Kapitel 4
Differentialrechnung in mehreren Variablen

4.1 Topologie des $\mathbb{R}^n$ und Stetigkeit von Funktionen

Gegenstand dieses Kapitels sind Funktionen in mehreren Variablen. Wir können die Definitionsbereiche solcher Funktionen als Teilmengen eines mehrdimensionalen Raumes auffassen. Deshalb schauen wir uns zunächst den zugrundeliegenden Raum $\mathbb{R}^n$ (für ein $n \in \mathbb{N}$) genauer an, und definieren solche Begriffe wie Nachbarschaft und Konvergenz von Folgen. Identifiziert man Punkte im n-dimensionalen Raum mit ihren Ortsvektoren, liefert die Norm auf $\mathbb{R}^n$ einen Abstandsbegriff, eine sogenannte Metrik. Man definiert nämlich als Abstand zwischen $p, q \in \mathbb{R}^n$:

$$\operatorname{dist}(p, q) := ||p - q|| \, .$$

Die offene Kugel vom Radius r um $p \in \mathbb{R}^n$ ist definiert als

$$K_r(p) := \{q \in \mathbb{R}^n \mid ||p - q|| < r\} \, .$$

Im Fall $n = 1$ ist eine solche Kugel nichts anderes als ein offenes Intervall. Im Fall $n > 1$ gibt es ausserdem viele weitere interessante Umgebungen von Punkten, weil die Vielfalt an möglichen Figuren im mehrdimensionalen natürlich wesentlich grösser ist. Deshalb führt man folgenden Begriff ein:

4.1.1 Definition Eine Teilmenge $U \subset \mathbb{R}^n$ heisst *offen*, falls zu jedem Punkt $p \in U$ ein $\epsilon > 0$ existiert, so dass $K_\epsilon(p) \subset U$. Eine Teilmenge $A \subset \mathbb{R}^n$ heisst *abgeschlossen*, wenn das Komplement der Menge $U := \mathbb{R}^n \setminus A$ in $\mathbb{R}^n$ offen ist.

Zum Beispiel ist jede offene Kugel selbst offen, aber auch jede beliebige Vereinigung von offenen Kugeln. Allgemeiner sind endliche Durchschnitte und beliebige Vereinigungen offener Mengen wieder offen. Die leere Menge und der ganze Raum $\mathbb{R}^n$ sind sowohl offen als auch abgeschlossen. Abgeschlossen sind zum Beispiel auch endliche Punktmengen. Beliebige Durchschnitte und endliche Vereinigungen abgeschlossener Mengen sind wieder abgeschlossen.

Im eindimensionalen Fall sind offene Teilmengen nichts anderes als disjunkte Vereinigungen offener Intervalle. Im zweidimensionalen Fall aber gibt es wesentlich mehr offene Teilmengen als nur die disjunkten Vereinigungen von offenen Kreisscheiben, wie zum Beispiel das Innere von einfach geschlossenen Kurven.

4.1.2 Definition Sei jetzt M eine beliebige Teilmenge des $\mathbb{R}^n$. Ein Punkt $p \in \mathbb{R}^n$ wird als *Randpunkt* von M bezeichnet, wenn jede Kugel $K_r(p)$ um p ($r > 0$ beliebig) sowohl M als auch das Komplement $\mathbb{R}^n \setminus M$ schneidet. Die Gesamtheit aller Randpunkte bilden den *Rand* der Teilmenge M und dieser Rand wird mit ∂M

bezeichnet. Aus der Definition ergibt sich sofort, dass der Rand von M und der Rand des Komplementes von M miteinander übereinstimmen. Weiter definiert man den *Abschluss* von M als

$$\overline{M} := M \cup \partial M\,.$$

Der Abschluss von M ist die kleinste abgeschlossene Teilmenge von $\mathbb{R}^n$, die M enthält.

Zum Beispiel ist der Rand der offenen Kugel $K_r(p)$ um $p \in \mathbb{R}^n$ gerade die Kugeloberfläche

$$\partial K_r(p) = \{q \in \mathbb{R}^n \mid ||p-q|| = r\}\,.$$

Der Abschluss der offenen Kugel ist die sogenannte abgeschlossene Kugel

$$\overline{K_r(p)} = \{q \in \mathbb{R}^n \mid ||p-q|| \leq r\}\,.$$

Der Rand der Kugel und allgemeiner jeder Rand ∂M einer Teilmenge M ist abgeschlossen, denn es handelt sich um den Durchschnitt von zwei abgeschlossenen Mengen $\partial M = \overline{M} \cap \overline{(\mathbb{R}^n \setminus M)}$.

