

Mathematik für Wirtschaftswissenschaftler

Vorlesungsprogramm für den 14. 06. 2007

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, SS 2007)

Kapitel 5: Analysis der Funktionen mehrerer Variablen

5.1 Funktionen von mehreren Veränderlichen, partielle Ableitungen und Elastizitäten

Die meisten ökonomischen Vorgänge hängen von mehreren Parametern ab und werden dementsprechend mathematisch modelliert durch Funktionen von mehreren reellen Variablen. Funktionen von einer Veränderlichen erhält man in solchen Modellen nur, indem man die Diskussion auf einen Parameter konzentriert, der als unabhängige Variable angesehen wird, und alle übrigen Parameter als konstant annimmt (die sog. *c.p.-Bedingung* — “ceteris paribus”). Diese Annahme ist aber häufig unrealistisch, und es ist notwendig, das Funktionsverhalten zu analysieren, wenn mehrere Parameter gleichzeitig variieren. Man hat dann eine (**reelle**) **Funktion von mehreren reellen Veränderlichen** $f(x_1, \dots, x_n)$ zu betrachten, die für jede Spezifikation der Werte der n unabhängigen Variablen $x_1, \dots, x_n \in \mathbb{R}$ als “Eingabe” eine reelle Zahl $f(x_1, \dots, x_n)$ als “Ausgabe” liefert. Mathematisch gesprochen handelt es sich um eine Abbildung $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$, die jedem Element $x = (x_1, \dots, x_n)$ aus dem Definitionsbereich D , der eine (nichtleere) Teilmenge des n -dimensionalen Zahlenraums \mathbb{R}^n ist, eine reelle Zahl $f(x) = f(x_1, \dots, x_n) \in \mathbb{R}$ als Funktionswert zuordnet. Zur Notationsvereinfachung werden dabei die n unabhängigen Variablen zu einer Vektorvariablen $x = (x_1, \dots, x_n)$ zusammengefasst, die in $D \subset \mathbb{R}^n$ variiert. Bei einer niedrigen Anzahl der unabhängigen Variablen werden diese häufig nicht durchnummeriert, sondern mit x, y, z, \dots oder anderen Buchstaben bezeichnet; man schreibt also $f(x, y)$ (oder $f(u, v)$ etc.) für eine Funktion von zwei Veränderlichen und $f(x, y, z)$ (oder $f(r, s, t)$ etc.) für eine Funktion von drei Variablen. Bei dieser Notation bezeichnet dann jeder Buchstabe x, y, z, \dots eine reelle Variable, während x bei der Vektornotation $\mathbb{R}^n \supset D \ni x \mapsto f(x) \in \mathbb{R}$ für einen Satz (x_1, \dots, x_n) von n reellen Variablen steht. Aus dem Kontext ergibt sich jeweils, ob eine reelle oder eine vektorielle Variable gemeint ist. (Nötigenfalls kann man die vektoriellen Variablen durch eine Notation wie $\mathbf{x}, \mathbf{y}, \dots$ von den reellen Variablen unterscheiden.)

Wenn ein ökonomischer Sachverhalt nicht durch eine, sondern durch mehrere reelle Größen beschrieben wird, die von denselben reellen Parametern abhängen, so hat man $m \geq 2$ reelle Funktionen f_1, \dots, f_m von n reellen Variablen $x = (x_1, \dots, x_n)$ in demselben Definitionsbereich $\subset \mathbb{R}^n$ zu betrachten. Mathematisch gesprochen ist hierdurch eine Abbildung $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ definiert, die jedem $x \in D$ den vektoriellen Wert

$$\begin{aligned} F(x) &= (f_1(x), f_2(x), \dots, f_m(x)) \\ &= (f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) \in \mathbb{R}^m \end{aligned}$$

zuordnet. (Oft werden diese m -gliedrigen Vektoren auch als Spalten geschrieben; wir hatten ja vereinbart, dass (y_1, \dots, y_m) eine typographisch günstigere Notation für den m -gliedrigen Spaltenvektor mit Einträgen $y_1, \dots, y_m \in \mathbb{R}$ sein soll.) Man nennt F dann eine **Vektorfunktion von mehreren Veränderlichen** und f_1, f_2, \dots, f_m ihre **Komponentenfunktionen**. Diese sind **skalare Funktionen**, wie \mathbb{R} -wertige Funktionen auch genannt werden, um sie zu unterscheiden von “echten” Vektorfunktionen. (Streng genom-

men ist auch eine reellwertige Funktion eine “Vektorfunktion”, eben eine mit Werten im Vektorraum $\mathbb{R} = \mathbb{R}^1$. Alles was im Folgenden über Vektorfunktionen gesagt werden wird, gilt daher insbesondere auch für skalare Funktionen. Wir reservieren den Ausdruck “Vektorfunktion” aber im Allgemeinen für “echte” Vektorfunktionen.)

Eine \mathbb{R}^m -wertige Funktion F enthält dieselbe Information wie das m -tupel ihrer reellen Komponentenfunktionen f_1, \dots, f_m . Deshalb schreibt man einfach $F = (f_1, \dots, f_m)$, um anzuzeigen, dass $F(x) = (f_1(x), \dots, f_m(x))$ ist für alle x aus dem gemeinsamen Definitionsbereich D der Komponentenfunktionen f_i . Die Analysis der Vektorfunktion F läuft im Wesentlichen auf die Untersuchung ihrer reellen Komponentenfunktionen hinaus; deshalb werden wie uns im Folgenden überwiegend mit reellen Funktionen befassen. Bei einer kleinen Anzahl von Komponentenfunktionen verzichtet man oft auf ihre Nummerierung und schreibt also (f, g) bzw. (f, g, h) für die \mathbb{R}^2 -wertige Vektorfunktion mit den reellen Komponentenfunktionen f, g bzw. für die \mathbb{R}^3 -wertige Funktion (f, g, h) mit den skalaren Komponenten f, g, h . (Man könnte andererseits auch $f(x) = (f_1(x), \dots, f_m(x))$ für die Vektorfunktion mit Komponenten f_i schreiben, so dass f für eine Vektorfunktion steht. Doch bevorzugen wir im Allgemeinen Großbuchstaben, um Vektorfunktionen schon durch die Notation als solche kenntlich zu machen.)

Bei der Analyse einer vorgelegten Funktion von mehreren Veränderlichen sollte man sich — wie in der Linearen Algebra — zuerst Klarheit über die Formate von “Eingaben” und “Ausgaben” verschaffen. Man bestimmt oder spezifiziert also zuerst

- die Anzahl n der unabhängigen Variablen (Komponentenzahl der Eingabevektoren)
- und die Anzahl m der Komponentenfunktionen (Komponentenzahl der Funktionswerte; $m = 1$ bei skalaren Funktionen, $m > 1$ bei Vektorfunktionen).

Mit dem Definitionsbereich nimmt man es dagegen in der Analysis und insbesondere in der Wirtschaftsmathematik nicht so genau. Er wird oft nicht genau angegeben, sondern man unterstellt, dass irgendein für die betrachtete Funktionsvorschrift mathematisch oder ökonomisch sinnvoller Definitionsbereich gewählt ist, ohne diesen näher zu spezifizieren.

BEISPIELE: (1) Häufig auftretende *Definitionsbereiche* für Funktionen von n Variablen sind neben dem ganzen Zahlenraum \mathbb{R}^n die Menge $\mathbb{R}_{>0}^n$ aller *positiven Vektoren* bzw. die Menge $\mathbb{R}_{\geq 0}^n$ aller *nichtnegativen Vektoren*, wobei die Vektoren $x = (x_1, \dots, x_n)$ gemeint sind, deren Einträge x_j allesamt positiv bzw. alle nichtnegativ sind. Allgemeiner sind (Kartesische) *Produkte von Intervallen* $D = I_1 \times I_2 \times \dots \times I_n$, womit die Menge aller $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ bezeichnet ist, für die x_1 im ersten, x_2 im zweiten ... und x_n im letzten der gegebenen Intervalle $I_1, I_2, \dots, I_n \subset \mathbb{R}$ liegt. Haben alle Intervalle I_j eine endliche positive Länge, so nennt man D einen *Quader* mit den Kantenintervallen I_j bzw. einen *Würfel*, wenn die Längen alle gleich sind (und in der Zeichenebene \mathbb{R}^2 bzw. im Anschauungsraum \mathbb{R}^3 ist D dann ein Achsen-paralleler Quader bzw. Würfel).

(2) Die (nach den konstanten Funktionen) einfachsten Funktionen von n Veränderlichen sind die (**affin**) **linearen Funktionen**

$$\ell(x) = a_1x_1 + a_2x_2 + \dots + a_nx_n + b = \mathbf{a} \cdot x + b$$

mit Koeffizientenvektor $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ und Inhomogenität $b \in \mathbb{R}$ im skalaren Fall

bzw.

$$L(x) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n + b_n \end{pmatrix} = Ax + \mathbf{b}$$

mit Koeffizientenmatrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ und Inhomogenität $\mathbf{b} = (b_1, \dots, b_m) \in \mathbb{R}^m$ im vektoriellen Fall. Diese Funktionen waren der Gegenstand der Linearen Algebra (Kap. 3), und mit Hilfe der Linearen Algebra kann man für solche Funktionen alle interessierenden Fragen beantworten und alle relevanten Größen konkret ausrechnen. In der Realität sind aber ökonomische Abhängigkeiten selten linear; sie lassen sich höchstens lokal, d.h. auf geeignet kleinen Definitionsbereichen, einigermaßen gut durch lineare Funktionen approximieren. “Global” kann man gewisse qualitative Eigenschaften von abhängigen Variablen $f(x)$, z.B. superlineares Wachstum $f(x) \rightarrow \infty$ schneller als $|x| \rightarrow \infty$ oder das Erreichen einer endlichen “Sättigungsgrenze” $f(x) \rightarrow s$ bei $|x| \rightarrow \infty$, mit linearen Funktionen überhaupt nicht modellieren. (Hier ist $|x| = \sqrt{(x_1)^2 + \dots + (x_n)^2}$ die Euklidische Norm des Vektors $x \in \mathbb{R}^n$, also der Abstand des Punktes x zum Ursprung $0 \in \mathbb{R}^n$, und $|x| \rightarrow \infty$ bedeutet in $\mathbb{R}_{>0}^n$ zum Beispiel, dass mindestens ein Eintrag x_j von $x = (x_1, \dots, x_n)$ über alle Grenzen wächst.)

(3) Also muss man für die mathematische Beschreibung ökonomischer Vorgänge meistens nichtlineare Funktionen von mehreren Veränderlichen heranziehen. Die in gewisser Weise einfachsten nichtlinearen Funktionen sind **Polynomfunktionen von mehreren Veränderlichen**, darunter versteht man Linearkombinationen von sog. *Monomen*, d.h. Potenzprodukten $x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$ der n Variablen x_1, x_2, \dots, x_n mit Exponenten $k_j \in \mathbb{N}_0$ (wobei Faktoren $x_j^0 \equiv 1$ normalerweise nicht hingeschrieben werden). Eine konkrete Polynomfunktion von zwei reellen Veränderlichen x, y ist z.B.

$$p(x, y) = x^2 - 2xy^2 + 3y^3,$$

eine von drei reellen Variablen x, y, z ist

$$q(x, y, z) = 1 - 2z + x^2 - y^2 + 5xy^2z$$

und eine von n Veränderlichen x_1, \dots, x_n ist

$$r(x) = x_1 + 2x_2^2 + 3x_3^3 + \dots + nx_n^n - (x_1x_2 \cdot x_n)^2.$$

Der **Grad** der Polynomfunktion ist die größte auftretende Exponentensumme bei den Monomen, deren Linearkombination das Polynom ist. Der Grad von p ist also 3, der von q ist 4 und der von r ist $2n$. Die Polynomfunktionen vom Grad 0 sind die konstanten Funktionen, die vom Grad ≤ 1 sind genau die linearen Funktionen. Es gibt $\binom{k+n-1}{k}$ verschiedene Monome vom Grad k in n Variablen; die allgemeine Polynomfunktion vom Grad m in n Variablen ist daher eine Linearkombination von $\sum_{k=0}^m \binom{k+n-1}{k} = \binom{m+n}{m}$ verschiedenen Monomen (z.B. von 10, wenn $n = 2$ und $m = 3$, oder wenn $n = 3$ und $m = 2$). Eine *polynomiale Vektorfunktion* von n Veränderlichen ist eine Vektorfunktion, deren sämtliche Komponentenfunktionen (skalare) Polynome von n Variablen sind.

(4) Eine **rationale Funktion von mehreren Veränderlichen** ist eine Funktion, die als Quotient $r(x) = \frac{p(x)}{q(x)}$ von zwei Polynomfunktionen von n Veränderlichen dargestellt werden kann. Der maximale Definitionsbereich ist dann natürlich das Komplement der Menge aller Nennernullstellen $\{x \in \mathbb{R}^n : q(x) = 0\}$. Anders als im Fall $n = 1$ besteht diese Nullstellenmenge im Allgemeinen nicht aus isolierten Punkten. Das zeigt schon der Fall linearer Nennerfunktionen $\ell(x)$, die eine $(n - 1)$ -dimensionale (Hyper-)Ebene als Nullstellenmenge in \mathbb{R}^n haben. (Es kann aber auch vorkommen, dass der Nenner genau eine oder gar keine Nullstelle hat, wie etwa $q(x, y) = x^2 + y^2$ bzw. $q(x, y) = x^2 + y^2 + 1$ auf \mathbb{R}^2 .) Eine *rationale Vektorfunktion* von mehreren Veränderlichen ist eine, deren Komponentenfunktionen allesamt rational sind.

(5) Die allgemeinsten Funktionen von mehreren Veränderlichen, die in konkreten ökonomischen Aufgabenstellungen auftreten, sind **elementare Funktionen von mehreren Veränderlichen**. Darunter versteht man Funktionen, die durch einen "Rechenterm" $T(x_1, \dots, x_n)$ definiert werden können, der aus Konstanten (in \mathbb{R}), den reellen Variablen x_1, \dots, x_n und den elementaren Grundfunktionen (von einer Veränderlichen) mit denselben Operationen aufgebaut ist wie die Rechenterme, die elementaren Funktionen von einer Veränderlichen definieren, d.h. mit Addition, Subtraktion, Multiplikation und Division (durch Nenner $\neq 0$) und Verkettung. Eine *elementare Vektorfunktion* von mehreren Veränderlichen ist eine, deren Komponentenfunktionen skalare elementare Funktionen sind. Die Funktionswerte zu konkreten Eingaben x_1, \dots, x_n kann man also im Prinzip mit einem einfachen Taschenrechner, der über die elementaren Grundfunktionen verfügt, ausrechnen. Beim Aufbau der elementaren Funktionen von mehreren Veränderlichen ist auch das Einsetzen verschiedener skalarer elementarer Funktionen $f_1(x), \dots, f_m(x)$ für die Argumente einer elementaren (skalaren oder vektoriellen) Funktion $G(y_1, \dots, y_m)$ erlaubt, d.h. $G(f_1(x), \dots, f_m(x))$ ist dann auch wieder eine elementare Funktion von n Veränderlichen $x = (x_1, \dots, x_n)$.

Eine elementare Funktion von zwei Veränderlichen x, y ist z.B.

$$f(x, y) = x^y = e^{y \ln x} \quad \text{auf } \mathbb{R}_{>0} \times \mathbb{R}$$

und eine von drei Variablen x, y, z ist

$$g(x, y, z) = e^{-x^2 - y^2 - z^2} \ln \left(\frac{1 + x^2 e^y}{1 + y^2 e^z} \right).$$

Als *Produktionsfunktionen* $x(r_1, \dots, r_n)$, die den Output einer Produktion (in Mengeneinheiten) in Abhängigkeit vom Einsatz von Produktionsfaktoren r_1, \dots, r_n (in Faktoreinheiten; Arbeit, Kapital, Rohstoffe, Energie, ...) angeben, wird oft eine elementare Funktion ähnlich wie ein Monom angesetzt, jedoch nicht unbedingt mit ganzen Exponenten, nämlich eine sog. **Cobb–Douglas–Funktion**:

$$x(r_1, r_2, \dots, r_n) = r_1^{s_1} r_2^{s_2} \cdot \dots \cdot r_n^{s_n} = \exp(s_1 \ln r_1 + s_2 \ln r_2 + \dots + s_n \ln r_n) \quad \text{auf } \mathbb{R}_{>0}^n,$$

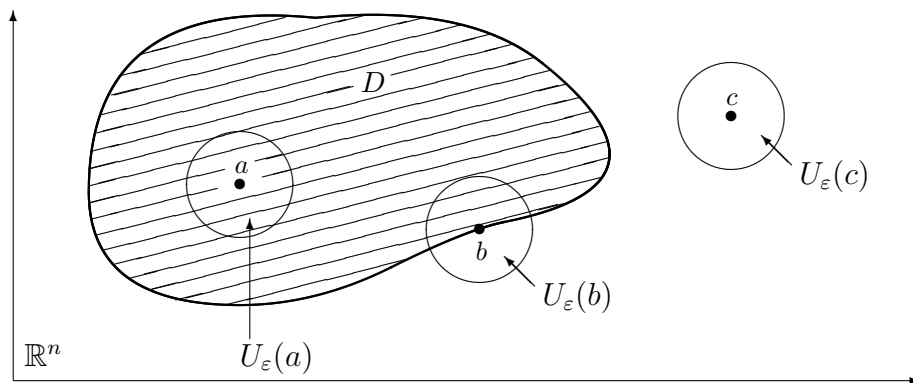
wobei die Exponenten s_j reelle Zahlen sind (meistens echt zwischen 0 und 1), oft mit Summe $s_1 + s_2 + \dots + s_n = 1$. Diese Summenbedingung bedeutet, dass eine Erhöhung des Faktoreinsatzes bei allen Produktionsfaktoren um denselben Prozentsatz $p\%$ auch eine Erhöhung des Outputs um diesen Prozentsatz $p\%$ bewirkt, also $x(qr_1, qr_2, \dots, qr_n) = qx(r_1, r_2, \dots, r_n)$ für $q = 1 + \frac{p}{100}$, und das ist oft eine sinnvolle Annahme. Auch Vielfache $c r_1^{s_1} r_2^{s_2} \cdot \dots \cdot r_n^{s_n}$ heißen "Cobb–Douglas–Funktionen". ■

In der Differentialrechnung für Funktionen f von mehreren Variablen geht es darum, die Änderungsraten in allen Richtungen zu bestimmen. Dazu müssen wir Werte $f(a)$ mit Werten $f(x)$ in allen nahe bei a gelegenen Punkten $x \in \mathbb{R}^n$ vergleichen und dann entsprechende Differenzenquotienten bilden. Voraussetzung dafür ist natürlich, dass die Funktion an Stellen x hinreichend nahe bei a definiert ist, dass also eine ganze "Umgebung" von a ganz im Definitionsbereich D der Funktion liegt. Für diese Eigenschaft eines Punktes $a \in D$ bzgl. der Menge $D \subset \mathbb{R}^n$ gibt es eine aus der Anschauung abgeleitete Terminologie, die wir nun zunächst einführen, weil wir sie im Folgenden benutzen werden. (Sie ist aber für das Verständnis der Differentiation bei mehreren Veränderlichen nicht wesentlich; es genügt, eine anschauliche Vorstellung von den Begriffen zu haben.)

DEFINITION: (i) Die (offene) **Kugel** in \mathbb{R}^n mit Mittelpunkt (Zentrum) $a \in \mathbb{R}^n$ und Radius $\varepsilon > 0$ ist die Menge $U_\varepsilon(a) = \{x \in \mathbb{R}^n : |x-a| < \varepsilon\}$ aller Punkte $x \in \mathbb{R}^n$, deren Euklidischer Abstand $|x-a| = \sqrt{(x_1-a_1)^2 + \dots + (x_n-a_n)^2}$ zum Punkt a kleiner als ε ist. Diese Kugel wird auch die **ε -Umgebung** des Punktes a in \mathbb{R}^n genannt.

(ii) Der Punkt $a \in \mathbb{R}^n$ heißt **innerer Punkt** der Menge $D \subset \mathbb{R}^n$, wenn eine ε -Umgebung $U_\varepsilon(a)$ ganz in D enthalten ist (mit beliebig kleinem, aber positiven ε). Er heißt dagegen **äußerer Punkt** von D , wenn eine ε -Umgebung $U_\varepsilon(a)$ keinen Punkt von D enthält, also ganz im Komplement $\mathbb{R}^n \setminus D$ liegt. Er heißt **Randpunkt** der Menge D , wenn er weder innerer noch äußerer Punkt ist, wenn also jede ε -Umgebung (mit noch so kleinem Radius ε) sowohl mindestens einen Punkt aus D enthält als auch mindestens einen, der nicht zu D gehört. Die Menge aller Randpunkte bildet den **Rand** von D und wird ∂D bezeichnet.

(iii) Die Menge $D \subset \mathbb{R}^n$ heißt **abgeschlossene Teilmenge** von \mathbb{R}^n , wenn sie all ihre Randpunkte enthält ($\partial D \subset D$), und sie heißt **offene Teilmenge**, wenn sie keinen ihrer Randpunkte enthält ($D \cap \partial D = \emptyset$), bzw. äquivalent, wenn alle Punkte von D innere Punkte sind. ■



innerer Punkt a , Randpunkt b und äußerer Punkt c von $D \subset \mathbb{R}^n$

BEMERKUNGEN und BEISPIELE:

(1) Der Rand des Komplements einer Teilmenge D von \mathbb{R}^n ist offenbar gleich dem Rand von D , $\partial(\mathbb{R}^n \setminus D) = \partial D$. Daher ist eine Teilmenge genau dann offen, wenn ihr Komplement abgeschlossen ist. Die meisten Teilmengen des \mathbb{R}^n sind aber weder offen noch abgeschlossen; sie enthalten eben einige ihrer Randpunkte, aber nicht alle.

(2) Der ganze Raum \mathbb{R}^n ist der einzige Definitionsbereich ohne Randpunkte und damit der einzige Definitionsbereich, der zugleich offen und abgeschlossen ist. (Nur noch die leere Menge ist offen und zugleich abgeschlossen; aber die leere Menge lassen wir als Definitionsbereich nicht zu.)

(3) Typische *abgeschlossene Definitionsbereiche* in \mathbb{R}^n sind Produkte von abgeschlossenen Intervallen, also z.B. der ganze Raum \mathbb{R}^n , ein *abgeschlossener Halbraum* $\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1}$, der Raum der nichtnegativen Vektoren $\mathbb{R}_{\geq 0}^n$ und sogenannte *abgeschlossene Quader* $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$. Ein abgeschlossener Bereich in \mathbb{R}^n ist auch die **abgeschlossene Kugel** $B_r(a) = \{x \in \mathbb{R}^n : |x-a| \leq r\}$ mit Mittelpunkt $a \in \mathbb{R}^n$ und Radius $r > 0$, also die Menge aller Punkte des \mathbb{R}^n , deren (Euklidischer) Abstand zu a nicht größer ist als r . Der Rand $\partial B_r(a)$ ist hier die Menge $S_r(a) = \{x \in \mathbb{R}^n : |x-a| = r\}$ aller Punkte mit Abstand zu a genau gleich r , genannt die (Abstands-) **Sphäre** mit Zentrum a und Radius r und offenbar in $B_r(a)$ enthalten. (Bei Dimension $n = 2$ sagt man *Kreisscheibe* statt "Kugel" und *Kreislinie* statt "Sphäre".)

