R}^n$:

$$\begin{aligned}
\text{Sei } f \in C^2(U) &\Rightarrow H_f \in C^2(U, \mathbb{R}^{n\times n}) \text{ mit allen Komponenten stetig}\\
&\Rightarrow \tilde{g} : \mathbb{R}^n \to \mathbb{R}, \tilde{g}(x) = (x-x_0)^\top H_f(x)(x-x_0) \text{ ist stetig}\\
&\Rightarrow g(x) = (x-x_0)^\top H_f(h(x))(x-x_0) \text{ ist stetig}\\
\text{ist } H_f(x_0) \text{ pos. def.} &\Rightarrow (x-x_0)^\top H_f(x_0)(x-x_0) > 0 \text{ für } x \neq x_0\\
&\Rightarrow (x-x_0)^\top H_f(h(x_0))(x-x_0) > 0 \text{ falls } z(x_0) \text{ nahe genug bei } x_0
\end{aligned}$$

Sei $h(x_0) = x_0 + \xi(x-x_0)$ dann gilt:

$$f(x) = f(x_0) + \underbrace{\underbrace{\nabla f(x_0)^\top}_{=0}(x-x_0)}_{=0} + \underbrace{\frac{1}{2}(x-x_0)^\top \underbrace{H_f(x_0+\xi(x-x_0))}_{\text{pos. def. in } U_\varepsilon(x_0)}(x-x_0)}_{>0\ \forall x \in U_\varepsilon(x_0)} \geqslant f(x_0)$$

1.6.17 Definition der Sattelpunkte

Sei $U = U_\varepsilon(x_0), f \in C^2(U), \nabla f(x_0) = 0, H_f(x_0)$ indefinit, dann besitzt die Funktion f bei x_0 einen **Sattelpunkt**.

Bemerkung: $\forall\ \varepsilon > 0$ gilt:

$$\exists\ x_1, x_2 \in U_\varepsilon(x_0) \text{ mit } f(x_1) > f(x_0) \ \underline{\text{und}}\ f(x_2) < f(x_0)$$

1.6.18 Satz für den Fall $n = 2$

Ist $f \in C^2(U)$ für eine Umgebung U von $x_0 = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ gilt weiter $f_x(a,b) = f_y(a,b) = 0$, dann ist x_0:

(i) eine lokale **Minimal**stelle, wenn
$f_{xx}(a,b) > 0$ (erster Hauptminor D_1) und
$f_{xx}(a,b)f_{yy}(a,b) - 2(f_{xy}(a,b))^2 > 0$ (zweiter Hauptminor D_2)

(ii) eine lokale **Maximal**stelle, wenn
$f_{xx}(a,b) < 0$ (erster Hauptminor D_1) und
$f_{xx}(a,b)f_{yy}(a,b) - 2(f_{xy}(a,b))^2 > 0$ (zweiter Hauptminor D_2)

(iii) ein **Sattelpunkt**, wenn
$f_{xx}(a,b) < 0$ (erster Hauptminor D_1) oder
$f_{xx}(a,b)f_{yy}(a,b) - 2(f_{xy}(a,b))^2 < 0$ (zweiter Hauptminor D_2)

1.6.19 Beispiel

Sei $f(x,y,z) = x^2 + xy + y^2 - \cos z \in C^2(\mathbb{R}^2)$:

$$\nabla f(x,y,z) = \begin{pmatrix} 2x+y \\ x+2y \\ \sin z \end{pmatrix} \overset{!}{=} 0$$

1.7 Extremstellen unter Nebenbedingungen und implizite Funktionen

Bisher: Optimierungsproblem ohne Nebenbedingungen: $\begin{cases} f(x,y) \to \min \\ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \end{cases}$

Lösbar in drei Schritten:

1. lokale Minimalstellen bestimmen
2. untersuche $f(x,y)$ für $\left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\| \to \infty$
3. vergleiche die Resultate aus 1. und 2.

