

Mathematik für Wirtschaftswissenschaftler

Vorlesungsprogramm für den 21. 06. 2007

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

5.3 Totales Differential und allgemeine Kettenregel

Bisher haben wir in diesem Kapitel nur die Änderung von Funktionswerten $F(x)$ betrachtet, wenn das Argument $\tilde{x} = x + hu$ auf einer Geraden durch x variiert wird (mit Richtungsvektor u und reellen Zahlen h von kleinem Betrag). Dies bedeutet, dass man feste Verhältnisse $(\tilde{x}_1 - x_1) : (\tilde{x}_2 - x_2) : \dots : (\tilde{x}_n - x_n) = u_1 : u_2 : \dots : u_n$ für die Änderungen der Komponenten annimmt. Im Allgemeinen gibt es dafür aber keinen vernünftigen Grund, sondern man muss *beliebige* Variationen \tilde{x} von x in Betracht ziehen und dafür die Änderung $F(\tilde{x}) - F(x)$ bestimmen. Ist der Abstand $|\tilde{x} - x|$ klein (und positiv), so ist diese Änderung näherungsweise gleich dem Produkt von $|\tilde{x} - x|$ mit der Richtungsableitung in Richtung des Einheitsvektors $u = \frac{1}{|\tilde{x} - x|}(\tilde{x} - x)$, und diese Richtungsableitung kann man gemäß 5.2 (bei "normalen" Funktionen) berechnen als Produkt $(\frac{\partial F}{\partial x}(x)) u$ der Ableitungsmatrix mit dem Richtungsvektor (was im skalaren Fall $F = f$ das Skalarprodukt $\nabla f(x) \cdot u$ des Gradienten mit dem Richtungsvektor ist). Wir sehen also, dass die Änderung $F(\tilde{x}) - F(x)$ der Funktionswerte näherungsweise durch das Produkt $(\frac{\partial F}{\partial x}(x))(\tilde{x} - x)$ der Ableitungsmatrix mit dem Änderungsvektor des Arguments gegeben ist.

Dieses Produkt ist nun eine *lineare* Funktion des Änderungsvektors, und damit sind wir bei der *Grundidee der Differentialrechnung*: Die Funktionsänderung $F(\tilde{x}) - F(x)$ wird für \tilde{x} nahe der (festen) Stelle x approximiert durch eine lineare Funktion $L(\tilde{x} - x) = A(\tilde{x} - x)$ der Argumentänderung, wobei A die Ableitungsmatrix von F an der Stelle x ist. Um aus dieser Idee einen brauchbaren Ableitungsbegriff zu entwickeln, muss man nur noch genauer sagen, wie gut diese Approximation ist, d.h. wie klein der Fehler $R(\tilde{x} - x) = F(\tilde{x}) - F(x) - L(\tilde{x} - x)$ ist. Da $L(\tilde{x} - x)$ für beliebige lineare Abbildungen L im Allgemeinen ein Vektor von derselben Größenordnung wie $|\tilde{x} - x|$ ist, hat die Approximation nur Sinn, wenn der Fehler von kleinerer Größenordnung ist als $|\tilde{x} - x|$, d.h. wenn $|R(\tilde{x} - x)|$ auch nach Division durch $|\tilde{x} - x|$ noch gegen Null geht bei $\tilde{x} \rightarrow x$. Dies ist tatsächlich der Fall; denn es gilt folgender

SATZ (beste lineare Approximation; Zuwachsformel): *Hat $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ an der inneren Stelle $x \in D$ stetige partielle Ableitungen, so gibt es genau eine (homogen) lineare Abbildung $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$, welche die Funktionsänderung bei x approximiert im Sinne folgender **Zuwachsformel**:*

$$\text{bzw.} \quad F(\tilde{x}) - F(x) = L(\tilde{x} - x) + R(\tilde{x} - x) \quad \text{für } \tilde{x} \in D$$

$$F(x + v) - F(x) = L(v) + R(v) \quad \text{für } x + v \in D,$$

wobei die Fehlergröße $|R(\tilde{x} - x)|$ noch nach Division mit $|\tilde{x} - x|$ gegen Null geht bei $\tilde{x} \rightarrow x$ (bzw. $\frac{1}{|v|}|R(v)| \rightarrow 0$ bei $|v| \rightarrow 0$). Diese beste lineare Approximation L ist gegeben durch Multiplikation der Ableitungsmatrix von F zu der Stelle x an die Argumentänderung,

$$L(v) = \left(\frac{\partial F}{\partial x}(x) \right) v \quad (= \nabla f(x) \cdot v \quad \text{im skalaren Fall } m = 1).$$

Die Aussage über die Fehlergröße bedeutet genau, dass es zu jeder gegebenen (noch so kleinen) Fehlerschranke $\varepsilon > 0$ ein $\delta > 0$ gibt, derart dass $|R(\tilde{x}-x)| \leq \varepsilon|\tilde{x}-x|$ ausfällt für alle $\tilde{x} \in \mathbb{R}^n$ mit $\max_{j=1}^n |\tilde{x}_j - x_j| \leq \delta$. Zum Beweis kann man zunächst annehmen, dass alle partiellen Ableitungen $\frac{\partial F}{\partial x_j}(x) = 0$ sind und $L = 0$ setzen (sonst subtrahiere einfach die durch die Ableitungsmatrix gegebene lineare Abbildung von F). Wegen der vorausgesetzten Stetigkeit der partiellen Ableitungen gibt es dann ein $\delta > 0$, derart dass $|\frac{\partial F}{\partial x_j}(\xi)| < \frac{1}{n}\varepsilon$ ist für $j = 1 \dots n$ an allen Stellen ξ mit $\max_{j=1}^n |\xi_j - x_j| < \delta$. Sind ξ und $\tilde{\xi} = \xi + te_j$ zwei solche Stellen, so folgt dann aus dem Schrankensatz (siehe 4.6) $|F(\tilde{\xi}) - F(\xi)| \leq \frac{1}{n}\varepsilon|t|$. (Der Schrankensatz wird hier bzgl. der j -ten Variablen angewandt; er gilt auch für vektorwertige Funktionen, wie man mit komponentenweiser Differentiation überlegen kann.) Das wendet man an mit $\xi^{[j]} = (\tilde{x}_1, \dots, \tilde{x}_j, x_{j+1}, \dots, x_n)$ und $\xi^{[j-1]} = (\tilde{x}_1, \dots, \tilde{x}_{j-1}, x_j, \dots, x_n)$ und erhält $|F(\tilde{x}) - F(x)| = \left| \sum_{j=1}^n [F(\xi^{[j]}) - F(\xi^{[j-1]})] \right| \leq \sum_{j=1}^n |F(\xi^{[j]}) - F(\xi^{[j-1]})| \leq \sum_{j=1}^n \frac{1}{n}\varepsilon|\tilde{x}_j - x_j| \leq \varepsilon|\tilde{x} - x|$ für alle \tilde{x} mit $\max_{j=1}^n |\tilde{x}_j - x_j| \leq \delta$. Die Eindeutigkeitsaussage gilt, weil aus der Zuwachsformel für zwei lineare Abbildungen L, \tilde{L} und $0 \neq v \in \mathbb{R}^n$ folgt $\frac{1}{|v|}|L(v) - \tilde{L}(v)| = \frac{1}{|tv|}|L(tv) - \tilde{L}(tv)| = \frac{1}{|tv|}|R(tv) - \tilde{R}(tv)| < 2\varepsilon$ für $t > 0$ hinreichend klein, wegen der Beliebigkeit von $\varepsilon > 0$ also $L(v) = \tilde{L}(v)$.

DEFINITION: Die homogen lineare Abbildung $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$, welche die Funktionsänderung von $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ bei der inneren Stelle $x \in D$ am besten approximiert im Sinne des obigen Satzes, heißt das (totale) **Differential** oder die **totale Ableitung** von F an der Stelle x und wird $d_x F$ notiert (oder $dF(x), DF(x), F'(x), \dots$). Die affine Abbildung $\tilde{x} \mapsto d_x F(\tilde{x} - x) + F(x)$, welche die Funktion $F(\tilde{x})$ selbst bei der Stelle x am besten approximiert, heißt die **beste (affin) lineare Approximation zu F bei x** . ■

DISKUSSION: 1) Zunächst sollte man sich über Folgendes im Klaren sein:

- *Das Differential $d_x F$ einer Funktion F an einer Stelle x ist eine homogen lineare Abbildung, und zwar von dem Vektorraum, in dem der Definitionsbereich von F liegt, in den Vektorraum, in dem F seine Werte nimmt.*

Wenn also $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ abbildet, so ist $d_x F$ eine homogen lineare Abbildung von \mathbb{R}^n nach \mathbb{R}^m . Eine solche Abbildung kann man bekanntlich (siehe Kap. 3) durch eine $m \times n$ -Matrix beschreiben in der Form $\mathbb{R}^n \ni v \mapsto Av \in \mathbb{R}^m$, wo rechts das Matrix-Vektor-Produkt von A mit dem Spaltenvektor v steht. Der obige Satz sagt uns:

