

26. Höhere Ableitungen

Im letzten Kapitel haben wir gesehen, wie man für Abbildungen zwischen mehrdimensionalen Räumen das Konzept der Differenzierbarkeit definieren und für differenzierbare Abbildungen die Ableitung berechnen kann. Analog zum eindimensionalen Fall ist eine der wichtigsten Anwendungen dieser Theorie, (lokale) Extremwerte von Funktionen zu bestimmen. Wir wollen in diesem Kapitel untersuchen, wie dies im Mehrdimensionalen funktioniert. Wie in Satz 11.18 werden wir dazu höhere Ableitungen und die Taylor-Formel benötigen, die wir daher zuerst behandeln müssen.

Um von Extremwerten überhaupt sprechen zu können, müssen wir natürlich Funktionswerte miteinander vergleichen können – was nur geht, wenn die Wertemenge der betrachteten Funktionen die Menge \mathbb{R} der reellen Zahlen ist. „Mehrdimensional“ bedeutet im Zusammenhang mit Extremwerten also nur, dass die *Definitionsmenge* unserer Abbildungen eine Teilmenge von \mathbb{R}^n ist.

Hier ist zunächst einmal die exakte Definition von Extrema, die im Prinzip genauso ist wie im Eindimensionalen in Definition 10.18.

Definition 26.1 (Extrema). Es seien $D \subset \mathbb{R}^n$, $f: D \rightarrow \mathbb{R}$ eine Funktion und $a \in D$. Man sagt, ...

- (a) f habe in a ein **(globales) Maximum**, wenn $f(a) \geq f(x)$ für alle $x \in D$.
- (b) f habe in a ein **lokales Maximum**, wenn es eine Umgebung U von a gibt mit $f(a) \geq f(x)$ für alle $x \in U$. Gilt sogar $f(a) > f(x)$ für alle $x \in U$ mit $x \neq a$, so nennt man das lokale Maximum **isoliert**.

Analog definiert man globale und lokale (isolierte) **Minima**. Hat f in a ein (globales, lokales, isoliertes) Maximum oder Minimum, so sagt man auch, dass f dort ein (globales, lokales, isoliertes) **Extremum** hat.

Einfach zu sehen ist genau wie im Eindimensionalen (siehe Lemma 10.20), dass die (erste) Ableitung einer Funktion ein einfaches *notwendiges* Kriterium für ein lokales Extremum liefert, das nicht auf dem Rand der Definitionsmenge liegt:

Lemma 26.2 (Notwendige Bedingung für Extrema). Es seien $D \subset \mathbb{R}^n$ und $f: D \rightarrow \mathbb{R}$ eine Abbildung. Hat f dann in einem Punkt $a \in \overset{\circ}{D}$ im Inneren der Definitionsmenge ein lokales Extremum und ist f dort differenzierbar, so gilt $f'(a) = 0$.

Wie im Eindimensionalen in Bemerkung 10.21 (a) werden Punkte $a \in \overset{\circ}{D}$ mit $f'(a) = 0$ als **kritische Punkte** bezeichnet.

Beweis. Es sei $i \in \{1, \dots, n\}$. Wegen $a \in \overset{\circ}{D}$ ist die reellwertige Funktion $g: t \mapsto f(a + te_i)$ einer reellen Variablen t in einer Umgebung von 0 definiert, und hat dort natürlich ebenfalls ein lokales Extremum. Nach Lemma 10.20 ist also

$$0 = g'(0) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t} = \partial_i f(a),$$

und damit wie behauptet $f'(a) = (\partial_1 f(a) \mid \dots \mid \partial_n f(a)) = 0$. □

Beispiel 26.3. Wir wollen die (lokalen) Extrema der Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, f(x) = (x_1 + x_2)^2 + e^{x_1^2}$$

finden. Nach Bemerkung 25.19 ist f differenzierbar mit

$$f'(x) = (2(x_1 + x_2) + 2x_1 e^{x_1^2} \quad 2(x_1 + x_2)).$$

Ein notwendiges Kriterium für ein lokales Extremum am Punkt $x \in \mathbb{R}^2$ ist also

$$2(x_1 + x_2) + 2x_1 e^{x_1^2} = 0 \quad \text{und} \quad 2(x_1 + x_2) = 0,$$

was offensichtlich nur für $x_1 = x_2 = 0$, also im Nullpunkt erfüllt ist. In der Tat kann man in diesem Fall einfach sehen, dass im Nullpunkt ein (sogar globales) Minimum vorliegt, denn für alle $x \in \mathbb{R}^2$ ist ja

$$f(x) = (x_1 + x_2)^2 + e^{x_1} \geq 0 + 1 = 1 = f(0).$$

Bemerkung 26.4. Bis hierher funktioniert die Extremwertberechnung also genau wie im Eindimensionalen. Es stellen sich nun natürlich sofort die folgenden Fragen:

- Wie kann man von einem Punkt $a \in \dot{D}$ mit $f'(a) = 0$ möglichst einfach überprüfen, ob dort wirklich ein lokales Extremum vorliegt, d. h. welche *hinreichenden* Kriterien für lokale Extrema gibt es? Wie bereits angekündigt werden wir hierfür höhere Ableitungen und eine mehrdimensionale Taylor-Entwicklung benötigen, die wir in diesem Kapitel untersuchen werden.
- Wie kann man Extrema am *Rand* der Definitionsmenge D finden, wenn D nicht offen ist? Betrachten wir z. B. die Funktion f aus Beispiel 26.3 auf dem abgeschlossenen Einheitskreis $K_1 = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$, so muss f dort nach Folgerung 24.35 irgendwo ein Maximum annehmen – und da wir außer dem Minimum bei 0 keine weiteren Stellen gefunden haben, an denen die Ableitung von f verschwindet, muss dieses Maximum auf dem Rand des Einheitskreises liegen. Im Gegensatz zum Eindimensionalen besteht dieser Rand jetzt aber nicht mehr nur aus zwei Punkten, sondern aus einem in diesem Fall selbst eindimensionalen Objekt, nämlich der Einheitskreislinie. Wir können also nicht mehr einfach alle Randpunkte in f einsetzen, um das Maximum zu finden, sondern brauchen ein geschickteres Verfahren, um Randextrema zu finden. Dies wird der Inhalt von Abschnitt ?? sein.