Bemerkung: aus $f(x,y) \to \min$ wird $-f(x,y) \to \max$ darum reicht es o.B.d.A. denn Fall $\to \min$ zu betrachten.

Jetzt: Optimierungsproblem mit Nebenbedingungen: $\begin{cases} f(x,y) \to \min \\ \begin{pmatrix} x \\ y \end{pmatrix} \in A \subset \mathbb{R}^2 \end{cases}$

Es ist bekannt, dass wenn $f \in C(A)$ und $A \neq \varnothing$ kompakt (abgeschlossen und beschränkt) dann existiert sicher eine Lösung (Satz von Weierstraß: f stetig und A kompakt $\Rightarrow$ es wird ein Minimum (Maximum) angenommen).

Es sind zwei Fälle möglich:

(i) globale Minimalstelle liegt in $\mathring{A}$ (im Inneren)

(ii) globale Minimalstelle liegt in ∂A (auf dem Rand)

Wenn zusätzlich $f \in C^2(A)$ gilt, kann die Minimalstelle wie folgt gefunden werden:

1. bestimme lokale Minimalstellen mit $\nabla f(x_0) = 0$ und $H_f(x_0)$ positiv definit in $\mathring{A}$
2. bestimme lokale Minimalstelle in ∂A
   Bemerkung: Eckpunkte müssen gesondert betrachtet werden, denn die Funktion kann an diesen Stellen nicht differenzierbar sein.
3. wähle das Minimum aus 1. und 2.

Wiederholung im Eindimensionalen

Eine Funktion $f : X$ (Definitionsbereich) $\to Y$ (Bildbereich) ist:

(i) **injektiv**, wenn

$$x_1, x_2 \in X \text{ mit } x_1 \neq x_2 \quad \Rightarrow \quad f(x_1) \neq f(x_2)$$

Bemerkung: Von zwei verschiedenen Punkten aus dem Definitionsbereich darf nicht auf den gleichen Punkt im Bildbereich abgebildet werden.

(ii) **surjektiv**, wenn

$$\forall\, y \in Y \quad \exists\, x \in X \quad \text{mit} \quad f(x) = y$$

Bemerkung: Für alle Bildpunkte existiert ein Punkt im Definitionsbereich.

(iii) **bijektiv**, wenn die Funktion f injektiv und surfjektiv ist.

Beispiel

Sei $d : C^1(\mathbb{R}) \to C(\mathbb{R})$ eine Funktion mit $f \mapsto f'$ (Ableitungsoperator für alle einmal stetig differenzierbaren Funktionen). Für diese Funktion d gilt:

(i) d ist <u>nicht</u> injektiv da:
Seien $f_1(x) = x, f_2(x) = x + 1$ zwei Funktionen aus $C^1(\mathbb{R})$. Für diese gilt: $f_1 \neq f_2$ aber $d(f_1) = f_1' = 1 = f_2' = d(f_2)$.

(ii) d ist surjektiv, denn nach dem Hauptsatz der Differential und Integralrechnung gilt: Alle stetigen Funktionen besitzen eine Stammfunktion.

1.7.1 Spezialfälle

(i) f ist linear und $f : \mathbb{R}^n \to \mathbb{R}^m : \quad f(x) = Ax \quad$ mit $A \in \mathbb{R}^{m \times n}$
Im Fall $m = n$ gilt: f ist bijektiv $\Leftrightarrow$ A ist invertierbar.

