9. Reelle Funktionen in mehreren Variablen

Kapitel 3 führte Funktionen allgemein ein. Kapitel 5 diskutierte dann reelle Funktionen, die eine reelle Zahl auf eine andere reelle Zahl abbilden. Kapitel 8 behandelte lineare Funktionen, die einen Vektor auf einen anderen Vektor abbilden. In diesem Kapitel werden wir nun Funktionen betrachten, die einen Vektor auf eine reelle Zahl abbilden.

9.1 Grundlagen

In den Wirtschaftswissenschaften gibt es nur wenige Grössen, welche nur von einer Einflussgrösse abhängen. Wenn C den Konsum einer ganzen Volkswirtschaft darstellt, hängt dieser unter anderem von den Preisen verschiedener Konsumgüter und den Einkommen der verschiedenen Haushalte ab. Die Nachfrage nach einem Produkt hängt in der Regel unter anderem vom Preis des Produkts, dem Preisindex (gemittelt über alle Waren) und der Qualität des Produkts ab.

Funktionen, welche einen Vektor auf eine reelle Zahl abbilden, werden reellwertige Funktionen in n Variablen oder reelle Funktionen in n Variablen genannt. Im Fliesstext werden wir diese Funktionen auch manchmal als reelle Funktionen in mehreren Variablen bezeichnen. Die meisten Definitionen und Ergebnisse in diesem Abschnitt stellen Erweiterungen der in Kapitel 5 diskutierten Konzepte dar.

9.1.1 Definition

Ziele dieses Unterkapitels
- Was ist eine reelle Funktion in n Variablen?
- Was versteht man unter einer affin-linearen, einer quadratischen, einer Cobb Douglas und einer Leontief Funktion in n Variablen?

Definition 9.1.1 — Reelle Funktion in n Variablen.
Seien $D \subseteq \mathbb{R}^n$, $Z \subseteq \mathbb{R}$. Dann heisst eine Funktion

$$f : D \to Z \ \text{ mit } \ \mathbf{x} \mapsto f(\mathbf{x})$$

reellwertige Funktion in n (unabhängigen reellen) Variablen oder kurz reelle Funktion in n Variablen.

$f(D)$ heisst auch im Fall einer reellen Funktion in n Variablen Bild von D unter f oder Wertebereich von f.

Setzt man in die Abbildungsvorschrift einer reellen Funktion in n Variablen einen Spaltenvektor $\mathbf{x}$ ein, ergeben sich wie schon bei den linearen Abbildungen doppelte Klammern. Beispielsweise ergibt Einsetzen von $\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ in die Funktion f den Ausdruck $f(\begin{pmatrix} 1 \\ 3 \end{pmatrix})$. Um dies zu vermeiden, schreibt man auch $f\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ oder $f(1,3)$.[1]

Wie schon im Fall einer reellen Funktion in einer Variablen kann man reelle Funktionen in n Variablen nur durch ihre Abbildungsvorschrift charakterisieren. Werden der Definitionsbereich und die Zielmenge nicht spezifiziert, ist hier stets der natürliche Definitionsbereich, d.h. die Menge aller $\mathbf{x} \in \mathbb{R}^n$, für welche die Abbildungsvorschrift definiert ist, und der zugehörige Wertebereich als Zielmenge anzunehmen.[2] Um klarzustellen, wie viele Komponenten das Argument hat, ist es in der Regel jedoch sehr hilfreich, den Definitionsbereich explizit anzugeben.

■ Beispiel 9.1.1 — Reelle Funktionen in n Variablen.
Modelliert man einen Konsumenten, der zwei Produkte P1 und P2 als Komplemente ansieht, wobei jeweils eine Mengeneinheit von P1 gemeinsam mit zwei Mengeneinheiten von P2 konsumiert wird, kann man den Nutzen aus x_1 Einheiten von P1 und x_2 Einheiten von P2 als Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = \min\left\{x_1, \frac{1}{2}x_2\right\}$$

modellieren. Ebenso kann diese Funktion die maximale Produktionsmenge wiedergeben, wenn für die Produktion einer Einheit jeweils eine Mengeneinheit von P1 und zwei Mengeneinheiten von P2 benötigt werden und x_1 Einheiten von P1 und x_2 Einheiten von P2 zur Verfügung stehen. Funktionen dieses Typs, bei denen das Minimum von Vielfachen mehrerer Variablen gebildet wird, nennt man auch Leontief Funktionen.

Auch die Funktion $f : (0, +\infty)^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = \max\{x_1x_2 - 1, 0\}$$

ist eine reelle Funktion in 2 Variablen, aber keine Leontief Funktion.

Aus der Volkswirtschaftslehre kennen Sie sicher auch Cobb Douglas Nutzenfunktionen wie z.B. $f : (0, +\infty)^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^{0.8}x_2^{0.2}.$$

[1] Derartige Schreibweisen werden auch oft im Kontext linearer Abbildungen verwendet. Wir haben in Kapitel 8 bewusst darauf verzichtet, um zu verdeutlichen, dass die Argumente $\mathbf{x}$ Spaltenvektoren darstellen.

[2] Der Definitionsbereich ist dabei stets als eine Teilmenge des $\mathbb{R}^n$ zu wählen, wobei n der Anzahl der verwendeten Variablen entspricht.

Auch diese Funktion ist eine reelle Funktion in 2 Variablen. ■

■ Beispiel 9.1.2 — Lineare und quadratische Funktionen in n Variablen.
In Kapitel 8 behandelten wir lineare Abbildungen, welche Vektoren aus dem $\mathbb{R}^n$ auf Vektoren des $\mathbb{R}^m$ abbilden. Im Fall $m = 1$ ist eine lineare Abbildung eine reelle Funktion in n Variablen. Zum Beispiel beschreiben die Abbildungsvorschriften

$$f(\mathbf{x}) = (-8,3)\mathbf{x} = -8x_1 + 3x_2$$

oder

$$f(\mathbf{x}) = (7,9,8)\mathbf{x} = 7x_1 + 9x_2 + 8x_3$$

reelle Funktionen in 2 bzw. 3 Variablen.

Der Graph der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = -8x_1 + 3x_2 + 10$$

entsteht durch eine Verschiebung des Graphen einer linearen Abbildung mit Abbildungsmatrix $(-8,3)$ um 10 nach oben. Da der Graph von f einen affinen Raum im $\mathbb{R}^3$ beschreibt, nennt man eine solche Funktion auch affin-lineare Funktion in 2 Variablen.

Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$$

hingegen ist keine affin-lineare Funktion, weil sie quadratische und gemischte Terme enthält. Man nennt eine solche Funktion auch eine quadratische Funktion in 2 Variablen. Ebenso sind die Funktionen $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1x_2$$

oder

$$f(\mathbf{x}) = x_1^2 + 4(x_2-1)^2 + 3 = x_1^2 + 4x_2^2 - 8x_2 + 7$$

quadratische Funktionen in 2 Variablen.

Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(\mathbf{x}) = x_1x_2^2 + 2x_2 - 5$$

hingegen ist keine quadratische Funktion in 2 Variablen.[3] ■

Ebenso wie im Fall einer Funktion in einer Variablen benennt man besonders wichtige reelle Funktionen in mehreren Variablen. Wir beschränken uns hier auf die Definition von vier Funktionstypen: affin-lineare und quadratische Funktionen, also (auf n Variablen verallgemeinerte) Polynome vom Grad 1 und 2, sowie die Cobb Douglas und die Leontief Funktion.

Definition 9.1.2 — Besondere Funktionen in n Variablen.
Sei $n \in \mathbb{N}$. Eine reelle Funktion f in n Variablen mit Abbildungsvorschrift

- $f(\mathbf{x}) = \mathbf{b}^T\mathbf{x} + c = \sum_{i=1}^n b_ix_i + c = b_1x_1 + b_2x_2 + \cdots + b_nx_n + c$ für $\mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$ heisst affin-lineare Funktion in n Variablen;
- $f(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j + \sum_{i=1}^n b_ix_i + c$ für $\mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$ und Matrix A der Ordnung n, welche nicht der Nullmatrix entspricht, heisst quadratische Funktion in n Variablen;

[3] Es handelt sich um ein Polynom vom Grad 3.

- $f(\mathbf{x}) = \prod_{i=1}^{n} x_i^{b_i} = x_1^{b_1} \cdot \dots \cdot x_n^{b_n}$ mit $\mathbf{x} \in (0, +\infty)^n$ und $\mathbf{b} \in \mathbb{R}^n$, mit $b_1 + \dots + b_n = 1$ und $b_i \geq 0$ für $i = 1, \dots, n$, heisst Cobb Douglas Funktion in n Variablen;
- $f(\mathbf{x}) = b_0 \min\limits_{i=1,\dots,n} \{b_i x_i\} = b_0 \min\{b_1 x_1, \dots, b_n x_n\}$ für $b_0 \in \mathbb{R}, \mathbf{b} \in \mathbb{R}^n$ heisst Leontief Funktion in n Variablen.

■ Beispiel 9.1.3 — Fortsetzung von Beispiel 9.1.2.
Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit Abbildungsvorschrift $f(\mathbf{x}) = (-8,3)\mathbf{x} = -8x_1 + 3x_2$ ist eine affin-lineare Funktion in 2 Variablen mit $\mathbf{b} = (-8,3)^T$ und $c = 0$. Ebenso ist die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = -8x_1 + 3x_2 + 10$$

eine affin-lineare Funktion in 2 Variablen mit $\mathbf{b} = (-8,3)^T$ und $c = 10$.

Die Funktion $f : \mathbb{R}^3 \to \mathbb{R}$ mit $f(\mathbf{x}) = (7,9,8)\mathbf{x} = 7x_1 + 9x_2 + 8x_3$ ist eine affin-lineare Funktion in 3 Variablen mit $\mathbf{b} = (7,9,8)^T$ und $c = 0$.

Für $A = \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{3}{2} & 1 \end{pmatrix}$, $\mathbf{b} = \mathbf{0}$ und $c = 0$ ist $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{3}{2} & 1 \end{pmatrix} \mathbf{x} = \mathbf{x}^T \begin{pmatrix} x_1 - \frac{3}{2}x_2 \\ -\frac{3}{2}x_1 + x_2 \end{pmatrix} = x_1^2 + x_2^2 - 3x_1x_2$$

eine quadratische Funktion in 2 Variablen. Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1 x_2$$

ist eine quadratische Funktion in 2 Variablen mit $A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$, $\mathbf{b} = \mathbf{0}$ und $c = 0$.

Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3 = x_1^2 + 4x_2^2 - 8x_2 + 7$$

ist eine quadratische Funktion in 2 Variablen mit $A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$, $\mathbf{b} = (0,-8)^T$ und $c = 7$. ■

Z Eine reelle Funktion in n Variablen bildet reelle Vektoren $\mathbf{x} \in D \subseteq \mathbb{R}^n$ auf reelle Zahlen $f(\mathbf{x}) \in Z \subseteq \mathbb{R}$ ab.

Sei A eine symmetrische Matrix der Ordnung n, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$. Dann hat eine affin-lineare reelle Funktion in n Variablen eine Abbildungsvorschrift der Form $f(\mathbf{x}) = \mathbf{b}^T\mathbf{x} + c$; eine quadratische Funktion in n Variablen eine Abbildungsvorschrift der Form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$; eine Cobb Douglas Funktion in n Variablen eine Abbildungsvorschrift der Form $f(\mathbf{x}) = \prod_{i=1}^{n} x_i^{b_i}$ mit $\mathbf{b} \in \mathbb{R}^n$, $b_1 + \dots + b_n = 1$ und $b_i \geq 0$ für $i = 1, \dots, n$, und eine Leontief Funktion in n Variablen eine Abbildungsvorschrift der Form $f(\mathbf{x}) = \min_{i=1,\dots,n}\{b_i x_i\}$.

9.1.2 Geometrische Darstellung

Ziele dieses Unterkapitels

- Was versteht man unter der Höhenline einer reellen Funktion f in n Variablen zum Niveau y?
- Was ist ein Vertikalschnitt von f durch $\mathbf{x}^0 \in \mathbb{R}^n$ in Richtung $\mathbf{r} \in \mathbb{R}^n$?

Der Graph einer reellen Funktion f in n Variablen ist die Menge aller Tupel $(\mathbf{x}, f(\mathbf{x}))$ mit $\mathbf{x} \in D \subseteq \mathbb{R}^n$. Bei reellen Funktionen in einer Variablen entspricht der Vektor $\mathbf{x}$ der Zahl x_1. Ein solcher Graph wird in der Regel in einem Koordinatensystem visualisiert, in dem die Variable x_1 horizontal und der Wert $y = f(\mathbf{x})$ vertikal abgetragen ist. Bei einer reellen Funktion in $n \geq 2$ Variablen ist eine visuelle Darstellung schwieriger. Im Fall $n = 2$ Variablen kann man den Graphen der Funktion als ein Gebirge in einem 3-dimensionalen Koordinatensystem darstellen.

■ Beispiel 9.1.4 — Darstellung reeller Funktionen in einem Koordinatensystem. Abbildungen 9.1, 9.2, 9.3, 9.4, 9.5 und 9.6 stellen Graphen der Funktionen in zwei Variablen aus Beispielen 9.1.1 und 9.1.2 dar.

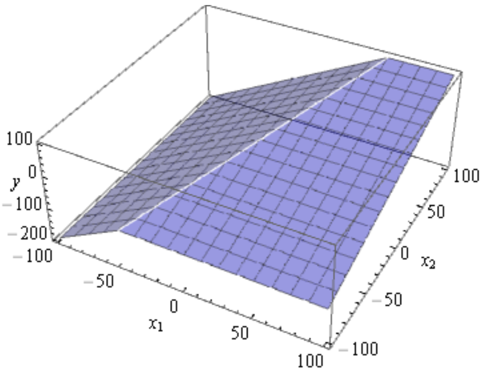

Abbildung 9.1: Der Graph von $f(\mathbf{x}) = \min\{x_1, \frac{1}{2}x_2\}$.

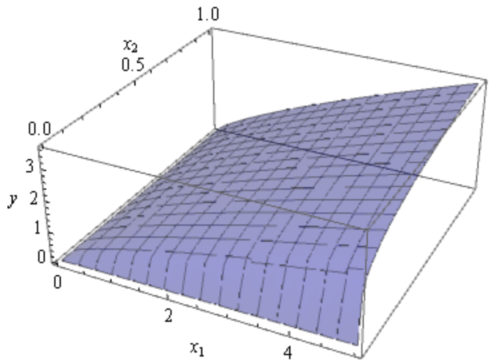

Abbildung 9.2: Der Graph der Cobb Douglas Funktion.

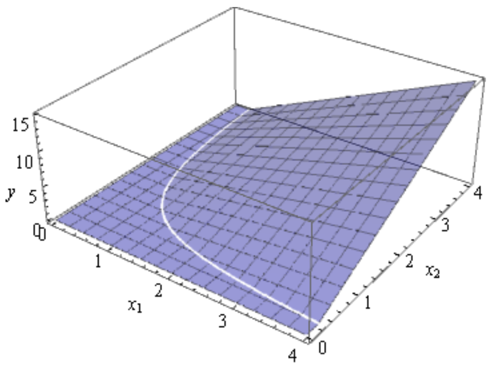

Abbildung 9.3: Der Graph von $f(\mathbf{x}) = \max\{x_1x_2 - 1, 0\}$.

Die Funktion $f(\mathbf{x}) = 7x_1 + 9x_2 + 8x_3$ aus Beispiel 9.1.2 ist eine reelle Funktion in drei Variablen. Ihr Graph besteht daher aus 4-Tupeln $(x_1, x_2, x_3, f(\mathbf{x}))$, die man nicht einfach als Gebirge visualisieren kann. ■

Je nach Form des Graphen und der dargestellten Perspektive können Darstellungen in einem dreidimensionalen Koordinatensystem mehr oder weniger aussagekräftig sein. Höhenlinien sind eine andere Art, die Informationen des dreidimensionalen Graphen in nur zwei Dimensionen darzustellen. Hierzu wird die y-Achse, welche die Funktionswerte reprä-

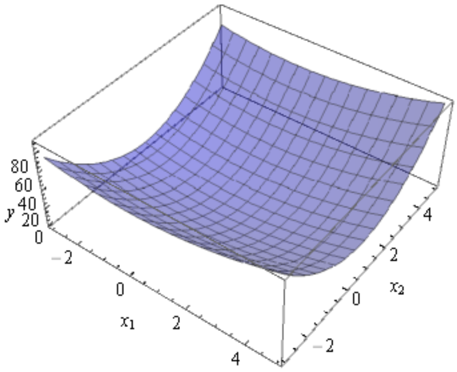

Abbildung 9.4: Der Graph von $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$.

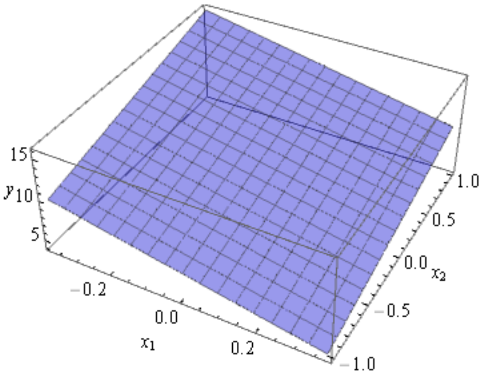

Abbildung 9.5: Der Graph der affin-linearen Funktion $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$.

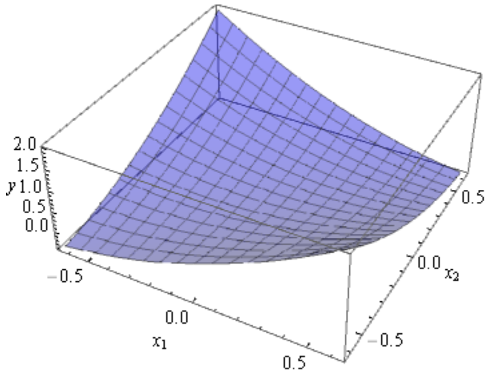

Abbildung 9.6: Der Graph der quadratischen Funktion $f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$.

sentiert, nicht gezeichnet. Eine Höhenlinie verbindet alle Punkte des Definitionsbereiches, welche auf den gleichen Funktionswert $y = f(\mathbf{x})$ abgebildet werden.[4]

Definition 9.1.3 — Höhenlinien.
Für eine reelle Funktion in n Variablen $f : D \to Z$ mit $D \subseteq \mathbb{R}^n$, und Funktionswert $y \in Z \subseteq \mathbb{R}$ heisst die Menge

$$N_y = \{\mathbf{x} \in D \mid f(\mathbf{x}) = y\}$$

Höhenlinie, Niveaumenge, Niveaulinie oder Isoquante zur Höhe bzw. zum Niveau y.

Dabei kann die Bezeichnung als Linie irreführend sein, da N_y nicht unbedingt eine Linie sein muss, sondern mehrere Linien oder ganze Flächen umfassen kann. Einige Graphikprogramme nutzen unterschiedliche Farben, um unterschiedliche Niveaus zu kennzeichnen (z.B. blau für kleine und gelb für grosse Werte).

■ **Beispiel 9.1.5 — Höhenlinien reeller Funktionen in 2 Variablen.**
Abbildungen 9.7, 9.8, 9.9, 9.10, 9.11 und 9.12 zeigen verschiedene Höhenlinien der Funktionen aus Beispielen 9.1.1 und 9.1.2. ■

Die Höhenlinie zur Höhe y kann man sich als horizontalen Schnitt durch den Graphen

[4] Derartige Höhenlinien von Gebirgen werden u.a. auch auf Wanderkarten verwendet.

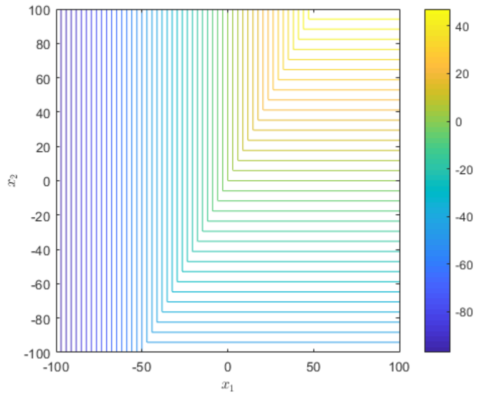

Abbildung 9.7: Höhenlinien der Funktion $f(\mathbf{x}) = \min\{x_1, \frac{1}{2}x_2\}$.

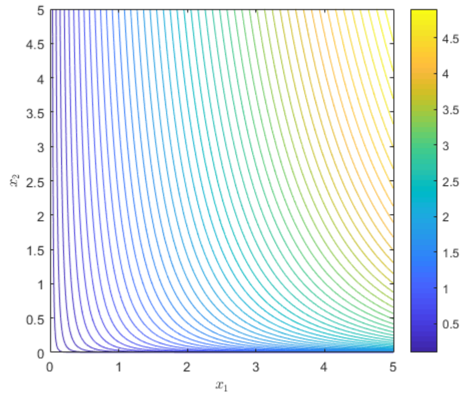

Abbildung 9.8: Höhenlinien der Cobb Douglas Funktion.

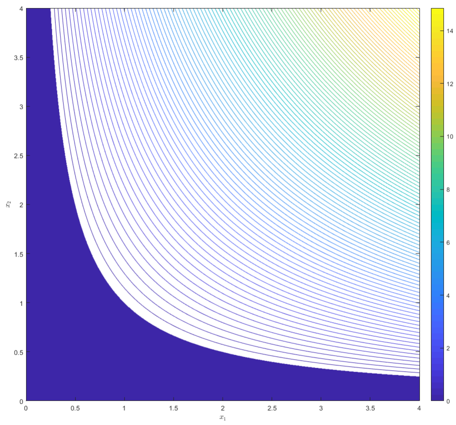

Abbildung 9.9: Höhenlinien der Funktion $f(\mathbf{x}) = \max\{x_1x_2 - 1, 0\}$.

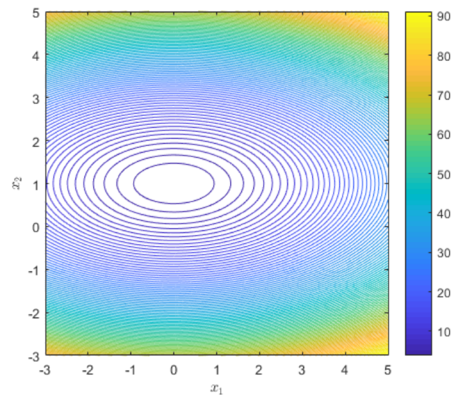

Abbildung 9.10: Höhenlinien der Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$.

von f auf der Höhe y vorstellen. Schneidet man durch den Punkt $(\mathbf{x}^0, f(\mathbf{x}^0))$ vertikal durch den Graphen von f, so entstehen sogenannte Vertikalschnitte. Der Einheitsvektor $\mathbf{r}$ mit $||\mathbf{r}|| = 1$ gibt dabei an, in welche Richtung der Schnitt durch den Punkt $(\mathbf{x}^0, f(\mathbf{x}^0))$ verläuft. Man bezeichnet den Einheitsvektor $\mathbf{r}$ in diesem Kontext daher als Richtung.

Definition 9.1.4 — Vertikalschnitte.
Sei $f : D \to Z$ eine reelle Funktion in n Variablen. Für einen Punkt $\mathbf{x}^0 \in \mathbb{R}^n$ und eine Richtung $\mathbf{r} \in \mathbb{R}^n$ heisst die Funktion

$$f_{\mathbf{x}^0,\mathbf{r}} : \{t \in \mathbb{R} \mid \mathbf{x}^0 + t\mathbf{r} \in D\} \to \mathbb{R} \text{ mit } f_{\mathbf{x}^0,\mathbf{r}}(t) = f(\mathbf{x}^0 + t\mathbf{r})$$

Vertikalschnitt von f durch bzw. ausgehend von $\mathbf{x}^0$ in Richtung $\mathbf{r}$. Die Variable des Vertikalschnitts t nennt man Schrittweite. Ist $\mathbf{r} = \mathbf{e}^i$, für $i = 1, \ldots, n$, spricht man auch von senkrechten Vertikalschnitten.

Der senkrechte Vertikalschnitt einer reellen Funktion in zwei Variablen in Richtung $\mathbf{r} = \mathbf{e}^1$

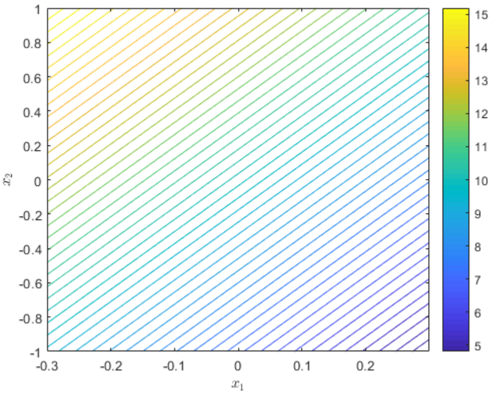

Abbildung 9.11: Höhenlinien der Funktion $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$.

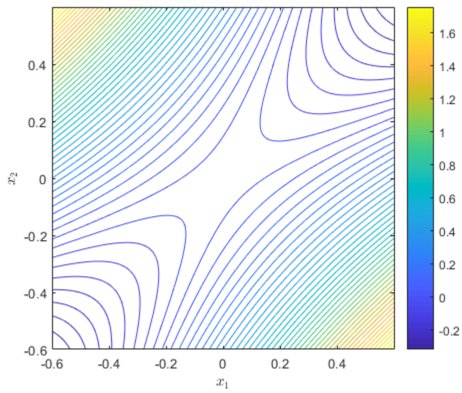

Abbildung 9.12: Höhenlinien der Funktion $f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$.

ausgehend von $\mathbf{x}^0 = (x_1^0, x_2^0)^T$ entspricht also der Funktion

$$f_{\mathbf{x}^0,\mathbf{e}^1}(t) = f(\mathbf{x}^0 + t\mathbf{e}^1) = f(x_1^0 + t\cdot 1, x_2^0 + t\cdot 0) = f(x_1^0 + t, x_2^0).$$

Man hält den Wert der zweiten Variablen x_2^0 fest und betrachtet die Funktion für verschiedene x_1-Werte in Abhängigkeit von t. Analog entspricht der senkrechte Vertikalschnitt in Richtung $\mathbf{e}^2$ einer Fixierung der ersten Variable auf den Wert x_1^0 mit verschiedenen Werten von $x_2 = x_2^0 + t$. Wir visualisieren Vertikalschnitte im Folgenden entweder als Teilmenge des Graphen von f oder als Graph der Funktion $f_{\mathbf{x}^0,\mathbf{r}}$. Im Fall $D = \mathbb{R}^n$ ist der Vertikalschnitt für alle $t \in \mathbb{R}$ definiert.

■ Beispiel 9.1.6 — Vertikalschnitte der Cobb Douglas Funktion.
Abbildung 9.13 zeigt den Vertikalschnitt der Cobb Douglas Funktion in 2 Variablen

$$f : (0, +\infty)^2 \to \mathbb{R} \text{ mit } f(x_1, x_2) = x_1^{0.8} x_2^{0.2}$$

in Richtung $\mathbf{r} = \mathbf{e}^1$ durch $\mathbf{x}^0 = (1,1)^T$. Die durch den Vertikalschnitt beschriebene Funktion entspricht hier

$$f_{\mathbf{x}^0,\mathbf{e}^1}(t) = f(\mathbf{x}^0 + t\mathbf{e}^1) = f(1 + t\cdot 1, 1 + t\cdot 0) = (1+t)^{0.8}(1)^{0.2} = (1+t)^{0.8},$$

wobei $t \in (-1, +\infty)$ gelten muss, damit $(1 + t\cdot 1, 1)^T \in (0, +\infty)^2$.

Abbildung 9.14 zeigt den Vertikalschnitt der gleichen Cobb Douglas Funktion in Richtung $\mathbf{r} = \mathbf{e}^2$ durch $\mathbf{x}^0 = (1,1)^T$. Der senkrechte Vertikalschnitt entspricht

$$f_{\mathbf{x}^0,\mathbf{e}^2}(t) = f(\mathbf{x}^0 + t\mathbf{e}^2) = f(1 + t\cdot 0, 1 + t\cdot 1) = (1)^{0.8}(1+t)^{0.2} = (1+t)^{0.2},$$

wobei wieder $t \in (-1, +\infty)$ gelten muss, damit $(1, 1 + t\cdot 1)^T \in (0, +\infty)^2$. Obige senkrechte Vertikalschnitte verändern jeweils nur eine Komponente und halten die andere konstant. Ein diagonaler Vertikalschnitt mit $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ verändert beide Komponenten gleichzeitig und in gleichen Anteilen. Der Diagonalschnitt ist in Abbildung 9.15 visualisiert und berechnet sich als

$$f_{\mathbf{x}^0,\mathbf{r}}(t) = f(\mathbf{x}^0 + t\mathbf{r}) = f(1 + t\cdot\frac{1}{\sqrt{2}}, 1 + t\cdot\frac{1}{\sqrt{2}}) = (1 + t\frac{1}{\sqrt{2}})^{0.8}(1 + t\frac{1}{\sqrt{2}})^{0.2} = 1 + t\frac{1}{\sqrt{2}},$$

wobei nun $t \in (-\sqrt{2}, +\infty)$ gelten muss, damit $(1 + t\frac{1}{\sqrt{2}}, 1 + t\frac{1}{\sqrt{2}})^T \in (0, +\infty)^2$. ■

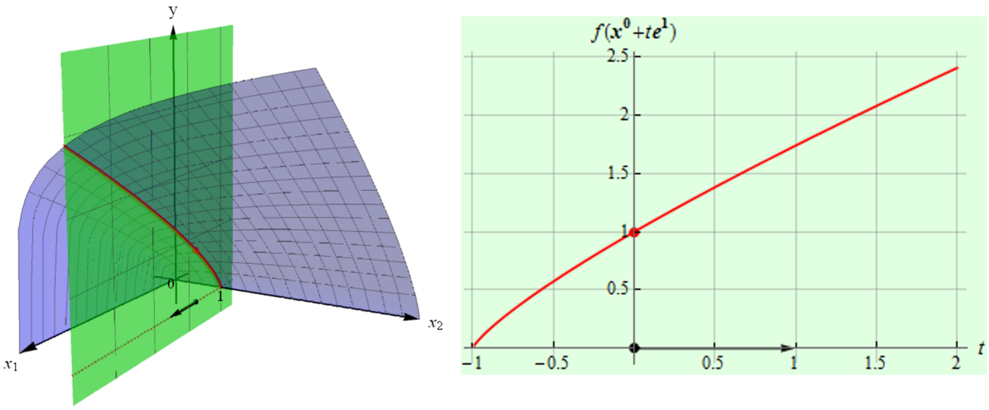

Abbildung 9.13: Vertikalschnitt der Cobb Douglas Funktion in Richtung $\mathbf{e}^1$ durch $\mathbf{x}^0 = (1,1)^T$.

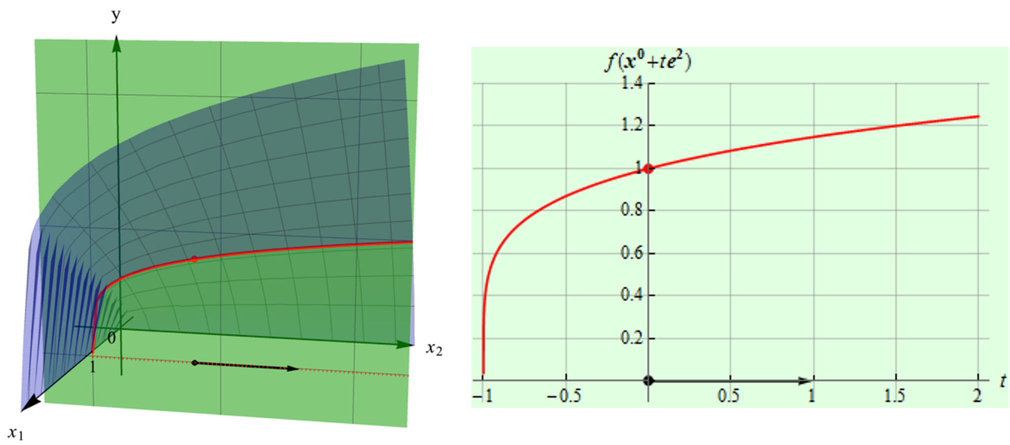

Abbildung 9.14: Vertikalschnitt der Cobb Douglas Funktion in Richtung $\mathbf{e}^2$ durch $\mathbf{x}^0 = (1,1)^T$.

■ Beispiel 9.1.7 — Vertikalschnitte quadratischer Funktionen.
Die Funktion

$$f:\mathbb{R}^2 \to \mathbb{R} \text{ mit } f(x_1,x_2) = x_1^2 + 4(x_2-1)^2 + 3$$

ist eine quadratische Funktion in 2 Variablen. Die entsprechenden Vertikalschnitte durch $\mathbf{x}^0 = \mathbf{0}$ mit Richtungen $\mathbf{e}^1, \mathbf{e}^2$ und $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ sind hier

$$\begin{aligned}
f_{\mathbf{0},\mathbf{e}^1}(t) &= f(\mathbf{0}+t\mathbf{e}^1) = f(0+t\cdot 1, 0+t\cdot 0) = t^2+7,\\
f_{\mathbf{0},\mathbf{e}^2}(t) &= f(\mathbf{0}+t\mathbf{e}^2) = f(0+t\cdot 0, 0+t\cdot 1) = 4(t-1)^2+3 \text{ und}\\
f_{\mathbf{0},\mathbf{r}}(t) &= f(\mathbf{0}+t\mathbf{r}) = f\left(0+t\cdot\frac{1}{\sqrt{2}}, 0+t\cdot\frac{1}{\sqrt{2}}\right) = \frac{1}{2}t^2 + 4\left(\frac{1}{\sqrt{2}}t-1\right)^2 + 3
\end{aligned}$$

und in Abbildungen 9.16, 9.17 und 9.18 dargestellt.

Abbildungen 9.19, 9.20 und 9.21 zeigen Vertikalschnitte der quadratischen Funktion

$$f:\mathbb{R}^2 \to \mathbb{R} \text{ mit } f(x_1,x_2) = x_1^2 + x_2^2 - 3x_1x_2$$

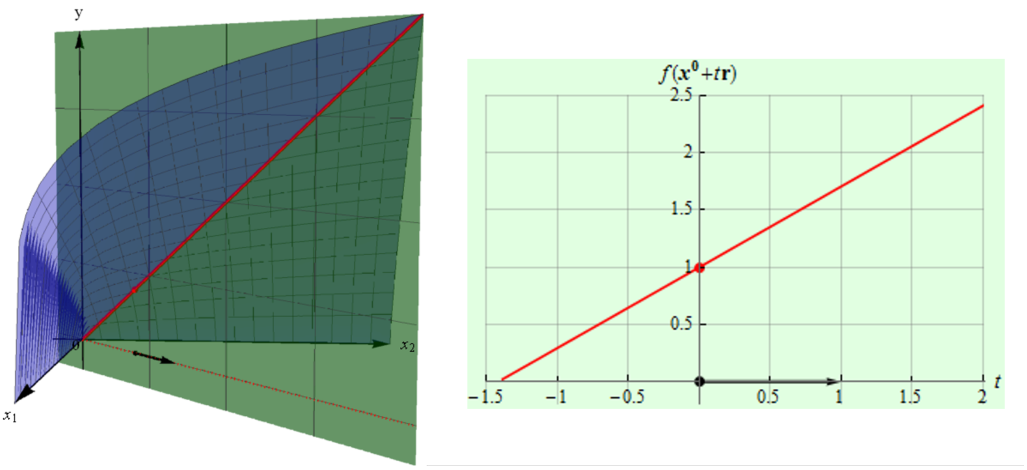

Abbildung 9.15: Vertikalschnitt der Cobb Douglas Funktion in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ durch $\mathbf{x}^0 = (1,1)^T$.

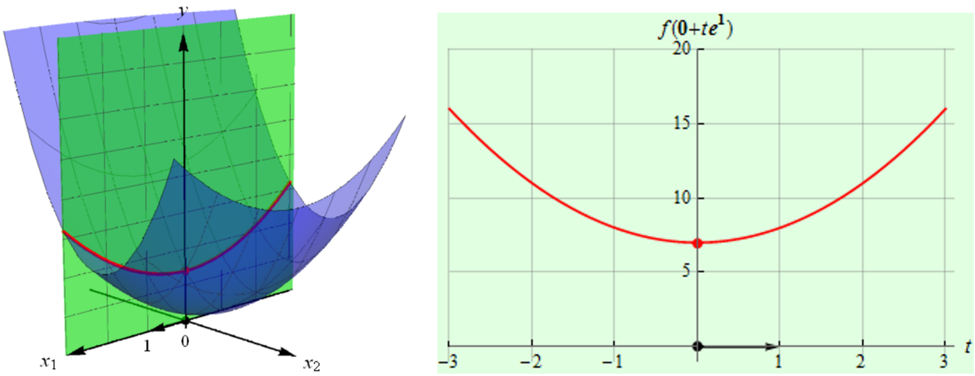

Abbildung 9.16: Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^2 + 3$ in Richtung $\mathbf{e}^1$ durch $\mathbf{0}$.

aus Beispiel 9.1.2 durch $\mathbf{x}^0 = (1,0)^T$ in Richtungen $\mathbf{e}^1, \mathbf{e}^2$ und $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$. Sie berechnen sich als

$$\begin{aligned}
f_{\mathbf{x}^0,\mathbf{e}^1}(t) &= f(\mathbf{x}^0 + t\mathbf{e}^1) = f(1+t\cdot 1, 0+t\cdot 0) = (1+t)^2 \\
f_{\mathbf{x}^0,\mathbf{e}^2}(t) &= f(\mathbf{x}^0 + t\mathbf{e}^2) = f(1+t\cdot 0, 0+t\cdot 1) = 1+t^2-3t \text{ und} \\
f_{\mathbf{x}^0,\mathbf{r}}(t) &= f(\mathbf{x}^0 + t\mathbf{r}) = f\left(1+t\cdot\frac{1}{\sqrt{2}}, 0+t\cdot\frac{1}{\sqrt{2}}\right) \\
&= \left(1+t\frac{1}{\sqrt{2}}\right)^2 + \left(t\frac{1}{\sqrt{2}}\right)^2 - 3\left(1+t\frac{1}{\sqrt{2}}\right)t\frac{1}{\sqrt{2}} = 1-\frac{1}{2}t^2 - t\frac{1}{\sqrt{2}}.
\end{aligned}$$

■

Im Fall von $n > 2$ Variablen ist die Darstellung des Graphen einer Funktion deutlich schwieriger. Vertikalschnitte in Richtung $\mathbf{r}$ kann man aber auch hier von einem beliebigen

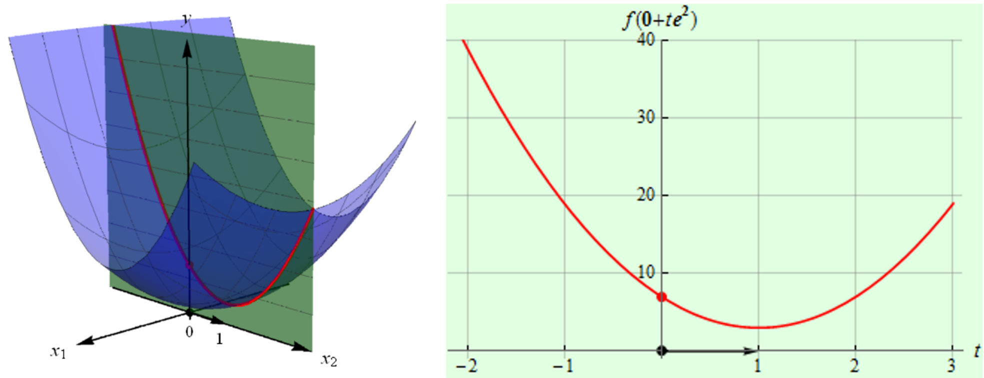

Abbildung 9.17: Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ in Richtung $\mathbf{e}^2$ durch $\mathbf{0}$.

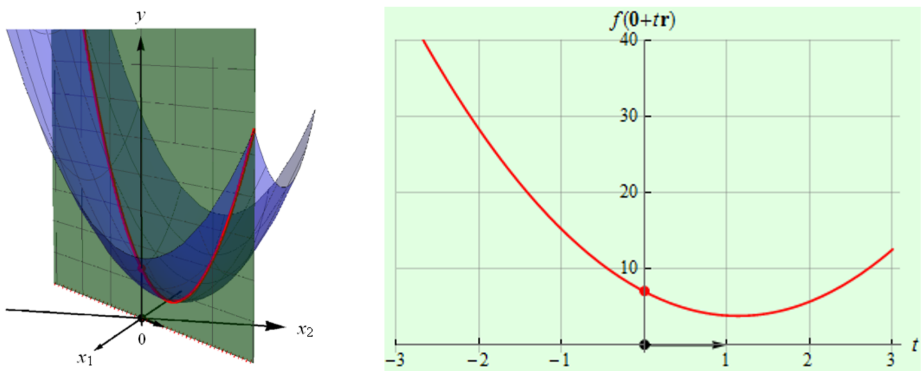

Abbildung 9.18: Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ durch $\mathbf{0}$.

Punkt $\mathbf{x}^0 \in D$ aus darstellen. Aus einem solchen Graphen kann man ablesen, wie sich die Funktion f vom Punkt $\mathbf{x}^0$ ausgehend entlang der Richtung $\mathbf{r}$ verhält.

■ Beispiel 9.1.8 — Vertikalschnitte einer reellen Funktion in 3 Variablen.
Für die Funktion mit Abbildungsvorschrift $f(\mathbf{x}) = 7x_1 + 9x_2 + 8x_3$ aus Beispiel 9.1.2 zeigt Abbildung 9.22 senkrechte Vertikalschnitte, also Vertikalschnitte in Richtung der Einheitsvektoren, ausgehend von $\mathbf{x}^0 = \mathbf{0}$. Das linke Bild zeigt dabei den Graphen des Vertikalschnitts $f_{\mathbf{0},\mathbf{e}^1}(t) = f(t\mathbf{e}^1) = 7t$, das mittlere den Graphen von $f_{\mathbf{0},\mathbf{e}^2}(t) = f(t\mathbf{e}^2) = 9t$ und das rechte den Graphen von $f_{\mathbf{0},\mathbf{e}^3}(t) = f(t\mathbf{e}^3) = 8t$. ■

Z Die Höhenlinie der reellen Funktion f in n Variablen zum Niveau $y \in Z \subseteq \mathbb{R}$ ist die Menge aller Argumente, welche auf diesen Funktionswert abgebildet werden, $N_y = \{\mathbf{x} \in D \mid f(\mathbf{x}) = y\}$.

Ist f eine reelle Funktion in n Variablen, dann heisst

$$f_{\mathbf{x}^0,\mathbf{r}} : \{t \in \mathbb{R} \mid \mathbf{x}^0 + t\mathbf{r} \in D\} \to \mathbb{R} \text{ mit } f_{\mathbf{x}^0,\mathbf{r}}(t) = f(\mathbf{x}^0 + t\mathbf{r})$$

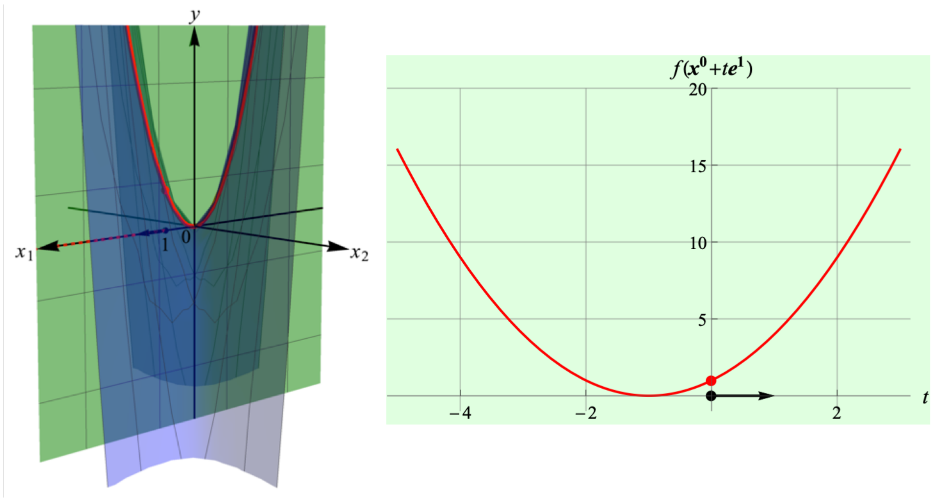

Abbildung 9.19: Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$ in Richtung $\mathbf{r} = \mathbf{e}^1$ durch $\mathbf{x}^0 = (1,0)^T$.