4.1.3 Bemerkung *Eine Teilmenge $A \subset \mathbb{R}^n$ ist genau dann abgeschlossen, wenn sie alle ihre Randpunkte enthält. Eine Teilmenge $U \subset \mathbb{R}^n$ ist genau dann offen, wenn sie keinen ihrer Randpunkte enthält. Eine abgeschlossene Teilmenge kann niederdimensional sein, aber eine offene, nichtleere Teilmenge hat immer die volle Dimension n.*

4.1.4 Definition Man nennt eine offene (oder abgeschlossene) Teilmenge $M \subset \mathbb{R}^n$ *zusammenhängend*, wenn es keine Zerlegung von M in zwei disjunkte offene (bzw. abgeschlossene) nichtleere Teilmengen gibt. Dabei heissen zwei Teilmengen *disjunkt*, wenn ihr Durchschnitt leer ist.

Und hier der letzte wichtige topologische Grundbegriff:

4.1.5 Definition Eine abgeschlossene Teilmenge $A \subset \mathbb{R}^n$ heisst *kompakt*, wenn sie zusätzlich beschränkt ist, das heisst, wenn es eine Schranke $S \in \mathbb{R}$ gibt mit

$$||p|| \leq S \quad \text{für alle } p \in A.$$

Zum Beispiel sind abgeschlossene Kugeln kompakt. Im eindimensionalen Fall sind dies gerade die abgeschlossenen Intervalle. Auch die folgende Menge ist kompakt

$$A = \{0\} \cup \bigcup_{n\in\mathbb{N}} \{p \in \mathbb{R}^n \mid ||p|| = \frac{1}{n}\}\,.$$

Sie besteht aus unendlich vielen Zusammenhangskomponenten.

4.1.6 Definition Man sagt, eine Folge von Punkten $(p_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ konvergiert gegen den Grenzwert $p \in \mathbb{R}^n$ genau dann, wenn die Abstände der Punkte p_k zu p gegen Null konvergieren, das heisst $\lim_{k\to\infty} ||p_k - p|| = 0$. Wie bei Zahlenfolgen, verwendet man in diesem Fall auch die Schreibweise $\lim_{k\to\infty} p_k = p$.

Man kann folgendes zeigen:

4.1.7 Bemerkung *Eine Folge von Punkten p_k in $\mathbb{R}^n$ konvergiert genau dann gegen den Grenzwert $p \in \mathbb{R}^n$, wenn die Folge der j-ten Koordinaten der Punkte p_k gegen die j-te Koordinate von p konvergiert und zwar für $j = 1, \dots, n$.*

Zum Beispiel konvergiert die Folge der Punkte $p_k = (\cos\frac{\varphi}{k}, \sin\frac{\varphi}{k})$ (für $\varphi \in [0, 2\pi]$ fest) auf dem Einheitskreis in $\mathbb{R}^2$ gegen den Punkt $p = (1, 0)$. Und die Folge der Punkte $q_k = (1 - \frac{1}{k})q$ (für $q \in \mathbb{R}^n$ fest) konvergiert gegen q.

Kommen wir nun zum Begriff der stetigen Funktion. Wir betrachten hier nur Funktionen, deren Definitionsbereich entweder offen oder Abschluss einer offenen Menge ist.

4.1.8 Definition Sei $U \subset \mathbb{R}^n$ eine offene oder abgeschlossene Teilmenge. Eine Funktion $f: U \to \mathbb{R}$ heisst *stetig* an der Stelle $p \in U$, wenn für jede Folge (p_n) aus Punkten in U mit $\lim_{n\to\infty} p_n = p$ gilt

$$\lim_{n\to\infty} f(p_n) = f(p)\,.$$

Wir nennen f stetig, wenn f an jeder Stelle des Definitionsbereichs stetig ist.

Das bedeutet, dass der Funktionswert an der Stelle $f(p)$ bereits durch das Verhalten der Funktion in der Nähe von p bestimmt ist. Die folgende Charakterisierung von Stetigkeit lässt sich so verstehen, dass "kleine" Änderungen der Variablen p nur zu "kleinen" Änderungen des Bildes $f(p)$ führen.

4.1.9 Satz *Eine Funktion $f: U \to \mathbb{R}$ auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ ist genau dann stetig in $p \in U$, wenn zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert mit $K_\delta(p) \subset U$ und*

$$|f(p) - f(q)| < \epsilon \quad \text{für alle } q \in K_\delta(p).$$

Für Funktionen mehrerer Variabler gelten entsprechende Rechenregeln für Stetigkeit wie im eindimensionalen Fall, das heisst Summen und Differenzen, sowie Produkte und Zusammensetzungen stetiger Funktionen sind wieder stetig. Zum Beispiel ist also folgende Funktion in drei Variablen stetig:

$$f(x, y, z) = x^2 + 5y^2 z - \exp(\sqrt{x^2 + y^2 + z^2}) \quad (x, y, z \in \mathbb{R})\,.$$

Einige Eigenschaften stetiger Funktionen im Bezug auf die topologischen Begriffe sind im folgenden Satz zusammengefasst:

4.1.10 Satz *Sei f eine stetige Funktion auf der offenen Teilmenge $D \subset \mathbb{R}^n$. Dann gilt folgendes:*

- *Ist $I = (a, b)$ ein offenes Intervall, so ist $f^{-1}(I) = \{p \in D \mid a < f(p) < b\}$ in $\mathbb{R}^n$ ebenfalls offen.*