(4) Typische *offene Definitionsbereiche* in \mathbb{R}^n sind Produkte von offenen Intervallen, also z.B. der ganze Raum \mathbb{R}^n , ein *offener Halbraum* $\mathbb{R}_{> 0} \times \mathbb{R}^{n-1}$, der Raum der positiven Vektoren $\mathbb{R}_{> 0}^n$ und *offene Quader* $]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_n, b_n[$. Ein offener Bereich in \mathbb{R}^n ist auch die **offene Kugel** $U_r(a) = \{x \in \mathbb{R}^n : |x-a| < r\}$ mit Mittelpunkt $a \in \mathbb{R}^n$ und Radius $r > 0$, also die Menge aller Punkte des \mathbb{R}^n , deren (Euklidischer) Abstand zu a echt kleiner ist als r . Der Rand $\partial U_r(a)$ ist derselbe wie bei der abgeschlossenen Kugel $B_r(a)$, also die Sphäre $S_r(a) = \{x \in \mathbb{R}^n : |x-a| = r\}$, hat aber mit $U_r(a)$ keinen Punkt gemeinsam, weshalb diese Kugel zu Recht "offen" genannt wird. ■

Nach diesem kleinen Exkurs in die Topologie (die Theorie, zu der die diversen qualitativen Konzepten der Lage von Punkten relativ zu Mengen gehören) können wir uns dem eigentlichen Gegenstand dieses Abschnitts zuwenden, den partiellen Ableitungen. Die Idee hierbei ist, dass man eine Funktion $f(x) = f(x_1, \dots, x_n)$ von $n \geq 2$ Veränderlichen als Funktion einer einzigen reellen Variablen x_j auffassen kann, indem man die anderen Variablen x_i mit $i \neq j$ als feste Parameter betrachtet, was in den Wirtschaftswissenschaften bekanntlich **c.p.-Bedingung** genannt wird ("ceteris paribus", d.h. die übrigen [Variablen] werden gleich [konstant] gehalten). Man studiert dann also für $j = 1 \dots n$ die sog. *partiellen Funktionen*

$$\varphi(x_j) := f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n)$$

als Funktionen jeweils einer reellen Variablen x_j bei festen Werten der anderen Variablen oder auch die Funktionen

$$\psi(t) := f(x_1, \dots, x_{j-1}, x_j+t, x_{j+1}, \dots, x_n) = f(x + te_j)$$

der reellen Variablen t bei festen Werten von x_1, \dots, x_n ; das läuft auf Dasselbe hinaus. (Zur Erinnerung: e_j ist der j -te *kanonische Basisvektor* in \mathbb{R}^n , der als j -ten Eintrag 1 hat und alle anderen Einträge 0; für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ und $t \in \mathbb{R}$ gilt also $x + te_j = (x_1, \dots, x_{j-1}, x_j+t, x_{j+1}, \dots, x_n)$.)

Die partiellen Ableitungen der Funktion $f(x_1, \dots, x_n)$ von mehreren Veränderlichen sind nun nichts anderes als die Ableitungen nach einer der Variablen x_j bei c.p.-Bedingung, also die Ableitungen der partiellen Funktionen nach der einen reellen Variablen, von der sie jeweils abhängen. Insofern ist dies kein neues Konzept im Vergleich zum Ableitungsbegriff aus Kap. 4, und dort haben wir ja auch schon nebenbei von partiellen Ableitungen gesprochen.

DEFINITION (i) Für eine reelle Funktion $\mathbb{R}^n \supset D \rightarrow \mathbb{R}$ von n Veränderlichen, eine innere Stelle $x = (x_1, \dots, x_n)$ von D und $j = 1 \dots n$ wird die **j -te partielle Ableitung an der Stelle x** erklärt als die Ableitung von f nach der j -ten Variablen an der Stelle x_j bei festen Werten der anderen Variablen x_i ($i \neq j$; c.p.-Bedingung), also als der Grenzwert von Differenzenquotienten

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(x + he_j) - f(x)}{h} &:= \frac{d}{dt} \Big|_{t=0} f(x + te_j) \\ &= \frac{d}{dt} \Big|_{t=0} f(x_1, \dots, x_{j-1}, x_j + t, x_{j+1}, \dots, x_n) \\ &= \frac{d}{dx_j} [f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n)], \end{aligned}$$

sofern dieser Grenzwert existiert. Der Wert dieser partiellen Ableitung an der Stelle x wird dann notiert

$$\frac{\partial f}{\partial x_j}(x) \quad \text{oder} \quad \frac{\partial}{\partial x_j} f(x), \quad \frac{\partial}{\partial x_j} \Big|_x f, \quad \partial_j f(x), \quad \partial_{e_j} f(x), \quad f_{x_j}(x), \dots$$

Ist D offen und $\frac{\partial f}{\partial x_j}(x)$ an allen Stellen $x \in D$ definiert, so heißt f **partiell nach der j -ten Variablen differenzierbar** auf D , und die Funktion

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

wird die **j -te partielle Ableitung(sfunktion)** von f auf D genannt.

(ii) Ist $F = (f_1, \dots, f_m)$ eine \mathbb{R}^m -wertige Vektorfunktion auf $D \subset \mathbb{R}^n$ mit reellen Komponentenfunktionen f_1, \dots, f_m , so erklärt man die **j -te partielle Ableitung der Vektorfunktion an der Stelle x** (innerer Punkt von D) als den Vektor mit den partiellen Ableitungen der Komponentenfunktionen an dieser Stelle als Einträgen, also als den (Spalten-)Vektor

$$\frac{\partial F}{\partial x_j}(x) = \left(\frac{\partial f_1}{\partial x_j}(x), \dots, \frac{\partial f_m}{\partial x_j}(x) \right) \in \mathbb{R}^m,$$

sofern alle f_i an der Stelle x partiell nach x_j differenzierbar sind. Entsprechend heißt die Vektorfunktion

$$\frac{\partial F}{\partial x_j} = \left(\frac{\partial f_1}{\partial x_j}, \dots, \frac{\partial f_m}{\partial x_j} \right) : D \rightarrow \mathbb{R}^m$$

die **j -te partielle Ableitung(sfunktion) der Vektorfunktion F** auf D . ■

Die Voraussetzung in (i), dass x innerer Punkt bzw. der Definitionsbereich D offen ist, stellt sicher, dass die partiellen Funktionen auf einem Intervall $]x_j - \varepsilon, x_j + \varepsilon[$ mit $\varepsilon > 0$ um die jeweils relevante Stelle $x_j \in \mathbb{R}$ definiert sind, so dass die Differenzenquotienten für alle genügend kleine Werte von $|h| > 0$ gebildet werden können. Der Wert $\frac{\partial f}{\partial x_j}(x)$ einer partiellen Ableitung hängt nicht nur vom Wert der Variablen x_j ab, nach der differenziert wurde, sondern natürlich auch von den Werten der anderen (festgehaltenen) Variablen. Daher ist die partielle Ableitung $\frac{\partial f}{\partial x_j}(x_1, \dots, x_n)$ eine Funktion von *allen* n unabhängigen Veränderlichen. Das Attribut “partiell” drückt übrigens aus, dass mit einer solchen Ableitung nicht das gesamte Änderungsverhalten der Funktion beschrieben werden kann, sondern nur ein gewisser Teilaspekt davon, nämlich die Änderung der Funktionswerte, wenn man nur *eine* der unabhängigen Variablen variiert. (Mit den später definierten “Richtungsableitungen” und mit der “totalen Ableitung” wird dagegen das Änderungsverhalten der Funktion bei kleinen simultanen Änderungen der Werte von *allen* unabhängigen Variablen erfasst.)

DISKUSSION: 1) Zur *Notation*: Man verwendet $\frac{\partial}{\partial x_j}$ mit dem kyrillischen Buchstaben ∂ , um anzuzeigen, dass die Funktion noch von anderen Variablen abhängt als die, nach der differenziert wird, und dass das Ergebnis wieder als Funktion von allen Variablen aufgefasst wird, die im Spiel sind. Stattdessen $\frac{d}{dx_j}$ zu schreiben, wäre aber auch nicht verkehrt (und wird tatsächlich geschrieben, wenn die anderen Variablen während der gesamten Rechnung oder Diskussion als konstante, unveränderliche Parameter angesehen werden). Für die unabhängigen Variablen x_1, \dots, x_n und die abhängige Variable f werden je nach Kontext andere Buchstabensymbole gewählt. Bei einer Funktion $f(x, y)$ schreibt man dann z.B. $\frac{\partial f}{\partial x}(x, y)$ bzw. $\frac{\partial f}{\partial y}(x, y)$ für die partiellen Ableitungen nach der ersten bzw. zweiten Variablen, und bei einer Funktion $g(r, s, t)$ von drei Variablen entsprechend $\frac{\partial g}{\partial r}(r, s, t)$, $\frac{\partial g}{\partial s}(r, s, t)$, $\frac{\partial g}{\partial t}(r, s, t)$ für ihre partiellen Ableitungen.

In der Wirtschaftsmathematik sind die Buchstabensymbole für die Variablen oft durch den ökonomischen Kontext vorgegeben. Zum Beispiel kann x dort auch für die abhängige Variable (also für die Funktion) stehen. Bei Produktionsfunktionen etwa gibt $x(r_1, \dots, r_n)$ den Output (in Mengeneinheiten) in Abhängigkeit von den Faktoreinsätzen r_1, \dots, r_n (in Einheiten für die einzelnen Produktionsfaktoren) an; dann hat man also partielle Ableitungen $\frac{\partial x}{\partial r_1}, \dots, \frac{\partial x}{\partial r_n}$. Werden konkret die beiden Produktionsfaktoren Arbeit und Kapital betrachtet, so schreibt man $x(A, K)$ für die Produktionsfunktion in Abhängigkeit von diesen Faktoren und $\frac{\partial x}{\partial A}$, $\frac{\partial x}{\partial K}$ für die partiellen Ableitungen. Ist eine partielle Ableitung durch einen Rechterm $T(x) = T(x_1, \dots, x_n)$ gegeben und soll an einer bestimmten Stelle $a = (a_1, \dots, a_n)$ ausgewertet werden, so kann man das durch $T(x)|_{x=a}$ oder $T(x_1, \dots, x_n)|_{x_1=a_1, \dots, x_n=a_n}$ kenntlich machen. Flexibilität in der Notation ist zweckmäßig; es kommt nur darauf an, eindeutig auszudrücken, nach welcher Variablen differenziert werden soll, welche Variablen während des Ableitungsprozesses als konstant anzunehmen sind und an welcher Stelle die partielle Ableitung auszuwerten ist, wenn eine konkrete Stelle im Definitionsbereich gemeint ist.

2) *Geometrische Interpretation*: Der Wert der partiellen Ableitung $\frac{\partial f}{\partial x_j}(x)$ einer reellen Funktion an einer Stelle x im Definitionsbereich $D \subset \mathbb{R}^n$ gibt die **Steigung (des Graphen) von f an der Stelle x in Richtung des j -ten kanonischen Basisvektors e_j** an. Man stellt sich $\text{Graph}(f) = \{(x, f(x)) : x \in D\}$ als Oberfläche eines Gebirges über D vor (also über der Zeichenebene, wenn $n = 2$ ist), wobei $f(x)$ die Höhe über jedem Punkt der Basis D angibt. Bewegt man sich dann in der "Basis" D vom Punkt x aus geradlinig in verschiedenen Richtungen, so hat man natürlich auf der Gebirgsoberfläche im Allgemeinen auch verschiedene Steigungen in diesen Richtungen zu verzeichnen. Die Steigungen an der Stelle x in Richtung der (in positivem Sinn durchlaufenen) Koordinatenachsen sind die Werte der partiellen Ableitungen $\frac{\partial f}{\partial x_j}(x)$. Das ergibt sich aus der Interpretation der Ableitungswerte als Graphensteigung bei Funktionen von einer Veränderlichen.

3) Eine (nicht nur) *ökonomische Interpretation*: Der Ableitungswert $\frac{\partial f}{\partial x_j}(x)$ gibt die **Änderungsrate oder Änderungsgeschwindigkeit** der Funktion f an der Stelle x **bzgl. der j -ten Variablen** an (bei c.p.-Bedingung). Das ist genau wie bei der Ableitung von Funktionen einer einzigen Veränderlichen (und partielle Ableitungen sind ja Ableitungen nach *einer* Veränderlichen). Vereinfacht gesagt gilt:

- **Der Wert der partiellen Ableitung $\frac{\partial f}{\partial x_j}(x)$ gibt näherungsweise an, um wieviele Einheiten sich der Funktionswert $f(x)$ ändert, wenn man die Variable x_j um eine Einheit vergrößert und die Werte der anderen Variablen unverändert lässt.**

Jedenfalls gilt das, wenn eine Einheit der j -ten Variablen so klein ist, dass man die Änderungsrate der partiellen Funktion zwischen x_j und x_j+1 als (fast) konstant annehmen kann. Dann ist nämlich der Ableitungswert $\frac{\partial f}{\partial x_j}(x)$ ungefähr gleich dem Differenzenquotienten $\frac{f(x+e_j)-f(x)}{1}$. Genauer gilt Folgendes:

- Die Änderung $f(x+he_j) - f(x)$ des Funktionswert bei Erhöhung von x_j um h Einheiten (c.p.) ist ungefähr gleich $\frac{\partial f}{\partial x_j}(x)$, wenn $|h|$ klein ist,

wobei die Übereinstimmung von $f(x+he_j)$ mit $f(x)+h\frac{\partial f}{\partial x_j}(x)$ um so besser ist, je kleiner $|h|$ ist. (Dies ist nichts anderes als die Aussage, dass $\frac{\partial f}{\partial x_j}(x)$ der Limes der Differenzenquotienten $\frac{1}{h}[f(x+he_j) - f(x)]$ ist.)

- Das **Vorzeichen der partiellen Ableitung** $\frac{\partial f}{\partial x_j}(x)$ bestimmt dabei die Richtung der Änderung der Funktionswerte.

Bei positiver Ableitung $\frac{\partial f}{\partial x_j}(x) > 0$ nimmt der Wert $f(x)$ zu, wenn man x_j (geringfügig) vergrößert, und bei negativer Ableitung $\frac{\partial f}{\partial x_j}(x) < 0$ nimmt er in diesem Falle ab. (Wenn man x_j geringfügig verkleinert, so ist es natürlich gerade umgekehrt. Im Fall $\frac{\partial f}{\partial x_j}(x) = 0$ ist die Änderungsrate Null und die Änderung des Funktionswerts jedenfalls vernachlässigbar winzig gegenüber einer geringfügigen Änderung des Arguments x_j bei c.p.-Bedingung; über die Richtung der Änderung, d.h. Zunehmen oder Abnehmen des Funktionswerts, kann man in dieser Situation aber keine generelle Aussage machen.)

4) *Ökonomische Terminologie:* Diese ist ähnlich wie bei der Differentialrechnung für Funktionen von einer Veränderlichen, nur meist mit dem Zusatz "partiell". So heißt $\frac{\partial f}{\partial x_j}$ die **partielle Grenzfunktion** oder **partielle Marginalfunktion** zu f bzgl. der Variablen x_j . Bei einer Produktionsfunktion $x(A, K)$ spricht man z.B. von den (*partiellen*) *Grenzproduktivitäten* $\frac{\partial x}{\partial A}$, $\frac{\partial x}{\partial K}$ der Produktionsfaktoren Arbeit bzw. Kapital. Die Grenzproduktivität $\frac{\partial x}{\partial K}(A, K)$ gibt also an, um wieviele Produktionseinheiten sich der Output x erhöht (wenn $\frac{\partial x}{\partial K}(A, K) > 0$), wenn man eine zusätzliche Kapitaleinheit einsetzt und den Arbeitseinsatz unverändert lässt. In anderem Zusammenhang wird gesprochen von *partiellen Grenzkosten*, *partiellem Grenzgewinn* bzw. *partieller marginaler Konsumquote* jeweils bzgl. einer von mehreren ökonomischen Variablen, von denen die Kosten, der Gewinn bzw. der Konsum abhängen. Allgemein versteht man unter **Partialanalyse** die Untersuchung einer von mehreren Variablen abhängenden ökonomischen Funktion in Abhängigkeit von jeweils nur einer Variablen bei c.p.-Bedingung. Die **partielle Marginalanalyse** bezieht sich dabei auf die Änderungsraten der Funktion bzgl. der einzelnen Variablen bei c.p.-Bedingung, also auf die partiellen Ableitungen. (Die Partialanalyse gibt aber meistens weder mathematisch noch ökonomisch ein hinreichendes Verständnis der untersuchten Funktion — dazu mehr in 5.2.)

5) Für die partielle Differentiation von Vektorfunktionen gilt:

- Die partiellen Ableitungen $\frac{\partial F}{\partial x_j}$ von Vektorfunktionen F bildet man komponentenweise;
- sie haben Werte in demselben Vektorraum wie die Vektorfunktion F selbst.

Die erste Aussage bedeutet

$$\frac{\partial F}{\partial x_j} = \left(\frac{\partial f_1}{\partial x_j}, \frac{\partial f_2}{\partial x_j}, \dots, \frac{\partial f_m}{\partial x_j} \right) \quad \text{für } F = (f_1, f_2, \dots, f_m),$$

und so haben wir ja die partiellen Ableitungen von Vektorfunktionen erklärt. Die partielle Differentiation von Vektorfunktionen ist also nicht schwieriger als die von skalaren Funktionen (und als die Differentiation von reellen Funktionen einer einzigen Variablen), nur aufwendiger: Wenn F Werte in \mathbb{R}^m hat, so muss man eben m reelle Komponentenfunktionen partiell ableiten. Die zweite Aussage ist, dass die partielle Ableitungsfunktionen einer Vektorfunktion F ebenso viele Komponentenfunktionen haben wie F selbst. Eine Vektorfunktion mit $m \geq 2$ Komponentenfunktionen partiell abzuleiten und als Ergebnis eine skalare Funktion oder eine Vektorfunktion mit einer anderen Zahl von Komponentenfunktionen zu erhalten, oder bei Auswertung an einer bestimmten Stelle einen Vektor mit nicht genau m Einträgen — das ist schlimmer als ein falsches Ergebnis aufgrund von Rechenfehlern; nicht einmal das Format des Ergebnisses ist dann korrekt!

Die ökonomische Interpretation der partiellen Ableitungswerte $\frac{\partial F}{\partial x_j}(x) \in \mathbb{R}^m$ von Vektorfunktionen ist im Übrigen dieselbe wie in 4), d.h. $h \frac{\partial F}{\partial x_j}(x)$ gibt die ungefähre Änderung des Vektors $F(x)$ in \mathbb{R}^m an, wenn man x_j um h geringfügig erhöht ($h > 0$) oder erniedrigt ($h < 0$) bei c.p.-Bedingung.

6) Es gibt noch eine andere, mehr geometrische Interpretation der partiellen Ableitungen von Vektorfunktionen. Man kann nämlich $\frac{\partial F}{\partial x_j}(x) \in \mathbb{R}^m$ als Tangentenvektor der Kurve im m -dimensionalen Raum interpretieren, die man erhält, wenn man in $F(x_1, \dots, x_n)$ die j -te Variable laufen lässt und die anderen fixiert. Unter einer **Kurve** in \mathbb{R}^m versteht man eine Abbildung $I \ni t \mapsto c(t) = (c_1(t), \dots, c_m(t)) \in \mathbb{R}^m$ von einem Intervall $I \subset \mathbb{R}$ (das *Parameterintervall* der Kurve). Man stellt sich dabei t als laufende Zeit vor und $c(t)$ als zeitlich veränderlichen Punkt in \mathbb{R}^m , der die Kurve durchläuft. Die Ableitung einer solchen Kurve an einer Stelle $t = t_0$ ist der Limes $c'(t_0) = \lim_{t \rightarrow t_0} \frac{1}{t-t_0} [c(t) - c(t_0)]$. Dabei wird wieder komponentenweise differenziert, es ist also

$$c'(t) = (c'_1(t), \dots, c'_m(t)) \in \mathbb{R}^m.$$

Der Vektor $c'(t_0)$ ist ein Grenzwert von (um Faktoren $\frac{1}{t-t_0}$ gestreckten) Sehnenvektoren $c(t) - c(t_0)$ der Kurve, deren variabler Endpunkt $c(t)$ bei dem Grenzübergang gegen den festen Endpunkt $c(t_0)$ auf der Kurve strebt, also ein Tangentenvektor an die Kurve im Punkt $c(t_0)$. Die Euklidische Norm $|c'(t_0)|$ kann dabei als Momentangeschwindigkeit zur Zeit t_0 beim Durchlaufen der Kurve gemäß dem "Fahrplan" $c(t)$ gedeutet werden und die Richtung von $c'(t_0)$ (wenn nicht der Nullvektor) als momentane Richtung der Bewegung zur Zeit t_0 .

Im Zusammenhang mit partiellen Ableitungen von \mathbb{R}^m -wertigen Vektorfunktionen F ist die relevante Kurve eine partielle Funktion $c(t) = F(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n) \in \mathbb{R}^m$ und der partielle Ableitungswert $\frac{\partial F}{\partial x_j}(x)$ ist der Tangentenvektor $c'(t_0)$ dieser Kurve zur Zeit $t_0 = x_j$. Somit hat man eine Interpretation der Vektoren $\frac{\partial F}{\partial x_j}(x)$ als Tangentenvektoren an Kurven in \mathbb{R}^m , die durch die Vektorfunktion F definiert sind. Diese physikalisch-geometrische Deutung ist aber für die Ökonomie nicht von Bedeutung. ■

BEISPIELE: (1) $f(x, y) = x^2(1 + y) \implies \frac{\partial f}{\partial x}(x, y) = 2x(1 + y), \quad \frac{\partial f}{\partial y}(x, y) = x^2.$

(Für die Berechnung von $\frac{\partial f}{\partial x}$ haben wir y als Konstante zu betrachten und nach x abzuleiten; bei der Berechnung von $\frac{\partial f}{\partial y}$ ist es umgekehrt.) Für die Funktion kann auch ein Rechterm stehen, ohne dass ein Funktionsname vergeben ist:

$$\begin{aligned} \frac{\partial}{\partial x} e^{xy} &= ye^{xy}, & \frac{\partial}{\partial y} e^{xy} &= xe^{xy}; \\ \frac{\partial}{\partial r} r^s &= sr^{s-1} & (\text{Ableitung einer Potenzfunktion, } r > 0), \\ \frac{\partial}{\partial s} r^s &= r^s \ln r & (\text{Ableitung einer Exponentialfunktion, } r > 0). \end{aligned}$$

Die Verwendung der Notation ∂_j für die partielle Ableitung nach der j -ten Variablen empfiehlt sich nur, wenn ein Funktionsname vergeben ist und die Veränderlichen eindeutig nummeriert sind; sonst gibt es Unklarheiten. Zum Beispiel ist $\partial_1 e^{tz} = te^{tz}$, wenn $z = x_1$ die erste und $t = x_2$ die zweite Variable ist, aber $\partial_1 e^{tz} = ze^{tz}$ wenn man die Variablen $t = x_1, z = x_2$ umgekehrt nummeriert hat.

(2) Die Cobb–Douglas–Funktion

$$f(x, y, z) = \sqrt{x} \sqrt[3]{y} \sqrt[4]{z^3} \quad \text{auf } \mathbb{R}_{>0}^3$$

hat die partiellen Ableitungen

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y, z) &= \frac{1}{2} x^{-1/2} y^{1/3} z^{4/3} = \frac{1}{2} \frac{f(x, y, z)}{x}, \\ \frac{\partial f}{\partial y}(x, y, z) &= \frac{1}{3} x^{1/2} y^{-2/3} z^{4/3} = \frac{1}{3} \frac{f(x, y, z)}{y}, \\ \frac{\partial f}{\partial z}(x, y, z) &= \frac{3}{4} x^{1/2} y^{1/3} z^{1/3} = \frac{3}{4} \frac{f(x, y, z)}{z}. \end{aligned}$$

Der natürliche Definitionsbereich ist hier $\mathbb{R}_{>0}^3$, womit wir die Menge aller positiven Vektoren in \mathbb{R}^3 bezeichnet haben, also die Menge aller (x, y, z) mit $x > 0, y > 0$ und $z > 0$. Man erkennt das System: Für allgemeine Cobb–Douglas–Funktionen

$$f(x_1, x_2, \dots, x_n) = c x_1^{r_1} x_2^{r_2} \cdot \dots \cdot x_n^{r_n} \quad \text{auf } \mathbb{R}_{>0}^n$$

sind die partiellen Ableitungen gegeben durch

$$\frac{\partial f}{\partial x_j}(x_1, x_2, \dots, x_n) = \frac{r_j}{x_j} f(x_1, x_2, \dots, x_n).$$

(3) Die (Kartesischen) **Koordinatenfunktionen** auf \mathbb{R}^n sind die Funktionen $x \mapsto x_i$, die jedem Vektor $x \in \mathbb{R}^n$ seine reellen Einträge zuordnen. Man nennt die Funktion $f(x_1, \dots, x_n) = x_i$ auch die *Projektion auf die i -te Komponente* und bezeichnet sie meist einfach mit x_i . Für die partiellen Ableitungen der n Koordinatenfunktionen gilt

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij} = \begin{cases} 1, & \text{wenn } i = j, \\ 0, & \text{wenn } i \neq j, \end{cases} \quad \text{an jeder Stelle in } \mathbb{R}^n.$$

Für **affin lineare Funktionen** $\ell(x_1, \dots, x_n) = a_1 x_1 + \dots + a_n x_n + b$ ist allgemeiner

$$\frac{\partial \ell}{\partial x_j}(x_1, \dots, x_n) = a_j \quad \text{konstant gleich dem } j\text{-ten Koeffizienten.}$$

Im vektoriellen Fall $L(x) = Ax + \mathbf{b}$ mit $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ und $\mathbf{b} \in \mathbb{R}^m$ ist

$$\frac{\partial L}{\partial x_j}(x) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = A e_j \quad \text{konstant gleich der } j\text{-ten Koeffizientenspalte.}$$

(4) Ein **Monom** $x_1^{k_1} x_2^{k_2} \cdot \dots \cdot x_n^{k_n}$ ist eine Cobb–Douglas–Funktion mit nichtnegativen ganzen Exponenten. Die Ableitungsformel aus (2) gilt dafür auf ganz \mathbb{R}^n ,

$$\frac{\partial}{\partial x_j} x_1^{k_1} x_2^{k_2} \cdot \dots \cdot x_n^{k_n} = k_j x_1^{k_1} \cdot \dots \cdot x_j^{k_j-1} \cdot \dots \cdot x_n^{k_n} = \frac{k_j}{x_j} x_1^{k_1} x_2^{k_2} \cdot \dots \cdot x_n^{k_n},$$

wobei im Falle $k_j = 0$ natürlich $x_j^{k_j} = 1$ und $\frac{k_j}{x_j} x_1^{k_1} x_2^{k_2} \cdot \dots \cdot x_n^{k_n} = 0$ interpretiert wird, auch wenn $x_j = 0$ ist. Damit kann man dann auch jede **Polynomfunktion** von mehreren Veränderlichen problemlos partiell differenzieren; denn eine solche Funktion ist ja eine Linearkombination von Monomen, und die Ableitung nach einer Variablen x_j kann mit der Summen- und Faktorregel für Differentiation von Funktionen einer Veränderlichen einfach durch entsprechende Ableitung der Monome berechnet werden. Weil jede partielle Ableitung eines Monoms wieder ein Monom ist (oder die Nullfunktion) mit um 1 reduziertem Grad, erkennt man:

- Die partiellen Ableitungen einer Polynomfunktion sind wieder Polynomfunktionen,
- und zwar mit um 1 kleinerem Grad (sofern sie nicht die Nullfunktion sind).