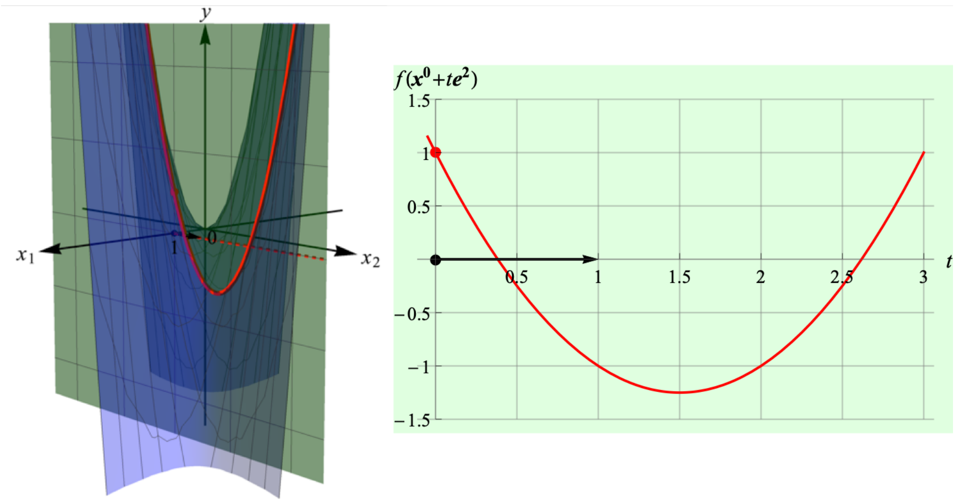

Abbildung 9.20: Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$ in Richtung $\mathbf{r} = \mathbf{e}^2$ durch $\mathbf{x}^0 = (1,0)^T$.

Vertikalschnitt durch $\mathbf{x}^0 \in \mathbb{R}^n$ in Richtung $\mathbf{r} \in \mathbb{R}^n$.

9.1.3 Grenzwert der Funktion an einer Stelle

Ziele dieses Unterkapitels

- Wie definiert man den Grenzwert einer reellen Funktion in mehreren Variablen an einer Stelle?

Wir definieren den Grenzwert einer reellen Funktion in mehreren Variablen an einer Stelle $\mathbf{x}^0$ als den Funktionswert, dem man sich nähert, wenn sich der Vektor der Variablen $\mathbf{x}$ der Stelle $\mathbf{x}^0$ beliebig nähert, also $f(\mathbf{x})$ für $\mathbf{x} \to \mathbf{x}^0$.

Im Sinne der Definition 5.2.5 des Grenzwerts für Funktionen in einer Variablen ist der Grenzwert gleich $a \in \mathbb{R}$, symbolisch $\lim_{\mathbf{x} \to \mathbf{x}^0} f(\mathbf{x}) = a$, wenn es zu jeder maximalen Abweichung ε von a eine δ-Umgebung um $\mathbf{x}^0$ gibt, so dass alle Funktionswerte in dieser

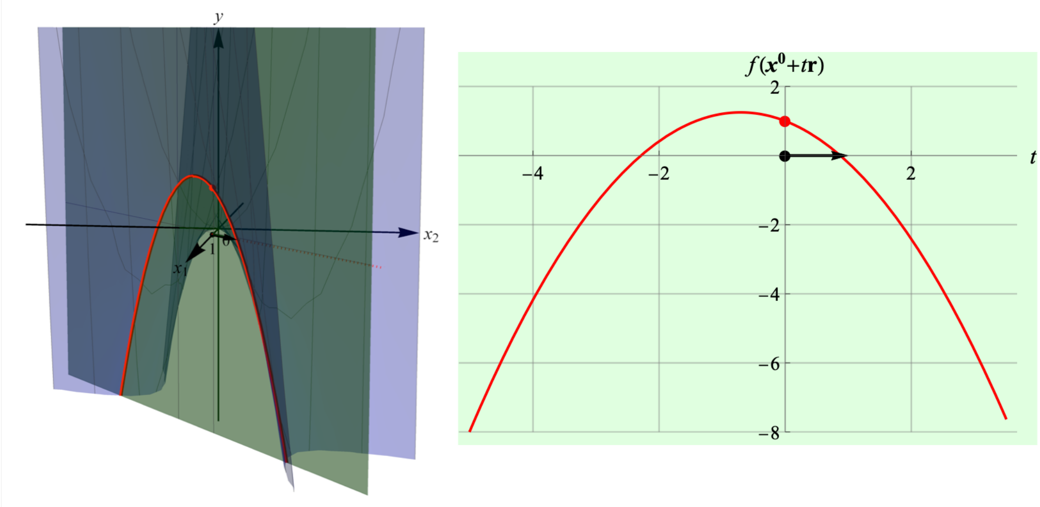

Abbildung 9.21: Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$ in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ durch $\mathbf{x}^0 = (1,0)^T$.

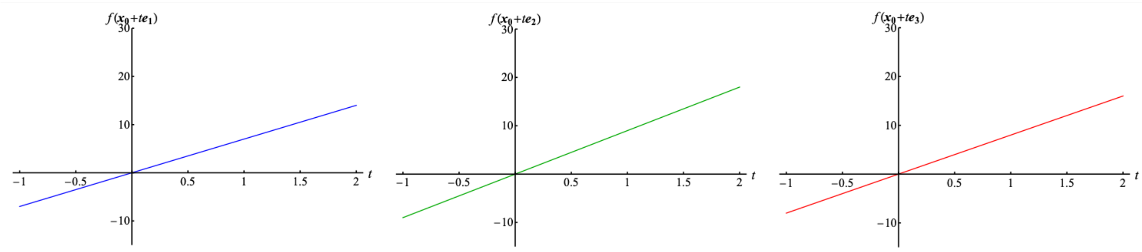

Abbildung 9.22: Senkrechte Vertikalschnitte der Funktion $f(\mathbf{x}) = 7x_1 + 9x_2 + 8x_3$.

Umgebung weniger als die vorgegebene maximale Abweichung ε von a entfernt sind. Die Definition der δ-Umgebung $U(\mathbf{x}^0, \delta)$ von $\mathbf{x}^0$ findet man in Definition 2.2.6. Bei einer Funktion in einer Variablen ist die δ-Umgebung von $x^0 \in \mathbb{R}$ einfach durch das Intervall $(x^0 - \delta, x^0 + \delta)$ gegeben. In diesem Intervall befinden sich alle Werte, die weniger als δ von x^0 entfernt sind. Auch im $\mathbb{R}^n$ beschreibt die δ-Umgebung von $\mathbf{x}^0$ alle Punkte, die weniger als δ von $\mathbf{x}^0$ entfernt sind. Im $\mathbb{R}^2$ entspricht diese Menge einem Kreisinneren und im $\mathbb{R}^3$ einem Kugelinneren, wobei der zugrundeliegende Kreis bzw. die Kugel einen Radius δ haben.

Definition 9.1.5 — Der Grenzwert einer reellen Funktion in n Variablen.
Es sei $f : D \to Z$ mit $D \subseteq \mathbb{R}^n$ und $Z \subseteq \mathbb{R}$ eine reelle Funktion in n Variablen und $\mathbf{x}^0 \in \mathbb{R}^n$, wobei für jede δ-Umgebung von $\mathbf{x}^0$ gilt, dass $U(\mathbf{x}^0, \delta) \cap D \neq \{\}$. Die Funktion f hat an der Stelle $\mathbf{x}^0$ den Grenzwert $a \in \mathbb{R}$,

$$\lim_{\mathbf{x} \to \mathbf{x}^0} f(\mathbf{x}) = a,$$

wenn es zu jeder beliebig kleinen Zahl $\varepsilon > 0$ eine Zahl $\delta > 0$ gibt, so dass für alle $\mathbf{x} \in (U(\mathbf{x}^0, \delta) \cap D) \setminus \{\mathbf{x}^0\}$ gilt:

$$|f(\mathbf{x}) - a| < \varepsilon.$$

Damit $a \in \mathbb{R}$ der Grenzwert von f für $\mathbf{x} \to \mathbf{x}^0$ ist, muss es also zu jeder maximalen Abweichung ε ein δ geben, so dass alle Funktionswerte von Punkten innerhalb der δ-Umgebung von $\mathbf{x}^0$ weniger als ε von a entfernt sind.

■ Beispiel 9.1.9 — Grenzwert einer quadratischen Funktion.
Betrachten wir die reelle Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$$

an der Stelle $\mathbf{x}^0 = (0,1)^T$. Aufgrund der Rechenregeln für Grenzwerte von reellen Funktionen in einer Variablen vermuten wir, dass

$$\lim_{\mathbf{x} \to \mathbf{x}^0} f(\mathbf{x}) = 0^2 + 4(1-1)^2 + 3 = 3.$$

Will man dies mithilfe von Definition 9.1.5 nachweisen, muss man zeigen, dass es zu jeder (maximalen) Abweichung ε eine δ-Umgebung von $\mathbf{x}^0$ gibt, in der alle Funktionswerte weniger als ε von 3 entfernt sind. In anderen Worten muss man zu jedem ε ein δ so bestimmen können, dass

$$|f(\mathbf{x}) - 3| < \varepsilon \qquad \text{für alle } \mathbf{x} \in U(\mathbf{x}^0, \delta).$$

Die Entfernung des Funktionswerts von der Zahl 3 kann man mithilfe der Dreiecksungleichung, Satz 2.2.1, wie folgt nach oben abschätzen:

$$|f(\mathbf{x}) - 3| = |x_1^2 + 4(x_2 - 1)^2| \leq |x_1^2| + 4|(x_2 - 1)^2|.$$

Wir wollen nun diese Entfernung für alle Elemente einer δ-Umgebung weiter abschätzen. Hierfür betrachten wir erneut Definition 2.2.6 einer δ-Umgebung. Demnach gilt

$$\mathbf{x} \in U(\mathbf{x}^0, \delta) \Leftrightarrow \sqrt{x_1^2 + (x_2 - 1)^2} < \delta,$$

alle Stellen $\mathbf{x}$ in der δ-Umgebung von $\mathbf{x}^0$ sind weniger als δ von $\mathbf{x}^0$ entfernt. Für Stellen innerhalb der δ-Umgebung von $\mathbf{x}^0$ gilt also unter anderem $|x_1| < \delta$ und $|x_2 - 1| < \delta$. Wir erhalten also

$$|f(\mathbf{x}) - 3| \leq |x_1^2| + 4|(x_2 - 1)^2| = \underbrace{|x_1|^2}_{<\delta^2} + 4\underbrace{|x_2 - 1|^2}_{<\delta^2} < 5\delta^2.$$

Will man beispielsweise eine Umgebung um $\mathbf{x}_0$ angeben, in welcher alle Funktionswerte weniger als $\varepsilon = 0.5$ von 3 entfernt sind, könnte man obige Schätzung nutzen, um aus

$$5\delta^2 \leq 0.5 \Leftrightarrow \delta \leq \sqrt{0.1}$$

zu schliessen, dass alle Funktionswerte in einer $\sqrt{0.1}$-Umgebung von $\mathbf{x}^0$ Funktionswerte zwischen 2.5 und 3.5 haben müssen, vgl. Abbildung 9.23.

Wählt man also für gegebenes $\varepsilon > 0$ den Wert $\delta = \sqrt{\frac{\varepsilon}{5}}$, so gilt für alle $\mathbf{x} \in U(\mathbf{x}^0, \delta)$

$$|f(\mathbf{x}) - 3| = |x_1^2 + 4(x_2 - 1)^2| < 5\delta^2 = \varepsilon.$$

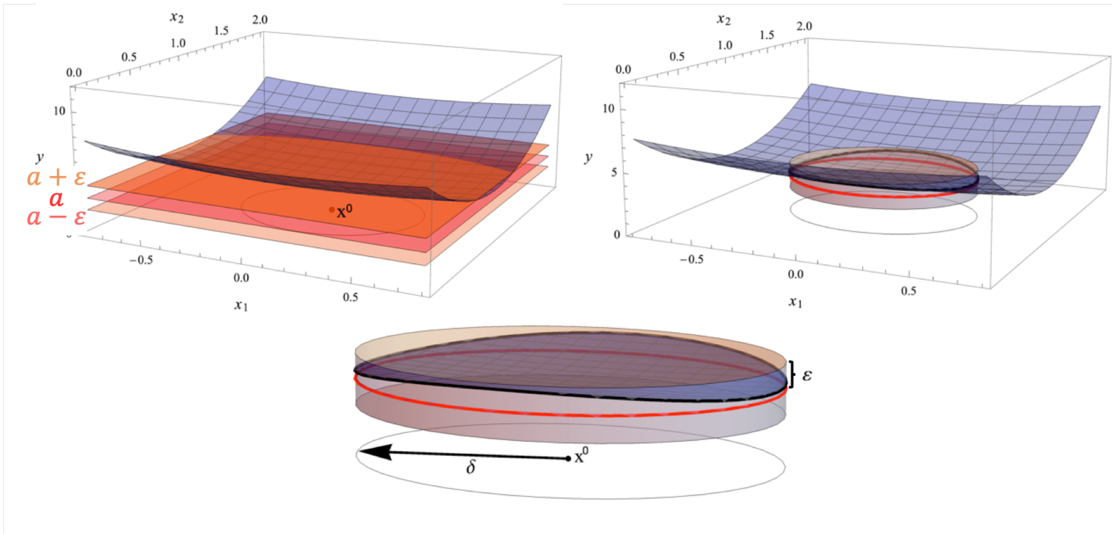

Abbildung 9.23: Der Grenzwert von $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ an der Stelle $\mathbf{x}^0 = (0,1)^T$.

Für jedes ε sind also alle Funktionswerte von Punkten in einer $\delta = \sqrt{\frac{\varepsilon}{5}}$-Umgebung weniger als ε von 3 entfernt. Es gilt somit

$$|f(\mathbf{x}) - 3| < \varepsilon \quad \text{für alle } x \in U(\mathbf{x}^0, \delta) \text{ bzw. } \lim_{\mathbf{x}\to\mathbf{x}^0} f(\mathbf{x}) = 3 = f(\mathbf{x}^0).$$

Der Grenzwert der reellen Funktion für $\mathbf{x} \to \mathbf{x}^0$ ist gleich dem Funktionswert $f(\mathbf{x}^0) = 3$, vgl. Abbildung 9.23.

■

Wie schon im Fall einer reellen Funktion in nur einer Variablen muss die Funktion an einer Stelle $\mathbf{x}^0$ nicht definiert sein.

■ Beispiel 9.1.10 — Grenzwert der Cobb Douglas Funktion für $\mathbf{x} \to \mathbf{0}$.
Für die Cobb Douglas Nutzenfunktion $f: (0, +\infty)^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^{0.8} x_2^{0.2}$ gilt beispielsweise

$$\lim_{\mathbf{x}\to\mathbf{0}} f(\mathbf{x}) = 0.$$

Denn für $\delta = \varepsilon$ ist der Funktionswert an jeder Stelle einer δ-Umgebung von $\mathbf{0}$, welche im Definitionsbereich liegt, kleiner als $\varepsilon > 0$:

$$|f(\mathbf{x}) - 0| = x_1^{0.8} x_2^{0.2} < \delta^{0.8} \delta^{0.2} = \delta = \varepsilon \text{ für alle } \mathbf{x} \in (U(\mathbf{0}, \delta) \cap (0, +\infty)^2).$$

Der Funktionswert an der Stelle $\mathbf{0}$ ist jedoch nicht definiert, da $\mathbf{0}$ nicht im Definitionsbereich liegt. ■

Bei reellen Funktionen in einer Variablen haben wir die Existenz eines Grenzwerts an einer Stelle x^0 auf die Existenz des links- und rechtsseitigen Grenzwerts zurückführen können, vgl. Satz 5.2.4. Bei reellen Funktionen in mehreren Variablen kann man es sich leider nicht so einfach machen, da man sich innerhalb der δ-Umgebung um $\mathbf{x}^0$ nun $\mathbf{x}^0$ aus verschiedenen Richtungen annähern kann.

■ **Beispiel 9.1.11 — Grenzwert einer anderen Funktion für $\mathbf{x} \to \mathbf{0}$.**
In diesem Beispiel versuchen wir den Grenzwert

$$\lim_{\mathbf{x}\to\mathbf{x}^0} h(\mathbf{x})$$

der Funktion $h : \mathbb{R}^2 \setminus \{\mathbf{0}\} \to \mathbb{R}$ mit

$$h(\mathbf{x}) = \frac{x_1 x_2}{x_1^2 + x_2^2}$$

zu berechnen, oder zu entscheiden, dass dieser nicht existiert. Abbildung 9.24 zeigt Höhenlinien dieser Funktion.

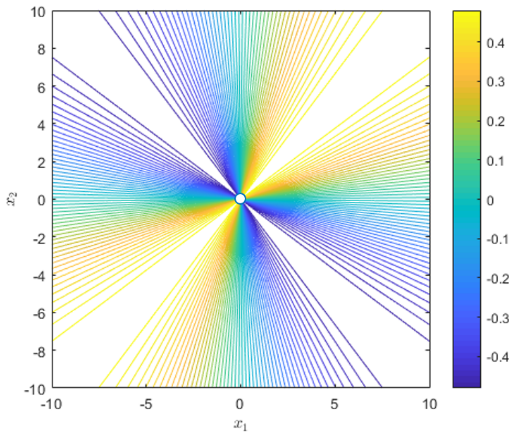

Abbildung 9.24: Höhenlinien der Funktion h.

Der Grenzwert an der Stelle $\mathbf{0}$ hat den Wert a, wenn es zu jedem ε eine δ-Umgebung um $\mathbf{0}$ gibt, in welcher alle Punkte des Definitionsbereichs Funktionswerte haben, welche nicht weiter als ε von a entfernt sind. Betrachten wir hierfür eine beliebige δ-Umgebung um $\mathbf{0}$. Für jedes $\delta > 0$ liegen in der δ-Umgebung sowohl die Punkte $(\frac{\delta}{2}, 0)^T$ als auch $(\frac{\delta}{2}, \frac{\delta}{2})^T$ mit

$$h\left(\frac{\delta}{2}, 0\right) = \frac{0}{(\frac{\delta}{2})^2} = 0 \text{ und } h\left(\frac{\delta}{2}, \frac{\delta}{2}\right) = \frac{\frac{\delta}{2} \cdot \frac{\delta}{2}}{(\frac{\delta}{2})^2 + (\frac{\delta}{2})^2} = \frac{1}{2}.$$

Für kleine Werte von ε, wie z.B. $\varepsilon = 0.1$, kann es also keinen Wert $a \in \mathbb{R}$ geben, so dass beide Werte weniger als ε von a entfernt sind. Die Funktion hat keinen Grenzwert an der Stelle $\mathbf{0}$, siehe Abbildung 9.25. Das Problem bei dieser vermeintlichen Grenzwertbildung kann man hier im Beispiel auch mithilfe von Vertikalschnitten veranschaulichen, siehe Abbildung 9.26: Für ein festes $x_2 = 0$ ist die Funktion h stets gleich 0. Betrachten wir h ausgehend von $\mathbf{0}$ in Richtung $\mathbf{e}^1$ ergibt sich die Funktion $h_{\mathbf{0},\mathbf{e}^1}(t) = 0$. Auch h ausgehend von $\mathbf{0}$ in Richtung $\mathbf{e}^2$ ist stets gleich 0. Betrachtet man die Funktion h ausgehend von $\mathbf{0}$ in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$, ergibt sich jedoch

$$h_{\mathbf{0},\mathbf{r}}(t) = h(\mathbf{0} + t\mathbf{r}) = \frac{1}{2} \text{ für } t \neq 0.$$

■

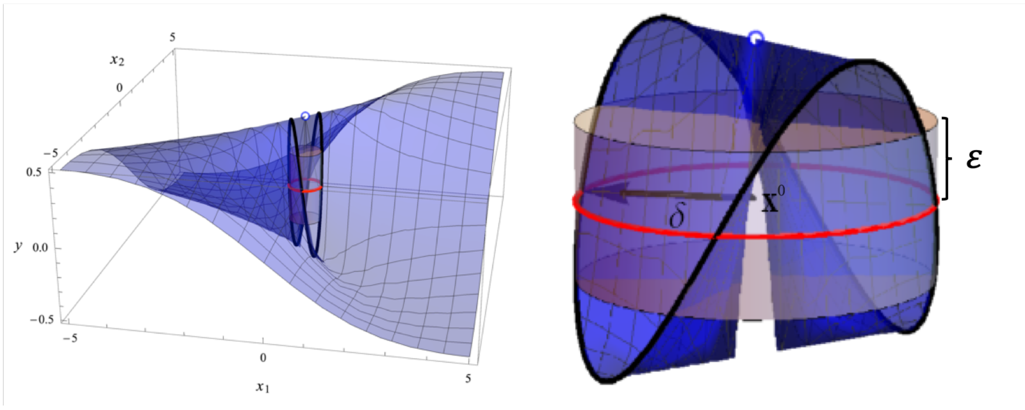

Abbildung 9.25: Das Verhalten von h nahe $\mathbf{x} = (0,0)^T$.

Analog zu Funktionen in einer Variable kann man auch uneigentliche Grenzwerte an einer Stelle definieren. Wir verzichten hier jedoch auf eine formale Definition und Beispiele, da wir diese Konzepte im Folgenden nicht verwenden.

Z $f : D \to Z$ mit $\mathbf{x}^0 \in D \subseteq \mathbb{R}^n$ und $Z \subseteq \mathbb{R}$ hat an der Stelle $\mathbf{x}^0$ den Grenzwert $a \in \mathbb{R}$, $\lim_{\mathbf{x}\to\mathbf{x}^0} f(\mathbf{x}) = a$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass $|f(\mathbf{x}) - a| < \varepsilon$ für alle $\mathbf{x} \in (U(\mathbf{x}^0, \delta) \cap D) \setminus \{\mathbf{x}^0\}$.

9.1.4 Stetigkeit

Ziele dieses Unterkapitels

- Wann heisst eine reelle Funktion in mehreren Variablen stetig an einer Stelle?
- Wann heisst eine reelle Funktion in mehreren Variablen stetig?
- Wie kann man Stetigkeit einer reellen Funktion in mehreren Variablen einfach überprüfen?

Wie im Fall einer Funktion in einer Variablen nennen wir eine Funktion f stetig an der Stelle $\mathbf{x}^0$, wenn der Grenzwert an dieser Stelle existiert und dem Funktionswert entspricht.

Definition 9.1.6 — Stetigkeit einer Funktion.
Es sei $f : D \to Z$ mit $D \subseteq \mathbb{R}^n$ und $Z \subseteq \mathbb{R}$ eine reelle Funktion in n Variablen und $\mathbf{x}^0 \in D$. Man sagt, f ist an der Stelle $\mathbf{x}^0$ stetig, wenn

$$\lim_{\mathbf{x}\to\mathbf{x}^0} f(\mathbf{x}) = f(\mathbf{x}^0).$$

Die Funktion f heisst stetig, falls f stetig an der Stelle $\mathbf{x}^0$ ist für alle $\mathbf{x}^0 \in D$.

Stetigkeit bedeutet auch hier, dass kleine Änderungen in den Werten der Variablen nur zu kleinen Änderungen des Funktionswertes führen.

■ **Beispiel 9.1.12 — Überprüfung der Stetigkeit.**
Betrachten wir die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$$

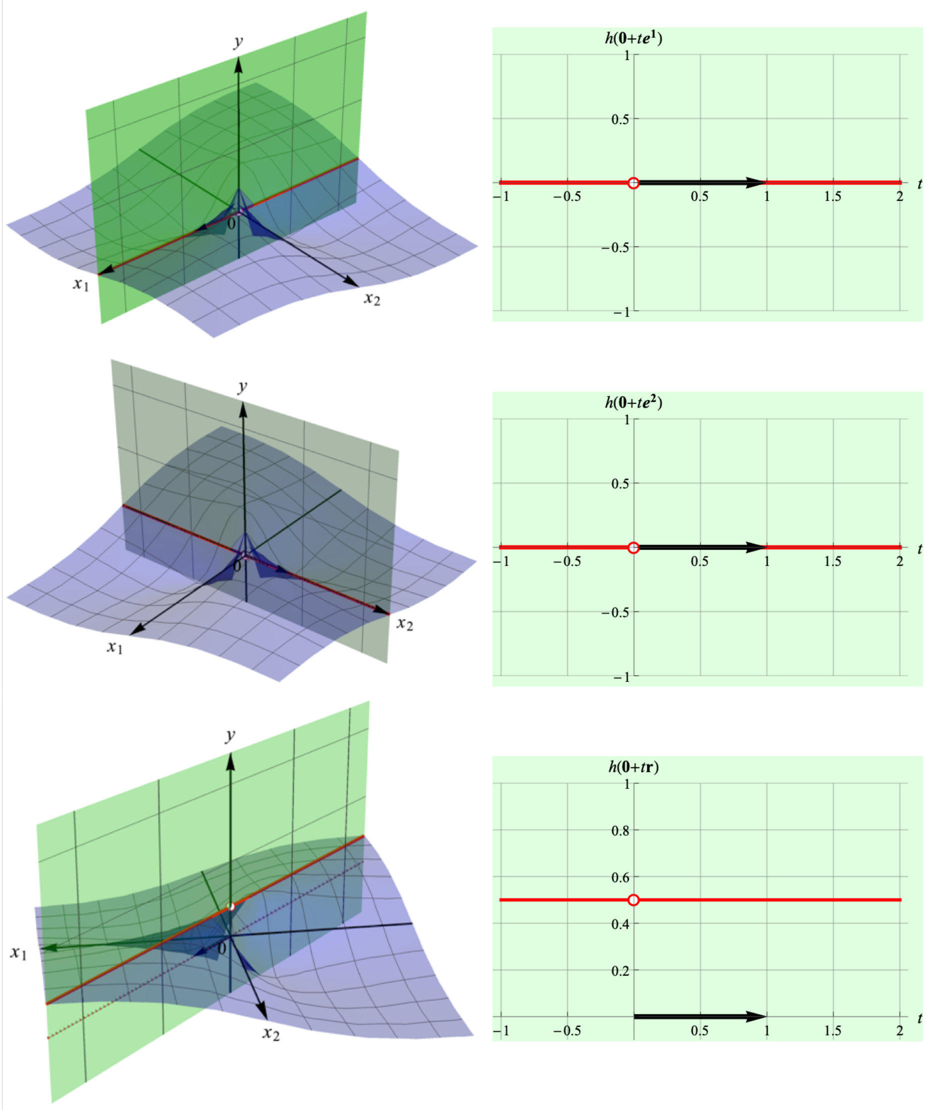

Abbildung 9.26: Vertikalschnitte von h an der Stelle $\mathbf{x} = \mathbf{0}$.

an der Stelle $\mathbf{x}^0 = (0,1)^T$. Aus Beispiel 9.1.9 wissen wir, dass

$$\lim_{\mathbf{x}\to\mathbf{x}^0} f(\mathbf{x}) = 3 = f(\mathbf{x}^0).$$

Die Funktion ist an der Stelle $\mathbf{x}^0$ stetig. ■

Wie im Fall von reellen Funktionen in einer Variablen verifiziert man die Stetigkeit einer reellen Funktion in mehreren Variablen in der Regel dadurch, dass man die Funktion auf eine einfache Funktion zurückführt. Dabei können wir analog zum Fall $n = 1$ nutzen, dass folgende Verknüpfungen Stetigkeit erhalten:

Satz 9.1.1 — Verknüpfungen stetiger Funktionen.
Sind die beiden reellen Funktionen $f : D \to Z_1$ und $g : D \to Z_2$ in n Variablen mit $D \subseteq \mathbb{R}^n$ und $Z_1, Z_2 \subseteq \mathbb{R}$ stetig, so gilt

- $f+g$, $f-g$, $f\cdot g$, $\max\{f,g\}$ und $\min\{f,g\}$ sind stetig auf D,

- $\left(\frac{f}{g}\right)$ ist stetig auf $\{\mathbf{x} \in D \mid g(\mathbf{x}) \neq 0\}$.

Ist h eine stetige reelle Funktion in einer Variablen mit einem Definitionsbereich $g(D)$, so ist die Komposition $h \circ g$ ebenfalls eine reelle Funktion in n Variablen und stetig.

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.2, Satz 5.3). ■

Zudem sind viele uns bekannte Funktionen wie beispielsweise affin-lineare, quadratische, Cobb Douglas und Leontief Funktionen, stetig.

Satz 9.1.2 — Stetigkeit einiger Funktionen.
Die affin-lineare, die quadratische, die Cobb Douglas und die Leontief Funktion in n-Variablen gemäss Definition 9.1.2 sind stetige Funktionen auf ihrem natürlichen Definitionsbereich.

Beweis. Den Beweis mithilfe von Satz 9.1.1 findet man in Kall (1982, Seiten 148-149). ■

Im obigen Beispiel 9.1.12 hätte man also auch einfach darauf verweisen können, dass eine quadratische Funktion überall stetig ist.

■ Beispiel 9.1.13 — Eine Funktion mit Unstetigkeitsstelle.
Wir betrachten die Funktion $h : \mathbb{R}^2 \to \mathbb{R}$ mit

$$h(\mathbf{x}) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{für } \mathbf{x} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Die Funktionen f und g mit Abbildungsvorschriften $f(\mathbf{x}) = x_1 x_2$ und $g(\mathbf{x}) = x_1^2 + x_2^2$ sind quadratische Funktionen in 2 Variablen und damit stetig. Somit ist laut Satz 9.1.1 ihr Quotient stetig für alle $\mathbf{x}$ mit $g(\mathbf{x}) = x_1^2 + x_2^2 \neq 0$. Da $x_1^2 + x_2^2$ genau dann 0 ist, wenn $\mathbf{x} = \mathbf{0}$, ist sichergestellt, dass die Funktion stetig ist für alle $\mathbf{x} \neq \mathbf{0}$.

Aus Beispiel 9.1.11 wissen wir, dass der Grenzwert $\lim_{\mathbf{x} \to \mathbf{0}} h(\mathbf{x})$ nicht existiert. An der Stelle $\mathbf{x} = \mathbf{0}$ ist die Funktion h also auch nicht stetig. ■

Im Fall von reellen Funktionen in zwei Variablen können wir uns Stetigkeit ähnlich vorstellen wie im Fall reeller Funktionen in einer Variablen: Hat der Graph der Funktion keine Sprünge, so ist die Funktion stetig.

Z

Gilt $\lim_{\mathbf{x} \to \mathbf{x}^0} f(\mathbf{x}) = f(\mathbf{x}^0)$, so ist die reelle Funktion f in mehreren Variablen an der Stelle $\mathbf{x}^0$ stetig. Ist f stetig an jeder Stelle $\mathbf{x}^0$ des Definitionsbereichs, so heisst f stetig.

Affin-lineare, quadratische, Cobb Douglas und Leontief Funktionen in mehreren Variablen sind stetig.

Summen, Differenzen, Produkte, das Maximum, das Minimum sowie die Komposition stetiger Funktionen in mehreren Variablen sind stetig. Zudem ist der Quotient stetiger Funktionen in mehreren Variablen an allen Stellen stetig, an welchen der Nenner ungleich 0 ist.

9.2 Differenzierbarkeit

Für reelle Funktionen in einer Variable beschreibt die Ableitung die Rate, mit welcher sich die Funktion verändert, wenn sich das Argument ändert. Wir definierten sie als den Grenzwert des Differenzenquotienten, also des Quotienten aus der Veränderung des Funktionswerts $f(x+\Delta x)-f(x)$ und der Veränderung von x, also Δx, für $\Delta x \to 0$.

Um den Begriff der Ableitung und den hiermit verbundenen Begriff der Differenzierbarkeit auf reelle Funktionen in mehreren Variablen zu übertragen, müssen wir klarstellen, welche Arten der Veränderung von $\mathbf{x}$, einer Variable mit mehreren Komponenten, wir betrachten. Hierbei beginnen wir mit dem einfachsten Fall, dem Fall dass sich nur eine Komponente verändert.

9.2.1 Partielle Ableitungen und der Gradient

Ziele dieses Unterkapitels

- Was versteht man unter der partiellen Ableitung einer reellen Funktion in mehreren Variablen nach x_i?
- Was versteht man unter dem Gradienten einer reellen Funktion in mehreren Variablen?
- Wie kann man eine Tangentialhyperebene einer partiell differenzierbaren reellen Funktion in mehreren Variablen an einer Stelle berechnen?

Wir motivieren den Begriff der partiellen Ableitung mit einem Beispiel.

■ Beispiel 9.2.1 — Veränderungen der Cobb Douglas Funktion.
Wir betrachten erneut die Cobb Douglas Funktion $f:(0,+\infty)^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^{0.8}x_2^{0.2}$. Tabelle 9.1 zeigt eine Liste von Funktionswerten einiger Stellen, Abbildung 9.27 zeigt den Graphen der Funktion mit den aufgelisteten Funktionswerten. Ausgehend von $x_1 = 1$

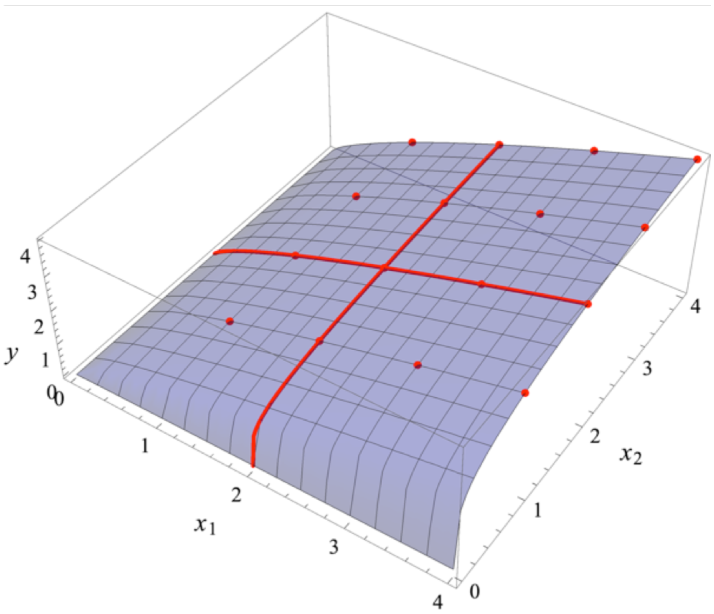

Abbildung 9.27: Graph der Cobb Douglas Funktion.

$f(\mathbf{x})$	$x_1 = 1$	$x_1 = 2$	$x_1 = 3$	$x_1 = 4$
$x_2 = 1$	1.0	$2^{0.8} \approx 1.7$	$3^{0.8} \approx 2.4$	$4^{0.8} \approx 3.0$
$x_2 = 2$	$2^{0.2} \approx 1.1$	2.0	$3^{0.8}2^{0.2} \approx 2.8$	$4^{0.8}2^{0.2} \approx 3.5$
$x_2 = 3$	$3^{0.2} \approx 1.2$	$2^{0.8}3^{0.2} \approx 2.2$	3.0	$4^{0.8}3^{0.2} \approx 3.8$
$x_2 = 4$	$4^{0.2} \approx 1.3$	$2^{0.8}4^{0.2} \approx 2.3$	$3^{0.8}4^{0.2} \approx 3.2$	4.0

Tabelle 9.1: Funktionswerte der Cobb Douglas Funktion.

entspricht für festes $x_2 = 2$ der Nutzen einer zusätzlichen Einheit von x_1 dem Differenzenquotient $\frac{f(2,2)-f(1,2)}{1} \approx 0.9$. Der zusätzliche Nutzen einer weiteren Einheit von x_1 ist nur noch $\frac{f(3,2)-f(2,2)}{1} \approx 0.8$. Erhöht man x_1 erneut um eine Mengeneinheit, ist der zusätzliche Nutzen ungefähr 0.7. Betrachten wir die Funktion f für festes $x_2 = 2$ als Funktion in der Variablen x_1, ergibt sich der Graph in Abbildung 9.28 (links oben).[5] Aus dem Graphen erahnt man, dass die Funktion f für festes $x_2 = 2$ monoton steigt, der Betrag der Steigung aber fällt. Um diese Intuition zu präzisieren, betrachten wir die Funktion für festes x_2 und leiten die so entstandene Funktion in einer Variablen nach x_1 ab. Für festes $x_2 = 2$ ergibt sich diese Ableitung als $0.8x_1^{0.8-1}2^{0.2} > 0$. Da man hier nur Veränderungen in einer Variablen, nämlich x_1, betrachtet und die anderen Variablen der Funktion fixiert, bezeichnet man diese Ableitung auch als die partielle Ableitung von f nach x_1. Die partielle Ableitung von f nach x_1 an der Stelle $(1,2)^T$ ist also beispielsweise $0.8 \cdot 1^{0.8-1}2^{0.2} = 0.8 \cdot 2^{0.2}$.

Ebenso scheint der Grenznutzen von x_2 für festes $x_1 = 2$, also die Ableitung der Funktion f für festes x_1, stets positiv aber fallend, vgl. Abbildung 9.28 (rechts). Wir bestimmen die partielle Ableitung von f nach x_2 in Beispiel 9.2.2. ■

Betrachtet man alle Variablen der Funktion f ausser x_i als Parameter, und leitet dann nach x_i ab, spricht man von der partiellen Ableitung von f nach x_i.

Definition 9.2.1 — Partielle Ableitung.
Sei $D \subseteq \mathbb{R}^n$ offen. Eine reelle Funktion $f: D \to Z$ in n Variablen heisst an der Stelle $\mathbf{x}^0 \in D$ partiell nach x_i, $i = 1, \ldots, n$ differenzierbar, wenn der Grenzwert

$$\lim_{\Delta x \to 0} \frac{f(\mathbf{x}^0 + \Delta x \cdot \mathbf{e}^i) - f(\mathbf{x}^0)}{\Delta x}$$

existiert. Im Falle der Existenz nennt man den Grenzwert die partielle Ableitung, symbolisch $\frac{\partial f(\mathbf{x}^0)}{\partial x_i}$. Die Funktion f heisst partiell differenzierbar nach x_i, $i \in \{1, \ldots, n\}$, wenn sie an jeder Stelle $\mathbf{x}^0 \in D$ partiell nach x_i differenzierbar ist. Die Funktion mit Abbildungsvorschrift

$$f_{x_i}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i}$$

nennt man die partielle Ableitung von f nach x_i. Ist eine Funktion f nach allen Variablen $x_i, i = 1, \ldots, n$ differenzierbar, so nennt man f auch partiell differenzierbar.

Abbildung 9.29 visualisiert den Differenzenquotient, dessen Grenzwert die partielle Ableitung nach x_1 darstellt, am Beispiel der Cobb Douglas Funktion. Wir demonstrieren

[5] Diese Betrachtungsweise entspricht einem Vertikalschnitt in Richtung $\mathbf{e}^1$ ausgehend von $\mathbf{x}^0 = (0,2)$, siehe Abbildung 9.27.

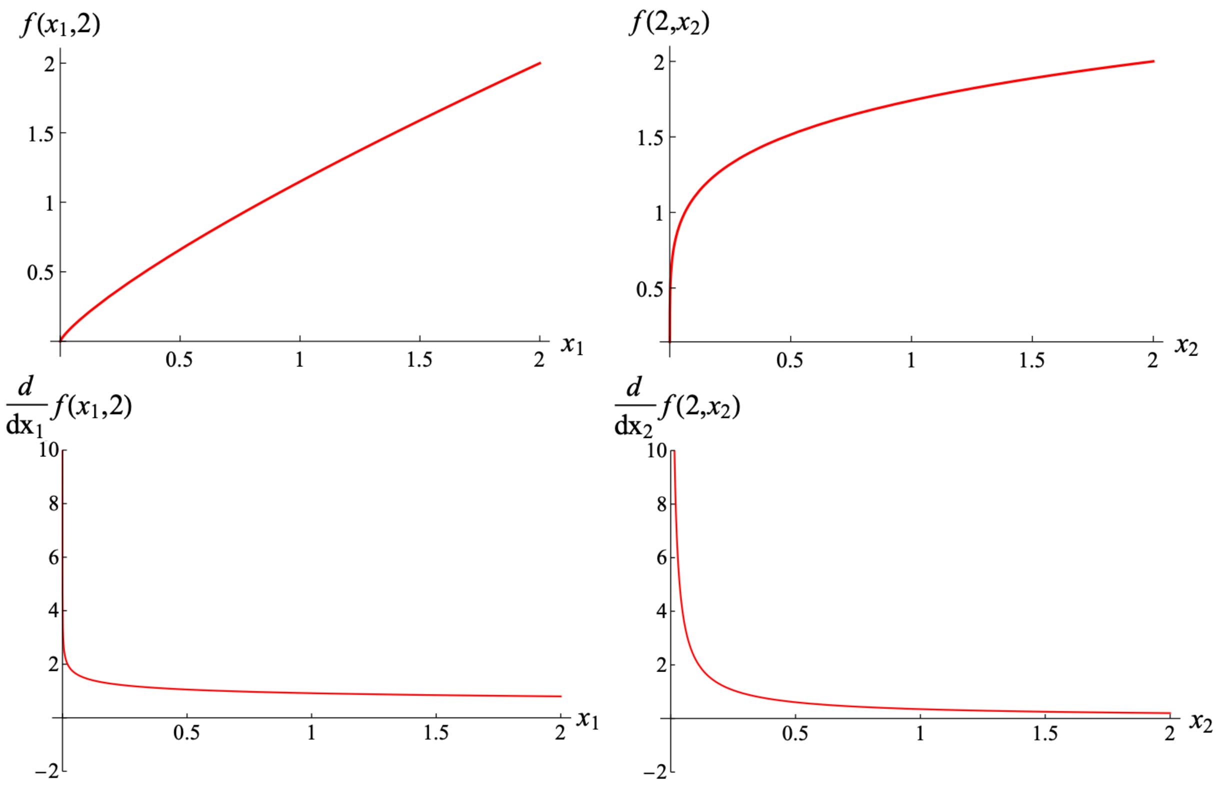

Abbildung 9.28: Grenznutzen der Cobb Douglas Funktion.

die Bildung partieller Ableitungen in einigen Beispielen.

■ Beispiel 9.2.2 — Partielle Ableitung der Cobb Douglas Funktion.
Berechnet man in Beispiel 9.2.1 die partielle Ableitung der Cobb Douglas Funktion $f : (0,+\infty)^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^{0.8} x_2^{0.2}$ nach x_1, so behandelt man x_2 wie eine Konstante und leitet nach x_1 ab. Die partielle Ableitung nach x_1 ist dann gleich

$$f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} = \frac{\partial x_1^{0.8} x_2^{0.2}}{\partial x_1} = 0.8 x_1^{0.8-1} x_2^{0.2} = 0.8 x_1^{-0.2} x_2^{0.2} = 0.8 \frac{x_2^{0.2}}{x_1^{0.2}} = 0.8 \left(\frac{x_2}{x_1}\right)^{0.2}.$$

Um die partielle Ableitung der Funktion nach x_2 zu berechnen, behandelt man x_1 wie eine Konstante und leitet nach x_2 ab. Die partielle Ableitung nach x_2 ergibt dann

$$f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2} = \frac{\partial x_1^{0.8} x_2^{0.2}}{\partial x_2} = 0.2 x_1^{0.8} x_2^{0.2-1} = 0.2 x_1^{0.8} x_2^{-0.8} = 0.2 \frac{x_1^{0.8}}{x_2^{0.8}} = 0.2 \left(\frac{x_1}{x_2}\right)^{0.8}.$$

■

■ Beispiel 9.2.3 — Partielle Ableitung einer quadratischen Funktion.
Wir berechnen nun die partiellen Ableitungen der quadratischen Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^2 + 3$. Der Term $4(x_2-1)^2 + 3$ verändert sich nicht, wenn nur x_1 verändert wird. Somit ist die partielle Ableitung dieses Terms nach x_1 gleich 0. Der Term x_1^2 abgeleitet nach x_1 ist $2x_1$. Wir erhalten somit

$$f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} = \frac{\partial (x_1^2 + 4(x_2-1)^2 + 3)}{\partial x_1} = 2x_1.$$

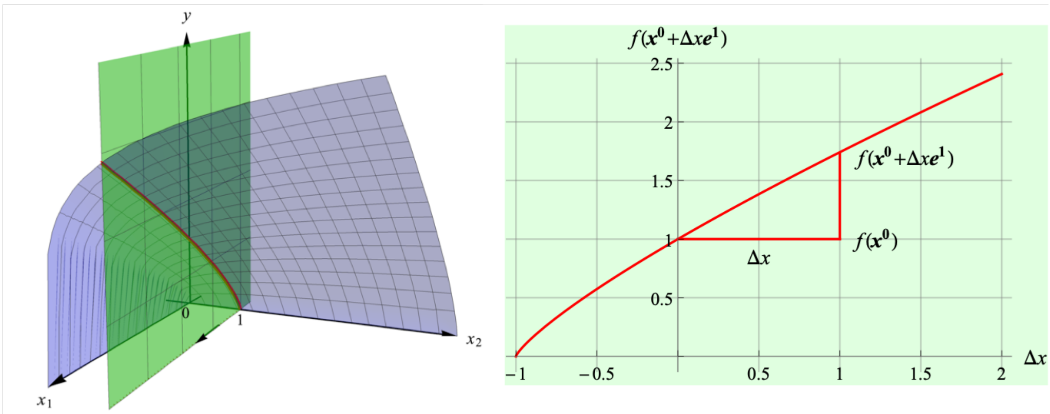

Abbildung 9.29: Visualisierung der partiellen Ableitung.