- *Ist $I = [a, b]$ ein abgeschlossenes Intervall, enthalten in $f(D)$, so ist das Urbild $f^{-1}(I) = \{p \in D \mid a \le f(p) \le b\}$ in $\mathbb{R}^n$ ebenfalls abgeschlossen. Insbesondere ist für jedes $c \in \mathbb{R}$ die Menge $f^{-1}(c) = \{p \in D \mid f(p) = c\}$ abgeschlossen.*

- *Ist $U \subset D$ zusammenhängend, dann ist auch $f(U)$ zusammenhängend.*
- *Ist $K \subset D$ kompakt, dann ist auch $f(K)$ kompakt.*

4.1.11 Beispiel Die Funktion $f(x, y) = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2$ ist stetig. Das Urbild eines Wertes $c > 0$, also die Teilmenge, definiert durch die Gleichung $f(x, y) = c$, ist eine Ellipse und als solche abgeschlossen. Die Teilmenge, definiert durch die Ungleichung $f(x, y) < c$, ist offen und zwar handelt es sich hier um das Innere der Ellipse.

Für stetige Funktionen auf kompakten Teilmengen gelten Sätze, die die Aussagen über stetige Funktionen auf abgeschlossenen Intervallen verallgemeinern.

4.1.12 Satz *Sei $K \subset \mathbb{R}^n$ eine nichtleere kompakte Teilmenge und sei $f : K \to \mathbb{R}$ stetig. Dann nimmt f auf K Maximum und Minimum an, das heisst, es gibt Punkte $p, q \in K$ mit*

$$f(p) \le f(v) \le f(q) \quad \text{für alle } v \in K.$$

Ist $f : U \subset \mathbb{R}^n \to \mathbb{R}$, so versteht man unter dem Graphen von f die folgende Teilmenge des $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$:

$$\mathrm{Graph}(f) := \{(p, f(p)) \mid p \in U\} \subset \mathbb{R}^n \times \mathbb{R}\,.$$

4.1.13 Beispiele

- Ist $f : \mathbb{R}^2 \to \mathbb{R}$ linear, definiert durch $f(x, y) = ax + by$ ($a, b \in \mathbb{R}$ konstant), so ist der zugehörige Graph die Ebene in $\mathbb{R}^3$, definiert durch die Gleichung $z = ax + by$.
- Ist $f(x, y) = \sqrt{x^2 + y^2} = ||(x, y)||$, so ist der Graph ein auf die Spitze gestellter, nach oben offener Kegel.
- Der Graph der Funktion $f(x, y) = x^2 + y^2$ ist ein Paraboloid.
- Der Graph der Funktion $f(x, y) = \dfrac{1}{x^2 + y^2}$ (für $(x, y) \ne (0, 0)$) hat die Gestalt eines Trichters mit nach oben im Unendlichen geöffneter Spitze.
- Der Graph der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, definiert durch $f(x, y) = 1 - (x^2 - 1)^2 - y^2$ ist ein Gebirge im $\mathbb{R}^3$ mit zwei Berggipfeln.
- Die Funktion $f(x, y) = \begin{cases} 1 & \text{für } x \le y \\ 2 & \text{für } x > y \end{cases}$ ist für Punkte mit $x = y$ nicht stetig. Der Graph hat eine Abbruchkante über der Linie $x = y$.
- Die Funktion $f(x, y) = |x - y|$ ist überall stetig, allerdings ist der Graph oberhalb der Linie $x = y$ gefalzt.

Man kann den Verlauf einer Funktion $f : U \subset \mathbb{R}^n \to \mathbb{R}$ auch im Ausgangsraum $\mathbb{R}^n$ graphisch darstellen, indem man dort die *Niveaumengen* einzeichnet. Unter der Niveaumenge zur Zahl $c \in \mathbb{R}$ versteht man die Menge

$$N_c := \{p \in U \mid f(p) = c\}\,.$$

Ist f stetig, so sind alle Niveaumengen von f (wie oben bemerkt) abgeschlossene Teilmengen des $\mathbb{R}^n$. Für $n = 2$ sind die Niveaumengen in der Regel Linien, die sogenannten Höhenlinien. Aber je nach Wahl von c können die Niveaumengen natürlich auch leer sein, nur aus einzelnen Punkten bestehen oder wie im sechsten Beispiel oben zweidimensional sein.