(ii) Allgemeinerer Fall z.b. $f : [0, \infty) \times [-\pi, \pi) \to \mathbb{R}^2 : \quad f(r, \varphi) = r \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$

f ist surjektiv aber <u>nicht</u> injektiv, denn $f(0, \varphi) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \forall\, \varphi \in [-\pi, \pi)$

1.7.2 Bemerkung

a. Folgende Aussagen sind äquivalent:

(i) A ist invertierbar (oder regulär oder nicht singulär)

(ii) $Ax = b$ besitzt $\forall\, b \in \mathbb{R}^n$ eine eindeutige Lösung $x \in \mathbb{R}^n$

(iii) $\det A \neq 0$

b. Sei $f : \mathbb{R}^n \to \mathbb{R}^m$ mit $x_0 \in \mathbb{R}^n : \quad f(x) = f(x_0) + A(x - x_0) \quad$ mit $\det A \neq 0$
Die Umkehrfunktion f^{-1} kann durch Äquivalenzumformungen gebildet werden:

$$\begin{aligned} & f(x_0) + A(x - x_0) \overset{!}{=} y \\ \Leftrightarrow \quad & A(x - x_0) = y - f(x_0) \\ \Leftrightarrow \quad & x - x_0 = A^{-1}(y - f(x_0)) \\ \Leftrightarrow \quad & x = x_0 + A^{-1}(y - f(x_0)) = f^{-1}(y) \end{aligned}$$

Für die Ableitung der Umkehrfunktion gilt: $\frac{\partial}{\partial y} f^{-1}(y) = A^{-1}$

Kommentar: Differenzierbare Funktionen sind in einer hinreichend kleinen ε-Umgebung im Prinzip linear (nicht ganz korrekt, aber sehr anschaulich).

1.7.3 Satz über die Umkehrfunktion

Sei $f \in C^1(G; \mathbb{R}^n)$ mit $G \subset \mathbb{R}^n$ ein Gebiet, $x_0 \in G$ mit $\det f'(x_0) = \det J_f(x_0) \neq 0$. Dann gilt in einer geeigneten offenen Umgebung U um x_0, dass

(i) $V = f(U)$ ist offen (das Bild V von U ist offen) und $\det f'(x) \neq 0 \quad \forall\, x \in U$

(ii) $f : U \to V$ ist bijektiv, das heißt: $\exists\, f^{-1} : V \to U$ (die Funktion ist lokal invertierbar)

Beweisidee: Da f stetig differenzierbar ist, gilt:

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + r(x)\|x - x_0\| \text{ mit } r(x) \overset{x \to x_0}{\to} 0 \\ &\approx f(x_0) + f'(x_0)(x - x_0) \\ &= f(x_0) + A(x - x_0) \text{ mit } \det A \neq 0 \end{aligned}$$

Wenn $\approx$ ein $=$ wäre so könnte die Umkehrfunktion $f^{-1}(y)$ einfach durch Äquivalenzumformungen bestimmt werden
$\Rightarrow$ In einer hinreichend kleinen ε-Umgebung kann $\approx$ als $=$ angenommen werden.

(iii) $\left(f^{-1}\right)'(y) = \left(f'\left(f^{-1}(y)\right)\right)^{-1} \quad \Leftrightarrow \quad J_{f^{-1}}(y) = J_f^{-1}\left(f^{-1}(y)\right)$

Beweisidee: Wenn $f^{-1} : V \to U$ existiert, dann ist f^{-1} auch stetig diff'bar.
Beispielskizze aus dem Eindimensionalen:

$$\begin{aligned} & f(x) = e^x, \quad f'(x) = e^x, \quad \exists\, f^{-1}(y) = \log y \\ \Rightarrow \quad & y = f\left(f^{-1}(y)\right) \\ \overset{\substack{\text{Ableiten mit} \\ \text{Kettenregel}}}{\Rightarrow} \quad & 1 = f'\left(f^{-1}(y)\right) \cdot (f^{-1})'(y) \\ \Rightarrow \quad & (f^{-1})'(y) = \log' y = \frac{1}{f'\left(f^{-1}(y)\right)} = \frac{1}{e^{\log y}} = \frac{1}{y} \end{aligned}$$

Beispiel übertragen auf den allgemeinen mehrdimensionalen Fall:

$$\begin{aligned} & y = f\left(f^{-1}(y)\right) \\ \overset{\substack{\text{Ableiten mit} \\ \text{Kettenregel}}}{\Rightarrow} \quad & 1 = \underbrace{f'\left(f^{-1}(y)\right)}_{J_{f'}(f^{-1}(y))} \cdot \frac{d}{dy} f^{-1}(y) \\ \Rightarrow \quad & \frac{d}{dy} f^{-1}(y) = \left(f'\left(f^{-1}(y)\right)\right)^{-1} \end{aligned}$$

1.7.4 Polarkoordinaten

...

