

3 Analysis in mehreren Variablen

9 Differentialrechnung bei mehreren Variablen

9.1 Funktionen zwischen Tupelräumen

9.1.1 Die betrachteten Funktionentypen

Wir werden in diesem Kapitel Funktionen (Synonym für Abbildungen, s.7.1.1) folgenden Typs untersuchen:

$\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}$	Typ (1) (reellwertige Funktionen)
$\mathbb{R} \supseteq D \xrightarrow{f} \mathbb{R}^n$	Typ (2)
$\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^n$	Typ (3)
$\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m$	Typ (4) (allgemeiner Typ)

Übereinkunft

Wir schreiben in diesem Kapitel die n -Tupel $x \in \mathbb{R}^n$ meist, wie in der Analysis üblich, als $x = (x_1, x_2, \dots, x_n)$ und nur selten als einspaltigen Vektor, wie wir es in der Linearen Algebra getan haben.

Anmerkung

Bei Funktionen aus den Anwendungen in der Wirtschaftstheorie wird der Definitionsbereich D oft die Teilmenge

$$\mathbb{R}_{\geq 0}^n := \{ x \in \mathbb{R}^n \mid x_i \geq 0 \text{ für alle Koordinaten } x_i \text{ von } x \}$$

sein.

Wir nennen $\mathbb{R}_{\geq 0}^n$ den nicht-negativen Quadranten von $\mathbb{R}^n$. (Ein “Quadrant” ist es nur im $\mathbb{R}^2$, im $\mathbb{R}^3$ ist es ein “Oktant” und im $\mathbb{R}^3$ die entsprechende Verallgemeinerung. In Ermangelung eines entsprechenden allgemeinen Ausdrucks bleiben wir beim Wort “Quadrant”.)

Die Teilmenge

$$\mathbb{R}_{> 0}^n := \{ x \in \mathbb{R}^n \mid x_i > 0 \text{ für alle Koordinaten } x_i \text{ von } x \}$$

heißt entsprechend der positive Quadrant im $\mathbb{R}^n$.

Beispiele (zur Motivation):

1 Lineare Funktionen

Lineare Funktionen des allgemeinen Typs (4) haben wir in 7.1. behandelt:

$$\text{Für } A \in \mathbb{R}^{m \times n} \text{ hat man } f_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m, x \longmapsto A \cdot x \quad (\text{s.7.1.2})$$

Spezielle Funktionen des Typs (1) sind die (linearen) Zielfunktionen der linearen Programmierung:

$$f = f_Z : \mathbb{R}^n \longrightarrow \mathbb{R}, x \longmapsto z_1 x_1 + \dots + z_n x_n = Z \cdot x \text{ mit } Z = (z_1 \dots z_n) \in \mathbb{R}^{1 \times n}$$

- 2 Hier Beispiele für reellwertige Funktionen, wie sie in den Wirtschaftswissenschaften benutzt werden.

(i) In allgemeinen Wirtschaftstheorien arbeitet man oft mit Funktionen in drei Variablen:

$f(K, A, t)$, wo $K = \underline{\text{Kapital}}$, $A = \underline{\text{Arbeit}}$, $t = \text{Parameter für den technischen Fortschritt}$

(ii) Beliebte Funktionen in Wirtschaftstheorien sind Funktionen wie

$$f(x, y, z) = x^\alpha \cdot y^\beta \cdot z^\gamma \text{ mit } x, y, z \in \mathbb{R}, 0 \leq \alpha, \beta, \gamma, 1 = \alpha + \beta + \gamma$$

(iii) In der Wirtschaftstheorie arbeitet man auch mit sog. "Nutzenfunktionen". Das sind reellwertige Funktionen $f(x_1, x_2, \dots, x_n)$ in n Variablen, die so interpretiert werden:

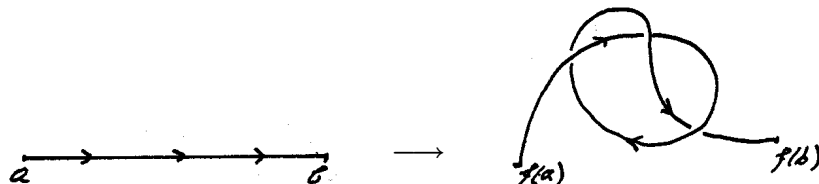
Die Argumente $x = (x_1, \dots, x_n)$ stehen für Güterbündel (s.6.1.1 Anwendungen (2)).

Die Funktion f wird benutzt, um den Nutzen von solchen Bündeln zu "vergleichen":

$f(x_1, \dots, x_n) > f(y_1, \dots, y_n)$ bedeutet: Das Bündel x wird dem Bündel y vorgezogen.

(iv) Der Mittelwert $(a_1, \dots, a_n) \mapsto \bar{a} := \frac{1}{n} \sum_{i=1}^n a_i$ und die Varianz $(a_1, \dots, a_n) \mapsto \frac{1}{n} \sum_{i=1}^n a_i^2$ sind Funktionen des Typs (1).

- 3 Funktionen des Typs (2) werden oft als Kurven interpretiert.



Mit solchen Kurven lassen sich etwa Bahnen von Himmelssonden modellieren.

- 4 Funktionen des Typs (3) treten als Koordinatentransformationen auf. Zum Beispiel:

$$\mathbb{R}^2 \setminus \{0\} =: D \xrightarrow{f} \left\{ \begin{pmatrix} r \\ \alpha \end{pmatrix} \in \mathbb{R}^2 \mid r > 0, 0 \leq \alpha < 2\pi \right\} \subset \mathbb{R}^2$$

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} r = \|x\| \\ \alpha = \text{der Winkel } \alpha, \text{ für den } \frac{1}{\|x\|} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \end{pmatrix}$$

Der Übergang von x zu diesem $f(x)$ heißt Übergang zu Polarkoordinaten.



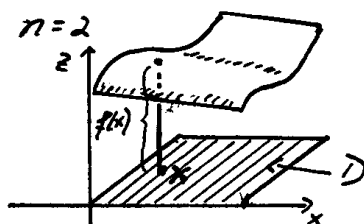
9.1.2 Anschauliche Vorstellung 1 : Der Graph einer reellwertigen Funktion

 $n=1$:

Man hat das übliche Schaubild (den Graph):

Der Graph von $f : I \rightarrow \mathbb{R}$ ist die Kurve

$$\left\{ \begin{pmatrix} x \\ f(x) \end{pmatrix} \mid x \in I \right\} \text{ im } \mathbb{R}^2$$

 $n=2$:

Man definiert den Graphen von

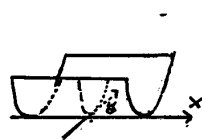
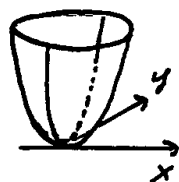
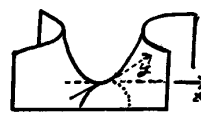
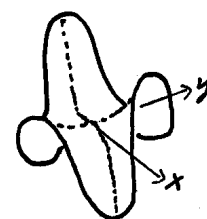
 $f : D \rightarrow \mathbb{R} \quad (D \subseteq \mathbb{R}^2)$ durch

$$G(f) = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid \begin{pmatrix} x \\ y \end{pmatrix} \in D, z = f \left(\begin{pmatrix} x \\ y \end{pmatrix} \right) \right\}$$



Stellt man sich ein Stück "abstrakter" Erdoberfläche vor, parametrisiert durch x = Längengrad, y = Breitengrad, und ist $z = h(x, y)$ = Höhe von (x, y) über dem Meeresspiegel, so ist der Graph die "echte" Erdoberfläche (mit all ihren Zerklüftungen). Siehe Bild.

Skizzen einiger weiterer Graphen:


 $f(x, y) = y^2$
 "Rinne"

 $f(x, y) = x^2 + y^2$
 Rotationsparaboloid

 $f(x, y) = x^2 - y^2$
 Sattelfläche

 $f(x, y) = x^3 - 3xy^2$
 "Affensattel"

Anmerkung:

In den Dimensionen $n \geq 3$ ist der Graph eine Teilmenge des $\mathbb{R}^{n+1}$, $n+1 \geq 4$ und kann direkt nicht mehr vorgestellt werden. In der Dimension 3 kann man noch die Höhenlinien des folgenden Abschnitts zur Veranschaulichung hinzuziehen.

9.1.3 Anschauliche Vorstellung 2 : Die Höhenlinien

Bezeichnung (Isoquanten)

Sei $D \subseteq \mathbb{R}^n$ und $f : D \rightarrow \mathbb{R}$ eine reelle Funktion auf D .

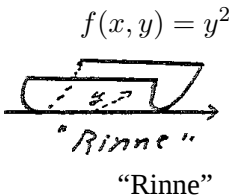
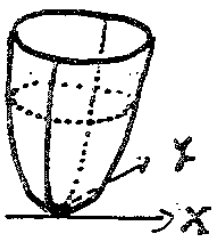
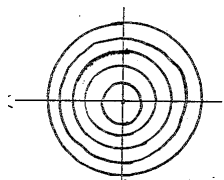
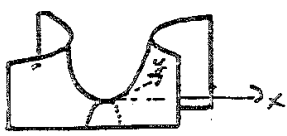
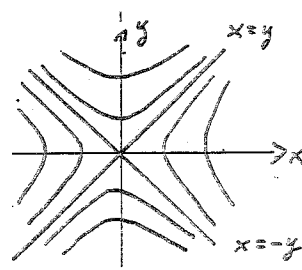
Die Punktmenge $I_c = \{x \in D \mid f(x) = c\}$, $c \in \mathbb{R}$, heißen **Höhenlinien** für $n = 2$, **Höhenflächen** für $n = 3$ und **Isoquanten** von f für allgemeines n .

Manchmal sagt man "Höhenlinien" der Einfachheit halber auch im allgemeinen Fall.

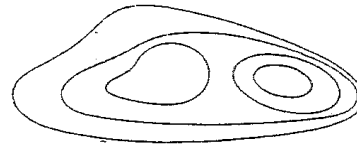
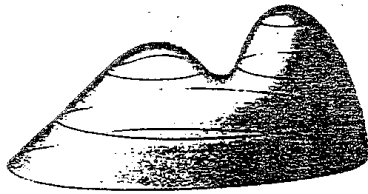
Anschaulich: Beim Vergleich des Schaubilds von f mit dem Relief einer Landschaft entsprechen die I_c 's den üblichen Höhenlinien, wie sie auf Karten eingezeichnet sind.

Höhenflächen im $\mathbb{R}^3$ muß man sich i.a. als gekrümmte Flächen vorstellen.

Beispiele bei früher betrachteten f :

<u>Funktion</u>	<u>Höhenlinien</u>
$f(x) = z_1 x_1 + \dots + z_n x_n$ linear (Gewinn-, Kostenfunktion)	Die Isogewinn- bzw. Isokostengeraden $f(x) = c$
$f(x, y) = y^2$  "Rinne"	 I_c = die beiden Parallelen zur y -Achse mit $y = +\sqrt{c}$ und $y = -\sqrt{c}$ für $c \neq 0$ und die (x -Achse für $c = 0$)
$f(x, y) = x^2 + y^2$ 	 Die I_c , $c > 0$, sind die konzentrischen Kreise um 0 mit Radius $\sqrt{c}$
$f(x, y) = x^2 - y^2$ 	 Für $c \neq 0$ sind die I_c Hyperbeln (mit zwei Ästen) mit den Asymptoten $x - y = 0$ und $x + y = 0$ Für $c = 0$ ist I_c das Koordinatenkreuz.

Noch eine abstrakte Landschaft:



9.1.4 Stetigkeit. Offene Mengen.

Bezeichnung:

Für $x \in \mathbb{R}^n$, $\varepsilon > 0$, heißt

$$B_\varepsilon(x) = \{y \in \mathbb{R}^n \mid \|y - x\| < \varepsilon\}$$

die (offene) ε -Kugel oder die ε -Umgebung von x .

Für die Anschauungsdimensionen:

Im $\mathbb{R}^1$: Intervall mit Mittelpunkt x und Länge 2ε .