26.A Die mehrdimensionale Taylor-Entwicklung

Zur Einführung höherer Ableitungen, die wir für die Taylor-Entwicklung benötigen werden, gibt es zwei ganz unterschiedliche Herangehensweisen:

Bemerkung 26.5 (Höhere Ableitungen). Es seien $D \subset \mathbb{R}^n$ offen und $f: D \rightarrow \mathbb{R}^m$ eine Funktion.

- Ist f differenzierbar, so ist die Ableitung von f nach den Bemerkungen 25.4 (c) und 25.6 eine Funktion $f': D \rightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{m \times n}$, also selbst wieder eine Funktion zwischen (offenen Teilmengen von) normierten Räumen – für die wir ja definiert haben, was Differenzierbarkeit bedeutet. Es liegt also eigentlich nichts näher als zu sagen, dass f zweimal differenzierbar heißt, wenn f' selbst wieder differenzierbar ist, und f'' dann als die Ableitung von f' zu definieren. Da f' eine Abbildung von $D \subset \mathbb{R}^n$ nach $\text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ ist, ist f'' nach Bemerkung 25.6 dann eine Abbildung

$$f'': D \rightarrow \text{Hom}(\mathbb{R}^n, \text{Hom}(\mathbb{R}^n, \mathbb{R}^m)).$$

Dies kann man natürlich fortsetzen, und erhält (im Fall der Existenz) z. B. als dritte Ableitung von f eine Abbildung

$$f''': D \rightarrow \text{Hom}(\mathbb{R}^n, \text{Hom}(\mathbb{R}^n, \text{Hom}(\mathbb{R}^n, \mathbb{R}^m))).$$

Auch wenn dies eigentlich die natürlichste Art der Definition höherer Ableitungen im mehrdimensionalen Fall ist, sollte hieraus schon ersichtlich sein, dass es sich dabei allein schon von der Notation her nicht unbedingt um die bequemste Art handelt. Wir werden im Folgenden daher einen anderen Ansatz verfolgen:

- Da die Spalten der Ableitungsmatrix von f gerade die partiellen Ableitungen sind, können wir zur Betrachtung k -facher Ableitungen auch einfach mehrfache partielle Ableitungen der Form $\partial_{i_1} \cdots \partial_{i_k} f$ betrachten. Diese haben im Gegensatz zum Ansatz aus (a) den großen Vorteil, im Fall ihrer Existenz selbst wieder Funktionen von D nach \mathbb{R}^m zu sein – ihr Nachteil besteht lediglich darin, dass es nicht nur eine, sondern n^k solcher k -fachen partiellen Ableitungen gibt, da wir bei jeder der k Differentiationen wieder eine der n Variablen auswählen müssen.

Man kann allerdings zumindest hoffen, dass es bei derartigen mehrfachen partiellen Ableitungen nicht auf die Reihenfolge der Ableitungen ankommt. Der folgende Satz besagt, dass dies in den meisten Fällen auch wirklich so ist.

66

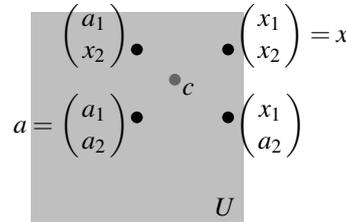
Satz 26.6 (Satz von Schwarz: Vertauschbarkeit partieller Ableitungen). *Es seien $D \subset \mathbb{R}^n$ offen und $f: D \rightarrow \mathbb{R}^m$ eine Abbildung, für die die zweifachen partiellen Ableitungen $\partial_i \partial_j f$ für gewisse $i, j \in \{1, \dots, n\}$ auf D existieren und in einem Punkt $a \in D$ stetig sind. Dann gilt $\partial_i \partial_j f(a) = \partial_j \partial_i f(a)$.*

Beweis. Wir können ohne Einschränkung $m = 1$ annehmen, da man sonst jede Komponentenfunktion von f einzeln betrachten kann. Weiterhin können wir natürlich $i \neq j$ voraussetzen und alle Variablen außer x_i und x_j als konstante Parameter betrachten. Nach Umbenennung der Variablen können wir uns daher auf den Fall $n = 2$, $i = 1$ und $j = 2$ beschränken.

Da D offen ist, können wir eine offene Kugel U um a in der Maximumsnorm wählen, die noch ganz in D liegt. Für einen Punkt $x \in U$ mit $x_1 \neq a_1$ und $x_2 \neq a_2$ betrachten wir nun den „doppelten Differenzenquotienten“

$$Q(x) := \frac{f\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) - f\left(\begin{smallmatrix} a_1 \\ x_2 \end{smallmatrix}\right) - f\left(\begin{smallmatrix} x_1 \\ a_2 \end{smallmatrix}\right) + f\left(\begin{smallmatrix} a_1 \\ a_2 \end{smallmatrix}\right)}{(x_1 - a_1)(x_2 - a_2)},$$

der sich aus den Funktionswerten von f an den vier Eckpunkten des Rechtecks wie im Bild rechts ergibt. Wir wollen zeigen, dass $Q(x)$ für $x \rightarrow a$ sowohl gegen $\partial_1 \partial_2 f(a)$ als auch gegen $\partial_2 \partial_1 f(a)$ konvergiert, woraus dann die Behauptung folgt.