1.7.5 Beispiel

$f : \mathbb{R}^2 \to \mathbb{R}^2$ mit $f(x,y) = \begin{pmatrix} x - y^2 \\ 2y \end{pmatrix}$

$$\Rightarrow f'(x,y) = J_f(x,y) = \begin{pmatrix} 1 & -2y \\ 0 & 2 \end{pmatrix} \quad \text{mit} \quad \det J_f \neq 0$$

Direktes Bestimmen der Umkehrfunktion (im Allgemeinen will man dies vermeiden):

$$\begin{aligned}
\begin{pmatrix} u \\ v \end{pmatrix} \overset{!}{=} f(x,y) \quad &\Leftrightarrow \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x - y^2 \\ 2y \end{pmatrix} \\
&\Rightarrow \quad v = 2y \quad \Leftrightarrow \quad y = \frac{v}{2} \\
&\Rightarrow \quad u = x - \left(\frac{v}{2}\right)^2 \quad \Leftrightarrow \quad x = u + \frac{v^2}{4} \\
&\Rightarrow \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u + \frac{v^2}{4} \\ \frac{v}{2} \end{pmatrix} = f^{-1}(u,v)
\end{aligned}$$

Damit gilt für die Ableitung der Umkehrfunktion:

$$\left(f^{-1}(u,v)\right)' = J_{f^{-1}}(u,v) = \begin{pmatrix} 1 & \frac{v}{2} \\ 0 & \frac{1}{2} \end{pmatrix}$$

Alternativ mit dem Satz über die Umkehrfunktion:

$$\left(f^{-1}\right)'(u,v) = \left(f'\left(f^{-1}(u,v)\right)\right)^{-1} = J_f^{-1}\left(u + \frac{v^2}{4}, \frac{v}{2}\right) = \begin{pmatrix} 1 & -v \\ 0 & 2 \end{pmatrix}^{-1} \overset{\text{Inv. mit Gauß best.}}{=} \begin{pmatrix} 1 & \frac{v}{2} \\ 0 & \frac{1}{2} \end{pmatrix}$$

1.7.6 Kugelkoordinaten

Sei $f : \mathbb{R}^3 \to \mathbb{R}^3$

$$f(r,\varphi,\theta) = \begin{pmatrix} r\cos\varphi\cos\theta \\ r\sin\varphi\cos\theta \\ r\sin\theta \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

1.7.7 Korollar: Gebietstreue

Ist $G \subset \mathbb{R}^n$ offen und sei $f : G \to \mathbb{R}^n$ eine Funktion mit $f \in C^1(G;\mathbb{R}^n)$, dann ist $f(G)$ offen. Ist G außerdem ein Gebiet, so ist $f(G)$ auch ein Gebiet.
Fazit: Für f^{-1} kann man $\tilde{G} = f(G)$ definieren und $f^{-1} : \tilde{G} \to \mathbb{R}^n$ betrachten. Dabei hat $\tilde{G}$ die gleichen Eigenschaften wie G.