Im $\mathbb{R}^2$: Kreisfläche mit Mittelpunkt x und Radius ε .

Im $\mathbb{R}^3$: Kugel mit Mittelpunkt x und Radius ε , etc.

Definition 1 (Stetigkeit):

Sei $D \subseteq \mathbb{R}^n$, $f : D \rightarrow \mathbb{R}^m$ eine Abbildung, $a \in D$.

f heißt stetig in a , wenn gilt:

Für alle $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß für alle $x \in B_\delta(a) \cap D \subseteq \mathbb{R}^n$ gilt:

$$f(x) \in B_\varepsilon(f(a)).$$

Wir haben die Definition nur notiert, um zu zeigen, daß sich die Definition von der einfachen Situation mit einer Variablen (siehe die $\varepsilon - \delta$ -Definition der Stetigkeit in 3.3.1) problemlos auf den allgemeinen Fall der Funktionen vom Typ (4) übertragen läßt.

Weitergehend werden wir Stetigkeit nicht untersuchen. Das Phänomen der Stetigkeit ist in den $\mathbb{R}^n$, $n \geq 2$, trotz der formal gleichen Definition viel komplexer als für $n = 1$.

Definition 2:

$D \subset \mathbb{R}^n$ heißt offen: $\iff$ Für alle $x \in D$ gibt es ein $\varepsilon > 0$, so daß noch $B_\varepsilon(x) \subseteq D$.

Definition 3:

Sei $D \subseteq \mathbb{R}^n$. Ein $x \in D$ heißt ein **innerer Punkt** von D , wenn es ein $\varepsilon > 0$ gibt mit $B_\varepsilon(x) \subseteq D$. Die Menge aller inneren Punkte von D heißt das **Innere** von D . Ein Punkt aus D , der kein innerer Punkt ist, heißt ein **Randpunkt** von D . Die Menge aller Randpunkte von D , heißt der **Rand** von D .

Schreibweisen: $\overset{\circ}{D} :=$ das Innere von D , $\partial D :=$ der Rand von D .

Beispiele

- (i) Ist $D = I$ ein Intervall im $\mathbb{R}^1 = \mathbb{R}$, so stimmt das hier definierte Innere mit dem Innern des Intervalls überein, wie es in 4.2.3 definiert wurde.
- (ii) Sei $D = \mathbb{R}_{\geq 0}^n$ der nicht-negative Quadrant. Dann ist $\overset{\circ}{D} = \mathbb{R}_{> 0}^n$ der positive Quadrant und $\partial D = \{x \in D \mid x_i = 0 \text{ für mindestens ein } i \text{ mit } 1 \leq i \leq n\}$.
- (iii) Bei den in 8.1.2 und 8.4.3 gezeichneten zulässigen Bereichen K von Linearen-Optimierungs-Problemen ist der Rand ∂K von K die Vereinigung der Randgeradenstücke (der “Seiten”) von K und das Innere $\overset{\circ}{K}$ ist das schraffierte Innere des Bereiches ohne die Seiten. /bin/sh: schraffierte: command not found

Bemerkung

Das Innere $\overset{\circ}{D}$ eines $D \subset \mathbb{R}^n$ ist offen. (Die leere Menge gilt als offen.)

9.2 Differenzieren reellwertiger Funktionen auf dem $\mathbb{R}^n$

Gemeint sind Funktionen des Typs $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}$.

9.2.1 Die partielle Ableitung

Definition 1 (partielle Ableitung):

Sei $f : D \rightarrow \mathbb{R}$ eine Funktion und sei x ein Punkt im Innern $\overset{\circ}{D}$ von D .
Für $i \in \{1, 2, \dots, n\}$ sei $e_i = (0, \dots, 0, \underset{\uparrow}{1}, 0, \dots, 0)$ der i -te Einheitsvektor.

Betrachte den Limes

$$\lim_{h \rightarrow 0} \frac{f(x + h \cdot e_i) - f(x)}{h} \stackrel{!}{=} \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x)}{h}$$

Existiert dieser Limes und ist gleich α , so heißt f in x nach der i -ten Variablen partiell differenzierbar. Der Limes α heißt die i -te partielle Ableitung von f in x .

Schreibweisen: $\alpha =: \frac{\partial f}{\partial x_i}(x) =: f_{x_i}(x) =: f_i(x) =: D_i f(x)$.

Ist $\frac{\partial f}{\partial x_i}(x)$ definiert für alle $x \in D$, so heißt die Funktion

$$\frac{\partial f}{\partial x_i} : D \rightarrow \mathbb{R}, \mapsto \frac{\partial f}{\partial x_i}(x)$$

die i -te partielle Ableitung von f , f heißt partiell differenzierbar in D nach der i -ten Koordinate (“nach x_i ” oder kurz “nach i ”). Ist zudem $\frac{\partial f}{\partial x_i}$ stetig, so heißt f stetig partiell differenzierbar.

Merkregel:

Beim partiellen Ableiten nach i wird angenommen, die Variablen $x_j, j \neq i$, seien konstant, und nur die i -te Variable x_i sei variabel. Dann hat man eine Funktion vorliegen, die nur von einer Variablen, nämlich von x_i , abhängt. Diese Funktion wird dann nach $x := x_i$ abgeleitet wie gewohnt.

Bemerkung: Ist x auf dem Rand von ∂D von D , so existieren eventuell die entsprechenden einseitigen Limes $\lim_{h \rightarrow 0^-}$ oder $\lim_{h \rightarrow 0^+}$ (s. 3.3.3) statt des Limes $\lim_{h \rightarrow 0}$ in der Definition. Als partielle Ableitungen nimmt man dann diese "einseitigen Ableitungen".

Beispiel für so eine Situation: $D := \mathbb{R}_{\geq 0}^n$ und x ein Punkt aus D , bei dem (mindestens) eine Koordinate 0 ist.

Beispiele (für partielle Ableitungen):

$$(i) f(x, y, z) = x^3 \cdot y^4 \cdot z^2 - x \cdot y$$

$$\frac{\partial f}{\partial x}(x, y, z) = 3x^2 \cdot y^4 \cdot z^2 - y,$$

$$\frac{\partial f}{\partial y}(x, y, z) = 4x^3 y^3 \cdot z^2 - x,$$

$$\frac{\partial f}{\partial z}(x, y, z) = 2x^3 \cdot y^4 \cdot z$$

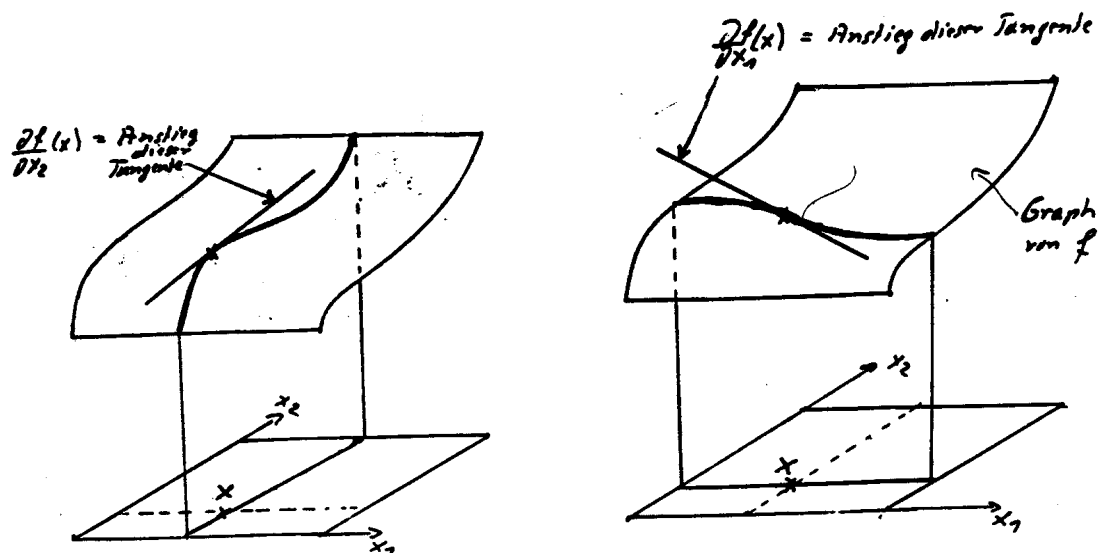
$$(ii) f(x, y, z) = x^2 \cdot y \cdot z - e^{2x} \cdot e^{y^2} \cdot e^{x \cdot \sin z}$$

$$\frac{\partial f}{\partial x}(x, y, z) = 2xy \cdot z - 2e^{2x} \cdot e^{y^2} \cdot e^{x \cdot \sin z} - \sin z \cdot e^{2x} \cdot e^{y^2} \cdot e^{x \cdot \sin z}$$

$$\frac{\partial f}{\partial z}(x, y, z) = x^2 \cdot y - x \cdot \cos z \cdot e^{2x} \cdot e^{y^2} \cdot e^{x \cdot \sin z}$$

$$\frac{\partial f}{\partial y}(x, y, z) = x^2 \cdot z - 2y \cdot e^{2x} \cdot e^{y^2} \cdot e^{x \cdot \sin z}$$

Veranschaulichung



Im "Schaubildraum" $\mathbb{R}^n \times \mathbb{R} = \{(x, z) \mid x \in \mathbb{R}^n, z \in \mathbb{R}\}$ betrachte die Parallelen zur x_i -Achse und zur z -Achse im Punkte $(x, 0)$, $x \in D \subset \mathbb{R}^m$. Die von diesen beiden Geraden erzeugte Ebene

E schneidet das Schaubild $G_f = \{(x, f(x)) \mid x \in D\}$ in einer Kurve. In E mit der Parallelen zur x_i -Achse als reeller Achse bekommt man so das Schaubild einer üblichen reellen Funktion mit x_i als reellem Parameter. Der Anstieg der Tangente dieses Schaubilds im Punkt $x_i = i$ -te Koordinate von x ist dann $\frac{\partial f}{\partial x_i}(x)$.

Definition 2: (höhere (“gemischte”) partielle Ableitungen)

Die i -te Ableitung $\frac{\partial f}{\partial x_i}$ als Funktion sei definiert.

Sei $j \in \{1, 2, \dots, n\}$.

Ist dann $\frac{\partial f}{\partial x_i}$ (in x) nach der j -ten Variablen partiell differenzierbar, so heißt die partielle Ableitung

$$\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) (x) =: \frac{\partial^2 f}{\partial x_j \partial x_i} (x) =: f_{x_i x_j}(x) =: D_j D_i f(x)$$

die partielle Ableitung 2. Ordnung (2. partielle Ableitung) nach i und nach j in x .

Entsprechend sind höhere “gemischte” partielle Ableitungen betrachtet, etwa

$$\frac{\partial^3}{\partial x_3 (\partial x_1)^2} := \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_1} \left(\frac{\partial f}{\partial x_1} \right) \right), \quad \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i} \text{ usw.}$$

Beispiel:

$$(iii) f(x, y, z) = x^3 \cdot y^2 \cdot \cos z$$

$$\frac{\partial^3 f}{\partial z \cdot (\partial x)^2} (x, y, z) = \frac{\partial}{\partial z} (6x \cdot y^2 \cdot \cos z) = -6x \cdot y^2 \cdot \sin z$$

Anmerkung

Die andere Reihenfolge der Indizes bei der Schreibweise $f_{x_i x_j}(x)$ kommt daher, daß nach Definition $f_{x_i x_j}(x) := (f_{x_i})_{x_j}$ ist.

Diese leichte Irritation durch die Bezeichnung ist aber folgenlos, da die gemischte 2. partielle Ableitung nach der Tatsache in 9.2.3 bei allen gebräuchlichen Funktionen unabhängig von der Reihenfolge ist.

9.2.2 Gradient und Hesse-Matrix

Sei $D \subset \mathbb{R}^n$ offen, $f : D \rightarrow \mathbb{R}$ sei in x partiell differenzierbar nach allen Variablen $i = 1, \dots, n$. (Ist D von Anfang an nicht offen, so kann man sich auf die “offene Situation” zurückziehen, indem man die Funktion auf das Innere von D einschränkt.)