- *Die Matrixdarstellung des Differentials $d_x F$ ist die Ableitungsmatrix $\left(\frac{\partial F}{\partial x}(x)\right)$ von F an der Stelle x , also ist $d_x F(v) = \left(\frac{\partial F}{\partial x}(x)\right)v \in \mathbb{R}^m$ für $v \in \mathbb{R}^n$.*

Damit ist die *Berechnung des Differentials* zurückgeführt auf die Berechnung der Einträge der Ableitungsmatrix, also der partiellen Ableitungen $\frac{\partial f_i}{\partial x_j}(x)$ aller Komponentenfunktionen f_i von F nach allen Variablen x_j an der Stelle x . Für die Anwendung des Differentials auf einem beliebigen Vektor v zeigt die Zuwachsformel $\frac{1}{t}[F(x + tv) - F(x)] = \frac{1}{t}d_x F(tv) + \frac{1}{t}R(tv) = d_x F(v) + |v|\frac{1}{|tv|}R(tv)$ für $0 \neq v \in \mathbb{R}^n$ und $t \in \mathbb{R}_{\neq 0}$, wobei die linke Seite gegen $\partial_v F(x)$ strebt bei $t \rightarrow 0$ nach Definition der Richtungsableitungen und die rechte gegen $d_x F(v)$ wegen der Kleinheitseigenschaft des Restes in der Zuwachsformel. Also hat man folgenden Zusammenhang zwischen Differential und Richtungsableitungen, der sich auch aus der in 5.2 schon gefundenen Formel $\partial_v f(x) = \left(\frac{\partial F}{\partial x}(x)\right)v$ ergibt:

- *Das Differential ist die Richtungsableitung der Funktion aufgefasst als (lineare) Funktion der Richtungsvektoren, d.h. $d_x F(v) = \partial_v F(x)$.*

2) Im *skalaren Fall* $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ist die Ableitungsmatrix von f im Wesentlichen der Gradient. Genauer ist die Ableitungsmatrix $\left(\frac{\partial f}{\partial x}(x)\right)$ die einzeilige Matrix mit denselben Einträgen $\frac{\partial f}{\partial x_j}(x)$ wie der Gradient $\nabla f(x)$ (den wir uns aber als Spaltenvektor vorstellen). Und die Multiplikation dieser Matrix von links an einen Spaltenvektor ist gleich dem Skalarprodukt des Gradienten mit diesem Vektor, also kann das Differential in diesem Fall durch den Gradienten ausgedrückt werden:

$$d_x F(v) = \nabla f(x) \bullet v.$$

3) Die mathematische (und auch ökonomische) *Bedeutung von Differential und Ableitungsmatrix* besteht darin, dass $d_x F$ die beste lineare Approximation der Funktionsänderung im Sinne der **Zuwachsformeln** ist:

$$F(\tilde{x}) - F(x) \approx d_x F(\tilde{x} - x) = \left(\frac{\partial F}{\partial x}(x)\right) (\tilde{x} - x), \quad F(x+v) - F(x) \approx d_x F(v) = \left(\frac{\partial F}{\partial x}(x)\right) v,$$

bzw. im skalaren Fall

$$f(\tilde{x}) - f(x) \approx \nabla f(x) \bullet (\tilde{x} - x), \quad f(x+v) - f(x) \approx \nabla f(x) \bullet v.$$

Diese Approximation ist umso besser, je kleiner der Abstand der Argumente $|\tilde{x} - x|$ bzw. die Größe der Argumentänderung $|v|$ ist. Genauer ist der Approximationsfehler von kleinerer Größenordnung als $|\tilde{x} - x|$ bzw. $|v|$, d.h. auch nach Division durch $|\tilde{x} - x|$ bzw. $|v|$ unterschreitet er noch jede beliebig klein gegebene Fehlerschranke, wenn nur die Argumentänderung klein genug ist. Dabei werden hier keinerlei Vorgaben für die Änderungen $\tilde{x}_j - x_j = v_j$ der einzelnen Komponenten gemacht, d.h. der Approximationsfehler strebt nach Division durch $|\tilde{x} - x|$ gegen Null, wenn sich \tilde{x} dem Punkt $x \in \mathbb{R}^n$ in ganz beliebiger Weise nähert, auf einer Geraden wie bei den Richtungsableitungen, auf einer Kurve wie etwa eine Parabel oder eine Spirale oder überhaupt auf irgendeine Weise, bei der die Abstände $|\tilde{x} - x|$ gegen Null gehen und die Richtungen $\frac{1}{|\tilde{x} - x|}(\tilde{x} - x)$ ganz beliebig sind.

4) Symbolisch wird die Zuwachsformel oft

$$dF = \frac{\partial F}{\partial x_1} dx_1 + \frac{\partial F}{\partial x_2} dx_2 + \dots + \frac{\partial F}{\partial x_n} dx_n$$

geschrieben, wobei gesagt wird, dass das Differential dF die “unendlich kleine Veränderung” sei, die der Funktionswert bei “unendlich kleinen Veränderungen” dx_j der Argumente erfahre. Es gibt aber keine “unendlich kleinen Größen“ außer Null, und deshalb sind solche Formulierungen eigentlich sinnlos und unverständlich. Einen Sinn erhält die Sache, wenn man kleine (aber nicht “unendlich kleine”) Änderungen $\Delta x_j = \tilde{x}_j - x_j$ der Komponenten von x betrachtet und die obige Gleichung als näherungsweise Beschreibung der Änderung $\Delta F(x) = F(\tilde{x}) - F(x)$ des Funktionswertes versteht:

$$\Delta F(x) \approx \frac{\partial F}{\partial x_1}(x) \Delta x_1 + \frac{\partial F}{\partial x_2}(x) \Delta x_2 + \dots + \frac{\partial F}{\partial x_n}(x) \Delta x_n$$

Das ist nichts anderes als die Zuwachsformel, wenn man man noch die genauere Aussage über die Größenordnung des Fehlers bei dieser Approximation hinzunimmt. Man sieht:

- Die näherungsweise Änderung des Funktionswertes bei einer Änderung des Argument x ist insgesamt die Summe der Änderungen, die der Funktionswert bei entsprechender Änderung der einzelnen Komponenten x_j von x erfahren würde.

Das liegt natürlich an der Linearität des Differentials. Für $\tilde{x} - x = \sum_{j=1}^n (\tilde{x}_j - x_j) e_j$ ist $F(\tilde{x}) - F(x) \approx d_x F(x)(\tilde{x} - x) = \sum_{j=1}^n (\tilde{x}_j - x_j) d_x F(e_j)$, und $d_x F(e_j) = \partial_{e_j} F(x)$ ist gerade die j -te partielle Ableitung.

Übrigens erhält die symbolische Formel $dF = \frac{\partial F}{\partial x_1} dx_1 + \dots + \frac{\partial F}{\partial x_n} dx_n$ doch einen mathematisch exakten Sinn, wenn man dx_j als Differential der j -ten Koordinatenfunktion $x \mapsto x_j$ interpretiert. Dann ist $d_x x_j$ die lineare Abbildung von \mathbb{R}^n nach \mathbb{R} , die dem j -ten kanonischen Basisvektor den Wert 1 und den anderen kanonischen Basisvektoren 0 zuordnet, und die symbolische Formel wird bei Einsetzen einer Stelle x eine korrekte Gleichung für lineare Abbildungen von \mathbb{R}^n nach \mathbb{R}^m (wenn F \mathbb{R}^m -wertig ist). Die Matrixdarstellung der linearen Abbildung $d_x x_j$ ist der j -te kanonische n -gliedrige Zeilenvektor, und die Gleichung $dF = \frac{\partial F}{\partial x_1} dx_1 + \dots + \frac{\partial F}{\partial x_n} dx_n$ sagt dann im Prinzip nur aus, dass die Ableitungsmatrix von $F = (f_1, \dots, f_m)$ spaltenweise zusammengesetzt ist aus den Vektoren $\frac{\partial F}{\partial x_j} = (\frac{\partial f_1}{\partial x_j}, \dots, \frac{\partial f_m}{\partial x_j})$. Diese Tautologie ist aber nicht die inhaltliche Bedeutung des Differentials, genauso wenig wie die formale Zusammenstellung aller partiellen Ableitungen aller Komponentenfunktionen zur Ableitungsmatrix. Die eigentliche Bedeutung ist die lineare Approximation der Funktionsänderungen durch die Zuwachsformeln zusammen mit der Aussage über die Größenordnung des Approximationsfehlers.