4.2 Partielle Ableitungen

Sei jetzt zunächst $U \subset \mathbb{R}^2$ offen und $f\colon U \to \mathbb{R}$ eine Funktion in 2 Variablen x und y. Wählen wir einen Punkt $p = (x_0, y_0)$ in U aus. Betrachten wir zunächst y als einen festen Parameter, und zwar konstant gleich y_0. Ist jetzt $f(x, y_0)$, aufgefasst als Funktion von x, an der Stelle x_0 nach x differenzierbar, dann wird die entsprechende Ableitung als *partielle Ableitung von f nach x* an der Stelle p bezeichnet. Es gilt:

$$\partial_x f(p) = \lim_{t\to 0} \frac{f(x_0 + t, y_0) - f(x_0, y_0)}{t} .$$

Die partielle Ableitung nach x gibt also die Änderungsrate der Funktion f in bezug auf die Variable x an. Entsprechend ist die *partielle Ableitung nach y* definiert:

$$\partial_y f(p) := \lim_{t\to 0} \frac{f(x_0, y_0 + t) - f(x_0, y_0)}{t} .$$

Hier wird also umgekehrt x als ein Parameter aufgefasst und konstant gleich x_0 gesetzt, während y als variabel gedacht ist. Hier ein erstes Beispiel:

4.2.1 Beispiel Die Funktion $f(x, y) = x^2 \exp(4y)$ hat an der Stelle $p = (x, y)$ die partiellen Ableitungen

$$\partial_x f(x, y) = 2x \exp(4y) \quad \text{und} \quad \partial_y f(x, y) = 4x^2 \exp(4y) .$$

Um diese Ableitungen zu berechnen, fassen wir jeweils eine der beiden Variablen als festen Parameter auf und leiten nach der anderen, als frei gedachten Variablen ab.

Sei nun allgemeiner $U \subset \mathbb{R}^n$ und f eine Funktion in n Variablen. Fixieren wir wieder einen Punkt $p = (p_1, \ldots, p_n)$ in U.

4.2.2 Definition Die *partielle Ableitung* von f bei $p = (p_1, \ldots, p_n)$ nach der Variablen x_j ($j = 1, \ldots, n$) ist folgender Grenzwert (falls er existiert):

$$\partial_{x_j} f(p) := \lim_{t\to 0} \frac{f(p_1, \ldots, p_{j-1}, p_j + t, p_{j+1}, \ldots, p_n) - f(p)}{t} .$$

Es gibt dafür auch die folgenden Schreibweisen:

$$\partial_{x_j} f(p) = \partial_j f(p) = \frac{\partial f}{\partial x_k}(p) .$$

Wenn diese Ableitungen existieren, nennt man f bei p *partiell differenzierbar*.

Um die partielle Ableitung einer Funktion nach einer bestimmten Variablen zu berechnen, kann man alle übrigen Variablen im definierenden Ausdruck als konstant betrachten und dann die bekannten Rechenregeln für Ableitungen von Funktionen einer Variablen anwenden.

4.2.3 Beispiel Sei $f(x,y,z) = x^2 e^{5y} + \sin^2 z$ für $x,y,z \in \mathbb{R}$. Dann sind

$$\partial_x f(x,y,z) = 2xe^{5y}\,, \quad \partial_y f(x,y,z) = 5\,x^2 e^{5y}\,, \quad \partial_z f(x,y,z) = 2\sin(z)\cos(z)\,.$$

Bei Funktionen in zwei Variablen kann man sich die Bedeutung der partiellen Ableitung auch graphisch veranschaulichen. Schauen wir uns dazu ein weiteres Beispiel an.

4.2.4 Beispiel Die Funktion $f(x,y) = x^2 + y^2$ (für $x,y \in \mathbb{R}$) hat die partiellen Ableitungen $\partial_x f(x,y) = 2x$ und $\partial_y f(x,y) = 2y$. Im Punkt $p = (0,1)$ beispielsweise ist $\partial_y f(p) = 2$. Dies ist gerade die Ableitung bei $y = 1$ der Funktion $g(y) = y^2$, die wir aus f erhalten, wenn wir $x = 0$ einsetzen. Der Graph der Funktion g entsteht aus dem Graphen von f (einem Paraboloid), indem wir mit der Ebene, definiert durch $x = 0$, schneiden. Und die partielle Ableitung $\partial_y f(p)$ gibt die Steigung der Tangente an den Graphen von f in p oberhalb der y-Richtung an.

Entsprechend stimmt $\partial_x f(p) = 0$ mit der Ableitung der Funktion $h(x) = x^2 + 1$ bei $x = 0$ überein, die wir aus f erhalten, wenn wir $y = 1$ einsetzen. Der Graph von h ergibt sich aus dem Graphen von f durch Schneiden mit der Ebene, definiert durch $y = 1$. Die Schnittfigur ist eine Parabel mit Scheitelpunkt über p, also ist die Tangentensteigung dort gleich Null.

Hier ein Beispiel einer Funktion, die nicht überall partiell differenzierbar ist.