(5) Wir betrachten zwei **Nachfrage–Funktionen** (*Preis–Absatz–Funktionen*) $x_1(p_1, p_2)$, $x_2(p_1, p_2)$ (in Mengeneinheiten) in Abhängigkeit von den Preisen p_1, p_2 (in Geldeinheiten) für die beiden betrachteten Produkte:

$$x_1(p_1, p_2) = 0.5 \cdot \frac{p_2^{0.7}}{p_1}, \quad x_2(p_1, p_2) = p_1^{1/2} - 1.5p_2 + 1.65.$$

Die partiellen Ableitungen sind:

$$\begin{aligned} \frac{\partial x_1}{\partial p_1} &= -0.5 \frac{p_2^{0.7}}{p_1^2}, & \frac{\partial x_1}{\partial p_2} &= 0.35 \frac{1}{p_1 p_2^{0.3}} \\ \frac{\partial x_2}{\partial p_1} &= \frac{1}{2p_1^{1/2}}, & \frac{\partial x_2}{\partial p_2} &= -1.5 \text{ konstant.} \end{aligned}$$

Fassen wir beide Funktionen zu einer Vektorfunktion $x(p_1, p_2) = \begin{pmatrix} x_1(p_1, p_2) \\ x_2(p_1, p_2) \end{pmatrix}$ zusammen, so können wir schreiben:

$$\frac{\partial x}{\partial p_1} = \begin{pmatrix} -0.5 p_1^{-2} p_2^{0.7} \\ 0.5 p_1^{-1/2} \end{pmatrix}, \quad \frac{\partial x}{\partial p_2} = \begin{pmatrix} 0.35 p_1^{-1} p_2^{-0.3} \\ -1.5 \end{pmatrix}.$$

Die Vorzeichen der partiellen Ableitungen haben hier ökonomische Bedeutung: Natürlich nimmt die Nachfrage nach einem Gut ab, wenn sein Preis erhöht wird; deshalb sind $\frac{\partial x_1}{\partial p_1}$ und $\frac{\partial x_2}{\partial p_2}$ negativ. Die “kreuzweise” gebildeten partiellen Ableitungen $\frac{\partial x_1}{\partial p_2}$ und $\frac{\partial x_2}{\partial p_1}$ sind dagegen positiv, d.h. die Nachfrage nach dem zweiten Produkt steigt, wenn der Preis für das erste erhöht wird und umgekehrt. Man sagt in diesem Falle, dass es sich um **substitutive Güter** handelt; bei Preiserhöhung für das eine weichen die Konsumenten auf das jeweils andere aus. Anders ist die Sachlage bei sog. **komplementären Gütern** die nur zusammen gekauft werden und deshalb auf Preisänderungen mit Nachfrageänderung in derselben Richtung reagieren.

Die **Erlösfunktion** für beide Güter zusammen ist, wenn der Absatz gleich der Nachfrage ist, das Skalarprodukt des Absatzvektors $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ mit dem Preisvektor $p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$, als Funktion der Preise ausgedrückt also

$$E(p_1, p_2) = \begin{pmatrix} x_1(p_1, p_2) \\ x_2(p_1, p_2) \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = 0.5 p_2^{0.7} + p_2 (p_1^{1/2} - 1.5 p_2 + 1.65).$$

Für die *partiellen Grenzerlöse* ergibt sich dann

$$\frac{\partial E}{\partial p_1} = \frac{1}{2}p_1^{-1/2}p_2, \quad \frac{\partial E}{\partial p_2} = p_1^{1/2} + 0.35p_2^{-0.3} - 3p_2 + 1.65.$$

Die partiellen Ableitungen der Nachfragefunktionen und des Erlöses geben an, um wieviel sich diese Größen jeweils ändern, wenn p_1 bzw. p_2 um eine (kleine) Geldeinheit erhöht wird. Betrachten wir konkret $p_1 = 4$ und $p_2 = 1$, so bewirkt eine Preiserhöhung von p_1 um h Geldeinheiten (bei festem Preis p_2) eine Verminderung der Nachfrage x_1 um ca. $h\frac{1}{32} = 0.03125h$ Mengeneinheiten, eine Erhöhung der Nachfrage x_2 um $h\frac{1}{4} = 0.25h$ Mengeneinheiten und eine Erhöhung des Erlöses um $4h$. Erhöhung von p_2 um h (bei festem Preis p_1) resultiert dagegen in einer Erhöhung der Nachfrage x_1 um $h0.35\frac{1}{4} = 0.0875h$, einer Verringerung der Nachfrage x_2 um $1.5h$ und einer Steigerung des Erlöses um h (alles nur näherungsweise gültig, wenn $h > 0$ klein ist).

Man erkennt hier deutlich, dass die Partialanalyse nicht ausreicht, um die Sachlage ökonomisch beurteilen zu können; denn was man natürlich eigentlich wissen möchte ist, wie sich der Erlös bei *simultanen* Änderungen der Preise p_1 und p_2 verändert und wie man die Preisveränderungen zu koppeln hat, um die größtmögliche Erlössteigerung zu erzielen. Das kann durch Analyse der einzelnen partiellen Ableitungen nicht beantwortet werden, weil diese ja nur die Änderungsrate bei Erhöhung *eines* Preises und Fixierung des anderen angeben. (Wir werden aber in 5.2 sehen, wie man die Antwort durch simultane Betrachtung beider partiellen Ableitungen finden kann.) ■

BEMERKUNGEN (Rechenregeln für partielle Ableitungen):

1) Bei skalaren Funktionen ergeben sich die Rechenregeln für partielle Ableitungen unmittelbar aus den Rechenregeln für Differentiation von Funktionen einer Veränderlichen (siehe 4.3); denn auch bei partiellen Ableitungen wird ja nur nach *einer* Variablen differenziert. Man hat also (für skalare Funktionen f, g von $x = (x_1, \dots, x_n)$ und $r \in \mathbb{R}$):

Summenregel	$\frac{\partial}{\partial x_j}(f \pm g) = \frac{\partial f}{\partial x_j} \pm \frac{\partial g}{\partial x_j},$	}	Linearität des Ableitungs- operators $\frac{\partial}{\partial x_j}$
Faktorregel	$\frac{\partial}{\partial x_j}(rf) = r \frac{\partial f}{\partial x_j},$		
Produktregel	$\frac{\partial}{\partial x_j}(fg) = \frac{\partial f}{\partial x_j}g + f \frac{\partial g}{\partial x_j},$		
Quotientenregel	$\frac{\partial}{\partial x_j} \frac{f}{g} = \frac{1}{g^2} \left(g \frac{\partial f}{\partial x_j} - f \frac{\partial g}{\partial x_j} \right)$		(wo $g \neq 0$).

Diese Regeln gelten an allen Stellen $x \in \mathbb{R}^n$, wo f und g auf einer Umgebung definiert sind und die partiellen Ableitungen existieren. Außerdem haben wir noch die

spezielle Kettenregel $\frac{\partial}{\partial x_j}(h \circ f) = (h' \circ f) \frac{\partial f}{\partial x_j}$

für Funktionen h von *einer* reellen Variablen. Dies gilt an der Stelle $x \in \mathbb{R}^n$, wenn $\frac{\partial f}{\partial x_j}(x)$ und $h'(f(x))$ existieren. Bei Auswertung an einer solchen Stelle x lautet die spezielle Kettenregel also:

$$\frac{\partial}{\partial x_j}(h(f(x))) = (h'(f(x))) \frac{\partial f}{\partial x_j}(x).$$

(Später werden wir eine allgemeinere Kettenregel zeigen, die auch für Vektorfunktionen F statt f gilt; deshalb heißt die Kettenregel hier "speziell".)

2) Für *Vektorfunktionen* gelten Summen- und Faktorregel wörtlich wie in 1), nur dass jetzt die Funktionen Werte in \mathbb{R}^m haben. Die **Produktregel für Vektorfunktionen** gilt in der Form

$$\frac{\partial}{\partial x_j}(F * G) = \frac{\partial F}{\partial x_j} * G + F * \frac{\partial G}{\partial x_j}$$

sogar für beliebige Produkte $*$: $\mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^m$, wenn F \mathbb{R}^p -wertig ist und G \mathbb{R}^q -wertig, also $(F * G)(x) = F(x) * G(x) \in \mathbb{R}^m$. Ein solches allgemeines Produkt ist eine bilineare Abbildung $(u, v) \mapsto u * v$, die jedem Paar von Vektoren $u \in \mathbb{R}^p$, $v \in \mathbb{R}^q$ einen Vektor $u * v \in \mathbb{R}^m$ zuordnet. (Bilinearität bedeutet, dass die Distributivgesetze $(ru + \tilde{r}\tilde{u}) * v = r(u * v) + \tilde{r}(\tilde{u} * v)$ und $u * (sv + \tilde{s}\tilde{v}) = s(u * v) + \tilde{s}(u * \tilde{v})$ gelten.) Man kann hier zum Beispiel an das Skalarprodukt $\mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$, $(u, v) \mapsto u \cdot v$ denken, an das Matrix-Vektor-Produkt $\mathbb{R}^{m \times n} \times \mathbb{R}^n \rightarrow \mathbb{R}^m$, $(A, x) \mapsto Ax$, an das Matrix-Produkt $\mathbb{R}^{l \times m} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{l \times n}$, $(A, B) \mapsto AB$, oder an die Multiplikation mit Skalaren $\mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, $(r, x) \mapsto rx$.

3) Zur **Existenz der partiellen Ableitungen** ist zu sagen, dass diese für jede auf einer offenen Menge $D \subset \mathbb{R}^n$ erklärte elementare, also durch eine “geschlossene Formel” definierbare, Funktion von mehreren Veränderlichen überall auf D gebildet werden können. Das ergibt sich aus den Rechenregeln in 1), aus der komponentenweisen Differentiation von Vektorfunktionen und aus der später gezeigten allgemeinen Kettenregel. Die partiellen Ableitungen kann man mit den Techniken der Differentialrechnung einer Veränderlichen auch konkret ausrechnen, und sie sind wieder elementar (so dass man sie beliebig oft weiter partiell differenzieren kann).

- *Jede elementare Funktion von mehreren Veränderlichen auf einem offenen Definitionsbereich ist dort nach allen Variablen partiell differenzierbar,*
- *und die partiellen Ableitungen sind wieder elementare Funktionen, die mit den Rechenregeln der Differentialrechnung berechnet werden können.*

Funktionen mit denen man ökonomische Sachverhalte modelliert, sind oft — aber nicht immer — elementar. Nichtelementare Funktionen ergeben sich, wenn man die Definition der Funktionsvorschrift durch unterschiedliche Formeln auf verschiedenen Teilgebieten des Definitionsbereichs vornimmt, wenn man also verschiedene elementare Funktionen zu einer Funktion zusammensetzt. In der Regel sind solche “stückweise elementaren Funktionen” an der Grenze zwischen Teilgebieten mit verschiedenen definierenden Termen nicht nach allen Variablen partiell differenzierbar, auch wenn die verschiedenen Funktionsstücke dort “stetig zusammenpassen”. Zum Beispiel ist $xe^{|y|}$ zusammengesetzt aus xe^y für $y \geq 0$ und xe^{-y} für $y \leq 0$; auf der Grenze liegen hier die Stellen $(x, 0)$, und dort existiert zwar $\frac{\partial}{\partial x}(xe^{|y|}) = e^{|y|}$, aber nicht $\frac{\partial}{\partial y}(xe^{|y|})$ (außer für $x = 0$). ■

Wie bei Funktionen einer Veränderlichen in 3.4 hat man auch im Fall von mehreren Variablen eine Variante des Konzepts der partiellen Ableitung, die *partielle Elastizität*, die besonders für ökonomische Anwendungen wichtig ist. Hierbei fragt man nicht, um wieviele Einheiten sich der Funktionswert $f(x)$ ändert, wenn man eine Variable x_j um eine Einheit erhöht (bei c.p.-Bedingung), sondern man interessiert sich für die *relative Änderung des Funktionswertes $f(x)$ in Abhängigkeit von der* relativen Änderung einer unabhängigen Variablen x_j bei Festhalten aller anderen. (Statt “relative Änderung” kann man auch *prozentuale Änderung* sagen.)

Die relative Änderung des Funktionswertes $f(x)$ beim Übergang von x_j zu $x_j + h$ und Festlassen der Werte x_i mit $i \neq j$ ist der Quotient $\frac{f(x+he_j)-f(x)}{f(x)}$, die relative Änderung des Arguments x_j ist $\frac{x_j+h}{x_j} = \frac{h}{x_j}$. Setzt man beides ins Verhältnis und führt den Grenzübergang $h \rightarrow 0$ aus, so erhält man

$$\lim_{h \rightarrow 0} \frac{f(x+he_j) - f(x)}{f(x)} : \frac{h}{x_j} = \lim_{h \rightarrow 0} \frac{x_j}{f(x)} \frac{f(x+he_j) - f(x)}{h} = \frac{x_j}{f(x)} \frac{\partial f}{\partial x_j}(x).$$

Das ist also für kleine Änderungen von x_j c.p. näherungsweise das Verhältnis der prozentualen Änderung des Funktionswertes zur prozentualen Änderung des Argumentes x_j .

DEFINITION: Die (**partielle**) **Elastizität bzgl. der j -ten Variablen** einer Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ an der inneren Stelle x von D ist (wenn $\frac{\partial f}{\partial x_j}(x)$ existiert)

$$\varepsilon_{f,x_j}(x) = \frac{x_j}{f(x)} \frac{\partial f}{\partial x_j}(x) \stackrel{\text{symbolisch}}{=} \frac{\frac{\partial f}{\partial x_j}(x)}{\frac{f(x)}{x_j}}.$$

Hierbei wird im Allgemeinen vorausgesetzt, dass x_j und $f(x)$ positiv sind. Existiert $\frac{\partial f}{\partial x_j}$ auf D , so heißt $D \ni x \mapsto \varepsilon_{f,x_j}(x)$ die **j -te partielle Elastizitätsfunktion** von f . ■

DISKUSSION: 1) Der wesentliche Vorteil der partiellen Elastizitäten gegenüber den partiellen Ableitungen ist (siehe 3.4).

- Die partielle Elastizität $\varepsilon_{f,x_j}(x)$ ist unabhängig von der Wahl der Einheiten, mit denen die Größe der Variablen x_j und der Funktionswerte $f(x)$ gemessen werden.

Elastizitäten sind daher unmittelbar vergleichbar, ohne dass Umrechnungen auf andere Einheiten nötig sind. Da eine Änderung der Variablen um 1% im Allgemeinen als kleine Änderung gelten kann, darf man sagen:

- Die partielle Elastizität $\varepsilon_{f,x_j}(x)$ gibt an, um wieviel Prozent sich der Funktionswert $f(x)$ ändert, wenn der Wert von x_j um 1% erhöht wird (c.p.).

Genauer ist es so, dass $f(x)$ sich um ca. $p \cdot \varepsilon_{f,x_j}(x)$ Prozent ändert, wenn x_j um p Prozent verändert wird (c.p.) und $|px_j|$ klein ist, wobei die Güte der Approximation $f(x + \frac{p}{100}x_j e_j) - f(x) \approx \frac{p}{100} \cdot \varepsilon_{f,x_j}(x) \cdot f(x)$ umso besser ist, je kleiner $|p|$ ist. Von relativen (oder prozentualen) Änderungen kann man nur bei positiven Größen sinnvoll sprechen; daher ist die Positivitätsvoraussetzung an $f(x)$ und x_j in der Definition.

2) Das **Vorzeichen der Elastizität** $\varepsilon_{f,x_j}(x)$ gibt an, ob der Funktionswert $f(x)$ bei geringer Vergrößerung von x_j c.p. zunimmt ($\varepsilon_{f,x_j}(x) > 0$) oder abnimmt ($\varepsilon_{f,x_j}(x) < 0$). Wegen $f(x) > 0$ und $x_j > 0$ hat nämlich $\varepsilon_{f,x_j}(x)$ dasselbe Vorzeichen wie die partielle Ableitung $\frac{\partial f}{\partial x_j}(x)$. (Es ist $\varepsilon_{f,x_j}(x) = 0$, genau wenn auch $\frac{\partial f}{\partial x_j}(x) = 0$ ist. In diesem Fall ist keine generelle Aussage über die Richtung der Funktionswertänderung möglich; sie ist aber jedenfalls marginal bei kleinen Änderungen von x_j .)

3) Wie bei Funktionen von einer Veränderlichen nennt man die skalare Funktion f an der Stelle x bzgl. der Variablen x_j (**partiell**) **elastisch**, wenn $|\varepsilon_{f,x_j}(x)| > 1$ ist, **unelastisch**, wenn $|\varepsilon_{f,x_j}(x)| < 1$ ist, und **ausgeglichene elastisch** (oder *proportional elastisch*), wenn $\varepsilon_{f,x_j}(x) = \pm 1$ ist. Das erste bedeutet, dass kleine relative Änderungen von x_j größere relative Änderungen von $f(x)$ bewirken, im zweiten Fall ist die prozentuale Änderung von $f(x)$ kleiner als die von x_j , im dritten Fall sind die relativen Änderungen von x_j und $f(x)$ ungefähr gleich groß (dem Betrage nach).

Im Sonderfall $\varepsilon_{f,x_j}(x) = 0$ ($\iff \frac{\partial f}{\partial x_j}(x) = 0$) bleibt $f(x)$ bei kleinen Änderungen von x_j praktisch unverändert; dann heißt f **vollkommen unelastisch** (oder *starr*) an der Stelle x bzgl. der Variablen x_j . Ein anderer Extremfall liegt vor, wenn kleine prozentuale Änderungen von x_j schon beliebig große relative Änderungen von $f(x)$ bewirken. Diese "instabile" Situation kann aber nur eintreten, wenn $\frac{\partial f}{\partial x_j}(x)$ nicht existiert und z.B. formal gleich $\pm\infty$ ist (d.h. der Limes der Differenzenquotienten $\frac{1}{h}[f(x + he_j) - f(x)]$ bei $h \rightarrow 0$ ist schon $\pm\infty$). Die Funktion f heißt in diesem Fall **vollkommen elastisch** bzgl. x_j an der Stelle x , und man setzt formal entsprechend $\varepsilon_{f,x_j}(x) = \pm\infty$.

4) Rechenregeln für partielle Elastizitäten ergeben sich unmittelbar aus den Rechenregeln für partielle Ableitungen bzw. aus den Rechenregeln für Elastizitäten von Funktionen einer einzigen Variablen in 3.4. Zum Beispiel gilt für Produkte bzw. Quotienten (positiver Funktionen)

$$\varepsilon_{fg,x_j} = \varepsilon_{f,x_j} + \varepsilon_{g,x_j}, \quad \varepsilon_{f/g,x_j} = \varepsilon_{f,x_j} - \varepsilon_{g,x_j}$$

und für Summen bzw. Differenzen (mit $f - g > 0$)

$$\varepsilon_{f \pm g, x_j} = \frac{f \varepsilon_{f, x_j} \pm g \varepsilon_{g, x_j}}{f \pm g}.$$

Für Verkettungen von f mit Funktionen $h: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ gilt schließlich

$$\varepsilon_{h \circ f, x_j} = \varepsilon_h(f(x)) \varepsilon_{f, x_j}(x).$$

5) Die partiellen Elastizitäten von Vektorfunktionen $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ erklärt man wie die partiellen Ableitungen komponentenweise, d.h. wenn F die Komponentenfunktionen f_1, \dots, f_m hat, so ist ε_{F, x_j} der (Spalten-)Vektor mit den partiellen Elastizitäten ε_{f_i, x_j} der Komponentenfunktionen als Einträgen:

$$\varepsilon_{F, x_j}(x) = \begin{pmatrix} \varepsilon_{f_1, x_j}(x) \\ \vdots \\ \varepsilon_{f_m, x_j}(x) \end{pmatrix} \in \mathbb{R}^m,$$

sofern x_j und alle Komponentenfunktionswerte $f_i(x)$ positiv sind (und die f_i partiell nach x_j differenzierbar sind an der Stelle x).

6) Die Ökonomische Terminologie ist ähnlich wie bei Funktionen von einer Veränderlichen, nur mit dem Zusatz "partiell". Als unabhängig angesehene Variable werden, besonders im Zusammenhang mit Produktionsfunktionen, als (*Input-*)**Faktoren** bezeichnet, was im Sinne eines "Faktors" gemeint ist, der Einfluss auf den Wert der abhängigen ökonomischen Größe, der *Output-Variablen*, hat (also nicht im Sinne eines Faktors, mit dem multipliziert wird). So spricht man bei einer Produktionsfunktion $x(r_1, \dots, r_n)$ von der *partiellen Output-Elastizität* bzgl. des j -ten Produktionsfaktors ε_{x, r_j} , und bei n Nachfragefunktionen $x_1(p_1, \dots, p_n), \dots, x_n(p_1, \dots, p_n)$ von der *Elastizitäten-Matrix der Nachfrage* bzgl. der Preise $(\varepsilon_{x_i, p_j})_{1 \leq i, j \leq n}$, worin der Eintrag in Zeile i und Spalte j die Elastizität der Nachfragefunktion $x_i(p_1, \dots, p_n)$ für das i -te Gut bzgl. des Preises p_j für das j -te Gut angibt. Im Fall $i \neq j$ wo also die Abhängigkeit der Nachfrage "über Kreuz" vom Preis für ein anderes Produkt beschrieben wird, ist die Rede von der *Kreuz-Preis-Elastizität der Nachfrage* ε_{f_i, x_j} , $i \neq j$. Die Güter Nr. i und Nr. j heißen, wie schon erklärt, *substitutiv*, wenn ihre Kreuz-Preis-Elastizitäten ε_{f_i, x_j} , ε_{f_j, x_i} beide positiv sind, und *komplementär*, wenn sie beide negativ sind. Bei zwei substitutiven Gütern hat eine Preisänderung für eines davon einen entgegengesetzt gerichteten Effekt auf die Nachfragefunktionen der beiden Güter, während bei komplementären Gütern eine Preisänderung bei einem davon zur gleichgerichteten Änderungen der Nachfrage für beide Güter führt. Die Elastizitäten ε_{f_i, x_i} , ε_{f_j, x_j} sind meistens negativ (Preiserhöhung verringert die Nachfrage).