Definition 1 (Gradient):

Das Tupel (der “Vektor”):

$$\text{grad } f(x) := \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) = (f_{x_1}(x), f_{x_2}(x), \dots, f_{x_n}(x))$$

heißt der Gradient von f in x .

Definition 2 (Hesse-Matrix):

Existieren alle partiellen Ableitungen 2. Ordnung in x , so heißt die (n, n) -Matrix

$$H_f(x) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(x) \right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}} = \begin{pmatrix} f_{x_1 x_1}(x) & f_{x_1 x_2}(x) & \dots & f_{x_1 x_n}(x) \\ f_{x_2 x_1}(x) & f_{x_2 x_2}(x) & \dots & f_{x_2 x_n}(x) \\ \vdots & \vdots & \dots & \vdots \\ f_{x_n x_1}(x) & f_{x_n x_2}(x) & \dots & f_{x_n x_n}(x) \end{pmatrix}$$

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

Beispiel: Sei $f(x_1, x_2, x_3) = x_1^2 x_2^3 x_3$. Es ist

$$\begin{aligned} \text{grad } f(x) &= (2x_1 x_2^3 x_3, 3x_1^2 x_2^2 x_3, x_1^2 x_2^3) \\ H_f(x) &= \begin{pmatrix} 2x_2^3 x_3 & 6x_1 x_2^2 x_3 & 2x_1 x_2^3 \\ 6x_1 x_2^2 x_3 & 6x_1^2 x_2 x_3 & 3x_1^2 x_2^2 \\ 2x_1 x_2^3 & 3x_1^2 x_2^2 & 0 \end{pmatrix} \end{aligned}$$

und konkret an der Stelle $x = (-1, 2, 3)$:

$$\text{grad } f(-1, 2, 3) = (-48, 36, 8), \quad H_f(-1, 2, 3) = \begin{pmatrix} 48 & -72 & -16 \\ -72 & 36 & 12 \\ -16 & 12 & 0 \end{pmatrix}$$

Einfaches, aber wichtiges allgemeines Beispiel des Gradienten:

Sei $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = z_1 x_1 + \dots + z_n x_n$ linear. Dann ist

$$\text{grad } f(x) = (z_1, z_2, \dots, z_n) \text{ für alle } x.$$

Beobachtung

Die Hesse-Matrix im Beispiel ist symmetrisch (s. 6.2.5). Das ist kein Zufall, sondern ein genereller Sachverhalt wegen folgender Tatsache:

Tatsache ("Lemma von Schwarz"):

Sei $f: D \rightarrow \mathbb{R}$ 2-mal stetig differenzierbar, d.h. alle partiellen Ableitungen bis zur Ordnung 2 existieren und sind stetig. Dann gilt:

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \quad \text{für alle } x \in D$$

und alle $i, j \in \{1, 2, \dots, n\}$.

Das heißt für jedes $x \in D$: Die Hesse-Matrix $H_f(x)$ ist eine symmetrische Matrix.

Hinweis

Bei den später behandelten Extremwertbestimmungen bei mehreren Variablen werden der Gradient die Rolle der ersten Ableitung und die Hesse-Matrix die Rolle der zweiten Ableitung übernehmen.

9.2.3 Richtungsableitungen

Vorbemerkung:

Erinnere die Definition

$$\lim_{h \rightarrow 0} \frac{f(x + h \cdot e_i) - f(x)}{h} =: \frac{\partial f}{\partial x_i}(x)$$

der partiellen Ableitung. Sie gibt den Anstieg der Funktion f in Richtung der positiven x_i -Achse wider.

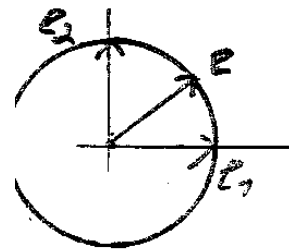
Nun gibt es noch all die anderen Richtungen, die nicht in Richtung einer Koordinatenachse zeigen. Man kann den Anstieg von f auch in diesen Richtungen definieren. Dazu eine formale Definition des Begriffs Richtung.

Bezeichnung:

Ein $e \in \mathbb{R}^n$ mit $\|e\| = 1$ heißt eine **Richtung** im $\mathbb{R}^n$.

Die Menge $\{e \in \mathbb{R}^n \mid \|e\| = 1\}$ aller Richtungen

wird oft mit S^{n-1} bezeichnet und **Einheitssphäre** genannt.



Beispiele im $\mathbb{R}^2$: (Man denke bei e an eine Kompaßnadel.)

Die Richtung $(1,0)$ ist die Richtung in positiver x -Achse ("Osten"), $(0,-1)$ "zeigt" in Richtung negativer y -Achse ("Süden"), $\frac{1}{\sqrt{2}}(1,1)$ zeigt in Richtung der positiven Hauptdiagonalen ("Nord-osten") usw.

Definition (Richtungsableitung):

Betrachte $f : D \rightarrow \mathbb{R}$ und sei e eine Richtung. Existiert der Limes

$$\lim_{h \rightarrow 0} \frac{f(x + h \cdot e) - f(x)}{h}$$

so wird er die **Richtungsableitung von f in Richtung e** genannt.

Schreibweise:
$$\lim_{h \rightarrow 0} \frac{f(x + h \cdot e) - f(x)}{h} =: \frac{\partial f}{\partial e}(x).$$

Wir werden in 9.3 als Anwendung der Kettenregel (s.dort) nachweisen, daß sich für die gebräuchlichen Funktionen die Richtungsableitungen in $x \in D$ auf folgende Weise aus dem Gradienten $\text{grad } f(x)$ berechnen lassen.

Tatsache :

Sei $D \subseteq \mathbb{R}^n$ offen. Sei $f : D \rightarrow \mathbb{R}$ stetig differenzierbar, d.h. alle partiellen Ableitungen in D existieren und sind stetig. Sei dann $a \in D$ und $e \in \mathbb{R}^n$, $\|e\| = 1$, sei eine Richtung. Dann gilt

$$\frac{\partial f}{\partial e}(a) = \langle \text{grad } f(a), e \rangle \text{ (Skalarprodukt).}$$

Beispiele:

(i) $e = e_i : \frac{\partial f}{\partial e}(x) = \frac{\partial f}{\partial x_i}(x)$

$$(ii) e = \frac{1}{\sqrt{2}}(1, 1) : \frac{\partial f}{\partial e}(x) = \frac{1}{\sqrt{2}} \frac{\partial f}{\partial x_1}(x) + \frac{1}{\sqrt{2}} \frac{\partial f}{\partial x_2}(x)$$

$$(iii) f(x, y, z) = x^2 \cdot z \cdot y^2 - \frac{2}{x \cdot y \cdot z}, \quad x = (1, \frac{1}{2}, 1)$$

Richtung: von x zum Punkt $(6, 3, 6)$.

Dann:

$$\text{grad} f(x) = (2xyz^2 + \frac{2}{x^2 \cdot y \cdot z}, x^2 z^2 + \frac{2}{xy^2 z}, 2x^2 yz + \frac{2}{xyz^2})$$

$$\text{grad} f(x) = (5, 9, 5)$$

$$\text{Der Richtungsvektor ist } e = \frac{(6, 3, 6) - (1, \frac{1}{2}, 1)}{\|(6, 3, 6) - (1, \frac{1}{2}, 1)\|} = \frac{(5, \frac{5}{2}, 5)}{\|(5, \frac{5}{2}, 5)\|} = (\frac{2}{3}, \frac{1}{3}, \frac{2}{3}).$$

Die gesuchte Richtungsableitung ist also:

$$\frac{\partial f}{\partial e}(x) = \langle (5, 9, 5), (\frac{2}{3}, \frac{1}{3}, \frac{2}{3}) \rangle = 3 + \frac{20}{3} = 9\frac{2}{3}.$$

9.2.4 Prägnante Eigenschaften des Gradienten

Wir beginnen mit einer Ungleichung zum Skalarprodukt, die eigentlich in den Abschnitt 6.3.5 gehört. (Sie wird beim Beweis der dort formulierten Dreiecksungleichung für die Norm gebraucht.)

Tatsache 1 (Die Cauchy-Schwarzsche-Ungleichung)

Für alle $x, y \in \mathbb{R}^n$ gilt

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

Dabei gilt die Gleichheit dann und nur dann, wenn x und y linear abhängig sind.

Der Beweis ist nicht schwer aber nicht trivial.

Als Anwendung erhält man:

Tatsache 2

Der Gradient zeigt in die Richtung, in der die Richtungsableitung maximal ist. Genauer:

Sei $a \in \overset{\circ}{D}$ und sei $\text{grad} f(a) \neq 0$. Sei $\gamma := \frac{\text{grad} f(a)}{\|\text{grad} f(a)\|}$ die "Richtung des Gradienten".

Dann gilt für alle Richtungen $e \in S^{n-1}$:

$$\frac{\partial f}{\partial e}(a) \leq \frac{\partial f}{\partial \gamma}(a) = \|\text{grad} f(a)\|$$

$$\text{Beweis: } \frac{\partial f}{\partial e}(a) = \langle \text{grad} f(a), e \rangle \leq \|\text{grad} f(a)\| \cdot \|e\| = \|\text{grad} f(a)\| = \frac{1}{\|\text{grad} f(a)\|}.$$

$$\langle \text{grad} f(a), \text{grad} f(a) \rangle = \langle \text{grad} f(a), \frac{\text{grad} f(a)}{\|\text{grad} f(a)\|} \rangle = \langle \text{grad} f(a), \gamma \rangle = \frac{\partial f}{\partial \gamma}(a).$$

Die Ungleichung dabei ist die Cauchy-Schwarzsche-Ungleichung, die darauf folgende Gleichung gilt, weil $\|e\| = 1$ ist. Der Rest sind Umformungen.

Eine ökonomische Deutung von Tatsache 2

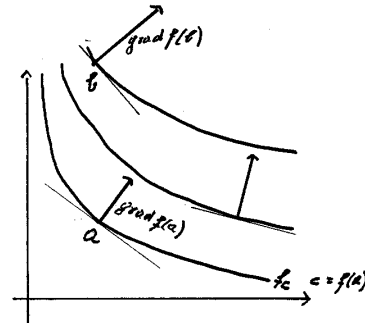
Ist f eine Produktionsfunktion, so hat man: Die Änderungsrate der Produktion ist am größten bei einer "Faktorvariation" in Richtung des Gradienten.

Eine anschauliche Eigenschaft des Gradienten ist die folgende:

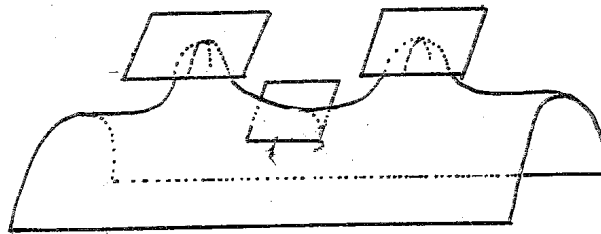
Tatsache 3

Der Gradient von f in a steht senkrecht auf der Isoquante von f durch a . (Das ist die Isoquante $I_{f(a)}$.)

Siehe die Vorlesung für Genauereres.

Ein einfaches Beispiel zum Verständnis:

Für lineare $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = z_1 x_1 + \dots + z_n x_n$ ist der Gradient gleich $(z_1, z_2, \dots, z_n)$ für alle x . Aus 6.3.3 wissen wir (dort in der Dimension 3), daß dieser Vektor senkrecht auf den Geraden steht, die durch $z_1 x_1 + \dots + z_n x_n = c$ definiert sind und diese Geraden sind die Isoquanten von f .