5) Eine Anwendung findet das Differential in der sog. **Fehlerrechnung**. Hier hat man einen geschätzten oder fehlerbehafteten Datensatz $x = (x_1, \dots, x_n)$ vorliegen und kennt die "wahren" Werte $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$ nicht, hat aber eine begründete Annahme über die Größe des Fehlervektors $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) = (\hat{x}_1 - x_1, \dots, \hat{x}_n - x_n) = \hat{x} - x$. Abzuschätzen ist dann der Fehler $\eta = \hat{y} - y = F(\hat{x}) - F(x)$ einer gemäß der Funktionsvorschrift F abhängigen Größe, der entsteht, wenn man F nicht am (unbekannten) "wahren" Wert \hat{x} auswertet, sondern in dem (bekannten) fehlerbehafteten Vektor x . Wenn hier der Argumentfehler $|\hat{x} - x| = |\varepsilon|$ klein angenommen werden darf, so gibt die Zuwachsformel für den Fehler der abhängigen Variablen den Näherungswert

$$\eta = \hat{y} - y \approx d_x F(\hat{x} - x) = d_x F(\varepsilon) = \left(\frac{\partial F}{\partial x}(x) \right) \varepsilon.$$

Man sieht, dass die Größe des Fehlervektors bei der abhängigen Variablen abhängt von der Größe der Einträge der Ableitungsmatrix von $F = (f_1, \dots, f_m)$ an der Stelle x .

Genauer kann man dazu sagen, wenn die Größenmessung für die Fehlervektoren festgelegt ist. Wählen wir z.B. das Maximum der Beträge der Einträge $|\varepsilon|_\infty = \max_{j=1}^n |\varepsilon_j|$, so können wir $|\eta_i| \approx \left| \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x) \varepsilon_j \right| \leq |\varepsilon|_\infty \sum_{j=1}^n \left| \frac{\partial f_i}{\partial x_j}(x) \right|$ abschätzen, wobei zuletzt Gleichheit eintritt, wenn alle Fehlerkomponenten $\varepsilon_j = \pm |\varepsilon|_\infty$ gleich groß sind und geeignete Vorzeichen haben. Es folgt dann, dass das Betragsmaximum $|\eta|_\infty$ der Komponenten des Fehlervektors bei der abhängigen Variablen (bestmöglich) abgeschätzt ist durch das Produkt $M \cdot |\varepsilon|_\infty$, wobei $M = \max_{i=1}^m \sum_{j=1}^n \left| \frac{\partial f_i}{\partial x_j}(x) \right|$ ist, also das *Maximum der Zeilenbetragssummen der Ableitungsmatrix*.

Nehmen wir dagegen die Euklidische Norm zur Messung der Fehlervektorgroße, so können wir mit der Cauchy-Schwarz-Ungleichung (s. 5.2) abschätzen $|\eta_i| \approx |\nabla f_i(x) \cdot \varepsilon| \leq |\nabla f_i(x)| \cdot |\varepsilon|$ und damit $|\eta|^2 \leq |\varepsilon|^2 \sum_{i=1}^m |\nabla f_i(x)|^2 = |\varepsilon|^2 \sum_{i=1}^m \sum_{j=1}^n \left(\frac{\partial f_i}{\partial x_j}(x) \right)^2$. Die Euklidische Länge des Fehlervektors $|\eta|$ bei der abhängigen Variablen ist also abgeschätzt durch das Produkt $M|\varepsilon|$, wobei nun M die *Quadratwurzel aus der Quadratsumme aller Einträge der Ableitungsmatrix* ist, also gewissermaßen die Euklidische Norm der Ableitungsmatrix. (Eine genauere Analyse zeigt, dass die kleinstmögliche Konstante M hier die Wurzel aus dem größten Eigenwert des Produkts $\left(\frac{\partial f_i}{\partial x_j}(x) \right)^T \left(\frac{\partial f_i}{\partial x_j}(x) \right)$ von der transponierten Ableitungsmatrix mit der Ableitungsmatrix ist.)

6) Die *geometrische Bedeutung* der bestapproximierenden affin linearen Abbildung zu einer differenzierbaren Funktion $F: D \rightarrow \mathbb{R}^m$ bei einer inneren Stelle $x \in D \subset \mathbb{R}^n$ ist, dass der Graph dieser affinen Abbildung der *Tangentialraum an den Graphen* von F im Punkte $(x, F(x)) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$ ist.

$$\tilde{y} = F(x) + d_x F(\tilde{x} - x)$$

ist also die Gleichung für die Punkte $(\tilde{x}, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ dieses Tangentialraums. Für die ökonomischen Anwendungen und das Verständnis des Differentials ist diese geometrische Interpretation aber nicht von Nutzen. Man kann sich den Graphen einer Funktion von mehreren Veränderlichen ohnehin nur im skalaren Fall $m = 1$ bei $n = 2$ Veränderlichen anschaulich vorstellen, als Fläche im dreidimensionalen Raum über dem Definitionsbereich; der Graph der besten linearen Approximation bei einer Stelle x ist dann die Tangentialebene an diese Fläche in dem senkrecht über x gelegenen Graphenpunkt.

7) Die Voraussetzung der Stetigkeit der partiellen Ableitungen im obigen Satz über die beste lineare Approximation ist, wie in 5.1 erläutert, für elementare Funktionen von mehreren Veränderlichen, also für alle konkreten ökonomischen Funktionen, an inneren Stellen ihres Definitionsbereichs immer erfüllt. In der Mathematik nimmt man die Gültigkeit einer Zuwachsformel $F(\tilde{x}) - F(x) = L(\tilde{x} - x) + R(\tilde{x} - x)$ mit einer linearen Abbildung $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $\frac{1}{|\tilde{x}-x|}R(\tilde{x} - x) \rightarrow 0$ bei $\tilde{x} \rightarrow x$ als Definition der (totalen) Differenzierbarkeit der Funktion $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ an der inneren Stelle x von D und des Differentials $d_x F = L$. Es folgt dann Existenz alle Richtungsableitungen an der Stelle x samt $\partial_v F(x) = d_x F(v)$, und die Matrixdarstellung von $L = d_x F$ ist durch die Ableitungsmatrix von F an der Stelle x gegeben. Die Existenz aller Richtungsableitungen ist allein noch kein vernünftiger Differenzierbarkeitsbegriff, selbst wenn man die lineare Abhängigkeit der Richtungsableitungen vom Richtungsvektor fordert, weil sie nicht ausschließt, dass sich die Funktion schlecht — evtl. sogar unstetig — verhält bei Annäherung an die Stelle x längs einer gekrümmten Kurve. Derartiges wird durch die Gültigkeit einer Zuwachsformel ausgeschlossen, und deshalb ist totale Differenzierbarkeit der “richtige” Differenzierbarkeitsbegriff.

8) Für positive stetig partiell (oder total) differenzierbare Funktionen von positiven Variablen x_1, \dots, x_n hat man eine analoge **relative Zuwachsformel**

$$\Delta_{\text{rel}} f(x) = \frac{f(\tilde{x}) - f(x)}{f(x)} \approx \mathcal{E}_f(x) \cdot \Delta_{\text{rel}} x = \sum_{j=1}^n \varepsilon_{f, x_j}(x) \frac{\tilde{x}_j - x_j}{x_j},$$

wo $\Delta_{\text{rel}} x = (\frac{\tilde{x}_1 - x_1}{x_1}, \dots, \frac{\tilde{x}_n - x_n}{x_n})$ der Vektor der relativen Argumentänderungen ist. Für vektorielle $F = (f_1, \dots, f_m)$ mit positiven Komponentenfunktionen und den Vektor $\Delta_{\text{rel}} F(x) = (\Delta_{\text{rel}} f_1(x), \dots, \Delta_{\text{rel}} f_m(x))$ der relativen Komponentenfunktionsänderungen gilt entsprechend mit der Elastizitätsmatrix an Stelle des Elastizitätsgradienten:

$$\Delta_{\text{rel}} F(x) \approx \left(\varepsilon_{f_i, x_j}(x) \right) \Delta_{\text{rel}} x.$$

Diese Zuwachsformeln ergeben sich einfach, indem man auf f bzw. F die (nichtrelative) Zuwachsformel anwendet und durch $f(x)$ dividiert. Der Fehler bei der relativen Zuwachsformel ist daher kleiner als $\varepsilon |\Delta_{\text{rel}} x|$, wenn alle relativen Argumentänderungen $|\frac{\tilde{x}_j - x_j}{x_j}|$ hinreichend klein sind.

Man kann die relative Zuwachsformel auch so interpretieren, dass statt $f(x_1, \dots, x_n)$ der Logarithmus $\varphi(\xi_1, \dots, \xi_n) = \ln f(e^{x_1}, \dots, e^{x_n})$ als Funktion der Logarithmen $\xi_j = \ln x_j$ der Argumente betrachtet wird. Die partiellen Ableitungen von φ sind dann gerade die Elastizitäten $\frac{\partial \varphi}{\partial \xi_j}(\xi) = \frac{1}{\varphi(\xi)} \frac{\partial f}{\partial x_j}(x) e^{\xi_j} = \frac{1}{f(x)} \frac{\partial f}{\partial x_j}(x) x_j = \varepsilon_{f, x_j}(x)$ und der Gradient von φ an der Stelle ξ ist gleich dem Elastizitätsgradienten $\mathcal{E}_f(x)$. Entsprechend ist im vektoriellen Fall $F = (f_1, \dots, f_m)$ die Ableitungsmatrix von $\Phi(\xi)$ mit den analog definierten Komponentenfunktionen $\varphi_i(\xi)$ zu f_i gleich der Elastizitäten-Matrix von F . Die relative Zuwachsformel für F ist also im Wesentlichen dasselbe wie die gewöhnliche Zuwachsformel für die logarithmisch (bei Argumenten und Wertekomponenten) reskalierte Funktion. ■

BEISPIELE: (1) $F(x, y, z) = (x^2(1+y), x^3e^{yz})$ ist eine \mathbb{R}^2 -wertige elementare Funktion auf \mathbb{R}^3 . Die Ableitungsmatrix an der Stelle $(x, y, z) \in \mathbb{R}^3$ ist

$$\begin{pmatrix} 2x(1+y) & x^2 & 0 \\ 3x^2e^{yz} & x^3ze^{yz} & x^3ye^{yz} \end{pmatrix}.$$

Diese 2×3 -Matrix stellt also die totale Ableitung $d_{(x,y,z)}F: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ dar. Für kleine Störungen (u, v, w) von (x, y, z) gilt dann

$$F(x+u, y+v, z+w) - F(x, y, z) \approx \begin{pmatrix} 2x(1+y) & x^2 & 0 \\ 3x^2e^{yz} & x^3ze^{yz} & x^3ye^{yz} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix},$$

wobei die Norm der Differenz der beiden Seiten von \approx für gegebenes $\varepsilon > 0$ kleiner ist als $\varepsilon\sqrt{u^2+v^2+w^2}$, wenn $|u|$, $|v|$ und $|w|$ genügend klein sind.