Mit einiger Erfahrung lässt sich die ökonomische Terminologie im Zusammenhang mit Elastizitäten entschlüsseln, so dass man den ausladenden Wortverbindungen entnehmen kann, welche Funktion $f(x)$ gemeint ist und um welche partielle Elastizität ε_{f,x_j} es sich handelt. Um vor Irrtümern zu bewahren, weisen wir darauf hin, dass

- die Elastizität immer von der abhängigen Variablen genommen wird, die in der Notation ε_{f,x_j} als erstes Subskript erscheint.

Wenn also von der “Preis–Elastizität der Nachfrage” gesprochen wird, so handelt es sich um die Elastizität der Nachfrage–Funktion bzgl. eines Preises. Die “Elastizität des Preises bzgl. des Angebots” bezieht sich dagegen auf eine Preis–Funktion, deren Elastizität gebildet wird bzgl. eines am Markt angebotenen Produktions–Outputs (für das Produkt, dessen Marktpreis betrachtet wird, oder “über Kreuz” für ein Konkurrenzprodukt). ■

BEISPIELE: (1) Wir betrachten eine *Produktionsfunktion*

$$x(A, K) = 2.5 \cdot A^{0.6} \cdot K^{0.4}$$

in Abhängigkeit von den Produktionsfaktoren Arbeit und Kapital. Die Berechnung der partiellen Elastizitäten ist einfach:

$$\begin{aligned}\varepsilon_{x,A} &= \frac{A \cdot \frac{\partial x}{\partial A}(A, K)}{x(A, K)} = \frac{A \cdot 0.6 \cdot x(A, K)/A}{x(A, K)} = 0.6 \text{ konstant,} \\ \varepsilon_{x,K} &= \frac{K \cdot \frac{\partial x}{\partial K}(A, K)}{x(A, K)} = \frac{K \cdot 0.4 \cdot x(A, K)/K}{x(A, K)} = 0.4 \text{ konstant.}\end{aligned}$$

Wenn also der Arbeitseinsatz um 1 % erhöht wird (bei gleichbleibendem Kapitaleinsatz), so nimmt der Produktions–Output um 0.6 % zu, und wenn der Kapitaleinsatz um 1 % erhöht wird (bei gleichbleibendem Arbeitseinsatz), so wächst der Output um 0.4 %.

(2) Die Produktionsfunktion in (1) ist eine Funktion vom Cobb–Douglas–Typ

$$f(x_1, x_2, \dots, x_n) = c x_1^{r_1} x_2^{r_2} \cdot \dots \cdot x_n^{r_n},$$

also in jeder Variablen eine Potenzfunktion multipliziert mit einer Konstanten. Nun ist aber die Elastizität einer Potenzfunktion t^r von einer Variablen gerade der Exponent; denn $\varepsilon_{t^r,t} = \frac{t}{t^r} \frac{dt^r}{dt} \equiv r$. Und die Elastizität einer Funktion ändert sich nicht, wenn man sie mit einem konstanten Faktor multipliziert. (Das ist gerade die Unabhängigkeit der Elastizitäten von der Wahl der Einheiten, mit denen die Größe der Funktionswerte gemessen wird.) Also sehen wir ohne weitere Rechnung

$$\varepsilon_{f,x_j} = r_j \text{ konstant}$$

und haben das allgemeine Ergebnis:

- Die partiellen Elastizitäten von Cobb–Douglas–Funktionen sind konstant gleich den Exponenten, mit denen die einzelnen unabhängigen Variablen auftreten.

Die Eigenschaft, konstante partielle Elastizitäten zu haben, charakterisiert übrigens die Cobb–Douglas–Funktionen auf $\mathbb{R}_{>0}^n$. Ist nämlich $\varepsilon_{f,x_1} \equiv r_1$ konstant, so folgt $\varepsilon_{f/x_1^{r_1}, x_1} = \varepsilon_{f,x_1} - \varepsilon_{x_1^{r_1}, x_1} \equiv r_1 - r_1 = 0$, also ist $\frac{\partial}{\partial x_1}(f(x)/x_1^{r_1}) \equiv 0$, d.h. die Funktion $g(x_2, \dots, x_n) = f(x)/x_1^{r_1}$ ist von x_1 unabhängig. Ist auch $\varepsilon_{f,x_2} \equiv r_2$ konstant, so folgt $\varepsilon_{g,x_2} \equiv r_2$ und daher $f(x)/(x_1^{r_1} x_2^{r_2}) = g(x_2, \dots, x_n)/x_2^{r_2} = h(x_3, \dots, x_n)$ u.s.w. bis zur Konstanz $f(x)/(x_1^{r_1} x_2^{r_2} \cdot \dots \cdot x_n^{r_n}) \equiv c$.

(3) Wir greifen das frühere Beispiel der **Nachfrage-Funktionen**

$$x_1(p_1, p_2) = 0.5 \cdot \frac{p_2^{0.7}}{p_1}, \quad x_2(p_1, p_2) = p_1^{1/2} - 1.5p_2 + 1.65.$$

auf. Die partiellen Elastizitäten sind:

$$\begin{aligned} \varepsilon_{x_1, p_1} &= -1 \text{ konstant}, & \varepsilon_{x_1, p_2} &= 0.7 \text{ konstant}, \\ \varepsilon_{x_2, p_1} &= \frac{p_1^{1/2}}{2x_2(p_1, p_2)}, & \varepsilon_{x_2, p_2} &= \frac{-1.5p_2}{x_2(p_1, p_2)}. \end{aligned}$$

Für die Erlösfunktion $E(p_1, p_2) = 0.5p_2^{0.7} + p_2(p_1^{1/2} - 1.5p_2 + 1.65)$ hatten wir die partiellen Grenzerlöse

$$\frac{\partial E}{\partial p_1} = \frac{1}{2}p_1^{-1/2}p_2, \quad \frac{\partial E}{\partial p_2} = 0.35p_2^{-0.3} + p_1^{1/2} - 3p_2 + 1.65.$$

errechnet, also sind die partiellen Elastizitäten der Erlösfunktion

$$\varepsilon_{E, p_1} = \frac{p_1^{1/2} p_2}{2E(p_1, p_2)}, \quad \varepsilon_{E, p_2} = \frac{p_2(0.35p_2^{-0.3} + p_1^{1/2} - 3p_2 + 1.65)}{E(p_1, p_2)}.$$

Berechnung der Elastizitäten bei $p_1 = 4$, $p_2 = 1$ zeigt, dass eine Erhöhung von p_1 um 1 % zu einer Verringerung der Nachfrage x_1 um 1 %, einer Erhöhung der Nachfrage x_2 um $\frac{1}{2.15} \% \approx 0.47 \%$ und einer Erlössteigerung um $\frac{1}{2.65} \% \approx 0.38 \%$ führt. Eine Steigerung von p_2 um 1 % bewirkt eine Erhöhung der Nachfrage x_1 um 0.7 %, eine Verringerung der Nachfrage x_2 um $\frac{6}{2.15} \% \approx 2.79 \%$ und ebenfalls eine Erlöserhöhung um $\frac{1}{2.65} \% \approx 0.38 \%$.

(4) Wir betrachten analog zu (3) die Nachfrage nach drei Gütern

$$\begin{aligned} x_1(p_1, p_2, p_3) &= 100 - 4.5p_1 + 1.5p_2 - 0.5p_3, \\ x_2(p_1, p_2, p_3) &= \frac{100}{50 + p_1 - 0.5p_3} - 1.5p_2, \\ x_3(p_1, p_2, p_3) &= 2.5^{-p_1} 3p_2 10^{-p_3}, \end{aligned}$$

in einem Bereich, in dem die drei Nachfragewerte positiv sind. Die Elastizitäten-Matrix ist (mit $p = (p_1, p_2, p_3)$)

$$(\varepsilon_{x_i, p_j}) = \begin{pmatrix} \frac{-4.5p_1}{x_1(p)} & \frac{1.5p_2}{x_1(p)} & \frac{-0.5p_3}{x_1(p)} \\ \frac{-100p_1}{(50+p_1-\frac{1}{2}p_3)^2 x_2(p)} & \frac{-1.5p_2}{x_2(p)} & \frac{50p_3}{(50+p_1-\frac{1}{2}p_3)^2 x_2(p)} \\ -p_1 \ln 2.5 & p_2 \ln 3 & -p_3 \ln 10 \end{pmatrix}.$$

(In der i -ten Zeile steht die Elastizität von x_i bzgl. p_j .) Hier sind die Kreuz-Preis-Elastizitäten ε_{x_2, p_3} und ε_{x_3, p_2} positiv, die Güter Nr. 2 und Nr. 3 also Substitutionsgüter. Die Kreuz-Preis-Elastizitäten ε_{x_1, p_3} und ε_{x_3, p_1} sind negativ, die Güter Nr. 1 und Nr. 3 also Komplementärgüter. Von den Kreuz-Preis-Elastizitäten ε_{x_1, p_2} , ε_{x_2, p_1} ist die erste positiv und die zweite negativ, also sind die Güter Nr. 1 und Nr. 2 weder substitutiv noch komplementär zueinander. ■

5.2 Richtungsableitungen und Richtungselastizitäten

Durch Partialanalyse, also die Betrachtung einer Funktion $f(x_1, \dots, x_n)$ von n Veränderlichen als Funktion von jeweils nur einer Variablen x_j bei fixierten Werten der anderen Variablen $x_i (i \neq j)$, bekommt man meistens nur ungenügende Information über die zu untersuchende Funktion f . Es kann ja sein, dass sich $f(x_1, \dots, x_n)$ bei simultaner Änderung von mehreren unabhängigen Variablen ganz anders verhält als wenn man nur eine unabhängige Veränderliche bei c.p.-Bedingung variiert. Geometrisch gesprochen entspricht die Betrachtung der partiellen Funktion $t \mapsto f(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n)$ der Analyse des Funktionsverhaltens in Richtung der j -ten Koordinatenachse. Man wird aber unter Umständen von der Funktion einen ganz anderen Eindruck gewinnen, wenn man sie nicht nur in Richtung der Koordinatenachsen ansieht, sondern auch in anderen Richtungen. Das zeigen schon ein ganz einfaches

BEISPIEL: Die lineare Funktion $f(x, y) = x + y$ von zwei Veränderlichen hat die Steigung $\frac{\partial f}{\partial x} \equiv 1$ und $\frac{\partial f}{\partial y} \equiv 1$ in Richtung der Koordinatenachsen. Hat sie damit in allen Richtungen Steigung 1 oder gibt es noch Richtungen mit größerer Steigung? Schauen wir die Funktion z.B. in Richtung der Diagonalen an, also in Richtung des Einheitsvektors $\frac{1}{\sqrt{2}}(1, 1)$, so erhalten wir für die entsprechenden Differenzenquotienten

$$\frac{1}{h} \left[f\left(x + \frac{h}{\sqrt{2}}, y + \frac{h}{\sqrt{2}}\right) - f(x, y) \right] = \frac{1}{h} \left[\frac{h}{\sqrt{2}} + \frac{h}{\sqrt{2}} \right] = \sqrt{2},$$

also ist die Steigung der Funktion in dieser Richtung $\sqrt{2}$. Wenn wir eine beliebige Richtung ins Auge fassen, repräsentiert durch einen Einheitsvektor (u, v) , so ergeben sich als Differenzenquotienten in dieser Richtung

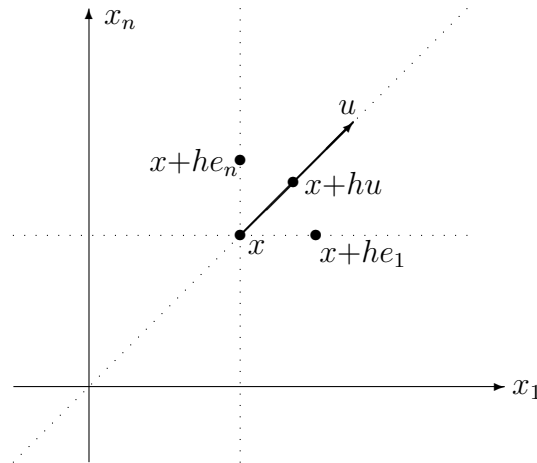
$$\frac{1}{h} [f(x+hu, y+hv) - f(x, y)] = \frac{1}{h} [hu + hv] = u + v,$$

und diese Steigungen können genau die Werte zwischen $-\sqrt{2}$ und $\sqrt{2}$ annehmen (weil $v^2 = 1 - u^2$ ist wegen $|(u, v)| = \sqrt{u^2 + v^2} = 1$ und weil die quadratische Gleichung $u \pm \sqrt{1 - u^2} = m \iff 1 - u^2 = (m - u)^2$ genau für $m^2 \leq 2$ eine Lösung u hat). Die Funktion hat also in verschiedenen Richtungen ganz verschiedene Steigungen, und $\sqrt{2}$ ist ihre größtmögliche Steigung. Allein mit Partialanalyse kann man dieses Funktionsverhalten, das im Übrigen typisch für differenzierbare Funktionen von mehreren Veränderlichen ist, natürlich nicht erkennen. ■

Die Aufgabe, die Richtungen maximaler Steigung einer Funktion $f(x)$ von mehreren Veränderlichen $x = (x_1, \dots, x_n)$ an einer gegebenen Stelle zu bestimmen, hat offensichtlich auch ökonomische Bedeutung: Wenn es sich z.B. um eine Gewinnfunktion handelt, deren Werte von den ökonomischen Variablen x_1, \dots, x_n abhängen, so ist man natürlich daran interessiert, die Variablen so einzustellen, dass der Funktionswert ein Maximum ist. Die Bestimmung des Maximums bei Funktionen von mehreren Variablen ist aber nicht einfach und oft nicht explizit möglich. In diesem Fall wird man wenigstens wissen wollen, in welche Richtung man einen aktuellen Wertesatz $x = (x_1, \dots, x_n)$ der unabhängigen Variablen verändern muss, so dass der Gewinn möglichst stark zunimmt. Mit anderen Worten: Wir suchen die *Richtung der größten Steigung* der Funktion f an der Stelle x .

Nun hat der höherdimensionale Raum \mathbb{R}^n viele Richtungen, und die Funktion f wird an der Stelle x in verschiedenen Richtungen verschiedene Steigungen haben; in manchen Richtungen wächst sie, in anderen fällt sie, in einer bestimmten Richtung wird sie vielleicht am stärksten wachsen. Nur in seltenen Ausnahmefällen wird es vorkommen, dass diese Richtung des stärksten Anstiegs gerade die Richtung eines kanonischen Basisvektors e_j ist, so dass die entsprechende Steigung als partielle Ableitung berechnet werden kann. Im Normalfall wird die Richtung u des stärksten Anstiegs an der Stelle x aber nicht achsenparallel sein, und dann können wir diese Richtung nur finden und die entsprechende Steigung berechnen, wenn wir die Funktion auch bei Veränderungen von x in dieser Richtung u studieren und nicht nur auf achsenparallelen Geraden durch x .

Die Quintessenz dieser Überlegungen ist: Es genügt nicht, partielle Ableitungen zu betrachten, also die Steigungen einer Funktion in den Achsenrichtungen, sondern wir müssen offenbar auch die *Ableitungen in beliebigen Richtungen* u erklären und berechnen, um ein so nahe liegendes Problem wie das der Richtung des stärksten Anstiegs angehen zu können. Aus dem Beispiel und aus den Formeln für die partiellen Ableitungen, also die Ableitungen in Richtung der kanonischen Basisvektoren, ist schon klar, wie das zu machen ist: Wir betrachten die Änderung der Funktionswerte $f(x+hu) - f(x)$, wenn wir x in Richtung u um ein Stück hu verschieben (mit $h \in \mathbb{R}_{\neq 0}$ und $|h|$ klein; wenn $h < 0$ ist, so geht diese Verschiebung natürlich in Richtung $-u$). Um die Steigung in Richtung u zu definieren, müssen wir die Änderung der Funktionswerte noch durch die Größe h dieser Verschiebung dividieren (jedenfalls ist das die Größe, wenn u ein Vektor der Länge 1 ist; in der folgenden Definition lassen wir aber Vektoren beliebiger Länge zu). Der Limes der "Richtungs-Differenzenquotienten" $\frac{1}{h}[f(x+hu) - f(x)]$ bei $h \rightarrow 0$, also die Ableitung von $f(x+tu)$ nach t an der Stelle $t=0$, ist dann die gesuchte Richtungsableitung.



DEFINITION: Die **Richtungsableitung** einer Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ an der Stelle $x \in D$ nach dem Vektor $u \in \mathbb{R}^n$ ist

$$\partial_u f(x) := \lim_{h \rightarrow 0} \frac{1}{h} [f(x+hu) - f(x)] = \left. \frac{d}{dt} \right|_{t=0} f(x+tu) \in \mathbb{R},$$

wenn x innerer Punkt von D ist und der Limes existiert. Ist u ein Einheitsvektor (also ein Vektor mit Euklidischer Norm 1), so heißt der Wert $\partial_u f(x)$ der Richtungsableitung an der Stelle x in Richtung u auch die **Steigung der Funktion in Richtung u** (auch $\frac{\partial f}{\partial u}(x)$ notiert statt $\partial_u f(x)$). Die **Richtungsableitung von Vektorfunktionen** $F = (f_1, \dots, f_m): \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ werden komponentenweise erklärt, also ist $\partial_u F(x) = (\partial_u f_1(x), \dots, \partial_u f_m(x)) \in \mathbb{R}^m$, wenn für jede Komponentenfunktion f_i die Richtungsableitung nach dem Vektor u an der Stelle x definiert ist. Ist D offen und $\partial_u f(x)$ bzw. $\partial_u F(x)$ an allen Stellen $x \in D$ definiert, so erhält man entsprechend die **Richtungsableitungsfunktion** $\partial_u f: D \ni x \mapsto \partial_u f(x) \in \mathbb{R}$ bzw. $\partial_u F: D \rightarrow \mathbb{R}^m$ auf D . ■

DISKUSSION: 1) Vergleich mit der Definition der partiellen Ableitungen zeigt:

- Die partiellen Ableitungen sind die Richtungsableitungen in Richtung der kanonischen Basisvektoren,

$$\frac{\partial f}{\partial x_j}(x) = \partial_{e_j} f(x) \quad \text{für } j = 1 \dots n.$$

2) *Geometrische Interpretation:* Die Richtungsableitung $\partial_u f(x)$ gibt, wenn u ein Einheitsvektor ist, die *Steigung des Funktionsgraphen* im Punkt $(x, f(x))$ in Richtung von u an. Wenn man sich den Graphen als Oberfläche eines Gebirges vorstellt, wobei $f(x)$ die Höhe über jedem Punkt x der Basis D angibt, so ist das diejenige Steigung, die ein auf dem Gebirge im Punkt $(x, f(x))$ befindlicher Wanderer feststellt, wenn er horizontal in Richtung u blickt. Obwohl es in der Definition nicht vorausgesetzt ist, spricht man von Richtungsableitungen im engeren Sinne meist nur, wenn der Vektor u die Euklidische Norm (Länge) $|u| = 1$ hat. (Zur Erinnerung: $|u|$ ist für Vektoren $u \in \mathbb{R}^n$ definiert als die Wurzel aus der Summe der Quadrate aller n Komponenten des Vektors, also $|u| = \sqrt{(u_1)^2 + \dots + (u_n)^2} = 1$.) Jedenfalls

- ist die Normierung der Vektoren auf Länge $|u| = 1$, unbedingt erforderlich, wenn man die Steigung $\partial_u f(x)$ in Richtung von u berechnen will;

denn wenn man von u zu einem Vektor su mit gleicher (oder entgegengesetzter, wenn $s < 0$) Richtung wie u übergeht, so ändert sich der Wert der Richtungsableitung auch um den Faktor $s \in \mathbb{R}$, d.h. es gilt $\partial_{su} f(x) = s \cdot \partial_u f(x)$. Um die Steigung in einer Richtung also eindeutig festzulegen, müssen wir für jede Richtung, d.h. für jeden Strahl $\mathbb{R}_{>0}v$ aus dem Ursprung des \mathbb{R}^n , einen eindeutigen Vektor wählen, der diese Richtung repräsentiert. Ein (*Ursprungs-*)*Strahl* besteht definitionsgemäß aus allen Vielfachen rv eines festen Vektors $v \neq 0$ mit positiven Faktoren r , und eine natürliche Wahl ist eben der eindeutige Einheitsvektor $u = \frac{1}{|v|}v$ auf diesem Strahl. Einheitsvektoren, also Vektoren $u \in \mathbb{R}^n$ der Euklidischen Länge $|u| = 1$, nennen wir daher auch **Richtungsvektoren** (im eigentlichen Sinn).

3) *Ökonomische Interpretation:* Für $u = (u_1, \dots, u_n) \in \mathbb{R}^n$ gilt:

- Der Wert der Richtungsableitung $\partial_u f(x)$ gibt an, um wieviele Einheiten sich der Funktionswert $f(x)$ ungefähr ändert, wenn man simultan die Werte der Variablen x_1, x_2, \dots bzw. x_n um u_1, u_2, \dots bzw. u_n Einheiten verändert.

Jedenfalls gilt das, wenn die Änderungen u_1, u_2, \dots, u_n der unabhängigen Variablen im Betrag als klein angesehen werden können. Genauer ist es so, dass die Änderung $f(x+hu) - f(x)$ des Funktionswertes beim Übergang von $x = (x_1, \dots, x_n)$ zu $x+hu = (x_1+hu_1, \dots, x_n+hu_n)$ approximativ gleich $h \cdot \partial_u f(x)$ ist, wobei die Approximation umso besser ausfällt, je kleiner $|h|$ ist. Das gilt auch, wenn u kein Einheitsvektor ist, und es wäre bei dieser ökonomischen Interpretation auch nicht zweckmäßig, die Definition der Richtungsableitungen auf Vektoren u der Norm 1 zu beschränken. Wenn man allerdings die **Änderungsrate (Änderungsgeschwindigkeit)** der Funktion **in Richtung eines Vektors** u angeben will, so ist die Normierung $|u| = 1$ erforderlich oder man muss $\partial_u f(x)$ ins Verhältnis zur Länge $|u|$ setzen:

- Die Änderungsrate (Änderungsgeschwindigkeit) der Funktion f an einer Stelle x in Richtung eines Vektors $0 \neq u \in \mathbb{R}^n$ ist der Wert der Richtungsableitung $\partial_u f(x)$, wenn u ein Einheitsvektor ist, und $\frac{1}{|u|}\partial_u f(x)$ im allgemeinen Fall.

Das **Vorzeichen der Richtungsableitung** $\partial_u f(x)$ bestimmt dabei, ob die Funktionswerte $f(x)$ bei kleiner Verschiebung der Stelle x in Richtung des Vektors u zunehmen ($\partial_u f(x) > 0$) oder abnehmen ($\partial_u f(x) < 0$). Im Fall $\partial_u f(x) = 0$ verändern sich die Funktionswerte bei solchen kleinen Verschiebungen praktisch überhaupt nicht.

4) Die Richtungsableitung $\partial_u f(x)$ ist die Ableitung der Funktion f auf der Geraden durch x mit Richtung von u , und diese Gerade hat dieselbe Dimension 1 wie die Zahlengerade. Daher läßt sich diese Ableitung auch als Ableitung einer Funktion von *einer* reellen Veränderlichen t berechnen, nämlich $\partial_u f(x) = \left. \frac{d}{dt} f(x+tu) \right|_{t=0}$ wie in der Definition. Eine Konsequenz ist:

- Für Richtungsableitungen gelten die Ableitungsrechenregeln aus der Differentialrechnung für Funktionen einer Veränderlichen, also die Summen- und Faktorregel, die Produktregel, die Quotientenregel und die Kettenregel, analog.