9.3 Extremwertbestimmung**9.3.1 Extremwerte ohne Nebenbedingung**

2 isolierte Maxima, 1 Sattelpunkt



ganze x -Achse besteht aus Minima (nicht isoliert)

Bezeichnung 1:

Sei $D \in \mathbb{R}^n$, $f: D \rightarrow \mathbb{R}$ und sei $a \in D$.

f hat in a ein lokales Maximum (isoliertes lokales Maximum).

$$\Leftrightarrow \begin{cases} \text{Es gibt } \varepsilon > 0, \text{ so daß gilt:} \\ \text{Für alle } x \in D \text{ mit } \|x - a\| < \varepsilon \text{ gilt } f(a) \geq f(x). \\ \text{(Bzw. im Fall des isolierten lokalen Maximums: Es ist} \\ f(a) > f(x) \text{ für alle } x \in D \text{ mit } x \neq a \text{ und } \|x - a\| < \varepsilon) \end{cases}$$

Eine analoge Definition gilt für lokale Minima.

Anschauliche Beispiele oben.

Satz:

Sei $D \in \mathbb{R}^n$ und $f : D \rightarrow \mathbb{R}$ sei 2-mal stetig differenzierbar. Sei $a \in \overset{\circ}{D}$ (dem Innern von D).

(1) Hat f in a ein Extremum, so ist $\text{grad}f(a) = 0$.

(2) Sei $\text{grad}f(a) = 0$. Dann:

- (i) $H_f(a)$ positiv definit $\implies \{ f \text{ hat in } a \text{ ein isoliertes lokales Minimum} \}$
- (ii) $H_f(a)$ negativ definit $\implies \{ f \text{ hat in } a \text{ ein isoliertes lokales Maximum} \}$
- (iii) $H_f(a)$ indefinit $\implies \{ f \text{ hat in } a \text{ keine Extremstelle} \}$

Im Vergleich zur Dimension 1:

Der Gradient übernimmt die Rolle der 1. Ableitung.

Die Hesse-Matrix übernimmt die Rolle der 2. Ableitung.

Dabei übernimmt die positive Definitheit die Rolle des “Positiv-Seins”, die negative Definitheit die Rolle des “Negativ-Seins”.

Die Definitheit der Hesse-Matrix untersucht man mit Hilfe des Hurwitz-Kriteriums in 7.7.4 3.

Musterbeispiele:

(i) $f(x, y) = x^2 + y^2$,

$$\text{grad}f(x, y) = (2x, 2y) = 0 \iff x = y = 0,$$

$$H_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \text{ allgemein und } H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \text{ sind positiv definit.}$$

Also: Isoliertes Minimum (s. Schaubild) in $a = \mathbf{0}$.

(ii) $f(x, y) = -x^2 - y^2$

$$\text{Auch hier } \text{grad}f(x, y) = 0 \iff x = y = 0.$$

Diesmal ist $H_f(0, 0) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$ negativ definit.

Ergebnis: Isoliertes Maximum in $\mathbf{0}$.

(iii) $f(x, y) = x^2 - y^2$, $\text{grad}f(x, y) = 0 \iff x = y = 0$

$$\text{In } (0, 0) : H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} =: A, \text{ ist indefinit, denn:}$$

$$\text{Für } x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ ist } x^t A x = 2 > 0 \text{ und für } y = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ ist } y^t A y = -2$$

Ergebnis: Kein Extremwert.

Bezeichnung 2:

Die Punkte a mit $\text{grad}f(a) = 0$ bezeichnet man auch als **kritische Punkte** von f .

Die kritischen Punkte im $\mathbb{R}^2$ mit indefiniter Hesse-Matrix werden **Sattelpunkte** genannt.

Weniger einfaches Beispiel:

(iv) $f(x, y) = -(x^2 + y^2)^2 + x^2 - y^2$

$$\frac{\partial f}{\partial x}(x, y) = -4x(x^2 + y^2) + 2x = -4x(x^2 + y^2 - \frac{1}{2})$$

$$\frac{\partial f}{\partial y}(x, y) = -4y(x^2 + y^2) - 2y = -4y(x^2 + y^2 + \frac{1}{2})$$

Dann: $\frac{\partial f}{\partial y}(x, y) = 0 \iff y = 0$, weil $(x^2 + y^2 + \frac{1}{2})$ immer > 0 , also $\neq 0$ ist. .

$$\frac{\partial f}{\partial x}(x, 0) = 0 \iff -4x^3 + 2x = 0 \iff x = 0 \text{ oder } x = \pm \frac{\sqrt{2}}{2}$$

Fazit: Die kritischen Punkte sind: $(0, 0)$, $(\frac{\sqrt{2}}{2}, 0)$ und $(-\frac{\sqrt{2}}{2}, 0)$.

$$\text{Hesse-Matrix } H_f(x, y) = \begin{pmatrix} -12x^2 - 4y^2 + 2 & -8xy \\ -8xy & -4x^2 - 12y^2 - 2 \end{pmatrix}$$

Dann:

$$H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \text{ indefinit (s. Beispiel (iii))}$$

$$H_f(\frac{\sqrt{2}}{2}, 0) = \begin{pmatrix} -4 & 0 \\ 0 & -4 \end{pmatrix} = H_f(-\frac{\sqrt{2}}{2}, 0)$$

Also: Ein Sattelpunkt in $(0, 0)$ und zwei relative Maxima in $(\frac{\sqrt{2}}{2}, 0)$ und $(-\frac{\sqrt{2}}{2}, 0)$.

(Das Schaubild sieht ähnlich aus wie das Anfangsbild dieses Abschnitts.)

Zur Beachtung: Die analoge "Warnung" wie in 4.2.3 ist auch hier angebracht:

Erstens: Die kritischen Punkte liefern erst einmal nur Kandidaten für Extremwerte.

Zweitens: Auf dem Rand ∂D von D können Extremstellen vorliegen, ohne daß der Gradient 0 ist, d.h. Randextrema müssen extra untersucht werden.

Es kommt hinzu, daß in der Dimension $n \geq 2$ der Rand eines D recht kompliziert sein kann.

Wir geben noch eine explizite Beschreibung der Definitheitsfrage im Falle einer $(2,2)$ -Matrix, weil dieser einfache Fall häufig vorkommt.

Definitheit der Hesse-Matrix bei zwei Variablen:

Sei $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ symmetrische $2, 2$ -Matrix. Dann:

$$A \text{ positiv definit} \iff a > 0 \text{ und } ad - b^2 > 0$$

$$A \text{ negativ definit} \iff a < 0 \text{ und } ad - b^2 > 0$$

Man kann zeigen: $ad - b^2 < 0 \iff A$ indefinit

$$\text{Für die Hessesche } H_f(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} = f_{xy} & f_{yy} \end{pmatrix}$$

(dabei $f_x := \frac{\partial f}{\partial x}$, $f_y := \frac{\partial f}{\partial y}$) heißt das:

$$H_f(x, y) \text{ positiv definit} \iff f_{xx} > 0 \text{ und } f_{xx} \cdot f_{yy} - f_{xy}^2 > 0$$

$$H_f(x, y) \text{ negativ definit} \iff f_{xx} < 0 \text{ und } f_{xx} \cdot f_{yy} - f_{xy}^2 > 0$$

$$f_{xx} \cdot f_{yy} - f_{xy}^2 < 0 \iff H_f \text{ indefinit.}$$

9.3.2 Extremwerte unter Nebenbedingungen

Situation:

Man hat eine Funktion $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}$, eine weitere Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ und ein $c \in \mathbb{R}$.

Bezeichnung:

Sei $M_c := \{x \in \mathbb{R}^n \mid g(x) = c\}$ die Isoquante von g zum Wert c . Sei $a \in D$. Man sagt:

$$f \text{ hat in } a \text{ ein lokales Maximum unter der Nebenbedingung } g(x) = c \iff \left\{ \begin{array}{l} \text{Es ist } g(a) = c, \text{ d.h. } a \in D \cap M_c \\ \text{und es gibt } \varepsilon > 0, \text{ so da\ss gilt:} \\ \text{Es ist } f(a) \geq f(x) \text{ f\"ur alle} \\ \text{\underline{ } } x \in D \cap M_c \text{ mit } \|x - a\| < \varepsilon \end{array} \right.$$

Also: Es ist $a \in M_c$ und zum Vergleich mit dem Wert von f in a werden nur die Werte von f in denjenigen $x \in D$ herangezogen, die in M_c liegen.

Bemerkung: Das Maximum unter der Nebenbedingung braucht kein Maximum f\"ur das gesamte D zu sein.

Eine analoge Definition hat man f\"ur **lokale Minima**.

Musterbeispiel aus der \u00d6konomie:

Es sei $f : \mathbb{R}_{\geq 0}^n = D \rightarrow \mathbb{R}$ eine "Nutzenfunktion". Das g sei in diesem Fall eine Kostenfunktion $g : D \rightarrow \mathbb{R}, x \mapsto p_1x_1 + p_2x_2 + \dots + p_nx_n$ und die Nebenbedingung ist eine "Budget-Gleichung" $p_1x_1 + p_2x_2 + \dots + p_nx_n = \beta$.

Gesucht: Das G\"uterb\"undel x maximalen Nutzens unter der Nebenbedingung $p_1x_1 + p_2x_2 + \dots + p_nx_n = \beta$, d.h. f\"ur das das Geld noch reicht.

Eine M\u00f6glichkeit zur L\u00f6sung solcher Probleme bietet der folgende Satz:

Satz :

Die Situation sei wie in der Bezeichnung.

Au\u00dferdem: Es sei a aus dem Innern $\overset{\circ}{D}$ von D und sowohl f als auch g seien nach allen Koordinaten stetig partiell differenzierbar.

Schlie\u00dflich: Es sei $\text{grad } g(a) \neq 0$.

Dann gilt:

Hat f in a ein lokales Extremum unter der Nebenbedingung $g(x) = c$, so folgt:

Es gibt ein $\lambda \in \mathbb{R}$ mit

$$\text{grad } f(a) = \lambda \cdot \text{grad } g(a) \quad (\text{d.h. grad } f(a) \text{ und grad } g(a) \text{ zeigen in dieselbe Richtung})$$

d.h. mit

$$\frac{\partial f}{\partial x_i}(a) = \lambda \cdot \frac{\partial g}{\partial x_i}(a) \quad \text{f\"ur alle } i = 1, \dots, n$$

Name: Das λ hei\u00dft ein **Lagrangescher Multiplikator** f\"ur die Extremstelle.

Beachte:

(1) Der Satz gibt wieder nur eine “notwendige Bedingung” für Extremstellen an. D.h. er liefert Kandidaten für Extremstellen. Ob tatsächlich ein Maximum oder Minimum vorliegt und welches von beiden, muß jeweils extra untersucht werden.

Auch die Bedingung “ $\text{grad } g(a) \neq 0$ ” muß sei, damit der Satz funktioniert.

(2) Um die Koordinaten $x_1, \dots, x_n$ der Extremstellen-Kandidaten und, als Hilfwert, das λ zu bestimmen, stehen folgende $n + 1$ Gleichungen zur Verfügung:

$$\frac{\partial f}{\partial x_i}(x) = \lambda \cdot \frac{\partial g}{\partial x_i}(x) \quad \text{für } i = 1, \dots, n \quad \text{und } g(x) = c.$$

Beispiel:

$f: \mathbb{R}_{>0}^2 \rightarrow \mathbb{R}, f(x, y) = x^2 + 4xy$, mit der Nebenbedingung $g(x, y) = x^2 \cdot y = c > 0$.

Kandidaten für Extremwerte sind diejenigen (x, y) , wo

$$(1) \quad (2x + 4y, 4x) = \text{grad } f(x, y) = \lambda \cdot \text{grad } g(x, y) = \lambda \cdot (2xy, x^2)$$

und: $(2) \quad g(x, y) = x^2 \cdot y = c > 0$

Aus (1) :

$$(3) \quad 2x + 4y = \lambda \cdot 2xy$$

$$(4) \quad \begin{aligned} 4x &= \lambda \cdot x^2 &\implies \lambda &= \frac{4}{x} \\ \xRightarrow{\text{aus (3)}} 2x + 4y &= 8y &\implies x &= 2y \end{aligned}$$

Eingesetzt in (2): $4y^3 = c \implies y = \sqrt[3]{\frac{c}{4}} \quad \text{und} \quad x = 2\sqrt[3]{\frac{c}{4}}$

Die Nebenbedingung kann für beliebig großes x erfüllt werden (wenn nur $y = \frac{c}{x^2}$). Also kann $f(x, y)$ beliebig groß werden. Andererseits ist $f(x, y) > 0$, d.h. nach unten beschränkt. Aus all dem läßt sich schließen, daß f in $a = (2\sqrt[3]{\frac{c}{4}}, \sqrt[3]{\frac{c}{4}})$ ein Minimum hat.