(2) Wir betrachten eine *Produktionsfunktion* $x(A, K) = 2A^{0.2}K^{0.8}$ in Abhängigkeit von den Produktionsfaktoren Arbeit A und Kapital K . Bei **partieller Faktorvariation**, d.h. Veränderung von A zu \tilde{A} bei festem K oder von K zu \tilde{K} bei festem A gilt

$$\begin{aligned} x(\tilde{A}, K) - x(A, K) &\approx \frac{\partial x}{\partial A}(A, K)(\tilde{A} - A) = 0.4 A^{-0.8} K^{0.8}(\tilde{A} - A), \\ x(A, \tilde{K}) - x(A, K) &\approx \frac{\partial x}{\partial K}(A, K)(\tilde{K} - K) = 1.6 A^{0.2} K^{-0.2}(\tilde{K} - K). \end{aligned}$$

Bei **totaler Faktorvariation**, d.h. simultaner Änderung von A und K ist die Output-Änderung ungefähr die Summe der partiellen Faktorvariationen (vgl. Punkt 4) oben:

$$\begin{aligned} x(\tilde{A}, \tilde{K}) - x(A, K) &\approx \frac{\partial x}{\partial A}(A, K)(\tilde{A} - A) + \frac{\partial x}{\partial K}(A, K)(\tilde{K} - K) \\ &= 0.4 A^{-0.8} K^{0.8}(\tilde{A} - A) + 1.6 A^{0.2} K^{-0.2}(\tilde{K} - K). \end{aligned}$$

Da der Fehler bei dieser Approximation von kleinerer Größenordnung ist als die Argumentänderungen $|\tilde{A} - A|$, $|\tilde{K} - K|$ wird er in der Praxis oft einfach ignoriert. Bei relativer partieller Faktorvariation von $a\%$ bei A bzw. $k\%$ bei K sind die relativen Output-Änderungen

$$\begin{aligned} \frac{x(\tilde{A}, K) - x(A, K)}{x(A, K)} &\approx \varepsilon_{x,A}(A, K) \cdot \frac{a}{100} = 0.2 \frac{a}{100}, \\ \frac{x(A, \tilde{K}) - x(A, K)}{x(A, K)} &\approx \varepsilon_{x,K}(A, K) \cdot \frac{k}{100} = 0.8 \frac{k}{100}, \end{aligned}$$

und bei simultaner Änderung beider Faktoren um diese Prozentsätze ist die relative Output-Änderung

$$\frac{x(\tilde{A}, \tilde{K}) - x(A, K)}{x(A, K)} \approx \varepsilon_{x,A}(A, K) \cdot \frac{a}{100} + \varepsilon_{x,K}(A, K) \cdot \frac{k}{100} = 0.2 \frac{a}{100} + 0.8 \frac{k}{100}.$$

(3) Ganz analog lässt sich bei Produktionsfunktionen $x(r_1, \dots, r_n)$ von mehreren Faktoren die Outputänderung bei kleinen Faktorvariationen näherungsweise beschreiben durch

$$\begin{aligned} x(\tilde{r}_1, \dots, \tilde{r}_n) - x(r_1, \dots, r_n) &\approx \nabla x(r) \cdot (\tilde{r} - r) \\ &= (\tilde{r}_1 - r_1) \frac{\partial x}{\partial r_1}(r_1, \dots, r_n) + \dots + (\tilde{r}_n - r_n) \frac{\partial x}{\partial r_n}(r_1, \dots, r_n), \end{aligned}$$

wobei rechts die Summe der Outputänderungen bei partieller Faktorvariation von r_j zu \tilde{r}_j c.p. für $j = 1 \dots n$ steht. Die relative Änderung des Output bei Veränderung von r_j um $p_j\%$ für $j = 1 \dots n$ ist

$$\begin{aligned} \frac{x(\tilde{r}_1, \dots, \tilde{r}_n) - x(r_1, \dots, r_n)}{x(r_1, \dots, r_n)} &\approx \mathcal{E}_x(r) \cdot \frac{p}{100} \\ &= \frac{p_1}{100} \varepsilon_{x, r_1}(r_1, \dots, r_n) + \dots + \frac{p_n}{100} \varepsilon_{x, r_n}(r_1, \dots, r_n). \end{aligned} \quad \blacksquare$$

Eine Anwendung des Konzepts der totalen Ableitung (Differentials) ist die folgende allgemeine Kettenregel für Funktionen von mehreren Veränderlichen. Diese Kettenregel kann man eigentlich nur mit dem Konzept der linearen Approximation richtig verstehen (und sich merken), und deswegen haben wir das Differential oben auch eingeführt (um außerdem bei dieser Gelegenheit auch noch gleich mit den in Büchern zur Wirtschaftsmathematik verbreiteten “unendlich kleinen Differentials” etwas aufzuräumen).

Zur Motivation betrachten wir folgende Situation: Ein Unternehmen fertigt m Produkte, deren Outputs durch m Produktionsfunktionen $x(r) = (x_1(r), \dots, x_m(r))$ in Abhängigkeit von n Produktionsfaktoren $r = (r_1, \dots, r_n)$ gegeben sind. Der beim Verkauf erzielbare Erlös sei durch eine Erlösfunktion $E(x) = E(x_1, \dots, x_m)$ beschrieben, die von den hergestellten (und, wie wir annehmen, auch absetzbaren) Produktmengen abhängt. Wir stellen uns vor, dass die produzierten Güter teilweise substitutiv und teilweise komplementär sind, so dass der Marktpreis für die einzelnen Produkte von den am Markt angebotenen Mengen mehrerer oder aller m Produkte abhängt. Wird nun der Faktoreinsatz von r zu $\tilde{r} = (\tilde{r}_1, \dots, \tilde{r}_n)$ (wenig) geändert, so erhält man eine entsprechende Änderung der Produktions-Outputs von x zu $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_m)$, den man näherungsweise mit der Zuwachsformel für die vektorielle Produktionsfunktion x berechnen kann zu $\tilde{x} - x \approx \left(\frac{\partial x}{\partial r}(x)\right) (\tilde{r} - r) = \sum_{j=1}^n (\tilde{r}_j - r_j) \frac{\partial x}{\partial r_j}(x)$. Interessiert ist man eigentlich aber an der mittelbar von der Output-Änderung abhängenden Erlösänderung $E(\tilde{x}) - E(x)$. Nun ist mit $|\tilde{r} - r|$ auch $|\tilde{x} - x|$ klein, also erhält man die Erlösänderung näherungsweise mit der Zuwachsformel $E(\tilde{x}) - E(x) \approx \left(\frac{\partial E}{\partial x}(x)\right) (\tilde{x} - x) = \sum_{i=1}^m (\tilde{x}_i - x_i) \frac{\partial E}{\partial x_i}(x)$. Substituiert man in dieser Näherungsformel den zuvor für $\tilde{x} - x$ angegebenen Näherungsausdruck, so gelangt man zu

$$E(\tilde{x}) - E(x) \approx \left(\frac{\partial E}{\partial x}(x(r))\right) \left(\frac{\partial x}{\partial r}(r)\right) (\tilde{r} - r) = \sum_{i=1}^m \sum_{j=1}^n \frac{\partial E}{\partial x_i}(x(r)) \frac{\partial x_i}{\partial r_j}(r) (\tilde{r}_j - r_j),$$

d.h. die mittelbare Änderung des Erlöses ist in erster Näherung durch durch Multiplikation des Produkts der Ableitungsmatrizen zu E (an der Stelle $x(r)$) und x (an der Stelle r) an den Änderungsvektor $\tilde{r} - r$ gegeben. Dieselbe Überlegung führt zu dem