Dazu wenden wir zunächst den Mittelwertsatz 10.23 (a) auf die zwischen a_2 und x_2 definierte reellwertige Funktion

$$g: t \mapsto f\left(\begin{smallmatrix} x_1 \\ t \end{smallmatrix}\right) - f\left(\begin{smallmatrix} a_1 \\ t \end{smallmatrix}\right) \quad \text{mit} \quad g': t \mapsto \partial_2 f\left(\begin{smallmatrix} x_1 \\ t \end{smallmatrix}\right) - \partial_2 f\left(\begin{smallmatrix} a_1 \\ t \end{smallmatrix}\right)$$

an und erhalten ein c_2 zwischen a_2 und x_2 mit

$$Q(x) = \frac{g(x_2) - g(a_2)}{(x_1 - a_1)(x_2 - a_2)} = \frac{g'(c_2)}{x_1 - a_1} = \frac{\partial_2 f\left(\begin{smallmatrix} x_1 \\ c_2 \end{smallmatrix}\right) - \partial_2 f\left(\begin{smallmatrix} a_1 \\ c_2 \end{smallmatrix}\right)}{x_1 - a_1}.$$

Nun wenden wir den Mittelwertsatz erneut auf die zwischen a_1 und x_1 definierte Funktion

$$h: t \mapsto \partial_2 f\left(\begin{smallmatrix} t \\ c_2 \end{smallmatrix}\right) \quad \text{mit} \quad h': t \mapsto \partial_1 \partial_2 f\left(\begin{smallmatrix} t \\ c_2 \end{smallmatrix}\right)$$

an und erhalten so ein c_1 zwischen a_1 und x_1 mit

$$Q(x) = \frac{h(x_1) - h(a_1)}{x_1 - a_1} = h'(c_1) = \partial_1 \partial_2 f(c)$$

für ein $c \in U$ wie im Bild oben rechts, dessen Koordinaten zwischen a_1 und x_1 bzw. a_2 und x_2 liegen.

Mit $x \rightarrow a$ gilt nun auch $c \rightarrow a$, und damit erhalten wir wie behauptet wegen der vorausgesetzten Stetigkeit von $\partial_1 \partial_2 f$

$$\lim_{x \rightarrow a} Q(x) = \lim_{c \rightarrow a} \partial_1 \partial_2 f(c) = \partial_1 \partial_2 f(a).$$

Der Ausdruck $Q(x)$ ist aber symmetrisch in den beiden Koordinaten des Startraums, und daher ergibt sich mit genau dem gleichen Argument (nur indem man den Mittelwertsatz zuerst auf die erste und dann auf die zweite Koordinate anwendet) auch $\lim_{x \rightarrow a} Q(x) = \partial_2 \partial_1 f(a)$. Die Behauptung folgt damit aus der Eindeutigkeit des Grenzwerts. \square

Die Vertauschbarkeit partieller Ableitungen ist also gewährleistet, wenn die partiellen Ableitungen existieren und stetig sind. Dass man auf diese Zusatzforderung der Stetigkeit leider nicht verzichten kann, zeigt die folgende Aufgabe:

Aufgabe 26.7 (Partielle Ableitungen müssen nicht miteinander vertauschen). Zeige, dass die partiellen Ableitungen $\partial_1 \partial_2 f$ und $\partial_2 \partial_1 f$ der Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} x_1 x_2 \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} & \text{für } (x_1, x_2) \neq (0, 0), \\ 0 & \text{für } (x_1, x_2) = (0, 0) \end{cases}$$

zwar existieren, aber nicht übereinstimmen.

Um die Probleme solcher nicht miteinander vertauschbaren Ableitungen zu umgehen, wollen wir uns daher ab jetzt auf Abbildungen beschränken, deren partielle Ableitungen nicht nur existieren, sondern auch stetig sind.

Definition 26.8 (Mehrfach stetig differenzierbare Funktionen). Es sei $f: D \rightarrow \mathbb{R}^m$ eine Abbildung auf einer offenen Teilmenge $D \subset \mathbb{R}^n$. Für $r \geq 0$ sagt man, f sei **r -mal stetig differenzierbar** oder **r -mal stetig partiell differenzierbar**, wenn alle r -fachen partiellen Ableitungen $\partial_{i_1} \cdots \partial_{i_r} f$ für $i_1, \dots, i_r \in \{1, \dots, n\}$ auf D existieren und stetig sind. Man schreibt diese Ableitungen auch als

$$\frac{\partial^r f}{\partial x_{i_1} \cdots \partial x_{i_r}}$$

bzw. bei mehrfachen Ableitungen nach derselben Variablen in „Potenzschreibweise“, d. h. z. B. als

$$\partial_1^2 f \quad \text{oder} \quad \frac{\partial^2 f}{\partial x_1^2}$$

für die zweifache partielle Ableitung nach x_1 . Im Fall $m = 1$ eines eindimensionalen Zielraums schreibt man die Menge aller r -mal stetig differenzierbaren Funktionen auf D wie in Definition 11.7 als $C^r(D)$.

Bemerkung 26.9.

- In Bemerkung 25.20 hatten wir bereits gesehen, dass im Fall $r = 1$ die Begriffe „(einmal) stetig differenzierbar“ und „(einmal) stetig partiell differenzierbar“ zusammenfallen, wenn man sie auf die natürliche Art definiert. Man kann zeigen, dass dieselbe Aussage auch für höhere Ableitungen gilt, wenn man „ r -mal stetig differenzierbar“ als Stetigkeit der r -ten Ableitung wie in Bemerkung 26.5 (a) und „ r -mal stetig partiell differenzierbar“ wie in Definition 26.8 interpretiert. Da wir mit den höheren Ableitungen wie in Bemerkung 26.5 (a) hier nicht arbeiten werden, wollen wir dies hier allerdings nicht beweisen, und erwähnen diese Tatsache nur als Motivation dafür, dass wir die Begriffe „ r -mal stetig differenzierbar“ und „ r -mal stetig partiell differenzierbar“ oben als gleichwertig definiert haben.
- Ist f eine r -mal stetig differenzierbare Funktion wie in Definition 26.8, so sind nicht nur die partiellen Ableitungen der Stufe r , sondern auch die aller Stufen $k \leq r$ stetig: Da die partiellen Ableitungen $\partial_{i_1} \cdots \partial_{i_r} f$ nach Voraussetzung stetig sind, sind die $(r-1)$ -fachen partiellen Ableitungen $\partial_{i_2} \cdots \partial_{i_r} f$ stetig partiell differenzierbar, nach Satz 25.17 und Lemma 25.7 also auch total differenzierbar und damit stetig. Mit Induktion sind demnach dann auch alle niedrigeren partiellen Ableitungen stetig.
- Nach (b) und dem Satz 26.6 von Schwarz kommt es bei bis zu r -fachen partiellen Ableitungen einer r -fach stetig differenzierbaren Funktion nicht auf die Reihenfolge dieser Ableitungen an, da man in einem Ausdruck der Form $\partial_{i_1} \cdots \partial_{i_k} f$ (mit $k \leq r$) durch fortgesetztes Vertauschen zweier benachbarter partieller Ableitungen jede andere Reihenfolge erzeugen kann.