Ein weiteres einfaches Beispiel:

Gesucht: Extremwerte von $f(x, y) = x^2 + y^2$

unter der Nebenbedingung $g(x, y) = 2x + y = 4$.

Man hat $\text{grad } f = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \text{grad } g = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Bedingungen:

$$(1) \quad 2x = 2\lambda, \quad (2) \quad 2y = \lambda \quad \text{und} \quad (3) \quad 2x + y = 4$$

Daraus: $\lambda = x \xRightarrow{\text{in (2)}} y = \frac{x}{2} \xRightarrow{\text{in (3)}} \frac{5}{2}x = 4.$

Also: $x = \frac{8}{5}, y = \frac{4}{5}$. Somit ist $a = \begin{pmatrix} \frac{8}{5} \\ \frac{4}{5} \end{pmatrix}$ das einzige mögliche Extremum.

Da $x^2 + y^2$ für x, y mit großem Betrag beliebig groß wird, liegt ein Minimum vor.

Geometrische Deutung des letzten Beispiels:

Für $X = (x, y)$ ist $x^2 + y^2 = \|X\|^2 \stackrel{!}{=} \text{Quadrat des Abstands von } X \text{ vom Nullpunkt.}$

Also: $a = \begin{pmatrix} \frac{8}{5} \\ \frac{4}{5} \end{pmatrix}$ ist der Punkt auf der Geraden $2x + y = 4$, welcher von 0 den kleinsten Abstand hat!

Alternative Lösungsmethode mittels Substitution:

Unter Umständen kann man die Nebenbedingung $g(x) = c$ nach einer der Variablen, etwa nach x_n , „auflösen“, d.h. für die Punkte $x \in M_c$ kann man x_n , zumindest in einer Umgebung von x , als Funktion $x_n = \varphi(x_1, \dots, x_{n-1})$ darstellen.

(In unseren Beispielen z.B. geht das: Im ersten Beispiel ist $y = \frac{c}{x^2}$ und im zweiten Beispiel ist $y = 4 - 2x$.)

In diesem Fall kann man die Funktion $F(x_1, \dots, x_{n-1}) := f(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1}))$ betrachten. In ihr ist die Nebenbedingung „integriert“. Das Problem, Extremstellen für f unter der Nebenbedingung zu finden, ist transformiert in das Problem, Extremstellen von F ohne Nebenbedingung und bei nur noch $n - 1$ Variablen zu finden. Diese Problem löst man dann mit den Methoden aus 9.3.1 (oder, wenn $n = 2$ also $n - 1 = 1$ ist, mit den Methoden der (Schul-)Kurvendiskussion aus 4.2.3).

Etwa bei unserem ersten Beispiel: $y = \frac{c}{x^2}$ in f eingesetzt ergibt $F(x) = x^2 + \frac{4c}{x}$. Die Ableitung ist $F'(x) = 2x - \frac{4c}{x^2}$. Dies nullgesetzt ergibt $0 = 2x - \frac{4c}{x^2} \Leftrightarrow x^3 = 2c \Leftrightarrow x = \sqrt[3]{2c}$ ($= 2\sqrt[3]{\frac{c}{4}}$ wie oben errechnet; die letzte Gleichung bitte als kleine Übung in Wurzelrechnen nachprüfen).

Die zweite Ableitung $F''(x) = 2 + \frac{8c}{x^3}$ ist gleich $6 > 0$ für $x = \sqrt[3]{2c}$. Also hat man ein Minimum. Den y -Wert im Minimum erhält man durch Einsetzen von $x = \sqrt[3]{2c}$ in die Substitutionsgleichung $y = \frac{c}{x^2}$. Es ist $y = \sqrt[3]{\frac{c}{4}}$ wie auch oben errechnet.

Als Übung löse man das zweite Beispiel mit der Substitutionsmethode.

Alternative Formulierungen zum Satz

Viele Autoren schreiben die Nebenbedingung nur in der Form $g(x) = 0$. Das ist keine Einschränkung der Allgemeinheit. Denn man kann $g(x) = c$ als $g(x) - c = 0$ und dies als $\tilde{g}(x) = 0$ schreiben mit $\tilde{g} := g - c$ als neuer Nebenbedingungsfunktion.

Die Gradienten von g und $\tilde{g}$ sind dieselben.

Gegeben sei f und die Nebenbedingung $g(x) = 0$. Man kann die Funktion

$$f(x) + \lambda g(x) =: L(x, \lambda) = L(x_1, \dots, x_n, \lambda)$$

in den $n + 1$ Variablen $x_1, \dots, x_n$ und λ betrachten und registriert:

$$\text{grad } L(x, \lambda) = \mathbf{0} \iff \left\{ \begin{array}{ccc} D_1 f(x) + \lambda D_1 g(x) & = & 0 \\ \vdots & \ddots & \vdots \\ D_n f(x) + \lambda D_n g(x) & = & 0 \\ g(x) & = & 0 \end{array} \right\} \iff \left\{ \begin{array}{l} \text{grad } f(x) = -\lambda \text{ grad } g(x) \\ \text{und} \\ g(x) = 0 \end{array} \right.$$

Das sind (mit $-\lambda$ statt λ , was irrelevant ist) dieselben notwendigen Bedingungen wie sie im Satz für eine Extremstelle verlangt sind.

Der Satz lautet in dieser Formulierung also:

In einer Extremstelle gilt $\text{grad } L(x, \lambda) = \mathbf{0}$.

9.4 Die (totale) Ableitung

9.4.1 Die Definition der Ableitung

Erinnere an den Abschnitt 4.1.2 “Differenzierbarkeit und lineare Approximation”. Dort hatten wir die Differenzierbarkeit folgendermaßen charakterisiert:

$$\left. \begin{array}{l} f \text{ ist differenzierbar in } a \\ \text{mit Ableitung } \alpha \end{array} \right\} \iff \left\{ \begin{array}{l} \text{Es ist } f(x) = f(a) + \alpha \cdot (x - a) + r(x) \text{ mit} \\ \text{einem “Rest” } r(x), \text{ für den } \lim_{x \rightarrow x_0} \frac{r(x)}{x - x_0} = 0. \end{array} \right.$$

In Worten:

Die lineare Funktion $t(x) := f(x_0) + \alpha(x - x_0)$ approximiert $f(x)$ in der Nähe von x_0 besonders gut, und zwar so, daß für $x \rightarrow x_0$ nicht nur der Rest $r(x)$, sondern sogar $\frac{r(x)}{x - x_0}$ gegen 0 strebt.

Die lineare Funktion $t : \mathbb{R} \rightarrow \mathbb{R}, t(x) = f(x_0) + \alpha \cdot (x - x_0)$ war dann die Gleichung der Tangente an f bei x_0 .

Verallgemeinerung auf unsere Situation

Es stellt sich heraus, daß die zitierte Charakterisierung den besten Ansatz liefert, den Begriff der Ableitung einer Funktion auf unsere Funktionen mit mehreren Variablen zu verallgemeinern.

Wir machen dies gleich im allgemeinen Fall von Funktionen des Typs $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m$.

Definition (Differenzierbarkeit und Ableitung):

Sei $D \subseteq \mathbb{R}^n \xrightarrow{f} \mathbb{R}^m$ eine Funktion und sei $a \in \overset{\circ}{D}$. Dann:

$$\left. \begin{array}{l} f \text{ heißt} \\ \text{differenzierbar in } a \end{array} \right\} \iff \left\{ \begin{array}{l} \text{Es gibt eine lineare Abbildung } F_a : \mathbb{R}^n \rightarrow \mathbb{R}^m, \\ \text{so daß gilt: Ist } r : D \rightarrow \mathbb{R}^m \text{ die “Restfunktion”} \\ \text{aus der Gleichung } f(x) = f(a) + F_a(x - a) + r(x), \\ \text{so ist } \lim_{x \rightarrow a} \frac{\|r(x)\|}{\|x - a\|} = 0 \end{array} \right.$$

Gilt die rechte Seite, so heißt F_a die **Ableitung von f in a** .

(Auch die Bezeichnung **totales Differential** von f in a ist für F_a gebräuchlich.)

Schreibweisen: $F_a =: df_a =: Df_a$.

Vergleiche: Das ist formal die gleiche Definition wie bei Funktionen in einer Variablen. Anstelle der simpelsten linearen Funktion $\mathbb{R} \ni x \mapsto \alpha x$ hat man hier das F_a und weil man den Vektor $r(x)$ nicht durch den Vektor $x - a$ dividieren kann, bildet man erst die Norm von beiden. Das liefert reelle Zahlen und die kann man dividieren.

Der Limes in der Definition bedeutet folgendes:

Die Funktion $D \ni x \mapsto \begin{cases} \frac{\|r(x)\|}{\|x - a\|} & x \neq a \\ 0 & x = a \end{cases}$ ist stetig in a .

Bemerkung

Man kann zeigen, daß ein F_a wie auf der rechten Seite der Definition eindeutig ist, sofern es existiert. Daher ist die Definition von F_a als der Ableitung sinnvoll.

Zum Einprägen: Die Ableitung von f in a ist eine lineare Abbildung vom $\mathbb{R}^n$ in den $\mathbb{R}^m$!

Interpretation als lineare Approximation

Wie bei der Differenzierbarkeit bei Funktionen einer Variablen läßt sich die Differenzierbarkeit einer Funktion als Approximation durch eine lineare Funktion deuten:

Die "lineare" Funktion $t(x) := f(a) + F_a(x - a)$ approximiert f in der Nähe von a so gut, daß nicht nur der Rest $r(x)$, sondern sogar $\frac{\|r(x)\|}{\|x - a\|}$ gegen Null geht für $x \rightarrow a$. Zur anschaulichen Bedeutung von $t(x)$ siehe 9.4.4.

Die Existenz der Ableitung ist eine starke Eigenschaft. Wie in einer Variablen gilt

Satz (Stetigkeit differenzierbarer Funktionen)

$$f \text{ differenzierbar in } a \implies f \text{ stetig in } a.$$

Der Beweis ist nicht schwer.

9.4.2 Die Jakobi-Matrix

Bezeichnungswahl: In den nächsten Abschnitten werden wir die Tupel aus den Tupelräumen $\mathbb{R}^n$, wie in der Linearen Algebra, wieder als 1-spaltige Matrizen schreiben.

Sei $D \subseteq \mathbb{R}^n \xrightarrow{f} \mathbb{R}^m$ eine Funktion wie im Abschnitt zuvor. Man hat also:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto f(x) =: \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

Die i -te Koordinate von $f(x)$, haben wir dabei $f_i(x)$ genannt, $i = 1, \dots, m$.

Bemerkung: Dadurch werden m reellwertige Funktionen

$f_i : D \longrightarrow \mathbb{R}, x \longmapsto f_i(x) := i\text{-te Koordinate von } f(x), i = 1, \dots, m,$
definiert.

Bezeichnung:

Diese f_i heißen die **Koordinatenfunktionen** von f .

Man kann dann für alle $i = 1, \dots, m$ und alle $j = 1, \dots, n$ die j -te partielle Ableitung von f_i betrachten:

$$\frac{\partial f_i}{\partial x_j}(x) =: D_j f_i(x), \quad i = 1, \dots, m, j = 1, \dots, n$$

↙ wir übernehmen hier diese Schreibweise

Definition: Betrachte $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m, x \longmapsto \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$.

Sei dann x aus dem Innern von D
und alle f_i seien partiell differenzierbar nach allen Variablen.