SATZ (allgemeine Kettenregel): Ist $F: \mathbb{R}^n \supset D \rightarrow E \subset \mathbb{R}^m$ differenzierbar an der inneren Stelle $x \in D$ und $G: \mathbb{R}^m \supset E \rightarrow \mathbb{R}^l$ differenzierbar an der inneren Stelle $y = F(x)$ von E , so ist auch die Verkettung $G \circ F$ an der Stelle x differenzierbar, und das Differential $d_x(G \circ F)$ ist die Komposition der homogen linearen Abbildungen $d_y G: \mathbb{R}^m \rightarrow \mathbb{R}^l$ und $d_x F: \mathbb{R}^n \rightarrow \mathbb{R}^m$,

$$d_x(G \circ F) = (d_{F(x)}G)(d_x F): \mathbb{R}^n \rightarrow \mathbb{R}^l \quad \text{homogen linear.}$$

Mit anderen Worten: Die $l \times n$ -Ableitungsmatrix der Verkettung $G \circ F$ an der Stelle x ist das Produkt der $l \times m$ -Ableitungsmatrix der äußeren Funktion G an der Stelle $F(x)$ mit der $m \times n$ -Ableitungsmatrix der inneren Funktion F an der Stelle x ,

$$\left(\frac{\partial(G \circ F)}{\partial x}(x) \right) = \left(\frac{\partial G}{\partial y}(F(x)) \right) \left(\frac{\partial F}{\partial x}(x) \right).$$

Der Beweis ist nichts weiter als eine Ausarbeitung der folgenden Idee: Wenn $L(u) \approx F(x+u) - F(x)$ eine lineare Approximation der Änderung von F bei x ist und $M(v) \approx G(y+v) - G(y)$ eine lineare Approximation der Änderung von G bei $y = F(x)$, so wird die Hintereinanderausführung $M(L(u)) \approx G(y+L(u)) - G(y) \approx G(F(x+u)) - G(F(x))$ eine lineare Approximation zu $G \circ F$ bei x sein. Schreiben wir genauer $F(x+u) - F(x) = L(u) + |u|r(u)$ mit dem Differential $L = d_x F$ und einem Restterm $|r(u)| \rightarrow 0$ bei $|u| \rightarrow 0$ in \mathbb{R}^n sowie $G(y+v) - G(y) = M(v) + |v|s(v)$ mit dem Differential $d_y G$ und einem Restterm $|s(v)| \rightarrow 0$ bei $|v| \rightarrow 0$ in \mathbb{R}^m , so ergibt sich durch Einsetzen von $F(x)$ für y und $F(x+u) - F(x)$ für v (wenn $u \neq 0$):

$$G(F(x+u)) - G(F(x)) = M(L(u)) + |u| \left[M(r(u)) + \frac{|v|}{|u|} s(v) \right].$$

Ist C der Betrag des größten Eintrags der Matrixdarstellung von L , so gilt $|L(u)| \leq \sqrt{mn}C|u|$. Wählt man $|u|$ so klein, dass $|r(u)| \leq 1$ ist, so folgt $|v| = |F(x+u) - F(x)| = |L(u) + |u|r(u)| \leq (\sqrt{mn}C + 1)|u| = C_1|u|$. Analog sieht man $|M(r(u))| \leq C_2|r(u)|$ mit einer von M abhängigen endlichen Konstanten. Der Ausdruck $t(u) = [\dots]$ oben hat daher eine Norm $\leq C_2|r(u)| + C_1|s(v)|$, und das ist kleiner als eine beliebig gegebene Fehlerschranke $\varepsilon > 0$ für $|u|$ klein genug, weil erstens $|r(u)| \rightarrow 0$ geht bei $|u| \rightarrow 0$, weil zweitens $|s(v)| \rightarrow 0$ gilt bei $|v| \rightarrow 0$ und weil drittens wegen $|v| \leq C_1|u|$ mit $|u| \rightarrow 0$ auch $|v| \rightarrow 0$ geht. Daher ist $t(u)$ ein Term mit $|t(u)| \rightarrow 0$ bei $|u| \rightarrow 0$, und die obige Formel $G(F(x+u)) - G(F(x)) = (ML)(u) + |u|t(u)$ eine Zuwachsformel für $G \circ F$ mit Restterm von kleinerer Größenordnung als $|u|$. Das bedeutet aber gerade, dass $ML = (d_y G)(d_x F)$ das Differential von $G \circ F$ an der Stelle x ist.

DISKUSSION: 1) Ganz explizit mit Komponentenfunktionen $F = (f_1, \dots, f_m)$ und $G = (g_1, \dots, g_l)$ und partiellen Ableitungen geschrieben gibt die allgemeine Kettenregel die folgende Formel für die Einträge $\frac{\partial}{\partial x_j}(g_h \circ F)$ der Ableitungsmatrix von $G \circ F$ ($h = 1 \dots l, j = 1 \dots n$):

$$\frac{\partial(g_h \circ F)}{\partial x_j}(x) = \sum_{i=1}^m \frac{\partial g_h}{\partial y_i}(y) \frac{\partial f_i}{\partial x_j}(x) \quad (y = F(x))$$

bzw. ganz ausführlich

$$\begin{aligned} & \frac{\partial}{\partial x_j} [g_h(f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n))] \\ &= \frac{\partial g_h}{\partial y_1}(y_1, y_2, \dots, y_m) \frac{\partial f_1}{\partial x_j}(x_1, x_2, \dots, x_n) \\ &+ \frac{\partial g_h}{\partial y_2}(y_1, y_2, \dots, y_m) \frac{\partial f_2}{\partial x_j}(x_1, x_2, \dots, x_n) \quad (y_i = f_i(x_1, \dots, x_n)) \\ &\vdots \\ &+ \frac{\partial g_h}{\partial y_m}(y_1, y_2, \dots, y_m) \frac{\partial f_m}{\partial x_j}(x_1, x_2, \dots, x_n). \end{aligned}$$

- Man erhält die partielle Ableitung einer Verkettung nach der gewünschten Variablen, indem man die äußere Funktion nach jedem ihrer Argumente differenziert, diese partielle Ableitung dann mit der gewünschten partiellen Ableitung der zugehörigen inneren Funktion multipliziert und die so entstehenden Produkte aufsummiert.

2) Spezialfälle der Kettenregel: Ist $F = f$ eine skalare Funktion ($m = 1$) und g eine Funktion von einer einzigen Veränderlichen (skalar oder vektoriell), so gilt

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

Das ist die **spezielle Kettenregel** aus 5.1, die unmittelbar aus der Kettenregel der Differentialrechnung in einer Veränderlichen folgt (siehe 4.3). Ein anderer Spezialfall der allgemeinen Kettenregel liegt vor, wenn die innere Funktion nur von einer reellen Variablen abhängt ($n = 1$). Dann ist die innere Funktion $F(t) = c(t) = (c_1(t), \dots, c_m(t))$ eine Kurve in \mathbb{R}^m , und die allgemeine Kettenregel lautet hier:

$$\frac{d}{dt} G(c(t)) = c'_1(t) \frac{\partial G}{\partial y_1}(c(t)) + \dots + c'_m(t) \frac{\partial G}{\partial y_m}(c(t)) = (\partial_{c'(t)} G)(c(t)).$$

Das ist die Regel für die **Ableitung längs einer Kurve** $c(t)$, die wir schon in 5.2 angegeben hatten; man kann sie berechnen als Richtungsableitung in Richtung des Tangentenvektors $c'(t)$ der differenzierbaren Kurve.

3) Zur Notation: Hat man Variable x_1, \dots, x_n , abhängig davon Variable y_1, \dots, y_m und von diesen wiederum abhängig z_1, \dots, z_l , so wird die Kettenregel oft suggestiv

$$\frac{\partial z_h}{\partial x_j} = \sum_{i=1}^m \frac{\partial z_h}{\partial y_i} \frac{\partial y_i}{\partial x_j} = \frac{\partial z_h}{\partial y_1} \frac{\partial y_1}{\partial x_j} + \dots + \frac{\partial z_h}{\partial y_m} \frac{\partial y_m}{\partial x_j}$$

geschrieben oder in Matrix-Form

$$\left(\frac{\partial z_h}{\partial x_j} \right) = \left(\frac{\partial z_h}{\partial y_i} \right) \left(\frac{\partial y_i}{\partial x_j} \right).$$

Warnung: Hier steht aber z_h links und rechts nicht für dieselbe Funktion: Rechts ist die “unmittelbare Funktion” $z_h(y_1, \dots, y_m)$ gemeint, links aber die “mittelbare Funktion” $z_h(y_1(x_1, \dots, x_n), \dots, y_m(x_1, \dots, x_n))$. Daher ist, wenn $\frac{\partial z_h}{\partial x_j}$ links an der Stelle $x = (x_1, \dots, x_n)$ auszuwerten ist, rechts zwar $\frac{\partial y_i}{\partial x_j}$ auch an dieser Stelle zu nehmen, aber $\frac{\partial z_h}{\partial y_i}$ an der zugehörigen Stelle $y = (y_1(x), \dots, y_m(x))$. (Alles andere würde auch gar keinen Sinn machen: Man kann x im Allgemeinen schon deswegen nicht in $\frac{\partial z_h}{\partial y_i}$ einsetzen, weil diese partielle Ableitung eine Funktion von m Variablen ist wie $z_h(y_1, \dots, y_m)$, während $x = (x_1, \dots, x_n)$ ein Vektor mit n Komponenten ist.)