Aufgabe 26.10. Es sei f die in einer Umgebung des Ursprungs von \mathbb{R}^2 definierte reellwertige Funktion mit

$$f(x) = \frac{\cos(x_1^5 + x_2^5)}{1 - x_1^3 - x_2^3}.$$

Berechne mit einer geeigneten Potenzreihenentwicklung (und ohne Computer) die partielle Ableitung $\partial_1^{11} \partial_2^8 f(0)$.

Mit Hilfe dieser höheren Ableitungen können wir nun zur Verallgemeinerung der Taylor-Formel aus Satz 11.13 auf den mehrdimensionalen Fall kommen. Damit das darin auftretende Taylor-Polynom formal analog zum eindimensionalen Fall aussieht, führen wir dazu zunächst einige nützliche Schreibweisen ein.

Notation 26.11 (Multi-Indizes). Ein Element $I = (i_1, \dots, i_n)$ in \mathbb{N}^n nennt man einen **Multi-Index**. Für derartige Multi-Indizes definieren wir die suggestiven Notationen

$$\begin{aligned} |I| &:= i_1 + \dots + i_n, \\ I! &:= i_1! \dots i_n!, \\ \partial^I &:= \partial_1^{i_1} \dots \partial_n^{i_n}, \\ x^I &:= x_1^{i_1} \dots x_n^{i_n} \quad \text{für } x \in \mathbb{R}^n. \end{aligned}$$

Zusätzlich wird im Folgenden die Konstruktion $I + e_j = (i_1, \dots, i_{j-1}, i_j + 1, i_{j+1}, \dots, i_n)$ auftreten, wobei e_j mit $j \in \{1, \dots, n\}$ wie üblich für den j -ten Einheitsvektor steht.

Definition 26.12 (Taylor-Polynom). Es seien $D \subset \mathbb{R}^n$ offen, $r \in \mathbb{N}$, $f: D \rightarrow \mathbb{R}^m$ eine r -mal stetig differenzierbare Funktion, und $a \in D$. Dann heißt

$$T_{f,a}^r: D \rightarrow \mathbb{R}^m, x \mapsto \sum_{I: |I| \leq r} \frac{\partial^I f(a)}{I!} (x-a)^I$$

das **r -te Taylor-Polynom** von f mit Entwicklungspunkt a .

Wir werden dieses Taylor-Polynom nahezu ausschließlich im Fall $m = 1$ eines eindimensionalen Zielraums betrachten – dies ist natürlich auch der für die Untersuchung von Extremwerten relevante Fall. Zunächst einmal können wir dann feststellen, dass es mit unserer alten Definition des eindimensionalen Taylor-Polynoms aus Definition 11.10 (a) übereinstimmt, wenn $n = 1$ ist (und I damit aus nur einem Index besteht). Auch für allgemeine n gibt die Zahl $|I|$ in den Summanden offensichtlich an, wie oft f differenziert wird. Die Terme mit kleinem $|I|$ lassen sich auf die folgende Art mit Hilfe von Matrixprodukten auch anders schreiben:

Definition 26.13 (Hesse-Matrix). Es seien $D \subset \mathbb{R}^n$ offen und $f: D \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion, also $f \in C^2(D)$. Dann heißt für $a \in D$ die (nach Satz 26.6 symmetrische) Matrix der zweiten partiellen Ableitungen

$$Hf(a) := (\partial_i \partial_j f(a))_{i,j} \in \mathbb{R}^{n \times n}$$

die **Hesse-Matrix** von f in a .

Bemerkung 26.14 (Alternative Schreibweise der Taylor-Polynome vom Grad höchstens 2). Für die ersten Terme in den Taylor-Polynomen einer Funktion $f \in C^2(D)$ können die zugehörigen Multi-Indizes explizit angegeben werden:

- (a) Für $|I| = 0$ ist nur $I = (0, \dots, 0)$ möglich; der zugehörige Term im Taylor-Polynom ist einfach der Funktionswert $f(a)$ am Entwicklungspunkt.
- (b) Für $|I| = 1$ ist notwendigerweise $I = e_i = (0, \dots, 0, 1, 0, \dots, 0)$ mit der Eins an einer beliebigen Position $i \in \{1, \dots, n\}$. Die entsprechenden Summanden im Taylor-Polynom lassen sich also schreiben als

$$\sum_{i=1}^n \frac{\partial_i f(a)}{1!} (x_i - a_i) = f'(a) \cdot (x - a).$$

- (c) Für $|I| = 2$ haben die Multi-Indizes die Form $I = e_i + e_j = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$ mit Einsen an Positionen $i < j$ sowie $I = 2e_i = (0, \dots, 0, 2, 0, \dots, 0)$ mit einer 2 an Position

i. Im Taylor-Polynom führt dies zu den Termen

$$\begin{aligned} \sum_{i < j} \frac{\partial_i \partial_j f(a)}{1! \cdot 1!} (x_i - a_i)(x_j - a_j) + \sum_{i=1}^n \frac{\partial_i^2 f(a)}{2!} (x_i - a_i)^2 &= \frac{1}{2} \cdot \sum_{i,j=1}^n \partial_i \partial_j f(a) \cdot (x_i - a_i) \cdot (x_j - a_j) \\ &= \frac{1}{2} (x - a)^\top \cdot Hf(a) \cdot (x - a). \end{aligned}$$

Das zweite Taylor-Polynom einer zweimal stetig differenzierbaren Funktion $f: D \rightarrow \mathbb{R}$ an einem Entwicklungspunkt $a \in D$ ist also

$$T_{f,a}^2(x) = f(a) + f'(a) \cdot (x - a) + \frac{1}{2} (x - a)^\top \cdot Hf(a) \cdot (x - a).$$