4.2.5 Beispiel Sei $f(x,y) = |\sin(x) \cdot \sin(y)|$ für $x,y \in \mathbb{R}$. Der Graph von f erinnert an die Oberfläche einer Steppdecke. Im Nullpunkt existieren die partiellen Ableitungen und verschwinden, denn

$$\partial_x f(0,0) = \lim_{t\to 0} \frac{f(t,0) - f(0,0)}{t} = 0 \quad \text{und} \quad \partial_y f(0,0) = \lim_{t\to 0} \frac{f(0,t) - f(0,0)}{t} = 0\,.$$

Aber im Punkt $p = (\frac{\pi}{2}, 0)$ gibt es keine Ableitung nach y. Denn die Funktion $g(t) = f(\frac{\pi}{2}, t) = |\sin(t)|$ (für $t \in \mathbb{R}$) ist bei $t = 0$ nicht differenzierbar, der Graph hat von g hat dort eine Knickstelle, die auch sichtbar wird, wenn wir den Graphen von f mit der Ebene, definiert durch $x = \frac{\pi}{2}$, schneiden.

Die Existenz von partiellen Ableitungen kann von der Wahl des Koordinatensystems abhängen.

4.2.6 Beispiel Sei $f(x,y) = \sqrt{|xy|}$. Hier existieren die partiellen Ableitungen im Nullpunkt $\partial_x f(0,0) = 0$ und $\partial_y f(0,0) = 0$. Aber der Grenzwert

$$\lim_{t\to 0} \frac{f(t,t) - f(0,0)}{t} = \lim_{t\to 0} \frac{\sqrt{t^2}}{t} = \lim_{t\to 0} \frac{|t|}{t}$$

existiert nicht. Dies ist sozusagen die Ableitung in Richtung der Diagonalen $x = y$. Würden wir die Funktion f also in einem um $45°$ gedrehten Koordinatensystem darstellen, dann gäbe es die partiellen Ableitungen im Nullpunkt nicht.

4.3 Lokale Extrema und die Hessesche Form

Sei jetzt wieder $U \subset \mathbb{R}^n$ offen und $f\colon U \to \mathbb{R}$ eine Funktion. Unter einem lokalen Extremum der Funktion f verstehen wir folgendes:

4.3.1 Definition Die Funktion f hat an der Stelle $p \in U$ ein isoliertes lokales Maximum (bzw. Minimum), wenn es ein $\epsilon > 0$ gibt mit $f(q) < f(p)$ (bzw. $f(q) > f(p)$) für alle $p \neq q \in K_\epsilon(p) \subset U$. Man spricht von einem nichtisolierten Maximum bzw. Minimum, wenn statt der strikten Ungleichungen jeweils nur $\leq$ bzw. $\geq$ gelten.

Ein notwendiges Kriterium für lokale Extrema lautet:

4.3.2 Satz *Sei $f\colon U \to \mathbb{R}$ auf U partiell differenzierbar. Hat f an der Stelle $p \in U$ ein lokales Extremum, dann ist $\partial_{x_j} f(p) = 0$ für alle $j = 1, \dots, n$.*

Beweis. Nehmen wir an, f hat bei p ein lokales Maximum. Wählen wir jetzt eine Koordinatenrichtung e_j aus. Dann gilt insbesondere $f(p+te_j) \leq f(p)$ für genügend kleine t. Also hat die Funktion $g(t) = f(p+te_j)$ bei $t = 0$ ein lokales Maximum und daher folgt aus der eindimensionalen Theorie

$$g'(0) = \partial_{x_j} f(p) = \frac{d}{dt} f(p + te_j)|_{t=0} = 0\,.$$

Dies gilt für alle $j = 1, \dots, n$. q.e.d.

Diejenigen Punkte p, bei denen die partiellen Ableitungen verschwinden, sind also Kandidaten für lokale Extrema. Man nennt sie deshalb auch die *kritischen Punkte* von f. Ist p ein kritischer Punkt, in dem weder ein lokales Maximum noch ein lokales Minimum vorliegt, so spricht man von einem *Sattelpunkt*.

Wir fassen jetzt die partiellen Ableitungen von f an der Stelle p zu einem Vektor in $\mathbb{R}^n$ zusammen. Man spricht hier auch vom *Gradienten* von f an der Stelle p und verwendet die folgende Schreibweise:

$$\nabla f(p) = \begin{pmatrix} \partial_{x_1} f(p) \\ \vdots \\ \partial_{x_n} f(p) \end{pmatrix} \quad \text{für } p \in U.$$

4.3.3 Beispiele

1. $f(x, y) = x^2 + y^2$ für $x, y \in \mathbb{R}$. Dann ist $\nabla f(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$. Der Gradient von f verschwindet nur im Nullpunkt, und dort hat f ein isoliertes lokales (und absolutes) Minimum, denn $x^2 + y^2 > 0$ für alle $(x, y) \neq (0, 0)$.