1.7.8 Definition impliziter Funktionen

Sei $g : \mathbb{R}^{m+n} \to \mathbb{R}^m$ mit $n,m \in \mathbb{N}$ eine Funktion und sei

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m \text{ konstant}$$

Man betrachte die Gleichung

$$g(x,y) = \begin{pmatrix} g_1(x_1,...,x_n,y_1,...,y_m) \\ \vdots \\ g_m(x_1,...,x_n,y_1,...,y_m) \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ g_m \end{pmatrix}$$

und sagt

a. die Funktion g ist bei $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ **lokal nach y auflösbar**, wenn eine Funktion f in der Umgebung von $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ existiert mit:

   (i) $f(x_0) = y_0$ und
   (ii) $g(x, f(x)) = b$

b. g ist auf $A \subset \mathbb{R}^n$ **global nach y auflösbar**, wenn eine Funktion $f : A \to \mathbb{R}^m$

   (i) existiert und
   (ii) $g(x, f(x)) \quad \forall\, x \in A$ gilt

c. analog gilt die Definition für auflösbar nach x mit $g(f(y), y)$

1.7.9 Beispiel Einheitskreis

Sei $g : \mathbb{R}^2 \to \mathbb{R}$ (also $m = n = 1$) mit $g(x, y) = x^2 + y^2$ und $b = 1$. Die Funktion g ist auflösbar bei $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$, wenn $x_0 \neq 0$ (denn $y_0^2 = 1$ hat zwei Lösungen) gilt:

$$\begin{aligned} \text{Fall } y_0 > 0: &\quad f_{>0}(x) = y = \sqrt{1 - x^2} \\ \text{Fall } y_0 < 0: &\quad f_{<0}(x) = y = -\sqrt{1 - x^2} \end{aligned}$$

1.7.10 Hauptsatz über implizite Funktionen

Seien $x_0 \in \mathbb{R}^n, y, b \in \mathbb{R}^m$ für eine offene Umgebung G um $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ und sei $g \in C^1(G; \mathbb{R}^m)$ (stetig diff'bar). Gilt weiter $g(x_0, y_0) = b$ und $\det g_y(x_0, y_0) \neq 0$ so existiert eine offene Umgebung U_0 von x_0 und V_0 von y_0, so dass gilt:

(i) $\det g_y(x, y) \neq 0 \quad \forall\, x \in X_0, \forall\, y \in Y_0$

(ii) die Gleichung $g(x, y) = b$ besitzt eine **eindeutig bestimmte Auflösung**:
$f : U_0 \to V_0$ nach y mit $f(x_0) = y_0$ und $g(x, f(x)) = b \quad \forall\, x \in U_0$
außerdem ist diese Funktion f differenzierbar und es gilt:
$f'(x) = -\left(g_y\left(x_0, f(x_0)\right)\right)^{-1} g_x\left(x_0, f(x_0)\right) = -\left(g_y\left(x_0, y_0\right)\right)^{-1} g_x\left(x_0, y_0\right)$

(iii) ist $g \in C^r(G; \mathbb{R}^m)$ mit $r \geqslant 2$ so ist $f \in C^r(U_0; \mathbb{R}^m)$ und die höheren Ableitungen werden durch weiteres ableiten von f' in (ii) bestimmt.

Bemerkung:

$$g : \mathbb{R}^{m+n} \to \mathbb{R}^m \quad \text{mit} \quad x = \begin{pmatrix} x_1 \\ \vdots x_n \end{pmatrix} \in \mathbb{R}^n, \quad y = \begin{pmatrix} y_1 \\ \vdots y_m \end{pmatrix} \in \mathbb{R}^m$$

$$(x, y) \mapsto g(x, y) = \begin{pmatrix} g_1(x, y) \\ \vdots \\ g_m(x, y) \end{pmatrix} = \begin{pmatrix} g_1(x_1, ..., x_n, y_1, ..., y_m) \\ \vdots \\ g_m(x_1, ..., x_n, y_1, ..., y_m) \end{pmatrix}$$

$$\begin{aligned} g' = \frac{\partial}{\partial(x, y)} = J_g &= \begin{pmatrix} \frac{\partial}{\partial x_1} g_1 & \cdots & \frac{\partial}{\partial x_n} g_1 & \frac{\partial}{\partial y_1} g_1 & \cdots & \frac{\partial}{\partial y_m} g_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_1} g_m & \cdots & \frac{\partial}{\partial x_n} g_m & \frac{\partial}{\partial y_1} g_m & \cdots & \frac{\partial}{\partial y_m} g_m \end{pmatrix} \in \mathbb{R}^{m \times (n+m)} \\ &= \begin{pmatrix} \frac{\partial}{\partial x} g & \frac{\partial}{\partial y} g \end{pmatrix} \end{aligned}$$