Beispiel 26.15. Wir betrachten die Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto x_1 \sin x_2$$

und wollen dazu das zweite Taylor-Polynom um $a = 0$ berechnen. Dazu benötigen wir die ersten und zweiten partiellen Ableitungen

$$\begin{aligned} \partial_1 f(x) &= \sin x_2, \quad \partial_2 f(x) = x_1 \cos x_2 \\ \text{und } \partial_1^2 f(x) &= 0, \quad \partial_1 \partial_2 f(x) = \cos x_2, \quad \partial_2^2 f(x) = -x_1 \sin x_2, \end{aligned}$$

im Nullpunkt also

$$f(0) = \partial_1 f(0) = \partial_2 f(0) = \partial_1^2 f(0) = \partial_2^2 f(0) = 0 \quad \text{und} \quad \partial_1 \partial_2 f(0) = 1.$$

Das gesuchte Taylor-Polynom ist damit

$$\begin{aligned} T_{f,0}^2(x) &= \sum_{i_1+i_2 \leq 2} \frac{\partial_1^{i_1} \partial_2^{i_2} f(0)}{i_1! i_2!} x_1^{i_1} x_2^{i_2} \\ &= \underbrace{f(0)}_{=0} + \underbrace{\frac{\partial_1^1 f(0)}{1!}}_{=0} x_1^1 + \underbrace{\frac{\partial_2^1 f(0)}{1!}}_{=0} x_2^1 + \underbrace{\frac{\partial_1^2 f(0)}{2!}}_{=0} x_1^2 + \underbrace{\frac{\partial_1^1 \partial_2^1 f(0)}{1! 1!}}_{=1} x_1^1 x_2^1 + \underbrace{\frac{\partial_2^2 f(0)}{2!}}_{=0} x_2^2 \\ &= x_1 x_2. \end{aligned}$$

Alternativ ist wie in Bemerkung 26.14

$$f'(0) = (0 \ 0) \quad \text{und} \quad Hf(0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

und damit wieder

$$T_{f,0}^2(x) = 0 + (0 \ 0) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \frac{1}{2} (x_1 \ x_2) \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 x_2.$$

Das entscheidende Resultat über Taylor-Polynome ist nun, dass $T_{f,a}^r$ wie im Eindimensionalen als „beste Näherung“ der Funktion f im Punkt a durch ein Polynom vom Grad r angesehen werden kann. Präzise ausgedrückt ist die Differenz $f(x) - T_{f,a}^r(x)$ wie in Satz 11.13 ein (in der Regel kleines) Restglied, das genauso aussieht wie der nächste Term der Ordnung $r + 1$ der Taylor-Entwicklung, allerdings mit der Ableitung an einer Zwischenstelle berechnet statt am Entwicklungspunkt.

Satz 26.16 (Taylor-Formel). Es seien $D \subset \mathbb{R}^n$ offen und konvex, $r \in \mathbb{N}$ und $f \in C^{r+1}(D)$. Dann gibt es für alle $a, x \in D$ einen Punkt $c \in \overline{ax}$ auf der Verbindungsstrecke von a nach x , so dass

$$f(x) - T_{f,a}^r(x) = \sum_{|I|=r+1} \frac{\partial^I f(c)}{I!} (x - a)^I.$$

Wie in Satz 11.13 bezeichnet man diesen Ausdruck auch als **Restglied** der Taylor-Entwicklung.

Beweis. Wegen der Konvexität von D liegt die Verbindungsstrecke \overline{ax} ganz in D , so dass wir also die Funktion $g: [0, 1] \rightarrow \mathbb{R}$, $t \mapsto f(a + t(x - a))$ betrachten können, die die Einschränkung von f auf diese Strecke beschreibt. Wir werden nun zeigen, dass die eindimensionale Taylor-Formel aus Satz 11.13 angewendet auf g an der Stelle 1 mit Entwicklungspunkt 0 exakt die Behauptung unseres Satzes ist.

Dazu müssen wir die höheren Ableitungen von g berechnen: Wir zeigen für alle $t \in [0, 1]$ mit Induktion über $k = 0, \dots, r + 1$, dass

$$g^{(k)}(t) = k! \cdot \sum_{I:|I|=k} \frac{\partial^I f(a + t(x - a))}{I!} (x - a)^I. \quad (*)$$

Der Induktionsanfang für $k = 0$, also $g^{(0)}(t) = g(t) = f(a + t(x - a))$, ist dabei trivial. Wir nehmen nun an, dass diese Formel für ein $k \in \{0, \dots, r\}$ gilt, und müssen diesen Ausdruck erneut nach t differenzieren, um $g^{(k+1)}$ zu berechnen. Dazu bemerken wir zunächst, dass die Ableitung der Funktion $t \mapsto \partial^I f(a + t(x - a))$ nach der Kettenregel aus Satz 25.30 das Matrixprodukt der Ableitungen von $\partial^I f$ und $t \mapsto a + t(x - a)$ ist, also gleich

$$\sum_{j=1}^n \partial_j \partial^I f(a + t(x - a)) \cdot (x_j - a_j) = \sum_{j=1}^n \partial^{I+e_j} f(a + t(x - a)) \cdot (x_j - a_j).$$

Damit erhalten wir als Ableitung von (*), wobei wir wieder $I = (i_1, \dots, i_n)$ schreiben,

$$\begin{aligned} g^{(k+1)}(t) &= k! \cdot \sum_{j=1}^n \sum_{I:|I|=k} \frac{\partial^{I+e_j} f(a + t(x - a))}{I!} (x - a)^{I+e_j} \\ &= k! \cdot \sum_{j=1}^n \sum_{\substack{I:|I|=k+1 \\ i_j > 0}} \frac{\partial^I f(a + t(x - a))}{(I - e_j)!} (x - a)^I \quad (\text{Indexverschiebung } i_j \rightarrow i_j - 1) \\ &= k! \cdot \sum_{I:|I|=k+1} \left(\underbrace{\sum_{j:i_j > 0} i_j}_{=|I|=k+1} \right) \cdot \frac{\partial^I f(a + t(x - a))}{I!} (x - a)^I \quad (\text{Erweitern mit } i_j) \\ &= (k + 1)! \cdot \sum_{I:|I|=k+1} \frac{\partial^I f(a + t(x - a))}{I!} (x - a)^I, \end{aligned}$$

was die behauptete Formel (*) mit Induktion zeigt. Damit folgt nun mit der eindimensionalen Taylor-Formel für g

$$f(x) - T_{f,a}^r(x) = g(1) - \sum_{k=0}^r \frac{g^{(k)}(0)}{k!} \cdot 1^k = g(1) - T_{g,0}^r(1) \stackrel{11.13}{=} \frac{g^{(r+1)}(t)}{(r+1)!} \cdot 1^{r+1} \quad \text{für ein } t \in (0, 1),$$

was wegen (*) für $c := a + t(x - a)$ genau die Aussage des Satzes ist. \square

Die wichtigsten Fälle in Satz 26.16 sind sicher $r = 0$ und $r = 1$. Während $r = 1$ zur im nächsten Abschnitt betrachteten Extremwertberechnung führt, ergibt sich für $r = 0$ einfach die mehrdimensionale Entsprechung des Mittelwertsatzes 10.23 (a):