Die Matrix

$$J_f(x) = (D_j f_i(x))_{\substack{i=1,\dots,m \\ j=1,\dots,n}} = \begin{pmatrix} D_1 f_1(x) & D_2 f_1(x) & \dots & D_n f_1(x) \\ D_1 f_2(x) & D_2 f_2(x) & \dots & D_n f_2(x) \\ \vdots & \vdots & & \vdots \\ D_1 f_m(x) & D_2 f_m(x) & \dots & D_n f_m(x) \end{pmatrix}$$

heißt die Jakobi-Matrix von f in x .

Spezialfälle:

1 $m = 1$: Das heißt, f ist reellwertig (vom Typ (1)). Dann:

$$J_f(x) = (D_1 f(x) \quad D_2 f(x) \quad \dots \quad D_n f(x)) \stackrel{!}{=} (\text{grad } f(x))^t.$$

Also: Die Jakobi-Matrix einer reellwertigen Funktion ist die transponierte des Gradienten.

Anmerkung: Im Fall reellwertiger Funktionen sind also die Jakobi-Matrix und der Gradient im wesentlichen das gleiche. Der (feine) Unterschied in der Schreibweise wird z.B. ermöglichen, die Kettenregel besonders einfach und prägnant zu formulieren.

2 $n = 1$, d.h. Typ (2) : Der Fall der "Kurven".

Da wir dann nur ein Variable im Spiel haben, sind die Koordinatenfunktionen gewöhnliche reelle Funktionen. Es gibt nur eine Variable und deshalb nur eine "partielle" Ableitung, und das ist die gewöhnliche Ableitung aus dem Paragraphen 4. Also:

$$J_f(x) = \begin{pmatrix} \frac{d}{dx} f_1(x) \\ \vdots \\ \frac{d}{dx} f_m(x) \end{pmatrix} = \begin{pmatrix} f'_1(x) \\ \vdots \\ f'_m(x) \end{pmatrix} \quad (x \in \mathbb{R}) =: f'(x)$$

Bezeichnung:

Das m -Tupel $J_f(x) = f'(x)$ nennt man in diesem Fall den **Tangential-** oder **Geschwindigkeitsvektor** der "Kurve" f .

Im Punkt $f(x)$ angesetzt, ist der Vektor $f'(x)$ tangential zur Kurve f .

3 Eine Funktion vom Typ (3): Der Übergang zu den Polarkoordinaten (s. 9.1.1)

$$f: \mathbb{R}^2 \setminus \{0\} \longrightarrow \mathbb{R}^2, f(r, \varphi) = \begin{pmatrix} r \cdot \cos \varphi \\ r \cdot \sin \varphi \end{pmatrix}$$

In diesem Fall:

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

9.4.3 Jakobi-Matrix $\equiv$ Ableitung

In 7.1.2 haben wir den engen Zusammenhang zwischen Matizen A aus $\mathbb{R}^{m \times n}$ und lineare Abbildungen vom $\mathbb{R}^n$ in den $\mathbb{R}^m$ kennengelernt.

Zu Funktionen $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m$ haben wir bisher lineare Funktionen als Ableitung und Matrizen als Jakobi-Matrix eingeführt. Es stellt sich nun heraus, daß beide für alle gebräuchlichen Funktionen im Sinne der linearen Algebra dasselbe sind:

Satz

Gegeben seien die Funktion $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m$ und ei $a \in \overset{\circ}{D}$, dem Innern von D .

- (1) Ist f in a differenzierbar mit Ableitung Df_a , so sind alle Koordinatenfunktionen f_i , $i = 1, \dots, m$ von f nach allen Koordinaten partiell differenzierbar. Es existiert also die Jacobi-Matrix $J_f(a)$ und sie ist die Matrix der Ableitung, d.h. für alle $x \in \mathbb{R}^n$, ist

$$Df_a(x) = J_f(a) \cdot x$$

- (2) Ist f stetig differenzierbar in $\overset{\circ}{D}$, d.h. existieren dort alle partiellen Ableitungen aller Koordinatenfunktionen f_i und sind sie stetig, so ist f differenzierbar in (jedem) a und es gilt erneut

$$Df_a(x) = J_f(a) \cdot x \quad \text{für } x \in \mathbb{R}^n.$$

Anmerkung Die Zusammenhänge bei allgemeineren Voraussetzungen sind recht komplex. Die Existenz der Ableitung ist die stärkere Eigenschaft. Es gibt z.B. Funktionen, bei denen alle partiellen Ableitungen, also die Jakobi-Matrizen, in einem a existieren, wo aber f in a nicht differenzierbar ist.

9.4.4 Die Ableitung bei reellwertigen Funktionen

Situation

Die Funktion $f: \mathbb{R}^n \supseteq D \longrightarrow \mathbb{R}$ sei gegeben und differenzierbar in $a \in \overset{\circ}{D}$. Für $x \in \mathbb{R}^n$ betrachte

$$\begin{aligned} t: \mathbb{R}^n \longrightarrow \mathbb{R}, t(x) &:= f(a) + Df_a(x)(x-a) = f(a) + J_f(a) \cdot (x-a) \\ &\stackrel{!}{=} f(a) + \langle \text{grad } f(a), x-a \rangle. \end{aligned}$$

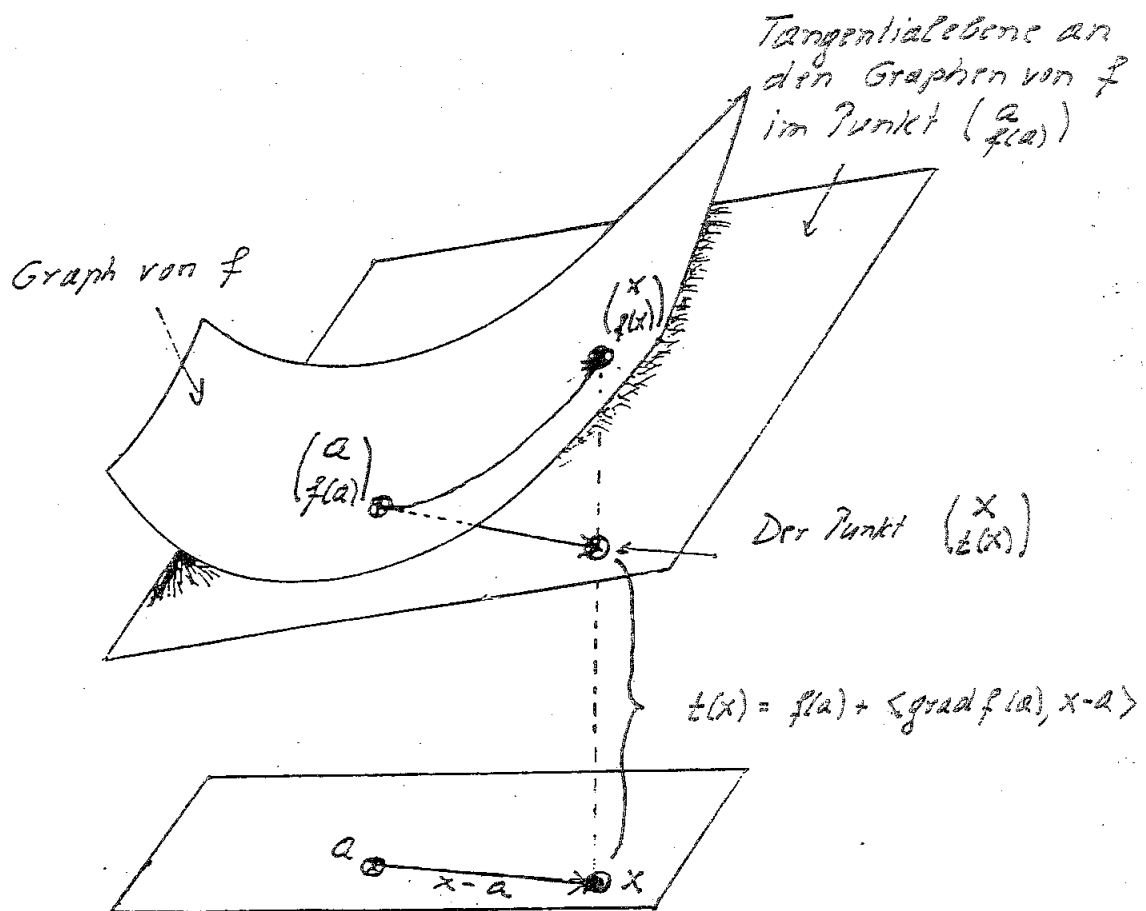
Bemerkung 1:

Das Schaubild von t ist durch die Gleichung $x_{n+1} = t(x_1, \dots, x_n)$, gegeben und beschreibt eine Hyperebene (s. Def.2 in 6.3.2) im $\mathbb{R}^{n+1}$, für $n = 2$ also eine Ebene im $\mathbb{R}^3$.

Bezeichnung

Dieses Schaubild heißt die **Tangential-Hyperebene** (**Tangentialebene** für $n = 2$) von f in a .

Vorstellung: Es ist die (Hyper-)Ebene durch den Punkt $(a, f(a)) \in \mathbb{R}^{n+1}$, an die sich das Schaubild der Funktion "anschmiegt". Siehe das folgende Bild.



Bemerkung 2:

Wie in der Dimension 1 werden die Punkte auf der Tangentialhyperebene als Näherungswerte für die $f(x)$, x nahe bei a , angesehen:

$$f(x) \approx f(a) + J_f(a) \cdot (x - a) = \sum_{i=1}^n D_i f(a) (x_i - a_i) = \langle \text{grad } f(a), x - a \rangle$$

oder für $\Delta f := f(x) - f(a)$ und $\Delta x := x - a$:

$$\Delta f \approx J_f(a) \cdot \Delta x = \langle \text{grad } f(a), \Delta x \rangle$$

↘ ungefähr, "in 1. Näherung"
(das alles für x nahe bei a)

9.4.5 Kettenregel

Bemerkung:

Wie in 3.2.2 für die gewöhnlichen reellen Funktionen kann man auch unsere allgemeineren Funktionen hintereinanderschalten (komponieren):

Gegeben seien die Funktionen $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m$ und $\mathbb{R}^m \supseteq D' \xrightarrow{g} \mathbb{R}^k$ und es sei $f(D) \subseteq D'$. Dann existiert die **Komposition** von f und g , das ist die Funktion

$$g \circ f : D \longrightarrow \mathbb{R}^k, a \longmapsto g \circ f(x) := g(f(x)).$$

In 4.1.5 haben wir mittels “Kettenregel” die Ableitung der Komposition bei gewöhnlichen (d.h. es ist $n = m = 1$) differenzierbaren Funktionen berechnet. In diesem Fall gilt für $x \in D$:

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x).$$

In Worten: Die Ableitung von $g \circ f$ in x ist gleich dem Produkt der Ableitung von g in $f(x)$ mit der Ableitung von f in x .

Es stellt sich heraus, daß wir diese Formulierung beinahe wörtlich auch im Fall mehrerer Variablen übernehmen können:

Satz (Kettenregel):

Gegeben seien $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}^m$, $\mathbb{R}^m \supseteq D' \xrightarrow{g} \mathbb{R}^k$ und es sei $f(D) \subseteq D'$.

Es seien $a \in \overset{\circ}{D}$ und $f(a) \in \overset{\circ}{D'}$. Dann:

Ist f differenzierbar in a und g differenzierbar in $f(a)$, so ist $g \circ f$ differenzierbar in a und für die Ableitung gilt

$$D(g \circ f)_a = Dg_{f(a)} \circ Df_a$$

In Worten: Die Ableitung der Komposition differenzierbarer Funktionen ist die Komposition der entsprechenden Ableitungen.

Für die Jakobi-Matrizen gilt die äquivalente Aussage:

$$J_{g \circ f}(a) = J_g(f(a)) \cdot J_f(a) \quad (\text{Matrizenprodukt})$$

Also: Die Jakobi-Matrix der Komposition differenzierbarer Funktionen ist das Produkt der entsprechenden Jakobi-Matrizen.

Zum Beweis benutzt man die Approximationsaussagen der Differenzierbarkeitsdefinition und muß einige Tatsachen über Grenzwerte bei Funktionen mehrerer Variablen sorgfältig beachten.