Beweisidee: Im Spezialfall, dass die Funktion g linear ist gilt: $g(x,y) = A\begin{pmatrix} x \\ y \end{pmatrix}$
Dabei besteht die Matrix $A \in \mathbb{R}^{m\times(m+n)}$ aus zwei Teilmatrizen $X \in \mathbb{R}^{m\times n}, Y \in \mathbb{R}^{m\times m}$ also gilt:

$$g(x,y) = \begin{pmatrix} X & Y \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = Xx + Yy \quad \text{mit} \quad \frac{\partial}{\partial x}g = X, \quad \frac{\partial}{\partial y}g = Y$$

Ist die Funktion g differenzierbar, so ist sie in einer hinreichend kleinen Umgebung näherungsweise linear, also gilt:

$$\begin{aligned} g(x,y) = b \quad &\Leftrightarrow \quad Xx + Yy = b \\ &\Leftrightarrow \quad Yy = b - Xx \overset{\det Y \neq 0}{=} y = Y^{-1}(b - Xx) = Y^{-1}b - Y^{-1}Xx = f(x) \\ &\Rightarrow \quad f'(x) = -Y^{-1}X = -(g_y)^{-1}g_x \end{aligned}$$

Bemerkung: für den Fall $m = n = 1$
Es gilt $g(x, f(x)) = b$ für $x \in U_0$ mit $g : \mathbb{R}^2 \to \mathbb{R}, f : \mathbb{R} \to \mathbb{R}, b \in \mathbb{R}$ konstant
Setze $h(x) = g(x, f(x))$ mit $h : \mathbb{R} \to \mathbb{R}$ also muss $h(x) = b, \forall x \in U_0$ gelten. Da b konstant ist, gilt für die Ableitung $h'(x) = 0$.

$$\begin{aligned} h'(x) &\overset{\substack{\text{Ketten-}\\\text{regel}}}{=} g'(x, f(x))\begin{pmatrix} x \\ f(x) \end{pmatrix}' = \nabla g(x, f(x))^\top \begin{pmatrix} 1 \\ f'(x) \end{pmatrix} \\ &= \begin{pmatrix} g_x(x, f(x)) & g_y(x, f(x)) \end{pmatrix}\begin{pmatrix} 1 \\ f(x) \end{pmatrix} = g_x(x, f(x)) + g_y(x, f(x))f'(x) = 0 \\ &\overset{g_y(x,f(x))\neq 0}{\Rightarrow} f'(x) = -\frac{g_x(x, f(x)}{g_y(x, f(x))} \overset{f(x_0)=y_0}{\Rightarrow} -\frac{g_x(x_0, y_0)}{g_y(x_0, y_0)} = f'(x_0) \end{aligned}$$

1.7.11 Lokale Extremstellen unter Nebenbedingungen

Seien $f, g_1, ..., g_m : G \subset \mathbb{R}^n \to \mathbb{R}$ mit G offen gegeben, sowie $b_1, ..., b_m \in \mathbb{R}$. Dann nennt man $x_0 \in G$ ein lokales Minimum (Maximum) unter den Nebenbedingungen $g_1(x) = b_1, ..., g_m(x) = b_m$ wen es eine offene Umgebung $U \subset G$ von x_0 gibt mit:

$$f(x) \overset{(\leqslant)}{\geqslant} f(x_0) \quad \forall\, x \in U \text{ mit } g(x) = b$$