Bei der partiellen Ableitung der Funktion nach x_2 wird x_1 als Konstante betrachtet und nach x_2 abgeleitet. Somit entspricht die partielle Ableitung

$$f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2} = \frac{\partial(x_1^2+4(x_2-1)^2+3)}{\partial x_2} = 8(x_2-1).$$

■

■ **Beispiel 9.2.4 — Partielle Ableitung einer affin-linearen Funktion.**
Wie im Fall reeller Funktionen in einer Variablen sind die partiellen Ableitungen einer affin-linearen Funktion in mehreren Variablen konstant. Beispielsweise ist die partielle Ableitung der affin-linearen Funktion mit Abbildungsvorschrift $f(\mathbf{x}) = -8x_1+3x_2+10$ nach x_1 gleich

$$f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} = \frac{\partial(-8x_1+3x_2+10)}{\partial x_1} = -8.$$

Die partielle Ableitung nach x_2 ist

$$f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2} = \frac{\partial(-8x_1+3x_2+10)}{\partial x_2} = 3.$$

■

■ **Beispiel 9.2.5 — Eine nicht partiell differenzierbare Funktion.**
Die Leontief Funktion aus Beispiel 9.1.1 mit Abbildungsvorschrift $f(\mathbf{x}) = \min\{x_1, \frac{1}{2}x_2\}$ ist nicht überall partiell differenzierbar. Beispielsweise gilt an der Stelle $\mathbf{x}^0 = (1,2)^T$

$$\frac{f(\mathbf{x}^0+\Delta x\mathbf{e}^1)-f(\mathbf{x}^0)}{\Delta x} = \frac{\min\{1+\Delta x, 1\}-1}{\Delta x} = \begin{cases} 0 & \text{für } \Delta x > 0 \\ 1 & \text{für } \Delta x < 0 \end{cases}.$$

Somit existiert $\lim_{\Delta x\to 0} \frac{f(\mathbf{x}^0+\Delta x\mathbf{e}^1)-f(\mathbf{x}^0)}{\Delta x}$ für $\mathbf{x}^0 = (1,2)^T$ nicht. Die partielle Ableitung nach x_1 existiert also an dieser Stelle nicht. Am Graphen der Funktion ist an dieser Stelle ein Knick erkennbar, vgl. Abbildung 9.1. ■

Fasst man die partiellen Ableitungen nach allen Variablen an einer Stelle zusammen, ergibt sich ein Vektor, den man als Gradienten von f bezeichnet.

Definition 9.2.2 — Der Gradient.
Sei $D \subseteq \mathbb{R}^n$ offen, und $f:\ D \to Z$ eine reelle und an der Stelle $\mathbf{x}^0$ partiell differenzierbare Funktion in n Variablen. Der Vektor

$$\operatorname{grad} f(\mathbf{x}^0) = \nabla f(\mathbf{x}^0) = \begin{pmatrix} \frac{\partial f(\mathbf{x}^0)}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x}^0)}{\partial x_n} \end{pmatrix}$$

heisst Gradient von f an der Stelle $\mathbf{x}^0$.

■ **Beispiel 9.2.6 — Gradient einer affin-linearen Funktion.**
Die partiellen Ableitungen einer affin-linearen Funktion sind stets konstant. Der Gradient der Funktion mit $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$ ist für alle $\mathbf{x}^0$ gleich

$$\operatorname{grad} f(\mathbf{x}^0) = \nabla f(\mathbf{x}^0) = \begin{pmatrix} -8 \\ 3 \end{pmatrix}.$$

Damit gilt unter anderem

$$\nabla f(1,1) = \nabla f(1,32) = \begin{pmatrix} -8 \\ 3 \end{pmatrix}.$$

■

■ **Beispiel 9.2.7 — Gradient der Cobb Douglas Funktion.**
Die partiellen Ableitungen der Cobb Douglas Funktion $f:(0,+\infty)^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^{0.8} x_2^{0.2}$ haben wir als

$$f_{x_1}(\mathbf{x}) = 0.8\left(\frac{x_1}{x_2}\right)^{0.2} \text{ und } f_{x_2}(\mathbf{x}) = 0.2\left(\frac{x_1}{x_2}\right)^{0.8}$$

bestimmt. Allgemein berechnet sich der Gradient dieser Funktion an einer Stelle $\mathbf{x}^0$ also als

$$\operatorname{grad} f(\mathbf{x}^0) = \nabla f(\mathbf{x}^0) = \begin{pmatrix} 0.8\left(\frac{x_1^0}{x_2^0}\right)^{0.2} \\ 0.2\left(\frac{x_1^0}{x_2^0}\right)^{0.8} \end{pmatrix}.$$

An der Stelle $\mathbf{x}^0 = (1,1)^T$ ist der Gradient damit

$$\operatorname{grad} f(1,1) = \nabla f(1,1) = \begin{pmatrix} 0.8\left(\frac{1}{1}\right)^{0.2} \\ 0.2\left(\frac{1}{1}\right)^{0.8} \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}.$$

Auch an der Stelle $\mathbf{x}^0 = (32,32)^T$ ist der Gradient

$$\operatorname{grad} f(32,32) = \nabla f(32,32) = \begin{pmatrix} 0.8\left(\frac{32}{32}\right)^{0.2} \\ 0.2\left(\frac{32}{32}\right)^{0.8} \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}.$$

An der Stelle $\mathbf{x}^0 = (1,32)^T$ ist der Gradient

$$\operatorname{grad} f(1,32) = \nabla f(1,32) = \begin{pmatrix} 0.8\left(\frac{32}{1}\right)^{0.2} \\ 0.2\left(\frac{1}{32}\right)^{0.8} \end{pmatrix} = \begin{pmatrix} \frac{8}{5} \\ \frac{1}{80} \end{pmatrix}.$$

Will man an der Stelle $\mathbf{x}^0 = (1,32)^T$ eine Tangente an den Vertikalschnitt in Richtung der x_1-Achse, also in Richtung $\mathbf{e}^1$ ausgehend von $\mathbf{x}^0$, legen, hat diese Tangente eine Steigung, die der partiellen Ableitung nach x_1 an dieser Stelle entspricht. Die Steigung der Tangente ist somit $f_{x_1}(\mathbf{x}^0) = \frac{8}{5}$. Ihre Geradengleichung ist

$$y = f(\mathbf{x}^0) + \frac{8}{5}(x_1 - 1) = 2 + \frac{8}{5}(x_1 - 1).$$

Will man an der Stelle $\mathbf{x}^0 = (1,32)^T$ eine Tangente an den Vertikalschnitt in Richtung $\mathbf{e}^2$ ausgehend von $\mathbf{x}^0$ legen, hat diese Tangente eine Steigung, die der partiellen Ableitung nach x_2 an dieser Stelle entspricht. Diese Steigung ist $f_{x_2}(\mathbf{x}^0) = \frac{1}{80}$. Die Geradengleichung der Tangente ist damit

$$y = f(\mathbf{x}^0) + \frac{1}{80}(x_2 - 32) = 2 + \frac{1}{80}(x_2 - 32).$$

Die beiden Tangenten zusammen spannen die sogenannte Tangentialebene von f an der Stelle $\mathbf{x}^0 = (1,32)^T$ auf,

$$t_{1,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0)(x_i - x_i^0) = 2 + \frac{8}{5}(x_1 - 1) + \frac{1}{80}(x_2 - 32),$$

vgl. Abbildung 9.30. Die Funktion $t_{1,\mathbf{x}^0}$ entspricht also der Verallgemeinerung einer Tangente für (partiell) differenzierbare Funktionen in $n = 2$ Variablen. Mithilfe obiger Formel

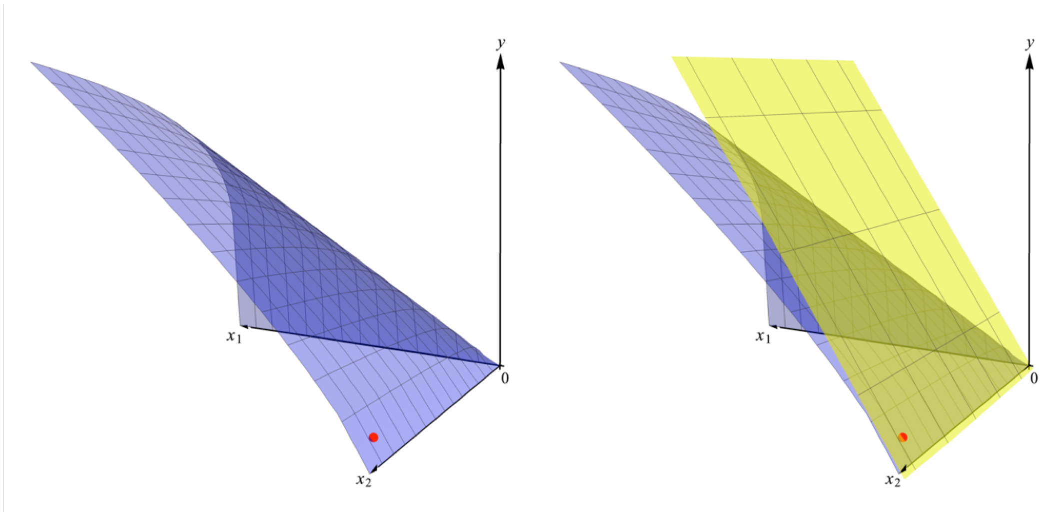

Abbildung 9.30: Tangentialebene der Cobb Douglas Funktion an der Stelle $\mathbf{x}^0 = (1,32)^T$.

ergibt sich die Tangentialebene von f an der Stelle $\mathbf{x}^0 = (1,1)^T$ als

$$\begin{aligned} t_{1,\mathbf{x}^0}(\mathbf{x}) &= 1 + 0.8(x_1 - 1) + 0.2(x_2 - 1) \\ &= 1 - 0.8 - 0.2 + 0.8x_1 + 0.2x_2 = 0.8x_1 + 0.2x_2, \end{aligned}$$

vgl. Abbildung 9.31. Der Graph von $t_{1,\mathbf{x}^0}$ entspricht auch geometrisch einer Ebene im $\mathbb{R}^3$, da die Menge aller Punkte (x_1, x_2, y) mit $y = t_{1,\mathbf{x}^0}(x_1, x_2)$ der Lösungsmenge der linearen Gleichung $0 = 0.8x_1 + 0.2x_2 - y$ entspricht, vgl. Satz 7.3.1. ■

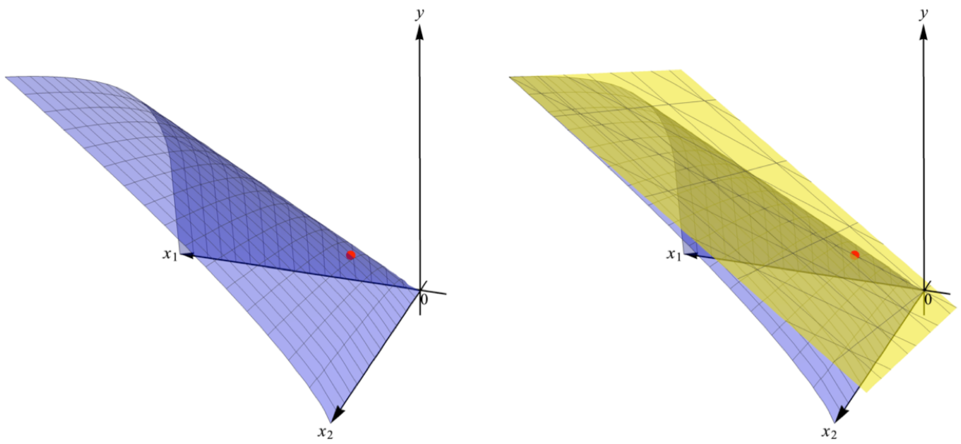

Abbildung 9.31: Tangentialebene der Cobb Douglas Funktion an der Stelle $\mathbf{x}^0 = (1,1)^T$.

So wie man mithilfe der Ableitung eine Tangente an den Graphen einer reellen Funktion in einer Variablen legen kann, kann man mithilfe der partiellen Ableitungen bzw. des Gradienten eine Tangentialebene an den Graphen einer reellen Funktion in zwei Variablen legen. Bei Funktionen in $n \geq 3$ Variablen spricht man von Tangentialhyperebenen.

Definition 9.2.3 — Tangentialebene und Tangentialhyperebene .
Sei $D \subseteq \mathbb{R}^n$ offen, $f : D \to Z$ eine partiell differenzierbare reelle Funktion in n Variablen und $\mathbf{x}^0 \in D$. Die Funktion

$$t_{1,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0)(x_i - x_i^0) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0)$$

beschreibt im Fall $n = 1$ eine Tangente, im Fall $n = 2$ eine Tangentialebene, und allgemein eine Tangentialhyperebene (an den Graphen) von f an der Stelle $\mathbf{x}^0$.

Abbildung 9.32 visualisiert schematisch eine Tangential(hyper)ebene im $\mathbb{R}^3$. Ihre Elemente sind Summen des Punkts $(x_1^0, x_2^0, f(x_1^0, x_2^0))^T$ und Elementen der linearen Hülle der Vektoren $(1,0,f_{x_1}(x_1^0,x_2^0))^T$ und $(0,1,f_{x_2}(x_1^0,x_2^0))^T$. Dabei beschreibt der Vektor $(1,0,f_{x_1}(x_1^0,x_2^0))^T$ den Anstieg um $f_{x_1}(x_1^0,x_2^0)$, wenn sich x_1 um 1 vergrössert und $(0,1,f_{x_2}(x_1^0,x_2^0))^T$ den Anstieg um $f_{x_2}(x_1^0,x_2^0)$, wenn sich x_2 um 1 vergrössert.

■ **Beispiel 9.2.8 — Konstruktion einer Tangentialebene.**
Will man die Tangentialebene von

$$f : \mathbb{R}^2 \to \mathbb{R} \text{ mit } f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$$

an der Stelle $\mathbf{x}^0 = (1,2)^T$ als Funktion beschreiben, ergibt sich mit

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} f_{x_1}(x_1^0, x_2^0) \\ f_{x_2}(x_1^0, x_2^0) \end{pmatrix} = \begin{pmatrix} 2x_1^0 \\ 8(x_2^0 - 1) \end{pmatrix}$$

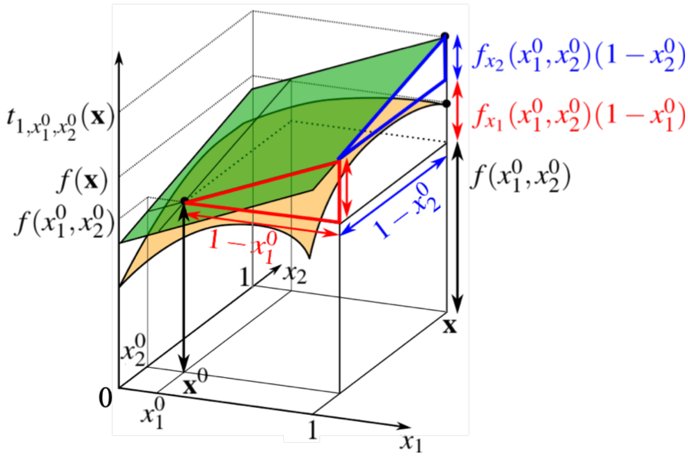

Abbildung 9.32: Die Tangentialebene.

die Funktion

$$\begin{aligned} t_{1,\mathbf{x}^0}(x_1,x_2) &= f(x_1^0,x_2^0) + f_{x_1}(x_1^0,x_2^0)(x_1 - x_1^0) + f_{x_2}(x_1^0,x_2^0)(x_2 - x_2^0) \\ &= f(1,2) + f_{x_1}(1,2)(x_1 - 1) + f_{x_2}(1,2)(x_2 - 2) \\ &= 8 + 2(x_1 - 1) + 8(x_2 - 2) = 2x_1 + 8x_2 - 10. \end{aligned}$$

Diese Tangentialebene ist in Abbildung 9.33 dargestellt. ■

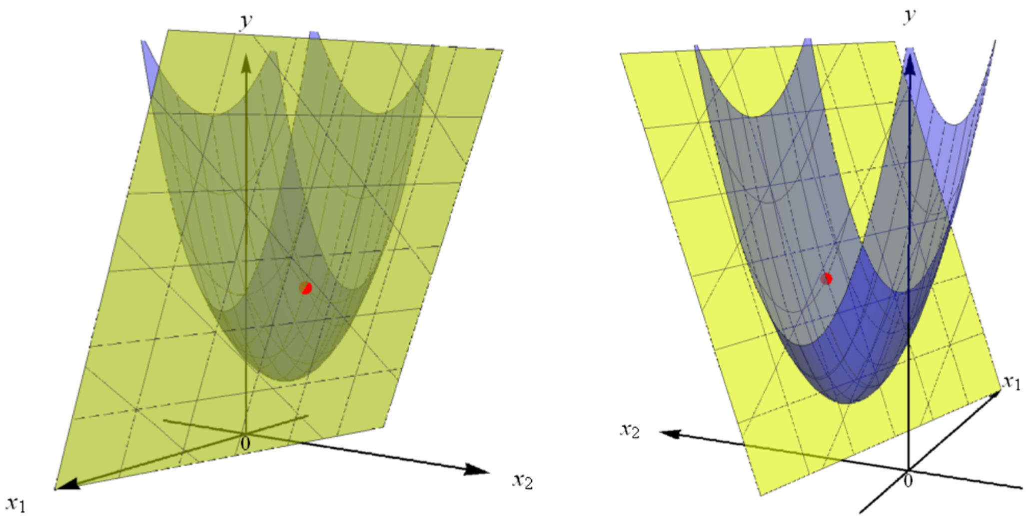

Abbildung 9.33: Die Tangentialebene der quadratischen Funktion an der Stelle $\mathbf{x}^0 = (1,2)^T$.

Z Ist f eine reelle Funktion in n Variablen, dann ist im Falle der Existenz die partielle Ableitung nach x_i die Funktion

$$f_{x_i}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{\Delta x \to 0} \frac{f(\mathbf{x} + \Delta x \cdot \mathbf{e}^i) - f(\mathbf{x})}{\Delta x}.$$

Der Gradient an der Stelle $\mathbf{x}^0$ ist der Spaltenvektor
$\operatorname{grad} f(\mathbf{x}^0) = \nabla f(\mathbf{x}^0) = \left(\frac{\partial f(\mathbf{x}^0)}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{x}^0)}{\partial x_n}\right)^T \in \mathbb{R}^n$.
Sei f eine partiell differenzierbare reelle Funktion in n Variablen und $\mathbf{x}^0$ im Definitionsbereich von f. Die Funktion

$$t_{1,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0)(x_i - x_i^0) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T (\mathbf{x} - \mathbf{x}^0)$$

beschreibt eine Tangentialhyperebene (an den Graphen) von f an der Stelle $\mathbf{x}^0$.

9.2.2 Stetig (partiell) differenzierbare Funktionen

Ziele dieses Unterkapitels

- Was ist eine stetig partiell differenzierbare reelle Funktion in mehreren Variablen?
- Ist jede partiell differenzierbare reelle Funktion in mehreren Variablen auch stetig?
- Ist jede stetig partiell differenzierbare reelle Funktion in mehreren Variablen auch stetig?

Ist eine reelle Funktion in einer Variablen an einer Stelle unstetig, so ist sie dort auch nicht differenzierbar. Folgendes Beispiel illustriert, dass eine reelle Funktion in mehreren Variablen an einer Stelle unstetig und dennoch (auch an dieser Stelle) nach allen Variablen partiell differenzierbar sein kann:

■ Beispiel 9.2.9 — Partielle Ableitungen einer unstetigen Funktion.
Wir betrachten erneut die Funktion

$$h(\mathbf{x}) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{für } \mathbf{x} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ 0 & \text{für } \mathbf{x} = \mathbf{0} \end{cases}$$

aus Beispiel 9.1.13. Für $x_2 \neq 0$ gilt $\mathbf{x} \neq \mathbf{0}$. Mithilfe der Quotientenregel erhalten wir damit die partielle Ableitung nach x_1 für festes $x_2 \neq 0$ als

$$\begin{aligned} h_{x_1}(\mathbf{x}) &= \frac{\partial h(\mathbf{x})}{\partial x_1} = \frac{\partial \frac{x_1 x_2}{x_1^2 + x_2^2}}{\partial x_1} = \frac{\partial \left(x_2 \cdot \frac{x_1}{x_1^2 + x_2^2}\right)}{\partial x_1} \\ &= x_2 \left[\frac{1(x_1^2 + x_2^2) - x_1(2x_1 + 0)}{(x_1^2 + x_2^2)^2}\right] = x_2 \left[\frac{-x_1^2 + x_2^2}{(x_1^2 + x_2^2)^2}\right] = \frac{x_2(x_2^2 - x_1^2)}{(x_1^2 + x_2^2)^2}. \end{aligned}$$

Für $\mathbf{x} = \mathbf{0}$ folgt aus der Definition der partiellen Ableitung, dass

$$h_{x_1}(\mathbf{0}) = \lim_{\Delta x \to 0} \frac{h(\mathbf{0} + \Delta x \cdot \mathbf{e}^1) - h(\mathbf{0})}{\Delta x} = \lim_{\Delta x \to 0} \frac{h(\Delta x \cdot \mathbf{e}^1) - h(\mathbf{0})}{\Delta x} = \lim_{\Delta x \to 0} \frac{0 - 0}{\Delta x} = 0.$$

Zusammenfassend ist also für alle $\mathbf{x} \in \mathbb{R}^2$

$$h_{x_1}(\mathbf{x}) = \begin{cases} \frac{x_2(x_2^2 - x_1^2)}{(x_1^2 + x_2^2)^2} & \text{für } \mathbf{x} \neq \mathbf{0} \\ 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Analog ergibt sich die partielle Ableitung der Funktion h nach der Variable x_2 als

$$h_{x_2}(\mathbf{x}) = \begin{cases} \frac{x_1(x_1^2-x_2^2)}{(x_1^2+x_2^2)^2} & \text{für } \mathbf{x} \neq \mathbf{0} \\ 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Die Funktion ist also an allen Stellen ihres Definitionsbereichs sowohl nach x_1 als auch nach x_2 partiell differenzierbar. Es handelt sich also um eine partiell differenzierbare Funktion.

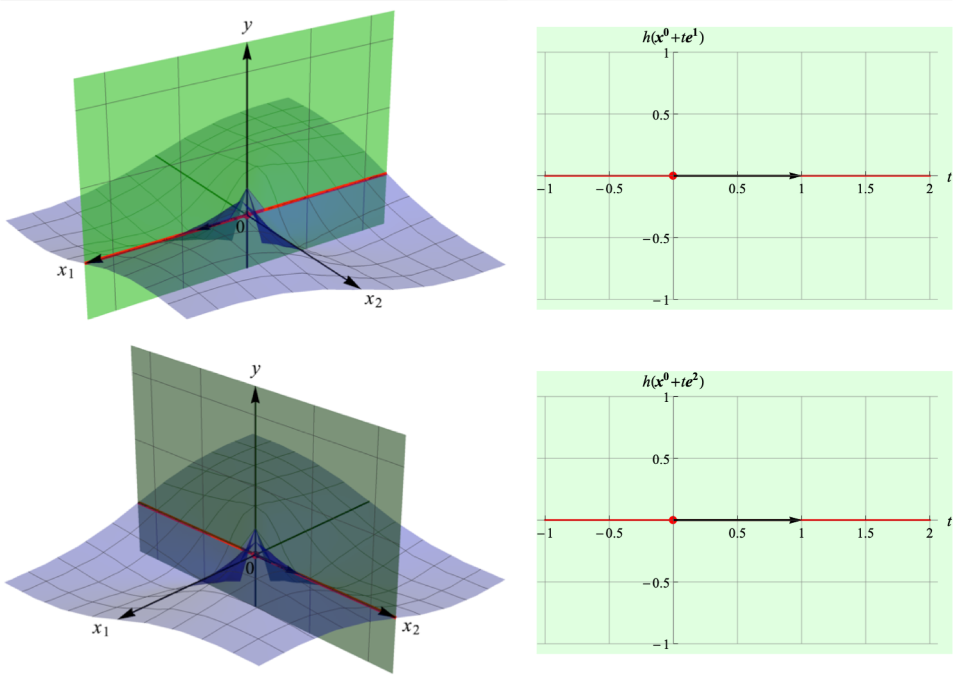

Abbildung 9.34: Senkrechte Vertikalschnitte von h durch $\mathbf{0}$.

In Beispiel 9.1.13 hatten wir gezeigt, dass h an der Stelle $\mathbf{0}$ nicht stetig ist. Dennoch ist die Funktion auch an dieser Stelle sowohl nach x_1 als auch nach x_2 partiell differenzierbar.

Dass dies kein Widerspruch ist, kann man sich wie folgt überlegen: Für festes $x_2 = 0$ ist h eine Funktion in x_1, die alle x_1 auf 0 abbildet, und damit insbesondere stetig. Ebenso ist die Funktion für festes $x_1 = 0$ eine Funktion in x_2, die alle x_2 auf 0 abbildet, und damit ebenfalls stetig, siehe Abbildung 9.34. Die Unstetigkeit an der Stelle $\mathbf{0}$ haben wir im Beispiel 9.1.13 gezeigt, indem man sich durch gleichzeitige Veränderung von x_1 und x_2 der Stelle $\mathbf{0}$ nähert. Partielle Ableitungen betrachten die Funktion nur für festes x_1 oder festes x_2. Somit kann die Funktion an der Stelle $\mathbf{0}$ partiell differenzierbar aber unstetig sein.

Somit kann man sogar an Unstetigkeitsstellen Tangentialebenen von h bilden. An der Stelle $\mathbf{0}$ ergibt sich hier beispielsweise

$$t_{1,\mathbf{0}}(x_1,x_2) = h(0,0) + h_{x_1}(0,0)(x_1-0) + h_{x_2}(0,0)(x_2-0) = 0.$$

Die Tangentialebene ist in Abbildung 9.35 visualisiert. ■

Wie im vorherigen Beispiel gesehen, ist nicht jede partiell differenzierbare Funktion auch stetig. Fordert man neben der Existenz der partiellen Ableitungen zusätzlich die Stetigkeit dieser partiellen Ableitungen, erhält man allerdings hinreichende Kriterien für die Stetigkeit. Folgende Definition und folgender Satz formalisieren dies:

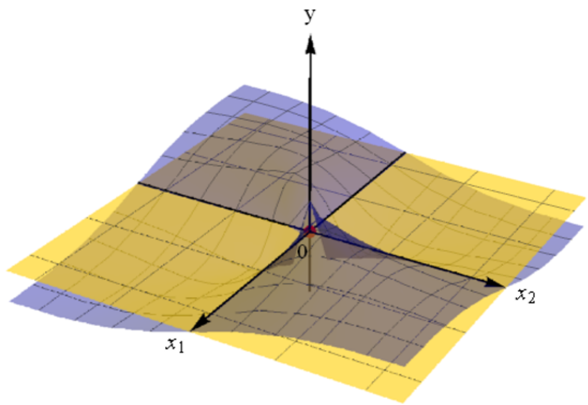

Abbildung 9.35: Die Tangentialebene von h durch $\mathbf{0}$.

Definition 9.2.4 — Stetige partielle Differenzierbarkeit.
Sei $D \subseteq \mathbb{R}^n$ offen und $f : D \to Z$ eine reelle Funktion in n Variablen. Sind die partiellen Ableitungen f_{x_i} für alle $i = 1, \ldots, n$ stetig an der Stelle $\mathbf{x}^0$, so heisst f stetig (partiell) differenzierbar in $\mathbf{x}^0$.

Sind die partiellen Ableitungen f_{x_i} für alle $i = 1, \ldots, n$ auf dem gesamten Definitionsbereich $D \subseteq \mathbb{R}^n$ stetig, so heisst f stetig (partiell) differenzierbar.

Satz 9.2.1 — Stetigkeit stetig partiell differenzierbarer Funktionen.
Eine stetig partiell differenzierbare reelle Funktion in n Variablen ist stetig.

Beweis. Einen Beweis findet man in Forster (1984, Kapitel 6, Satz 1). ■

Liest man obigen Satz als “f ist stetig partiell differenzierbar $\Rightarrow$ f ist stetig”, ist die Kontraposition offensichtlich “f ist nicht stetig $\Rightarrow$ f ist nicht stetig partiell differenzierbar”. Die Funktion aus Beispiel 9.2.9 ist nicht stetig und kann daher nicht stetig partiell differenzierbar sein. Wir zeigen dies in folgender Fortsetzung des Beispiels.

■ Beispiel 9.2.10 — Fortsetzung von Beispiel 9.2.9 (∗).
Wir betrachten erneut die Funktion

$$h(\mathbf{x}) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{für } \mathbf{x} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Wir zeigen, dass h_{x_1} an der Stelle $\mathbf{0}$ unstetig ist. Wäre h_{x_1} an der Stelle $\mathbf{0}$ stetig, würde es zu jedem $\varepsilon > 0$ ein $\delta > 0$ geben, so dass

$$|h_{x_1}(\mathbf{x}) - h_{x_1}(\mathbf{0})| = |h_{x_1}(\mathbf{x}) - 0| < \varepsilon \text{ für alle } \mathbf{x} \in U(\mathbf{0}, \delta).$$

Wir zeigen, dass es unter anderem für $\varepsilon = 1$ kein solches $\delta > 0$ gibt, indem wir erst zeigen, dass es kein $\delta \leq 1$ gibt und dann erklären, warum es auch kein $\delta > 1$ geben kann.

Für jedes $\delta \leq 1$ ist die Stelle $\mathbf{x} = (0, \frac{\delta}{2})^T$ nur $\frac{\delta}{2}$ von $\mathbf{0}$ entfernt. Die Stelle liegt also in der δ-Umgebung von $\mathbf{0}$ und es gilt $h_{x_1}(\mathbf{x}) = \frac{2}{\delta} \geq 2 > 1 = \varepsilon$. Für $\delta \leq 1$ kann man also in jeder solchen δ-Umgebung von $\mathbf{0}$ eine Stelle finden, für welche obige Ungleichung, welche man für Stetigkeit zeigen müsste, nicht erfüllt ist.

Da eine Umgebung $U(\mathbf{0},\delta)$ für $\delta > 1$ als Teilmenge u.a. die Umgebung $U(\mathbf{0},1)$ enthält und wir oben gezeigt haben, dass in jeder solchen Umgebung die Ungleichung an mindestens einer Stelle nicht erfüllt ist, kann es auch keine Umgebung $U(\mathbf{0},\delta)$ mit $\delta > 1$ geben, so dass die Ungleichung an allen Stellen der Umgebung erfüllt ist.

Es gibt also für $\varepsilon = 1$ keine δ-Umgebung mit $|h_{x_1}(\mathbf{x}) - 0| < 1 = \varepsilon$ für alle $\mathbf{x} \in U(\mathbf{0},\delta)$. Die partielle Ableitung ist nicht stetig. ■

Z Sind die partiellen Ableitungen einer partiell differenzierbaren reellen Funktion f in n Variablen für alle x_i, $i = 1,\ldots,n$, auf dem Definitionsbereich stetig, nennt man f auch stetig partiell differenzierbar.

Stetig partiell differenzierbare reelle Funktionen in mehreren Variablen sind auch stets stetig; partiell differenzierbare reelle Funktionen in mehreren Variablen sind nicht zwingend auch stetig.

9.2.3 Partielle Ableitungen höherer Ordnung und die Hesse-Matrix

Ziele dieses Unterkapitels

- Was versteht man unter partiellen Ableitungen zweiter Ordnung?
- Was ist die Hesse-Matrix?
- Was ist eine zweimal stetig partiell differenzierbare reelle Funktion in mehreren Variablen?
- Ist die Hesse-Matrix immer symmetrisch?

Ebenso wie man für reelle Funktionen in einer Variablen Ableitungen höherer Ordnung bilden kann, können partielle Ableitungen höherer Ordnung für reelle Funktionen in mehreren Variablen gebildet werden. Leitet man die partielle Ableitung nach x_i erneut nach x_i ab, so entsteht die zweite partielle Ableitung nach x_i. Leitet man die partielle Ableitung nach x_i nach einer anderen Variable x_j mit $j \neq i$ ab, so entsteht eine sogenannte gemischte Ableitung oder Kreuzableitung.

Definition 9.2.5 — Partielle Ableitung zweiter Ordnung und Hesse-Matrix. Sei f eine partiell differenzierbare reelle Funktion in n Variablen. Ist die partielle Ableitung nach x_i, $f_{x_i}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i}$, partiell differenzierbar nach x_j, so entstehen partielle Ableitungen zweiter Ordnung

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

Die Matrix, die alle partiellen Ableitungen zweiter Ordnung der Funktion an einer Stelle $\mathbf{x}^0$ enthält,

$$\nabla^2 f(\mathbf{x}^0) = H_f(\mathbf{x}^0) = \left(f_{x_i x_j}(\mathbf{x}^0)\right)_{ij} = \begin{pmatrix} f_{x_1x_1}(\mathbf{x}^0) & f_{x_1x_2}(\mathbf{x}^0) & \ldots & f_{x_1x_n}(\mathbf{x}^0) \\ f_{x_2x_1}(\mathbf{x}^0) & f_{x_2x_2}(\mathbf{x}^0) & \ldots & f_{x_2x_n}(\mathbf{x}^0) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1}(\mathbf{x}^0) & f_{x_nx_2}(\mathbf{x}^0) & \ldots & f_{x_nx_n}(\mathbf{x}^0) \end{pmatrix},$$

nennt man Hesse-Matrix der Funktion f an der Stelle $\mathbf{x}^0$.

Analog kann man auch partielle Ableitungen höherer Ordnung definieren. Wir berechnen die Hesse-Matrix einiger Funktionen an verschiedenen Stellen.

■ Beispiel 9.2.11 — Die Hesse-Matrix einer affin-linearen Funktion.
Für die affin-lineare Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$ wurden die ersten partiellen Ableitungen nach x_1 und x_2 in Beispiel 9.2.4 berechnet. Es gilt

$$f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} = -8 \quad \text{und} \quad f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2} = 3.$$

Die zweiten partiellen Ableitungen von $f(\mathbf{x})$ erhält man nun durch Ableiten der partiellen Ableitungen $f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1}$ und $f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2}$. Partielles Ableiten nach x_1 ergibt

$$f_{x_1x_1}(\mathbf{x}) = \frac{\partial f_{x_1}(\mathbf{x})}{\partial x_1} = \frac{\partial(-8)}{\partial x_1} = 0 \quad \text{und} \quad f_{x_2x_1}(\mathbf{x}) = \frac{\partial f_{x_2}(\mathbf{x})}{\partial x_1} = \frac{\partial(3)}{\partial x_1} = 0.$$

Leitet man $f_{x_1}(\mathbf{x})$ und $f_{x_2}(\mathbf{x})$ nach x_2 ab, ergibt sich

$$f_{x_1x_2}(\mathbf{x}) = \frac{\partial f_{x_1}(\mathbf{x})}{\partial x_2} = \frac{\partial(-8)}{\partial x_2} = 0 \quad \text{und} \quad f_{x_2x_2}(\mathbf{x}) = \frac{\partial f_{x_2}(\mathbf{x})}{\partial x_2} = \frac{\partial(3)}{\partial x_2} = 0.$$

Alle zweiten partiellen Ableitungen dieser affin-linearen Funktion sind 0, unabhängig von der betrachteten Stelle. Die Hesse-Matrix der Funktion fasst diese zweiten partiellen Ableitungen zusammen. Sie entspricht also an allen Stellen $\mathbf{x}$ der Nullmatrix:

$$H_f(\mathbf{x}) = \begin{pmatrix} f_{x_1x_1}(\mathbf{x}) & f_{x_1x_2}(\mathbf{x}) \\ f_{x_2x_1}(\mathbf{x}) & f_{x_2x_2}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

■

Bei allen affin-linearen Funktionen ist die Hesse-Matrix die Nullmatrix. Wir betrachten daher nun andere Typen von Funktionen:

■ Beispiel 9.2.12 — Die Hesse-Matrix einer quadratischen Funktion.
Als nächstes betrachten wir die quadratische Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ aus Beispiel 9.2.3. Auch in diesem Fall kennen wir bereits die ersten partiellen Ableitungen nach x_1 und x_2,

$$f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} = 2x_1 \quad \text{und} \quad f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2} = 8(x_2 - 1).$$

Daraus können wir wiederum die partiellen Ableitungen zweiter Ordnung berechnen. Zuerst leiten wir $f_{x_1}(\mathbf{x})$ und $f_{x_2}(\mathbf{x})$ partiell nach x_1 ab und erhalten

$$f_{x_1x_1}(\mathbf{x}) = \frac{\partial f_{x_1}(\mathbf{x})}{\partial x_1} = \frac{\partial(2x_1)}{\partial x_1} = 2 \quad \text{und} \quad f_{x_2x_1}(\mathbf{x}) = \frac{\partial f_{x_2}(\mathbf{x})}{\partial x_1} = \frac{\partial(8(x_2 - 1))}{\partial x_1} = 0.$$

Partielles Ableiten von $f_{x_1}(\mathbf{x})$ und $f_{x_2}(\mathbf{x})$ nach x_2 führt zu

$$f_{x_1x_2}(\mathbf{x}) = \frac{\partial f_{x_1}(\mathbf{x})}{\partial x_2} = \frac{\partial(2x_1)}{\partial x_2} = 0 \quad \text{und} \quad f_{x_2x_2}(\mathbf{x}) = \frac{\partial f_{x_2}(\mathbf{x})}{\partial x_2} = \frac{\partial(8(x_2 - 1))}{\partial x_2} = 8.$$

Auch hier ergeben sich zweite partielle Ableitungen, deren Werte unabhängig von der betrachteten Stelle sind. Die partiellen Ableitungen zweiter Ordnung der Funktion f an einer beliebigen Stelle $\mathbf{x} \in \mathbb{R}^2$ können wir in der Hesse-Matrix zusammenfassen:

$$H_f(\mathbf{x}) = \begin{pmatrix} f_{x_1x_1}(\mathbf{x}) & f_{x_1x_2}(\mathbf{x}) \\ f_{x_2x_1}(\mathbf{x}) & f_{x_2x_2}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

Die Hesse-Matrix dieser quadratischen Funktion ist eine Diagonalmatrix für alle $\mathbf{x} \in \mathbb{R}^2$. ∎

Auch bei einer quadratischen Funktion ist die Hesse-Matrix an jeder Stelle des Definitionsbereichs gleich. Im Allgemeinen ist dies jedoch nicht der Fall:

∎ Beispiel 9.2.13 — Die Hesse-Matrix der Cobb Douglas Funktion.
Aus Beispiel 9.2.2 wissen wir, dass die ersten partiellen Ableitungen der Cobb Douglas Funktion $f(\mathbf{x}) = x_1^{0.8}x_2^{0.2}$ als

$$f_{x_1}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} = 0.8\frac{x_2^{0.2}}{x_1^{0.2}} \quad \text{und} \quad f_{x_2}(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2} = 0.2\frac{x_1^{0.8}}{x_2^{0.8}}$$

gegeben sind. Berechnen wir hieraus die zweiten partiellen Ableitungen, erhalten wir durch partielles Ableiten von $f_{x_1}(\mathbf{x})$ und $f_{x_2}(\mathbf{x})$ nach x_1

$$f_{x_1x_1}(\mathbf{x}) = \frac{\partial f_{x_1}(\mathbf{x})}{\partial x_1} = \frac{\partial\left(0.8\frac{x_2^{0.2}}{x_1^{0.2}}\right)}{\partial x_1} = 0.8\cdot(-0.2)\frac{x_2^{0.2}}{x_1^{1.2}} = -0.16\frac{x_2^{0.2}}{x_1^{1.2}}$$

sowie

$$f_{x_2x_1}(\mathbf{x}) = \frac{\partial f_{x_2}(\mathbf{x})}{\partial x_1} = \frac{\partial\left(0.2\frac{x_1^{0.8}}{x_2^{0.8}}\right)}{\partial x_1} = \frac{0.16}{x_1^{0.2}x_2^{0.8}}.$$

Partielles Ableiten von $f_{x_1}(\mathbf{x})$ und $f_{x_2}(\mathbf{x})$ nach x_2 führt zu

$$f_{x_1x_2}(\mathbf{x}) = \frac{\partial f_{x_1}(\mathbf{x})}{\partial x_2} = \frac{\partial\left(0.8\frac{x_2^{0.2}}{x_1^{0.2}}\right)}{\partial x_2} = \frac{0.16}{x_1^{0.2}x_2^{0.8}}$$

und

$$f_{x_2x_2}(\mathbf{x}) = \frac{\partial f_{x_2}(\mathbf{x})}{\partial x_2} = \frac{\partial\left(0.2\frac{x_1^{0.8}}{x_2^{0.8}}\right)}{\partial x_2} = -0.16\frac{x_1^{0.8}}{x_2^{1.8}}.$$

Somit lautet die Hesse-Matrix an der Stelle $\mathbf{x}^0 = (1,1)^T$ beispielsweise

$$H_f(1,1) = \begin{pmatrix} f_{x_1x_1}(1,1) & f_{x_1x_2}(1,1) \\ f_{x_2x_1}(1,1) & f_{x_2x_2}(1,1) \end{pmatrix} = \begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix}.$$

An der Stelle $\mathbf{x}^0 = (32,32)^T$ ist die Hesse-Matrix

$$H_f(32,32) = \begin{pmatrix} f_{x_1x_1}(32,32) & f_{x_1x_2}(32,32) \\ f_{x_2x_1}(32,32) & f_{x_2x_2}(32,32) \end{pmatrix} = \begin{pmatrix} -0.005 & 0.005 \\ 0.005 & -0.005 \end{pmatrix}.$$

An der Stelle $\mathbf{x}^0 = (1,32)^T$ ist die Hesse-Matrix

$$H_f(1,32) = \begin{pmatrix} f_{x_1x_1}(1,32) & f_{x_1x_2}(1,32) \\ f_{x_2x_1}(1,32) & f_{x_2x_2}(1,32) \end{pmatrix} \approx \begin{pmatrix} -0.32 & 0.01 \\ 0.01 & -0.0003 \end{pmatrix}.$$

Berechnet man die Hesse-Matrix an verschiedenen Stellen $\mathbf{x}^0$, ergeben sich also verschiedene Matrizen. ■

Obwohl auch reelle Funktionen in mehreren Variablen unstetig oder nicht differenzierbar sein können, beschränken wir uns bei der Diskussion von Eigenschaften reeller Funktionen in n Variablen und ihrer Optimierung auf Funktionen mit stetigen ersten und zweiten partiellen Ableitungen. Wie im Fall von reellen Funktionen in einer Variable, nennt man derartige Funktionen auch zweimal stetig differenzierbar.

Definition 9.2.6 — Zweimal stetig differenzierbare Funktionen.
Sei $D \subseteq \mathbb{R}^n$ offen und sei $f : D \to Z$ eine stetig partiell differenzierbare reelle Funktion in n Variablen. Sind die zweiten partiellen Ableitungen von f für alle $i, j = 1, \ldots, n$ stetig, so heisst f zweimal stetig (partiell) differenzierbar.

Für zweimal stetig partiell differenzierbare reelle Funktionen in n Variablen ist die Reihenfolge, in der die Funktion zweimal partiell differenziert wird, unerheblich. In diesem Fall ist die Hesse-Matrix wie in obigen Beispielen symmetrisch.

Satz 9.2.2 — Symmetrie der Hesse-Matrix (Satz von Schwarz).
Sei $D \subseteq \mathbb{R}^n$ offen. Ist die reelle Funktion $f : D \to Z$ in n Variablen zweimal stetig differenzierbar, dann ist die Hesse-Matrix $H_f(\mathbf{x}^0)$ von f an jeder Stelle $\mathbf{x}^0 \in D$ symmetrisch.

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.7). ■

Die Hesse-Matrizen aus obigen Beispielen sind alle symmetrisch, da alle Funktionen zweimal stetig partiell differenzierbar sind. Sind die zweiten partiellen Ableitungen einer Funktion nicht stetig, muss die Hesse-Matrix nicht symmetrisch sein. Wir demonstrieren zur Vollständigkeit ein solches Beispiel.