Das bedeutet also (für $r \in \mathbb{R}$ und f, g, x, u wie oben und $h: \mathbb{R} \rightarrow \mathbb{R}$, wobei natürlich $g(x) \neq 0$ sein muss bei der Quotientenregel):

$$\begin{aligned}\partial_u(rf)(x) &= r\partial_u f(x), \\ \partial_u(f \pm g)(x) &= \partial_u f(x) \pm \partial_u g(x), \\ \partial_u(fg)(x) &= (\partial_u f)(x)g(x) + f(x)\partial_u g(x), \\ \partial_u(f/g)(x) &= \frac{g(x)\partial_u f(x) - f(x)\partial_u g(x)}{g(x)^2}, \\ \partial_u(h \circ f)(x) &= h'(f(x)) \cdot \partial_u f(x).\end{aligned}$$

5) Weil die Richtungsableitung $\partial_u F$ von Vektorfunktionen $F = (f_1, \dots, f_m)$ komponentenweise erklärt sind, gelten die Summen- und Faktorregel hierfür ebenfalls. Außerdem hat man auch eine Produktregel $\partial_u(F * G) = (\partial_u F) * G + F * (\partial_u G)$ für beliebige Produkte $*$: $\mathbb{R}^m \times \mathbb{R}^l \rightarrow \mathbb{R}^k$. Das ist ganz so wie in 5.1 bei den partiellen Ableitungen $\frac{\partial}{\partial x_j}(F * G) = \partial_{e_j}(F * G)$, und wir brauchen darauf nicht weiter einzugehen. Auch die in 5.1 gemachte Bemerkung, dass man sich bei der Differentiation von Vektorfunktion zu allererst über das Format des Ergebnisses im Klaren sein sollte, gilt hier genauso:

- Die Richtungsableitung $\partial_u F$ einer Vektorfunktion F hat Werte in demselben Vektorraum \mathbb{R}^m , in dem auch F seine Werte nimmt. ■

Wie bei den partiellen Ableitungen gibt es auch bei Richtungsableitungen eine von Einheitenwahlen unabhängige und daher besonders für ökonomische Anwendungen geeignete Variante. Dabei wird die relative (d.h. prozentuale) Änderung des Funktionswertes $f(x) = f(x_1, \dots, x_n)$ ins Verhältnis gesetzt zu den relativen Änderungen der unabhängigen Variablen x_j in Richtung eines gegebenen n -gliedrigen Vektors. Um dabei von der Einheitenwahl bei jeder einzelnen Variablen x_j unabhängig zu sein, dürfen wir aber nun nicht annehmen, dass die Zuwächse bei den x_j in vorgegebenen Verhältnissen zueinander stehen, wie das beim Übergang von x zu $x + hu$ der Fall ist, sondern wir müssen solche Verhältnisse $\frac{hu_1}{x_1} : \frac{hu_2}{x_2} : \dots : \frac{hu_n}{x_n} = v_1 : v_2 : \dots : v_n$ für die *relativen* Variablenzuwächse vorschreiben. Dies bedeutet, dass der Zuwachs bei x_j durch $hx_j v_j$ gegeben ist, dass wir also die Stelle x in Richtung des Vektors $u = (x_1 v_1, x_2 v_2, \dots, x_n v_n)$ variieren müssen. Beim Übergang von x zu $x + hu$ ist dann die relative Änderung von

x_j gleich $\frac{x_j+hu_j-x_j}{x_j} = hv_j$, d.h. x_j wird um $h \cdot v_j$ % geändert. Nehmen wir die Zahl $h \in \mathbb{R}_{\neq 0}$ als Maß für die Größe der relativen Änderung des vektoriellen Arguments x und setzen die relative Änderung von $f(x)$ dazu ins Verhältnis, so erhalten wir den Quotienten $\frac{f(x+hu)-f(x)}{hf(x)}$ mit dem Grenzwert $\frac{1}{f(x)}\partial_u f(x)$ als von Einheiten unabhängiges Maß der Funktionswertänderung zu dem gegebenen "Prozentsatzvektor" $v = (v_1, \dots, v_n)$.

DEFINITION: Die **Richtungselastizität** einer Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ an der inneren Stelle $x \in D$ bzgl. des Vektors $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ ist

$$\varepsilon_{f,v}(x) := \frac{\left. \frac{d}{dt} \right|_{t=0} f(x_1+tx_1v_1, \dots, x_n+tx_nv_n)}{f(x)} = \frac{\partial_u f(x)}{f(x)} \quad \text{mit } u = (x_1v_1, \dots, x_nv_n),$$

wenn x innerer Punkt von D ist und $\partial_u f(x)$ existiert. (Wie immer bei Elastizitäten wird hierbei natürlich auch vorausgesetzt, dass $f(x) > 0$ ist.) Die **Richtungselastizität von Vektorfunktionen** $F = (f_1, \dots, f_m): \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ wird komponentenweise erklärt, also ist $\varepsilon_{F,v}(x) = (\varepsilon_{f_1,v}(x), \dots, \varepsilon_{f_m,v}(x)) \in \mathbb{R}^m$, wenn die Elastizitäten $\varepsilon_{f_i,v}(x)$ definiert sind für $i = 1 \dots m$. Ist D offen und sind die Richtungselastizitäten bzgl. v an allen Stellen $x \in D$ definiert, so erhält man entsprechend **Richtungselastizitätsfunktionen** $\varepsilon_{f,v}: D \rightarrow \mathbb{R}$ bzw. $\varepsilon_{F,v}: D \rightarrow \mathbb{R}^m$. ■

DISKUSSION: 1) Für kanonische Basisvektoren $v = e_j$ und $x \in \mathbb{R}^n$ ist in der Definition $u = (x_1v_1, \dots, x_nv_n) = x_j e_j$ und $\partial_u f(x) = x_j \frac{\partial f}{\partial x_j}(x)$.

- Also sind die die Richtungselastizitäten bzgl. der der kanonischen Basisvektoren nichts anderes als die partiellen Elastizitäten:

$$\varepsilon_{f,x_j}(x) = \varepsilon_{f,e_j}(x) \quad \text{für } j = 1 \dots n;$$

2) **Ökonomische Interpretation:** Für eine positive Funktion von n Variablen und einen gegebenen "Prozentsatzvektor" $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ gilt:

- **Der Wert der Richtungselastizität** $\varepsilon_{f,v}(x)$ gibt an, um wieviel Prozent sich der Funktionswert $f(x)$ ungefähr ändert, wenn man simultan die Werte der Variablen x_1, x_2, \dots bzw. x_n um v_1 %, v_2 %, ... bzw. v_n % verändert.

Jedenfalls gilt das, wenn die Prozentsätze v_1, v_2, \dots, v_n der Änderung der unabhängigen Variablen im Betrag als klein angesehen werden können (gegenüber 100 %). Genauer ist es so, dass bei simultaner Veränderung der x_j um $h \cdot v_j$ %, also beim Übergang von $x = (x_1, \dots, x_n)$ zu $x + \frac{h}{100}u = (x_1 + \frac{h}{100}v_1x_1, \dots, x_n + \frac{h}{100}v_nx_n)$, die relative Änderung $\frac{f(x + (h/100)u) - f(x)}{f(x)}$ des Funktionswertes approximativ gleich $h \cdot \varepsilon_{f,v}(x)$ % ist, wobei die Approximation umso besser ausfällt, je kleiner $|h|$ ist.

Das **Vorzeichen der Richtungselastizität** $\varepsilon_{f,v}(x)$ bestimmt dabei, ob die Funktionswerte $f(x)$ zunehmen ($\varepsilon_{f,v}(x) > 0$) oder abnehmen ($\varepsilon_{f,v}(x) < 0$) bei kleiner Veränderung der Stelle x entsprechend den Komponentenverhältnissen des Prozentsatzvektors v . Im Fall $\varepsilon_{f,v}(x) = 0$ verändern sich die Funktionswerte bei solchen kleinen Änderungen von x praktisch überhaupt nicht.

3) Mit der Terminologie “elastisch”, “ausgeglichen elastisch”, “unelastisch” im Zusammenhang mit Richtungselastizitäten gibt es Schwierigkeiten: “Elastisch” würde man eine Funktion f an einer Stelle x bzgl. eines Vektors v nennen, wenn kleine relative Änderungen von x gemäß dem Prozentsatzvektor v größere relative Änderungen des Wertes $f(x)$ bewirken, d.h. wenn der Betrag der Elastizität $|\varepsilon_{f,v}(x)|$ die Größe des Prozentsatzvektors v übersteigt. Man müsste dazu also eine reelle Zahl $\|v\|$ als Größe des Prozentsatzvektors $v \in \mathbb{R}^n$ festlegen und überprüfen, ob $|\varepsilon_{f,v}(x)|$ größer oder kleiner als $\|v\|$ ist; aber für die Festlegung einer solchen Größenmessung bei Vektoren gibt es keine eindeutige sinnvolle Vorschrift.

Das Problem wird deutlich anhand der folgenden konkreten Frage: Wenn man jede Komponente eines Punktes $(x, y) \in \mathbb{R}_{>0}^2$ um 1% erhöht, ist das eine größere oder kleinere relative Änderung, als wenn man eine Komponente um 2% vergrößert und die andere unverändert lässt?

Bei den partiellen Elastizitäten haben wir im Prinzip 1 als Größe von $v = e_j$ definiert und die Größe der relativen Änderung beim Übergang von x_j zu $x_j + tx_j$ unter c.p.-Bedingung mit $|t|$ angesetzt; dann ist f elastisch bzgl. der j -ten Variablen, genau wenn $|\varepsilon_{f,x_j}(x)| > 1$ ist. Im allgemeinen Fall könnte man als Größe $\|v\|$ von v die Euklidische Norm $|v|$ nehmen, was aus Sicht ökonomischer Anwendungen wenig sinnvoll scheint, oder z.B. die Betragssumme $|v|_1 := \sum_{j=1}^n |v_j|$ aller Einträge oder auch ihr Maximum $|v|_\infty := \max_{j=1}^n |v_j|$, was in manchem ökonomischen Kontext nahe liegt. Die Größe der relativen Änderung von x beim simultanen Übergang von x_j zu $tv_j x_j$ für $j = 1 \dots n$ ist dann das $|t|$ -fache der festgelegten Größe $\|v\|$ von v . Bei allen drei Größenmessungen für v ergibt sich im Fall kanonischer Basisvektoren die Größe 1. Die Größe der relativen Änderung von $(x, y) \in \mathbb{R}_{>0}^2$ bei Erhöhung beider Komponenten um jeweils 1% wäre aber bei Wahl der Euklidischen Norm mit $|(1, 1)| = \sqrt{2}$ anzusetzen, bei Wahl der Summen-Norm mit $|(1, 1)|_1 = 2$ und bei Wahl der Maximum-Norm mit $|(1, 1)|_\infty = 1$. Dementsprechend wäre eine dabei festgestellte relative Änderung der Funktionswerte um $\sqrt{2}\%$ als elastisches Funktionsverhalten zu qualifizieren, wenn man die Maximum-Norm zu Grunde legt, als unelastisches Verhalten, wenn man die Summen-Norm festgelegt hat, und als ausgeglichen elastisches Verhalten, wenn man sich für die Euklidische Norm entschieden hat.

Die Konsequenz aus alledem ist, dass wir nicht von elastischem oder unelastischem Funktionsverhalten einer Funktion von n Variablen bzgl. eines allgemeinen Vektors $v \in \mathbb{R}^n$ sprechen dürfen (es sei denn, wir haben genau festgelegt, was damit gemeint ist). Die allgemeinen Richtungselastizitäten $\varepsilon_{f,v}(x)$ sind aber sinnvoll und haben ökonomische Relevanz gemäß 2).

4) Aus der Definition $\varepsilon_{f,v}(x) = \frac{1}{f(x)} \partial_u f(x)$ mit $u = (x_1 v_1, \dots, x_n v_n)$ und den Rechenregeln für Richtungsableitungen ergeben sich **Rechenregeln für Richtungselastizitäten**, die völlig analog zu den in 5.1 angegebenen für partielle Elastizitäten sind. (Man muss dort nur ε_{f,x_j} überall durch $\varepsilon_{f,v}$ ersetzen.) Wir ersparen uns die Details und vermerken nur:

- Für Richtungselastizitäten gelten die Ableitungsrechenregeln aus der Differentialrechnung für Funktionen einer Veränderlichen analog.

Für Vektorfunktionen $F = (f_1, \dots, f_m)$ mit positiven Komponentenfunktionen ergeben sich entsprechende Rechenregeln dann unmittelbar aus der komponentenweisen Definition $\varepsilon_{F,v} = (\varepsilon_{f_1,v}, \dots, \varepsilon_{f_m,v})$. ■

BEISPIELE: (1) Wir betrachten $f(x, y) = x^2(1 + y)$. Die Richtungsableitung in Richtung des diagonalen Einheitsvektors $u = \frac{1}{\sqrt{2}}(1, 1)$ ist dann

$$\begin{aligned}\partial_u f(x, y) &= \left. \frac{d}{dt} \right|_{t=0} f\left(x + \frac{t}{\sqrt{2}}, y + \frac{t}{\sqrt{2}}\right) = \left. \frac{d}{dt} \right|_{t=0} \left(x + \frac{t}{\sqrt{2}}\right)^2 \left(1 + y + \frac{t}{\sqrt{2}}\right) \\ &= \frac{1}{\sqrt{2}} 2x(1 + y) + x^2 \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} (2x + 2xy + x^2)\end{aligned}$$

und die Richtungsableitung in Richtung des Einheitsvektors $v = \left(\frac{3}{5}, \frac{4}{5}\right)$ ist

$$\begin{aligned}\partial_v f(x, y) &= \left. \frac{d}{dt} \right|_{t=0} f\left(x + \frac{3}{5}t, y + \frac{4}{5}t\right) = \left. \frac{d}{dt} \right|_{t=0} \left(x + \frac{3}{5}t\right)^2 \left(1 + y + \frac{4}{5}t\right) \\ &= \frac{3}{5} 2x(1 + y) + \frac{4}{5} x^2.\end{aligned}$$

Für allgemeine Vektoren $w = (w_1, w_2)$ ergibt sich analog

$$\partial_w f(x, y) = \left. \frac{d}{dt} \right|_{t=0} (x + tw_1)^2 (1 + y + tw_2) = w_1 2x(1 + y) + w_2 x^2.$$

Wollen wir näherungsweise bestimmen, um wieviel Prozent sich der Funktionswert an einer Stelle $(x, y) \in \mathbb{R}_{>0}$ ändert, wenn man x und y um jeweils 1% vergrößert, so haben wir die Richtungselastizität bzgl. des Vektors $v = (1, 1)$ zu berechnen. Dazu müssen wir $w = (xv_1, yv_2) = (x, y)$ betrachten und haben

$$\varepsilon_{f, (1,1)}(x, y) = \frac{1}{f(x, y)} \partial_w f(x, y) = \frac{2x^2(1 + y) + yx^2}{x^2(1 + y)} = \frac{2 + 3y}{1 + y} \quad (w = (x, y)).$$

Für $y = 1$ und beliebige $x > 0$ ist also die prozentuale Erhöhung des Funktionswerts ca. 2.5%. Dass dieses Ergebnis nicht von x abhängt, ergibt sich auch aus der Produktregel für Elastizitäten: $\varepsilon_{x^2(1+y), (1,1)} = (\varepsilon_{x, (1,1)})^2 \cdot \varepsilon_{1+y, (1,1)}$, wobei $\varepsilon_{x, (1,1)} = \frac{1}{x} \left. \frac{\partial}{\partial t} \right|_{t=0} (x + tx) \equiv 1$ ist und $\varepsilon_{1+y, (1,1)} = \frac{1}{1+y} \left. \frac{\partial}{\partial t} \right|_{t=0} (1 + x + ty) = \frac{y}{1+y}$ nur von y abhängt.

(2) Die Funktion $f(x, y, z) = x^3 e^{yz}$ hat in Richtung von $u = \frac{1}{\sqrt{3}}(1, 1, 1)$ die Ableitung

$$\partial_u f(x, y, z) = \left. \frac{d}{dt} \right|_{t=0} \left(x + \frac{1}{\sqrt{2}}t\right)^3 e^{(y+t/\sqrt{2})(z+t/\sqrt{2})} = \frac{1}{\sqrt{3}} e^{yz} (3x^2 + x^3 z + x^3 y),$$

und für allgemeine Vektoren $w = (w_1, w_2, w_3) \in \mathbb{R}^3$

$$\partial_w f(x, y, z) = \left. \frac{d}{dt} \right|_{t=0} (x + tw_1)^3 e^{(y+tw_2)(z+tw_3)} = (w_1 3x^2 + w_2 x^3 z + w_3 x^3 y) e^{yz}.$$

Erniedrigt man $x = 2$ um 1% und erhöht $y = 1$ um 2%, ohne den Wert von $z = \ln 2$ zu ändern, so ist die prozentuale Änderung des Funktionswerts $f(2, 1, \ln 2) = 16$ näherungsweise (da man hier gemäß Definition der Richtungselastizität nach dem Vektor $w = (x(-1), y2, z0) = (-2, 2, 0)$ zu differenzieren hat):

$$\varepsilon_{f, (-2, 2, 0)}(2, 1, \ln 2) = \frac{1}{16} \partial_{(-2, 2, 0)} f(2, 1, \ln 2) = -3 + 2 \ln 2 \approx -1.61.$$

Insbesondere nimmt der Funktionswert bei dieser Änderung ab, weil das Ergebnis negativ ist.

(3) Für lineare Funktionen $\ell(x) = \mathbf{a} \cdot x + b = a_1x - 1 + \dots + a_nx_n + b$ ist die Richtungsableitung bzgl. $u \in \mathbb{R}^n$:

$$\partial_u \ell(x) = \left. \frac{d}{dt} \right|_{t=0} \ell(x + tu) = \left. \frac{d}{dt} \right|_{t=0} [\mathbf{a} \cdot (x + tu) + b] = \left. \frac{d}{dt} \right|_{t=0} [t(\mathbf{a} \cdot u) + \mathbf{a} \cdot x + b] = \mathbf{a} \cdot u.$$

Für \mathbb{R}^m -wertige lineare Vektorfunktionen $L(x) = Ax + \mathbf{b}$ auf \mathbb{R}^n , $A \in \mathbb{R}^{m \times n}$ eine $m \times n$ -Matrix und $\mathbf{b} \in \mathbb{R}^m$, lautet die analoge Rechnung:

$$\partial_u L(x) = \left. \frac{d}{dt} \right|_{t=0} [A(x + tu) + \mathbf{b}] = \left. \frac{d}{dt} \right|_{t=0} [tAu + Ax + \mathbf{b}] = Au,$$

da ja komponentenweise differenziert wird. Das Resultat ist:

- Jede Richtungsableitung einer linearen Funktion ist konstant, und zwar ist
- im skalaren Fall $\partial_u \ell \equiv \mathbf{a} \cdot u$ das Skalarprodukt des Koeffizientenvektors mit dem Richtungsvektor
- und im vektoriellen Fall $\partial_u L(x) \equiv Au$ das Produkt der Koeffizientenmatrix mit dem Richtungsvektor.

(4) Der **Ortsvektor** an einer Stelle $x \in \mathbb{R}^n$ ist der Vektor $u = x$ (aufgefasst als ein Richtungsvektor). Die Richtungsableitung einer Funktion f an der Stelle x in dieser Richtung $u = x$ ist

$$\partial_x f(x) = \left. \frac{d}{dt} \right|_{t=0} f(x + tx) = \left. \frac{d}{dt} \right|_{t=0} f((1+t)x) = \left. \frac{d}{ds} \right|_{s=1} f(sx)$$

und heißt **Ableitung nach dem Ortsvektor**, in ökonomischer Terminologie auch **Skalenableitung** oder **Niveauableitung** der Funktion f an der Stelle x . Diese Ausdrücke kommen daher, dass man eine Änderung $x \mapsto sx$ mit positivem Faktor s als *Skalenvariation* oder *Niveauvariation* des Vektors x bezeichnet. Eine solche Änderung läuft nämlich darauf hinaus, dass man nur die Skala, mit der die Größe der Einträge des Vektors x gemessen wird, für alle Komponenten um denselben Faktor abändert bzw. dass man nur das "Niveau" des Vektors x , wie in der Ökonomie gelegentlich seine Euklidische Norm $|x| = \sqrt{(x_1)^2 + \dots + (x_n)^2}$ genannt wird, verändert, ohne die Verhältnisse zwischen seinen Einträgen zu ändern. Wenn die unabhängigen Variablen x_j im ökonomischen Kontext "Faktoren" heißen, wie etwa bei Produktionsfunktionen, so wird auch von einer *proportionalen Faktorkomplex-Variation* gesprochen, weil alle Einträge des Faktorkomplexes (Faktorkomplex) um denselben Proportionalitätsfaktor abgeändert werden.

Der Ortsvektor x ist *radial gerichtet*, d.h. er hat die Richtung des Pfeils vom Ursprung des \mathbb{R}^n zum Punkt x , er ist aber im Allgemeinen kein Einheitsvektor (außer wenn x den Abstand 1 vom Ursprung hat). Will man die Steigung für die Funktion f in radialer Richtung bestimmen, so muss man nach dem entsprechenden *radialen Einheitsvektor* $\frac{1}{|x|}x$ differenzieren (wenn $x \neq 0$). Dies ist die sog. **radiale Ableitung**, die sich an der Stelle $x \neq 0$ von der Skalen-Ableitung um den Faktor $\frac{1}{|x|}$ unterscheidet:

$$\partial_{\text{rad}} f(x) = \partial_{x/|x|} f(x) = \frac{1}{|x|} \partial_x f(x).$$

Sie hat die Interpretation der **Steigung in radialer Richtung** von der Funktion f an der Stelle $x \neq 0$ bzw. der **Änderungsrate in radialer Richtung**. Die Änderung des Funktionswertes $f(x)$ ist also ungefähr $h \cdot \partial_{\text{rad}} f(x)$, wenn man den Argumentvektor x durch einen kleinen radial gerichteten Zuwachs $\frac{h}{|x|}x$ der Länge $|h|$ abändert. Bei Änderung von x um den Zuwachs hx beträgt die Änderung von $f(x)$ dagegen ungefähr $h \cdot \partial_x f(x) = h|x| \cdot \partial_{\text{rad}} f(x)$.

Die Elastizität an der Stelle x bzgl. Skalenvariationen heißt **Skalen–Elastizität** oder **Niveau–Elastizität** von f an der Stelle x und wird $\sigma_f(x)$ bezeichnet. Da Skalen–Variationen bedeuten, dass alle Komponenten von x um denselben Prozentsatz abgeändert werden, ist dies nichts anderes als die Richtungselastizität bzgl. des Vektors $v = (1, 1, \dots, 1)$ mit lauter Komponenten 1.

$$\sigma_f(x) = \frac{1}{f(x)} \left. \frac{d}{ds} \right|_{s=1} f(sx) = \frac{1}{f(x)} \partial_x f(x) = \varepsilon_{f, (1,1,\dots,1)}(x).$$

Die *ökonomische Bedeutung der Skalen–Elastizität* ist, dass der Wert $\sigma_f(x)$ die ungefähre prozentuale Änderung von $f(x)$ angibt, wenn alle (positiv angenommenen) Komponenten x_j simultan um 1% vergrößert werden.

Konkret ist z.B. für eine Cobb–Douglas–Funktion $f(x) = cx_1^{r_1} x_2^{r_2} \cdot \dots \cdot x_n^{r_n}$ die Skalenableitung

$$\begin{aligned} |x| \partial_{\text{rad}} f(x) = \partial_x f(x) &= \left. \frac{d}{ds} \right|_{s=1} [c(sx_1^{r_1} (sx_2)^{r_2} \cdot \dots \cdot (sx_n)^{r_n})] \\ &= \left. \frac{d}{ds} \right|_{s=1} s^{r_1+r_2+\dots+r_n} f(x) = (r_1 + r_2 + \dots + r_n) f(x), \end{aligned}$$

und die Skalenelastizität ist

$$\sigma_f(x) = \frac{1}{f(x)} \partial_x f(x) = r_1 + r_2 + \dots + r_n \quad \text{konstant.} \quad \blacksquare$$

Man sieht an diesen Beispielen, dass man bei elementaren Funktionen alle Richtungsableitungen und Richtungselastizitäten mit Differentiation von Funktionen *einer* Veränderlichen konkret ausrechnen kann. Aber wenn wir die Richtung des stärksten Anstiegs der Funktion herausfinden wollen — das ist schließlich eine Motivation für das Ganze —, so müssen wir die Richtungsableitung in *jeder* Richtung berechnen und der Größe nach miteinander vergleichen. Nun gibt es unendlich viele Richtungen und wir müssten demnach unendlich viele Richtungsableitungen berechnen — eine kaum zu bewältigende Aufgabe.