4) Zuwachsformel und Fehlerfortpflanzung bei mittelbaren Funktionen: Die Zuwachsformel für die “mittelbare Funktion” $G(F(x))$ lautet unter Berücksichtigung der Kettenregel

$$\begin{aligned} G(F(\tilde{x})) - G(F(x)) &\approx (d_{F(x)}G)(d_x F)(\tilde{x} - x) = \left(\frac{\partial G}{\partial y}(F(x))\right)\left(\frac{\partial F}{\partial x}(x)\right)(\tilde{x} - x) \\ &= \sum_{i=1}^m \sum_{j=1}^n (\tilde{x}_j - x_j) \frac{\partial f_i}{\partial x_j}(x) \frac{\partial G}{\partial y_i}(F(x)). \end{aligned}$$

Entsprechendes gilt für den Fehler bei der mittelbaren Funktion, wenn man \tilde{x} als den “wahren Wert” der Vektorvariablen x ansieht und x als fehlerbehaftet. Dann ist $\varepsilon = \tilde{x} - x$ der Fehlervektor bei der unabhängigen Variablen, $\eta = F(\tilde{x}) - F(x) \approx (d_x F)\varepsilon$ der näherungsweise Fehler beim “unmittelbaren” Funktionswert $F(x)$ und $\zeta = G(F(\tilde{x})) - G(F(x)) \approx (d_{F(x)}G)\eta \approx (d_{F(x)}G)(d_x F)\varepsilon$ der approximative Fehler beim “mittelbaren” Funktionswert $G(F(x))$. Für den durch Anwendung der beiden Funktionen G nach F fortgepflanzten Anfangsfehler gilt also in erster Näherung:

$$\zeta \approx \left(\frac{\partial G}{\partial y}(F(x))\right)\eta \approx \underbrace{\left(\frac{\partial G}{\partial y}(F(x))\right)}_B \underbrace{\left(\frac{\partial F}{\partial x}(x)\right)}_A \varepsilon.$$

Man erkennt, dass die Größe des mittelbaren Fehlervektors von der Größe der Einträge des Produkts BA der beiden rechts stehenden Ableitungsmatrizen abhängt. Weiter oben haben wir erklärt, dass z.B. für das Maximum der Komponentenbeträge gilt $|\eta|_\infty \leq |A|_{\infty,1}|\varepsilon|_\infty$, also

$$|\zeta|_\infty \leq |B|_{\infty,1}|A|_{\infty,1}|\varepsilon|_\infty,$$

wobei $|A|_{\infty,1}$ die größte Zeilenbetragssumme der Matrix A ist und entsprechend $|B|_{\infty,1}$. Messen wir die Fehlervektorgroße mit der Euklidischen Norm, so gilt $|\eta| \leq |A|_{2,2}|\varepsilon|$, also

$$|\zeta| \leq |B|_{2,2}|A|_{2,2}|\varepsilon|,$$

wobei nun $|A|_{2,2}$ bzw. $|B|_{2,2}$ die Wurzel aus der Quadratsumme aller Einträge von A bzw. B bezeichnen. Die Fortpflanzungsformel zeigt, dass eine unter Umständen bessere Abschätzung

$$|\zeta|_\infty \leq |BA|_{\infty,1}|\varepsilon|_\infty \quad \text{oder} \quad |\zeta| \leq |BA|_{2,2}|\varepsilon|$$

möglich ist, in die nur die Größe der Einträge der Produktmatrix eingeht. Jedenfalls ist diese Abschätzung nicht schlechter, wie man sich überlegen kann, unter Umständen aber wesentlich besser (wenn nämlich A und B große Einträge haben, aber BA nur kleine — Extremfall $A \neq 0$, $B \neq 0$, aber $BA = 0$).

5) Die Kettenregel für die Verkettung von mehr als zwei differenzierbaren Funktionen gilt analog. Zum Beispiel ist das Differential von $H \circ G \circ F$ an der Stelle x die Hintereinanderausführung der homogen linearen Abbildungen $d_{G(F(x))}H$ nach $d_{F(x)}G$ nach $d_x F$ und die Ableitungsmatrix von $H \circ G \circ F$ ist entsprechend das Matrixprodukt der Ableitungsmatrizen von H , G und F an den jeweiligen Stellen:

$$\left(\frac{\partial(H \circ G \circ F)}{\partial x}\right) = \left(\frac{\partial H}{\partial z}(G(F(x)))\right)\left(\frac{\partial G}{\partial y}(F(x))\right)\left(\frac{\partial F}{\partial x}(x)\right).$$

Das ergibt sich einfach durch Klammerung $H \circ (G \circ F)$ und zweimalige Anwendung der Kettenregel. Die Aussagen in 4) über Zuwachsformel und Fehlerfortpflanzung gelten entsprechend auch bei Verkettung von mehr als zwei differenzierbaren Funktionen.

6) Für die partiellen Elastizitäten von Verkettungen haben wir (unter den üblichen Positivitätsannahmen):

$$\begin{aligned} \varepsilon_{g_h \circ F, x_j}(x) &= \frac{x_j}{g_h \circ F(x)} \frac{\partial(g_h \circ F)}{\partial x_j}(x) = \sum_{i=1}^m \frac{x_j}{g_h(F(x))} \frac{\partial g_h}{\partial y_i}(F(x)) \frac{\partial f_i}{\partial x_j}(x) \\ &= \sum_{i=1}^m \frac{f_i(x)}{g_h(F(x))} \frac{\partial g_h}{\partial y_i}(F(x)) \cdot \frac{x_j}{f_i(x)} \frac{\partial f_i}{\partial x_j}(x) = \sum_{i=1}^m \varepsilon_{g_h, y_i}(F(x)) \cdot \varepsilon_{f_i, x_j}(x). \end{aligned}$$

Das ist die **allgemeine Kettenregel für Elastizitäten**. Mit Elastizitäts-Matrizen ausgedrückt lautet sie

$$\left(\varepsilon_{g_h \circ F, x_j}(x) \right) = \left(\varepsilon_{g_h, y_i}(y) \right) \left(\varepsilon_{f_i, x_j}(x) \right) \quad (y = F(x)).$$

- Die Elastizitäten-Matrix der Verkettung $G \circ H$ von $G = (g_1, \dots, g_m)$ mit $F = (f_1, \dots, f_n)$ an der Stelle x ist das Produkt der Elastizitäten-Matrix von G an der Stelle $y = F(x)$ mit der Elastizitäten-Matrix von F an der Stelle x .

Man kann das auch direkt auf die allgemeine Kettenregel zurückführen, indem man die frühere Beobachtung verwendet, dass die Elastizitäten-Matrix einer Funktion gleich der Ableitungsmatrix der bei abhängigen und unabhängigen Variablen logarithmisch reskalierten Funktion ist. Suggestiv, aber weniger genau wird die Kettenregel für Elastizitäten so geschrieben

$$\left(\varepsilon_{z_h, x_j} \right) = \left(\varepsilon_{z_h, y_i} \right) \left(\varepsilon_{y_i, x_j} \right) \quad \text{bzw.} \quad \varepsilon_{z_h, x_j} = \sum_{i=1}^m \varepsilon_{z_h, y_i} \cdot \varepsilon_{y_i, x_j}.$$

Ökonomisch interpretiert ist $\varepsilon_{z_h, x_j}(x)$ für $z_h = g_h(F(x))$ der Prozentsatz, um den sich die mittelbar von x abhängende Variable z_h (ungefähr) ändert, wenn x_j c.p. um 1% erhöht wird. Dazu muss man für *alle* unmittelbar von x abhängigen Variablen $y_i = f_i(x)$ die Elastizitäten $\varepsilon_{z_h, y_i}(y)$ berechnen, mit der prozentualen Änderung der jeweiligen Variablen $\varepsilon_{y_i, x_j}(x)$ multiplizieren und die so entstehenden Produkte aufsummieren — das besagt die Kettenregel für Elastizitäten. ■