Folgerung 26.17 (Mittelwertsatz). *Es seien $D \subset \mathbb{R}^n$ offen und konvex, sowie $f: D \rightarrow \mathbb{R}$ eine stetig differenzierbare Funktion. Dann gibt es für alle $a, x \in D$ eine Zwischenstelle $c \in \overline{ax}$ mit*

$$f(x) - f(a) = f'(c) \cdot (x - a).$$

Beweis. Mit der Schreibweise aus Bemerkung 26.14 (a) und (b) ist die behauptete Gleichung genau die Taylor-Formel aus Satz 26.16 für $r = 0$. \square

Bemerkung 26.18. Bereits am Mittelwertsatz sieht man gut, dass die Voraussetzung einer Funktion mit nur eindimensionaler Wertemenge in Satz 26.16 wesentlich ist: Ist $f: D \rightarrow \mathbb{R}^m$ mit $D \subset \mathbb{R}^n$ eine stetig differenzierbare Abbildung mit Komponentenfunktionen f_1, \dots, f_m und sind $a, x \in D$ wie oben,

so können wir zwar durch Anwendung des gerade gezeigten Mittelwertsatzes auf f_1, \dots, f_m Punkte $c_1, \dots, c_m \in \overline{ax}$ mit

$$f_i(x) - f_i(a) = f'_i(c_i) \cdot (x - a)$$

finden – da die Komponentenfunktionen f_1, \dots, f_m nichts miteinander zu tun haben müssen, werden diese Punkte aber im Allgemeinen verschieden sein, so dass wir nicht erwarten können, ein gemeinsames $c \in \overline{ax}$ mit

$$f(x) - f(a) = f'(c) \cdot (x - a)$$

zu finden. Ganz explizit kann man dies z. B. an der Abbildung

$$f: \mathbb{R} \rightarrow \mathbb{R}^2, x \mapsto \begin{pmatrix} \cos x \\ \sin x \end{pmatrix} \quad \text{mit Ableitung} \quad f'(x) = \begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}$$

sehen: Für $a = 0$ und $x = 2\pi$ ist $f(x) - f(a) = 0$, aber es gibt offensichtlich kein $c \in [0, 2\pi]$ mit $0 = f'(c) \cdot (2\pi - 0)$.

Der Mittelwertsatz (und allgemeiner die Taylor-Formel aus Satz 26.16) gilt also nur für Funktionen nach \mathbb{R} – bzw. bei Funktionen nach \mathbb{R}^m nur für die Komponentenfunktionen einzeln. Für mehrdimensionale Wertebereiche können wir lediglich wie folgt eine dem Mittelwertsatz ähnliche Abschätzung angeben, die wir später noch mehrfach benötigen werden.

Folgerung 26.19. *Es seien $D \subset \mathbb{R}^n$ offen und konvex, sowie $f: D \rightarrow \mathbb{R}^m$ eine stetig differenzierbare Funktion mit Komponentenfunktionen f_1, \dots, f_m . Dann gilt für alle $a, x \in D$*

$$\|f(x) - f(a)\|_\infty \leq \|f'|_{\overline{ax}}\|_\infty \cdot \|x - a\|_\infty,$$

wobei mit $\|f'|_{\overline{ax}}\|_\infty$ analog zu Beispiel 24.40 (b) „das Maximum der Zeilensummennorm von f' auf der Strecke \overline{ax} “ gemeint ist, also

$$\|f'|_{\overline{ax}}\|_\infty := \max\{\|f'(c)\|_\infty : c \in \overline{ax}\} = \max\left\{\sum_{j=1}^n |\partial_j f_i(c)| : i \in \{1, \dots, m\}, c \in \overline{ax}\right\}.$$

Beweis. Beachte zunächst, dass das angegebene Maximum existiert, da die nach Voraussetzung stetigen Funktionen $\sum_{j=1}^n |\partial_j f_i|$ auf der kompakten Menge $\overline{ax} \subset D$ gemäß Folgerung 24.35 ein Maximum annehmen.

Nach dem Mittelwertsatz aus Folgerung 26.17 gibt es nun für alle $i = 1, \dots, m$ einen Punkt $c_i \in \overline{ax}$ mit $f_i(x) - f_i(a) = f'_i(c_i) \cdot (x - a)$, also insbesondere mit

$$\begin{aligned} |f_i(x) - f_i(a)| &= \left| \sum_{j=1}^n \partial_j f_i(c_i) \cdot (x_j - a_j) \right| \leq \sum_{j=1}^n |\partial_j f_i(c_i)| \cdot |x_j - a_j| \leq \sum_{j=1}^n |\partial_j f_i(c_i)| \cdot \|x - a\|_\infty \\ &\leq \|f'|_{\overline{ax}}\|_\infty \cdot \|x - a\|_\infty. \end{aligned}$$

Mit der Definition der Maximumsnorm auf \mathbb{R}^m folgt daraus die behauptete Aussage. \square

26.B Extremwerte

Als Nächstes betrachten wir nun den für Extremwertberechnungen wichtigen Fall $r = 1$ in der Taylor-Formel aus Satz 26.16, in dem das Restglied also aus den zweiten Ableitungen der betrachteten Funktion besteht. Wir hatten in Lemma 26.2 ja schon gesehen, dass an einem lokalen Extremum die Ableitung der betrachteten Funktion notwendigerweise gleich Null sein muss. Wie im eindimensionalen Fall in Satz 11.18 geben die zweiten Ableitungen auch hier nun an einer solchen Stelle oft Auskunft darüber, ob wirklich ein Extremum vorliegt und ob es sich dabei um ein Maximum oder Minimum handelt – und zwar abhängig davon, ob die Hesse-Matrix gemäß Definition 21.9 (b) und Bemerkung 21.10 (c) positiv definit, negativ definit oder indefinit ist.