Zum Glück aber sind bei “normalen” Funktionen die Richtungsableitungen $\partial_u f(x)$ an einer festen Stelle $x \in \mathbb{R}^n$ für die verschiedenen Vektoren $u \in \mathbb{R}^n$ nicht unabhängig voneinander, sondern wenn man sie für n linear unabhängige Vektoren berechnet hat, so kann man alle anderen Richtungsableitungen an derselben Stelle in einfacher Weise dadurch ausdrücken. Insbesondere kann man für “normale” Funktionen alle Richtungsableitungen an einer Stelle durch die partiellen Ableitungen an dieser Stelle ausdrücken. Um das Ergebnis zu motivieren, schreiben wir $f(x + hu) - f(x)$ als Summe von Differenzen

$$f(x_1 + hu_1, \dots, x_j + hu_j, x_{j+1}, \dots, x_n) - f(x_1 + hu_1, \dots, x_{j-1} + hu_{j-1}, x_j, \dots, x_n)$$

mit einem Unterschied allein beim j -ten eingesetzten Argument. Für kleine $|h|$ ist diese Differenz ungefähr gleich $hu_j \frac{\partial f}{\partial x_j}(x_1 + hu_1, \dots, x_{j-1} + hu_{j-1}, x_j, \dots, x_n)$, und wenn $\frac{\partial f}{\partial x_j}$ stetig ist, so ist dies wiederum ungefähr gleich $hu_j \frac{\partial f}{\partial x_j}(x_1, \dots, x_n)$, so dass Aufsummieren für $f(x + hu) - f(x)$ den Näherungswert $h \sum_{j=1}^n u_j \frac{\partial f}{\partial x_j}(x)$ gibt. Dies gilt umso genauer, je kleiner $|h|$ ist, so dass man für $\partial_u f(x) = \lim_{h \rightarrow 0} \frac{1}{h} [f(x + hu) - f(x)]$ die Formel $\sum_{j=1}^n u_j \frac{\partial f}{\partial x_j}(x)$ erhält. Diese Formel ist der wesentliche Inhalt des folgenden Satzes. Sie zeigt, wie man die Richtungsableitungen durch die partiellen Ableitungen ausdrücken kann, und sie lässt erkennen, dass die Richtungsableitungen linear vom Richtungsvektor abhängen, dass also $\partial_{(ru+sv)} f(x) = r \partial_u f(x) + s \partial_v f(x)$ gilt für $u, v \in \mathbb{R}^n$ und $r, s \in \mathbb{R}$.

Die skizzierte Herleitung lässt sich mathematisch streng durchführen, wenn die partiellen Ableitungen von f an der Stelle x stetig sind. Diese Bedingung ist für alle elementaren Funktionen von mehreren Veränderlichen und allgemeiner für alle im richtigen (hier nicht weiter erklärten) Sinne “differenzierbaren” Funktionen von mehreren Veränderlichen erfüllt. Für die Mathematik in den Wirtschaftswissenschaften bedeutet es, dass man die Gültigkeit des Satzes praktisch für alle dort behandelten Funktionen annehmen kann. Da Vektorfunktionen komponentenweise differenziert werden, gilt der Satz nicht nur für skalare Funktionen, sondern automatisch auch für Vektorfunktionen.

SATZ (Richtungsableitungen und partielle Ableitungen): *Hat die reelle Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ auf einer Umgebung des Punktes $x \in D$ stetige partielle Ableitungen, so existieren die Richtungsableitungen $\partial_u f(x)$ an der Stelle x bzgl. aller Vektoren $u \in \mathbb{R}^n$ und die Richtungsableitungen hängen linear vom Richtungsvektor ab. Sie lassen sich folglich als Linearkombinationen der partiellen Ableitungen von f an dieser Stelle darstellen, wobei die Koeffizienten gerade die Komponenten des Vektors $u = (u_1, \dots, u_n)$ sind:*

$$\partial_u f(x) = u_1 \frac{\partial f}{\partial x_1}(x) + u_2 \frac{\partial f}{\partial x_2}(x) + \dots + u_n \frac{\partial f}{\partial x_n}(x).$$

Dasselbe gilt für die Richtungsableitungen $\partial_u F(x) \in \mathbb{R}^m$ von Vektorfunktionen $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$, wenn deren Komponentenfunktionen auf einer Umgebung der Stelle x stetige partielle Ableitungen haben.

Der Nutzen des Satzes ist offensichtlich: Sobald man die partiellen Ableitungen berechnet hat, erhält man alle Richtungsableitungen praktisch gratis, ohne weiteren Rechenaufwand! Man muss ja nur die partiellen Ableitungen linear kombinieren mit den Einträgen des Richtungsvektors als Koeffizienten.

Für die Richtungselastizitäten folgt aus der Berechnungsformel $\varepsilon_{f,v}(x) = \frac{1}{f(x)} \partial_u f(x)$ mit $u = (x_1 v_1, \dots, x_n v_n)$ unmittelbar eine analoge Aussage:

KOROLLAR (Richtungselastizitäten und partielle Elastizitäten): *Auch die Richtungselastizitäten hängen linear vom Richtungsvektor ab. Folglich lassen sich die Richtungselastizitäten $\varepsilon_{f,v}(x)$ als Linearkombination der partiellen Elastizitäten darstellen, wobei die Koeffizienten gerade die Komponenten des Vektors $v = (v_1, \dots, v_n)$ sind:*

$$\varepsilon_{f,v}(x) = v_1 \varepsilon_{f,x_1}(x) + v_2 \varepsilon_{f,x_2}(x) + \dots + v_n \varepsilon_{f,x_n}(x).$$

Auch hier gilt dasselbe für die Richtungselastizitäten $\varepsilon_{f,v}(x) \in \mathbb{R}^m$ von \mathbb{R}^m -wertige Vektorfunktionen.

BEMERKUNGEN: 1) Ist u ein Richtungsvektor im engeren Sinne, also ein Einheitsvektor ($|u| = 1$), so heißen die Komponenten u_j auch die *Richtungskosinusse* von u , weil sie sich als Kosinus des Winkels zwischen der Richtung von u und einer Achsenrichtung interpretieren lassen. Damit lässt sich der Satz für skalare Funktionen f so umformulieren:

- Die Steigung von f an der Stelle x in Richtung u ist die Linearkombination der partiellen Ableitungen in x mit den Richtungskosinussen von u als Koeffizienten.

2) *Stetigkeit bei Funktionen von mehreren Veränderlichen* bedeutet anschaulich, dass sich die Funktionswerte beliebig wenig ändern, wenn man die Änderungen der Argumente hinreichend klein hält. Das wird genauso präzisiert wie bei Funktionen einer Veränderlichen: Eine Funktion $g: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ (oder $\rightarrow \mathbb{R}^m$) heißt stetig an der Stelle $x \in D$, wenn es zu jeder gegebenen Fehlerschranke $\varepsilon > 0$ ein $\delta > 0$ gibt, derart dass der Abstand der Funktionswerte $|g(\tilde{x}) - g(x)| < \varepsilon$ ausfällt, wenn immer \tilde{x} ein Punkt aus D ist mit Abstand $|\tilde{x} - x| < \delta$ zu x . In der Mathematik wird gezeigt, dass alle elementaren Funktionen von mehreren Veränderlichen auf ihren offenen Definitionsbereichen stetig sind (d.h. an jeder Stelle des Definitionsbereichs). Da elementare Funktionen auf offenen Mengen dort auch elementare partielle Ableitungen besitzen (siehe 5.1), ist die Stetigkeitsvoraussetzung des obigen Satzes für elementare Funktionen immer erfüllt. (Ohne die Stetigkeitsvoraussetzung für die partiellen Ableitungen oder eine alternative Ersatzbedingung gilt die Aussage des Satzes übrigens im Allgemeinen nicht.)

3) Mit der Stetigkeit bei Funktionen von mehreren Veränderlichen brauchen wir uns nicht näher zu befassen. Wir erwähnen nur noch folgende quantitative Verschärfung des Stetigkeitskonzepts, die leichter verständlich ist: Eine Funktion $g: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ (oder $\rightarrow \mathbb{R}^m$) heißt **dehnungsbeschränkt** mit **Dehnungsschranke** $L \in \mathbb{R}_{\geq 0}$ (oder *Lipschitz-stetig* mit *Lipschitz-Konstante* L), wenn gilt

$$|g(\tilde{x}) - g(x)| \leq L |\tilde{x} - x| \quad \text{für alle } x, \tilde{x} \in D.$$

Dann wird also der Abstand der Argumente durch Anwendung der Abbildung g höchstens um den Faktor L “gedehnt” (wenn $L \geq 1$ ist; im Fall $0 \leq L \leq 1$ wird er mindestens um den Faktor L “gestaucht”). Dehnungsbeschränkte Funktionen sind offensichtlich stetig (zu $\varepsilon > 0$ kann man $\delta = \frac{1}{L}\varepsilon$ in 2) wählen, bzw. $\delta > 0$ beliebig, wenn $L = 0$ ist), aber Dehnungsbeschränktheit ist eine viel stärkere Eigenschaft als bloße Stetigkeit. Man kann (mit dem Schrankensatz) zeigen, dass jede elementare Funktionen von mehreren Veränderlichen zumindest auf abgeschlossenen Kugeln in ihrem (offenen) Definitionsbereich denungsbeschränkt ist.

4) Das vor dem Satz angegebene Beweisargument lässt sich allgemeiner auch auf Differenzen $f(c(t)) - f(x)$ für Kurven $c(t)$ im Definitionsbereich $D \subset \mathbb{R}^n$ von f mit $c(t_0) = x$ anwenden. Eine Differenz $f(c_1(t), \dots, c_j(t), x_{j+1}, \dots, x_n) - f(c_1(t), \dots, c_{j-1}(t), x_j, \dots, x_n)$ mit einem Unterschied nur beim j -ten Argument ist $\approx (c_j(t) - x_j) \frac{\partial f}{\partial x_j}(x)$ wenn $|c(t) - x|$ klein ist. Division mit $(t - t_0)$ und Summation über $j = 1 \dots n$ gibt unter Berücksichtigung von $\frac{1}{t-t_0}[c_j(t) - x_j] = \frac{1}{t-t_0}[c_j(t) - c_0(t)] \rightarrow c'(t_0)$ bei $t \rightarrow 0$ dann folgendes Ergebnis für die **Ableitung einer Funktion längs einer Kurve**:

$$\left. \frac{d}{dt} \right|_{t=t_0} f(c(t)) = \lim_{t \rightarrow t_0} \frac{f(c(t)) - f(x)}{t - t_0} = c'_1(t_0) \frac{\partial f}{\partial x_1}(x) + \dots + c'_n(t_0) \frac{\partial f}{\partial x_n}(x) = \partial_{c'(t_0)} f(x),$$

wenn $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ (oder $\rightarrow \mathbb{R}^m$) an der Stelle $x \in D$ stetige partielle Ableitungen hat und $c:]\alpha, \beta[\rightarrow D$ eine Kurve ist mit $c(t_0) = x$ und mit Tangentenvektor $c'(t_0) \in \mathbb{R}^n$. Kurz kann man sagen:

- *Die Ableitung einer Funktion längs einer Kurve ist gleich der Richtungsableitung der Funktion in Richtung des Tangentenvektors der Kurve.*

Die Richtungsableitung $\partial_u f(x)$ selbst ist per Definition auch eine Ableitung längs einer Kurve, nämlich der speziellen “Kurve” $c(t) = x + tu$, die eine Gerade parametrisiert. Mit der hier angegebenen Verallgemeinerung kann man auch die Ableitung längs “gekrümmter Kurven” durch Richtungsableitungen und also durch partielle Ableitungen berechnen. ■

Die rechte Seite der Formel, mit der eine Richtungsableitung $\partial_u f(x)$ durch die partiellen Ableitungen $\frac{\partial f}{\partial x_j}(x)$ ausgedrückt wird, kann man als Skalarprodukt von u mit dem Vektor in \mathbb{R}^n schreiben, dessen Komponenten die partiellen Ableitungen von f an der Stelle x sind. Entsprechen kann man für \mathbb{R}^m -wertige Vektorfunktionen $F = (f_1, \dots, f_m)$ die Richtungsableitung $\partial_u F(x)$ als Produkt der $m \times n$ -Matrix $(\frac{\partial f_i}{\partial x_j}(x))$ der partiellen Ableitungen mit dem (Spalten-)Vektor u schreiben. Nun ist es zwar kein Informationsgewinn, wenn man die partiellen Ableitungen zu einem Vektor zusammenfasst, aber wir werden sehen, dass der so gebildete Vektor eine entscheidende Rolle bei der Lösung des Problems spielt, die Richtung der größten Steigung von f an der Stelle x zu bestimmen. Daher erhält dieser Vektor einen besonderen Namen. Mit "Differenzierbarkeit" an einer Stelle x meinen wir im Folgenden, dass alle Richtungsableitungen dort existieren und linear vom Richtungsvektor abhängen. Für elementare Funktionen ist das wie gesagt immer der Fall.

DEFINITION: (i) Ist die skalare Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ differenzierbar an der Stelle $x \in \mathbb{R}^n$, so heißt der (Spalten-)Vektor mit den partiellen Ableitungen von f an dieser Stelle als Komponenten, also der Vektor

$$\text{grad } f(x) = \nabla f(x) := \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} \in \mathbb{R}^n,$$

der **Gradient** von f an der Stelle x . Ist $\nabla f(x)$ für alle x aus der offenen Menge D definiert, so heißt die Vektorfunktion $\nabla f: D \rightarrow \mathbb{R}^n$ das **Gradientenfeld** zu f auf D .

(ii) Ist die Vektorfunktion $F = (f_1, \dots, f_m)$ an der Stelle $x \in \mathbb{R}^n$ differenzierbar, so heißt die $m \times n$ -Matrix aller partiellen Ableitungen der Komponentenfunktionen die **Ableitungsmatrix** (oder **Jacobi-Matrix** oder **Funktionalmatrix**) von F im Punkt x ,

$$\left(\frac{\partial F}{\partial x}(x) \right) := \left(\frac{\partial f_i}{\partial x_j}(x) \right)_{1 \leq i \leq m, 1 \leq j \leq n} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}.$$

(i') In der Situation (i) heißt der (Spalten-)Vektor mit den partiellen Elastizitäten von f an dieser Stelle als Komponenten, also der Vektor

$$\mathcal{E}_f(x) = \left(\varepsilon_{f,x_1}(x), \varepsilon_{f,x_2}(x), \dots, \varepsilon_{f,x_n}(x) \right) = \begin{pmatrix} \varepsilon_{f,x_1}(x) \\ \varepsilon_{f,x_2}(x) \\ \vdots \\ \varepsilon_{f,x_n}(x) \end{pmatrix} \in \mathbb{R}^n,$$

der **Elastizitätsgradient** (oder **Elastizitätsvektor**) der skalaren Funktion f im Punkt x .

(ii') In der Situation (ii) heißt die $m \times n$ -Matrix aller partiellen Elastizitäten der Komponentenfunktionen die **Elastizitätsmatrix** der Vektorfunktion F im Punkt x ,

$$\left(\varepsilon_{F,x}(x) \right) := \left(\varepsilon_{f_i,x_j}(x) \right)_{1 \leq i \leq m, 1 \leq j \leq n} = \begin{pmatrix} \varepsilon_{f_1,x_1}(x) & \varepsilon_{f_2,x_2}(x) & \dots & \varepsilon_{f_1,x_n}(x) \\ \varepsilon_{f_2,x_1}(x) & \varepsilon_{f_2,x_2}(x) & \dots & \varepsilon_{f_2,x_n}(x) \\ \vdots & \vdots & & \vdots \\ \varepsilon_{f_m,x_1}(x) & \varepsilon_{f_m,x_2}(x) & \dots & \varepsilon_{f_m,x_n}(x) \end{pmatrix}. \quad \blacksquare$$

Das Symbol “ ∇ ” in (i) ist ein künstlicher Buchstabe, der “Nabla” gesprochen wird. Man liest aber “ ∇f ” auch einfach als “Gradient von f ”. Wichtig ist folgende Konvention:

- In den Zeilen der Ableitungsmatrix [Elastizitätsmatrix] stehen alle partiellen Ableitungen [Elastizitäten] einer Komponentenfunktion, in den Spalten die partiellen Ableitungen [Elastizitäten] aller Komponentenfunktionen nach derselben Variablen.

Wie oben schon bemerkt, ist die Richtungsableitung $\partial_u f(x)$ gerade das Skalarprodukt von u mit dem Gradienten von f an der Stelle x und im vektoriellen Fall ist $\partial_u F(x)$ das Produkt der Ableitungsmatrix $\left(\frac{\partial F}{\partial x}(x)\right)$ mit dem Spaltenvektor u . Entsprechendes gilt für Elastizitäten. Wir halten das fest als

KOROLLAR: (i) Die Richtungsableitung der skalaren Funktion f an der Stelle $x \in \mathbb{R}^n$ nach dem Vektor $u \in \mathbb{R}^n$ ist das Skalarprodukt dieses Vektors mit dem Gradienten von f an der Stelle x ,

$$\partial_u f(x) = u \cdot \nabla f(x).$$

Die Richtungselastizität von f an der Stelle x bzgl. eines Vektors $v \in \mathbb{R}^n$ ist das Skalarprodukt dieses Vektors mit dem Elastizitätsgradienten an der Stelle x ,

$$\varepsilon_{f,v}(x) = v \cdot \mathcal{E}_f(x).$$

(ii) Die Richtungsableitung der Vektorfunktion F an der Stelle $x \in \mathbb{R}^n$ nach dem Vektor $u \in \mathbb{R}^n$ ist das Produkt der Ableitungsmatrix von F an der Stelle x mit dem Spaltenvektor u ,

$$\partial_u F(x) = \left(\frac{\partial F}{\partial x}(x)\right) u.$$

Die Richtungselastizität von F an der Stelle x bzgl. eines Vektors $v \in \mathbb{R}^n$ ist das Produkt der Elastizitätsmatrix von F an der Stelle x mit dem Spaltenvektor v ,

$$\varepsilon_{F,v}(x) = \left(\varepsilon_{F,x}(x)\right) v. \quad \blacksquare$$

Damit können wir jetzt auch das Problem der Bestimmung der Richtung mit größter Steigung von f angehen. Dazu setzen wir $v := \nabla f(x)$ und müssen dann unter allen Einheitsvektoren $u = (u_1, \dots, u_n)$ einen bestimmen, für den das Skalarprodukt $u \cdot v$, also die Steigung $\partial_u f(x)$ von f in Richtung u , am größten ist. Dazu nehmen wir zunächst an, dass auch $v = (v_1, \dots, v_n)$ ein Einheitsvektor ist, also $|v| = 1$. Aus der “binomischen Ungleichung” $ab \leq \frac{1}{2}(a^2 + b^2)$ (die wegen $a^2 + b^2 - 2ab = (a-b)^2 \geq 0$ gilt mit Gleichheit nur für $a = b$) folgt dann:

$$\begin{aligned} u \cdot v &= u_1 v_1 + u_2 v_2 + \dots + u_n v_n \\ &\leq \frac{1}{2}(u_1^2 + v_1^2) + \frac{1}{2}(u_2^2 + v_2^2) + \dots + \frac{1}{2}(u_n^2 + v_n^2) = \frac{1}{2}|u|^2 + \frac{1}{2}|v|^2 = 1, \end{aligned}$$

Letzteres weil $|u|^2$ bzw. $|v|^2$ die Summe der Quadrate aller Komponenten von u bzw. v ist und wir $|u| = 1 = |v|$ angenommen hatten. Diese Ungleichung $u \cdot v \leq 1$ für Einheitsvektoren heißt **Cauchy-Schwarz-Ungleichung**, und Gleichheit gilt dabei, wie die Herleitung zeigt, genau wenn $u = v$ ist. Wenn nun $\nabla f(x)$ kein Einheitsvektor ist, aber auch nicht der Nullvektor (d.h. mindestens eine der partiellen Ableitungen an der Stelle x ist nicht Null), so setzen wir $v := \frac{1}{|\nabla f(x)|} \nabla f(x)$ und haben damit den Einheitsvektor,

der dieselbe Richtung wie der Gradient hat. Die Cauchy-Schwarz-Ungleichung gibt dann für alle Einheitsvektoren u :

$$\partial_u f(x) = u \cdot \nabla f(x) = |\nabla f(x)|(u \cdot v) \leq |\nabla f(x)|$$

mit Gleichheit nur wenn $u = v$ ist, wenn also u die Richtung des Gradienten hat. Damit ist unser Problem gelöst: Die Steigung $\partial_u f(x)$ ist am größten, wenn u der Einheitsvektor in Richtung von $\nabla f(x)$ ist, in allen anderen Richtungen ist die Steigung kleiner. Und der Wert der größten Steigung ist gerade die Länge $|\nabla f(x)|$ des Gradienten. Wegen $\partial_{-u} f(x) = -\partial_u f(x)$ ist dann die Steigung in Richtung von $-u$ negativ und der kleinste Wert unter allen Steigungen von f an der Stelle x .

In Richtungen w orthogonal zur Richtung u maximaler Steigung (d.h. das Skalarprodukt $w \cdot u$ ist $= 0$, siehe 3.5) — und nur in diesen Richtungen — ist übrigens die Steigung Null; denn w ist dann ja auch orthogonal zu $\nabla f(x)$, also gilt $\partial_w f(x) = w \cdot \nabla f(x) = 0$.

Die Situation kann also an jeder fixierten Stelle x , an der die reelle Funktion f differenzierbar ist, folgendermaßen beschrieben werden: Entweder der Gradient von f an der Stelle x ist der Nullvektor, dann ist die Steigung von f an der Stelle x in allen Richtungen Null und die Funktion “in erster Näherung” konstant; oder der Gradient ist nicht der Nullvektor, und dann gibt es genau eine Richtung, nämlich die des Gradienten, in der die Steigung an der Stelle x positiv und am größten ist, und genau eine, in der die Steigung an dieser Stelle negativ und am kleinsten ist, nämlich die Gegenrichtung zum Gradienten, und genau in den dazu senkrechten Richtungen ist die Steigung der Funktion an dieser Stelle Null. Wir halten das fest in folgendem Satz, der dem Gradienten erst seine eigentliche Bedeutung verleiht:

SATZ (über den Gradienten und die maximale Steigung): Die reelle Funktion f von n Veränderlichen sei an der Stelle $x \in \mathbb{R}^n$ differenzierbar. Dann gilt:

(i) Die größte Steigung von f an der Stelle x ist die Euklidische Norm (Länge) des Gradienten an dieser Stelle, also gleich

$$|\nabla f(x)| = \sqrt{\left(\frac{\partial f}{\partial x_1}(x)\right)^2 + \left(\frac{\partial f}{\partial x_2}(x)\right)^2 + \dots + \left(\frac{\partial f}{\partial x_n}(x)\right)^2}.$$

Die kleinste Steigung an der Stelle x ist das Negative $-|\nabla f(x)|$ davon.

(ii) Ist $\nabla f(x)$ nicht der Nullvektor, so gibt es genau eine Richtung, in der f an der Stelle x die größte Steigung hat, und zwar ist dies die Richtung des Gradienten, d.h. der Einheitsvektor

$$\frac{1}{|\nabla f(x)|} \nabla f(x) \quad \text{ist der Richtungsvektor der größten Steigung von } f \text{ an der Stelle } x;$$

die eindeutige Richtung, in der f an der Stelle x die kleinste Steigung hat, ist dann die Gegenrichtung $-\frac{1}{|\nabla f(x)|} \nabla f(x)$ hierzu.

(iii) Genau in den zum Gradienten $\nabla f(x)$ orthogonalen Richtungen w hat f an der Stelle x die Steigung $\partial_w f(x) = 0$. ■

Für den Elastizitätsgradienten und die Richtungselastizitäten gelten ganz analoge Aussagen. Da man hierbei aber die Richtungselastizitäten $\varepsilon_{f,v}(x)$ bzgl. “Prozentsatzvektoren” v der Euklidischen Länge 1 betrachten muss, was ökonomisch nicht unbedingt sinnvoll ist, formulieren wir das nicht gesondert. Wir haben ja schon auf die Probleme hingewiesen, die es beim Vergleich von Elastizitäten in verschiedenen Richtungen gibt und bei der Definition, ob eine Funktion in einer bestimmten Richtung elastisch oder unelastisch ist.