1.7.12 Satz von Lagrange

Notwendige Bedingung für Extremstellen unter Nebenbedingungen.
Seien $f, g_1, ..., g_m \in C^1(U)$ für eine offene Umgebung U von $x_0 \in \mathbb{R}^n$ und seien $b_1, ..., b_m \in \mathbb{R}$ $(b \in \mathbb{R}^m)$. Ist $x_0 \in U$ eine lokale Extremstelle unter der Nebenbedingung $g_1(x_0) = b_1, ..., g_m(x_0) = b_m$ und sind die Gradienten $\nabla g_1(x_0), ..., \nabla g_m(x_0)$ linear unabhängig also $\det(\nabla g_1(x_0), ..., \nabla g_m(x_0)) \neq 0$, dann existieren die Konstanten $\lambda_1, ..., \lambda_m \in \mathbb{R}$ (Lagrange-Multiplikatoren) mit:

$$\nabla f(x_0) + \underbrace{\lambda_1 \nabla g_1(x_0) + ... + \lambda_m \nabla g_m(x_0)}_{\underbrace{\begin{pmatrix} \nabla g_1(x_0) & \dots & \nabla g_m(x_0) \end{pmatrix}}_{J_g^\top}\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}} = 0$$

Bemerkung: Der Satz von Lagrange enthält nur eine **notwendige** Bedingung: Es kann also Punkte geben mit $\nabla f(x_0) + \lambda \nabla g(x_0) = 0$ die keine Extremstelle sind (z.b. Sattelpunkte). Der Satz liefert also nur Kandidaten, welche dann durch einsetzen in die Funktion f weiter untersucht werden müssen.

Bemerkung: Anschaulich bedeutet die Bedingung $\nabla f(x_0) + \lambda \nabla g(x_0) = 0$, dass die Gradienten beider Funktionen im Punkt x_0 in die gleiche (oder entgegengesetzte) Richtung schauen müssen.

Beweis: für den Fall $n = 2, m = 1$

$f, g : U \subset \mathbb{R}^2 \to \mathbb{R}, b \in \mathbb{R}$ und sei $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ eine lokale Extremstelle unter der Nebenbedingung $g(x_0, y_0) = b$. Außerdem sei $\nabla g(x_0, y_0)$ linear unabhängig (im Fall $n = 2, m = 1$ bedeutet dies: nicht der Nullvektor).

$$
\begin{aligned}
\nabla g(x_0, y_0) &= \begin{pmatrix} g_x(x_0, y_0) \\ g_y(x_0, y_0) \end{pmatrix} \\
&\Rightarrow \quad g_x(x_0, y_0) \neq 0 \text{ oder } g_y(x_0, y_0) \neq 0 \overset{\text{o.B.d.A.}}{\Rightarrow} g_y(x_0, y_)) \neq 0 \\
&\Rightarrow \quad \exists \text{ eine lokale Auflösung } h \text{ nach } y \\
&\qquad h : U_0 \subset \mathbb{R} \to \mathbb{R} \text{ mit } h(x_0) = y_0 \text{ und } g(x, h(x)) = b
\end{aligned}
$$

Sei nun $\hat{f}(x) = f(x, h(x))$ mit $\hat{f} : U_0 \subset \mathbb{R} \to \mathbb{R}$ und betrachte $\hat{f}(x) \to \min / \max$:

$$
\begin{aligned}
0 \overset{!}{=} \hat{f}'(x) = \frac{d}{dx} f(x, h(x)) &\overset{\substack{\text{Ketten-}\\ \text{regel}}}{=} \nabla f(x, h(x))^\top \cdot \begin{pmatrix} 1 \\ h'(x) \end{pmatrix} \\
&= \begin{pmatrix} f_x(x, h(x)) & f_y(x, h(x)) \end{pmatrix} \begin{pmatrix} 1 \\ h'(x) \end{pmatrix} \\
&= f_x(x, h(x)) + f_y(x, h(x)) h'(x)
\end{aligned}
$$