Anmerkung Man erkennt, wie perfekt diese Verallgemeinerung ist. In gewissem Sinne gewinnt sogar das frühere Ergebnis bei einer Variablen durch die verallgemeinerte Fassung an Klarheit und Kontur.

9.4.6 Spezialfälle der Kettenregel

(Wir geben die Formulierungen für die Jakobi-Matrizen)

$$\boxed{1} \quad \begin{array}{ccc} \mathbb{R} & \xrightarrow{f} & \mathbb{R}^m & \xrightarrow{g} & \mathbb{R} \\ t & \mapsto & f(t) & \mapsto & g(f(t)) \end{array}, \quad \text{d.h. es ist } n = k = 1.$$

Es gilt in diesem Fall

$$\frac{d}{dt}(g \circ f)(t) = (g \circ f)'(t) = J_g(f(t)) \cdot J_f(t) = (D_1 g(f(t)) \dots D_m g(f(t))) \cdot \begin{pmatrix} \frac{df_1}{dt}(t) \\ \vdots \\ \frac{df_m}{dt}(t) \end{pmatrix}.$$

Das Ergebnis als Funktion geschrieben ist also:

$$(*) \quad \frac{d}{dt}(g \circ f) = \sum_{j=1}^m D_j g \cdot \frac{df_j}{dt} = \langle \text{grad } g, f' \rangle$$

Als Anwendung (Beweis der Tatsache aus 9.3.2):

Wir benutzen die jetzigen Bezeichnungen:

Es sei $\mathbb{R}^m \supseteq D' \xrightarrow{g} \mathbb{R}^k$ die Funktion, deren Richtungsableitungen wir berechnen wollen, und es sei $a \in D'$. Sei e eine Richtung im $\mathbb{R}^m$. Sei $] - \delta, \delta[=: D$ mit $\delta > 0$ ein Intervall in $\mathbb{R}$, so daß $a + h e \in D'$ ist für alle $h \in D$. Wir betrachten dann die Abbildung $f : D \rightarrow D' \subseteq \mathbb{R}^m$, $h \mapsto a + h e$, und schließlich die Komposition $g \circ f$.

Man stellt fest:

- (i) Es ist $g \circ f(h) = g(a + h e)$ und $g \circ f(0) = g(a)$. Daraus
- (ii) Es ist $\lim_{h \rightarrow 0} \frac{g(a + h \cdot e) - g(a)}{h} = (g \circ f)'(0)$.
- (iii) Es ist $f'(0) \stackrel{!}{=} e$.

Aus all dem und aus (*) folgt dann:

$$\frac{\partial g}{\partial e}(a) = \langle \text{grad } g(a), e \rangle$$

Das ist aber, modulo der Bezeichnung, die Aussage der Tatsache in 9.3.2.

$$\boxed{2} \quad \text{Kompositionen des Typs } \mathbb{R}^n \xrightarrow{f} \mathbb{R}^n \xrightarrow{g} \mathbb{R}.$$

Die Formel ist in diesem Fall:

$$(**) \quad J_{g \circ f}(x) = \begin{pmatrix} D_1 g(f(x)) & D_2 g(f(x)) & \dots & D_n g(f(x)) \end{pmatrix} \cdot \begin{pmatrix} D_1 f_1(x) & \dots & D_n f_1(x) \\ D_1 f_2(x) & \dots & D_n f_2(x) \\ \vdots & \vdots & \vdots \\ D_1 f_n(x) & \dots & D_n f_n(x) \end{pmatrix}$$

Das liefert für die einzelnen partiellen Ableitungen:

$$D_i(g \circ f)(x) = \sum_{j=1}^n D_j g(f(x)) \cdot D_i f_j(x).$$

Beispiel: $f(r, \varphi) = \begin{pmatrix} r \cdot \cos \varphi \\ r \cdot \sin \varphi \end{pmatrix}$ (Polarkoordinaten)

Hier

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

Schreibt man $\begin{cases} x_1 := r \cdot \cos \varphi & \text{für } f_1(r, \varphi) \\ x_2 := r \cdot \sin \varphi & \text{für } f_2(r, \varphi) \end{cases}$, so erhält man

$$\frac{\partial}{\partial r} g(x_1, x_2) = \frac{\partial g}{\partial x_1}(x) \cdot \cos \varphi + \frac{\partial g}{\partial x_2}(x) \cdot \sin \varphi$$

$$\frac{\partial}{\partial \varphi} g(x_1, x_2) = -\frac{\partial g}{\partial x_1} \cdot r \cdot \sin \varphi + \frac{\partial g}{\partial x_2} \cdot r \cdot \cos \varphi$$

9.5 Einige Anwendungen

9.5.1 Partielle Wachstumsraten und Elastizitäten

Bezeichnung:

Gegeben sei $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}$. Es sei f differenzierbar in x und $f(x) \neq 0$ für alle $x \in D$.

$\frac{\partial f}{\partial x_i}(x) \cdot \frac{1}{f(x)} =: r_{f,i}(x)$ heißt **partielle Wachstumsrate** nach der i -ten Variablen.

$\frac{\partial f}{\partial x_i}(x) \cdot \frac{x_i}{f(x)} =: \varepsilon_{f,i}(x)$ heißt **partielle Elastizität** nach der i -ten Variablen.

(Die partielle Ableitung $\frac{\partial f}{\partial x_i}(x)$ selbst nennt man manchmal auch **Änderungsrate** nach der i -ten Variablen.)

Beispiel: $f(x, y) = x^\alpha \cdot y^\beta$. Dann:

$$\varepsilon_{f,1}(x) = \alpha \cdot x^{\alpha-1} \cdot y^\beta \cdot \frac{x}{x^\alpha \cdot y^\beta} = \alpha \quad \text{und analog} \quad \varepsilon_{f,2}(x) = \beta.$$

Diverse Bezeichnungsweisen:

Für konkrete und spezielle Funktionen hat man oft auch spezielle, in der jeweiligen Situation aber meist unmißverständliche Namen für diese Elastizitäten. Zum Beispiel:

Sei $x(A, K)$ eine Produktionsfunktion, abhängig von den "Faktoren" Arbeit (A) und Kapital (K). Man gibt dann den $\varepsilon_{x,i}(x)$ auch die entsprechenden konkreten Namen:

$\varepsilon_{x,1}(x) = \frac{\partial x}{\partial A}(A, B) \cdot \frac{A}{x}$ heißt Elastizität der Produktion nach der Arbeit und

$\varepsilon_{x,2}(x) = \frac{\partial x}{\partial K}(A, B) \cdot \frac{K}{x}$ heißt Elastizität der Produktion nach dem Kapital.

Die Interpretation: Die Interpretation ist entsprechend der für einfache Elastizitäten, s. 4.1.4 (und Vorlesung).

9.5.2 Skalenelastizität

So heißt die Elastizität bei der sogenannten “proportionalen Faktorvariation”. Das ist, ausgehend von einem $a \neq 0$ im Definitionsbereich von f , eine Änderung der Variablen (= “Faktoren”) auf der Geraden $\mathbb{R} \cdot a$.

(Im Kontrast dazu steht die sogenannte “partielle Faktorvariation”, wo man die Variablen in Richtung der Parallelen $a + \mathbb{R} \cdot e_i$ zur x_i -Achse verändert und die zur partiellen Ableitung und zur partiellen Elastizität führt.)

Genauer (wir beschränken uns auf die gebräuchliche Situation):

Gegeben: $f : \mathbb{R}_{>0}^n \rightarrow \mathbb{R}$, und $a \in D$. Sei $\delta > 0$ so, daß $\lambda \cdot a \in D$ ist für alle λ mit $1 - \delta < \lambda < 1 + \delta$. Man kann dann betrachten

$$\varphi :]1 - \delta, 1 + \delta[\rightarrow \mathbb{R}, \lambda \mapsto \varphi(\lambda) := f(\lambda \cdot a)$$

Definition (Skalenelastizität)

Die **Skalenelastizität** $\varepsilon_{f,\lambda}(\lambda \cdot a)$ von f in $\lambda \cdot a$ ist definiert als die (gewöhnliche) Elastizität von φ in λ .

Für $\lambda = 1$ erhält man die **Skalenelastizität** $\varepsilon_{f,\lambda}(a)$ von f in a .

Weil $\varepsilon_{\varphi}(\lambda) = \frac{\varphi'(\lambda)}{\varphi(\lambda)} \cdot \lambda \underset{\varphi(\lambda)=f(\lambda a)}{=} \frac{\varphi'(\lambda)}{f(\lambda \cdot a)} \cdot \lambda$ gilt also in Formeln:

$$\varepsilon_{f,\lambda}(\lambda \cdot a) \underset{(*)}{=} \frac{\varphi'(\lambda)}{f(\lambda \cdot a)} \cdot \lambda \quad \text{und}$$

$$\varepsilon_{f,\lambda}(a) \underset{(*)}{=} \frac{\varphi'(1)}{f(a)}.$$

Anwendung der Kettenregel bringt uns detailliertere Informationen:

Betrachte die Abbildung

$$\sigma :]1 - \delta, 1 + \delta[\rightarrow \mathbb{R}, \lambda \mapsto \lambda \cdot a.$$

Dann gilt:

$$\varphi \text{ ist die Komposition } \varphi \stackrel{!}{=} f \circ \sigma \quad \text{und} \quad \frac{d\sigma}{d\lambda}(\lambda) = \sigma'(\lambda) \stackrel{!}{=} a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \text{für alle } \lambda \in I.$$

Die Kettenregel in der Form des Spezialfalls [1] liefert daher (mit den jetzigen Bezeichnungen) :

$$(f \circ \sigma)'(\lambda) \underset{(**)}{=} \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\lambda \cdot a) \cdot a_i$$

Die Gleichung (**) in die Gleichung (*) eingesetzt ergibt:

$$\varepsilon_{f,\lambda}(\lambda \cdot a) = \sum_{i=1}^n \underbrace{\frac{\frac{\partial f}{\partial x_i}(\lambda \cdot a)}{f(\lambda \cdot a)} \cdot \lambda \cdot a_i}_{\varepsilon_{f,i}(\lambda \cdot a)} = \sum_{i=1}^n \varepsilon_{f,i}(\lambda \cdot a).$$

Als Ergebnis erhalten wir den

Satz (Wicksell-Johnson-Theorem):

Die Skalenelastizität ist die Summe der partiellen Elastizitäten.

Oder als Formel: Für alle $a \in D$ ist

$$\varepsilon_{f,\lambda}(a) = \sum_{i=1}^n \varepsilon_{f,i}(a).$$

Bemerkung

Das Wicksell-Johnson-Theorem erlaubt es, die Skalenelastizität direkt für alle a zu definieren ohne Rekurs auf die Abbildung $\lambda \mapsto \lambda \cdot a$, nämlich als Summe der partiellen Elastizitäten.

Beispiel:

Betrachte die Funktion $f(x, y) = x^\alpha y^\beta$ mit $\alpha, \beta > 0$. Dann ist (s. dieses Beispiel in 9.5.1): $\varepsilon_{f,\lambda}(a) = \alpha + \beta$ für alle a , insbesondere ist $\varepsilon_{f,\lambda}(a)$ konstant in a .

Bezeichnung

Ist f eine (Produktions-)Funktion, so spricht man von

konstanten Skalenerträgen	(constant returns to scale)	wenn	$\varepsilon_{f,\lambda} = 1$ in ganz D ,
wachsenden Skalenerträgen	(increasing returns to scale)	wenn	$\varepsilon_{f,\lambda} > 1$ in ganz D ,
fallenden Skalenerträgen	(decreasing returns to scale)	wenn	$\varepsilon_{f,\lambda} < 1$ in ganz D .

(Interpretation in der Mikroökonomie.)

9.5.3 Homogene Funktionen

Definition: (homogene Funktionen):

Sei $\mathbb{R}^n \supseteq D \xrightarrow{f} \mathbb{R}$ und sei $r \in \mathbb{R}$.

f heißt homogen vom Grade r $\iff \left\{ \begin{array}{l} \text{für alle } x \in D \text{ und alle } \lambda \in \mathbb{R} \text{ mit} \\ \lambda x \in D \text{ gilt: } f(\lambda x) = \lambda^r f(x). \end{array} \right.$

Beispiel: $f(x, y) = x^\alpha \cdot y^\beta$ mit $\alpha + \beta = r$ ist homogen vom Grade r .