■ **Beispiel 9.2.14 — Ein pathologisches Beispiel (∗).**
Ein oft zitiertes und von Hermann Amandus Schwarz selbst erwähntes Beispiel ist die Funktion $h : \mathbb{R}^2 \to \mathbb{R}$ mit

$$h(\mathbf{x}) = \begin{cases} \dfrac{x_1x_2(x_1^2 - x_2^2)}{x_1^2 + x_2^2} & \text{für } \mathbf{x} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Für festes $x_2 \neq 0$ ist die partielle Ableitung nach x_1 mithilfe der Quotientenregel ermittelbar. Für festes $x_2 = 0$ ist die Funktion für alle x_1 gleich 0, somit ist die partielle Ableitung nach x_1 für $x_2 = 0$ gleich 0. Zusammenfassend ergibt sich

$$h_{x_1}(\mathbf{x}) = \begin{cases} \dfrac{x_2(x_1^4 + 4x_1^2x_2^2 - x_2^4)}{(x_1^2 + x_2^2)^2} & \text{für } \mathbf{x} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Betrachten wir diese partielle Ableitung für festes x_1 und leiten partiell nach x_2 ab, so erhalten wir mithilfe der Quotientenregel

$$h_{x_1x_2}(\mathbf{x}) = \frac{\partial^2 h(\mathbf{x})}{\partial x_2 \partial x_1} = \frac{x_1^6 - x_2^6 - 9x_1^2x_2^4 + 9x_1^4x_2^2}{\left(x_1^2+x_2^2\right)^3} \qquad \text{für } x_1 \neq 0.$$

Anschaulich beschreibt dieser Ausdruck, wie sich die Veränderung in x_1, also die partielle Ableitung nach x_1, verändert, wenn sich nur x_2 verändert.

Für $x_1 = 0$ gilt

$$h_{x_1}(0,x_2) = \left\{\begin{array}{ll} \dfrac{x_2(-x_2^4)}{(x_2^2)^2} = -x_2 & \text{für } x_2 \in \mathbb{R}\setminus\{0\} \\ \qquad 0 & \text{für } x_2 = 0. \end{array}\right\} = -x_2.$$

Damit ist die partielle Ableitung von $h_{x_1}(\mathbf{x})$ für $x_1 = 0$ nach x_2 stets -1. Zusammenfassend gilt

$$h_{x_1x_2}(\mathbf{x}) = \begin{cases} \frac{x_1^6-x_2^6-9x_1^2x_2^4+9x_1^4x_2^2}{\left(x_1^2+x_2^2\right)^3} & \text{für } \mathbf{x} \in \mathbb{R}^2\setminus\{\mathbf{0}\} \\ -1 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Betrachten wir nun die Funktion h für festes x_1, um die partiellen Ableitungen nach x_2 zu bilden. Im Fall $x_1 = 0$ gilt stets $h(0,x_2) = 0$. Für $x_1 \neq 0$ ist die partielle Ableitung nach x_2 mithilfe der Quotientenregel ermittelbar. Es ergibt sich

$$h_{x_2}(\mathbf{x}) = \begin{cases} \dfrac{-x_1(x_2^4+4x_1^2x_2^2-x_1^4)}{(x_1^2+x_2^2)^2} & \text{für } \mathbf{x} \in \mathbb{R}^2\setminus\{\mathbf{0}\} \\ \qquad 0 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

Für $x_2 = 0$ gilt

$$h_{x_2}(x_1,0) = \left\{\begin{array}{ll} \dfrac{-x_1(-x_1^4)}{(x_1^2)^2} = x_1 & \text{für } x_1 \in \mathbb{R}\setminus\{0\} \\ \qquad 0 & \text{für } x_1 = 0. \end{array}\right\} = x_1.$$

Damit ist die partielle Ableitung von $h_{x_2}(\mathbf{x})$ für $x_2 = 0$ nach x_1 stets 1 und

$$h_{x_2x_1}(\mathbf{x}) = \begin{cases} \frac{x_1^6-x_2^6-9x_1^2x_2^4+9x_1^4x_2^2}{\left(x_1^2+x_2^2\right)^3} & \text{für } \mathbf{x} \in \mathbb{R}^2\setminus\{\mathbf{0}\} \\ 1 & \text{für } \mathbf{x} = \mathbf{0}. \end{cases}$$

An der Stelle $\mathbf{x}^0 = \mathbf{0}$ gilt also, dass

$$-1 = \frac{\partial^2 h(\mathbf{0})}{\partial x_2 \partial x_1} \neq \frac{\partial^2 h(\mathbf{0})}{\partial x_1 \partial x_2} = 1.$$

Die Hesse-Matrix

$$H_f(\mathbf{0}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

ist nicht symmetrisch. Dies ist möglich, da f nicht zweimal stetig partiell differenzierbar ist. ∎

Z Bildet man die partielle Ableitung von f_{x_i} nach x_j ergibt sich die partielle Ableitung zweiter Ordnung $f_{x_i x_j}$. Die so entstehende Matrix aller partiellen Ableitungen zweiter Ordnung $H_f(\mathbf{x}^0) = (f_{x_i x_j}(\mathbf{x}^0))_{ij}$ nennt man die Hesse-Matrix der Funktion f.

Sind die zweiten partiellen Ableitungen einer reellen Funktion f in mehreren Variablen stetig, heisst f zweimal stetig partiell differenzierbar. Die Hesse-Matrix einer zweimal stetig partiell differenzierbaren Funktion ist an allen Stellen $\mathbf{x}^0$ des Definitionsbereichs symmetrisch. Für allgemeine Funktionen ist die Hesse-Matrix aber nicht immer symmetrisch.

9.2.4 Taylorpolynome ersten und zweiten Grades

Ziele dieses Unterkapitels

- Welche Gleichung beschreibt das Taylorpolynom erster Ordnung von f an der Stelle $\mathbf{x}^0$?
- Welche Gleichung beschreibt das Taylorpolynom zweiter Ordnung von f an der Stelle $\mathbf{x}^0$?

Die Tangentialhyperebene von f an der Stelle $\mathbf{x}^0$ wird laut Definition 9.2.3 von einer affin-linearen Funktion in mehreren Variablen beschrieben, welche die gleichen (ersten) partiellen Ableitungen wie die Funktion f an der Stelle $\mathbf{x}^0$ hat. Diese Funktion bezeichnet man auch als Taylorpolynom erster Ordnung.

Ebenso wie wir für reelle Funktionen in einer Variablen das Taylorpolynom erster Ordnung mithilfe der zweiten Ableitung zum Taylorpolynom zweiter Ordnung verallgemeinert haben, kann man auch für reelle Funktionen in mehreren Variablen mithilfe der zweiten partiellen Ableitungen Taylorpolynome zweiter Ordnung definieren. Das Taylorpolynom zweiter Ordnung von f an der Stelle $\mathbf{x}^0$ hat dabei die gleichen ersten und zweiten partiellen Ableitungen wie die Funktion f an dieser Stelle.

Definition 9.2.7 — Taylorpolynome ersten und zweiten Grades. Sei $D \subseteq \mathbb{R}^n$ offen und sei $f : D \to Z$ eine reelle Funktion in n Variablen, die an der Stelle $\mathbf{x}^0 \in D$ partiell differenzierbar ist. Die Funktion

$$t_{1,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T (\mathbf{x} - \mathbf{x}^0)$$

heisst Taylorpolynom erster Ordnung von f an der Stelle $\mathbf{x}^0$.

Existieren an der Stelle $\mathbf{x}^0 \in D$ die partiellen Ableitungen zweiter Ordnung von f, so heisst die Funktion

$$t_{2,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T (\mathbf{x} - \mathbf{x}^0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T H_f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0)$$

Taylorpolynom zweiter Ordnung von f an der Stelle $\mathbf{x}^0$.

Die Taylorpolynome an der Stelle $\mathbf{x}^0$ sind dabei so definiert, dass an der Stelle $\mathbf{x}^0$

$$\begin{aligned}
f(\mathbf{x}^0) &= t_{1,\mathbf{x}^0}(\mathbf{x}^0) = t_{2,\mathbf{x}^0}(\mathbf{x}^0), \\
\nabla f(\mathbf{x}^0) &= \nabla t_{1,\mathbf{x}^0}(\mathbf{x}^0) = \nabla t_{2,\mathbf{x}^0}(\mathbf{x}^0), \\
H_f(\mathbf{x}^0) &= H_{t_{2,\mathbf{x}^0}}(\mathbf{x}^0)
\end{aligned}$$

gilt. Das Taylorpolynom erster Ordnung hat also den gleichen Funktionswert und gleiche erste partielle Ableitungen wie f an der Stelle $\mathbf{x}^0$. Das Taylorpolynom zweiter Ordnung hat zusätzlich gleiche zweite partielle Ableitungen an der Stelle $\mathbf{x}^0$. Ebenso wie im Fall reeller Funktionen in einer Variablen kann man auch in diesem allgemeineren Fall Abschätzungen bezüglich des sogenannten Restglieds $f(\mathbf{x}) - t_{1,\mathbf{x}^0}(\mathbf{x})$ bzw. $f(\mathbf{x}) - t_{2,\mathbf{x}^0}(\mathbf{x})$ machen. Wir verzichten an dieser Stelle jedoch darauf, dies zu spezifizieren.

■ Beispiel 9.2.15 — Taylorpolynome ersten und zweiten Grades.
Betrachten wir die Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^2 + 3$ an der Stelle $\mathbf{x}^0 = (1,2)^T$, so gilt dort

$$f(\mathbf{x}^0) = 8 \text{ und } \nabla f(\mathbf{x}^0) = \begin{pmatrix} 2x_1^0 \\ 8(x_2^0 - 1) \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}.$$

Das Taylorpolynom erster Ordnung von f an der Stelle $\mathbf{x}^0 = (1,2)^T$ lautet demnach

$$\begin{aligned} t_{1,\mathbf{x}^0}(\mathbf{x}) &= f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0) \\ &= 8 + (2,8)\begin{pmatrix} x_1 - 1 \\ x_2 - 2 \end{pmatrix} \\ &= 8 + 2(x_1 - 1) + 8(x_2 - 2) = 2x_1 + 8x_2 - 10. \end{aligned}$$

Dieses Taylorpolynom erster Ordnung hat den gleichen Funktionswert und die gleichen ersten partiellen Ableitungen wie die Funktion f an der Stelle $\mathbf{x}^0 = (1,2)^T$. Das Taylorpolynom erster Ordnung von f an der Stelle $\mathbf{x}^0 = (0,1)^T$ ist hingegen

$$\begin{aligned} t_{1,\mathbf{x}^0}(\mathbf{x}) &= f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0) \\ &= 3 + (0,0)\begin{pmatrix} x_1 - 0 \\ x_2 - 1 \end{pmatrix} = 3. \end{aligned}$$

Abbildung 9.36 zeigt die beiden durch die Taylorpolynome beschriebenen Tangentialebenen an den Graphen von f.[6] Bildet man das Taylorpolynom zweiter Ordnung an der Stelle $\mathbf{x}^0 = (1,2)^T$, so addiert man den Term $\frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T H_f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0)$ zum Taylorpolynom erster Ordnung an der Stelle $\mathbf{x}^0 = (1,2)^T$. Es gilt damit

$$\begin{aligned} t_{2,\mathbf{x}^0}(\mathbf{x}) &= f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T H_f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0) \\ &= 8 + (2,8)\begin{pmatrix} x_1 - 1 \\ x_2 - 2 \end{pmatrix} + \frac{1}{2}(x_1 - 1, x_2 - 2)\begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}\begin{pmatrix} x_1 - 1 \\ x_2 - 2 \end{pmatrix} \\ &= 8 + 2(x_1 - 1) + 8(x_2 - 2) + \frac{1}{2}2(x_1 - 1)^2 + \frac{1}{2}8(x_2 - 2)^2 \\ &= 8 + 2x_1 - 2 + 8x_2 - 16 + x_1^2 - 2x_1 + 1 + 4x_2^2 - 16x_2 + 16 \\ &= x_1^2 + 4x_2^2 - 8x_2 + 7 = x_1^2 + 4(x_2^2 - 2x_2) + 7 \\ &= x_1^2 + 4(x_2^2 - 2x_2 + 1) - 4 + 7 = x_1^2 + 4(x_2^2 - 2x_2 + 1) + 3 \\ &= x_1^2 + 4(x_2 - 1)^2 + 3. \end{aligned}$$

Das Taylorpolynom zweiten Grades entspricht also f. Dies ist nicht überraschend, da f selbst ein Polynom zweiten Grades ist. ■

[6] Aufgrund seiner konvexen Form verlaufen beide Tangentialebenen unterhalb des Graphen von f. Wir diskutieren Konvexität von Funktionen mit mehreren Variablen in einem späteren Abschnitt.

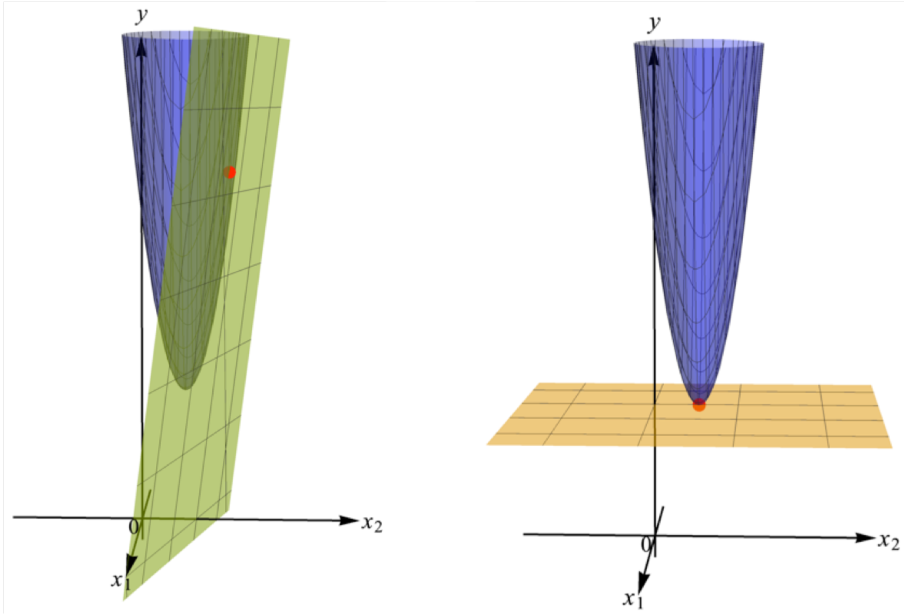

Abbildung 9.36: Tangentialebene der quadratischen Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ an den Stellen $\mathbf{x}^0 = (1,2)^T$ und $\mathbf{x}^0 = (0,1)^T$.

Quadratische Funktionen sind Polynome zweiten Grades. Taylorpolyonome zweiten Grades einer quadratischen Funktion entsprechen stets der Funktion selbst. Ebenso entsprechen Taylorpolynome ersten Grades einer affin-linearen Funktion stets der Funktion selbst. Im Allgemeinen entsprechen Taylorpolynome zweiten Gerades aber nicht der zugrundeliegenden Funktion. Wir illustrieren dies an folgendem Beispiel.

■ **Beispiel 9.2.16 — Taylorpolynome der Cobb Douglas Funktion.**
Wir betrachten erneut die Cobb Douglas Funktion mit $f(\mathbf{x}) = x_1^{0.8} x_2^{0.2}$ aus Beispiel 9.2.13. Für diese Funktion haben wir sowohl den Gradienten als auch die Hesse-Matrix bereits bestimmt. An der Stelle $\mathbf{x}^0 = (1,1)^T$ gilt

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} \text{ und } H_f(\mathbf{x}^0) = \begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix}.$$

Das Taylorpolynom ersten Grades an der Stelle $\mathbf{x}^0 = (1,1)^T$ ist damit gegeben durch

$$t_{1,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0) = 1 + (0.8, 0.2) \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} = 0.8x_1 + 0.2x_2$$

und beschreibt die Tangentialebene an dieser Stelle. Das Taylorpolynom zweiten Grades an der Stelle $\mathbf{x}^0 = (1,1)^T$ ist gegeben durch

$$\begin{aligned} t_{2,\mathbf{x}^0}(\mathbf{x}) &= f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T H_f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0) \\ &= 1 + (0.8, 0.2) \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} + \frac{1}{2}(x_1 - 1, x_2 - 1) \begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} \\ &= 0.8x_1 + 0.2x_2 - 0.08x_1^2 - 0.08x_2^2 + 0.16x_1x_2. \end{aligned}$$

Das Taylorpolynom zweiten Grades hat an der Stelle $\mathbf{x}^0 = (1,1)^T$ denselben Funktionswert wie f. Zudem stimmen an dieser Stelle die partiellen Ableitungen erster und zweiter Ordnung mit jenen der Funktion f überein. Es gilt

$$\begin{aligned}
f(\mathbf{x}^0) &= 1 = t_{1,\mathbf{x}^0}(\mathbf{x}^0) = t_{2,\mathbf{x}^0}(\mathbf{x}^0) \\
\nabla f(\mathbf{x}^0) &= \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} = \nabla t_{1,\mathbf{x}^0}(\mathbf{x}^0) = \nabla t_{2,\mathbf{x}^0}(\mathbf{x}^0) \\
H_f(\mathbf{x}^0) &= \begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix} = H_{t_{2,\mathbf{x}^0}}(\mathbf{x}^0).
\end{aligned}$$

Das Taylorpolynom zweiten Grades entspricht aber nicht der zugrundeliegenden Funktion f. ■

Z Existieren die ersten partiellen Ableitungen von f an der Stelle $\mathbf{x}^0$, dann heisst

$$t_{1,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0)$$

Taylorpolynom erster Ordnung von f an der Stelle $\mathbf{x}^0$. Existieren auch die zweiten partiellen Ableitungen von f an der Stelle $\mathbf{x}^0$, dann heisst

$$t_{2,\mathbf{x}^0}(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T(\mathbf{x} - \mathbf{x}^0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T H_f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0)$$

Taylorpolynom zweiter Ordnung von f an der Stelle $\mathbf{x}^0$.

9.2.5 Das totale Differential

Ziele dieses Unterkapitels

- Was versteht man unter dem totalen Differential einer reellen Funktion f in $\mathbf{x}^0$?

Analog zur Bezeichnung des Differentials für differenzierbare Funktionen in einer Variablen nutzt man die Bezeichnung totales Differential für stetig partiell differenzierbare reelle Funktionen, um die Veränderung des Funktionswertes anzugeben, welcher sich auf die ersten partiellen Ableitungen zurückführen lässt.

■ Beispiel 9.2.17 — Das totale Differential einer Cobb Douglas Funktion.

$$\Delta f(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}^0) = 33^{0.2}33^{0.8} - 32^{0.2}32^{0.8} = 1.$$

Die erste partielle Ableitung nach x_1 an der Stelle $\mathbf{x}^0$ ist 0.8. Die erste partielle Ableitung nach x_1 lässt also vermuten, dass sich aufgrund der Vergrösserung von x_1 um 1 der Funktionswert um $\frac{\partial f(\mathbf{x}^0)}{\partial x_1}\Delta x_1 = 0.8 \cdot 1 = 0.8$ vergrössert. Die erste partielle Ableitung nach x_2 an der Stelle $\mathbf{x}^0$ ist 0.2. Die durch diese Ableitung erklärte Differenz der Funktionswerte aufgrund der Veränderung von x_2 um Δx_2 ist $\frac{\partial f(\mathbf{x}^0)}{\partial x_2}\Delta x_2 = 0.2$. Zusammengenommen ergibt sich das totale Differential als

$$\nabla f(\mathbf{x}^0)^T \Delta\mathbf{x} = (0.8, 0.2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0.8 \cdot 1 + 0.2 \cdot 1 = 1.$$

Mit diesen Zahlenwerten ist das totale Differential gleich der Differenz der Funktionswerte. In der Regel beschreibt das totale Differential aber nur näherungsweise die Funktionswertveränderung (für kleine Veränderungen $\Delta\mathbf{x}$). Beispielsweise ist

$$\nabla f(1,32) = \begin{pmatrix} 0.8\frac{32^{0.2}}{1^{0.2}} \\ 0.2\frac{1^{0.8}}{32^{0.8}} \end{pmatrix} = \begin{pmatrix} \frac{8}{5} \\ \frac{1}{80} \end{pmatrix}.$$

Verändert man ausgehend von $\mathbf{x}^0 = (1,32)^T$ den Wert von x_1 um 1 und lässt x_2 konstant, so ist $\Delta\mathbf{x} = (1,0)^T$ und

$$\nabla f(1,32)^T \Delta\mathbf{x} = \left(\frac{8}{5}, \frac{1}{80}\right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{8}{5} = 1.6.$$

Die Veränderung des Funktionswerts ist jedoch

$$f(2,32) - f(1,32) = 2^{0.8}32^{0.2} - 1^{0.8}32^{0.2} = 2^{0.8}2 - 2 \approx 1.48 \neq 1.6.$$

■

Definition 9.2.8 — Totales Differential.
Eine stetig partiell differenzierbare reelle Funktion f in n Variablen nennt man auch total differenzierbar. Für $\Delta\mathbf{x} \in \mathbb{R}^n$ heisst dann der Ausdruck $\nabla f(\mathbf{x}^0)^T \Delta\mathbf{x}$ auch totales Differential von f in $\mathbf{x}^0$.

■ **Beispiel 9.2.18 — Eine nicht total differenzierbare Funktion.**
Da die Funktion h aus Beispiel 9.1.13 an der Stelle $\mathbf{0}$ nicht stetig ist, kann sie auch beliebig nahe der Stelle $\mathbf{0}$ vom Funktionswert $h(\mathbf{0}) = 0$ sehr verschiedene Werte annehmen. Es erscheint daher trotz eines Gradienten von

$$\nabla h(\mathbf{0}) = \mathbf{0}$$

wenig sinnvoll, Funktionswerte von Stellen nahe der Stelle $\mathbf{0}$ mithilfe von $\nabla h(\mathbf{0})^T \Delta\mathbf{x} = 0$ zu approximieren. Aus diesem Grund spricht man in diesem Fall auch nicht von einem totalen Differential. Die Funktion ist nicht total differenzierbar. ■

Z Ist f eine stetig partiell differenzierbare reelle Funktion, bezeichnet man $\nabla f(\mathbf{x}_0)^T \Delta\mathbf{x}$ auch als totales Differential von f in $\mathbf{x}^0$.

9.2.6 Totale Differenzierbarkeit und die Kettenregel (#)

Ziele dieses Unterkapitels

- Was bedeutet die Eigenschaft der totalen Differenzierbarkeit für den Zusammenhang zwischen dem totalen Differential und der Veränderung des Funktionswerts?
- Was besagt die verallgemeinerte Kettenregel?

Eigentlich ist der Begriff der totalen Differenzierbarkeit deutlich allgemeiner als in Definition 9.2.8 angedeutet. Er umfasst alle Funktionen, welche die im folgenden Satz beschriebene Eigenschaft besitzen. Diese Eigenschaft bezeichnet man als totale Differenzierbarkeit und trifft insbesondere auf stetig partiell differenzierbare reelle Funktionen zu:

Satz 9.2.3 — Totale Differenzierbarkeit stetig partiell differenzierbarer Funktionen. Sei $D \subseteq \mathbb{R}^n$ offen und $\mathbf{x}^0 \in D$. Ist $f : D \to Z$ eine stetig partiell differenzierbare reelle Funktion in n Variablen, dann gibt es eine reellwertige Funktion r und ein $\delta > 0$, so dass

$$f(\mathbf{x}^0 + \Delta\mathbf{x}) - f(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)^T \Delta\mathbf{x} + r(\Delta\mathbf{x}) \text{ für alle } \mathbf{x} \in U(\mathbf{x}^0, \delta)$$

$$\text{mit } \lim_{\Delta\mathbf{x} \to \mathbf{0}} \frac{r(\Delta\mathbf{x})}{||\Delta\mathbf{x}||} = 0.$$

Beweis. Einen Beweis findet man in Forster (1984, Kapitel 6, Satz 2). ■

In einer genügend kleinen Umgebung einer Stelle $\mathbf{x}^0$ ist die Veränderung des Funktionswerts einer stetig partiell differenzierbaren reellen Funktion also für kleine $\Delta\mathbf{x}$ näherungsweise gleich dem totalen Differential.

Für unsere Belange ist jedoch die Gleichsetzung von stetig partiell differenzierbaren und total differenzierbaren Funktionen völlig ausreichend.

Eine wichtige Rechenregel für partiell differenzierbare reelle Funktionen ist die sogenannte verallgemeinerte Kettenregel, welche wir im Folgenden motivieren wollen. Es handelt sich wie schon bei der bekannten Kettenregel der Differentialrechnung in einer Variablen um eine Rechenregel zur Bildung der Ableitung einer Komposition zweier Funktionen f und g. Im Gegensatz zur bekannten Rechenregel ist g nun allerdings eine Funktion, welche eine reelle Zahl auf einen Vektor mit n Komponenten abbildet, und f eine reelle Funktion in n Variablen. So wird durch die Komposition $f \circ g$ also eine reelle Zahl $t \in \mathbb{R}$ zunächst auf einen Vektor $g(t) \in \mathbb{R}^n$ abgebildet und dieser dann durch f wieder auf eine reelle Zahl $f(g(t)) \in \mathbb{R}$, vgl. Abbildung 9.37.

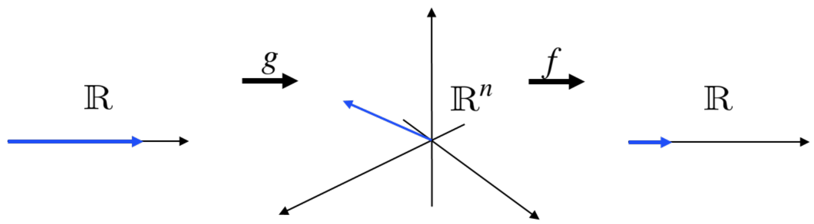

Abbildung 9.37: Darstellung der Verknüpfung von f und g.

■ Beispiel 9.2.19 — Die Ableitung verketteter Funktionen.
Prognosen für einen kleinen Flughafen, der nur von zwei Fluggesellschaften angeflogen wird, zeigen, dass das Verkehrsaufkommen der einen Fluggesellschaft quadratisch, das der anderen linear über die kommenden Jahre steigen wird. Die Funktion, welche die Zeit auf die Nachfrage abbildet, sei $g : \mathbb{R} \to \mathbb{R}^2$ mit $g(t) = (g_1(t), g_2(t)) = (t^2, 3t)$. Bei gegebener Nachfrage $\mathbf{x}$ entstehen Kosten von $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$.

Die Kosten zur Zeit t sind also mit $x_1 = g_1(t) = t^2$ und $x_2 = g_2(t) = 3t$

$$h(t) = (f \circ g)(t) = (g_1(t))^2 + 4(g_2(t) - 1)^2 + 3 = (t^2)^2 + 4(3t-1)^2 + 3 = t^4 + 36t^2 - 24t + 7.$$

Die Ableitung der Kosten nach der Zeit t ist damit

$$h'(t) = 4t^3 + 72t - 24.$$

Wir versuchen nun dieses Ergebnis zu erhalten ohne die Funktion $h(t) = t^4 + 36t^2 - 24t + 7$ bestimmen zu müssen. Hierfür nutzen wir, dass die partiellen Ableitungen von f nach x_1 und x_2 bereits bekannt sind. Es gilt

$$f_{x_1}(\mathbf{x}) = 2x_1 \text{ und } f_{x_2}(\mathbf{x}) = 8(x_2 - 1).$$

Mit $x_1 = g_1(t) = t^2$ und $x_2 = g_2(t) = 3t$ erkennt man, dass eine kleine Änderung von t die Werte von x_1 und x_2 unterschiedlich stark verändert: x_1 verändert sich mit Rate $g_1'(t) = 2t$ und x_2 mit Rate $g_2'(t) = 3$. Gewichtet man die partiellen Ableitungen entsprechend und beschreibt diese in Abhängigkeit von t, also als

$$f_{x_1}(g(t)) = 2t^2 \text{ und } f_{x_2}(g(t)) = 8(3t - 1),$$

ergibt sich das gleiche Ergebnis wie oben:

$$h'(t) = f_{x_1}(g(t)) \cdot \underbrace{g_1'(t)}_{\substack{\text{Veränderung von } x_1 \\ \text{in Abhängigkeit von } t}} + f_{x_2}(g(t)) \cdot \underbrace{g_2'(t)}_{\substack{\text{Veränderung von } x_2 \\ \text{in Abhängigkeit von } t}} = 2t^2 \cdot 2t + 8(3t - 1) \cdot 3 = 4t^3 + 72t - 24.$$

■

Sind die Ableitungen der Funktionen f und g bekannt, kann man die Ableitung der Komposition $h = f \circ g$ also auch direkt berechnen, ohne die Komposition zu bestimmen. Hierfür gewichtet man die partiellen Ableitungen von f an der Stelle $g(t)$ mit den entsprechenden Ableitungen von g. Es gilt:

Satz 9.2.4 — Verallgemeinerte Kettenregel.
Es seien $D \subseteq \mathbb{R}$ ein offenes Intervall, $Z_i \subseteq \mathbb{R}$ für $i = 1, \dots, n$, $g : D \to Z_1 \times \cdots \times Z_n \subseteq \mathbb{R}^n$ eine Funktion mit Abbildungsvorschrift

$$g(t) = \begin{pmatrix} g_1(t) \\ \vdots \\ g_n(t) \end{pmatrix},$$

wobei die Funktionen $g_i : D \to Z_i$ an der Stelle $t_0 \in D$ differenzierbare Funktionen sind, und $f : g(D) \to Z \subseteq \mathbb{R}$ eine total differenzierbare Funktion. Dann ist die Komposition $h = f \circ g : D \to Z$ an der Stelle t_0 differenzierbar und hat die erste Ableitung

$$\begin{aligned} h'(t_0) = (f \circ g)'(t_0) = \nabla f(g(t_0))^T \begin{pmatrix} g_1'(t_0) \\ \vdots \\ g_n'(t_0) \end{pmatrix} &= \sum_{i=1}^{n} f_{x_i}(g(t_0)) \cdot g_i'(t_0) \\ &= f_{x_1}(g(t_0)) \cdot g_1'(t_0) + \cdots + f_{x_n}(g(t_0)) \cdot g_n'(t_0). \end{aligned}$$

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.5). ■

Diese Regel nennt man auch verallgemeinerte Kettenregel, da sich die bekannte Kettenregel $f(g(x))' = f'(g(x))g'(x)$ im Fall $n = 1$ ergibt.

■ Beispiel 9.2.20 — Veranschaulichung der Verkettung.
In obigem Beispiel erhält man durch g zu jedem t ein Tupel $x_1 = g_1(t) = t^2$ und $x_2 = g_2(t) = 3t$ im $\mathbb{R}^2$. Für $t \in \mathbb{R}$ ergibt sich so ein Weg im $\mathbb{R}^2$. Durch die Komposition mit f werden nur Funktionswerte $f(g(t))$ dieser Tupel ausgewertet. Die Komposition beschreibt also die Funktionswerte von f entlang des durch g vorgegebenen Weges, vgl. Abbildung 9.38.

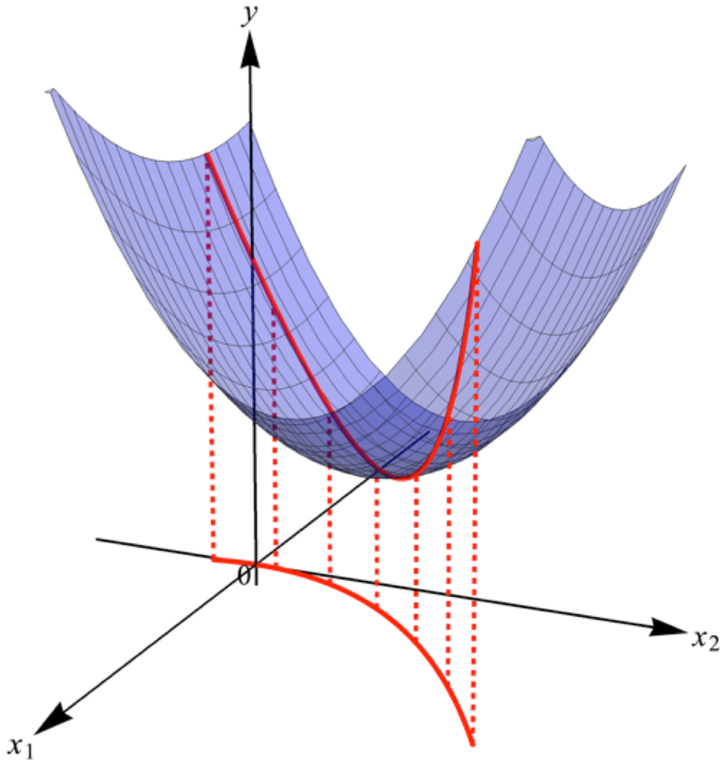

Abbildung 9.38: Die Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ entlang des Weges $(t^2, 3t)$.

■

Folgendes Beispiel verdeutlicht, dass der Vertikalschnitt einer Funktion f ausgehend von $\mathbf{x}^0$ in Richtung $\mathbf{r}$ auch immer als eine Verkettung der Funktion f mit einer Funktion g mit $g_i(t) = \mathbf{x}_i^0 + r_i t$ betrachtet werden kann.

■ Beispiel 9.2.21 — Vertikalschnitte und Verkettungen.
Verkettet man die Funktion $f(\mathbf{x}) = x_1^2 + x_2^2 - 3x_1x_2$ mit $g_1(t) = 1 \cdot t$ und $g_2(t) = 0 \cdot t = 0$ ergibt sich die Funktion $f(g(t)) = t^2$. Die Ableitung dieser Funktion ist $2t$. Das gleiche Ergebnis erhält man mit

$$f_{x_1}(\mathbf{x}) = 2x_1 - 3x_2 \text{ und } f_{x_2}(\mathbf{x}) = 2x_2 - 3x_1$$

aus

$$\begin{aligned}\nabla f(g(t))^T \begin{pmatrix} g_1'(t) \\ g_2'(t) \end{pmatrix} &= f_{x_1}(g(t)) \cdot g_1'(t) + f_{x_2}(g(t)) \cdot g_2'(t) \\ &= (2t - 3 \cdot 0) \cdot 1 + (2 \cdot 0 - 3t) \cdot 0 = 2t.\end{aligned}$$

Diese Funktion $f(g(t)) = t^2$ entspricht dem Vertikalschnitt der Funktion f ausgehend von $\mathbf{0}$ in Richtung $\mathbf{e}^1$. Betrachtet man den Vertikalschnitt von f ausgehend von $\mathbf{0}$ in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$, also $g_1(t) = \frac{1}{\sqrt{2}}t$ und $g_2(t) = \frac{1}{\sqrt{2}}t$, ergibt sich eine Funktion mit Ableitung

$$\begin{aligned}\nabla f(g(t))^T \begin{pmatrix} g_1'(t) \\ g_2'(t) \end{pmatrix} &= f_{x_1}(g(t)) \cdot g_1'(t) + f_{x_2}(g(t)) \cdot g_2'(t) \\ &= \left(2\frac{1}{\sqrt{2}}t - 3\frac{1}{\sqrt{2}}t\right)\frac{1}{\sqrt{2}} + \left(2\frac{1}{\sqrt{2}}t - 3\frac{1}{\sqrt{2}}t\right)\frac{1}{\sqrt{2}} = -t.\end{aligned}$$

Beide Vertikalschnitte sind in Abbildung 9.39 dargestellt.

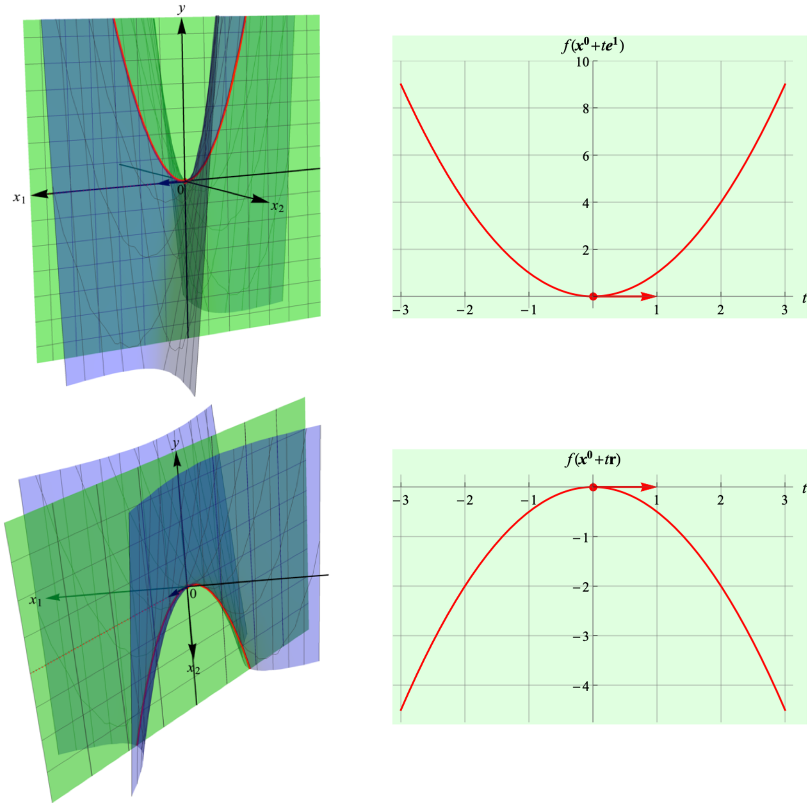

Abbildung 9.39: Vertikalschnitte von f durch $\mathbf{x}^0 = \mathbf{0}$.

■

Für eine total differenzierbare Funktion $f : D \to Z$ gibt es eine reellwertige Funktion r mit $\lim_{\Delta\mathbf{x}\to 0} r(\Delta\mathbf{x})/||\Delta\mathbf{x}|| = 0$ und ein $\delta > 0$, so dass

$$f(\mathbf{x}^0 + \Delta\mathbf{x}) - f(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)\Delta\mathbf{x} + r(\Delta\mathbf{x}) \text{ für alle } \mathbf{x} \in U(\mathbf{x}^0, \delta).$$

Die verallgemeinerte Kettenregel besagt für differenzierbare reelle Funktionen g_i in einer Variablen, $i = 1, \ldots, n$ mit offenem Definitionsbereich $D \subseteq \mathbb{R}$, welche eine reelle Funktion $g : D \to \mathbb{R}^n$ mit $g(t) = (g_1(t), \ldots, g_n(t))^T$ bilden, und eine total differenzierbare reelle Funktion in n Variablen $f : g(D) \to Z \subseteq \mathbb{R}$:

$$(f \circ g)'(t_0) = \sum_{i=1}^{n} f_{x_i}(g(t_0)) \cdot g_i'(t_0).$$

9.2.7 Die (erste) Richtungsableitung

Ziele dieses Unterkapitels

- Was versteht man unter einer Richtungsableitung einer Funktion f in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$?
- Wie kann man die Richtungsableitung einer Funktion in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$ mit Hilfe der partiellen Ableitungen bestimmen?
- Ausgehend von $\mathbf{x}^0$, wie findet man die Richtung $\mathbf{r}$ mit der grössten Richtungsableitung?

Die partielle Ableitung einer Funktion nach x_1 gibt die Ableitung der Funktion an, wenn alle anderen Variablen einem festen Wert zugeordnet werden und nur x_1 verändert wird. Bei einer reellen Funktion in zwei Variablen kann man sich die partielle Ableitung nach x_1 als Ableitung in Richtung der x_1-Achse für festes x_2 vorstellen. Wandert man also in einem durch die Funktion f aufgespannten Gebirge, so entspricht $f_{x_1}(\mathbf{x}^0) = \frac{\partial f(\mathbf{x}^0)}{\partial x_1}$ der Steigung der Tangentialebene an der Stelle $\mathbf{x}^0$, wenn man sich parallel zur x_1-Achse bewegt. Sie entspricht der Steigung der Tangente des Vertikalschnitts $f_{\mathbf{x}^0,\mathbf{e}^1}(t)$ in Richtung $\mathbf{e}^1$ bei $t = 0$. Analog kann man sich die partielle Ableitung nach x_2 als Ableitung in Richtung der x_2-Achse für festes x_1 vorstellen. Die partielle Ableitung $f_{x_2}(\mathbf{x}^0) = \frac{\partial f(\mathbf{x}^0)}{\partial x_2}$ entspricht der Steigung der Tangentialebene, wenn man sich parallel zur x_2-Achse bewegt. Damit entspricht sie der Steigung einer Tangente des Vertikalschnitts $f_{\mathbf{x}^0,\mathbf{e}^2}(t)$ in Richtung $\mathbf{e}^2$ bei $t = 0$. Richtungsableitungen verallgemeinern diese Idee nun für beliebige Richtungen.

Definition 9.2.9 — Die Richtungsableitung. Es sei $\mathbf{x}^0 \in D \subseteq \mathbb{R}^n$, D offen, $\mathbf{r} \in \mathbb{R}^n$ eine Richtung und $f : D \to Z$ eine stetig partiell differenzierbare Funktion. Dann heisst

$$f'_{\mathbf{r}}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{r}}(0) = \frac{d f_{\mathbf{x}^0,\mathbf{r}}}{dt}(0)$$

(erste) Richtungsableitung von f in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$.

In folgendem Beispiel illustrieren wir den Begriff.

■ Beispiel 9.2.22 — Richtungsableitung der Cobb Douglas Funktion.
Der Gradient der Cobb Douglas Funktion $f(x_1,x_2) = x_1^{0.8}x_2^{0.2}$ an der Stelle $\mathbf{x}^0 = (1,1)^T$ ist

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} f_{x_1}(\mathbf{x}^0) \\ f_{x_2}(\mathbf{x}^0) \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}.$$

Der Vertikalschnitt durch f in Richtung $\mathbf{e}^1$ durch $\mathbf{x}^0 = (1,1)^T$ ist

$$f_{\mathbf{x}^0,\mathbf{e}^1}(t) = (1+t)^{0.8}.$$

Die Ableitung dieses senkrechten Vertikalschnitts ist

$$f'_{\mathbf{x}^0,\mathbf{e}^1}(t) = 0.8(1+t)^{-0.2}.$$

Für $t = 0$ gilt also

$$f'_{\mathbf{e}^1}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{e}^1}(0) = 0.8 = f_{x_1}(\mathbf{x}^0).$$

Dies ist nicht überraschend: Die partielle Ableitung nach x_i beschreibt die Veränderungsrate des Funktionswerts, wenn nur x_i verändert wird und alle anderen Komponenten fixiert werden. Der senkrechte Vertikalschnitt in Richtung $\mathbf{e}^i$ fixiert ebenfalls alle Komponenten ausser x_i. Die Ableitung des Vertikalschnitts in Richtung $\mathbf{e}^i$ gibt somit ebenfalls die Veränderungsrate des Funktionswerts an, wenn nur x_i verändert wird und alle anderen Komponenten fixiert werden. So gilt hier z.B. auch

$$f_{\mathbf{x}^0,\mathbf{e}^2}(t) = (1+t)^{0.2} \text{ und } f'_{\mathbf{x}^0,\mathbf{e}^2}(t) = 0.2(1+t)^{-0.8} \text{ bzw. } f'_{\mathbf{e}^2}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{e}^2}(0) = 0.2 = f_{x_2}(\mathbf{x}^0).$$

Man sagt auch: Die Richtungsableitung von f in Richtung $\mathbf{e}^1$ an der Stelle $\mathbf{x}^0$ entspricht der partiellen Ableitung $f_{x_1}(\mathbf{x}^0)$. Die Richtungsableitung von f in Richtung $\mathbf{e}^2$ an der Stelle $\mathbf{x}^0$ entspricht der partiellen Ableitung $f_{x_2}(\mathbf{x}^0)$.

Der Vertikalschnitt in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ durch $\mathbf{x}^0$ ist

$$f_{\mathbf{x}^0,\mathbf{r}}(t) = 1 + \frac{1}{\sqrt{2}}t \text{ mit } f'_{\mathbf{x}^0,\mathbf{r}}(t) = \frac{1}{\sqrt{2}}.$$

Die Richtungsableitung ist also

$$f'_{\mathbf{r}}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{r}}(0) = \frac{1}{\sqrt{2}} = 0.8 \cdot \frac{1}{\sqrt{2}} + 0.2 \cdot \frac{1}{\sqrt{2}},$$

siehe Abbildung 9.40.