Satz 26.20 (Extremwertkriterium). *Es seien $D \subset \mathbb{R}^n$ offen und $f: D \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion. Ferner sei $a \in D$ ein kritischer Punkt von f , also $f'(a) = 0$. Ist dann die Hesse-Matrix $Hf(a) \dots$*

(a) *positiv definit, so hat f in a ein isoliertes lokales Minimum.*

- (b) negativ definit, so hat f in a ein isoliertes lokales Maximum.
 (c) indefinit, so hat f in a kein lokales Extremum.

Beweis. Nach eventuellem Verkleinern von D auf eine Kugel mit Mittelpunkt a können wir nach Beispiel 24.22 (c) annehmen, dass D konvex ist. Die Taylor-Formel aus Satz 26.16 für $r = 1$ in der Notation von Bemerkung 26.14 liefert dann zu jedem $x \in D$ ein $c \in \overline{ax} \subset D$ mit

$$f(x) = f(a) + f'(a) \cdot (x-a) + \frac{1}{2} (x-a)^\top \cdot Hf(c) \cdot (x-a).$$

Da nach Voraussetzung $f'(a) = 0$ gilt, müssen wir also untersuchen, ob der Ausdruck

$$f(x) - f(a) = \frac{1}{2} (x-a)^\top \cdot Hf(c) \cdot (x-a) \quad (*)$$

für x in einer Umgebung von a immer positiv ist, immer negativ ist, oder wechselndes Vorzeichen hat.

- (a) Nach dem Hurwitz-Kriterium aus Satz 21.42 ist die Hesse-Matrix $Hf(c)$ für ein $c \in D$ genau dann positiv definit, wenn

$$\det((\partial_i \partial_j f(c))_{i,j=1,\dots,k}) > 0 \quad \text{für alle } k = 1, \dots, n.$$

Da f nach Voraussetzung zweimal stetig differenzierbar ist, sind diese Determinanten nun stetige Funktionen in c , und damit ist die Menge

$$U := \{c \in D : Hf(c) \text{ ist positiv definit}\} = \bigcap_{k=1}^n \left\{ c \in D : \det((\partial_i \partial_j f(c))_{i,j=1,\dots,k}) > 0 \right\}$$

nach Lemma 23.34 (a) und Beispiel 24.18 als endlicher Durchschnitt offener Mengen offen und enthält nach Voraussetzung den Punkt a . Nach evtl. erneutem Verkleinern von D können wir daher $D \subset U$ annehmen.

Dann ist $Hf(c)$ in (*) aber stets positiv definit, und damit $f(x) - f(a) > 0$, d. h. $f(x) > f(a)$ für alle $x \neq a$. Also hat f in a ein isoliertes lokales Minimum.

- (b) folgt genauso wie (a), indem wir die Bedingungen an U gemäß Satz 22.32 (b) so abändern, dass sie der negativen Definitheit entsprechen.
 (c) Wir zeigen zunächst, dass f in a kein lokales Maximum hat.

Da $Hf(a)$ indefinit ist, gibt es einen Vektor $v \in \mathbb{R}^n$ mit $v^\top \cdot Hf(a) \cdot v > 0$. Wie in (a) sind die Einträge der Hesse-Matrix $Hf(c)$ stetige Funktionen in c , und damit ist die Menge

$$U := \{c \in D : v^\top \cdot Hf(c) \cdot v > 0\}$$

nach Beispiel 24.18 eine offene Umgebung von a . Wie oben können wir nach evtl. Verkleinern von D also wieder $D \subset U$ annehmen. Dann ist in (*) aber für alle $t > 0$, so dass $x := a + tv$ (und damit auch c) noch in D liegt,

$$f(x) - f(a) = \frac{1}{2} t^2 v^\top \cdot Hf(c) \cdot v > 0,$$

und damit $f(x) > f(a)$. Also hat f in a kein lokales Maximum.

Natürlich zeigt man mit Hilfe eines Vektors $w \in \mathbb{R}^n$ mit $w^\top \cdot Hf(a) \cdot w < 0$ analog, dass f in a auch kein lokales Minimum hat. \square

Bemerkung 26.21. Beachte, dass das Kriterium aus Satz 26.20 nicht in jedem Fall entscheiden kann, ob an einem kritischen Punkt wirklich ein Extremum vorliegt: Ist die Hesse-Matrix zwar nicht positiv oder negativ definit, aber semidefinit, so trifft keiner der drei Fälle von Satz 26.20 zu. In diesem Fall könnte man nun (analog zum Eindimensionalen in Satz 11.18) höhere als zweite Ableitungen von f betrachten und dafür ähnliche Kriterien beweisen. Da diese höheren Ableitungen im Mehrdimensionalen aufgrund der Vielzahl der partiellen Ableitungen recht kompliziert zu untersuchen sind, werden wir dies hier aber nicht weiter ausführen.

Bemerkung 26.22 (Geometrische Interpretation des Extremwertkriteriums). Die Aussage von Satz 26.20 ist auch leicht anschaulich zu verstehen: Nach Bemerkung 26.14 erhalten wir aus der Taylor-Formel für $r = 2$ um einen Punkt a mit $f'(a) = 0$ ja die Näherung

$$f(x) \approx f(a) + \frac{1}{2} (x-a)^\top \cdot Hf(a) \cdot (x-a).$$

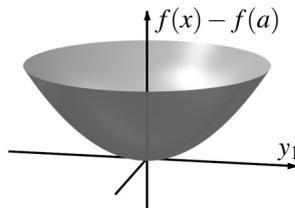
Zu der symmetrischen Hesse-Matrix $Hf(a)$ können wir nun nach dem Trägheitssatz von Sylvester aus Satz 22.35 bzw. Bemerkung 22.36 eine invertierbare Matrix $T \in GL(n, \mathbb{R})$ finden, so dass

$$T^\top \cdot Hf(a) \cdot T = \text{diag}(\underbrace{1, \dots, 1}_k, \underbrace{-1, \dots, -1}_l, 0, \dots, 0) =: D$$