2. $f(x, y) = xy$ für $x, y \in \mathbb{R}$. Hier ist $\nabla f(x, y) = \begin{pmatrix} y \\ x \end{pmatrix}$. Wiederum verschwindet der Gradient nur im Nullpunkt. Dort hat f aber weder ein lokales Minimum noch ein lokales Maximum, sondern einen Sattelpunkt. Denn zu jedem $\epsilon > 0$ finden wir Punkte $p_\epsilon = \frac{1}{2}(\epsilon, \epsilon)$ und $q_\epsilon = \frac{1}{2}(\epsilon, -\epsilon)$ in $K_\epsilon(0)$ mit $f(p_\epsilon) = \frac{1}{4}\epsilon^2 > 0$ und $f(q_\epsilon) = -\frac{1}{4}\epsilon^2 < 0$. In Richtung der Winkelhalbierenden liegt also ein lokales Minimum, in Richtung der Antidiagonalen ein lokales Maximum vor.

3. Ein Sattelpunkt kann auch eine andere Gestalt haben. Die Funktion $f(x,y) = x^3 + y^2$ zum Beispiel hat im Nullpunkt ebenfalls einen Sattelpunkt. Aber hier haben wir in y-Richtung ein lokales Minimum und in x-Richtung einen (eindimensionalen) Sattel, so dass der Graph einem Sessel ähnelt.

4. $f(x,y) = x^3 - \frac{3}{2}x^2 + y^2$ für $x,y \in \mathbb{R}$. Hier ist $\nabla f(x,y) = \begin{pmatrix} 3(x^2-x) \\ 2y \end{pmatrix}$, es gibt also zwei kritische Punkte, nämlich $(0,0)$ und $(1,0)$. Im Nullpunkt liegt ein Sattelpunkt vor (denn in x-Richtung haben wir hier ein lokales Maximum und in y-Richtung ein lokales Minimum). An der Stelle $(1,0)$ befindet sich ein lokales Minimum, denn sowohl in x-Richtung, als auch in y-Richtung ist hier ein lokales Minimum.

5. $f(x,y) = 1 - (x^2-1)^2 - y^2$ für $x,y \in \mathbb{R}$. Dann ist
$$\nabla f(x,y) = \begin{pmatrix} -4(x^2-1)x \\ -2y \end{pmatrix}.$$
Hier gibt es drei kritische Punkte, nämlich $(\pm 1, 0)$ und $(0,0)$. In den Punkten $(\pm 1, 0)$ hat f jeweils ein isoliertes lokales Maximum. Denn offenbar ist $f(x,y) \le 1$, und Gleichheit gilt genau dann, wenn $x^2 = 1$ und $y = 0$ ist. Im Nullpunkt liegt ein Sattelpunkt vor. Denn für $0 < t^2 < 1$ ist einerseits $f(t,0) = 1 - (t^2-1)^2 > 0$ und andererseits $f(0,t) = -t^2 < 0$.

6. $f(x,y) = \cos x$ für $x,y \in \mathbb{R}$. Der Gradient lautet $\nabla f(x,y) = \begin{pmatrix} -\sin x \\ 0 \end{pmatrix}$, kritische Stellen sind also die Punkte $p_k = (k\pi, y)$ ($k \in \mathbb{Z}$, $y \in \mathbb{R}$). Ist k gerade, so liegt bei p_k ein nichtisoliertes Maximum vor. Ist k ungerade, so hat f bei p_k ein nichtisoliertes Minimum.

Mithilfe der zweiten Ableitungen kann man - wie bei Funktionen in einer Variablen - in vielen Fällen entscheiden, ob an einer bestimmten kritischen Stelle ein lokales Maximum, ein lokales Minimum oder ein Sattelpunkt vorliegt. Wir nehmen dazu jetzt an, die ersten partiellen Ableitungen von f seien wiederum partiell differenzierbare Funktionen. Durch nochmaliges partielles Ableiten erhält man die zweiten partiellen Ableitungen an der Stelle $p \in U$:
$$\partial_{x_j}\partial_{x_k} f(p) := \partial_{x_j}(\partial_{x_k} f(q))\Big|_{q=p} \quad \text{und} \quad \partial^2_{x_k} f(p) := \partial_{x_k}(\partial_{x_k} f(q))\Big|_{q=p}.$$
Für die Wahl der Zahlenpaare (j,k) gibt es insgesamt n^2 Möglichkeiten und entsprechend viele zweite partielle Ableitungen, die zu einer quadratischen Matrix zusammengestellt werden.

4.3.4 Definition Die $n \times n$-Matrix
$$H_f(p) := \big(\partial_{x_j}\partial_{x_i} f(p)\big)_{i,j=1,\ldots,n}$$
wird als *Hessesche Matrix* von f bei p bezeichnet.

Nehmen wir jetzt zusätzlich an, dass die zweiten partiellen Ableitungen von f überall stetig sind. Man schreibt dafür $f \in C^2(U)$.

4.3.5 Lemma *Sei $f \in C^2(U)$. Dann gilt $\partial_{x_i}\partial_{x_j} f(p) = \partial_{x_j}\partial_{x_i} f(p)$ für alle i, j. Das heisst, die Matrix $H_f(p)$ ist symmetrisch.*

Auf den Beweis verzichten wir hier. Schauen wir uns nochmals die vorher betrachteten Beispiele an.