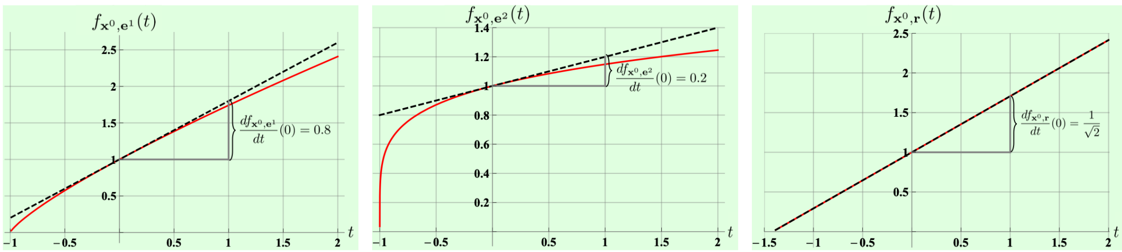

Abbildung 9.40: Vertikalschnitt von f durch $\mathbf{x}^0 = (1,1)^T$ in Richtung $\mathbf{e}^1$, $\mathbf{e}^2$ und $\mathbf{r}$.

■

In allen Berechnungen des vorherigen Beispiels zur Berechnung der Richtungsableitung einer reellen Funktion in zwei Variablen in Richtung $\mathbf{r} = (r_1, r_2)^T$ gilt

$$f'_{\mathbf{r}}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{r}}(0) = r_1 f_{x_1}(\mathbf{x}^0) + r_2 f_{x_2}(\mathbf{x}^0).$$

Folgender Satz besagt, dass diese Berechnungsvorschrift auch allgemein gilt:

Satz 9.2.5 — Berechnung der Richtungsableitung.
Sei $D \subseteq \mathbb{R}^n$ offen. Ist $f : D \to Z$ eine stetig partiell differenzierbare reelle Funktion in n Variablen, $\mathbf{x}^0 \in D$ und $\mathbf{r} \in \mathbb{R}^n$ eine Richtung, dann gilt

$$f'_{\mathbf{r}}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{r}}(0) = \nabla f(\mathbf{x}^0)^T \mathbf{r} = \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0) r_i.$$

Beweis (#): Die Aussage ist ein Spezialfall der verallgemeinerten Kettenregel mit $g(t) = \mathbf{x}^0 + t\mathbf{r}$. ■

Wir demonstrieren Richtungsableitungen an einigen Beispielen:

■ Beispiel 9.2.23 — Richtungsableitungen einer quadratischen Funktion.
Für die quadratische Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^2 + 3$ ist der Gradient an der Stelle $\mathbf{x}^0 = (1,0)^T$ als

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2 \\ -8 \end{pmatrix}$$

gegeben. Die Richtungsableitung in Richtung $\mathbf{e}^1$ ist also gleich

$$f'_{\mathbf{e}^1}(\mathbf{x}^0) = f'_{\mathbf{x}^0,\mathbf{e}^1}(0) = f_{x_1}(\mathbf{x}^0) = 2.$$

Der Vertikalschnitt von f ausgehend von $\mathbf{x}^0$ in Richtung $\mathbf{e}^1$ ist

$$f_{\mathbf{x}^0,\mathbf{e}^1}(t) = (1+t)^2 + 4(0-1)^2 + 3 = (1+t)^2 + 7.$$

Leitet man diesen Vertikalschnitt nach t ab, ergibt sich

$$f'_{\mathbf{x}^0,\mathbf{e}^1}(t) = 2(1+t) \text{ und } f'_{\mathbf{x}^0,\mathbf{e}^1}(0) = 2 = f'_{\mathbf{e}^1}(\mathbf{x}^0).$$

Die Richtungsableitung von f in Richtung $\mathbf{e}^1$ an der Stelle $\mathbf{x}^0$ entspricht $f'_{\mathbf{x}^0,\mathbf{e}^1}(t)$ an der Stelle $t = 0$.

Die Richtungsableitung von f in Richtung $\mathbf{e}^2$ an der Stelle $\mathbf{x}^0$ kann man ebenfalls über die Ableitung des Vertikalschnitts

$$f_{\mathbf{x}^0,\mathbf{e}^2}(t) = (1)^2 + 4(t-1)^2 + 3 = 4(t-1)^2 + 4$$

an der Stelle $t = 0$ oder direkt aus f als $f_{x_2}(\mathbf{x}^0) = -8$ bestimmen.

Die Richtungsableitung von f an der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ ist laut Satz 9.2.5

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \frac{1}{\sqrt{2}} \cdot f_{x_1}(\mathbf{x}^0) + \frac{1}{\sqrt{2}} \cdot f_{x_2}(\mathbf{x}^0) = \frac{1}{\sqrt{2}}(2) + \frac{1}{\sqrt{2}}(-8) = -\frac{6}{\sqrt{2}}.$$

Auch dies kann man einfach mithilfe der Ableitung des Vertikalschnitts

$$f_{\mathbf{x}^0,\mathbf{r}}(t) = \left(1 + \frac{1}{\sqrt{2}}t\right)^2 + 4\left(\frac{1}{\sqrt{2}}t - 1\right)^2 + 3$$

an der Stelle $t = 0$ überprüfen.

Bewegt man sich in entgegengesetzter Richtung, verändert die Ableitung ihr Vorzeichen. Die Richtungsableitung von f an der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{r} = \left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)^T$ ist

$$f'_{\mathbf{r}}(\mathbf{x}^0) = -\frac{1}{\sqrt{2}} \cdot f_{x_1}(\mathbf{x}^0) - \frac{1}{\sqrt{2}} \cdot f_{x_2}(\mathbf{x}^0) = \frac{6}{\sqrt{2}}.$$

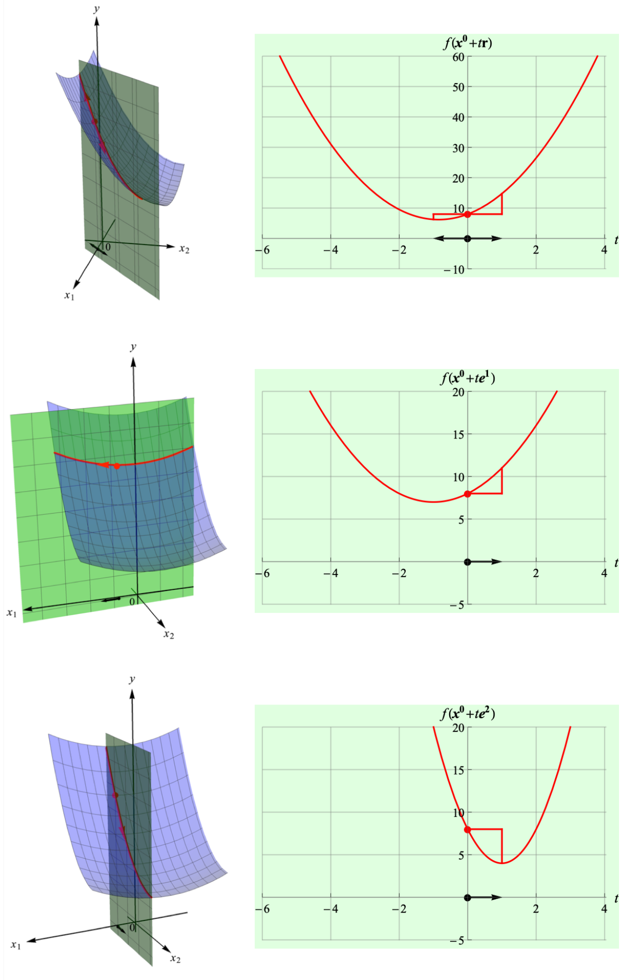

Abbildung 9.41: Richtungsableitungen der quadratischen Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$.

Der Vertikalschnitt in diese Richtung hat die Form

$$f_{\mathbf{x}^0,\mathbf{r}}(t) = \left(1 - \frac{1}{\sqrt{2}}t\right)^2 + 4\left(-\frac{1}{\sqrt{2}}t - 1\right)^2 + 3 \text{ mit } f'_{\mathbf{x}^0,\mathbf{r}}(0) = \frac{6}{\sqrt{2}}.$$

Die Richtungsableitungen sind in Abbildung 9.41 graphisch dargestellt.

■

■ **Beispiel 9.2.24 — Richtungsableitungen einer affin-linearen Funktion.**
Die Richtungsableitung der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$ in Richtung $\mathbf{r}$ an einer Stelle $\mathbf{x}^0$ ist

$$f'_{\mathbf{r}}(\mathbf{x}^0) = r_1 f_{x_1}(\mathbf{x}^0) + r_2 f_{x_2}(\mathbf{x}^0) = r_1 \cdot (-8) + r_2 \cdot (3).$$

Bei obiger affin-linearen Funktion sind die Richtungsableitungen also an jeder Stelle gleich. Sie unterscheiden sich aber natürlich je nach Richtung $\mathbf{r}$: Mit $\mathbf{r} = \mathbf{e}^1$ ist $f'_{\mathbf{e}^1}(\mathbf{x}^0) = f_{x_1}(\mathbf{x}^0) = -8$ und mit $\mathbf{r} = \mathbf{e}^2$ ist $f'_{\mathbf{e}^2}(\mathbf{x}^0) = f_{x_2}(\mathbf{x}^0) = 3$ für alle $\mathbf{x}^0 \in \mathbb{R}^2$. In Richtung

$\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ gilt

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \frac{1}{\sqrt{2}} \cdot (-8) + \frac{1}{\sqrt{2}} \cdot 3 = \frac{-5}{\sqrt{2}}.$$

Geht man in Richtung des Gradienten von f, das heisst $\mathbf{r} = \left(\frac{-8}{\sqrt{73}}, \frac{3}{\sqrt{73}}\right)^T$, ergibt sich

$$f'_{\mathbf{r}}(\mathbf{x}) = \frac{-8}{\sqrt{73}} \cdot (-8) + \frac{3}{\sqrt{73}} \cdot 3 = \frac{73}{\sqrt{73}} = \sqrt{73}.$$

In Abbildung 9.42 stellen Pfeile in verschiedenen Farben die verschiedenen Richtungen $\mathbf{r}$ ausgehend von $\mathbf{x}^0$ innerhalb der x_1-x_2-Ebene dar. Der blaue Pfeil repräsentiert $\mathbf{e}^1$, der rote Pfeil $\mathbf{e}^2$, der orangene Pfeil die Diagonale $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ und der lila Pfeil die Richtung des Gradienten $\left(\frac{-8}{\sqrt{73}}, \frac{3}{\sqrt{73}}\right)^T$. All diese Pfeile haben eine Länge von 1. Um besser erkennen zu können, wie steil die Funktion in diese verschiedenen Richtungen ansteigt, sind zusätzlich die Bilder der Pfeile gestrichelt auf dem Graphen der Funktion eingezeichnet. Man erkennt, dass die Funktion in Richtung des Gradienten am steilsten ansteigt.

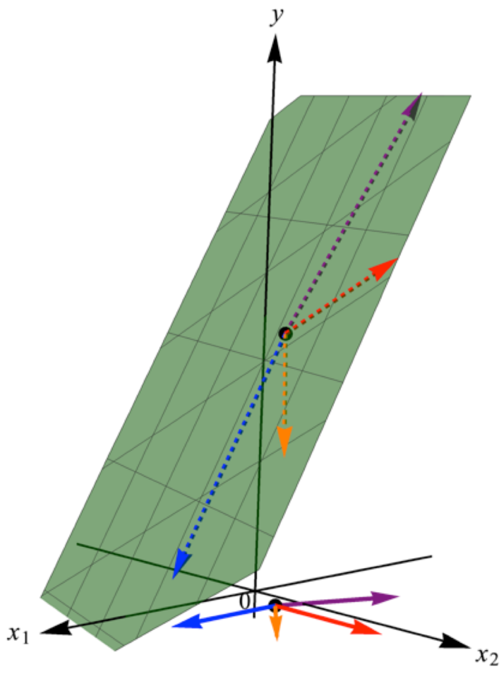

Abbildung 9.42: Darstellung der Steigungen in verschiedene Richtungen.

■

In obigem Beispiel der affin-linearen Funktion sieht man, dass der Wert der Richtungsableitung stark variieren kann. Der grösste Wert ergibt sich, wenn man die Funktion in Richtung des Gradienten betrachtet. Dieser Zusammenhang gilt allgemein und wird in der Optimierung häufig genutzt: der Gradient zeigt stets in die Richtung des steilsten Anstiegs.

Satz 9.2.6 — Gradient als Richtung des steilsten Anstiegs.
Sei $D \subseteq \mathbb{R}^n$ offen, $f : D \to Z$ eine stetig partiell differenzierbare reelle Funktion in n Variablen und $\mathbf{x}^0 \in D$. Für alle Richtungen $\mathbf{r} \in \mathbb{R}^n$ gilt mit $\tilde{\mathbf{r}} = \frac{\nabla f(\mathbf{x}^0)}{||\nabla f(\mathbf{x}^0)||}$, dass

$$f'_{\tilde{\mathbf{r}}}(\mathbf{x}^0) \geq f'_{\mathbf{r}}(\mathbf{x}^0).$$

Beweis (#): Da der Nullvektor keine Richtung hat, beschränken uns in diesem Beweis auf den Fall $\nabla f(\mathbf{x}^0) \neq \mathbf{0}$ mit $||\nabla f(\mathbf{x}^0)|| \neq 0$.

Dabei betrachten wir zuerst den Fall $n \in \{2,3\}$. Die Richtungsableitung an der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{r}$ ist

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \left(\nabla f(\mathbf{x}^0)\right)^T \mathbf{r} = ||\nabla f(\mathbf{x}^0)|| \cdot ||\mathbf{r}|| \cdot \cos(\phi),$$

wobei ϕ den Winkel zwischen der Richtung $\mathbf{r}$ und dem Gradienten beschreibt, vgl. Satz 6.1.1. Da für jede Richtung $||\mathbf{r}|| = 1$ gilt, kann man dies zu

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \left(\nabla f(\mathbf{x}^0)\right)^T \mathbf{r} = ||\nabla f(\mathbf{x}^0)|| \cdot \cos(\phi)$$

vereinfachen. Somit ist die Richtungsableitung in Richtung $\mathbf{r}$ gleich der Länge des Gradienten multipliziert mit dem Kosinus des Winkels, welcher die Richtung mit dem Gradienten einschliesst. Die Länge des Gradienten wird von der Richtung nicht beeinflusst. Die grösstmögliche Richtungsableitung ergibt sich also durch den Winkel, der einen Kosinus von 1 erzielt. Dies ist der Winkel 0. Die maximale Richtungsableitung wird also in der Richtung erzielt, welche der Richtung des Gradienten entspricht. Alle anderen Richtungen haben eine kleinere Richtungsableitung.

Für allgemeines n kann man wie folgt vorgehen:[7] Da die Norm eines Vektors stets positiv oder gleich 0 ist, und die Norm einer Richtung $\mathbf{r}$ gleich 1 ist, gilt für die Norm des Vektors $||\nabla f(\mathbf{x}^0)||\mathbf{r} - \nabla f(\mathbf{x}^0)$:

$$\begin{aligned}
0 &\leq ||(||\nabla f(\mathbf{x}^0)||\mathbf{r} - \nabla f(\mathbf{x}^0))||^2 \\
&= \sum_{i=1}^{n} (||\nabla f(\mathbf{x}^0)||r_i - f_{x_i}(\mathbf{x}^0))^2 \\
&= \sum_{i=1}^{n} ||\nabla f(\mathbf{x}^0)||^2 (r_i)^2 - 2||\nabla f(\mathbf{x}^0)|| \cdot r_i \cdot f_{x_i}(\mathbf{x}^0) + f_{x_i}(\mathbf{x}^0)^2 \\
&= ||\nabla f(\mathbf{x}^0)||^2 \sum_{i=1}^{n} (r_i)^2 - 2||\nabla f(\mathbf{x}^0)|| \cdot \sum_{i=1}^{n} (r_i \cdot f_{x_i}(\mathbf{x}^0)) + \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0)^2 \\
&= ||\nabla f(\mathbf{x}^0)||^2 - 2||\nabla f(\mathbf{x}^0)||\nabla f(\mathbf{x}^0)^T \mathbf{r} + ||\nabla f(\mathbf{x}^0)||^2 \\
&= 2||\nabla f(\mathbf{x}^0)||^2 - 2||\nabla f(\mathbf{x}^0)||\nabla f(\mathbf{x}^0)^T \mathbf{r}.
\end{aligned}$$

Ist $||\nabla f(\mathbf{x}^0)|| \neq 0$ ergibt sich aus obiger Ungleichung

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)^T \mathbf{r} \leq ||\nabla f(\mathbf{x}^0)|| = \nabla f(\mathbf{x}^0)^T \frac{\nabla f(\mathbf{x}^0)}{||\nabla f(\mathbf{x}^0)||} = f'_{\tilde{\mathbf{r}}}(\mathbf{x}^0).$$

■

[7] Wir beweisen hier einen Spezialfall der sogenannten Cauchy-Schwarzschen Ungleichung, aus welcher die Behauptung dann direkt folgt.

Ist eine reelle Funktion nicht stetig partiell differenzierbar, so müssen nicht alle Richtungsableitungen existieren.

Beispiel 9.2.25 — Eine nicht stetig partiell differenzierbare Funktion.
Wir betrachten erneut die Funktion

$$h(\mathbf{x}) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{für } \mathbf{x} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ 0 & \text{für } \mathbf{x} = \mathbf{0} \end{cases}$$

aus Beispiel 9.2.9. Diese Funktion ist partiell, aber nicht stetig partiell differenzierbar.

Der Vertikalschnitt $h_{\mathbf{0},\mathbf{e}^1}(t)$ durch $\mathbf{0}$ in Richtung $\mathbf{e}^1$ ist $h(\mathbf{0}+t\mathbf{e}^1) = 0$ für alle t. Der Vertikalschnitt in Richtung $\mathbf{e}^1$ durch $\mathbf{0}$ ist also differenzierbar mit Ableitung $h'_{\mathbf{0},\mathbf{e}^1}(t) = 0$.

Der Vertikalschnitt $h_{\mathbf{0},\mathbf{r}}(t)$ durch $\mathbf{0}$ in Richtung der Diagonalen $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ ist $h(\mathbf{0}+t\mathbf{r}) = \frac{1}{2}$ für alle $t \neq 0$ und $h(\mathbf{0}+0\cdot\mathbf{r}) = 0$. Der Vertikalschnitt in Richtung $\mathbf{r}$ durch $\mathbf{0}$ ist also nicht differenzierbar. Es existiert keine Richtungsableitung in Richtung $\mathbf{r}$.

Da nur für stetig partiell differenzierbare reelle Funktionen sichergestellt ist, dass der Vertikalschnitt der Funktion durch jede Stelle im Definitionsbereich in jede Richtung differenzierbar ist, werden Richtungsableitungen nur für diese Funktionen definiert. ■

Z Die Richtungsableitung einer stetig partiell differenzierbaren reellen Funktion f in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$, $f'_{\mathbf{r}}(\mathbf{x}^0)$, entspricht der Ableitung des Vertikalschnitts der Funktion in Richtung $\mathbf{r}$ durch $\mathbf{x}^0$ für $t = 0$, $f'_{\mathbf{r},\mathbf{x}^0}(0)$. Es gilt

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0) r_i.$$

Der Gradient zeigt stets in die Richtung des grössten Anstiegs, also ist $\mathbf{r} = \nabla f(\mathbf{x}^0)/||\nabla f(\mathbf{x}^0)||$ die Richtung des grössten Anstiegs.

9.2.8 Die zweite Richtungsableitung

Ziele dieses Unterkapitels

- Was versteht man unter der zweiten Richtungsableitung einer reellen Funktion f in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$?
- Wie kann man die zweite Richtungsableitung einer reellen Funktion in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$ mithilfe der Hesse-Matrix der Funktion an dieser Stelle bestimmen?

Schneidet man den Graphen von f an einer Stelle $\mathbf{x}^0$ in Richtung $\mathbf{r}$ durch, entsteht ein Vertikalschnitt. An der Stelle $\mathbf{x}^0 + t\mathbf{r}$, also t Längeneinheiten von $\mathbf{x}^0$ in Richtung $\mathbf{r}$ entfernt, ist der Funktionswert $f_{\mathbf{x}^0,\mathbf{r}}(t) = f(\mathbf{x}^0 + t\mathbf{r})$. Die erste Ableitung der Funktion $f_{\mathbf{x}^0,\mathbf{r}}(t)$ entspricht der Summe der mit r_i gewichteten partiellen Ableitungen an der Stelle $\mathbf{x}^0 + t\mathbf{r}$,

$$f'_{\mathbf{x}^0,\mathbf{r}}(t) = \nabla f(\mathbf{x}^0 + t\mathbf{r})^T \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0 + t\mathbf{r}) r_i,$$

vgl. Satz 9.2.5. Betrachtet man die partielle Ableitung f_{x_i} als Funktion, ist die erste Richtungsableitung der partiellen Ableitung f_{x_i} folglich

$$(f_{x_i})'_{\mathbf{r}}(\mathbf{x}^0) = \nabla f_{x_i}(\mathbf{x}^0+t\mathbf{r})^T \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}^0+t\mathbf{r}) r_j.$$

So ergibt sich die Ableitung der ersten Richtungsableitung als

$$\begin{aligned} f''_{\mathbf{x}^0,\mathbf{r}}(t) = \left(f'_{\mathbf{x}^0,\mathbf{r}}(t)\right)' &= \left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0+t\mathbf{r}) r_i\right)' = \sum_{i=1}^{n} (f_{x_i}(\mathbf{x}^0+t\mathbf{r}))' r_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}^0+t\mathbf{r}) r_j\right) r_i = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}^0+t\mathbf{r}) r_j r_i \\ &= \mathbf{r}^T H_f(\mathbf{x}^0+t\mathbf{r})\mathbf{r}, \end{aligned}$$

wobei $H_f(\mathbf{x}^0+t\mathbf{r})$ die Hesse-Matrix der Funktion f an der Stelle $\mathbf{x}^0+t\mathbf{r}$ beschreibt.

■ **Beispiel 9.2.26 — Zweite Richtungsableitung der Cobb Douglas Funktion.**
Die Hesse-Matrix der Cobb Douglas Funktion $f(x_1,x_2) = x_1^{0.8} x_2^{0.2}$ an der Stelle $\mathbf{x}^0 = (1,1)^T$ ist

$$H_f(\mathbf{x}^0) = \begin{pmatrix} f_{x_1x_1}(\mathbf{x}^0) & f_{x_1x_2}(\mathbf{x}^0) \\ f_{x_2x_1}(\mathbf{x}^0) & f_{x_2x_2}(\mathbf{x}^0) \end{pmatrix} = \begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix}.$$

Der Vertikalschnitt durch f in Richtung $\mathbf{e}^1$ durch $\mathbf{x}^0 = (1,1)^T$ ist

$$f_{\mathbf{x}^0,\mathbf{e}^1}(t) = (1+t)^{0.8} \text{ mit } f'_{\mathbf{x}^0,\mathbf{e}^1}(t) = 0.8(1+t)^{-0.2}, f''_{\mathbf{x}^0,\mathbf{e}^1}(t) = -0.16(1+t)^{-1.2}.$$

Für $t = 0$ gilt also

$$f''_{\mathbf{x}^0,\mathbf{e}^1}(0) = -0.16 = f_{x_1x_1}(\mathbf{x}^0).$$

Dies ist nicht überraschend: Die zweite partielle Ableitung nach x_i entspricht der zweiten Ableitung einer Funktion, welche x_2 fixiert und x_1 variabel lässt. Der senkrechte Vertikalschnitt in Richtung $\mathbf{e}^1$ fixiert ebenfalls alle Komponenten ausser x_1. Die zweite Ableitung des Vertikalschnitts in Richtung $\mathbf{e}^1$ enspricht daher der zweiten partiellen Ableitung $f_{x_1x_1}$. Analog gilt

$$f_{\mathbf{x}^0,\mathbf{e}^2}(t) = (1+t)^{0.2} \text{ mit } f'_{\mathbf{x}^0,\mathbf{e}^2}(t) = 0.2(1+t)^{-0.8}, f''_{\mathbf{x}^0,\mathbf{e}^2}(t) = -0.16(1+t)^{-1.8}.$$

Der Vertikalschnitt in Richtung der Diagonalen $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ durch $\mathbf{x}^0$ ist

$$f_{\mathbf{x}^0,\mathbf{r}}(t) = 1 + \frac{1}{\sqrt{2}} t \text{ mit } f'_{\mathbf{x}^0,\mathbf{r}}(t) = \frac{1}{\sqrt{2}}, f''_{\mathbf{x}^0,\mathbf{r}}(t) = 0.$$

In diese Richtung ist die zweite Ableitung des Vertikalschnitts gleich 0, vgl. auch Abbildung 9.40. Gewichtet man die Einträge der Hesse-Matrix entsprechend obiger Formel, ergibt sich ebenfalls:

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \begin{pmatrix} -0.16\frac{1}{\sqrt{2}} + 0.16\frac{1}{\sqrt{2}} \\ 0.16\frac{1}{\sqrt{2}} - 0.16\frac{1}{\sqrt{2}} \end{pmatrix} = 0.$$

■

Zusammenfassend kann also die zweite partielle Ableitung des Vertikalschnitts $f_{\mathbf{x}^0,\mathbf{r}}(t)$ mithilfe der Hesse-Matrix von f bestimmt werden. Analog zur ersten Richtungsableitung nennen wir diese zweite Ableitung die zweite Richtungsableitung von f in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$:

Definition 9.2.10 — Die zweite Richtungsableitung.
Es sei $\mathbf{x}^0 \in D \subseteq \mathbb{R}^n$, D offen, $\mathbf{r} \in \mathbb{R}^n$ eine Richtung und $f : D \to \mathbb{R}$ eine zweimal stetig partiell differenzierbare Funktion. Dann heisst

$$f''_{\mathbf{r}}(\mathbf{x}^0) = f''_{\mathbf{x}^0,\mathbf{r}}(0) = \frac{d^2 f_{\mathbf{x}^0,\mathbf{r}}}{dt^2}(0)$$

zweite Richtungsableitung von f (an der Stelle $\mathbf{x}^0$) in Richtung $\mathbf{r} \in \mathbb{R}^n$.

Satz 9.2.7 — Berechnung der zweiten Richtungsableitung.
Sei $D \subseteq \mathbb{R}^n$ offen. Ist $f : D \to Z$ eine zweimal stetig partiell differenzierbare reelle Funktion in n Variablen, $\mathbf{x}^0 \in D$ und $\mathbf{r} \in \mathbb{R}^n$ eine Richtung, dann gilt

$$f''_{\mathbf{r}}(\mathbf{x}^0) = f''_{\mathbf{x}^0,\mathbf{r}}(0) = \frac{d^2 f_{\mathbf{x}^0,\mathbf{r}}}{dt^2}(0) = \mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r}.$$

Beweis (#)*:* Die Aussage ist erneut ein Spezialfall der verallgemeinerten Kettenregel (angewandt auf die erste Richtungsableitung) mit $g(t) = \mathbf{x}^0 + t\mathbf{r}$. ■

■ Beispiel 9.2.27 — Fortsetzung von Beispiel 9.2.23.
Für die quadratische Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ gilt für alle Stellen $\mathbf{x}^0$:

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

In Beispiel 9.2.23 hatten wir bereits erste Richtungsableitungen dieser Funktion in Richtungen der kanonischen Einheitsvektoren und der Diagonalen an einer Stelle $\mathbf{x}^0$ bestimmt. Die entsprechenden Vertikalschnitte sind in Abbildung 9.41 dargestellt. In den Graphen scheinen die Vertikalschnitte konvex zu sein. Wir vermuten daher positive zweite Ableitungen der Vertikalschnitte, also auch positive zweite Richtungsableitungen.

Die zweite Richtungsableitung an der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{r} = \mathbf{e}^1$ berechnet sich laut Satz 9.2.7 also als

$$f''_{\mathbf{e}^1}(\mathbf{x}^0) = (1,0)H_f(\mathbf{x}^0)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2.$$

Die zweite Richtungsableitung an der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{e}^1$ entspricht der zweiten partiellen Ableitung $f_{x_1x_1}(\mathbf{x}^0)$. Statt mithilfe der Hesse-Matrix kann man diese zweite Richtungsableitung auch erhalten, indem man den Vertikalschnitt von f ausgehend von der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{e}^1 = (1,0)^T$ explizit als

$$f_{\mathbf{x}^0,\mathbf{e}^1}(t) = (x_1^0 + t)^2 + 4(x_2^0 - 1)^2 + 3$$

bestimmt und nach t ableitet,

$$\begin{aligned} ((x_1^0+t)^2 + 4(x_2^0-1)^2 + 3)' &= 2(x_1^0 + t) \\ ((x_1^0+t)^2 + 4(x_2^0-1)^2 + 3)'' &= \left(2(x_1^0+t)\right)' = 2. \end{aligned}$$

Die zweite Richtungsableitung in Richtung $\mathbf{e}^1$ an der Stelle $\mathbf{x}^0$ entspricht dieser zweiten Ableitung an der Stelle $t = 0$.

Die zweite Richtungsableitung an einer Stelle $\mathbf{x}^0$ in Richtung $\mathbf{e}^2$ berechnet sich als

$$f''_{\mathbf{e}^2}(\mathbf{x}^0) = (0,1)H_f(\mathbf{x}^0)\begin{pmatrix}0\\1\end{pmatrix} = 8.$$

Die zweite Ableitung des entsprechenden Vertikalschnitts

$$f_{\mathbf{x}^0,\mathbf{e}^2}(t) = (x_1^0)^2 + 4(x_2^0 + t - 1)^2 + 3$$

nach t hat für alle t und damit auch für $t = 0$ ebenfalls den Wert 8.

Die zweite Richtungsableitung an einer Stelle $\mathbf{x}^0$ in Richtung der Diagonalen $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ berechnet sich laut Satz 9.2.7 als

$$\begin{aligned} f''_{\mathbf{r}}(\mathbf{x}^0) &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) H_f(\mathbf{x}^0) \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix}2 & 0\\ 0 & 8\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} \\ &= 2\cdot\frac{1}{2} + 8\cdot\frac{1}{2} = 1 + 4 = 5. \end{aligned}$$

Dieser Wert entspricht der zweiten Ableitung des Vertikalschnitts durch $\mathbf{x}^0$ in Richtung der Diagonalen. ■

■ **Beispiel 9.2.28 — Zweite Richtungsableitungen einer affin-linearen Funktion.** Da die Hesse-Matrix der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$ der Nullmatrix entspricht, vgl. Beispiel 9.2.11, sind die zweiten Richtungsableitungen an allen Stellen des Definitionsbereichs in alle Richtungen gleich 0. ■

Z Die zweite Richtungsableitung einer reellen Funktion f in Richtung $\mathbf{r}$ an der Stelle $\mathbf{x}^0$, $f''_{\mathbf{r}}(\mathbf{x}^0)$, entspricht der zweiten Ableitung des Vertikalschnitts der Funktion in Richtung $\mathbf{r}$ durch $\mathbf{x}^0$ für $t = 0$. Ist f zweimal stetig partiell differenzierbar, gilt

$$f''_{\mathbf{r}}(\mathbf{x}^0) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}^0) r_i r_j = \mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r}.$$

9.3 Eigenschaften reeller Funktionen in mehreren Variablen

Wir versuchen nun bekannte Werkzeuge für reelle Funktionen in einer Variablen zu nutzen, um das Verhalten einer reellen Funktion in n Variablen zu beschreiben.

9.3.1 Monotonie

Ziele dieses Unterkapitels

- Wann nennt man eine reelle Funktion in n Variablen monoton steigend bzw. fallend in Richtung $\mathbf{r}$?
- Wie kann man die Monotonie einer stetig partiell differenzierbaren reellen Funktion in Richtung $\mathbf{r}$ überprüfen?

Der Vertikalschnitt $f_{\mathbf{x}^0,\mathbf{e}^i}(t) = f(\mathbf{x}^0 + t\mathbf{e}^i)$ beschreibt die Funktion f in n Variablen ausgehend von einer Stelle $\mathbf{x}^0$ in eine Richtung $\mathbf{e}^i$. Dabei ist $f_{\mathbf{x}^0,\mathbf{e}^i}(t)$ eine Funktion in einer Variablen, welche alle Komponenten ausser x_i fixiert. Diese Funktion kann wie jede andere reelle Funktion in einer Variablen auf Monotonie und andere Eigenschaften untersucht werden. Ist $f_{\mathbf{x}^0,\mathbf{e}^i}(t)$ für jede Stelle $\mathbf{x}^0$ im Definitionsbereich monoton steigend, so heisst f monoton steigend in der Variable x_i.

Definition 9.3.1 — Monotonie.
Eine reelle Funktion $f : D \to Z$ in n Variablen heisst (streng) monoton steigend in Richtung $\mathbf{r} \in \mathbb{R}^n$, wenn der Vertikalschnitt $f_{\mathbf{x}^0,\mathbf{r}}(t)$ von f durch $\mathbf{x}^0$ in Richtung $\mathbf{r}$ für alle $\mathbf{x}^0 \in D \subseteq \mathbb{R}^n$ (streng) monoton steigend ist. Die Funktion f heisst (streng) monoton fallend in Richtung $\mathbf{r}$, wenn $f_{\mathbf{x}^0,\mathbf{r}}(t)$ für alle $\mathbf{x}^0 \in D \subseteq \mathbb{R}^n$ (streng) monoton fallend ist.

Ist f monoton steigend bzw. fallend in Richtung $\mathbf{e}^i$, sagt man auch, dass f in der Variable x_i steigt bzw. fällt.

Die Richtungsableitungen beschreiben die Ableitungen der Vertikalschnitte. Daher gilt:

Satz 9.3.1 — Monotoniekriterium.
Sei $D \subseteq \mathbb{R}^n$ offen und konvex. Ist die reelle Funktion $f : D \to Z$ in n Variablen stetig partiell differenzierbar und $\mathbf{r} \in \mathbb{R}^n$ eine Richtung, dann gilt:

- $f'_{\mathbf{r}}(\mathbf{x}^0) \geq 0$ für alle $\mathbf{x}^0 \in D \Leftrightarrow f$ ist monoton steigend in Richtung $\mathbf{r}$.
- $f'_{\mathbf{r}}(\mathbf{x}^0) > 0$ für alle $\mathbf{x}^0 \in D \Rightarrow f$ ist streng monoton steigend in Richtung $\mathbf{r}$.
- $f'_{\mathbf{r}}(\mathbf{x}^0) \leq 0$ für alle $\mathbf{x}^0 \in D \Leftrightarrow f$ ist monoton fallend in Richtung $\mathbf{r}$.
- $f'_{\mathbf{r}}(\mathbf{x}^0) < 0$ für alle $\mathbf{x}^0 \in D \Rightarrow f$ ist streng monoton fallend in Richtung $\mathbf{r}$.

Beweis. Der Beweis folgt direkt aus Satz 5.5.1 über die Monotonie von reellen Funktionen. ∎

Ist f stetig partiell differenzierbar, kann $f_{\mathbf{x}^0,\mathbf{r}}(t)$ mithilfe der ersten Richtungsableitung untersucht werden. Wir demonstrieren dies in einem Beispiel.

Beispiel 9.3.1 — Beispiel einer monotonen Funktion.
Für die Cobb Douglas Funktion $f : (0,+\infty)^2 \to \mathbb{R}$ mit $f(x_1,x_2) = x_1^{0.8}x_2^{0.2}$ gilt an der Stelle $\mathbf{x}^0 \in D = (0,+\infty)^2$

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 0.8\left(\frac{x_2^0}{x_1^0}\right)^{0.2} \\ 0.2\left(\frac{x_1^0}{x_2^0}\right)^{0.8} \end{pmatrix}.$$

Für $\mathbf{r} = \mathbf{e}^1$ gilt

$$f'_{\mathbf{e}^1}(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)^T \mathbf{e}^1 = \left(0.8\left(\frac{x_2^0}{x_1^0}\right)^{0.2}, 0.2\left(\frac{x_1^0}{x_2^0}\right)^{0.8}\right)\begin{pmatrix}1\\0\end{pmatrix} = 0.8\left(\frac{x_2^0}{x_1^0}\right)^{0.2} > 0.$$

Also ist $f'_{\mathbf{e}^1}(\mathbf{x}^0) > 0$ für alle $\mathbf{x}^0 \in D$. Die Cobb Douglas Funktion ist in Richtung $\mathbf{e}^1$ streng monoton steigend.

Für eine allgemeine Richtung $\mathbf{r} \in \mathbb{R}^2$ gilt

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)^T \mathbf{r} = \left(0.8\left(\frac{x_2^0}{x_1^0}\right)^{0.2}, 0.2\left(\frac{x_1^0}{x_2^0}\right)^{0.8}\right)\begin{pmatrix}r_1\\r_2\end{pmatrix} = 0.8r_1\left(\frac{x_2^0}{x_1^0}\right)^{0.2} + 0.2r_2\left(\frac{x_1^0}{x_2^0}\right)^{0.8}.$$

Da für alle $\mathbf{x}^0 \in D$ der Quotient $\frac{x_2^0}{x_1^0} > 0$ und damit auch $\frac{x_1^0}{x_2^0} > 0$ ist, gilt für alle Richtungen mit $r_1 \geq 0, r_2 \geq 0$:

$$f'_{\mathbf{r}}(\mathbf{x}^0) \geq 0.$$

Die Funktion f ist also in alle Richtungen $\mathbf{r}$ mit $r_1, r_2 \geq 0$ monoton steigend. In Richtungen $\mathbf{r}$ mit $r_1, r_2 > 0$ gilt sogar

$$f'_{\mathbf{r}}(\mathbf{x}^0) > 0.$$

Die Funktion f ist also streng monoton steigend in diesen Richtungen. ■

Z Sei $D \subseteq \mathbb{R}^n$ offen. Die reelle Funktion $f : D \to Z$ in n Variablen heisst (streng) monoton steigend bzw. fallend in Richtung $\mathbf{r}$, wenn $f_{\mathbf{x}^0,\mathbf{r}}(t)$ für alle $\mathbf{x}^0 \in D \subseteq \mathbb{R}^n$ (streng) monoton steigend bzw. fallend ist.

Ist f stetig partiell differenzierbar, kann die Monotonie mithilfe der ersten Richtungsableitung $f'_{\mathbf{r}}(\mathbf{x}^0)$ überprüft werden.

9.3.2 Konvexität

Ziele dieses Unterkapitels

- Wann heisst eine reelle Funktion in n Variablen konvex, streng konvex, konkav bzw. streng konkav?
- Wie kann man die Konvexität einer stetig partiell differenzierbaren reellen Funktion mithilfe der Richtungsableitungen überprüfen?

In diesem Abschnitt beschränken wir uns auf Funktionen mit konvexem Definitionsbereich D. Für zwei Stellen $\mathbf{x}^1 \in D$ und $\mathbf{x}^2 \in D$ gilt also auch $\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2 \in D$ für alle $\alpha \in [0,1]$, vgl. Abbildung 9.43 und Kapitel 2.2. Analog zu der Charakterisierung reeller Funktionen

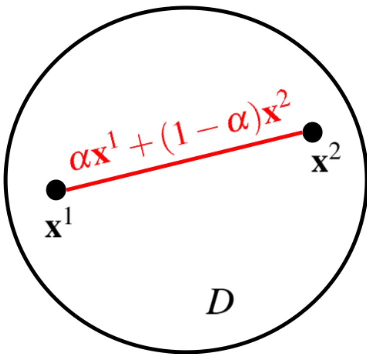

Abbildung 9.43: Darstellung einer konvexen Menge D.

in einer Variablen in Kapitel 5 nennt man eine reelle Funktion in n Variablen konvex, wenn ihr Epigraph, also die Menge aller Punkte oberhalb des Graphen, konvex ist. Anders ausgedrückt heisst eine reelle Funktion mit konvexem Definitionsbereich D in n Variablen konvex, wenn jede Verbindungslinie zweier Punkte auf oder oberhalb des Graphen wieder auf oder oberhalb des Graphen liegt, siehe Abbildung 9.44.

Bei einer streng konvexen Funktion liegen alle Verbindungslinien zweier Punkte oberhalb des Graphen auch oberhalb des Graphen, bei einer konvexen auf oder oberhalb des Graphen. Bei einer streng konkaven Funktion liegen alle Verbindungslinien zweier Punkte unterhalb des Graphen auch unterhalb des Graphen, bei einer konkaven auf oder unterhalb des Graphen.

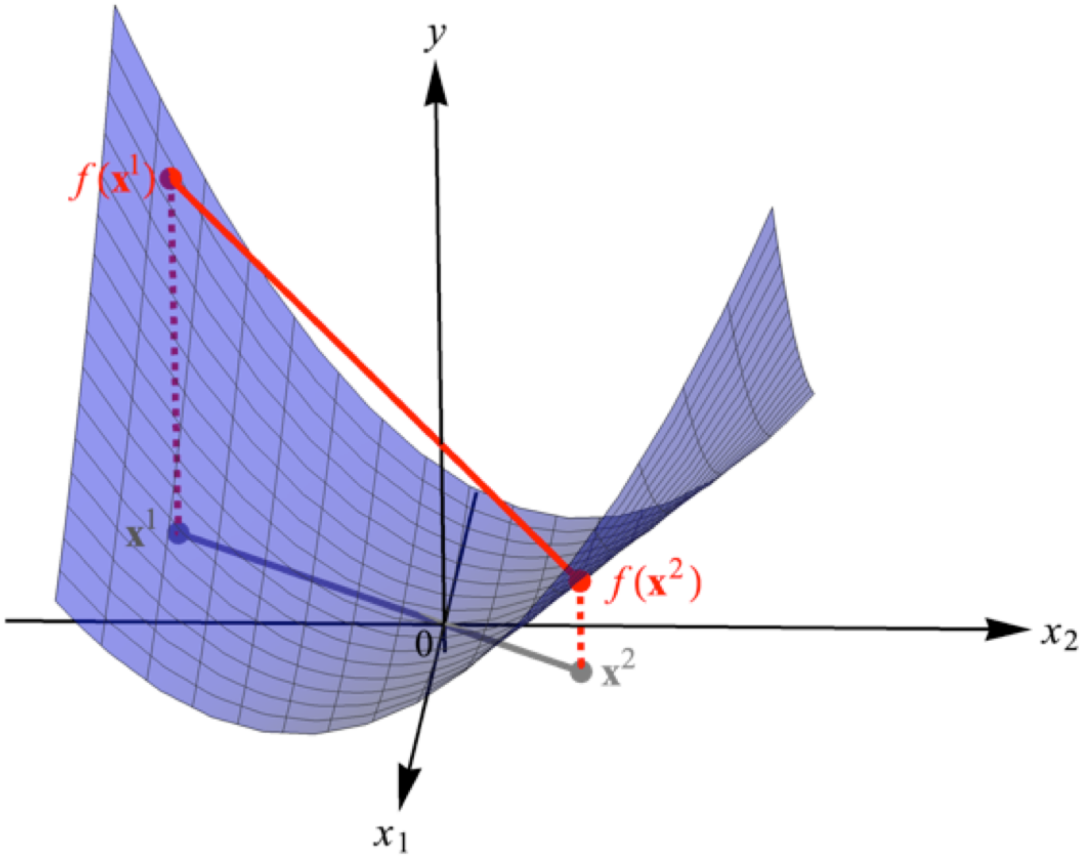

Abbildung 9.44: Konvexe Funktion.

Definition 9.3.2 — Konvexität einer reellen Funktion in n Variablen.
Sei f eine reelle Funktion in n Variablen $f : D \to Z$ mit konvexem Definitionsbereich $D \subseteq \mathbb{R}^n$. Gilt für alle $\mathbf{x}^1, \mathbf{x}^2 \in D$ mit $\mathbf{x}^1 \neq \mathbf{x}^2$ und alle $\alpha \in (0,1)$,

- $f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) \leq \alpha f(\mathbf{x}^1) + (1-\alpha)f(\mathbf{x}^2)$, dann heisst f konvex;
- $f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) < \alpha f(\mathbf{x}^1) + (1-\alpha)f(\mathbf{x}^2)$, dann heisst f streng konvex;
- $f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) \geq \alpha f(\mathbf{x}^1) + (1-\alpha)f(\mathbf{x}^2)$, dann heisst f konkav;
- $f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) > \alpha f(\mathbf{x}^1) + (1-\alpha)f(\mathbf{x}^2)$, dann heisst f streng konkav.

Dabei ergibt sich wie im Fall $n = 1$ direkt aus der Definition der Konkavität, dass eine Funktion f genau dann (streng) konkav ist, wenn $-f$ (streng) konvex ist.

Abbildung 9.45 zeigt links oben eine streng konvexe Funktion, oben in der Mitte eine streng konkave Funktion und rechts oben eine Funktion in 2 Variablen, die weder konvex noch konkav ist. Links unten zeigt sie eine konvexe Funktion, welche nicht streng konvex ist, und rechts unten eine konkave Funktion, welche nicht streng konkav ist. Ist eine Funktion konvex, so verläuft sie stets oberhalb jeder Tangentialhyperebene; ist sie konkav verläuft sie stets unterhalb jeder Tangentialhyperebene.