BEISPIELE: (1) Wir betrachten die Funktionen $F(x, y, z) = (x^2(1+y), x^3e^{yz})$ und $G(u, v) = (u^2 - v^2, 2uv)$. Dann bildet ist Verkettung $G \circ F$ eine Vektorfunktion mit zwei Komponenten (weil die äußere Funktion G zwei Komponentenfunktionen g_1, g_2 hat) von drei Variablen (weil die innere Funktion von drei Variablen abhängt), explizit gegeben durch Einsetzen der Komponenten f_1, f_2 von F für u und v :

$$G \circ F(x, y, z) = (x^4(1+y)^2 - x^6e^{2yz}, 2x^5(1+y)e^{yz}).$$

Die 2×3 -Ableitungsmatrix von $G \circ F$ kann man nun als Produkt der Ableitungsmatrizen von G an der Stelle $u = f_1(x, y, z), v = f_2(x, y, z)$ mit der Ableitungsmatrix von F an der Stelle (x, y, z) berechnen:

$$\begin{aligned} \left(\frac{\partial(G \circ F)}{\partial(x, y, z)} \right) &= \begin{pmatrix} 2u & -2v \\ 2v & 2u \end{pmatrix} \begin{pmatrix} 2x(1+y) & x^2 & 0 \\ 3x^2e^{yz} & x^3ze^{yz} & x^3ye^{yz} \end{pmatrix} \quad \begin{pmatrix} u = x^2(1+y), \\ v = x^3e^{yz} \end{pmatrix} \\ &= \begin{pmatrix} 2x^2(1+y) & -2x^3e^{yz} \\ 2x^3e^{yz} & 2x^2(1+y) \end{pmatrix} \begin{pmatrix} 2x(1+y) & x^2 & 0 \\ 3x^2e^{yz} & x^3ze^{yz} & x^3ye^{yz} \end{pmatrix} \\ &= \begin{pmatrix} 4x^3(1+y)^2 - 6x^5e^{2yz} & 2x^4(1+y) - 2x^6ze^{2yz} & -2x^6ye^{2yz} \\ 10x^4(1+y)e^{yz} & 2x^5(1+z+yz)e^{yz} & 2x^5y(1+y)e^{yz} \end{pmatrix}. \end{aligned}$$

Natürlich kann man die partiellen Ableitungen auch direkt mit der angegebenen expliziten Formel für $G \circ F$ ausrechnen; das Resultat ist — bei korrekter Rechnung — dasselbe.

(2) Die Ableitung der Funktion F aus (1) längs der Kurve $c(t) = (\sqrt{t}, 1/t, t^2)$ ($t > 0$) ist

$$\begin{aligned} \frac{d}{dt}F(c(t)) &= \left(\frac{\partial F}{\partial(x, y, z)}(c(t)) \right) c'(t) = c'_1(t) \frac{\partial F}{\partial x}(c(t)) + c'_2(t) \frac{\partial F}{\partial y}(c(t)) + c'_3(t) \frac{\partial F}{\partial z}(c(t)) \\ &= \frac{1}{2\sqrt{t}} \begin{pmatrix} 2\sqrt{t}(1+1/t) \\ 3te^t \end{pmatrix} - \frac{1}{t^2} \begin{pmatrix} t \\ t^3\sqrt{te^t} \end{pmatrix} + 2t \begin{pmatrix} 0 \\ \sqrt{te^t} \end{pmatrix} = \begin{pmatrix} 1 \\ (t+\frac{3}{2})\sqrt{te^t} \end{pmatrix}, \end{aligned}$$

was man durch Differentiation von $F(c(t)) = (t(1+1/t), t\sqrt{te^t}) = (t, t^{3/2}e^t)$ bestätigt. Der letzte Rechenweg (erst einsetzen, dann differenzieren) muss nicht immer einfacher sein, wie hier. Außerdem gibt die Kettenregel, wenn man erst einmal die Ableitungsmatrix von F berechnet hat, die Ableitung von F längs beliebiger Kurven $c(t)$ durch Berechnung von $c'(t)$ und eine simple Matrix–Vektor–Multiplikation.

(3) Wir betrachten eine Produktion von m Produkten. Der Output des i -ten Produkts wird beschrieben durch eine *Produktionsfunktion* $x_i(r_1, \dots, r_n)$ in Abhängigkeit vom Einsatz der Produktionsfaktoren r_1, \dots, r_n . (Nicht jeder Faktor r_j muss in jeder Produktionsfunktion x_i vorkommen; zum Beispiel ist denkbar, dass r_j den Einsatz eines Rohstoffs beschreibt, der für die Herstellung des i -ten Produkts gar nicht benötigt wird. In einem solchen Fall ist eben $\frac{\partial x_i}{\partial r_j} = 0$ überall. Unter Umständen ist es besser, die für Produktion des i -ten Produkts eingesetzte Menge des j -ten Faktors mit r_{ij} zu bezeichnen, also $x_i(x_{i1}, x_{i2}, \dots)$ zu schreiben unabhängig von den Faktoreinsatzmengen r_{hj} mit $h \neq i$. Hier stellen wir uns dann alle $r_{11}, r_{12}, \dots, r_{m1}, r_{m2}, \dots$ in eine Reihe r_1, r_2, \dots, r_n geschrieben vor.)

Die Kosten der Produktion sind bestimmt durch die Kosten für die eingesetzten Faktormengen, was durch eine *Kostenfunktion* $K(r_1, \dots, r_n)$ beschrieben wird. Im einfachsten Fall ist das eine lineare Funktion $k_1 r_1 + \dots + k_n r_n + k_0$. Der *Erlös* für produzierte und auch abgesetzte Produktmengen x_i ist $E(x_1, \dots, x_m) = x_1 p_1(x_1, \dots, x_m) + \dots + x_m p_m(x_1, \dots, x_m)$, wobei der *Marktpreis* p_i im einfachsten Fall konstant ist (wenn das Produkt bei vollständiger Konkurrenz am Gütermarkt abgesetzt wird), in komplizierteren Fällen aber sowohl von der angebotenen Menge x_i des i -ten Produkts monoton fallend abhängt als auch von den angebotenen Mengen x_h , $h \neq i$, einiger oder aller anderen hergestellten Güter abhängig ist (wobei p_i fallend ist bzgl. x_h , wenn die Güter Nr. i und Nr. h substitutiv sind, dagegen wachsend bei Komplementarität dieser Güter).

Will man den **optimalen Faktoreinsatz** bestimmen, so ist zunächst die *Gewinnfunktion* in Abhängigkeit vom Einsatz der Produktionsfaktoren aufzustellen:

$$\begin{aligned} G(r_1, \dots, r_n) &= E(x_1(r_1, \dots, r_n), \dots, x_m(r_1, \dots, r_n)) - K(r_1, \dots, r_n) \\ &= E(x(r)) - K(r) = x(r) \cdot p(x(r)) - K(r), \end{aligned}$$

wobei $r = (r_1, \dots, r_n)$ der Faktoreinsatzvektor, $x(r) = (x_1(r), \dots, x_m(r))$ der zugehörige Produktionsvektor und $p(x) = (p_1(x), \dots, p_m(x))$ der durch den Output–Vektor x bestimmte Preisvektor ist. Die Gewinnänderung bei Erhöhung des j -ten Faktoreinsatzes um eine (kleine) Einheit ist gegeben durch $\frac{\partial G}{\partial r_j}(r)$, was mit der Produktregel $\frac{\partial E}{\partial x_i}(x) = \frac{\partial}{\partial x_i}(x \cdot p(x)) = p_i(x) + x \cdot \frac{\partial p}{\partial x_i}(x)$ und der Kettenregel wie folgt berechnet wird:

$$\begin{aligned} \frac{\partial G}{\partial r_j}(r) &= \sum_{i=1}^m \frac{\partial E}{\partial x_i}(x(r)) \frac{\partial x_i}{\partial r_j}(r) - \frac{\partial K}{\partial r_j}(r) \\ &= \sum_{i=1}^m \left[p_i(x(r)) + x(r) \cdot \frac{\partial p}{\partial x_i}(x(r)) \right] \frac{\partial x_i}{\partial r_j}(r) - \frac{\partial K}{\partial r_j}(r). \end{aligned}$$

Bei Gewinn-maximierendem Faktoreinsatz hat der Gewinn ein Maximum bzgl. jedes einzelnen Faktors, also ist dann $\frac{\partial G}{\partial r_j}(r) = 0$ für $j = 1 \dots n$. (Das ist das notwendige Kriterium für innere Extremstellen bei Funktionen von einer Veränderlichen r_j , siehe 4.5 und auch 5.4; dabei ist vorausgesetzt, dass der Faktoreinsatz nicht durch Kapazitätsgrenzen eingeschränkt ist.) Für den optimalen Faktoreinsatz gelten also die folgenden Gleichungen, deren Aufstellung **Grenzproduktivitätstheorie** genannt wird:

$$\sum_{i=1}^m \frac{\partial E}{\partial x_i}(x(r)) \frac{\partial x_i}{\partial r_j}(r) = \frac{\partial K}{\partial r_j}(r) \quad \text{für } j = 1 \dots n.$$

In ökonomischem Jargon wird dieses Ergebnis so formuliert:

- *Bei optimalem Faktoreinsatz sind für jeden Faktor die Grenzkosten gleich dem Gesamtwert aller mit ihrem Grenzerlös bewerteten Grenzproduktivitäten des Faktors.*

Die Grenzkosten bzgl. des j -ten Produktionsfaktors sind dabei $\frac{\partial K}{\partial r_j}(r)$; sie geben die Kostensteigerung bei Erhöhung des Faktoreinsatzes r_j um eine (kleine) Einheit an. Die Grenzproduktivität von x_i bzgl. r_j ist $\frac{\partial x_i}{\partial r_j}(r)$ und beschreibt die entsprechende Erhöhung des i -ten Outputs. Der Grenzerlös $\frac{\partial E}{\partial x_i}(x(r))$ ist der Preis, zu dem eine zusätzliche Outputeinheit des i -ten Produkts am Markt abgesetzt werden kann, also ist "die mit ihrem Grenzerlös bewertete Grenzproduktivität" $\frac{\partial E}{\partial x_i}(x(r)) \frac{\partial x_i}{\partial r_j}(r)$ der zusätzliche Erlös für das i -te Produkt bei Erhöhung von r_j um eine Einheit. Die Grenzproduktivitätstheorie besagt, dass im Gewinn-Maximum der zusätzliche Gesamterlös bei Erhöhung eines Faktoreinsatzes c.p. gleich den zusätzlichen Kosten ist, sofern kein Faktoreinsatz an einer Kapazitätsgrenze erfolgt. Dieses ökonomische "Gleichgewichtsgesetz" ist mathematisch gesehen nichts anderes als die notwendige Bedingung für Extremstellen in Verbindung mit der Kettenregel $\frac{\partial}{\partial r_j} E(x(r)) = \sum_{i=1}^m \frac{\partial E}{\partial x_i}(x(r)) \frac{\partial x_i}{\partial r_j}(r)$.