4.3.6 Beispiele

1. $f(x,y) = x^2 + y^2$ für $x, y \in \mathbb{R}$. Dann ist $\nabla f(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$ und $H_f(x,y) = \begin{pmatrix} \partial_x^2 f & \partial_y \partial_x f \\ \partial_x \partial_y f & \partial_y^2 f \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

2. $f(x,y) = xy$ für $x, y \in \mathbb{R}$. Hier ist $\nabla f(x,y) = \begin{pmatrix} y \\ x \end{pmatrix}$ und $H_f(x,y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

3. $f(x,y) = 1 - (x^2-1)^2 - y^2$ für $x, y \in \mathbb{R}$. Dann erhalten wir $\nabla f(x,y) = \begin{pmatrix} -4(x^2-1)x \\ -2y \end{pmatrix}$ und $H_f(x,y) = \begin{pmatrix} -12x^2+4 & 0 \\ 0 & -2 \end{pmatrix}$.

Um nun das notwendige Kriterium für lokale Extrema formulieren zu können, brauchen wir die im vorigen Kapitel untersuchten Eigenschaften quadratischer Formen. Erinnern wir an die Begriffe:

4.3.7 Definition Eine symmetrische Matrix A ist positiv definit, wenn all ihre Eigenwerte positiv sind. Sie ist negativ definit, wenn all ihre Eigenwerte negativ sind. Hat die Matrix A sowohl negative als auch positive Eigenwerte, dann nennen wir A indefinit.

Man kann folgendes zeigen:

4.3.8 Bemerkung *Eine symmetrische invertierbare 2×2-Matrix A ist genau dann indefinit, wenn $\det A < 0$ ist. Sie ist positiv (bzw. negativ) definit, wenn $\det A > 0$ und $\mathrm{Spur}(A) > 0$ (bzw. $\mathrm{Spur}(A) < 0$) ist.*

Eine symmetrische invertierbare 3×3-Matrix A ist genau dann positiv definit, wenn $\det A > 0$, $\det B > 0$ und $a_{11} > 0$ ist. Hier bezeichnet B diejenige 2×2-Matrix, die aus A durch Streichen der letzten Zeile und Spalte entsteht.

Hier nun das gewünschte Kriterium:

4.3.9 Satz *Sei $f \in C^2(U)$ für eine offene Teilmenge $U \subset \mathbb{R}^n$. Sei $p \in U$ eine kritische Stelle von f. Dann gilt:*

1. *Ist die Hessesche Matrix $H_f(p)$ positiv definit, so hat f bei p ein isoliertes lokales Minimum.*

2. *Ist die Hessesche Matrix $H_f(p)$ negativ definit, so hat f bei p ein isoliertes lokales Maximum.*

3. *Ist $H_f(p)$ indefinit, so hat f bei p kein lokales Extremum, sondern einen Sattelpunkt.*

Für $n = 1$ ist $\nabla f(a) = (f'(a))$, kritische Stellen sind also gerade die Nullstellen von f'. Ausserdem hat die Hessesche Matrix dann den Typ 1×1 und ist genau dann positiv (bzw. negativ) definit, wenn $f''(a)$ positiv (bzw. negativ) ist. Also verallgemeinert dieser Satz das bekannte Kriterium für Funktionen einer Variablen.

Den Beweis des Satzes werden wir später nachtragen. Wenden wir das Kriterium hier zunächst auf die bereits erwähnten Beispiele an.

4.3.10 Beispiele 1. $f(x, y) = x^2 + y^2$ für $x, y \in \mathbb{R}$. Die Hessesche Matrix im Nullpunkt lautet $H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. Diese Matrix hat den doppelten Eigenwert 2, ist also positiv definit. Deshalb hat f im Nullpunkt ein isoliertes Minimum, wie wir bereits oben direkt gesehen haben.

2. $f(x, y) = xy$ für $x, y \in \mathbb{R}$. Die einzige kritische Stelle ist wiederum der Nullpunkt und $H_f(0,0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Diese Matrix hat die Eigenwerte ± 1, ist also indefinit. Und tatsächlich hat f im Nullpunkt einen Sattelpunkt.

3. $f(x, y) = 1 - (x^2 - 1)^2 - y^2$ für $x, y \in \mathbb{R}$. Die Hessesche Matrix an einer Stelle (x, y) lautet hier $H_f(x, y) = \begin{pmatrix} -12x^2 + 4 & 0 \\ 0 & -2 \end{pmatrix}$. Für den Nullpunkt erhalten wir $H_f(0,0) = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$. Diese Matrix ist indefinit, denn sie hat die Eigenwerte 4 und -2. Also hat f im Nullpunkt einen Sattelpunkt. An den beiden anderen kritischen Stellen haben wir $H_f(\pm 1, 0) = \begin{pmatrix} -8 & 0 \\ 0 & -2 \end{pmatrix}$. Hier ist die Hessesche Matrix negativ definit und deshalb hat f dort jeweils isolierte Maxima, in Übereinstimmung mit dem früheren Ergebnis.