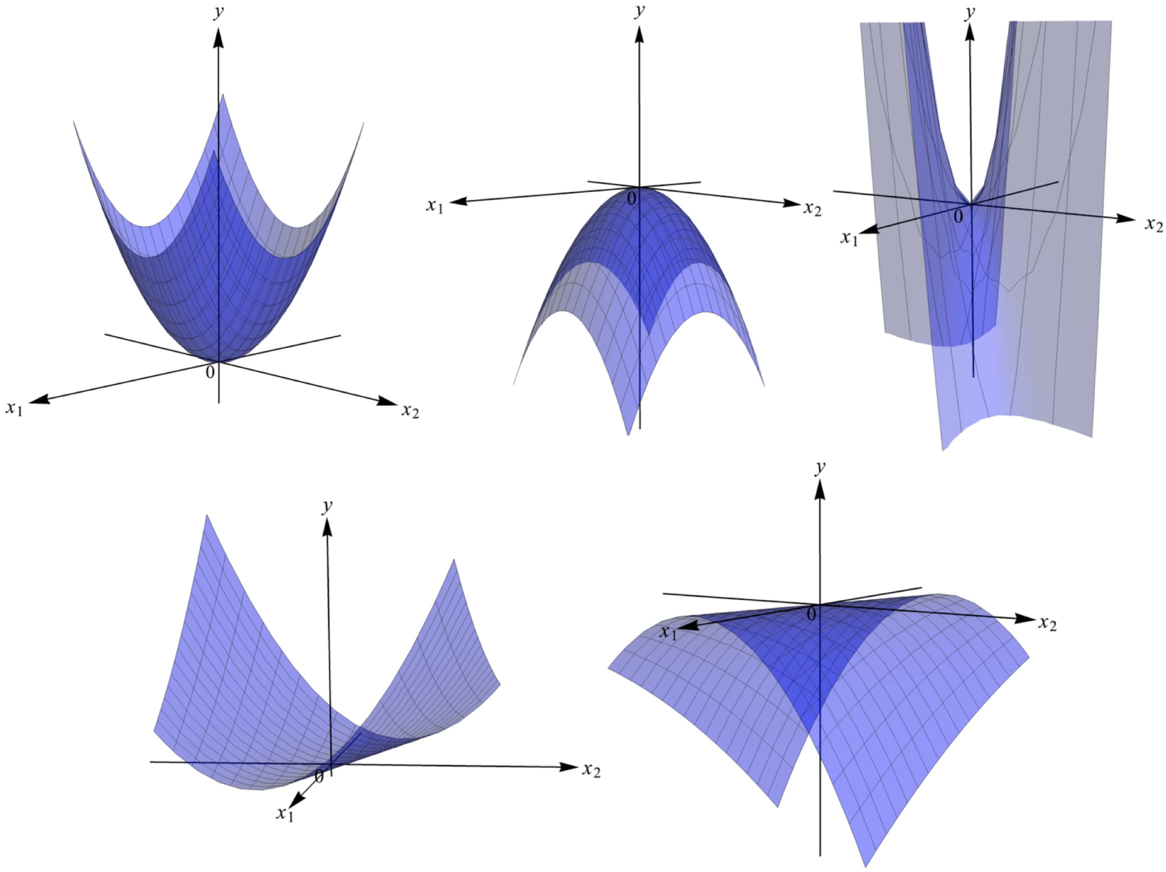

Abbildung 9.45: Darstellung verschiedener Funktionen in 2 Variablen.

Satz 9.3.2 — Tangentialebenen konvexer Funktionen.
Sei $D \subseteq \mathbb{R}^n$ eine offene, konvexe Menge. Die stetig partiell differenzierbare Funktion $f : D \to Z$ ist genau dann konvex, wenn für alle $\mathbf{x}^0 \in D$ gilt:

$$f(\mathbf{x}) \geq f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T (\mathbf{x} - \mathbf{x}^0) \text{ für alle } \mathbf{x} \in D.$$

f ist genau dann konkav, wenn für alle $\mathbf{x}^0 \in D$ gilt:

$$f(\mathbf{x}) \leq f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T (\mathbf{x} - \mathbf{x}^0) \text{ für alle } \mathbf{x} \in D.$$

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.10). ■

Wie zuvor versuchen wir die Eigenschaft der Konvexität über den Vertikalschnitt, also die Funktion $f_{\mathbf{x}^0,\mathbf{r}}(t) = f(\mathbf{x}^0 + t\mathbf{r})$ in einer Variablen, zu überprüfen.

Satz 9.3.3 — Konvexität einer Funktion und ihrer Vertikalschnitte.
Sei f eine reelle Funktion in n Variablen $f : D \to Z$ mit konvexem Definitionsbereich $D \subseteq \mathbb{R}^n$. Dann gilt:

- f ist (streng) konvex $\Leftrightarrow$ $f_{\mathbf{x}^0,\mathbf{r}}$ ist (streng) konvex für alle $\mathbf{x}^0 \in D$ und Richtungen $\mathbf{r} \in \mathbb{R}^n$;

- f ist (streng) konkav $\Leftrightarrow f_{\mathbf{x}^0,\mathbf{r}}$ ist (streng) konkav für alle $\mathbf{x}^0 \in D$ und Richtungen $\mathbf{r} \in \mathbb{R}^n$.

Beweis. Wir beweisen den Fall f konvex $\Leftrightarrow f_{\mathbf{x}^0,\mathbf{r}}$ konvex für alle $\mathbf{x}^0 \in D$ und Richtungen $\mathbf{r} \in \mathbb{R}^n$. Die anderen Fälle können analog gezeigt werden.

Wir zeigen zunächst, dass aus der Konvexität der Vertikalschnitte die Konvexität der Funktion folgt. Ist $f_{\mathbf{x}^0,\mathbf{r}}$ für alle $\mathbf{x}^0 \in D$ und Richtungen $\mathbf{r}$ konvex, so gilt laut Definition 9.3.2

$$\begin{aligned} f_{\mathbf{x}^0,\mathbf{r}}(\alpha t_1 + (1-\alpha)t_2) &\leq \alpha f_{\mathbf{x}^0,\mathbf{r}}(t_1) + (1-\alpha) f_{\mathbf{x}^0,\mathbf{r}}(t_2) \text{ bzw.} \\ f(\mathbf{x}^0 + (\alpha t_1 + (1-\alpha)t_2)\mathbf{r}) &= f(\alpha(\mathbf{x}^0 + t_1\mathbf{r}) + (1-\alpha)(\mathbf{x}^0 + t_2\mathbf{r})) \\ &\leq \alpha f(\mathbf{x}^0 + t_1\mathbf{r}) + (1-\alpha) f(\mathbf{x}^0 + t_2\mathbf{r}) \end{aligned}$$

für alle $\mathbf{x}^0 + t_1\mathbf{r}, \mathbf{x}^0 + t_2\mathbf{r} \in D$. Alle Punkte $\mathbf{x}^1$, $\mathbf{x}^2$ aus dem Definitionsbereich lassen sich für ein geeignetes $\mathbf{x}^0$ darstellen als $\mathbf{x}^1 = \mathbf{x}^0 + t_1\mathbf{r}$ und $\mathbf{x}^2 = \mathbf{x}^0 + t_2\mathbf{r}$. Es gilt also auch

$$f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) \leq \alpha f(\mathbf{x}^1) + (1-\alpha) f(\mathbf{x}^2).$$

Die Funktion f ist in diesem Fall also konvex.

Umgekehrt folgt aus

$$f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) \leq \alpha f(\mathbf{x}^1) + (1-\alpha) f(\mathbf{x}^2)$$

für alle $\mathbf{x}^1, \mathbf{x}^2 \in D$ mit $\mathbf{x}^1 = \mathbf{x}^0 + t_1\mathbf{r}$ und $\mathbf{x}^2 = \mathbf{x}^0 + t_2\mathbf{r}$ direkt

$$f_{\mathbf{x}^0,\mathbf{r}}(\alpha t_1 + (1-\alpha)t_2) \leq \alpha f_{\mathbf{x}^0,\mathbf{r}}(t_1) + (1-\alpha) f_{\mathbf{x}^0,\mathbf{r}}(t_2).$$

■

Die Funktion f ist daher (streng) konvex, wenn die Vertikalschnitte in Richtung $\mathbf{r}$ ausgehend von der Stelle $\mathbf{x}^0$, $f_{\mathbf{x}^0,\mathbf{r}}(t) = f(\mathbf{x}^0 + t\mathbf{r})$, für alle Richtungen und alle Stellen (streng) konvex ist. Auch umgekehrt sind alle Vertikalschnitte $f_{\mathbf{x}^0,\mathbf{r}}$ (streng) konvex, wenn f (streng) konvex ist.

Wir demonstrieren dies zunächst am Beispiel bevor wir das Vorgehen der Überprüfung in einem Satz zusammenfassen.

▪ Beispiel 9.3.2 — Konvexität einer quadratischen Funktion.
In Beispiel 9.2.15 betrachteten wir eine quadratische Funktion mit Abbildungsvorschrift $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^2 + 3$. An der Stelle $\mathbf{x}^0 = (x_1^0, x_2^0)^T \in \mathbb{R}^2$ ist der Gradient

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2x_1^0 \\ 8(x_2^0 - 1) \end{pmatrix}$$

und die Hesse-Matrix

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

In Beispiel 9.2.23 untersuchten wir $f_{\mathbf{x}^0,\mathbf{r}}$, die Vertikalschnitte von f ausgehend von einer Stelle $\mathbf{x}^0$ in verschiedene Richtungen $\mathbf{r}$. Die Ableitung dieser Funktion ergibt die erste Richtungsableitung in Richtung $\mathbf{r}$

$$f'_{\mathbf{r}}(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)^T\mathbf{r} = 2x_1^0 r_1 + 8(x_2^0 - 1) r_2.$$

Die zweite Richtungsableitung in Richtung $\mathbf{r}$ ist

$$f''_{\mathbf{r}}(\mathbf{x}^0) = \mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} = 2r_1^2 + 8r_2^2.$$

An der Stelle $\mathbf{x}^0$ ist die zweite Richtungsableitung für jede beliebige Richtung $\mathbf{r}$ streng grösser als 0, weil

$$\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} = 2r_1^2 + 8r_2^2 > 0 \text{ für alle Richtungen } \mathbf{r}, \text{ da } \mathbf{r} \neq \mathbf{0}.$$

Damit ist die Funktion $f_{\mathbf{x}^0,\mathbf{r}}$ für alle $\mathbf{x}^0$ und Richtungen $\mathbf{r}$ streng konvex. Aus obigem Satz folgt somit, dass f streng konvex ist und für alle $\mathbf{x}^1, \mathbf{x}^2 \in D$ gilt:

$$f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) < \alpha f(\mathbf{x}^1) + (1-\alpha) f(\mathbf{x}^2).$$

■

Aus Kapitel 5 wissen wir, dass der Vertikalschnitt $f_{\mathbf{x}^0,\mathbf{r}}$ genau dann konvex ist, wenn die zweiten Ableitungen von $f_{\mathbf{x}^0,\mathbf{r}}$ grösser oder gleich 0 sind. Sind die zweiten Ableitungen echt grösser als 0, ist $f_{\mathbf{x}^0,\mathbf{r}}$ streng konvex. Umgekehrt kann eine streng konvexe Funktion jedoch an einigen Punkten auch eine zweite Ableitung von 0 haben. Der Vertikalschnitt ist genau dann konkav, wenn seine zweiten Ableitungen kleiner oder gleich 0 sind. Sind die zweiten Ableitungen kleiner als 0, ist $f_{\mathbf{x}^0,\mathbf{r}}$ streng konkav. Ist f zweimal stetig partiell differenzierbar, ist die zweite Ableitung des Vertikalschnitts an der Stelle $\mathbf{x}^0$ in Richtung $\mathbf{r}$ gleich

$$f''_{\mathbf{r}}(\mathbf{x}^0) = \mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \quad \text{für alle } \mathbf{r},$$

vgl. Definition 9.2.10. Folgender Satz nutzt den Zusammenhang zwischen zweiten Ableitungen und Konvexität sowie die Tatsache, dass eine Funktion f genau dann (streng) konvex bzw. konkav ist, wenn alle Vertikalschnitte an allen Stellen (streng) konvex bzw. konkav sind.

Satz 9.3.4 — Konvexitätskriterium.
Sei D eine offene konvexe Menge, $f: D \to Z$ eine zweimal stetig differenzierbare reelle Funktion in n Variablen und $H_f(\mathbf{x}^0)$ die Hesse-Matrix von f an der Stelle $\mathbf{x}^0 \in D$. Dann gilt:

- $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \geq 0$ für alle $\mathbf{x}^0 \in D$, $\mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \Leftrightarrow f$ ist konvex in D;
- $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} > 0$ für alle $\mathbf{x}^0 \in D$, $\mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \Rightarrow f$ ist streng konvex in D;
- $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \leq 0$ für alle $\mathbf{x}^0 \in D$, $\mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \Leftrightarrow f$ ist konkav in D;
- $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} < 0$ für alle $\mathbf{x}^0 \in D$, $\mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \Rightarrow f$ ist streng konkav in D.

Beweis. Für Vektoren $\mathbf{r} \in \mathbb{R}^n$ mit Norm 1 folgen die Bedingungen für die zweite Richtungsableitung $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r}$ direkt aus Definition 9.2.10 und Satz 9.2.7 sowie dem Konvexitätskriterium für zweimal differenzierbare Funktionen in einer Variablen, Satz 5.5.12.

Da zudem für jeden Vektor $\mathbf{r} \in \mathbb{R}^n, \mathbf{r} \neq \mathbf{0}$ und $\mathbf{x}^0 \in D$ gilt, dass

$$\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} > 0 \Leftrightarrow \underbrace{\frac{1}{||\mathbf{r}||}\mathbf{r}^T}_{\tilde{\mathbf{r}}^T} H_f(\mathbf{x}^0) \underbrace{\frac{1}{||\mathbf{r}||}\mathbf{r}}_{\tilde{\mathbf{r}}^T} > 0,$$

ist es unerheblich, ob nur Richtungen, also Vektoren mit Norm 1, oder beliebige Vektoren mit Ausnahme des Nullvektors geprüft werden. ■

Aus der Konvexität einer Funktion in alle Richtungen $\mathbf{r}$ kann man also schliessen, dass die Funktion f selbst konvex sein muss.[8] Hierbei ist es wichtig, dass wirklich alle Richtungen $\mathbf{r}$ betrachtet werden. Dies illustrieren wir in folgendem Beispiel.

■ Beispiel 9.3.3 — Wichtigkeit alle Richtungen zu prüfen.
Wir betrachten erneut die Funktion mit Abbildungsvorschrift $f(\mathbf{x}) = x_1^2 - 3x_1x_2 + x_2^2$. Ausgehend von $\mathbf{x}^0 = \mathbf{0}$ in Richtung $\mathbf{r} = \mathbf{e}^1$ ist $f_{\mathbf{0},\mathbf{e}^1}(t) = t^2$. Auch in Richtung $\mathbf{r} = \mathbf{e}^2$ gilt $f_{\mathbf{0},\mathbf{e}^2}(t) = t^2$. In Richtungen $\mathbf{r} = \mathbf{e}^1$ und $\mathbf{r} = \mathbf{e}^2$ ist die Funktion also ausgehend von $\mathbf{x}^0 = \mathbf{0}$ konvex, vgl. Abbildung 9.46. Betrachtet man jedoch einen diagonalen Vertikalschnitt in

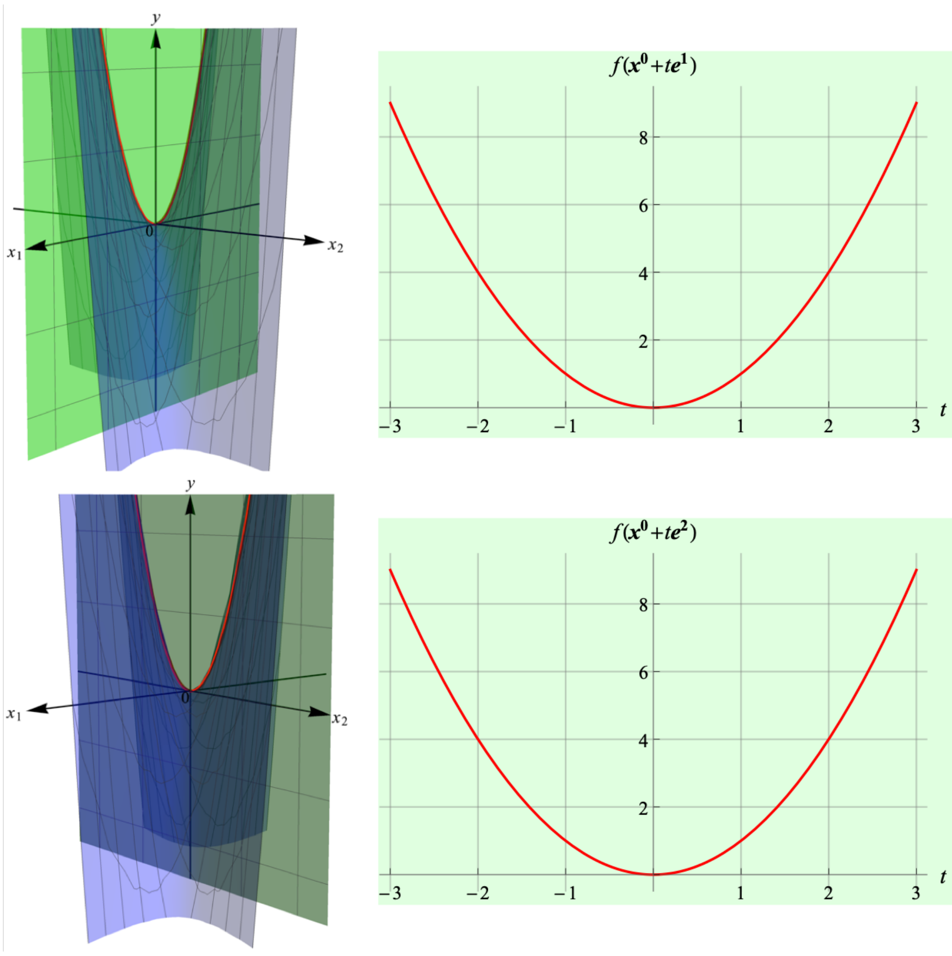

Abbildung 9.46: Senkrechte Vertikalschnitte der Funktion $f(\mathbf{x}) = x_1^2 - 3x_1x_2 + x_2^2$ durch $\mathbf{x}^0 = \mathbf{0}$.

Richtung $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$, ergibt sich eine konkave Funktion, vgl. Abbildung 9.47. Dies kann man auch wie folgt mithilfe der Hesse-Matrix nachprüfen: Der Gradient von f an der Stelle $\mathbf{x}^0 = (x_1^0, x_2^0)^T$ ist gegeben durch

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2x_1^0 - 3x_2^0 \\ 2x_2^0 - 3x_1^0 \end{pmatrix}.$$

[8] Aus dem Beweis von Satz 9.3.4 ergibt sich, dass es unerheblich ist, ob man alle Vektoren $\mathbf{r} \neq \mathbf{0}$ oder alle Richtungen (also Vektoren mit Norm 1) prüft.

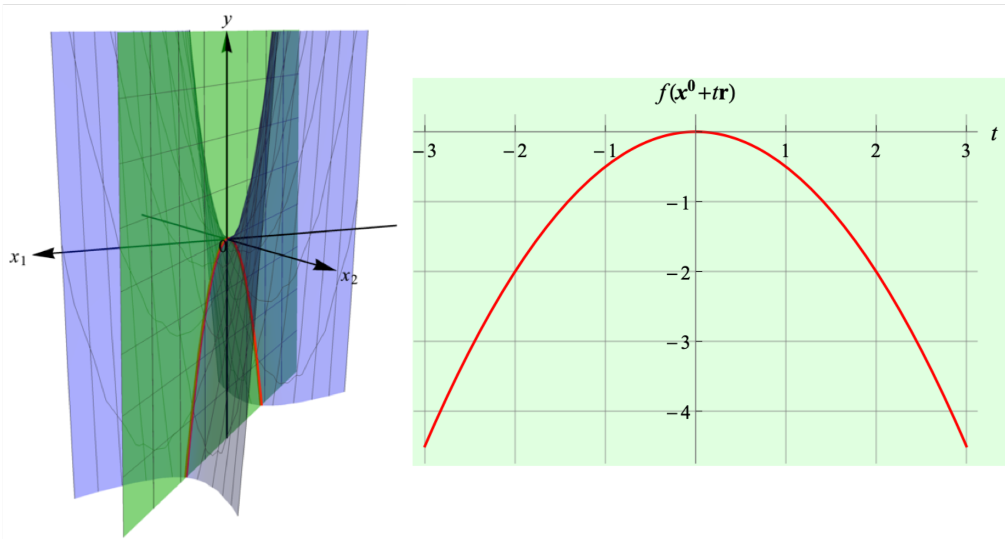

Abbildung 9.47: Diagonaler Vertikalschnitt der Funktion $f(\mathbf{x}) = x_1^2 - 3x_1x_2 + x_2^2$ durch $\mathbf{x}^0 = \mathbf{0}$.

Hieraus ergibt sich die Hesse-Matrix von f als

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}.$$

Für $\mathbf{r} = \mathbf{e}^1$ gilt nun

$$(\mathbf{e}^1)^T H_f(\mathbf{x}^0)\mathbf{e}^1 = (1,0)\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2 > 0$$

und für $\mathbf{r} = \mathbf{e}^2$

$$(\mathbf{e}^2)^T H_f(\mathbf{x}^0)\mathbf{e}^2 = (0,1)\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 2 > 0.$$

Dennoch ist die Funktion f nicht konvex, da man zum Beispiel für $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$

$$\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = -1 < 0$$

erhält. Man sieht also, dass die geforderte Ungleichung zwar für die kanonischen Einheitsvektoren $\mathbf{e}^1$ und $\mathbf{e}^2$, also in Richtung der Koordinatenachsen gilt, aber nicht in Richtung des Vektors $\mathbf{r} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$. Die Funktion ist also weder konvex noch konkav. ■

Laut Satz 9.3.4 kann man also aus der Tatsache, dass die zweite Richtungsableitung von f in alle Richtungen grösser oder gleich 0 ist, schliessen, dass f konvex ist und umgekehrt. Ist die zweite Richtungsableitung in alle Richtungen (streng) positiv, so ist auch f streng konvex. Wie schon im Fall einer Funktion in nur einer Variablen kann man aber umgekehrt nicht schliessen, dass die zweite Richtungsableitung einer streng konvexen Funktion stets (streng) positiv ist. Wir zeigen dies in folgendem Beispiel.

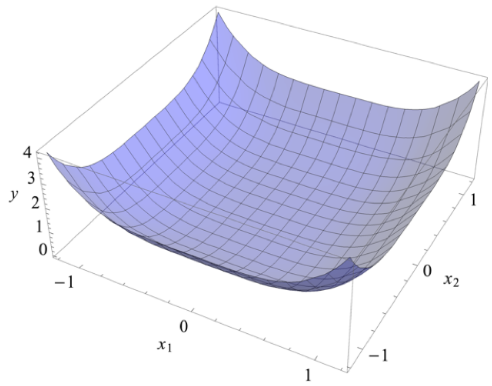

Abbildung 9.48: Graph der Funktion $f(\mathbf{x}) = (x_1)^4 + (x_2)^4$.

■ Beispiel 9.3.4 — Eine streng konvexe Funktion.
Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = (x_1)^4 + (x_2)^4$ ist in Abbildung 9.48 dargestellt. Der Gradient und die Hesse-Matrix von f an der Stelle $\mathbf{x}^0 = (x_1^0, x_2^0)$ sind[9]

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 4(x_1^0)^3 \\ 4(x_2^0)^3 \end{pmatrix} \text{ und } H_f(\mathbf{x}^0) = \begin{pmatrix} 12(x_1^0)^2 & 0 \\ 0 & 12(x_2^0)^2 \end{pmatrix}.$$

Die zweite Richtungsableitung in Richtung $\mathbf{r}$ berechnet sich an der Stelle $\mathbf{x}^0$ als

$$\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} = 12(r_1)^2(x_1^0)^2 + 12(r_2)^2(x_2^0)^2 \geq 0 \text{ für alle } \mathbf{r} \neq \mathbf{0}.$$

Allerdings ist an der Stelle $\mathbf{x}^0 = (0,0)^T$ die zweite Richtungsableitung für alle $\mathbf{r}$ gleich 0. Wir können aus Satz 9.3.4 also schliessen, dass f konvex ist. Wir können aus Satz 9.3.4 jedoch nicht schliessen, dass f streng konvex ist. Überprüft man für $\mathbf{x}^1 \neq \mathbf{x}^2$ die Ungleichung der Definition 9.3.2 direkt und nutzt, dass die Funktion von $\mathbb{R} \to \mathbb{R}$ mit Abbildungsvorschrift $y = x^4$ streng konvex ist, ergibt sich

$$\begin{aligned} f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) &= (\alpha x_1^1 + (1-\alpha)x_1^2)^4 + (\alpha x_2^1 + (1-\alpha)x_2^2)^4 \\ &< \alpha(x_1^1)^4 + (1-\alpha)(x_1^2)^4 + \alpha(x_2^1)^4 + (1-\alpha)(x_2^2)^4 \\ &= \alpha((x_1^1)^4 + (x_2^1)^4) + (1-\alpha)((x_1^2)^4 + (x_2^2)^4) \\ &= \alpha f(\mathbf{x}^1) + (1-\alpha)f(\mathbf{x}^2). \end{aligned}$$

Die Funktion f ist also streng konvex. ■

So wie im Fall reeller Funktionen in $n = 1$ Variable eine affin-lineare Funktion sowohl konvex als auch konkav ist, bezeichnet man eine affin-lineare Funktion auch in mehreren Variablen allgemein sowohl als konvex als auch als konkav.

■ Beispiel 9.3.5 — Konvexität einer affin-linearen Funktion.
In Beispiel 9.2.11 hatten wir gezeigt, dass die affin-lineare Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = -8x_1 + 3x_2 + 10$ die Hesse-Matrix

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

[9] Um Verwechslungen mit dem hochgestellten Index 0 an der Stelle $\mathbf{x}^0$ zu vermeiden, nutzen wir Klammern um Potenzen zu bilden. So ist beispielsweise $(x_2^0)^2$ das Quadrat von x_2^0 und x_2^2 die zweite Komponente von $\mathbf{x}^2$.

hat. Es gilt also für alle $\mathbf{r} \in \mathbb{R}^n$ an jeder Stelle $\mathbf{x}^0 \in D$:

$$\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} = 0.$$

Damit gilt $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \geq 0$ für alle Richtungen $\mathbf{r} \in \mathbb{R}^n$ und Stellen $\mathbf{x}^0 \in D$. Die Funktion ist also konvex. Zudem gilt $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \leq 0$ für alle Richtungen $\mathbf{r} \in \mathbb{R}^n$ und Stellen $\mathbf{x}^0 \in D$. Die Funktion ist also auch konkav. ∎

Z Eine reelle Funktion in n Variablen heisst

$$\left.\begin{array}{l}\text{konvex}\\ \text{streng konvex}\\ \text{konkav}\\ \text{streng konkav}\end{array}\right\}, \text{ wenn } f(\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2) \left\{\begin{array}{c}\leq\\ <\\ \geq\\ >\end{array}\right\} \alpha f(\mathbf{x}^1) + (1-\alpha) f(\mathbf{x}^2).$$

Sei $D \subseteq \mathbb{R}^n$ offen und konvex. Für eine zweimal stetig partiell differenzierbare reelle Funktion f in n Variablen mit $f: D \to Z$ gilt:

$$f''_{\mathbf{r}}(\mathbf{x}^0) = \mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \left\{\begin{array}{c}\geq\\ \leq\end{array}\right\} 0 \text{ für alle } \mathbf{x}^0 \in D, \mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \Leftrightarrow f \text{ ist } \left\{\begin{array}{l}\text{konvex}\\ \text{konkav}\end{array}\right. .$$

$$f''_{\mathbf{r}}(\mathbf{x}^0) = \mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \left\{\begin{array}{c}>\\ <\end{array}\right\} 0 \text{ für alle } \mathbf{x}^0 \in D, \mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \Rightarrow f \text{ ist } \left\{\begin{array}{l}\text{streng konvex}\\ \text{streng konkav}\end{array}\right. .$$

9.3.3 Definitheit von Matrizen zur Prüfung von Konvexität

Ziele dieses Unterkapitels

- Wann nennt man eine symmetrische Matrix positiv definit, positiv semidefinit, negativ definit, negativ semidefinit und wann indefinit?
- Welcher Zusammenhang besteht zwischen den Eigenwerten einer symmetrischen Matrix und der Klassifikation der Matrix als positiv definit, positiv semidefinit, negativ definit, negativ semidefinit oder indefinit?
- Was versteht man unter Hauptunterdeterminanten einer symmetrischen Matrix?
- Welcher Zusammenhang besteht zwischen den Hauptunterdeterminanten einer symmetrischen Matrix und der Klassifikation der Matrix als positiv definit, positiv semidefinit, negativ definit, negativ semidefinit oder indefinit?

Dem Ausdruck $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} > 0$ für alle $\mathbf{r} \neq \mathbf{0}$ kommt bei reellen Funktionen in n Variablen also eine Rolle zu, die vergleichbar ist mit der Rolle einer positiven zweiten Ableitung. Man sagt auch die Hesse-Matrix $H_f(\mathbf{x}^0)$ ist positiv definit, wenn $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} > 0$ für alle $\mathbf{r} \neq \mathbf{0}$, sie ist positiv semidefinit, wenn $\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} \geq 0$ für alle $\mathbf{r} \neq \mathbf{0}$. Analog definiert man negativ (semi-)definite Hesse-Matrizen.

Um die Verbindungen zu Kapiteln 6-8 zu verdeutlichen, nennen wir die betrachtete Matrix in folgenden Definitionen und Sätzen A. All diese Definitionen und Sätze gelten allgemein für quadratische Matrizen. Im Kontext der Diskussion von reellen Funktionen in n Variablen ist hier aber stets die Hesse-Matrix gemeint.

Definition 9.3.3 — Definitheit von Matrizen.
Sei A eine symmetrische $n \times n$-Matrix. A heisst

- **positiv semidefinit**, $A \succcurlyeq 0$, wenn $\mathbf{r}^T A\mathbf{r} \geq 0$ für alle $\mathbf{r} \neq \mathbf{0}$;
- **positiv definit**, $A \succ 0$, wenn $\mathbf{r}^T A\mathbf{r} > 0$ für alle $\mathbf{r} \neq \mathbf{0}$;

- **negativ semidefinit**, $A \preceq 0$, wenn $\mathbf{r}^T A\mathbf{r} \leq 0$ für alle $\mathbf{r} \neq \mathbf{0}$;
- **negativ definit**, $A \prec 0$, wenn $\mathbf{r}^T A\mathbf{r} < 0$ für alle $\mathbf{r} \neq \mathbf{0}$;
- **indefinit**, wenn A weder positiv definit, positiv semidefinit, negativ definit, noch negativ semidefinit ist.

Jede positiv definite Matrix ist also auch positiv semidefinit, jede negativ definite Matrix ist auch negativ semidefinit, aber nicht umgekehrt, vgl. Abbildung 9.49. Die einzige Matrix, welche sowohl positiv als auch negativ semindefinit ist, ist die Nullmatrix.

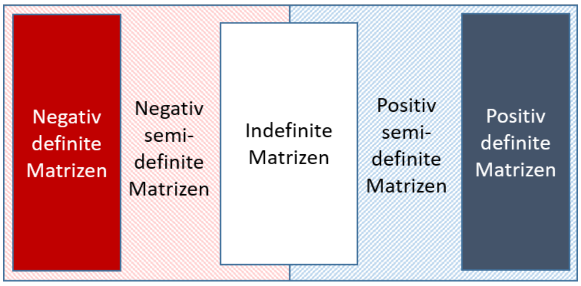

Abbildung 9.49: Definitheit symmetrischer Matrizen.

Eigenwerte zur Prüfung der Definitheit

Das Vorzeichen von $\mathbf{r}^T A\mathbf{r}$ für alle $\mathbf{r} \neq \mathbf{0}$ nachzuprüfen, ist für grosse Matrizen aufwendig. Oft kann man an den Vorzeichen der Eigenwerte einfacher ablesen, ob eine Matrix positiv definit ist.

Satz 9.3.5 — Eigenwerte und Definitheit.
Sei A eine symmetrische $n \times n$-Matrix mit Eigenwerten $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Dann ist A

- positiv semidefinit genau dann, wenn $\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$;
- positiv definit genau dann, wenn $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$;
- negativ semidefinit genau dann, wenn $\lambda_1, \lambda_2, \ldots, \lambda_n \leq 0$;
- negativ definit genau dann, wenn $\lambda_1, \lambda_2, \ldots, \lambda_n < 0$;
- indefinit genau dann, wenn mindestens ein Eigenwert positiv und einer negativ ist.

Beweis (#):
Wir nutzen hier die im Zusatzkapitel 8.4.4 vorgestellte Theorie: Wir bezeichnen die orthogonale Matrix der Eigenvektoren als V und die Diagonalmatrix der zugehörigen Eigenwerte als L. Definieren wir $\mathbf{r} = V\mathbf{x}$ für $\mathbf{x} \neq \mathbf{0}$, dann gilt

$$\begin{aligned}
&\mathbf{r}^T A\mathbf{r} > 0 \text{ für alle } \mathbf{r} \neq \mathbf{0} \\
&\qquad \Leftrightarrow (V\mathbf{x})^T A(V\mathbf{x}) = \mathbf{x}^T V^T AV\mathbf{x} = \mathbf{x}^T L\mathbf{x} = \sum_{i=1}^{n} \lambda_i x_i^2 > 0 \text{ für alle } \mathbf{x} \neq \mathbf{0} \\
&\qquad \Leftrightarrow \lambda_1, \ldots, \lambda_n > 0.
\end{aligned}$$

Die letzte Äquivalenz, $\sum_{i=1}^n \lambda_i x_i^2 > 0$ für alle $\mathbf{x} \neq \mathbf{0} \Leftrightarrow \lambda_1, \ldots, \lambda_n > 0$, erklärt sich wie folgt: Wenn $\sum_{i=1}^n \lambda_i x_i^2 > 0$ für alle $\mathbf{x} \neq \mathbf{0}$ gilt, dann gilt dies insbesondere für $\mathbf{x} = \mathbf{e}^i$ mit $i \in \{1, \ldots, n\}$. Setzt man also $\mathbf{x} = \mathbf{e}^i$ so sieht man, dass $\lambda_i > 0$ gelten muss. Damit gilt die Implikation $\sum_{i=1}^n \lambda_i x_i^2 > 0$ für alle $\mathbf{x} \neq \mathbf{0} \Rightarrow \lambda_1, \ldots, \lambda_n > 0$. Dass die umgekehrte Implikation $\lambda_1, \ldots, \lambda_n > 0 \Rightarrow \sum_{i=1}^n \lambda_i x_i^2 > 0$ für alle $\mathbf{x} \neq \mathbf{0}$ gilt, zeigt man in zwei Schritten:

Offensichtlich gilt für $\lambda_1, \ldots, \lambda_n > 0$, dass $\lambda_i x_i^2 \geq 0$ für alle $i \in \{1, \ldots, n\}$. Da $\mathbf{x} \neq \mathbf{0}$, ist mindestens eine Komponente von $\mathbf{x}$ nicht 0. Damit gibt es ein $i \in \{1, \ldots, n\}$ mit $\lambda_i x_i^2 > 0$ und somit $\sum_{i=1}^n \lambda_i x_i^2 > 0$ für alle $\mathbf{x} \neq \mathbf{0}$.

Damit ist A genau dann positiv definit, wenn alle Eigenwerte echt grösser als null sind. Die anderen Fälle kann man analog zeigen. ■

Wir demonstrieren die Anwendung der Kriterien aus Satz 9.3.5 anhand von Beispielen.

■ Beispiel 9.3.6 — Überprüfung der Definitheit mittels Eigenwerten.
Betrachten wir erneut die Hesse-Matrix der quadratischen Funktion aus Beispiel 9.3.2 an einer beliebigen Stelle $\mathbf{x}^0$,

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

Wie zeigten dort, dass

$$\mathbf{r}^T H_f(\mathbf{x}^0)\mathbf{r} = 2r_1^2 + 8r_2^2 > 0$$

für alle $\mathbf{r} \neq \mathbf{0}$ gilt. Die Matrix ist also positiv definit. Um das Kriterium aus Satz 9.3.5 anzuwenden, benötigen wir zunächst die Eigenwerte dieser Matrix. Diese ergeben sich aus den Nullstellen des charakteristischen Polynoms

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 0 \\ 0 & 8-\lambda \end{vmatrix} = (2-\lambda)(8-\lambda).$$

Die Nullstellen des Polynoms $(2-\lambda)(8-\lambda)$ sind 2 und 8. Damit sind die Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = 8$. Wegen $\lambda_1, \lambda_2 > 0$ kann man also auch anhand von Satz 9.3.5 schliessen, dass die Hesse-Matrix positiv definit ist.

Multipliziert man alle Einträge der Matrix $H_f(\mathbf{x}^0)$ mit (-1) ergibt sich die Matrix

$$\begin{pmatrix} -2 & 0 \\ 0 & -8 \end{pmatrix}$$

mit Eigenwerten $\lambda_1 = -2$ und $\lambda_2 = -8$. Wegen $\lambda_1, \lambda_2 < 0$ kann man schliessen, dass diese Matrix negativ definit sein muss. Es gilt also

$$\mathbf{r}^T \begin{pmatrix} -2 & 0 \\ 0 & -8 \end{pmatrix} \mathbf{r} < 0$$

für alle $\mathbf{r} \neq \mathbf{0}$. ■

■ Beispiel 9.3.7 — Die Hesse-Matrix aus Beispiel 9.3.3.
Wir untersuchen die Hesse-Matrix der Funktion aus Beispiel 9.3.3 an einer beliebigen Stelle $\mathbf{x}^0$,

$$\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}.$$

Auch hier können wir die Eigenwerte mithilfe der charakteristischen Gleichung $\det(A - \lambda I) = 0$ mit

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & -3 \\ -3 & 2-\lambda \end{vmatrix} = (2-\lambda)(2-\lambda) - 9 = 4 - 4\lambda + \lambda^2 - 9 = \lambda^2 - 4\lambda - 5$$

berechnen. Mit der Mitternachtsformel ist $\lambda^2 - 4\lambda - 5 = 0$ für $\lambda_1 = 5 > 0$ und $\lambda_2 = -1 < 0$. Die Matrix ist nach Satz 9.3.5 also indefinit. Es gibt demnach mindestens einen Vektor $\mathbf{r} \neq \mathbf{0}$ mit

$$\mathbf{r}^T \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} \mathbf{r} < 0$$

und mindestens einen weiteren Vektor $\mathbf{r} \neq \mathbf{0}$ mit

$$\mathbf{r}^T \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} \mathbf{r} > 0.$$

■

■ **Beispiel 9.3.8 — Die Hesse-Matrix der Cobb Douglas Funktion.**
Für die Hesse-Matrix der Cobb Douglas Funktion aus Beispiel 9.2.13 an der Stelle $(1,1)^T$,

$$\begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix},$$

berechnen wir analog

$$\begin{aligned} \begin{vmatrix} -0.16-\lambda & 0.16 \\ 0.16 & -0.16-\lambda \end{vmatrix} &= (-0.16-\lambda)^2 - 0.0256 \\ &= 0.0256 + 0.32\lambda + \lambda^2 - 0.0256 = \lambda^2 + 0.32\lambda \\ &= \lambda(\lambda + 0.32). \end{aligned}$$

Die Nullstellen sind $\lambda_1 = 0$ und $\lambda_2 = -0.32 < 0$. Die Matrix ist also negativ semidefinit. ■

■ **Beispiel 9.3.9 — Die Matrix aus Beispiel 8.4.9.**
In Beispiel 8.4.9 berechneten wir die Eigenwerte der Matrix

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

als $\lambda_1 = 2 > 0$ und $\lambda_2 = \lambda_3 = -1 < 0$. Diese Matrix ist also indefinit. ■

■ **Beispiel 9.3.10 — Zwei weitere Matrizen.**
Die Eigenwerte der Matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

sind $\lambda_1 = 1$, $\lambda_2 = 0$ und $\lambda_3 = -1$. Da auch diese Matrix sowohl negative als auch positive Eigenwerte hat, ist sie indefinit.

Die Eigenwerte der Matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

hingegen sind $\lambda_1 = 1$, $\lambda_2 = 0$ und $\lambda_3 = 1$. Diese Matrix ist positiv semidefinit. ■

Hauptunterdeterminanten zur Prüfung der Definitheit

Auch Eigenwerte sind nicht immer einfach zu berechnen. Determinanten stellen einfacher zu berechnende Kenngrössen von Matrizen dar. Besondere Determinanten, die Hauptunterdeterminanten, stellen eine weitere Art dar, Matrizen auf ihre Definitheitseigenschaften zu untersuchen. Hierbei definiert man die n Hauptunterdeterminanten einer symmetrischen $n \times n$-Matrix wie folgt:

Definition 9.3.4 — Hauptunterdeterminanten.
Sei $A = (a_{ij})$ eine symmetrische $n \times n$-Matrix. Dann heisst $\det(U_1) = a_{11}$ erste Hauptunterdeterminante und für $i = 2, \ldots, n$

$$\det(U_i) = \det \begin{pmatrix} a_{11} & \ldots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \ldots & a_{ii} \end{pmatrix}$$

die i-te Hauptunterdeterminante von A.

Somit ist also beispielsweise die zweite Hauptunterdeterminante der $n \times n$-Matrix A $\det(U_2) = a_{11}a_{22} - a_{12}a_{21}$. Ein hilfreicher, aber relativ schwer zu beweisender Satz ist folgender:

Satz 9.3.6 — Hauptunterdeterminantenkriterium.
Sei A eine symmetrische $n \times n$-Matrix. Dann gilt:

- A ist positiv semidefinit $\Rightarrow \det(U_i) \geq 0$ für alle $i = 1, \ldots, n$;
- A ist positiv definit $\Leftrightarrow \det(U_i) > 0$ für alle $i = 1, \ldots, n$;
- A ist negativ semidefinit $\Rightarrow \det(U_i) \geq 0$ für alle geraden $i = 2, 4, \ldots$ und $\det(U_i) \leq 0$ für alle ungeraden $i = 1, 3, \ldots$, also $\det(U_i)(-1)^i \geq 0$;
- A ist negativ definit $\Leftrightarrow \det(U_i) > 0$ für alle geraden $i = 2, 4, \ldots$ und $\det(U_i) < 0$ für alle ungeraden $i = 1, 3, \ldots$, also $(-1)^i \det(U_i) > 0$ für alle $i = 1, \ldots, n$;

Beweis. Einen Beweis findet man in Fischer (1995, Seite 302). ■

Wir demonstrieren diesen Satz an Beispielen:

▪ Beispiel 9.3.11 — Überprüfung der Definitheit mittels Hauptunterdeterminanten.
Wir betrachten erneut die (positiv definite) Hesse-Matrix der quadratischen Funktion aus Beispiel 9.3.6,

$$\begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

Um das Kriterium aus Satz 9.3.6 anzuwenden, berechnen wir die Hauptunterdeterminanten:

$$\det(U_1) = 2, \; \det(U_2) = 2 \cdot 8 - 0 \cdot 0 = 16.$$

Da $\det(U_1) > 0$ und $\det(U_2) > 0$ können wir mit Satz 9.3.6 daraus schliessen, dass die Matrix A positiv definit ist. Dies stimmt natürlich mit dem Ergebnis aus Beispiel 9.3.6 überein. ▪

■ Beispiel 9.3.12 — Die Hesse-Matrix der Cobb Douglas Funktion.
Wir wissen bereits aus Beispiel 9.3.8, dass die Matrix

$$\begin{pmatrix} -0.16 & 0.16 \\ 0.16 & -0.16 \end{pmatrix}$$

negativ semidefinit ist. Für die Hauptunterdeterminanten muss $\det(U_1) \leq 0$ und $\det(U_2) \geq 0$ gelten. Dies bestätigt sich durch Nachrechnen, da

$$\det(U_1) = -0.16,\ \det(U_2) = (-0.16)\cdot(-0.16) - 0.16\cdot 0.16 = 0.$$

■

Eine negativ semidefinite Matrix hat stets für gerade i Hauptunterdeterminanten $\det(U_i)$, die positiv oder gleich 0 sind, und für ungerade i Hauptunterdeterminanten $\det(U_i)$, die negativ oder 0 sind. Allerdings kann man aus diesen Eigenschaften der Hauptunterdeterminanten im Allgemeinen nicht umgekehrt schliessen, dass eine Matrix positiv oder negativ semidefinit sein muss, wie man in folgendem Beispiel erkennen kann:

■ Beispiel 9.3.13 — Die Matrizen aus Beispiel 9.3.10.
Die Matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

hat Hauptunterdeterminanten

$$\det(U_1) = 1,\ \det(U_2) = 0 \text{ und } \det(U_3) = 0.$$

Wir wissen aus Beispiel 9.3.10, dass diese Matrix positiv semidefinit ist.