DISKUSSION: 1) Manche Autoren verwenden das Wort “Gradient” auch für die (transponierte) Ableitungsmatrix einer Vektorfunktion. Dem schließen wir uns nicht an, weil dadurch die entscheidende Bedeutung des Gradienten als Richtung und Größe der größten Steigung vernebelt wird. Bei einer Vektorfunktion $F = (f_1, \dots, f_m)$ hat jede Komponentenfunktion f_i einen Gradienten ∇f_i . Aber diese Gradienten müssen an einer Stelle x des Definitionsbereichs keineswegs dieselbe Richtung haben, d.h. die Richtung des stärksten Anwachsens ist für verschiedene Komponentenfunktionen im Allgemeinen verschieden. Es macht bei einer \mathbb{R}^m -wertigen Funktion (mit $m \geq 2$) überhaupt keinen Sinn, von ihrer “Steigung” an irgendeiner Stelle in irgendeiner Richtung zu sprechen. Daher halten wir uns an das Prinzip:

- *Der Gradient ist nur für skalare Funktionen erklärt.*

2) *Ökonomische Interpretation des Gradienten:* Der Gradient $\nabla f(x)$ einer skalaren Funktion f , z.B. einer Gewinnfunktion, gibt an, wie man den Satz der unabhängigen Variablen $x = (x_1, \dots, x_n)$ ändern muss, um den größten Zuwachseffekt bei der abhängigen Variablen $f(x) = f(x_1, \dots, x_n)$ zu erzielen. Dabei bewirkt die Abänderung des Vektors x um einen Vektor $hu = \frac{h}{|\nabla f(x)|} \nabla f(x)$ der Länge $h > 0$ in Richtung von $\nabla f(x)$ näherungsweise einen Funktionszuwachs $f(x + hu) - f(x) \approx h \cdot |\nabla f(x)| = h \sqrt{(\frac{\partial f}{\partial x_1}(x))^2 + \dots + (\frac{\partial f}{\partial x_n}(x))^2}$, und diese Näherung ist um so besser je kleiner $h > 0$ ist. In entgegengesetzter Richtung wird der größte Verkleinerungseffekt für den Funktionswert erzielt, $f(x - hu) - f(x) \approx -h|\nabla f(x)|$. Ist $\nabla f(x) = 0$, so ist die Änderung des Funktionswertes bei kleinen Verschiebungen von x in beliebigen Richtungen marginal (“von erster Ordnung Null”).

3) Die Vektoren $w \in \mathbb{R}^n$, die zum Gradienten $\nabla f(x) \neq 0$ einer Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ an einer inneren Stelle x von D orthogonal sind, haben eine geometrische Bedeutung: Es sind genau die Tangentenvektoren an die Niveau-Menge von f im Punkt x . Unter der **Niveau-Menge** einer Funktion f auf D zu einem Wert y versteht man die Menge aller Punkte aus D , an denen f den Wert y annimmt. Ist $y = f(x)$ der Funktionswert an einer Stelle $x \in D$, so spricht man auch von der **Niveau-Menge durch den Punkt x** , das ist also die Menge $\{\tilde{x} \in D : f(\tilde{x}) = f(x)\}$ aller Punkte aus D , an denen die Funktion denselben Wert hat wie an der Stelle x . Die Namensgebung rührt daher, dass man sich f als Höhenfunktion eines Gebirges über der Basis D vorstellt; die Niveau-Mengen bestehen dann aus Punkten der Basis, über denen das Gebirge gleiche Höhe (“Niveau”) hat, sie entsprechen also den Höhenlinien auf einer Landkarte. In der Ökonomie nennt man die Niveau-Mengen die **Isoquanten** der Funktion f , weil sie aus Stellen mit demselben Funktionswert bestehen (gleiche “Quantität”).

Ist f eine skalare Funktion von n Veränderlichen, so werden die Isoquanten durch eine reelle Gleichung $f(x) = y$ beschrieben, es handelt sich also typischerweise um Kurven, wenn $n = 2$ ist, um Flächen im Fall $n = 3$ und allgemein um sog. Hyperflächen, d.h. Gebilde der Dimension $n-1$ im n -dimensionalen Raum. Dementsprechend spricht man von *Niveau-Kurven* oder *Isofunktionslinien* bzw. von *Niveau-Flächen* oder *Isofunktionsflächen*. (“Typischerweise” bedeutet, dass es nicht immer so sein muss wie erwartet. Die Niveau-Mengen können auch einmal eine “falsche” Dimension haben, z.B. ein Punkt, eine Kurve oder eine Vollkugel im dreidimensionalen Raum sein statt der erwarteten Fläche.) Die verschiedenen Niveau-Mengen schneiden sich nicht und der Definitionsbereich ist aus ihnen zusammengesetzt wie ein Material aus Fasern niedrigerer Dimension. In der Mathematik heißen die Niveau-Mengen daher auch die **Fasern** einer Abbildung.

Eine Kurve $c(t)$, die ganz in einer Niveau-Menge einer Funktion f verläuft, nennen wir eine (parametrisierte) **Kurve konstanten Niveaus** für diese Funktion. Äquivalent ist, dass $f(c(t))$ konstant ist als Funktion des Kurvenparameters t . Die Ableitung $\frac{d}{dt}f(c(t))$ der Funktion längs der Kurve ist also Null. Da diese Ableitung, wie weiter oben bemerkt, als Richtungsableitung $(\partial_{c'(t)}f)(c(t))$ in Richtung des Tangentenvektors $c'(t)$ der Kurve berechnet werden kann (wenn f stetige partielle Ableitungen hat) und da diese Richtungsableitung wiederum als Skalarprodukt $c'(t) \cdot \nabla f(c(t))$ mit dem Gradienten darstellbar ist (wenn f eine skalare Funktion ist), gilt

$$\frac{d}{dt}f(c(t)) = 0 \text{ für alle } t \iff c'(t) \cdot \nabla f(c(t)) = 0 \text{ für alle } t.$$

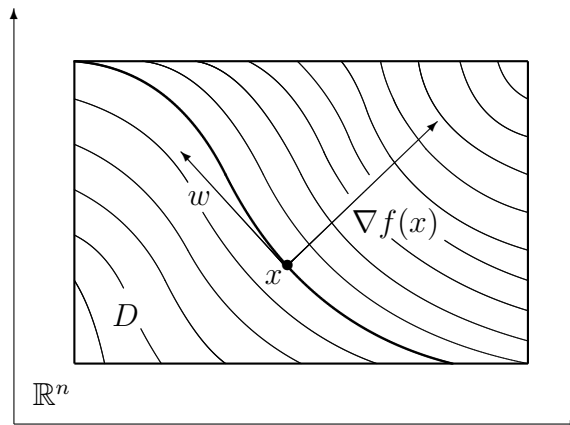
Da das Verschwinden des Skalarprodukts zwischen zwei Vektoren bedeutet, dass sie aufeinander senkrecht stehen (siehe 3.5), kann man dieses Ergebnis so formulieren:

- *Das Gradientenfeld steht senkrecht auf den Niveau-Mengen von f .*

Genauer gesagt:

- *Eine Kurve $c(t)$ ist genau dann eine Kurve konstanten Niveaus für die reelle Funktion f , wenn sie orthogonal zum Gradientenfeld von f verläuft.*

Die Kurven konstanten Niveaus verlaufen in den Niveau-Mengen, fallen aber nur im 2-dimensionalen Fall $n = 2$, der in der Abbildung dargestellt ist, mit den Niveau-Mengen der Funktion f zusammen. Es lässt sich zeigen (mit dem späteren Satz über implizite Funktionen), dass jeder zum Gradienten $\nabla f(x)$ an einer Stelle $x \in D$ senkrechte Vektor w der Tangentenvektor $w = c'(t_0)$ einer Kurve konstanten Niveaus $c(t)$ mit $c(t_0) = x$ ist, vorausgesetzt $\nabla f(x) \neq 0$. In diesem Sinne gilt für $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$:



- *Die zum Gradienten $\nabla f(x) \neq 0$ orthogonalen Vektoren $w \in \mathbb{R}^n$ sind genau die Tangentenvektoren im Punkte x an die Niveau-Menge von f durch den Punkt x .*

Man kann diese Ergebnisse noch weiter veranschaulichen, indem man f als Höhenfunktion eines Gebirges über der Grundfläche D interpretiert. Ein Wanderer auf diesem Gebirge, der sich senkrecht über dem Basispunkt x befindet, also an der Stelle $(x, f(x))$, und in alle horizontalen Richtungen blickt, stellt unterschiedlich steile An- und Abstiege in verschiedenen Richtungen fest. Die Richtung des steilsten Anstiegs ist die des Gradienten $\nabla f(x)$ und die Steigung in dieser Richtung ist die Länge des Gradienten $|\nabla f(x)|$. (Ausnahme: $\nabla f(x) = 0$; siehe die folgende Bemerkung.) Die Niveau-Mengen der Höhenfunktion f durch den Punkt x sind hier die Höhenlinien des Gebirges. In Richtung der Höhenlinien ist die Steigung der Gebirgsoberfläche Null, und wenn die Steigung in einem Punkt nicht in allen Richtungen Null ist, so verläuft die Höhenlinie durch diesen Punkt orthogonal zur Richtung des steilsten Anstiegs.

4) Die Nullstellen des Gradientenfeldes einer differenzierbaren reellen Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ nennt man ihre **kritischen Punkte** (oder *stationäre Stellen*). Das sind also die Stellen $x \in D$ in denen die partiellen Ableitungen $\frac{\partial f}{\partial x_j}$ simultan für $j = 1 \dots n$ verschwinden. Dann gilt auch $\partial_u f(x) = u \cdot \nabla f(x) = 0$ für alle $u \in \mathbb{R}^n$, d.h. die Steigung an der Stelle x ist in allen Richtungen Null. Bei der obigen Interpretation von f als Höhenfunktion eines Gebirges ist das z.B. der Fall im höchsten Punkt einer Kuppe, im tiefsten Punkt einer Senke oder in einem Sattelpunkt des Gebirges. Die kritischen Punkte von f sind Kandidaten für (lokale) Maximum- oder Minimumstellen der Funktion; darauf basieren die Methoden der Extremstellenbestimmung bei Funktionen von mehreren Veränderlichen, die wir später besprechen. Aber nicht jeder kritische Punkt muss eine (lokale) Extremstelle sein — es gibt schließlich auch Sattelpunkte. Bei Vektorfunktionen $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ definiert man als kritische Punkte die Stellen $x \in D$, in denen die Ableitungsmatrix $\left(\frac{\partial f_i}{\partial x_j}(x)\right)$ nicht den maximal möglichen Rang hat, also einen Rang $< \min\{m, n\}$. (Nur im skalaren Fall $m = 1$ oder im Fall einer Kurve $n = 1$ ist das äquivalent dazu, dass alle partiellen Ableitungen an der Stelle x simultan verschwinden.)

5) *Rechenregeln für die Gradientenbildung* lassen sich aus den Rechenregeln für partielle Differentiation herleiten. Zum Beispiel hat man *Linearität des Gradientenoperators*, d.h.

$$\nabla(rf + sg) = r\nabla f + s\nabla g$$

für differenzierbare $f, g: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ und $r, s \in \mathbb{R}$. Außerdem gilt eine *Produktregel für Gradienten*

$$\nabla(fg) = g\nabla f + f\nabla g,$$

die sich aus $\frac{\partial}{\partial x_j}(fg) = \frac{\partial f}{\partial x_j}g + f\frac{\partial g}{\partial x_j}$ ergibt. Auch die Bildung der Ableitungsmatrix von Vektorfunktionen ist linear, d.h. es gilt dafür eine Summen- und eine Faktorregel, und bei Bedarf kann man sich auch Produktregeln überlegen, die für die verschiedenen Matrix \times Vektor- bzw. Matrix \times Matrix-Produkte gelten. ■

Mit den obigen Sätzen ist Berechnung von Richtungsableitungen und maximaler Steigung auf die Berechnung des Gradienten zurückgeführt und damit auf die Berechnung der partiellen Ableitungen einer skalaren Funktion (die ja die Komponenten des Gradientenfeldes sind). Das geht also mit Differentialrechnung für Funktionen von *einer* Veränderlichen (bei c.p.-Bedingung) und vereinfacht die Berechnungen enorm. Dazu nun einige

BEISPIELE (*Berechnung von maximaler Steigung und Richtungsableitungen bzw. Richtungselastizitäten mit dem Gradienten bzw. Elastizitätsgradienten*):

(1) Wir greifen die Funktion $f(x, y) = x^2(1+y)$ von zwei Veränderlichen wieder auf, für die wir früher schon die Richtungsableitungen direkt mit Anwendung der Definition der Richtungsableitung ausgerechnet haben. Die partiellen Ableitungen sind leicht zu bestimmen (das wurde auch in 5.1 schon gemacht), und wir erhalten damit das Gradientenfeld

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)\right) = (2x(1+y), x^2).$$

Für die Ableitung in Richtung eines beliebigen Vektors $u = (u_1, u_2)$ ergibt sich damit

$$\partial_u f(x, y) = u \cdot \nabla f(x, y) = u_1 2x(1+y) + u_2 x^2,$$

ein Ergebnis, das wir früher schon durch Differentiation von $f(x+tu_1, y+tu_2)$ nach t an der Stelle $t = 0$ viel umständlicher berechnet haben, wobei die lineare Abhängigkeit des Ergebnisses vom Richtungsvektor gar nicht bemerkt oder eher als Zufall eingestuft wurde, während wir nun wissen, dass es sich um eine allgemeine Gesetzmäßigkeit handelt.

Betrachten wir konkret die Stelle $(x, y) = (3, 1)$, so haben wir $\nabla f(3, 1) = (12, 9) = 3(4, 3)$, die Richtung größter Steigung an dieser Stelle ist also $\frac{1}{5}(4, 3)$ und die maximale Steigung ist $|\nabla f(3, 1)| = 15$. Die Funktionswertänderung bei Erhöhung von $x = 3$ und $y = 1$ um dieselbe betragskleine Größe h ist dann näherungsweise $f(3+h, 1+h) - f(3, 1) \approx h \partial_{(1,1)} f(3, 1) = h(1, 1) \cdot (12, 9) = 21h$. In radialer Richtung $\frac{1}{\sqrt{10}}(3, 1)$ ist die Steigung an dieser Stelle $\partial_{\text{rad}} f(x, y) = \frac{1}{\sqrt{10}}(3, 1) \cdot (12, 9) = \frac{45}{\sqrt{10}}$ und die Ableitung nach dem Ortsvektor (Skalenableitung) ist $\partial_{(3,1)} f(3, 1) = 45$. An allgemeinen Stellen (x, y) sind Skalenableitung und radiale Ableitung (wenn $(x, y) \neq (0, 0)$) gegeben durch

$$\partial_{(x,y)} f(x, y) = (x, y) \cdot \nabla f(x, y) = x^2(3y+2), \quad \partial_{\text{rad}} f(x, y) = \frac{x^2(3y+2)}{\sqrt{x^2 + y^2}}.$$

In Richtung des Vektors $w = (-3, 4)$ hat die Funktion an der Stelle $(3, 1)$ Steigung Null, weil dieser Vektor orthogonal zum Gradienten ist, $(-3, 4) \cdot (12, 9) = 0$. Für die Ableitung der Funktion längs der Kurve $c(t) = (3/t, 2t^2 - 1)$ ($t > 0$) mit Tangentenvektor $c'(t) = (-3/t^2, 4t)$ ergibt sich

$$\frac{d}{dt} f(c(t)) = c'(t) \cdot \nabla f(c(t)) = \left(\frac{-3}{t^2}, 4t \right) \cdot \left(12t, \frac{9}{t^2} \right) = 0,$$

also ist $c(t)$ eine Kurve konstanten Niveaus für die Funktion. Tatsächlich ist $c(1) = (3, 1)$, die Kurve parametrisiert die Niveaulinie der Funktion durch den Punkt $(3, 1)$ und der Vektor $w = (-3, 4) = c'(1)$ ist gerade der Tangentenvektor der Kurve in diesem Punkt.

(2) Die Diskussion der Funktion $f(x, y)$ aus 1) lässt sich ganz analog für Richtungselastizitäten durchführen (auf dem Definitionsbereich $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$). Der Elastizitätsgradient ist

$$\mathcal{E}_f(x, y) = (\varepsilon_{f,x}(x, y), \varepsilon_{f,y}(x, y)) = \left(\frac{x \frac{\partial f}{\partial x}(x, y)}{f(x, y)}, \frac{y \frac{\partial f}{\partial y}(x, y)}{f(x, y)} \right) = \left(2, \frac{y}{1+y} \right),$$

die Richtungselastizität bzgl. eines Vektors $v = (v_1, v_2)$ also

$$\varepsilon_{f,v}(x, y) = v \cdot \mathcal{E}_f(x, y) = 2v_1 + v_2 \frac{y}{1+y}.$$

Insbesondere ist die Skalenelastizität gegeben durch

$$\sigma_f(x, y) = \varepsilon_{f,(1,1)}(x, y) = 2 + \frac{y}{1+y}$$

und dies ist auch gleich $\frac{1}{f(x,y)} \frac{d}{ds} \Big|_{s=1} f(sx, sy) = \frac{1}{f(x,y)} \partial_{(x,y)} f(x, y)$, was ja allgemein gilt und hier zur Bestätigung mit (1) verifiziert werden kann.

Man sieht, dass der Elastizitätsgradient im Allgemeinen eine andere Richtung hat als der Gradient. An der Stelle $(x, y) = (3, 1)$ haben wir z.B. $\mathcal{E}_f(3, 1) = (2, \frac{1}{2})$ mit Richtung $\frac{1}{\sqrt{17}}(4, 1)$, aber $\nabla f(3, 1) = (12, 9)$ mit Richtung $\frac{1}{5}(4, 3)$. Erhöht man $x = 3$ um 2% und vermindert $y = 1$ um 1%, so erhöht sich der Funktionswert $f(3, 1) = 18$ ungefähr um $(2, -1) \cdot (2, \frac{1}{2}) = 3.5$ Prozent. Bei Verminderung von $x = 3$ um 1% und Erhöhung von $y = 1$ um 2% ist die relative Änderung des Funktionswertes dagegen etwa $(-1, 2) \cdot (2, \frac{1}{2}) = -1$ Prozent; insbesondere verkleinert sich der Funktionswert. Verändert man $x = 3$ und $y = 1$ um denselben (kleinen) Prozentsatz $p\%$, so ist die relative Zunahme des Funktionswertes näherungsweise gleich $p \cdot \sigma_f(3, 1) \% = p \cdot 2.5\%$. Unter allen Einheitsvektoren $v \in \mathbb{R}^2$ ist der Einheitsvektor in Richtung des Elastizitätsgradienten $\frac{1}{\sqrt{17}}(4, 1)$ derjenige mit der größten Richtungselastizität $\varepsilon_{f,v}(3, 1)$. Ökonomisch sinnvoller ist unter Umständen die Frage, für welchen Prozentsatzvektor $v = (v_1, v_2)$ mit $\max\{|v_1|, |v_2|\} \leq 1$ bzw. mit $|v_1| + |v_2| \leq 1$ die relative Änderung des Funktionswertes an der Stelle $(3, 1)$ am größten ist: Die Antwort, die man aus $\varepsilon_{f,v}(3, 1) = 2v_1 + \frac{1}{2}v_2$ abliest, lautet $v = (1, 1)$ bei der ersten Bedingung an v bzw. $v = (1, 0)$ bei der zweiten.

(3) Für die Funktion $f(x, y, z) = x^3 e^{yz}$ haben wir

$$\nabla f(x, y, z) = x^2 e^{yz} (3, xz, xy), \quad \mathcal{E}_f(x, y, z) = (3, yz, yz)$$

und können damit sofort — und einfacher als früher — alle Richtungsableitungen und Richtungselastizitäten durch Skalarproduktbildung berechnen:

$$\partial_u f(x, y, z) = x^2 e^{yz} (3u_1 + xzu_2 + xyu_3), \quad \varepsilon_{f,v}(x, y, z) = 3v_1 + yz(v_2 + v_3).$$

An der Stelle $(x, y, z) = (2, 1, \ln 2)$ mit $f(2, 1, \ln 2) = 16$ ist die Skalenableitung bzw. Slanelastizität z.B.

$$\partial_{(2,1,\ln 2)} f(2, 1, \ln 2) = 16(3 + 2 \ln 2), \quad \sigma_f(2, 1, \ln 2) = 3 + 2(\ln 2).$$

Erhöht man alle drei Komponenten von $(2, 1, \ln 2)$ um dieselbe kleine Größe $h > 0$, so erhöht sich der Funktionswert um ca. $h \cdot \partial_{(1,1,1)}(2, 1, \ln 2) = 8(5 + 2 \ln 2)h$; erhöht man dagegen alle drei Komponenten um denselben Prozentsatz $p\%$, so ändert sich der Funktionswert um ca. $(3 + 2 \ln 2)p\%$.

Für die Kurve $c(t) = (e^{-t^2/3}, t, t)$ ist der Tangentenvektor $c'(t) = (-\frac{2}{3}te^{-t^2/3}, 1, 1)$ orthogonal zu $\nabla f(c(1)) = e^{t^2/3}(3, te^{-t^2/3}, te^{-t^2/3})$, es handelt sich also um eine Kurve konstanten Niveaus bzgl. der Funktion f (und zwar ist das Niveau konstant gleich $f(c(0)) = 1$). Bei der Kurve $b(t) = (3t, e^{3t^2/2}, e^{3t^2/2})$ hat dagegen der Tangentenvektor $b'(t) = (3, 3te^{3t^2/2}, 3te^{3t^2/2})$ für alle t die Richtung des Gradienten $\nabla f(b(t)) = (3t)^2 \exp(e^{3t^2})b'(t)$, die Kurve verläuft an jeder Stelle also in Richtung des stärksten Anstiegs der Funktion, es handelt sich um eine sog. **Gradientenlinie**.

(4) Für eine *Cobb–Douglas–Funktion* $f(x) = cx_1^{r_1} x_2^{r_2} \dots x_n^{r_n}$ gilt für alle $u, v \in \mathbb{R}^n$ an Stellen $x \in \mathbb{R}_{>0}^n$ (also $x_j > 0$ für $j = 1 \dots n$):

$$\begin{aligned} \nabla f(x) &= f(x) \left(\frac{r_1}{x_1}, \frac{r_2}{x_2}, \dots, \frac{r_n}{x_n} \right), \\ \partial_u f(x) &= u \cdot \nabla f(x) = \left(\frac{u_1 r_1}{x_1} + \frac{u_2 r_2}{x_2} + \dots + \frac{u_n r_n}{x_n} \right) f(x), \\ \partial_x f(x) &= x \cdot \nabla f(x) = (r_1 + r_2 + \dots + r_n) f(x), \\ \partial_{\text{rad}} f(x) &= \frac{1}{|x|} x \cdot \nabla f(x) = \frac{r_1 + r_2 + \dots + r_n}{|x|} f(x), \\ \mathcal{E}_f(x) &= (r_1, r_2, \dots, r_n) \in \mathbb{R}^n \quad \text{konstant}, \\ \varepsilon_{f,v}(x) &= r_1 v_1 + r_2 v_2 + \dots + r_n v_n \quad \text{konstant}, \\ \sigma_f(x) &= r_1 + r_2 + \dots + r_n \quad \text{konstant}. \end{aligned}$$

Für die früher betrachtet konkrete *Produktionsfunktion* $x(A, K) = 2.5 \cdot A^{0.6} \cdot K^{0.4}$ in Abhängigkeit von den Produktionsfaktoren Arbeit und Kapital haben wir z.B. $\mathcal{E}_x(A, K) \equiv (0.6, 0.4)$. In 5.1 konnte wir mit Berechnung der partiellen Elastizitäten nur sagen, wie sich $x(A, K)$ prozentual ändert, wenn wir nur einen der Produktionsfaktoren verändern und den anderen Wert beibehalten. Nun aber sehen wir aus der linearen Abhängigkeit der Richtungselastizitäten vom Prozentsatzvektor, dass diese Änderung $(0.6p + 0.4q)\%$ beträgt, wenn man A um $p\%$ und simultan K um $q\%$ ändert (kleine Prozentsätze unterstellt; diese dürfen auch negativ sein, d.h. der Faktoreinsatz wird reduziert).

Die Konstanz des Elastizitätsgradienten, also aller partiellen Elastizitäten (und Richtungselastizitäten), charakterisiert übrigens die Funktionen vom Cobb–Douglas–Typ, wie wir in 5.1 gezeigt haben. Es gibt aber viele weitere Funktionen mit konstanter Skalen–Elastizität (die sog. homogenen Funktionen, s.u.).