4. Sei jetzt $f(x, y) = -x^3 + xy + y^2$ für $x, y \in \mathbb{R}$. Der Gradient von f lautet $\nabla f(x, y) = \begin{pmatrix} -3x^2 + y \\ x + 2y \end{pmatrix}$. Er verschwindet genau dann, wenn $x = -2y$ und $-12y^2 + y = 0$ sind. Die Funktion f hat also zwei kritische Punkte: $p_1 = (0, 0)$ und $p_2 = (-\frac{1}{6}, \frac{1}{12})$. Die Hessematrix an der Stelle (x, y) lautet:

$$H_f(x, y) = \begin{pmatrix} -6x & 1 \\ 1 & 2 \end{pmatrix}.$$

Für den Nullpunkt erhalten wir

$$H_f(0,0) = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

Diese Matrix hat die Determinante -1, ist also indefinit. Also liegt im Nullpunkt ein Sattelpunkt vor. Für den zweiten kritischen Punkt ist

$$H_f(p_2) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Hier ist die Determinante ist gleich $+1$ und der Eintrag in der oberen linken Ecke ist ebenfalls positiv. Also ist $H_f(p_2)$ positiv definit und an der Stelle p_2 liegt ein Minimum vor.

5. Für $f(x,y) = \cos x$ ist $\nabla f(x,y) = \begin{pmatrix} -\sin x \\ 0 \end{pmatrix}$ und $H_f(x,y) = \begin{pmatrix} -\cos x & 0 \\ 0 & 0 \end{pmatrix}$. An den kritischen Stellen $p_k = (k\pi, 0)$ ist $H_f(k\pi, 0) = \begin{pmatrix} -\cos k\pi & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} (-1)^{k+1} & 0 \\ 0 & 0 \end{pmatrix}$. Die Hessesche Matrix ist hier also weder positiv noch negativ definit, noch indefinit, und über diesen Fall macht der Satz keine Aussage.

4.3.11 Bemerkung *Ist p ein kritischer Punkt von f und ist $H_f(p)$ positiv semidefinit, das heisst, sind sämtliche Eigenwerte von $H_f(p)$ grösser oder gleich Null und ist mindestens ein Eigenwert positiv, dann kann f bei p ein isoliertes oder nichtisoliertes lokales Minimum oder einen Sattelpunkt haben. Aber ein lokales Maximum ist ausgeschlossen.*

4.3.12 Beispiele

- Die Funktion $f(x,y) = x^2 + y^3$ hat im Nullpunkt einen Sattelpunkt und $H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$.
- Die Funktion $f(x,y) = x^2 + y^4$ hat im Nullpunkt ein isoliertes Minimum, und wiederum ist $H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$.

Man kann die beschriebene Methode zur Bestimmung der lokalen Extrema einer Funktion in mehreren Variablen nun einsetzen, um damit mehrdimensionale Optimierungsaufgaben zu lösen. Hier dazu ein Beispiel:

4.3.13 Beispiel Nehmen wir an, zur Verpackung von Speiseeis zu jeweils 1000ml werde eine quaderförmige Schachtel verwendet, und man sucht nun dasjenige Format, bei dem am wenigsten Materialkosten anfallen. Das bedeutet: Man sucht nach demjenigen Quader mit Seitenlängen x, y, z und Gesamtvolumen $xyz = 1000$, der die kleinste Oberfläche hat. Die Oberfläche ist die Summe der 6 Seitenflächen, also

$$g(x,y,z) = 2(xy + yz + xz)\,.$$

Mit der Bedingung $xyz = 1000$ kann man die Variable z eliminieren, indem man $z = \frac{1000}{xy}$ einsetzt. Es ergibt sich eine Funktion in zwei Variablen, deren Minimum im Bereich $x, y > 0$ gesucht wird:

$$f(x,y) = 2(xy + (x+y) \cdot \frac{1000}{xy})\,.$$

Jetzt berechnen wir erst die kritischen Stellen von f:

$$\nabla f(x,y) = \begin{pmatrix} 2y - 2000/x^2 \\ 2x - 2000/y^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

genau dann, wenn $x^2y = 1000 = y^2x$. Weil $x, y > 0$ sind, folgt $x = y$ und $x^3 = 1000$. Es gibt also im Bereich $x, y > 0$ nur einen kritischen Punkt bei $x = y = 10$, und die entsprechende Schachtel ist in diesem Fall ein Würfel der Seitenlänge 10cm. Mit der Hessematrix überprüfen wir jetzt noch, dass es sich wirklich um ein Minimum von f handelt.

$$H_f(x,y) = \begin{pmatrix} 4000/x^3 & 2 \\ 2 & 4000/y^3 \end{pmatrix}, \quad \text{also} \quad H_f(10,10) = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}.$$

Die Hessematrix beim kritischen Punkt ist positiv definit, f hat also dort tatsächlich ein isoliertes Minimum.