Nach dem Hauptsatz über implizite Funktionen gilt $h'(x) = -\frac{g_x(x,h(x))}{g_y(x,h(x))}$, damit gilt weiter:

$$
0 \overset{!}{=} f_x(x, h(x)) - f_y(x, h(x)) \frac{g_x(x, h(x))}{g_y(x, h(x))}
$$

Bei x_0 gilt ja gerade $h(x_0) = y_0$):

$$
0 \overset{!}{=} f_x(x_0, y_0) - f_y(x_0, y_0) \frac{g_x(x_0, y_0)}{g_y(x_0, y_0)} = f_x(x_0, y_0) g_y(x_0, y_0) - f_y(x_0, y_0) g_x(x_0, y_0)
$$

Also sind die Gradienten $\nabla f, \nabla g$ linear abhängig.

1.7.13 Bespiele

a. $\begin{cases} f(x, y) = x + y \to \min / \max \\ g(x, y) = x^2 + y^2 = b = 1 \end{cases}$

(i) Die Nebenbedingung $g(x, y) = b$ beschreibt eine kompakte (beschränkt und abgeschlossene) Menge $B = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } x^2 + y^2 = 1 \right\}$ und da f stetig ist gilt:

$\overset{\text{Satz von Weierstraß}}{\Rightarrow} \exists\, x_m, x_M \in B$ mit $f(x_m) \leqslant f(x) \leqslant f(x_M) \quad \forall\, x \in B$
(Die Funktion f nimmt auf B ein Minimum und ein Maximum an.)

(ii) Für ein x_0 (also x_m oder x_M) muss gelten, dass:
$\exists\, \lambda \in \mathbb{R}$ mit $\nabla f(x_0) + \lambda \nabla g(x_0) = 0$
Man erhält also eine Menge $\tilde{B} \subset B$ welche die Lagrange-Bedingung erfüllen:

$$\nabla g(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \nabla f(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{aligned}
& \nabla f(x,y) + \lambda \nabla g(x,y) \overset{!}{=} 0 \\
\Leftrightarrow \quad & \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\Leftrightarrow \quad & \left.\begin{array}{l} \lambda \cdot 2x = -1 \\ \lambda \cdot 2y = -1 \end{array}\right\} \begin{array}{l} \text{2 Gleichungen mit 3 Unbekannten} \\ \Rightarrow \text{1 Freiheitsgrad (durch } \lambda \text{ beschrieben)} \end{array}
\end{aligned}$$

Fall $x \neq 0, y \neq 0$ $\quad -\dfrac{1}{2x} = \lambda = -\dfrac{1}{2y} \quad \Rightarrow \quad x = y$

Fall $x = 0, y \neq 0$ $\quad 2 \cdot 0 \cdot \lambda = -1 \Rightarrow 0 = -1 \Rightarrow$ Widerspruch, Fall nicht möglich
$\Rightarrow$ dasselbe gilt für die Fälle $x \neq 0, y = 0$ und $x = 0, y = 0$

$$\overset{\text{Einetzen in die Nebenbedingung}}{\Rightarrow} \quad b = 1 = g(x,y) = x^2 + y^2 \overset{x=y}{\Rightarrow} 1 = x^2 + x^2 \Leftrightarrow x_{1,2} = \pm\frac{1}{\sqrt{2}}$$

$$\Rightarrow \quad \tilde{B} = \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\}$$

$$\overset{\text{Satz von Weierstraß}}{\Rightarrow} \quad x_m, x_M \in \tilde{B}$$

(iii) Prüfen der Elemente aus $\tilde{B}$:

$$f(\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \sqrt{2} \quad \Rightarrow \quad \max$$
$$f(\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} = -\sqrt{2} \quad \Rightarrow \quad \min$$

b. $\left\{ \begin{array}{l} f(x,y) = x^2 + y^2 \to \min / \max \\ g(x,y) = x + y = b = 1 \end{array} \right.$

Die Menge $B = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } x + y = 1 \right\}$ ist nicht kompakt, darum gilt hier der Satz von Weierstraß nicht! In diesem konkreten Fall existiert kein Maximum.