Denn:

$$f(\lambda x, \lambda y) = \lambda^\alpha x^\alpha \cdot \lambda^\beta y^\beta = \lambda^{\alpha+\beta} \cdot x^\alpha y^\beta = \lambda^r f(x).$$

Die hier und auch schon früher als Beispiel betrachteten Potenzfunktionen sind wichtig zur Demonstration grundlegender Zusammenhänge ökonomischer Theorien und haben einen klassischen Namen:

Bezeichnung (Cobb–Douglas-Funktionen):

Sei $f: \mathbb{R}_{>0}^n \rightarrow \mathbb{R}$, $f(x) = f(x_1, \dots, x_n) := c \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \dots \cdot x_n^{\alpha_n}$, mit $\alpha_1 + \alpha_2 + \dots + \alpha_n = r$ und einer Konstanten $c > 0$.

Dann heißt f eine **(allgemeine) Wicksell-Cobb-Douglas-Funktion**.

Mit **Cobb-Douglas-Funktion** schlechthin bezeichnet man meist die mit $r = 1$.

Wicksell-Cobb-Douglas-Funktionen sind homogen vom Grade r (Beweis wie im Beispiel).

Die Eulerschen Homogenitätsrelationen:

Vorausgesetzt: $f : D \rightarrow \mathbb{R}$ sei homogen vom Grade r und es sei $x \in D$.

Betrachte ein Intervall I in $\mathbb{R}$ mit $1 \in I$ und $\lambda \cdot x \in D$ für alle $\lambda \in I$ und die zusammengesetzte Abbildung φ definiert durch $\varphi := f \circ \sigma$, wo

$$\begin{array}{ccccc} \varphi : I & \xrightarrow{\sigma} & D & \xrightarrow{f} & \mathbb{R} \\ \lambda & \mapsto & \lambda \cdot x & \mapsto & f(\lambda \cdot x) \end{array}, \text{ also } \varphi(\lambda) := f(\lambda \cdot x)$$

Es gilt dann:

$$f(\lambda \cdot x) = \lambda^r \cdot f(x)$$

Beide Seiten nach λ abgeleitet, links nach der Kettenregel, ergibt:

$$(*) \quad \sum_{i=1}^n D_i f(\lambda \cdot x) \cdot x_i = \frac{df}{d\lambda}(\lambda \cdot x) = \frac{d}{d\lambda} \lambda^r \cdot f(x) = r \cdot \lambda^{r-1} \cdot f(x)$$

Für $\lambda = 1$ liefert das den:

Satz

Unter den bisherigen Voraussetzungen gilt:

$$r \cdot f(x) = \sum_{i=1}^n D_i f(x) \cdot x_i \quad (\text{Eulersche Homogenitätsrelation})$$

Dividiert man noch durch $f(x)$ (für $f(x) \neq 0$), so erhält man

$$r = \sum_{i=1}^n \varepsilon_{f,i}(x) \quad (\text{Eulersche Homogenitätsrelation für die partielle Elastizität}).$$

9.5.4 Herleitung ökonomischer Gesetzmäßigkeiten mittels Lagrange-Theorie

Gegeben: Güterbündel $x = (x_1, \dots, x_n) \in \mathbb{R}^n \geq 0$

Preisvektor $p = (p_1, \dots, p_n)$ mit $p_i > 0$, für alle i

(p_i = Preis pro Einheit von x_i)

Nutzenfunktion: $f : \mathbb{R}^n \rightarrow \mathbb{R}$

Budgetbeschränkung: $\sum x_i p_i = c = g(x)$

Gesucht: Maximum von f unter der Budgetbeschränkung als Nebenbedingung.

Die Theorie der Extremstellen unter Nebenbedingungen in 9.3.2 legt die Vorgehensweise nahe:

Man bestimme die x mit

$$\text{grad } f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} = \lambda \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} = \lambda \cdot \text{grad } g(x).$$

D.h. mit

$$\frac{\partial f}{\partial x_i}(x) = \lambda \cdot p_i \text{ bzw. } \lambda = \frac{\frac{\partial f}{\partial x_i}(x)}{p_i}, \quad i = 1, \dots, n.$$

Ergebnis: Die Quotienten rechts in der letzten Gleichung müssen alle gleich sein.

Wir haben folgendes “Gesetz” hergeleitet:

Satz (sogenanntes 2. Gossensches Gesetz):

Mit den bisherigen Bezeichnungen und Voraussetzungen gilt:

In einem x_m mit maximalem Nutzen ist das Verhältnis

$$\frac{\frac{\partial f}{\partial x_i}(x_m)}{p_i} = \frac{\text{Grenznutzen nach } x_i}{\text{Preis pro Einheit von } x_i} \quad \text{für alle } i \text{ dasselbe}. \quad \text{---}$$

Oder in anderer Form: Es ist

$$\frac{\frac{\partial f}{\partial x_i}(x_m)}{p_i} = \frac{\frac{\partial f}{\partial x_j}(x_m)}{p_j} \implies \text{für alle } 1 \leq i, j \leq n.$$

Bringt man in diesen Gleichungen die Preise auf eine Seite, so erhält man folgende Variante:

Im Punkte x_m maximalen Nutzens gilt für alle i, j die Gleichheit folgender Verhältnisse:

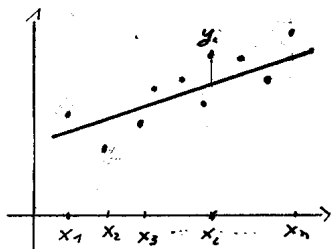
$$\frac{\frac{\partial f}{\partial x_i}(x_m)}{\frac{\partial f}{\partial x_j}(x_m)} = \frac{\text{Grenznutzen nach } x_i}{\text{Grenznutzen nach } x_j} = \frac{p_i}{p_j}$$

9.5.5 Ausgleichsgerade

Dies ist eine Anwendung der Extremwertbestimmung (ohne Nebenbedingungen) in der Statistik.

Ausgangslage:

Bei einer Erhebung festgestellt (bzw. bei einem Experiment gemessen) wurden für gewisse Argumente $x_1, x_2, \dots, x_n \in \mathbb{R}$ die Meßdaten $y_1, y_2, \dots, y_n$.



Gesucht:

Eine Gerade $g(x) = ax + b$, so daß

$$Q(a, b) = \sum_{i=1}^n (g(x_i) - y_i)^2 = \sum_{i=1}^n (ax_i + b - y_i)$$

minimal wird.

Die Theorie der Extremwertbestimmung gibt vor:

Zu suchen sind die (a, b) mit $\text{grad } Q(a, b) = \mathbf{0}$ d.h. mit $\frac{\partial Q}{\partial a} = 0 = \frac{\partial Q}{\partial b}$, also mit:

$$\begin{aligned}\frac{\partial Q}{\partial a} &= \sum_{i=1}^n 2(ax_i + b - y_i) \cdot x_i = 2 \cdot \left(\left(\sum_{i=1}^n x_i^2 \right) \cdot a + \left(\sum_{i=1}^n x_i \right) \cdot b - \sum_{i=1}^n x_i y_i \right) = 0 \\ \frac{\partial Q}{\partial b} &= \sum_{i=1}^n 2 \cdot (ax_i + b - y_i) = 2 \cdot \left(\left(\sum_{i=1}^n x_i \right) \cdot a + n \cdot b - \left(\sum_{i=1}^n y_i \right) \right) = 0\end{aligned}$$

Dabei: Die $\sum x_i$, $\sum x_i^2$, $\sum y_i$, $\sum x_i y_i$ sind durch das Experiment gegebene und also konstante Zahlen.

Man erhält ein Gleichungssystem aus zwei linearen Gleichungen für die beiden Unbekannten a und b :

$$\begin{aligned}\left(\sum_{i=1}^n x_i \right) \cdot a + n \cdot b &= \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i^2 \right) \cdot a + \left(\sum_{i=1}^n x_i \right) \cdot b &= \sum_{i=1}^n x_i y_i\end{aligned}$$

(Name für dieses Gleichungssystem: **System der Normalgleichungen**).

Daraus lassen sich a und b bestimmen.

Anmerkung Es lohnt sich nicht, die nicht ganz übersichtlichen Koeffizienten dieser Gleichungen auswendig zu lernen. Einfacher ist es, sich die Herleitung zu merken und sich die Koeffizienten selbst zu bestimmen.

9.6 Aufgaben

Aufgabe 1. Gegeben sei die Funktion:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(x, y) = x^4 - 4x^2y^2 + y^4.$$

Betrachten Sie die davon hergeleiteten Funktionen $f_{y=0}$ und $f_{x=0}$, die jeweils die Einschränkung von f auf die x -Achse bzw. die y -Achse darstellen:

$$\begin{aligned} f_{y=0} : \mathbb{R} &\rightarrow \mathbb{R}, & f_{y=0}(x) &= f(x, 0) = x^4 \\ f_{x=0} : \mathbb{R} &\rightarrow \mathbb{R}, & f_{x=0}(y) &= f(0, y) = y^4. \end{aligned}$$

Zeigen Sie, dass sowohl $f_{y=0}$ als auch $f_{x=0}$ an der Stelle $x = 0$ (bzw. $y = 0$) ein Minimum haben. Können Sie daraus schließen, dass die Funktion f an der Stelle $(x, y) = (0, 0)$ auch ein Minimum hat? Überprüfen Sie Ihre Aussage durch Einsetzen verschiedener Punkte.

Aufgabe 2. Bestimmen Sie Gradient und Hessematrix für folgende Funktionen:

$$\begin{aligned} f_1 : \mathbb{R}^2 &\rightarrow \mathbb{R} & (x, y) &\mapsto 3xy + 5y^2 + 2 \\ f_2 : \mathbb{R}^2 &\rightarrow \mathbb{R} & (x, y) &\mapsto \sin(xy) \\ f_3 : \mathbb{R}^3 &\rightarrow \mathbb{R} & (x, y, z) &\mapsto x \cdot \sin(yz) - e^{x^2+z} \end{aligned}$$

Aufgabe 3. Bestimmen Sie Gradient und Hessematrix von (vgl. Aufgabe 1):

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(x, y) = x^4 - 4x^2y^2 + y^4.$$

Zeigen Sie, dass die Hessematrix am Punkt $(0, 0)$ indefinit ist.

Aufgabe 4. Bestimmen Sie die relativen Extrema der Funktion:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto x^3 + 3x^2y + 3y^2x - 3x$$

Aufgabe 5. Eine Firma stellt aus den Produktionsfaktoren a und k ein Produkt P her. Die Produktionsfunktion sei

$$x(a, k) = 20a^{\frac{1}{4}}k^{\frac{1}{2}},$$

und das Endprodukt lasse sich zum fixen Marktpreis von $p = 10$ EUR absetzen. Eine Arbeitsstunde a kostet 5 EUR, eine Einheit k des Kapitalstocks kostet 10 EUR. Welcher Faktoreinsatz maximiert den Gewinn, wenn keine weiteren Kosten anfallen? Wie hoch ist dieser?

Aufgabe 6. Berechnen Sie die partiellen Elastizitäten sowie die Skalenelastizitäten folgender Funktionen:

a) $f(x, y) = e^{x^2+y}$

b) $f(x, y, z) = x^2yz^3$

Aufgabe 7. Betrachten Sie folgende Produktionsfunktion in Abhängigkeit der Faktoren Arbeit und Kapital:

$$f(A, K) = (A^{-\frac{1}{2}} + 2K^{-\frac{1}{2}})^{-2}.$$

Zeigen Sie, dass diese Funktion Skalanelastizität 1 besitzt.

Aufgabe 8. Bestimmen Sie mögliche Extrema der Funktion

$$f(x, y, z) = (x - 1)^2 + y^2 + z^2$$

unter der Nebenbedingung

$$x + 2y + 2z = 10$$

(vgl. auch Klausuraufgabe 15 von letztem Jahr).