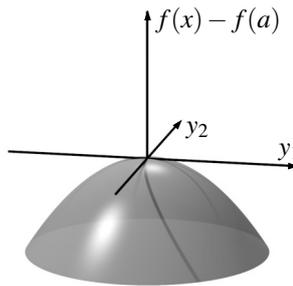
gilt, wobei k und l die Anzahl der positiven bzw. negativen Eigenwerte von $Hf(a)$ ist. Mit der linearen Koordinatentransformation $x = a + Ty$ wird die obige Näherungsformel dann zu

$$f(x) \approx f(a) + \frac{1}{2} y^\top \cdot D \cdot y = f(a) + \frac{1}{2} (y_1^2 + \dots + y_k^2 - y_{k+1}^2 - \dots - y_{k+l}^2).$$

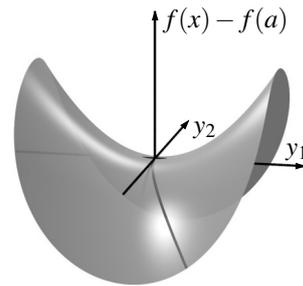
Wie in den Bildern unten im zweidimensionalen Fall dargestellt, kann man hieraus leicht das lokale Verhalten von f um a ablesen: Im positiv definiten Fall $k = n$ ist die Differenz $f(x) - f(a)$ als Summe von Quadraten für $x \neq a$ positiv, und damit liegt dort dann ein lokales Minimum vor. Der Schnitt des Graphen von f mit der y_1 - oder y_2 -Koordinatenachse ist in diesem Fall näherungsweise eine nach oben geöffnete quadratische Parabel. Für eine negativ definite Hesse-Matrix mit $l = n$ haben wir analog ein Maximum.



$Hf(a)$ positiv definit
 $f(x) - f(a) \approx \frac{1}{2} (y_1^2 + y_2^2)$
 Minimum



$Hf(a)$ negativ definit
 $f(x) - f(a) \approx \frac{1}{2} (-y_1^2 - y_2^2)$
 Maximum



$Hf(a)$ indefinit
 $f(x) - f(a) \approx \frac{1}{2} (y_1^2 - y_2^2)$
 Sattelpunkt

Im indefiniten Fall hat f auf den verschiedenen Geraden durch a manchmal ein Maximum und manchmal ein Minimum – man sagt auch, dass in diesem Fall ein **Sattelpunkt** vorliegt. Ist die Hesse-Matrix schließlich weder positiv definit, negativ definit noch indefinit, ist sie also z. B. einfach die Nullmatrix, so sagt die obige Näherung einfach $f(x) \approx f(a)$ für manche oder sogar alle x , und in diesem Fall müsste man sich diese „näherungsweise Gleichheit“ noch genauer anschauen, um entscheiden zu können, ob $f(x)$ nun etwas größer oder etwas kleiner ist als $f(a)$.

68

Beispiel 26.23. Wir wollen die (lokalen) Extrema der Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto x_1^3 + 3x_1^2 + x_2^2$$

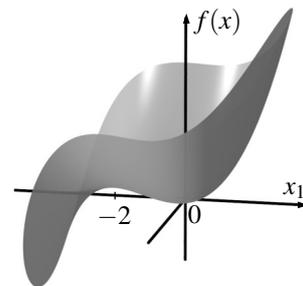
(siehe Bild rechts) berechnen. Dazu bestimmen wir zunächst gemäß Lemma 26.2 die kritischen Punkte als mögliche Kandidaten für Extrema, also die Stellen mit Ableitung Null: Es ist

$$f'(x) = (3x_1^2 + 6x_1 \quad 2x_2),$$

und damit $f'(x) = 0$ genau dann, wenn

$$3x_1(x_1 + 2) = 0 \quad \text{und} \quad 2x_2 = 0,$$

woraus sich die kritischen Punkte $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$ ergeben.



Für unser Extremwertkriterium müssen wir jetzt die Hesse-Matrix berechnen: Es ist

$$Hf(x) = \begin{pmatrix} \partial_1 \partial_1 f(x) & \partial_1 \partial_2 f(x) \\ \partial_2 \partial_1 f(x) & \partial_2 \partial_2 f(x) \end{pmatrix} = \begin{pmatrix} 6x_1 + 6 & 0 \\ 0 & 2 \end{pmatrix}.$$

Im Ursprung ist also

$$Hf \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix},$$

womit Hf dort nach Beispiel 21.15 (a) positiv definit ist. Damit hat f im Ursprung nach Satz 26.20 (a) ein isoliertes lokales Minimum (wie auch im Bild ersichtlich ist). Am anderen kritischen Punkt ist dagegen

$$Hf \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -6 & 0 \\ 0 & 2 \end{pmatrix},$$

und damit ist Hf dort nach Beispiel 21.15 (a) indefinit. Es liegt an dieser Stelle nach Satz 26.20 (c) also kein lokales Extremum, sondern ein Sattelpunkt vor.

Aufgabe 26.24. Bestimme alle lokalen Minima und Maxima der Funktionen

- (a) $f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto x_1^3 + x_2^3 + 3x_1x_2$;
 (b) $g: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto x_2(x_2 - \cos x_1 - 1) + \cos x_1$.

Gib zusätzlich von der Funktion g das zweite Taylor-Polynom mit Entwicklungspunkt $\begin{pmatrix} \pi/2 \\ 0 \end{pmatrix}$ an.

Aufgabe 26.25. Zeige, dass die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto (x_2 - x_1^2)(x_2 - 2x_1^2)$ keine lokalen Extrema hat, dass die Einschränkung von f auf jede Gerade durch den Ursprung aber ein lokales Minimum in 0 besitzt.

Aufgabe 26.26. Für gegebene Punkte $a_1, \dots, a_k \in \mathbb{R}^n$ betrachten wir die Summe der Abstandsquadrate

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \sum_{i=1}^k \|x - a_i\|_2^2.$$

Bestimme alle lokalen und globalen Extrema von f .

Aufgabe 26.27. Es sei $f: \mathbb{R}^n \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion, für die die Hesse-Matrix an jedem Punkt positiv definit ist. Zeige, dass f höchstens ein lokales Extremum besitzt.