Die Matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

ist laut Beispiel 9.3.10 indefinit. Sie hat jedoch genau die gleichen Hauptunterdeterminanten wie obige positiv semidefinite Matrix:

$$\det(U_1) = 1,\ \det(U_2) = 0 \text{ und } \det(U_3) = 0.$$

Es ist also $\det(U_i) \geq 0$ für $i = 1,2,3$, aber die Matrix ist nicht positiv semidefinit. ■

Mithilfe der Hauptunterdeterminanten kann man also indefinite Matrizen nicht immer von positiv oder negativ semidefiniten Matrizen unterscheiden. Ist jedoch eine Hauptunterdeterminante $\det(U_i)$ für gerades i negativ, so kann die Matrix laut Satz 9.3.6 weder positiv noch negativ (semi-)definit sein. Daher kann man in diesem Fall eindeutig schliessen, dass es sich um eine indefinite Matrix handelt.

■ Beispiel 9.3.14 — Fortsetzung von Beispiel 9.3.7.
Untersuchen wir die Hesse-Matrix der Funktion aus Beispiel 9.3.3,

$$\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix},$$

mithilfe der Hauptunterdeterminanten, ergibt sich

$$\det(U_1) = 2,\ \det(U_2) = 2 \cdot 2 - (-3) \cdot (-3) = -5.$$

Da $\det(U_2) < 0$ ist, kann die Matrix laut Satz 9.3.6 weder positiv semidefinit noch negativ semidefinit sein. Folglich ist die Matrix auch nicht positiv oder negativ definit. Die Matrix ist somit indefinit, wie wir auch in Beispiel 9.3.7 gezeigt haben. ▪

▪ Beispiel 9.3.15 — Die Matrix aus Beispiel 9.3.9.
Die Matrix

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

hat Hauptunterdeterminanten

$$\det(U_1) = 0,\ \det(U_2) = -1,\ \det(U_3) = 2.$$

Auch hier ist also eine Hauptunterdeterminante $\det(U_i)$ für einen geraden Index i negativ. Somit kann man direkt schliessen, dass die Matrix indefinit ist. ▪

Wendet man obige Kriterien auf die Hesse-Matrix an, kann man feststellen, ob die Funktion konvex oder konkav ist. Denn zusammenfassend gilt:

Satz 9.3.7 — Definitheit und Konvexität.
Sei $D \subseteq \mathbb{R}^n$ eine offene, konvexe Menge, $f : D \to Z$ eine zweimal stetig differenzierbare Funktion in n Variablen und $H_f(\mathbf{x}^0)$ die Hesse-Matrix von f an der Stelle $\mathbf{x}^0 \in D$. Dann gilt:

- $H_f(\mathbf{x}^0)$ ist positiv semidefinit für alle $\mathbf{x}^0 \in D \Leftrightarrow f$ ist konvex;
- $H_f(\mathbf{x}^0)$ ist positiv definit für alle $\mathbf{x}^0 \in D \Rightarrow f$ ist streng konvex;
- $H_f(\mathbf{x}^0)$ ist negativ semidefinit für alle $\mathbf{x}^0 \in D \Leftrightarrow f$ ist konkav;
- $H_f(\mathbf{x}^0)$ ist negativ definit für alle $\mathbf{x}^0 \in D \Rightarrow f$ ist streng konkav.

Beweis. Der Satz folgt direkt aus Definition 9.3.3 und Satz 9.3.4. ■

Z Ist $A \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix, dann:

$$A \text{ heisst } \left\{ \begin{matrix} \text{positiv semidefinit} \\ \text{positiv definit} \\ \text{negativ semidefinit} \\ \text{negativ definit} \end{matrix} \right\}, \text{ wenn für alle } \mathbf{r} \in \mathbb{R}^n, \mathbf{r} \neq \mathbf{0} \text{ gilt: } \mathbf{r}^T A \mathbf{r} = \left\{ \begin{matrix} \geq 0 \\ > 0 \\ \leq 0 \\ < 0. \end{matrix} \right.$$

Ist A weder positiv definit, positiv semidefinit, negativ definit noch negativ semidefinit, heisst A indefinit.

$$\text{Alle Eigenwerte von } A \text{ sind } \left\{ \begin{matrix} \geq 0 \\ > 0 \\ \leq 0 \\ < 0 \end{matrix} \right\} \Leftrightarrow A \text{ ist } \left\{ \begin{matrix} \text{positiv semidefinit} \\ \text{positiv definit} \\ \text{negativ semidefinit} \\ \text{negativ definit} \end{matrix} \right.$$

A hat mindestens einen Eigenwert > 0 und mindestens einen Eigenwert $< 0 \Leftrightarrow A$ ist indefinit.

Die i-te Hauptunterdeterminante einer symmetrischen Matrix ist die Determinante der Matrix, welche durch Streichen aller Zeilen und Spalten mit Inddizes grösser als i entsteht.

$$\left.\begin{array}{r} A \text{ positiv definit} \\ A \text{ negativ definit} \end{array}\right\} \Leftrightarrow \text{ für alle } i \text{ ist } \left\{\begin{array}{l} i\text{-te Hauptdeterminante} \\ i\text{-te Hauptdeterminante } \cdot (-1)^i \end{array}\right\} > 0$$

$$\left.\begin{array}{r} A \text{ positiv semidefinit} \\ A \text{ negativ semidefinit} \end{array}\right\} \Rightarrow \text{ für alle } i \text{ ist } \left\{\begin{array}{l} i\text{-te Hauptdeterminante} \\ i\text{-te Hauptdeterminante } \cdot (-1)^i \end{array}\right\} \geq 0$$

9.4 Extremwertbestimmung

In diesem Abschnitt befassen wir uns mit der Bestimmung von lokalen und globalen Extrema von zweimal stetig partiell differenzierbaren Funktionen in n Variablen.

9.4.1 Definition globaler und lokaler Extrema

Ziele dieses Unterkapitels

- Was versteht man unter dem Infimum und dem Supremum einer reellen Funktion in n Variablen?
- Was ist das globale Maximum und was ist das globale Minimum einer reellen Funktion in n Variablen?
- Was ist ein lokales Maximum und was ist ein lokales Minimum einer reellen Funktion in n Variablen?

Hierfür definieren wir zunächst, was globale Extrema sind. Die Terminologie ist durchgehend eine Erweiterung der Definitionen aus Kapitel 5 über reelle Funktionen in einer Variable.

Definition 9.4.1 — Supremum, Infimum und globale Extrema.
Sei $f : D \to Z$ eine reelle Funktion in n Variablen. Ist der Wertebereich $f(D) \subseteq Z$ nach oben beschränkt, so nennt man die kleinste obere Schranke

$$\sup f(D) = \sup_{\mathbf{x} \in D} f(\mathbf{x})$$

das Supremum von f. Gibt es ein $\mathbf{x}^{\max} \in D$ mit $f(\mathbf{x}^{\max}) = \sup_{\mathbf{x} \in D} f(\mathbf{x})$, so heisst

$$\max f(D) = \max_{\mathbf{x} \in D} f(\mathbf{x}) = f(\mathbf{x}^{\max})$$

globales Maximum von f und $\mathbf{x}^{\max}$ globale Maximalstelle. Die Menge aller globalen Maximalstellen wird durch $\arg\max_{\mathbf{x} \in D} f(\mathbf{x})$ symbolisiert. Ist $f(D)$ nicht nach oben beschränkt, schreibt man $\sup_{\mathbf{x} \in D} f(\mathbf{x}) = +\infty$.

Ist der Wertebereich $f(D)$ nach unten beschränkt, so nennt man die grösste untere Schranke

$$\inf f(D) = \inf_{\mathbf{x} \in D} f(\mathbf{x})$$

das Infimum von f. Gibt es ein $\mathbf{x}^{\min} \in D$ mit $f(\mathbf{x}^{\min}) = \inf_{\mathbf{x} \in D} f(x)$, so heisst

$$\min f(D) = \min_{\mathbf{x} \in D} f(\mathbf{x}) = f(\mathbf{x}^{\min})$$

globales Minimum von f und $\mathbf{x}^{\min}$ globale Minimalstelle. Die Menge aller globalen Minimalstellen wird durch $\arg\min_{\mathbf{x} \in D} f(\mathbf{x})$ symbolisiert. Ist $f(D)$ nicht nach unten beschränkt, schreibt man $\inf_{\mathbf{x} \in D} f(\mathbf{x}) = -\infty$.

Ein globales Maximum oder globales Minimum wird auch als globales Extremum oder Extremwert bezeichnet, analog ist eine (globale) Extremalstelle eine (globale) Minimal- oder (globale) Maximalstelle.

An einer globalen Extremalstelle hat die Funktion einen Wert, der grösser oder kleiner ist als (bzw. zumindest gleich ist wie) die Funktionswerte aller anderen Stellen im Definitionsbereich. Im Gegensatz zu globalen Extremalstellen ergeben sich an lokalen Extremalstellen Funktionswerte, die grösser oder kleiner sind als (bzw. zumindest gleich sind wie) an allen Stellen einer Umgebung innerhalb des Definitionsbereichs, vgl. Abbildung 9.50. Dabei nutzen wir erneut die in Kapitel 2.2 definierte ε-Umgebung einer

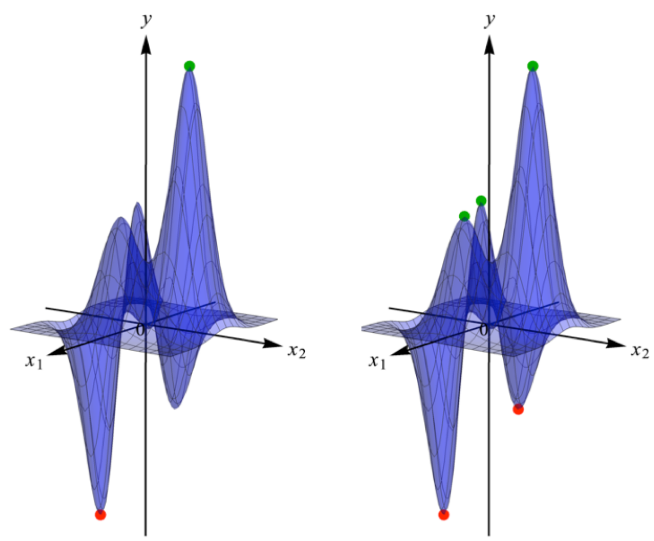

Abbildung 9.50: Globale (links) und lokale (rechts) Minima (rot) und Maxima (grün) einer Funktion.

Stelle $\mathbf{x}^0$, d.h. die Menge aller Stellen, welche weniger als ε von $\mathbf{x}^0$ entfernt sind. Auch bei reellen Funktionen in n Variablen sind lokale Extrema einfacher zu bestimmen als globale.

Definition 9.4.2 — Lokale Extrema und Extremalstellen.
Es sei $f : D \to Z$ eine reelle Funktion in n Variablen, $\mathbf{x}^0 \in D$ und $U(\mathbf{x}^0, \varepsilon)$ eine ε-Umgebung der Stelle $\mathbf{x}^0$.
Die Funktion f hat in $\mathbf{x}^0$ ein lokales Maximum $f(\mathbf{x}^0)$, falls ein $\varepsilon > 0$ existiert, so dass

$$f(\mathbf{x}^0) \geq f(\mathbf{x}) \text{ für alle } \mathbf{x} \in U(\mathbf{x}^0, \varepsilon) \cap D.$$

In diesem Fall nennt man $\mathbf{x}^0$ lokale Maximalstelle von f.
Die Funktion f hat in $\mathbf{x}^0$ ein lokales Minimum $f(\mathbf{x}^0)$, falls ein $\varepsilon > 0$ existiert, so dass

$$f(\mathbf{x}^0) \leq f(\mathbf{x}) \text{ für alle } \mathbf{x} \in U(\mathbf{x}^0, \varepsilon) \cap D.$$

In diesem Fall nennt man $\mathbf{x}^0$ lokale Minimalstelle von f. Lokales Extremum ist der Sammelbegriff für ein lokales Maximum oder Minimum. Eine lokale Extremalstelle ist eine lokale Minimal- oder Maximalstelle.

Ein lokales Extremum ist also der grösste oder kleinste Funktionswert in einer Umgebung einer Stelle $\mathbf{x}^0$, ein globales Extremum ist der grösste oder kleinste Funktionswert im Definitionsbereich. Wie im Fall einer reellen Funktion in einer Variable ist daher jedes globale Extremum ein lokales Extremum, aber nicht jedes lokale ein globales.

Z Das Supremum der reellen Funktion $f : D \to Z$ in n Variablen ist $\sup_{\mathbf{x}\in D} f(\mathbf{x}) = \sup f(D)$, das Infimum $\inf_{\mathbf{x}\in D} f(\mathbf{x}) = \inf f(D)$.

Gibt es eine Stelle $\mathbf{x}^{\max} \in D$ mit $f(\mathbf{x}^{\max}) = \sup_{\mathbf{x}\in D} f(\mathbf{x})$, so heisst $f(\mathbf{x}^{\max})$ auch globales Maximum von f und $\mathbf{x}^{\max}$ globale Maximalstelle. Gibt es eine Stelle $\mathbf{x}^{\min} \in D$ mit $f(\mathbf{x}^{\min}) = \inf_{\mathbf{x}\in D} f(\mathbf{x})$, so heisst $f(\mathbf{x}^{\min})$ auch globales Minimum von f und $\mathbf{x}^{\min}$ globale Minimalstelle.

Gibt es innerhalb einer ε-Umgebung um eine Stelle $\mathbf{x}^0$ keine Stelle mit einem grösseren (kleineren) Funktionswert, heisst $\mathbf{x}^0$ auch lokale Maximalstelle (Minimalstelle) und $f(\mathbf{x}^0)$ lokales Maximum (Minimum).

9.4.2 Identifikation möglicher lokaler Extrema

Ziele dieses Unterkapitels

- Was besagt das Kriterium von Fermat für stetig partiell differenzierbare Funktionen in n Variablen? Was versteht man unter stationären Stellen einer reellen Funktion in n Variablen?
- Was ist ein Sattelpunkt?

Aus der Definition lokaler Extrema folgt unmittelbar, dass eine reelle Funktion in n Variablen an einer Stelle $\mathbf{x}^0$ nur dann ein lokales Extremum haben kann, wenn der Vertikalschnitt durch $\mathbf{x}^0$ in jeder Richtung für $t = 0$ ein lokales Extremum hat, vgl. Abbildung 9.51. In

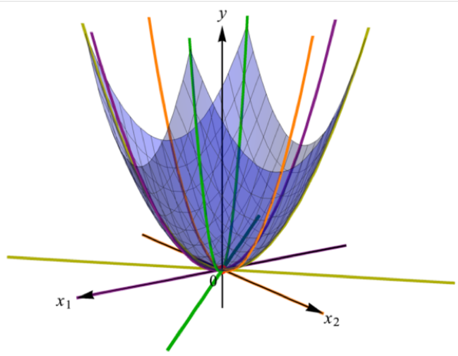

Abbildung 9.51: Lokales Minimum einer Funktion.

anderen Worten: Eine reelle Funktion in n Variablen kann an einer Stelle $\mathbf{x}^0$ nur dann ein lokales Extremum haben, wenn $f_{\mathbf{x}^0,\mathbf{r}}(t)$ für alle Richtungen $\mathbf{r}$ an der Stelle $t = 0$ ein Extremum hat. Aus dem Kriterium von Fermat, Satz 5.7.4, folgt somit für stetig partiell differenzierbare Funktionen, dass die Richtungsableitung $f'_{\mathbf{r}}(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)^T \mathbf{r}$ an dieser Stelle für alle Richtungen $\mathbf{r}$ gleich 0 sein muss. Dies gilt nur, wenn der Gradient dem Nullvektor entspricht, also wenn $\nabla f(\mathbf{x}^0) = \mathbf{0}$ gilt. Es folgt:

Satz 9.4.1 — Notwendiges Kriterium erster Ordnung (Kriterium von Fermat). Sei f eine stetig partiell differenzierbare Funktion in n Variablen mit offenem Definitionsbereich $D \subseteq \mathbb{R}^n$. Ist $\mathbf{x}^0 \in D$ eine lokale Extremalstelle, dann gilt $\nabla f(\mathbf{x}^0) = \mathbf{0}$.

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.8). ■

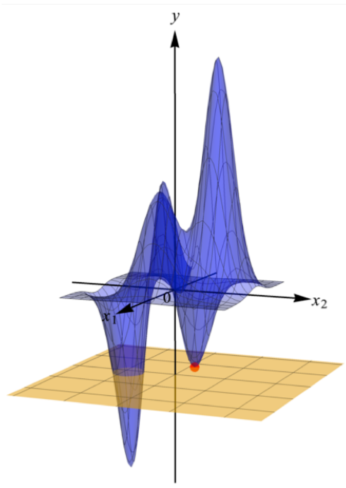

Abbildung 9.52: Visualisierung stationäre Stelle.

Wir nennen diese Stellen mit einem Gradienten, der gleich dem Nullvektor ist, auch stationäre Stellen, siehe Abbildung 9.52.

Definition 9.4.3 — Stationäre Stelle. Sei f eine reelle Funktion in n Variablen. Gilt $\nabla f(\mathbf{x}^0) = \mathbf{0}$, so nennt man $\mathbf{x}^0$ eine stationäre Stelle.

■ **Beispiel 9.4.1 — Bestimmung der stationären Stellen.** Möchte man die stationären Stellen der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^3 + x_2^3 - 3x_1x_2$$

bestimmen, so berechnet man zuerst den Gradienten. Die partiellen Ableitungen sind gegeben durch

$$f_{x_1}(\mathbf{x}) = 3x_1^2 - 3x_2,\ f_{x_2}(\mathbf{x}) = 3x_2^2 - 3x_1.$$

Zur Bestimmung der stationären Stellen müssen wir also das Gleichungssystem

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 3(x_1^0)^2 - 3(x_2^0) \\ 3(x_2^0)^2 - 3(x_1^0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

lösen. Dieses Gleichungssystem ist durch die auftretenden quadratischen Terme nicht linear. Man erkennt aber aus der zweiten Gleichung, dass

$$\begin{aligned} 3(x_2^0)^2 - 3(x_1^0) &= 0 & \Leftrightarrow \\ 3(x_2^0)^2 &= 3(x_1^0) & \Leftrightarrow \\ (x_2^0)^2 &= (x_1^0). \end{aligned}$$

Durch Einsetzen von $(x_1^0) = (x_2^0)^2$ in die erste Gleichung erhält man

$$\begin{aligned} 3(x_2^0)^4 - 3(x_2^0) &= 0 & \Leftrightarrow \\ 3(x_2^0)^4 &= 3(x_2^0) & \Leftrightarrow \\ (x_2^0)^4 &= (x_2^0). \end{aligned}$$

Daraus ergeben sich die Lösungen $x_2^0 = 0$ und $x_2^0 = 1$. Für $x_2^0 = 0$ folgt $x_1^0 = (x_2^0)^2 = 0^2 = 0$, für $x_2^0 = 1$ folgt $x_1^0 = (x_2^0)^2 = 1^2 = 1$. Die stationären Stellen von f sind also $(0,0)^T$ und $(1,1)^T$. ■

■ **Beispiel 9.4.2 — Die stationären Stellen der Funktion aus Beispiel 9.2.15.**
Um die stationären Stellen der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$ zu bestimmen, berechnet man die partiellen Ableitungen,

$$f_{x_1}(\mathbf{x}) = 2x_1,\ f_{x_2}(\mathbf{x}) = 8x_2 - 8.$$

Die stationären Stellen sind also durch das Gleichungssystem

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2(x_1^0) \\ 8(x_2^0) - 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

bzw.

$$\begin{aligned} 2(x_1^0) &= 0 \\ 8(x_2^0) - 8 &= 0 \end{aligned}$$

mit Lösung $x_1^0 = 0$ und $x_2^0 = 1$ gegeben. Die einzige stationäre Stelle von f ist also $(0,1)^T$.
■

Wie im Fall reeller Funktionen in einer Variablen beschreibt das Kriterium von Fermat nur eine notwendige, keine hinreichende Bedingung für lokale Extremalstellen. An stationären Stellen können, müssen aber keine lokalen Extrema vorliegen.

Zweimal stetig partiell differenzierbare Funktionen können an stationären Stellen lokale Extrema oder Sattelpunkte haben. Wir definieren hierbei Sattelpunkte als Punkte, an denen eine stationäre Stelle aber keine lokale Extremalstelle vorliegt. An einer Stelle mit einem Sattelpunkt gibt es also in jeder Umgebung Stellen mit grösseren Funktionswerten und Stellen mit kleineren Funktionswerten als der Funktionswert an dieser Stelle.

Definition 9.4.4 — Sattelpunkt.
Sei $\mathbf{x}^0$ eine stationäre Stelle von f. Liegt an der Stelle $\mathbf{x}^0$ kein lokales Extremum vor, spricht man von einem Sattelpunkt an der Stelle $\mathbf{x}^0$.

Z

Ist $\mathbf{x}^0$ eine lokale Extremalstelle einer stetig partiell differenzierbaren Funktion in n Variablen mit offenem Definitionsbereich D und ist $\mathbf{x}^0 \in D$ eine lokale Extremalstelle, dann ist laut dem Kriterium von Fermat $\mathbf{x}^0$ eine stationäre Stelle, d.h. $\nabla f(\mathbf{x}^0) = \mathbf{0}$.

Liegt an einer stationären Stelle $\mathbf{x}^0$ keine lokale Extremalstelle vor, hat f einen Sattelpunkt an der Stelle $\mathbf{x}^0$.

9.4.3 Bestimmung lokaler Extrema

Ziele dieses Unterkapitels

- Wie kann man anhand der Hesse-Matrix einer zweimal stetig partiell differenzierbaren Funktion schliessen, ob eine stationäre Stelle eine lokale Maximal- oder Minimalstelle ist?
- Welcher Zusammenhang besteht zwischen lokalen Extrema einer Funktion f und der Funktion $-f$?

Wie bei rellen Funktionen in einer Variable, vgl. Satz **??**, sind Informationen der ersten und zweiten Ableitungen in vielen Fällen ausreichend, um lokale Maxima oder Minima zu finden. Mithilfe des Gradienten und der Hesse-Matrix kann man folgende Aussagen treffen:

Satz 9.4.2 — Hinreichende Kriterien lokaler Extrema. Sei $D \subseteq \mathbb{R}^n$ offen, $f : D \to Z$ eine zweimal stetig differenzierbare Funktion in n Variablen und $\mathbf{x}^0 \in D$ eine stationäre Stelle mit $\nabla f(\mathbf{x}^0) = \mathbf{0}$ und Hesse-Matrix $H_f(\mathbf{x}^0)$. Dann gilt:

- $H_f(\mathbf{x}^0)$ positiv definit $\Rightarrow$ f hat in $\mathbf{x}^0$ ein lokales Minimum;
- $H_f(\mathbf{x}^0)$ negativ definit $\Rightarrow$ f hat in $\mathbf{x}^0$ ein lokales Maximum;
- $H_f(\mathbf{x}^0)$ ist indefinit $\Rightarrow$ f hat in $\mathbf{x}^0$ einen Sattelpunkt.

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.9). ■

Ist eine Hesse-Matrix positiv semidefinit aber nicht positiv definit oder negativ semidefinit aber nicht negativ definit, ist anhand der ersten und zweiten Ableitungen alleine keine Aussage über $\mathbf{x}^0$ möglich. Wir demonstrieren die aus dem Satz folgenden Schlüsse an Beispielen:

■ **Beispiel 9.4.3 — Lokale Extrema einer quadratischen Funktion.** Wir betrachten die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^2 + 3$. Der Gradient von f ist an der Stelle $\mathbf{x}^0$ gegeben durch

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2(x_1^0) \\ 8(x_2^0) - 8 \end{pmatrix}.$$

In Beispiel 9.4.2 zeigten wir, dass die einzige stationäre Stelle $(0,1)^T$ ist. Die Hesse-Matrix von f an der Stelle $(0,1)^T$ ist gegeben durch

$$H_f(0,1) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

In Beispiel 9.3.6 wurden die Eigenwerte als 2 und 8 bestimmt und so geschlossen, dass diese Matrix positiv definit ist. Die Funktion f hat an der Stelle $(0,1)^T$ also ein lokales Minimum. ■

■ **Beispiel 9.4.4 — Lokale Extrema einer anderen quadratischen Funktion.** Betrachten wir erneut die quadratische Funktion f aus Beispiel 9.4.3. Der Gradient der Funktion $g = -f$, also $g : \mathbb{R}^2 \to \mathbb{R}$ mit $g(\mathbf{x}) = -x_1^2 - 4(x_2 - 1)^2 - 3$ ist an der Stelle $\mathbf{x}^0$ gegeben durch

$$\nabla g(\mathbf{x}^0) = \begin{pmatrix} -2(x_1^0) \\ -8(x_2^0) + 8 \end{pmatrix}.$$

Die einzige stationäre Stelle von g ist $(0,1)^T$. An dieser Stelle ist die Hesse-Matrix von g gegeben durch

$$H_g(0,1) = \begin{pmatrix} -2 & 0 \\ 0 & -8 \end{pmatrix}.$$

Diese Matrix ist negativ definit. Die Funktion g hat an der Stelle $(0,1)^T$ also ein lokales Maximum, siehe Abbildung 9.53.

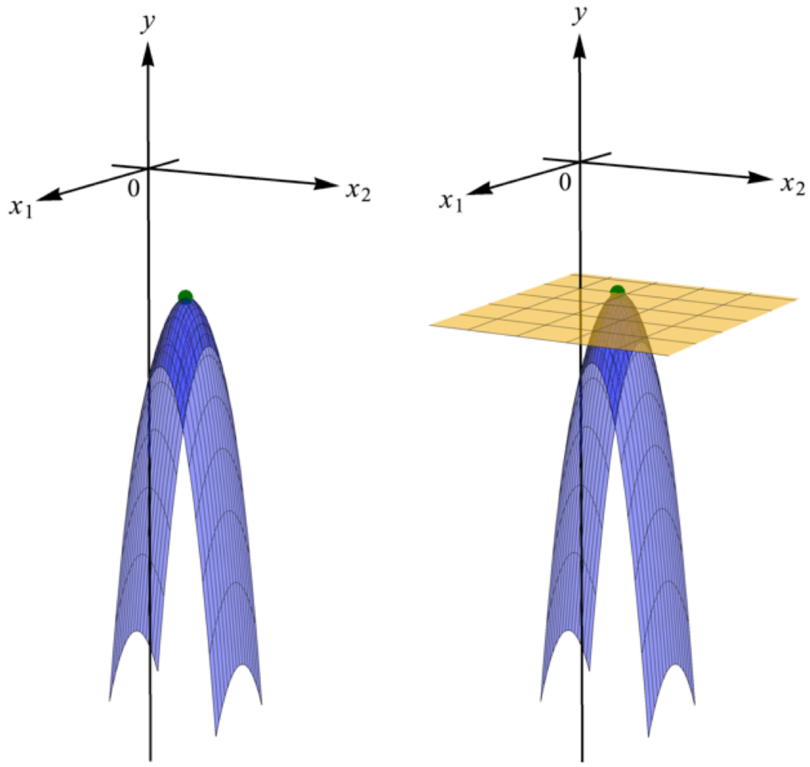

Abbildung 9.53: Lokales Maximum der Funktion $g(\mathbf{x}) = -x_1^2 - 4(x_2-1)^2 - 3$.

■

Was wir im Beispiel gesehen haben, gilt auch allgemein für reelle Funktionen in n Variablen: Hat die Funktion f ein lokales Maximum an der Stelle $\mathbf{x}^0$, so hat die Funktion $-f$ an der Stelle $\mathbf{x}^0$ ein lokales Minimum, und umgekehrt.

Satz 9.4.3 — Min-Max Dualität.
Die reelle Funktion $f : D \to Z$ in n Variablen mit $D \subseteq \mathbb{R}^n$ hat genau dann ein lokales Maximum an der Stelle $\mathbf{x}^0 \in D$, wenn die reelle Funktion $-f$ an der Stelle $\mathbf{x}^0$ ein lokales Minimum hat. Die reelle Funktion $f : D \to Z$ in n Variablen hat genau dann ein lokales Minimum an der Stelle $\mathbf{x}^0 \in D$, wenn $-f$ an der Stelle $\mathbf{x}^0$ ein lokales Maximum hat.

Hat f ein globales Maximum, gilt $\min_{\mathbf{x}\in D}(-f(\mathbf{x})) = -\max_{\mathbf{x}\in D} f(\mathbf{x})$. Hat f ein globales Minimum, gilt $\max_{\mathbf{x}\in D}(-f(\mathbf{x})) = -\min_{\mathbf{x}\in D} f(\mathbf{x})$.

Beweis. Der Beweis folgt direkt aus Definition 9.4.1. ■

Wir illustrieren Satz 9.4.2 über hinreichende Bedingungen in einem dritten Beispiel:

■ **Beispiel 9.4.5 — Eine Funktion mit einem Sattelpunkt.**
Betrachten wir die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(\mathbf{x}) = x_1 x_2$. Der Gradient von f an der Stelle $\mathbf{x}^0$ ist

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} x_2^0 \\ x_1^0 \end{pmatrix}.$$

Die einzige stationäre Stelle ist $\mathbf{0}$. Die Hesse-Matrix von f ist dort

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

Diese Matrix ist indefinit, da die zweite Hauptunterdeterminante $\det(U_2) = 0 \cdot 0 - 1 \cdot 1 = -1$ negativ ist. An der Stelle $\mathbf{0}$ liegt demnach ein Sattelpunkt vor, siehe Abbildung 9.54. ■

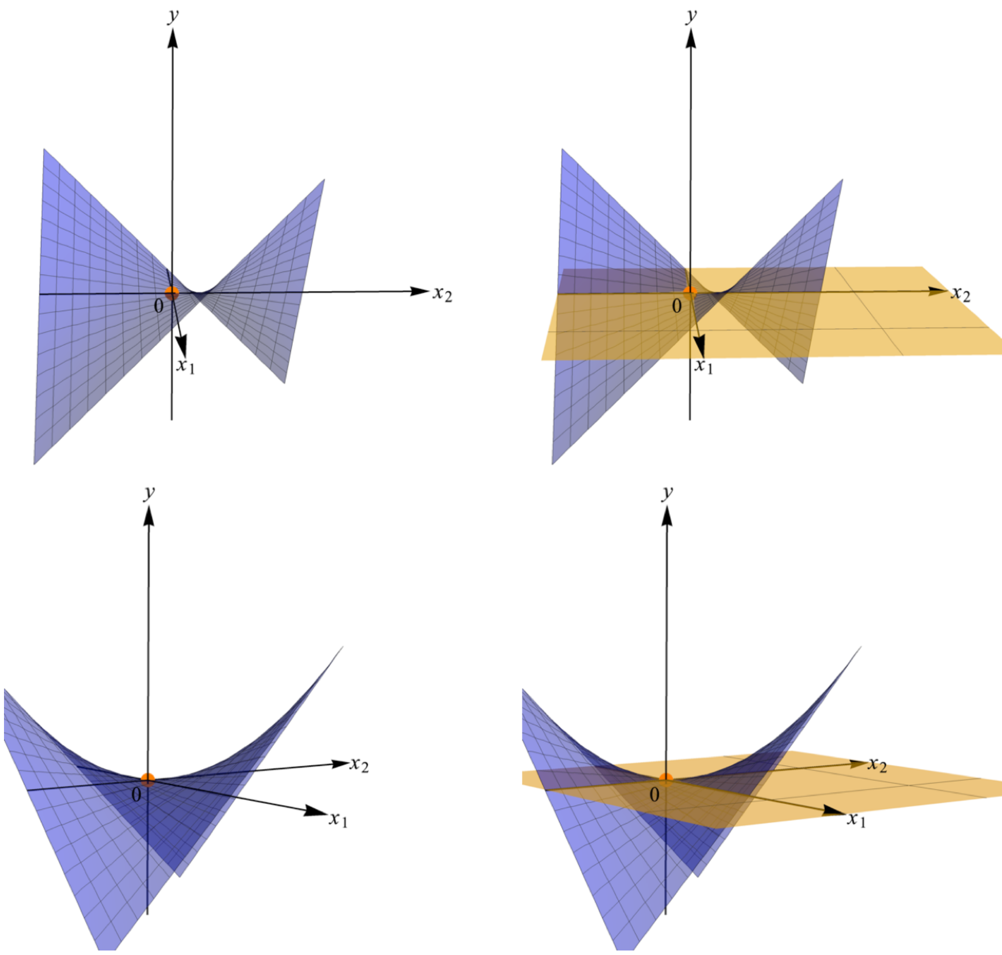

Abbildung 9.54: Sattelpunkt der Funktion $f(\mathbf{x}) = x_1 x_2$.

Wie in obigem Beispiel ist die Berechnung der Hauptunterdeterminanten zur Klassifikation der Hesse-Matrix bei reellen Funktionen in zwei Variablen besonders einfach. Denn hier muss nur die Determinante $\det(H_f(\mathbf{x}^0))$ berechnet werden. In diesem Fall reduziert sich Satz 9.4.2 auf folgende Aussage:

Satz 9.4.4 — Hinreichende Kriterien lokaler Extrema im Fall $n = 2$.
Sei $D \subseteq \mathbb{R}^2$ eine offene Menge, $f : D \to Z$ eine zweimal stetig differenzierbare Funktion in 2 Variablen und $\mathbf{x}^0 \in D$ eine stationäre Stelle mit $\nabla f(\mathbf{x}^0) = \mathbf{0}$ und Hesse-Matrix $H_f(\mathbf{x}^0)$. Dann gilt:

- Falls $\det(H_f(\mathbf{x}^0)) > 0$, hat f an der Stelle $\mathbf{x}^0$ ein lokales Extremum. Gilt ausserdem
  - $f_{x_1x_1}(\mathbf{x}^0) > 0$, so ist es ein lokales Minimum,
  - $f_{x_1x_1}(\mathbf{x}^0) < 0$, so ist es ein lokales Maximum.
- Gilt $\det(H_f(\mathbf{x}^0)) < 0$, dann hat f an der Stelle $\mathbf{x}^0$ kein lokales Extremum sondern einen Sattelpunkt.
- Gilt $\det(H_f(\mathbf{x}^0)) = 0$, dann kann anhand der ersten und zweiten partiellen Ableitungen im Punkt $\mathbf{x}^0$ nicht entschieden werden, ob an der Stelle $\mathbf{x}^0$ ein Extremum oder ein Sattelpunkt vorliegt.

Beweis. Da f zweimal stetig differenzierbar ist, ist die Hesse-Matrix symmetrisch und es gilt

$$\det(H_f(\mathbf{x}^0)) = f_{x_1x_1}(\mathbf{x}^0)f_{x_2x_2}(\mathbf{x}^0) - (f_{x_1x_2}(\mathbf{x}^0))^2.$$

Ist $f_{x_1x_1}(\mathbf{x}^0) = 0$, so gilt also $\det(H_f(\mathbf{x}^0)) = -(f_{x_1x_2}(\mathbf{x}^0))^2 \leq 0$. Ist $\det(H_f(\mathbf{x}^0)) > 0$, muss also $f_{x_1x_1}(\mathbf{x}^0) \neq 0$ gelten. Die anderen Aussagen ergeben sich hiermit direkt aus Satz 9.3.6 und 9.4.2. ■

Unabhängig davon, ob man sich auf Satz 9.4.2 oder Satz 9.4.4 beruft, kann jedoch im Falle einer positiv oder negativ semidefiniten Matrix keine Aussage bezüglich des Vorhandenseins eines lokalen Extremums oder Sattelpunktes getroffen werden. Wir zeigen im Folgenden zwei sehr unterschiedliche Beispiele, die aber beide an einer stationären Stelle $\mathbf{x}^0$ eine positiv semidefinite Hesse-Matrix haben.

■ **Beispiel 9.4.6 — Lokales Minimum bei positiv semidefiniter Hesse-Matrix.**
Wir betrachten die reelle Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^4 + 3$ in zwei Variablen. Der Gradient von f an einer Stelle $\mathbf{x}^0$ ist gegeben durch

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2(x_1^0) \\ 16((x_2^0) - 1)^3 \end{pmatrix}.$$

Der Gradient ist genau dann $\mathbf{0}$, wenn $x_1^0 = 0$ und $x_2^0 = 1$ gilt. Die einzige stationäre Stelle ist also $(0, 1)^T$. Die Hesse-Matrix von f ist

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 48((x_2^0) - 1)^2 \end{pmatrix}.$$

Ausgewertet an der Stelle $(0, 1)^T$ ergibt sich

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

Da $\det(H_f(0, 1)) = 2 \cdot 0 - 0 \cdot 0 = 0$ gilt, können wir mithilfe von Satz 9.4.4 keine Aussage treffen, ob sich an der Stelle $(0, 1)^T$ ein lokales Maximum, ein lokales Minimum oder ein Sattelpunkt befindet. Die Eigenwerte sind 0 und 2. Somit ist die Matrix positiv semidefinit. Auch Satz 9.4.2 hilft in diesem Fall nicht weiter.

Betrachtet man die Abbildungsvorschrift $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^4 + 3$ genauer, erkennt man leicht, dass die beiden ersten Summanden stets positiv oder 0 sind. Es gilt also $f(\mathbf{x}) \geq 3$ für alle $\mathbf{x} \in \mathbb{R}^2$. Da $f(0,1) = 3$, gilt also $f(\mathbf{x}) \geq 3 = f(0,1)$ für alle $\mathbf{x} \in \mathbb{R}^2$. An der Stelle $(0,1)^T$ liegt somit ein globales, und damit auch ein lokales Minimum vor, vgl. auch Abbildung 9.55. ■

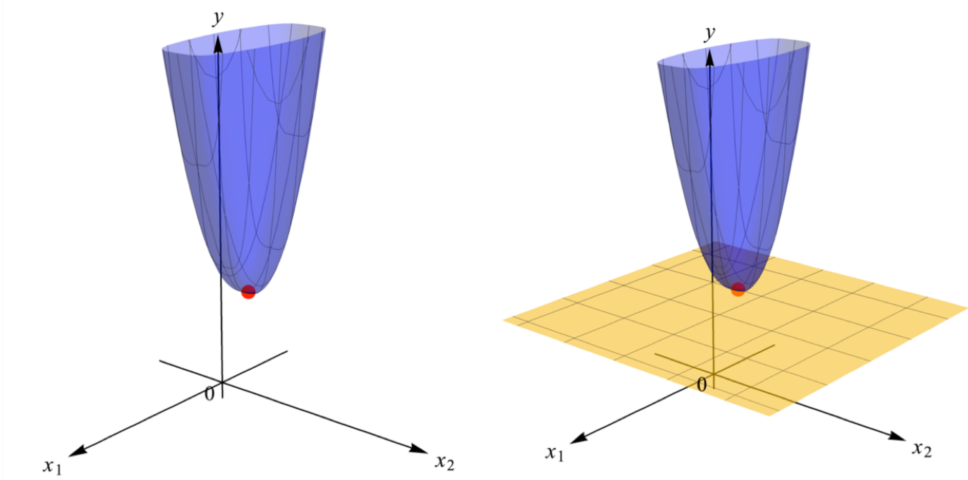

Abbildung 9.55: Minimum der Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^4 + 3$.

■ Beispiel 9.4.7 — Sattelpunkt bei positiv-semidefiniter Hesse-Matrix.
Betrachten wir die reelle Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^3 + 3$ in zwei Variablen. Der Gradient von f an einer Stelle $\mathbf{x}^0$ ist

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 2(x_1^0) \\ 12((x_2^0)-1)^2 \end{pmatrix}.$$

Der einzige stationäre Punkt ist erneut $(0,1)^T$. Die Hesse-Matrix von f ist

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 24((x_2^0)-1) \end{pmatrix}.$$

An der Stelle $(0,1)^T$ ergibt sich

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

Wie im vorherigen Beispiel ist die Hesse-Matrix mit Eigenwerten 0 und 2 positiv semidefinit. Da $\det(H_f(0,1)) = 2 \cdot 0 - 0 \cdot 0 = 0$ gilt, können wir auch hier mithilfe von Satz 9.4.4 keine Aussage treffen, ob es sich um ein lokales Maximum, ein lokales Minimum oder einen Sattelpunkt handelt. In Abbildung 9.56 erahnt man, dass es sich um einen Sattelpunkt handelt. Formal kann man wie folgt argumentieren, dass es sich um einen Sattelpunkt handeln muss: In einer ε-Umgebung der Stelle $(0,1)^T$ findet man für beliebig kleine $\varepsilon > 0$ stets eine Stelle $(0, 1+\frac{\varepsilon}{2})^T \in U((0,1)^T, \varepsilon)$ und eine andere Stelle $(0, 1-\frac{\varepsilon}{2})^T \in U((0,1)^T, \varepsilon)$. Wertet man die Funktion an der Stelle $(0, 1+\frac{\varepsilon}{2})^T$ aus, ergibt sich ein grösserer Funktionswert als an der stationären Stelle,

$$f\left(0, 1+\frac{\varepsilon}{2}\right) = 3 + \frac{\varepsilon^3}{2} > 3 = f(0,1).$$

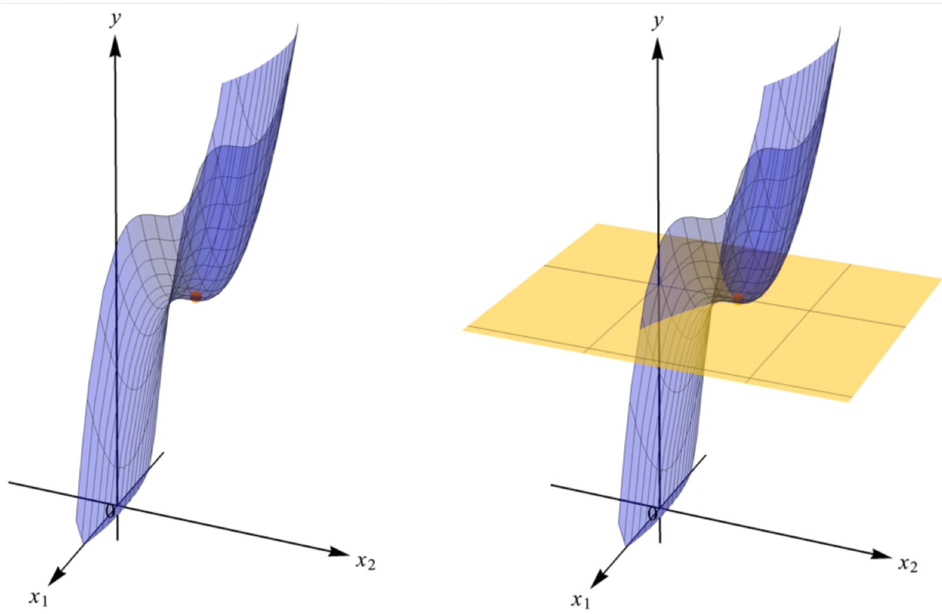

Abbildung 9.56: Sattelpunkt der Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2-1)^3 + 3$.

An der Stelle $(0,1)^T$ kann also kein lokales Maximum vorliegen. Wertet man die Funktion an der Stelle $(0, 1-\frac{\varepsilon}{2})^T$ aus, ergibt sich ein kleinerer Funktionswert als an der stationären Stelle,

$$f\left(0, 1-\frac{\varepsilon}{2}\right) = 3 - \frac{\varepsilon^3}{2} < 3 = f(0,1).$$

An der stationären Stelle $(0,1)^T$ kann also kein lokales Minimum vorliegen. In jeder ε-Umgebung dieser stationären Stelle findet man sowohl eine Stelle mit einem grösseren Funktionswert als auch eine Stelle mit einem kleineren Funktionswert. An der Stelle $(0,1)^T$ liegt somit ein Sattelpunkt vor. ■

Die Funktion im ersten Beispiel hat an einer globalen Minimalstelle eine positiv semidefinite Hesse-Matrix, die Funktion im zweiten Beispiel hat an einer Stelle mit einem Sattelpunkt ebenfalls eine positiv semidefinite Hesse-Matrix (Die Hesse-Matrizen sind sogar gleich). Das Entscheidungskriterium der Hesse-Matrix hilft in dieser Situation nicht weiter. Um zu zeigen, dass die Funktion im zweiten Beispiel an der stationären Stelle einen Sattelpunkt hat, haben wir gezeigt, dass in jeder Umgebung der stationären Stelle sowohl Stellen mit grösseren als auch Stellen mit kleineren Funktionswerten existieren. Diese Stellen befanden sich auf dem Vertikalschnitt von f ausgehend von der stationären Stelle in Richtung $\mathbf{e}^2$.