(5) Für skalare *affin lineare Funktionen* $\ell(x) = \mathbf{a} \cdot x + b$ auf \mathbb{R}^n ist der Gradient

$$\nabla \ell(x) = \mathbf{a} \in \mathbb{R}^n \quad \text{konstant gleich dem Koeffizientenvektor.}$$

Dies ergibt sich aus der Berechnung der partiellen Ableitungen in 5.1. Für vektorielle lineare Funktionen $L(x) = Ax + \mathbf{b}$ ist entsprechend die Ableitungsmatrix $\left(\frac{\partial L}{\partial x}(x)\right)$ an allen Stellen $x \in \mathbb{R}^n$ gleich der Koeffizientenmatrix $A \in \mathbb{R}^{m \times n}$. Die Konstanz des Gradienten bzw. der Ableitungsmatrix charakterisiert übrigens die linearen Funktionen. Ist nämlich z.B. $\nabla f \equiv \mathbf{a} \in \mathbb{R}^n$ konstant, so folgt $\frac{d}{dt}[f(tx) - t\mathbf{a} \cdot x] = x \cdot \nabla f(tx) - \mathbf{a} \cdot x \equiv 0$, also ist $\mathbb{R} \ni t \mapsto f(tx) - t\mathbf{a} \cdot x$ konstant gleich $f(0)$ und damit $f(x) = \mathbf{a} \cdot x + f(0)$.

(6) Für homogene *quadratische Funktionen*

$$q(x) = x \cdot Ax = \frac{1}{2} \sum_{i,j=1}^n a_{ij} x_i x_j$$

auf \mathbb{R}^n mit symmetrischer Koeffizientenmatrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ sind die partiellen Ableitungen

$$\frac{\partial q}{\partial x_k} = \frac{1}{2} \sum_{i,j=1}^n a_{ij} \left(\frac{\partial x_i}{\partial x_k} x_j + x_i \frac{\partial x_j}{\partial x_k} \right) = \frac{1}{2} \sum_{j=1}^n a_{kj} x_j + \frac{1}{2} \sum_{i=1}^n a_{ik} x_i = \sum_{j=1}^n a_{kj} x_j,$$

wobei $\frac{\partial x_i}{\partial x_k} \equiv 0$ für $i \neq k$ (und $= 1$ für $i = k$) und die Symmetrie $a_{ik} = a_{ki}$ benutzt wurde. Somit ist der Gradient an der Stelle x

$$\nabla q(x) = Ax \quad \text{Produkt der Koeffizientenmatrix mit } x,$$

und das Gradientenfeld ∇q ist die homogene lineare Vektorfunktion auf \mathbb{R}^n mit Koeffizientenmatrix A . Für allgemeine *quadratische Polynomfunktionen*

$$p(x) = \frac{1}{2} \sum_{i,j=1}^n a_{ij} x_i x_j + \sum_{j=1}^n b_j x_j + c$$

gilt dann, wenn $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^n$ und $c \in \mathbb{R}$:

$$\nabla p(x) = Ax + \mathbf{b}.$$

(7) Wir greifen das frühere Beispiel der **Nachfrage-Funktionen**

$$x_1(p_1, p_2) = 0.5 \cdot \frac{p_2^{0.7}}{p_1}, \quad x_2(p_1, p_2) = p_1^{1/2} - 1.5p_2 + 1.65.$$

auf, die wir zu einer Vektorfunktion $x(p) = (x_1(p_1, p_2), x_2(p_1, p_2))$ zusammenfassen. Die Ableitungsmatrix und Elastizitäts-Matrix haben wir in 5.1 schon angegeben:

$$\left(\frac{\partial x_i}{\partial p_j} \right) = \begin{pmatrix} -0.5 p_1^{-2} p_2^{0.7} & 0.35 p_1^{-1} p_2^{-0.3} \\ 0.5 p_1^{-1/2} & -1.5 \end{pmatrix}, \quad \left(\varepsilon_{x_i, p_j} \right) = \begin{pmatrix} -1 & 0.7 \\ \frac{p_1^{1/2}}{2 x_2(p_1, p_2)} & \frac{-1.5 p_2}{x_2(p_1, p_2)} \end{pmatrix}.$$

Für die zugehörige Erlösfunktion $E(p_1, p_2) = 0.5p_2^{0.7} + p_2(p_1^{1/2} - 1.5p_2 + 1.65)$ ist

$$\begin{aligned}\nabla E(p) &= \left(\frac{1}{2}p_1^{-1/2}p_2, p_1^{1/2} + 0.35p_2^{-0.3} - 3p_2 + 1.65 \right), \\ \mathcal{E}_x(p) &= \left(\frac{p_1^{1/2}p_2}{2E(p_1, p_2)}, \frac{p_2(0.35p_2^{-0.3} + p_1^{1/2} - 3p_2 + 1.65)}{E(p_1, p_2)} \right).\end{aligned}$$

Betrachten wir wie früher konkret die Preise $p_1 = 4$ und $p_2 = 1$, so haben wir:

$$\left(\frac{\partial x_i}{\partial p_j}(4, 1) \right) = \begin{pmatrix} 0.03125 & 0.0875 \\ 0.25 & -1.5 \end{pmatrix}, \quad \left(\varepsilon_{x_i, p_j}(4, 1) \right) = \begin{pmatrix} -1 & 0.7 \\ \frac{1}{2.15} & \frac{6}{2.15} \end{pmatrix},$$

$$\nabla E(4, 1) = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \quad \mathcal{E}_E(4, 1) = \frac{1}{2.65} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Mit diesen Informationen können wir nun sofort durch einfache Matrix \times Vektor-Multiplikation bzw. Skalarproduktbildung berechnen, wie sich der Nachfragevektor und der Erlös ändern, wenn die Preise p_1, p_2 simultan (wenig) geändert werden bzw. wie sich die Nachfragen x_1, x_2 und der Erlös dabei prozentual ändern. Bei einer Preisänderung von $p_1 = 4$ um h Geldeinheiten und von $p_2 = 1$ um k Geldeinheiten ergibt sich die ungefähre Änderung des Nachfragevektors $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ durch Multiplikation der Ableitungsmatrix von links an $\begin{pmatrix} h \\ k \end{pmatrix}$ und die Erlösänderung ist ca. $4h + k$ (wenn $|h|$ und $|k|$ klein sind). Bei einer Änderung von $p_1 = 4$ um $p\%$ und von $p_2 = 1$ um $q\%$ ergeben sich die prozentualen Änderungen der Komponenten des Nachfragevektors durch Multiplikation der Elastizitäts-Matrix von links an $\begin{pmatrix} p \\ q \end{pmatrix}$ und die relative Erlösänderung ist $\frac{1}{2.65}(p + q)\%$.

Entsprechend kann man bei dem früheren Beispiel von drei Nachfragefunktionen in Abhängigkeit von drei Preisen vorgehen, das am Ende von 5.1 vorgestellt wurde, und allgemein auch bei n Nachfragefunktionen in Abhängigkeit von den Preisen für n betrachtete Güter. Wir führen das nicht weiter aus; die ökonomische Bedeutung von Gradient und Ableitungsmatrix bzw. Elastizitäts-Gradient und Elastizitäts-Matrix sollte an diesem Beispiel hinreichen klar geworden sein.

(8) Da die Skalanelastizität $\sigma_f(x)$ die Richtungselastizität bzgl. des Vektors v mit lauter Einträgen 1 ist, $v = (1, 1, \dots, 1)$, gilt:

$$\sigma_f(x) = \varepsilon_{f, x_1}(x) + \varepsilon_{f, x_2}(x) + \dots + \varepsilon_{f, x_n}(x).$$

- Die Skalen-Elastizität ist die Summe der partiellen Elastizitäten.

Dieser einfache Sachverhalt, der sich aus der Darstellung $v = \sum_{j=1}^n e_j$ von v als Summe der kanonischen Basisvektoren und der linearen Abhängigkeit der Elastizitäten vom Richtungsvektor ergibt, heißt in der Wirtschaftswissenschaft *Wicksell-Johnson-Theorem* (s. Tietze, S. 321). ■

Funktionen, deren Skalen-Elastizität konstant ist, wie bei den Cobb-Douglas-Funktionen im vorangehenden Beispiel (4), treten bei der mathematischen Modellierung ökonomischer Situationen häufig auf. Der mathematische Fachbegriff für diese Eigenschaft ist "Homogenität", die wir nun am Ende dieses Abschnitts noch einführen und erläutern:

DEFINITION: Eine Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ (oder auch $\rightarrow \mathbb{R}^m$) heißt **positiv homogene Funktion**, wenn es einen Exponenten $r \in \mathbb{R}$ gibt mit

$$sx \in D \quad \text{und} \quad f(sx) = s^r f(x) \quad \text{für alle } x \in D \text{ und } s \in \mathbb{R}_{>0},$$

d.h. man kann positive Skalare $s > 0$ aus der Funktion mit der r -ten Potenz “herausziehen”. Der Exponent r heißt dann der **Homogenitätsgrad**. Die Funktion heißt **homogene Funktion** vom Homogenitätsgrad $k \in \mathbb{Z}$, wenn sogar für alle $t \in \mathbb{R}_{\neq 0}$ und $x \in D$ gilt $tx \in D$ und $f(tx) = t^k f(x)$. (Da hierbei auch Potenzen von negativen Zahlen t zu bilden sind, muss man sich auf ganze Exponenten beschränken.) ■

DISKUSSION 1) Die Bedingung, dass mit x auch alle Vielfachen sx mit positivem Faktor s in D liegen, bedeutet geometrisch, dass der Definitionsbereich ein **Kegel** in \mathbb{R}^n ist, d.h. eine Vereinigung von Ursprungsstrahlen. (Falls sogar $tx \in D$ ist für alle $t \neq 0$ und $x \in D$, so nennt man D einen *Doppelkegel*; das ist ein Kegel, der mit jedem Ursprungsstrahl auch den Strahl in Gegenrichtung enthält.) Typische Kegel sind \mathbb{R}^n selbst, $\mathbb{R}^n \setminus \{0\}$, eine Halbraum $\mathbb{R}_{>0} \times \mathbb{R}^{n-1}$ oder der Raum $\mathbb{R}_{>0}^n$ der Vektoren mit lauter positiven Komponenten (der *positive Quadrant* in \mathbb{R}^2 , der *positive Oktant* in \mathbb{R}^3, \dots).

2) Da für eine vom Grad r positiv homogene Funktion $f(x) = |x|^r f(\frac{1}{|x|}x)$ ist, wenn $x \neq 0$, gilt:

- Eine positiv homogene Funktion ist an allen Stellen $\neq 0$ schon bestimmt durch ihre Werte in den Einheitsvektoren ihres Definitionsbereichs.

Umgekehrt gibt es zu jeder gegebenen Funktion φ auf einer Teilmenge $\Delta \subset \mathbb{S}^{n-1}$ der Menge aller Einheitsvektoren im \mathbb{R}^n (das ist die sog. *Einheitssphäre* \mathbb{S}^{n-1} in \mathbb{R}^n) und zu jedem gegebenen Exponenten $r \in \mathbb{R}$ genau eine vom Grad r positiv homogene Funktion f auf der Vereinigung D aller Ursprungsstrahlen $\mathbb{R}_{>0}u$ durch Einheitsvektoren $u \in \Delta$, derart dass f auf Δ mit φ übereinstimmt, nämlich $f(x) = |x|\varphi(\frac{1}{|x|}x)$. Man kann also die Werte einer positiv homogenen Funktion auf den Einheitsvektoren und ihren Homogenitätsgrad beliebig vorgeben.

(3) Positive Homogenität vom *Homogenitätsgrad Null* bedeutet natürlich einfach, dass die Funktion auf jedem Strahl $\mathbb{R}_{>0}u = \{su : s > 0\}$ im Definitionsbereich konstant ist. Solche Funktionen nennt man auch **radial konstant**.

4) *Homogenität in der Ökonomie* tritt oft aufgrund des Skalierungsverhaltens ökonomischer Funktionen auf. Das Skalierungsverhalten bezieht sich auf die Änderung des Funktionswertes, wenn man alle unabhängigen Variablen um denselben Faktor $s > 0$ verändert (z.B. durch Übergang zu einer anderen Maßeinheit “reskaliert”). Oft ist einfach $f(sx_1, \dots, sx_n) = s f(x_1, \dots, x_n)$, also f positiv homogen vom Homogenitätsgrad 1. Die simultane und gleiche Vervielfachung des Einsatzes aller “Faktoren” x_j führt dann zu derselben Vervielfachung des “Outputs” $f(x)$. Funktionen vom Homogenitätsgrad 1 nennt man auch *linear homogen*. (Das müssen aber keine linearen Funktionen sein; das Wort “linear” bedeutet hier nur, dass $f(sx)$ eine lineare Funktion des Parameters $s \in \mathbb{R}_{>0}$ ist.) In komplizierteren Fällen reagiert die abhängige Größe $f(x_1, \dots, x_n)$ auf eine Vervielfachung aller Argumente x_j mit demselben beliebigen Faktor s , indem sie sich nicht mit diesem Faktor vervielfacht, sondern eben mit einer Potenz s^r ; dann liegt positive Homogenität vom Grad r vor. ■

BEISPIELE: (1) Homogen lineare Funktionen $\ell(x) = \mathbf{a} \cdot x$ bzw. $L(x) = Ax$ mit $\mathbf{a} \in \mathbb{R}^n$ bzw. $A \in \mathbb{R}^{m \times n}$ sind homogen auf \mathbb{R}^n vom Homogenitätsgrad 1.

(2) Ein *Monom* $x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}$ ($k_j \in \mathbb{N}_0$) ist homogen auf \mathbb{R}^n vom Homogenitätsgrad $k = k_1 + k_2 + \cdots + k_n$. Dasselbe gilt für jede Linearkombination solcher Monome mit derselben Exponentensumme, und solche Linearkombinationen heißen vom Grad k **homogene Polynomfunktionen**. Zum Beispiel ist $x^2y + 2y^3 - 3xy^2$ eine vom Grad 3 homogene Polynomfunktion von zwei Veränderlichen und $x^4 + 2x^2y^2 - 3xy^2z + yz^3$ eine vom Grad 4 homogene Polynomfunktion von drei Variablen.

(4) Der Quotient $\frac{p(x)}{q(x)}$ zweier homogenen Polynomfunktionen p, q der Grade k, l ist eine vom Grad $k - l$ *homogene rationale Funktion* auf dem Kegel $\{x \in \mathbb{R}^n : q(x) \neq 0\}$. Zum Beispiel ist $\frac{x^2 - y^2}{x^2 + x^2}$ vom Grad 0 homogen auf $\mathbb{R}_{\neq(0,0)}^2$ und $\frac{x^3 + 2xyz}{x^2 + z^2}$ vom Grad 1 homogen auf $\{(x, y, z) \in \mathbb{R}^3 : x \neq 0 \text{ oder } z \neq 0\}$. Allgemeiner sind natürlich auch Produkte bzw. Quotienten von (positiv) homogenen Funktionen beliebiger Homogenitätsgrade wieder homogen, wobei die Summe bzw. Differenz der Homogenitätsgrade gleich dem Homogenitätsgrad des Produkts bzw. Quotienten ist. Die Summe von zwei homogenen Funktionen ist aber im Allgemeinen nur dann wieder homogen, wenn beide denselben Homogenitätsgrad haben.

(4) Eine *Cobb–Douglas–Funktion* $f(x) = c x_1^{r_1} x_2^{r_2} \cdots x_n^{r_n}$ mit Exponenten $r_j \in \mathbb{R}$ ist auf $\mathbb{R}_{>0}^n$ positiv homogen vom Homogenitätsgrad $r = r_1 + r_2 + \cdots + r_n$. Funktionen dieses Typs sind sogar *partiell (positiv) homogen* in jeder einzelnen Variablen x_j vom Grad r_j , d.h.

$$f(x_1, \dots, x_{j-1}, s x_j, x_{j+1}, \dots, x_n) = s^{r_j} f(x_1, \dots, x_n)$$

gilt für alle (x_1, \dots, x_n) im Definitionsbereich und für alle $s > 0$. Umgekehrt ist jede partiell homogene Funktion f auf $\mathbb{R}_{>0}^n$ eine Cobb–Douglas–Funktion, weil dann $f(x_1, x_2, \dots, x_n) = x_1^{r_1} f(1, x_2, \dots, x_n) = x_1^{r_1} x_2^{r_2} f(1, 1, x_3, \dots, x_n) = \cdots = x_1^{r_1} x_2^{r_2} \cdots x_n^{r_n} c$ gilt mit $c = f(1, 1, \dots, 1)$. ■

SATZ (Charakterisierung homogener Funktionen): Sei $D \subset \mathbb{R}_{\neq 0}^n$ ein offener Kegel, und $f: D \rightarrow \mathbb{R}$ (oder $\rightarrow \mathbb{R}^m$) habe stetige partielle Ableitungen auf D . Dann sind für jeden Exponenten $r \in \mathbb{R}$ folgende Aussagen äquivalent:

(i) f ist positiv homogen vom Grad r , also $f(sx) = r^s f(x)$ für $s > 0$;

(ii) Die Skalen–Ableitung $\partial_x f(x) = \frac{d}{ds} \Big|_{s=1} f(sx)$ reproduziert f bis auf den Faktor r :

$$\partial_x f(x) = x_1 \frac{\partial f}{\partial x_1}(x) + \cdots + x_n \frac{\partial f}{\partial x_n}(x) = r f(x) \quad (\text{Eulersche Homogenitätsrelation});$$

(iii) für die radiale Ableitung gilt auf D : $|x| \partial_{\text{rad}} f(x) = r f(x)$;

(iv) die Skalen–Elastizität $\sigma_f(x) = \frac{1}{f(x)} \frac{d}{ds} \Big|_{s=1} f(sx)$ ist konstant r auf D .

Bei der letzten Aussage haben wir implizit – wie immer bei Elastizitäten – vorausgesetzt, dass f eine positive Funktion ist. Die Äquivalenz der Aussagen (ii) – (iv) wissen wir bereits, da $\sigma_f(x) = \frac{1}{f(x)} \partial_x f(x) = \frac{|x|}{f(x)} \partial_{\text{rad}} f(x)$ ist. Aus (i) folgt (iii) mit $\partial_x f(x) = \frac{d}{ds} \Big|_{s=1} f(sx) = \frac{d}{ds} \Big|_{s=1} s^r f(x) = r f(x)$. Gilt (iii), so folgt mit der Produktregel $\frac{d}{ds} (s^{-r} f(sx)) = -r s^{-r-1} f(sx) + s^{-r} \frac{d}{ds} f(sx) \equiv 0$, weil $\frac{d}{ds} f(sx) = \frac{d}{dt} \Big|_{t=0} f(sx + tx) = \partial_x f(sx) = \frac{1}{s} \partial_{sx} f(sx) = \frac{1}{s} r f(sx)$ ist. Daher ist dann $0 < s \mapsto s^{-r} f(sx)$ konstant gleich $f(x)$, und das ist (i).

BEMERKUNGEN: 1) Ist f (positiv) homogen vom Grad r , so sind die Ableitungen von f homogen von um 1 vermindertem Grad $r-1$. Das gilt für alle Arten von Ableitungen und folgt für Richtungsableitungen z.B. aus $(\partial_u f)(sx) = \frac{d}{dt} \Big|_{t=0} f(sx + tu) = \frac{d}{dt} \Big|_{t=0} s^r f(x + \frac{t}{s}u) = s^r \partial_{(1/s)u} f(x) = s^{r-1} \partial_u f(x)$.

2) Ist f positiv homogen auf \mathbb{R}^n vom Grad $r > 1$ mit $f(0) = 0$, so verschwinden alle Ableitungen von f in 0 wegen $\frac{1}{s}[f(su) - f(0)] = \frac{1}{|s|}|s|^r f(\pm u) = |s|^{r-1} f(\pm u) \rightarrow 0$ bei $s \rightarrow 0$.

3) Die auf ganz \mathbb{R}^n vom Grad $k \in \mathbb{N}_0$ (positiv) homogenen Funktionen f , die im Nullpunkt stetige partielle Ableitungen bis zur Ordnung k haben, sind genau die vom Grad k homogenen Polynomfunktionen. Für solche f gilt nämlich $\frac{d^k}{ds^k} \Big|_0 f(su) = (\partial_u)^k f(0)$ (k -fache Richtungsableitung in Richtung des Vektors $u \in \mathbb{R}^n$), wobei links $\frac{d^k}{ds^k} \Big|_0 s^k f(u) = k! f(u)$ herauskommt und rechts wegen der linearen Abhängigkeit der Richtungsableitung vom Richtungsvektor eine vom Grad k homogene Polynomfunktion $p(u)$ in der Variablen $u \in \mathbb{R}^n$. Mit $x = u$ folgt $f(x) = \frac{1}{k!} p(x)$. Es gibt aber viele nichtpolynomiale mit einem Grad $k \in \mathbb{N}$ positiv homogene Funktionen auf \mathbb{R}^n , die im Nullpunkt gar nicht oder nicht beliebig oft partiell differenzierbar sind, z.B. $f(x) = |x_j|$ oder $f(x) = |x| = \sqrt{(x_1)^2 + \dots + (x_n)^2}$.

4) Das Wachstum positiv homogener Funktionen $f(x)$ bei $|x| \rightarrow \infty$ ist höchstens wie $|x|^r$, wenn r der Homogenitätsgrad ist. Genauer gilt, wenn $c \leq f(u) \leq C$ für alle Einheitsvektoren ist: $c|x|^r \leq |x|^r f(\frac{1}{|x|}x) = f(x) \leq C|x|^r$ für alle $x \neq 0$. Ist die untere Schranke $c > 0$ und $r > 0$, so wächst daher $f(x)$ bei $|x| \rightarrow \infty$ mindestens wie die r -te Potenz $|x|^r$ des Abstands zum Ursprung. Ist dagegen $c \geq 0$ und $r < 0$, so strebt $f(x)$ bei $|x| \rightarrow \infty$ mindestens so schnell gegen Null wie die reziproke Potenz $1/|x|^{|r|}$.

Auf einfachen Wachstumseigenschaften homogener Funktionen basieren auch "ökonomische Gesetze" wie z.B. das Gesetz vom schließlich abnehmenden Durchschnittsertrag: Ist $f: \mathbb{R}_{>0}^n \rightarrow \mathbb{R}_{>0}$ eine vom Grad $0 < r < 1$ positiv homogene "Ertragsfunktion", die auf Einheitsvektoren beschränkt ist, so strebt für jedes $j = 1 \dots n$ der j -te partielle Durchschnittsertrag gegen Null d.h.

$$\frac{1}{x_j} f(x) \rightarrow 0 \quad \text{bei} \quad x_j \rightarrow \infty,$$

sofern die Variablen x_i mit $i \neq j$ beim Grenzübergang festgehalten werden oder jedenfalls nicht schneller als x_j gegen ∞ streben. Ist nämlich $f(x) \leq C < \infty$ eine obere Schranke für $|x| = 1$, so folgt $0 < \frac{1}{x_j} f(x) = \frac{1}{x_j} |x|^r f(\frac{1}{|x|}x) \leq \frac{1}{x_j} |x|^r C \leq \frac{1}{x_j} (\sqrt{n} B x_j)^r C \rightarrow 0$ bei $x_j \rightarrow \infty$ (beachte $r < 1$), wenn beim Grenzübergang $0 < x_i \leq B x_j$ bleibt für $i \neq j$ mit einer Konstanten $1 \leq B < \infty$, also $|x|^2 = (x_1)^2 + \dots + (x_n)^2 \leq n B^2 (x_j)^2$ und $|x| \leq \sqrt{n} B x_j$. Die ökonomische Bedeutung des Gesetzes ist, dass unter den gemachten Annahmen der Durchschnittsertrag bei übertriebenem Faktoreinsatz Null wird.

Beispiele homogener Funktionen, welche das Gesetz des schließlich abnehmenden Durchschnittsertrags erfüllen, sind **homogene neoklassische Funktionen**, d.h. $f(x_1, \dots, x_n)$ ist positiv und bzgl. jeder einzelnen Variablen $x_j > 0$ wachsend und konkav (bei c.p.-Bedingung) und f ist homogen auf $\mathbb{R}_{>0}^n$ mit einem Homogenitätsgrad $0 < r < 1$. Dazu gehören z.B. Cobb-Douglas-Funktionen

$$c x_1^{r_1} x_2^{r_2} \cdot \dots \cdot x_n^{r_n} \quad \text{mit allen } r_j > 0, r_1 + r_2 + \dots + r_n < 1, c > 0,$$

und auch Funktionen der Form

$$[c_1(x_1)^p + c_2(x_2)^p + \dots + c_n(x_n)^p]^{r/p} \quad \text{mit allen } c_j > 0, p > 0, 0 < r < 1. \quad \blacksquare$$