Wir demonstrieren die Bestimmung lokaler Extrema in einem weiteren Beispiel:

■ Beispiel 9.4.8 — Bestimmung aller lokalen Extrema.
Betrachten wir erneut die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\mathbf{x}) = x_1^3 + x_2^3 - 3x_1x_2$$

mit stationären Stellen $\mathbf{0}$ und $(1,1)^T$, vgl. Beispiel 9.4.1.

Die Hesse-Matrix der Funktion an einer beliebigen Stelle $\mathbf{x}^0$ ist

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 6(x_1^0) & -3 \\ -3 & 6(x_2^0) \end{pmatrix}.$$

An der Stelle $\mathbf{0}$ ist die Hesse-Matrix also

$$H_f(\mathbf{0}) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}.$$

Die zweite Hauptunterdeterminante dieser Matrix ist $\det(H_f(\mathbf{0})) = 0 \cdot 0 - (-3) \cdot (-3) = -9 < 0$. Damit ist die Matrix indefinit. An der Stelle $\mathbf{0}$ befindet sich ein Sattelpunkt. An der Stelle $(1,1)^T$ ist die Hesse-Matrix

$$H_f(1,1) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

Die zweite Hauptunterdeterminante dieser Matrix ist $\det(H_f(1,1)) = 6 \cdot 6 - (-3) \cdot (-3) = 27 > 0$. Zudem ist die erste Hauptunterdeterminante gleich $6 > 0$. Damit ist die Matrix positiv definit. An der Stelle $(1,1)^T$ befindet sich somit ein lokales Minimum, siehe Abbildung 9.57. ■

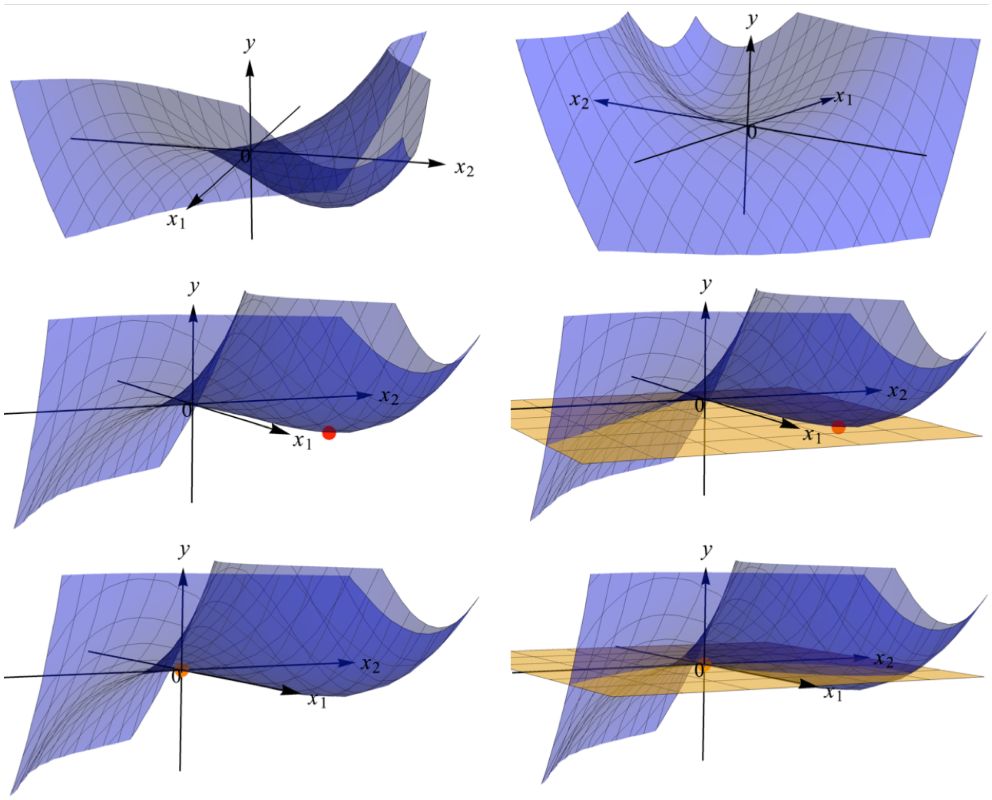

Abbildung 9.57: Sattelpunkt und lokales Minimum der Funktion $f(\mathbf{x}) = x_1^3 + x_2^3 - 3x_1x_2$.

Z Ist f eine zweimal stetig partiell differenzierbare Funktion mit stationärer Stelle $\mathbf{x}^0$, so kann man schliessen:

$$H_f(\mathbf{x}^0) \left\{ \begin{array}{l} \text{positiv definit} \\ \text{negativ definit} \\ \text{indefinit} \end{array} \right\} \Rightarrow \mathbf{x}^0 \text{ ist eine } \left\{ \begin{array}{l} \text{lokale Minimalstelle} \\ \text{lokale Maximalstelle} \\ \text{Stelle mit Sattelpunkt.} \end{array} \right.$$

Die Funktion f hat an der Stelle $\mathbf{x}^0$ genau dann ein lokales Minimum (Maximum), wenn die Funktion $-f$ an dieser Stelle ein lokales Maximum (Minimum) hat.

9.4.4 Lokale Extrema von Vertikalschnitten

Ziele dieses Unterkapitels

- Was kann man über die Definitheit der Hesse-Matrix einer zweimal stetig partiell differenzierbaren Funktion an einer lokalen Maximalstelle bzw. Minimalstelle aussagen?
- Kann man aus der Tatsache, dass jeder Vertikalschnitt von f durch $\mathbf{x}^0$ an der Stelle $t = 0$ ein lokales Extremum hat, schliessen, dass f an der Stelle $\mathbf{x}^0$ ein lokales Extremum hat?

Aus der Definition lokaler Extrema folgt, dass an einer lokalen Maximalstelle (Minimalstelle) $\mathbf{x}^0$ der Vertikalschnitt durch $\mathbf{x}^0$ in jede Richtung betrachtet auch ein lokales Maximum (Minimum) haben muss. Zusammen mit Definition 9.2.10 und Satz 9.2.7 zur zweiten Richtungsableitung gilt zudem:

Satz 9.4.5 — Notwendige Kriterien lokaler Extrema zweiter Ordnung.
Sei $D \subseteq \mathbb{R}^n$ offen, $f : D \to Z$ eine zweimal stetig differenzierbare Funktion in n Variablen und $\mathbf{x}^0 \in D$ eine stationäre Stelle, d.h. mit $\nabla f(\mathbf{x}^0) = \mathbf{0}$ und Hesse-Matrix $H_f(\mathbf{x}^0)$. Dann gilt:

- f hat in $\mathbf{x}^0$ eine lokales Minimum $\Rightarrow H_f(\mathbf{x}^0)$ ist positiv semidefinit;
- f hat in $\mathbf{x}^0$ eine lokales Maximum $\Rightarrow H_f(\mathbf{x}^0)$ ist negativ semidefinit.

Beweis. Einen Beweis findet man Geiger, Kanzow (1999, Satz 2.2). ∎

Umgekehrt kann man jedoch nicht alleine durch Betrachtung der verschiedenen Richtungsableitungen auf das Vorliegen lokaler Minima oder Maxima schliessen, wie folgendes Beispiel zeigt:

■ Beispiel 9.4.9 — Eine weitere positiv-semidefinite Hesse-Matrix (∗).
Wir betrachten die Funktion mit Abbildungsvorschrift $f(\mathbf{x}) = x_1^4 - 3x_1^2x_2 + x_2^2$. Der Gradient und die Hesse-Matrix von f an der Stelle $\mathbf{x}^0 = (x_1^0, x_2^0)^T$ sind

$$\nabla f(\mathbf{x}^0) = \begin{pmatrix} 4(x_1^0)^3 - 6(x_1^0)(x_2^0) \\ 2(x_2^0) - 3(x_1^0)^2 \end{pmatrix} \quad \text{und} \quad H_f(\mathbf{x}^0) = \begin{pmatrix} 12(x_1^0)^2 - 6(x_2^0) & -6(x_1^0) \\ -6(x_1^0) & 2 \end{pmatrix}.$$

Stationäre Stellen sind damit Lösungen des Gleichungssystems

$$\begin{aligned} 4(x_1^0)^3 - 6(x_1^0)(x_2^0) &= 0 \\ 2(x_2^0) - 3(x_1^0)^2 &= 0. \end{aligned}$$

Aus der zweiten Gleichung folgt $x_2^0 = \frac{3}{2}(x_1^0)^2$. Setzt man dies in die erste Gleichung ein, erhält man die Gleichung

$$4(x_1^0)^3 - 6(x_1^0)\left(\frac{3}{2}(x_1^0)^2\right) = 4(x_1^0)^3 - 9(x_1^0)^3 = -5(x_1^0)^3 = 0,$$

also $x_1^0 = 0$. Erneutes Einsetzen in $x_2^0 = \frac{3}{2}(x_1^0)^2$ liefert dann ebenfalls $x_2^0 = 0$. Die einzige stationäre Stelle ist also $\mathbf{0}$.

Die Hesse-Matrix von f an dieser stationären Stelle ist

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

Da $\det(H_f(\mathbf{0})) = 0 \cdot 2 - 0 \cdot 0 = 0$ gilt, können wir mithilfe von Satz 9.4.4 keine Aussage treffen, ob es sich um ein lokales Maximum, ein lokales Minimum oder einen Sattelpunkt handelt. Die Abbildungsvorschrift eines Vertikalschnitts der Funktion ausgehend von $\mathbf{0}$ in eine beliebige Richtung $\mathbf{r}$ ist

$$f_{\mathbf{0},\mathbf{r}}(t) = r_1^4 t^4 - 3r_1^2 r_2 t^3 + r_2^2 t^2$$

mit

$$f'_{\mathbf{0},\mathbf{r}}(t) = 4r_1^4 t^3 - 9r_1^2 r_2 t^2 + 2r_2^2 t \text{ und } f''_{\mathbf{0},\mathbf{r}}(t) = 12r_1^4 t^2 - 18r_1^2 r_2 t + 2r_2^2.$$

Da $f'_{\mathbf{0},\mathbf{r}}(0) = 0$ gilt, hat jeder Vertikalschnitt durch $\mathbf{0}$ an der Stelle $t = 0$ eine stationäre Stelle. Zudem ist die zweite Ableitung für $r_2 \neq 0$ gleich $f''_{\mathbf{0},\mathbf{r}}(0) > 0$. Aus dem hinreichenden Kriterium zweiter Ordnung für reelle Funktionen in einer Variable, Satz 5.7.7, kann man also schliessen, dass an der Stelle $t = 0$ ein lokales Minimum des Vertikalschnitts durch $\mathbf{0}$ in Richtung $\mathbf{r}$ vorliegt. Für $r_2 = 0$ gilt $f'_{\mathbf{0},\mathbf{r}}(t) = f''_{\mathbf{0},\mathbf{r}}(t) = 0$. Anhand der ersten beiden Ableitungen kann also keine Aussage über die stationäre Stelle getroffen werden. Die dritte und vierte Ableitung berechnet sich als

$$\begin{aligned}
f^{(3)}_{\mathbf{0},\mathbf{r}}(t) &= 24r_1^4 t - 18r_1^2 r_2 && \text{und damit } f^{(3)}_{\mathbf{0},\mathbf{r}}(0) = 0 \\
f^{(4)}_{\mathbf{0},\mathbf{r}}(t) &= 24r_1^4 && \text{und damit } f^{(4)}_{\mathbf{0},\mathbf{r}}(0) = 24r_1^4 > 0 \text{ für } r_1 \neq 0.
\end{aligned}$$

Für $r_1 \neq 0$ kann man also aus dem hinreichenden Kriterium n-ter Ordnung für reelle Funktionen in einer Variable, Satz 5.7.8, schliessen, dass an der Stelle $t = 0$ ein lokales Minimum des Vertikalschnitts durch $\mathbf{0}$ in Richtung $\mathbf{r}$ vorliegt, wenn $r_1 \neq 0$ ist.

Ist $\mathbf{r} \in \mathbb{R}^2$ eine Richtung, so ist $\mathbf{r} \neq \mathbf{0}$. Zusammenfassend hat der Vertikalschnitt von f durch $\mathbf{0}$ also in jeder Richtung ein lokales Minimum an der Stelle $t = 0$ mit Wert $f(\mathbf{0}) = 0$. Beispielsweise ist der Vertikalschnitt durch $\mathbf{0}$ in Richtung $\mathbf{r} = \mathbf{e}^2$ gleich $f_{\mathbf{x}^0,\mathbf{r}}(t) = f(0,t) = t^2 > 0$ für alle $t \neq 0$. Somit ist $\mathbf{0}$ ein globales Minimum des Vertikalschnitts in Richtung $\mathbf{e}^2$.

Überraschenderweise liegt an dieser Stelle $\mathbf{x}^0 = \mathbf{0}$ jedoch kein lokales Minimum der Funktion f vor. Um sich davon zu überzeugen, kann man in einem ersten Schritt einige Funktionswerte an Stellen $\mathbf{x}$ nahe $\mathbf{0}$ berechnen. So ergibt sich beispielsweise $f(1, 1.5) = -\frac{5}{4} < 0$, $f(\frac{1}{2}, \frac{3}{8}) = -\frac{5}{64} < 0$ oder $f(\frac{1}{20}, \frac{3}{800}) = -\frac{5}{640000} < 0$. Diese Werte wecken eine gewisse Skepsis gegenüber der Behauptung, dass sich an der Stelle $\mathbf{0}$ ein lokales Minimum mit $f(\mathbf{0}) = 0$ befinden soll. Um zu zeigen, dass an der Stelle $\mathbf{0}$ kein lokales Minimum ist, muss man aber zeigen, dass in jeder noch so kleinen ε-Umgebung von $\mathbf{0}$ eine Stelle mit einem Funktionswert kleiner als $f(\mathbf{0}) = 0$ existiert. Um diese Stelle zu finden, formen wir die Abbildungsvorschrift von f um:

$$f(\mathbf{x}) = x_1^4 - 3x_1^2 x_2 + x_2^2 = 2x_1^4 - 3x_1^2 x_2 + x_2^2 - x_1^4 = (x_1^2 - x_2)(2x_1^2 - x_2) - x_1^4.$$

An dieser Form erkennt man, dass die Funktion negative Werte annimmt, wenn einer der Faktoren des Produkts im letzten Ausdruck positiv und der andere negativ ist, also wenn man x_2 so wählt, dass

$$x_1^2 < x_2 < 2x_1^2$$

gilt. Wählt man also $x_2 = 1.5x_1^2$, so gilt stets

$$f(x_1, 1.5x_1^2) = x_1^4 - 3x_1^2(1.5x_1^2) + (1.5x_1^2)^2 = -1.25x_1^4 < 0.$$

Da es in jeder ε-Umgbung eine Stelle mit $x_2 = 1.5x_1^2$ gibt,[10] existiert also in jeder ε-Umgebung von $\mathbf{0}$ eine Stelle, welche einen kleineren Funktionswert hat als $f(\mathbf{0}) = 0$. Die Stelle $\mathbf{0}$ ist somit kein lokales Minimum von f.[11] Zudem findet man beispielsweise auf dem Vertikalschnitt durch $\mathbf{0}$ in Richtung $\mathbf{e}^2$ Stellen in jeder ε-Umgbung um $\mathbf{0}$, welche grössere Funktionswerte als $f(\mathbf{0}) = 0$ haben. Die Stelle $\mathbf{0}$ ist somit kein lokales Maximum von f.

Zusammenfassend liegt an der Stelle $\mathbf{0}$ kein lokales Extremum sondern ein Sattelpunkt der Funktion vor. ■

Z Die Hesse-Matrix einer zweimal stetig partiell differenzierbaren Funktion an einer lokalen Minimalstelle ist stets positiv semidefinit. Umgekehrt liegt nicht an jeder stationären Stelle mit positiv semidefiniter Hesse-Matrix ein lokales Minimum vor. Die Hesse-Matrix einer zweimal stetig partiell differenzierbaren Funktion an einer lokalen Maximalstelle ist stets negativ semidefinit. Umgekehrt liegt nicht an jeder stationären Stelle mit negativ semidefiniter Hesse-Matrix ein lokales Maximum vor.

Aus den Extrema der Vertikalschnitte von f durch $\mathbf{x}^0$ kann man nicht direkt auf ein lokales Extremum von f schliessen.

9.4.5 Bestimmung globaler Extrema

Ziele dieses Unterkapitels

- Welche Stellen sind im Allgemeinen Kandidaten für lokale und globale Extrema einer reellen Funktion in n Variablen?
- Was kann man über stationäre Stellen konvexer oder konkaver reeller Funktionen in n Variablen aussagen?

Wie schon bei reellen Funktionen in einer Variable trifft das Kriterium von Fermat nur eine Aussage für offene Definitionsbereiche und stetig partiell differenzierbare Funktionen. Es trifft keine Aussage über Stellen, die auf dem Rand des Definitionsbereichs liegen. Es trifft allgemein auch keine Aussage über Stellen, an denen die Funktion nicht differenzierbar ist. Auf der Suche nach lokalen und globalen Extremalstellen müssen also prinzipiell auch hier drei Arten von Stellen untersucht werden:

1. Stationäre Stellen (mit $\nabla f(\mathbf{x}^0) = \mathbf{0}$);
2. Innere Stellen, an denen f nicht stetig partiell differenzierbar ist;
3. Randstellen des Definitionsbereichs (sofern sie zum Definitionsbereich gehören).

Wir hatten dies bereits bei der Diskussion reeller Funktionen in einer Variable kennengelernt. Die Untersuchung unstetiger oder nicht differenzierbarer Funktionen ist in der Regel sehr aufwändig. Um die Unterschiede zu Funktionen in einer reellen Variable zu betonen

[10] Beispielsweise die Stelle $(0.5\varepsilon, 0.375\varepsilon^2)^T \in U(\mathbf{0}, \varepsilon)$ (für $\varepsilon < 1$) mit $f(0.5\varepsilon, 0.375\varepsilon^2) = (0.5\varepsilon)^4 - 3(0.5\varepsilon)^2 0.375\varepsilon^2 + (0.375\varepsilon^2)^2 = -\frac{5}{64}\varepsilon^4 < 0$.

[11] Anhand der Ungleichung $x_1^2 < x_2 < 2x_1^2$ erkennt man, dass man sich der Stelle $\mathbf{0}$ auf einem "krummem Weg" nähern muss, um durchgehend negative Funktionswerte zu finden, wogegen ein Vertikalschnitt sich immer auf einem "geraden Weg" nähert.

ohne unnötige Komplikationen zu betrachten, beschränken wir uns in diesem Kapitel auf Funktionen mit stetigen zweiten partiellen Ableitungen und offenen Definitionsbereichen.

Einfacher gestaltet sich die Überprüfung einer konvexen bzw. einer konkaven Funktion: Eine konvexe Funktion hat an einer stationären Stelle stets ein globales (und damit auch lokales) Minimum. Eine konkave Funktion hat an einer stationären Stelle stets ein globales (und damit auch lokales) Maximum.

Satz 9.4.6 — Globale Extrema konvexer und konkaver Funktionen.
Seien $D \subseteq \mathbb{R}^n$ eine offene, konvexe Menge und $f : D \to Z$ eine stetig partiell differenzierbare Funktion in n Variablen.

- Ist f konvex, dann besitzt f in $\mathbf{x}^0 \in D$ ein globales Minimum genau dann, wenn $\nabla f(\mathbf{x}^0) = \mathbf{0}$ gilt.
- Ist f konkav, dann besitzt f in $\mathbf{x}^0 \in D$ ein globales Maximum genau dann, wenn $\nabla f(\mathbf{x}^0) = \mathbf{0}$ gilt.

Beweis. Einen Beweis findet man in Kall (1982, Satz 5.11, Satz 5.13). ■

Ist die Hesse-Matrix also an allen Stellen des Definitionsbereichs positiv semidefinit, so befindet sich an jeder stationären Stelle ein globales Minimum; ist sie an allen Stellen des Definitionsbereichs negativ semidefinit, befindet sich dort ein Maximum.

■ Beispiel 9.4.10 — Fortsetzung von Beispiel 9.4.6.
In Beispiel 9.4.6 haben wir die Funktion $f(\mathbf{x}) = x_1^2 + 4(x_2 - 1)^4 + 3$ analysiert. Die Hesse-Matrix von f ist für alle $\mathbf{x}^0 \in D$

$$H_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 48(x_2^0 - 1)^2 \end{pmatrix}.$$

Diese Hesse-Matrix hat die Eigenwerte $2 \geq 0$ und $48(x_2^0 - 1)^2 \geq 0$. Damit ist die Hesse-Matrix für alle $\mathbf{x}^0$ positiv semidefinit und die Funktion f eine konvexe Funktion. Die Funktion f hat eine stationäre Stelle bei $(0,1)^T$. An dieser Stelle muss somit ein globales und damit auch ein lokales Minimum vorliegen. ■

Z Kandidaten für lokale und globale Extrema einer reellen Funktion in n Variablen sind im Allgemeinen stationäre Stellen, Stellen, an welchen die Funktion nicht stetig partiell differenzierbar ist und Randstellen des Definitionsbereichs, wenn sie Elemente des Definitionsbereichs darstellen.

Stationäre Stellen einer konvexen Funktion stellen stets globale Minima dar. Stationäre Stellen einer konkaven Funktion stellen stets globale Maxima dar.

9.5 Mehrfachintegrale

Ebenso wie wir in Kapitel 5 Integrale von Funktionen in einer Variable gebildet haben, können Integrale von Funktionen in mehreren Variablen gebildet werden. Bei der Bildung des Integrals über x_i werden (analog zu partiellen Ableitungen) alle Variablen ausser x_i konstant gehalten. Da das Integral von f über x_i dann von den Werten der anderen Variablen abhängt, spricht man von Parameterintegralen.[12] In folgenden Definitionen und Sätzen beschränken wir uns auf Integrale von Funktionen in $n = 2$ Variablen. Die Definitionen und Sätze lassen sich jedoch einfach für $n > 2$ erweitern.

9.5.1 Parameterintegrale

Ziele dieses Unterkapitels

- Was versteht man unter dem (Parameter-)Integral einer stetigen reellen Funktion (in zwei Variablen) f über x_i?

Definition 9.5.1 — Parameterintegral.
Für eine stetige reelle Funktion (in zwei Variablen) $f : [a_1, b_1] \times [a_2, b_2] \to \mathbb{R}$ ist

$$F_1(x_2) = \int_{a_1}^{b_1} f(\mathbf{x})\ dx_1$$

das (Parameter-)Integral der Funktion f über x_1 (für festes x_2) und

$$F_2(x_1) = \int_{a_2}^{b_2} f(\mathbf{x})\ dx_2$$

das (Parameter-)Integral der Funktion f über x_2 (für festes x_1).

■ Beispiel 9.5.1 — Berechnung von Parameterintegralen.
Die Funktion $f : [0,3] \times [0,2] \to \mathbb{R}$ mit $f(\mathbf{x}) = -2x_1^2 + e^{x_2}$ ist für alle $x_1 \in [0,3]$ und $x_2 \in [0,2]$ definiert. Das (Parameter-)Integral von f über x_1 ist

$$\begin{aligned} F_1(x_2) &= \int_0^3 -2x_1^2 + e^{x_2}\ dx_1 = \left[\frac{-2}{3}x_1^3 + x_1 e^{x_2}\right]_0^3 \\ &= -18 + 3e^{x_2} + 0 - 0 = 3e^{x_2} - 18. \end{aligned}$$

Das (Parameter-)Integral von f über x_2 ist

$$\begin{aligned} F_2(x_1) &= \int_0^2 -2x_1^2 + e^{x_2}\ dx_2 = \left[-2x_1^2 x_2 + e^{x_2}\right]_0^2 \\ &= -2x_1^2 \cdot 2 + e^2 + 0 - 1 = -4x_1^2 + e^2 - 1. \end{aligned}$$

Beide (Parameter-)Integrale werden in den Abbildungen 9.58 und 9.59 dargestellt. ■

Z Für eine stetige reelle Funktion (in zwei Variablen) $f : [a_1, b_1] \times [a_2, b_2] \to \mathbb{R}$ ist $\int_{a_i}^{b_i} f(\mathbf{x})\ dx_i$ das (Parameter-)Integral der Funktion f über x_i.

9.5.2 Doppelintegrale

[12] Ein logischer Name an dieser Stelle wäre "partielles Integral" in Anlehnung an die partielle Ableitung. Da partielle Integration jedoch schon für die Umkehrung der Produktregel genutzt wird, wäre diese Benennung irreführend.

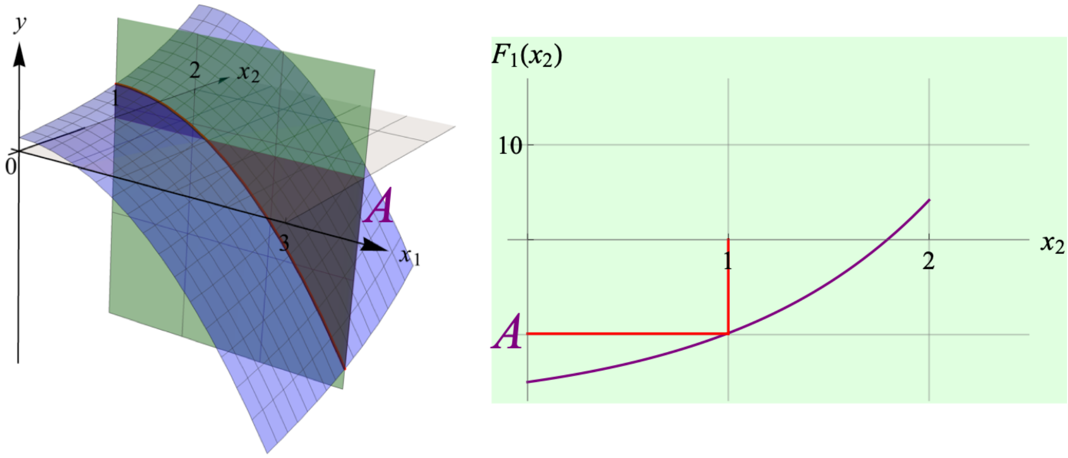

Abbildung 9.58: Darstellung des (Parameter-)Integrals von f über x_1.

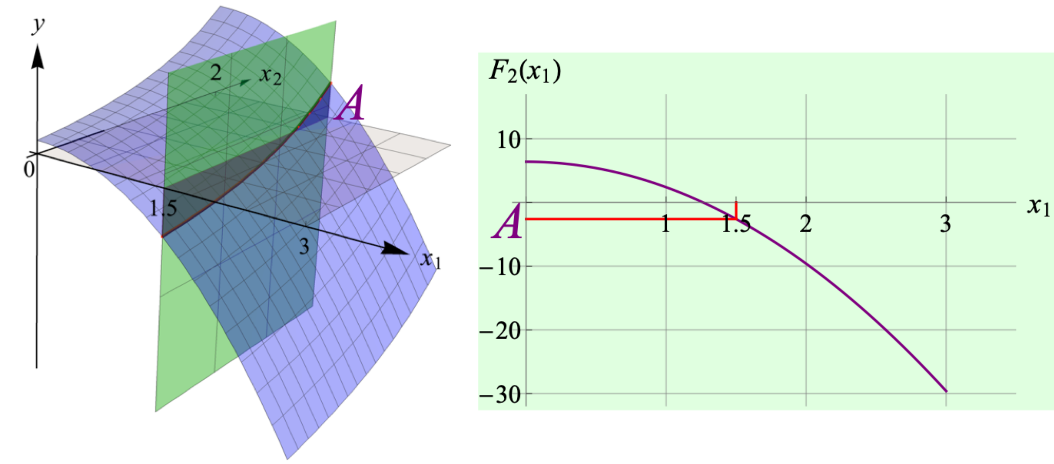

Abbildung 9.59: Darstellung des (Parameter-)Integrals von f über x_2.

Ziele dieses Unterkapitels

- Was versteht man unter einem Doppelintegral einer stetigen reellen Funktion in zwei Variablen über einer Fläche $[a_1, b_1] \times [a_2, b_2]$?
- Ist das Integral des Produkts (der Summe, der Differenz oder des Quotienten) zweier Funktionen in je einer Variablen gleich dem Produkt (der Summe, der Differenz oder dem Quotienten) der Integrale der Funktionen in je einer Variablen?

Da das Integral einer stetigen Funktion selbst wieder stetig und damit integrierbar ist, kann man sowohl die Funktion $F_1(x_2)$ nach x_2 als auch die Funktion $F_2(x_1)$ nach x_1 integrieren. Diese Integrale nennt man auch Doppelintegrale.

Definition 9.5.2 — Doppelintegral.
Für eine stetige reelle Funktion (in zwei Variablen) $f : [a_1, b_1] \times [a_2, b_2] \to \mathbb{R}$ mit

Parameterintegralen $F_1(x_2)$ und $F_2(x_1)$ nennt man

$$\int_{a_2}^{b_2} F_1(x_2)\,dx_2 = \int_{a_2}^{b_2}\left(\int_{a_1}^{b_1} f(\mathbf{x})\,dx_1\right)\,dx_2 \text{ bzw.}$$
$$\int_{a_1}^{b_1} F_2(x_1)\,dx_1 = \int_{a_1}^{b_1}\left(\int_{a_2}^{b_2} f(\mathbf{x})\,dx_2\right)\,dx_1$$

das Doppelintegral der Funktion f über $[a_1,b_1]\times[a_2,b_2]$.

Oft werden die Klammern hierbei auch weggelassen, da mehrfache Integrale immer von innen nach aussen gelesen werden, so wie auch obige Klammern nahelegen.

■ Beispiel 9.5.2 — Berechnung von Doppelintegralen.
Berechnen wir die Doppelintegrale der Funktion f aus Beispiel 9.5.1, erhalten wir als Integral über x_2 des Parameterintegrals $F_1(x_2)$ über x_1

$$\int_0^2 \underbrace{\int_0^3 (-2x_1^2+e^{x_2})\,dx_1}_{F_1(x_2)}\,dx_2 = \int_0^2 (3e^{x_2}-18)\,dx_2 = [3e^{x_2}-18x_2]_0^2$$
$$= 3e^2 - 18\cdot 2 - 3e^0 + 0$$
$$= 3e^2 - 39.$$

Das Integral über x_1 des Parameterintegrals $F_2(x_1)$ über x_2 ist

$$\int_0^3 \underbrace{\int_0^2 (-2x_1^2+e^{x_2})\,dx_2}_{F_2(x_1)}\,dx_1 = \int_0^3 (-4x_1^2+e^2-1)\,dx_1$$
$$= \left[-4\frac{x_1^3}{3}+e^2x_1-x_1\right]_0^3$$
$$= -4\frac{3^3}{3}+e^2 3-3+4\frac{0^3}{3}-e^2 0+0$$
$$= 3e^2-39.$$

Das Doppelintegral, bei dem man zuerst bezüglich x_1 und dann bezüglich x_2 integriert, entspricht also dem Doppelintegral, bei dem man zuerst bezüglich x_2 und dann bezüglich x_1 integriert. Allerdings muss man hierbei stets vorsichtig sein und jeweils über die zu x_1 bzw. x_2 gehörigen Intervalle integrieren. ■

Bei stetigen Funktionen ist die Reihenfolge der Integration dabei stets egal. Es gilt:

Satz 9.5.1 — Doppelintegrale (Satz von Fubini).
Für eine stetige reelle Funktion (in zwei Variablen) $f:[a_1,b_1]\times[a_2,b_2]\to\mathbb{R}$ sind die (Parameter-)Integrale $F_1(x_2)=\int_{a_1}^{b_1} f(\mathbf{x})\,dx_1$ und $F_2(x_1)=\int_{a_2}^{b_2} f(\mathbf{x})\,dx_2$ stetig und es gilt

$$\int_{a_2}^{b_2} F_1(x_2)\,dx_2 = \int_{a_2}^{b_2}\int_{a_1}^{b_1} f(\mathbf{x})\,dx_1\,dx_2 = \int_{a_1}^{b_1} F_2(x_1)\,dx_1 = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(\mathbf{x})\,dx_2\,dx_1.$$

Beweis. Einen Beweis findet man in Forster (1984, Kapitel 9, Satz 3). ■

Wir demonstrieren diese Aussage nochmals an dem Beispiel einer Funktion mit Abbildungsvorschrift $f(\mathbf{x}) = \sqrt{x_1 x_2}$.

■ Beispiel 9.5.3 — Ein weiteres Doppelintegral.
Wir betrachten die Funktion $f : [0,4] \times [0,6] \to \mathbb{R}$ mit $f(\mathbf{x}) = \sqrt{x_1 x_2}$. Die Funktion ist in Abbildung 9.60 dargestellt, sie ist für $x_1 \in [0,4]$ und $x_2 \in [0,6]$ definiert. Das Doppelintegral

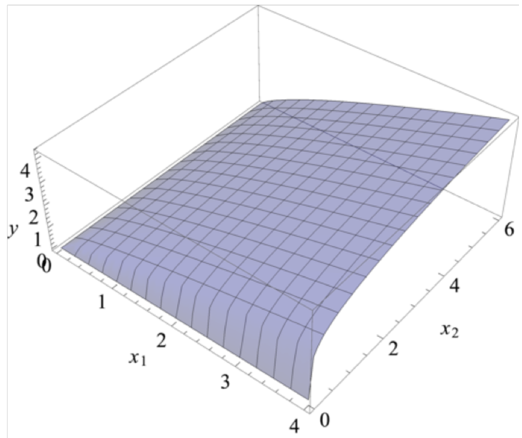

Abbildung 9.60: Die Funktion $f(\mathbf{x}) = \sqrt{x_1 x_2}$.

kann man hier wie folgt berechnen:

$$\int_0^6 \int_0^4 \sqrt{x_1 x_2}\, dx_1\, dx_2 = \int_0^6 \left[\frac{2}{3} x_1^{\frac{3}{2}} \sqrt{x_2}\right]_0^4 dx_2$$
$$= \int_0^6 \frac{16}{3} \sqrt{x_2}\, dx_2 = \left[\frac{16}{3}\frac{2}{3} x_2^{\frac{3}{2}}\right]_0^6 = \frac{192\sqrt{6}}{9} = \frac{64\sqrt{6}}{3}.$$

Analog zum Fall von nur einer einer Variablen entspricht das Ergebnis hier dem Volumen unterhalb des Graphen im Rechteck $[0,4] \times [0,6]$, siehe Abbildung 9.61. ■

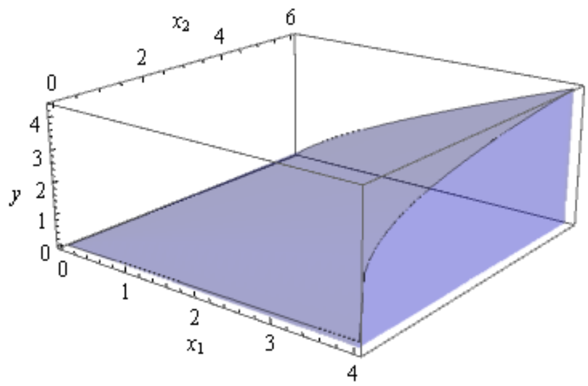

Abbildung 9.61: Volumen unterhalb des Graphen von f im Rechteck $[0,4] \times [0,6]$.

Integrale reeller Funktionen in zwei Variablen mit Definitionsbereich $[a_1,b_1]\times[a_2,b_2]$ und einem Wertebereich innerhalb von $[0,+\infty)$, lassen sich als das Volumen interpretieren, das oben vom Graphen der Funktion und unten vom Rechteck $[a_1,b_1]\times[a_2,b_2]$ begrenzt wird. Werden nur negative Werte angenommen, entspricht das Integral auch hier betragsmässig dem Volumen, jedoch mit negativem Vorzeichen.

Eine Besonderheit des vorherigen Beispiels spielt in der Statistik unter dem Schlagwort "unabhängige Zufallsvariablen" eine wichtige Rolle: Hat $f(\mathbf{x})$ die Form $g_1(x_1)g_2(x_2)$, so kann das Doppelintegral dieser Funktion einfach als Produkt der Integrale von g_1 und g_2 berechnet werden. Wir verdeutlichen zunächst die Idee erneut im Beispiel, bevor wir sie formal festhalten.

■ Beispiel 9.5.4 — Fortsetzung von Beispiel 9.5.3.
Betrachten wir die Funktion $f:[0,4]\times[0,6]\to\mathbb{R}$ mit $f(\mathbf{x})=\sqrt{x_1x_2}=\sqrt{x_1}\sqrt{x_2}$ aus Beispiel 9.5.3, so gilt $f(\mathbf{x})=g_1(x_1)g_2(x_2)$ mit $g_1:[0,4]\to\mathbb{R}$, $g_1(x_1)=\sqrt{x_1}$ und $g_2:[0,6]\to\mathbb{R}$, $g_2(x_2)=\sqrt{x_2}$. Das Integral von g_1 ist dabei

$$\int_0^4 \sqrt{x_1}\,dx_1=\left[\frac{2}{3}x_1^{\frac{3}{2}}\right]_0^4=\frac{16}{3},$$

das Integral von g_2 ist

$$\int_0^6 \sqrt{x_2}\,dx_2=\left[\frac{2}{3}x_2^{\frac{3}{2}}\right]_0^6=\frac{12\sqrt{6}}{3}.$$

Als Produkt dieser Werte ergibt sich

$$\int_0^4 \sqrt{x_1}\,dx_1\cdot\int_0^6 \sqrt{x_2}\,dx_2=\frac{16}{3}.\frac{12\sqrt{6}}{3}=\frac{192\sqrt{6}}{9}=\frac{64\sqrt{6}}{3},$$

was dem Wert von

$$\int_0^6\int_0^4 \sqrt{x_1x_2}\,dx_1\,dx_2=\frac{192\sqrt{6}}{9}=\frac{64\sqrt{6}}{3}$$

entspricht. ■

Satz 9.5.2 — Integration des Produkts zweier Funktionen.
Sind g_1 und g_2 stetige reelle Funktionen in einer Variablen, wobei g_1 stetig ist auf $[a_1,b_1]$ und g_2 stetig ist auf $[a_2,b_2]$, dann gilt

$$\begin{aligned}&\int_{a_2}^{b_2}\int_{a_1}^{b_1} g_1(x_1)\cdot g_2(x_2)\,dx_1\,dx_2\\ =&\left(\int_{a_1}^{b_1} g_1(x_1)\,dx_1\right)\cdot\left(\int_{a_2}^{b_2} g_2(x_2)\,dx_2\right).\end{aligned}$$

Beweis. Mithilfe der Integrationsregel $\int_a^b \alpha f(x)\,dx=\alpha\int_a^b f(x)\,dx$ aus Kapitel 5 ergibt sich

$$\begin{aligned}
&\int_{a_2}^{b_2}\int_{a_1}^{b_1} g_1(x_1)\cdot g_2(x_2)\; dx_1\; dx_2\\
&=\int_{a_2}^{b_2} g_2(x_2)\cdot\left(\int_{a_1}^{b_1} g_1(x_1)\; dx_1\right)\; dx_2\\
&=\left(\int_{a_1}^{b_1} g_1(x_1)\; dx_1\right)\cdot\left(\int_{a_2}^{b_2} g_2(x_2)\; dx_2\right)
\end{aligned}$$

■

Das Doppelintegral eines Produkts zweier stetiger Funktionen die jeweils nur von x_1 bzw. x_2 abhängen ist also gleich dem Produkt zweier Integrale, wovon eines nur über x_1 und das andere nur über x_2 integriert.

■ Beispiel 9.5.5 — Das Integral des Produkts.
Definieren wir die Funktionen $g_1 : [0,3] \to \mathbb{R}, g_1(x_1) = -2x_1^2$ mit

$$\int_0^3 (-2x_1^2)\; dx_1 = -18$$

und $g_2 : [0,2] \to \mathbb{R}, g_2(x_2) = e^{x_2}$ mit

$$\int_0^2 e^{x_2}\; dx_2 = e^2 - 1,$$

so ist das Doppelintegral der Funktion $h : [0,3] \times [0,2] \to \mathbb{R}$ mit

$$h(x_1,x_2) = g_1(x_1)g_2(x_2) = -2x_1^2 e^{x_2}$$

gleich

$$\begin{aligned}
\int_0^3\int_0^2 (-2x_1^2e^{x_2})\; dx_2\; dx_1 &= \int_0^3 \left[(-2x_1^2e^{x_2})\right]_0^2\; dx_1\\
&= \int_0^3 (-2x_1^2(e^2-1))\; dx_1\\
&= \left[-\frac{2}{3}x_1^3(e^2-1)\right]_0^3\\
&= -\frac{2}{3}3^3(e^2-1)+\frac{2}{3}0^3(e^2-1)\\
&= -18(e^2-1) = \left(\int_0^3 (-2x_1^2)\; dx_1\right)\left(\int_0^2 e^{x_2}\; dx_2\right).
\end{aligned}$$

Das Integral des Produkts entspricht also dem Produkt der Integrale. ■

■ Beispiel 9.5.6 — Das Integral der Summe.
Betrachten wir die Summe der Funktionen $g_1 : [0,3] \to \mathbb{R}, g_1(x_1) = -2x_1^2$ mit

$$\int_0^3 (-2x_1^2)\; dx_1 = -18$$

und $g_2 : [0,2] \to \mathbb{R}, g_2(x_2) = e^{x_2}$ mit

$$\int_0^2 e^{x_2}\, dx_2 = e^2 - 1,$$

so haben wir das Doppelintegral der Funktion $f : [0,3] \times [0,2] \to \mathbb{R}$ mit $f(\mathbf{x}) = g_1(x_1) + g_2(x_2) = -2x_1^2 + e^{x_2}$ als

$$\int_0^3 \underbrace{\int_0^2 (-2x_1^2 + e^{x_2})\, dx_2}_{F_2(x_1)}\, dx_1 = 3e^2 - 39$$

berechnet. Die Summe der Integrale über jeweils nur eine Variable entspricht hier (natürlich) nicht dem Integral von f. Es gilt aber:

$$\begin{aligned}\int_0^3 \int_0^2 (-2x_1^2 + e^{x_2})\, dx_2\, dx_1 &= \int_0^3 \underbrace{\int_0^2 (-2x_1^2)\, dx_2}_{=-4x_1^2}\, dx_1 + \int_0^3 \underbrace{\int_0^2 e^{x_2}\, dx_2}_{=e^2-1}\, dx_1 \\ &= \int_0^3 (-4x_1^2)\, dx_1 + \int_0^3 (e^2-1)\, dx_2 = -36 + 3(e^2-1) \\ &= 3e^2 - 39.\end{aligned}$$

■

Es ist also zu beachten, dass Satz 9.5.2 nur für Produkte, nicht für Summen, Differenzen, oder Quotienten gilt. So gilt allgemein zwar für Summen

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} (g_1(x_1) + g_2(x_2))\, dx_1\, dx_2 = \int_{a_2}^{b_2} \int_{a_1}^{b_1} g_1(x_1)\, dx_1\, dx_2 + \int_{a_2}^{b_2} \int_{a_1}^{b_1} g_2(x_2)\, dx_1\, dx_2,$$

aber

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} g_1(x_1) + g_2(x_2)\, dx_1\, dx_2 \neq \int_{a_1}^{b_1} g_1(x_1)\, dx_1 + \int_{a_2}^{b_2} g_2(x_2)\, dx_2.$$

Z

Für eine stetige reelle Funktion $f : [a_1,b_1] \times [a_2,b_2] \to \mathbb{R}$ mit (Parameter-)Integralen $F_1(x_2)$ und $F_2(x_1)$ nennt man

$$\int_{a_2}^{b_2} F_1(x_2)\, dx_2 = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(\mathbf{x})\, dx_1 \right) dx_2 = \int_{a_1}^{b_1} F_2(x_1)\, dx_1 = \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(\mathbf{x})\, dx_2 \right) dx_1$$

Doppelintegral der Funktion f.

Sind f und g stetige reelle Funktionen in jeweils einer Variablen, wobei f stetig ist auf $[a_1,b_1]$ und g stetig ist auf $[a_2,b_2]$, dann gilt $\int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1)g(x_2)\, dx_2 dx_1 = \int_{a_1}^{b_1} f(x_1)\, dx_1 \cdot \int_{a_2}^{b_2} g(x_2)\, dx_2$. Diese Vereinfachung kann nicht auf Summen, Differenzen oder Quotienten übertragen werden.