(4) Gelegentlich kommt es vor, dass *eine Funktion mittelbar und unmittelbar von denselben Variablen abhängt*. Damit ist gemeint, dass $f(x_1, \dots, x_n)$ unter Bedingungen an die Variablen x_1, \dots, x_n betrachtet wird, welche es ermöglichen, einige von ihnen durch die anderen auszudrücken. Nehmen wir etwa an, dass x_j für $j = k+1 \dots n$ als Funktion der ersten k Variablen ausgedrückt ist

$$x_j = \varphi_j(x_1, \dots, x_k) \quad (j = k+1 \dots n),$$

so hat man nun die Funktion

$$g(x_1, \dots, x_k, \varphi_{k+1}(x_1, \dots, x_k), \dots, \varphi_n(x_1, \dots, x_k))$$

zu betrachten, die nur noch von x_1, \dots, x_k abhängt, und zwar "unmittelbar" über die ersten k Argumente von f und "mittelbar" über die letzten $n-k$ Argumente. Die partielle Ableitung $\frac{\partial g}{\partial x_i}$ wird für $i = 1 \dots k$ nun mit der Kettenregel berechnet, wobei für die inneren Funktionen x_h einfach $\frac{\partial x_h}{\partial x_i} \equiv 0$ ist, wenn $h \neq i$, bzw. $\frac{\partial x_h}{\partial x_i} \equiv 1$, wenn $h = i$. Das Resultat ist:

$$\begin{aligned} \frac{\partial g}{\partial x_i}(x_1, \dots, x_k) &= \frac{\partial f}{\partial x_i}(x_1, \dots, x_k, \varphi_{k+1}(x_1, \dots, x_k), \dots, \varphi_n(x_1, \dots, x_k)) \\ &+ \sum_{j=k+1}^n \frac{\partial f}{\partial x_j}(x_1, \dots, x_k, \varphi_{k+1}(x_1, \dots, x_k), \dots, \varphi_n(x_1, \dots, x_k)) \frac{\partial \varphi_j}{\partial x_i}(x_1, \dots, x_k). \end{aligned}$$

- *Die partielle Ableitung einer Funktion nach einer Variablen, von der sie mittelbar und unmittelbar abhängt, ist die Summe der "unmittelbaren partiellen Ableitung" und aller gemäß Kettenregel berechneten "mittelbaren partiellen Ableitungen".*

Das gilt für skalare und (komponentenweise) auch für Vektorfunktionen.

(5) Ein ökonomisches Beispiel zu (4) ist ein *Dyopol*, bei dem es zwei Anbieter für ein Produkt gibt und der Marktpreis von der Summe $x = x_1 + x_2$ der beiden angebotenen Mengen abhängt, also durch eine Funktion $p(x) = p(x_1 + x_2)$ von einer Variablen gegeben ist. Die Gewinnfunktion des ersten Anbieters ist dann

$$G(x_1, x_2) = x_1 \cdot p(x_1 + x_2) - K_1(x_1),$$

wobei $K_1(x_1)$ seine Kosten für die Herstellung von x_1 Einheiten des Produkts sind. Reagiert der zweite Anbieter auf die Mengenzpolitik des ersten in der Weise, dass er sein Angebot ausschließlich nach der Angebotsmenge des ersten Anbieters richtet (diese Annahme ist natürlich eine grobe Vereinfachung), so wird dies durch eine *Reaktionsfunktion* $x_2 = R_2(x_1)$ beschrieben, und die Gewinnfunktion des ersten Anbieters kann als Funktion seines Angebots x_1 alleine aufgefasst werden:

$$G_1(x_1) = G(x_1, R_2(x_1)) = x_1 p(x_1 + R_2(x_1)) - K_1(x_1).$$

Für Gewinn-Maximierung beim ersten Anbieter lautet die notwendige Bedingung:

$$0 = \frac{dG_1}{dx_1}(x_1) = \underbrace{\frac{\partial G}{\partial x_1}(x_1, R_2(x_1))}_{\text{unmittelbare Abl. nach } x_1} + \underbrace{\frac{\partial G}{\partial x_2}(x_1, R_2(x_1)) \frac{dR_2}{dx_1}(x_1)}_{\text{mittelbare Ableitung nach } x_1}.$$

Hätte der erste Anbieter seinen Grenzgewinn ohne Rücksicht auf die Reaktion seines Konkurrenten ermittelt, so hätte er seinen optimalen Output stattdessen aus der Gleichung

$$\frac{\partial G}{\partial x_1}(x_1, x_2) = 0$$

bei bekanntem und als konstant angenommenem Konkurrenz-Output x_2 berechnet. Man sieht, dass die Optimalitätsbedingung bei Berücksichtigung der Reaktion des zweiten Anbieters nicht dieselbe ist: Links tritt noch ein Summand hinzu, der von $\frac{\partial G}{\partial x_2}$ abhängt und dadurch die Abhängigkeit des Gewinns beim ersten Anbieter vom Output des zweiten Anbieters berücksichtigt, und zwar um so stärker, je größer dessen "Grenzreaktion" $\frac{dR_2}{dx_1}(x_1)$ ist. Die letztgenannte Größe beschreibt die Änderung der Angebotspolitik des zweiten Anbieters in Abhängigkeit vom Angebot x_1 des ersten und heißt *Reaktionskoeffizient* des zweiten Anbieters bzgl. des ersten. In der Praxis ist das Problem natürlich, dass man die Reaktionsfunktion und ihre Ableitung, den Reaktionskoeffizienten, nicht genau kennt, sondern schätzen muss, um das eigene Angebot unter Berücksichtigung der gesamten Marktsituation optimal gestalten zu können.

Ähnliche, aber komplexere, Überlegungen lassen sich für ein *Oligopol* von $n \geq 3$ Anbietern desselben Produkts durchführen. Dann hat man einen Satz von n Reaktionsfunktionen $x_j = R_j(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$, die im einfachsten Fall nur von der Summe $x_1 + \dots + x_{j-1} + x_{j+1} + \dots + x_n$ abhängen. Da sich mit x_1 nun x_2, \dots, x_n ändern, ist der Reaktionskoeffizient des j -ten bzgl. des ersten Anbieters hier nicht einfach die unmittelbare partielle Ableitung $\frac{\partial R_j}{\partial x_1}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$. Vielmehr muss man aus den Reaktionsgleichungen für $j = 2 \dots n$ die Werte von x_2, \dots, x_n als Funktion der Variablen x_1 bestimmen, um die Marktreaktion auf eine Änderung bei x_1 darzustellen, und dann $R_j(x_1, x_2(x_1), \dots, x_{j-1}(x_1), x_{j+1}(x_1), \dots, x_n(x_1))$ mit der Kettenregel nach x_1 ableiten, um den Reaktionskoeffizienten des j -ten Anbieters bzgl. Änderungen bei x_1 zu erhalten. ■

Wenn man bei einer differenzierbaren Funktion F die Änderung $F(\tilde{x}) - F(x)$ durch die lineare Approximation $d_x F(\tilde{x} - x)$ ersetzt, so macht man (wenn F nichtlinear ist) im Allgemeinen einen Fehler. Dieser Fehler ist zwar von kleinerer Größenordnung als die Größe $|\tilde{x} - x|$ der Änderung des Arguments, aber das ist nur eine qualitative Aussage, die für konkrete Werte von x, \tilde{x} keine konkrete Fehlerabschätzung und daher auch keine exakte Abschätzung für $|F(\tilde{x}) - F(x)|$ ermöglicht. Dafür sind *exakte Schranken* für die Werteänderung $|F(\tilde{x}) - F(x)|$ erforderlich, und solche liefert der folgende

SATZ (Schranken-Satz): *Hat $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ stetige partielle Ableitungen und liegt die Strecke von x nach \tilde{x} in D , so gilt für die (Euklidische) Norm der Wertedifferenz*

$$|F(\tilde{x}) - F(x)| \leq L|\tilde{x} - x|,$$

wenn an allen Stellen dieser Strecke die Richtungsableitung von F in Richtung der Strecke höchstens die Norm L hat.

Im skalaren Fall $F = f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ ist dies einfach der Schrankensatz für die reelle Funktion $\varphi(t) = f(x + t(\tilde{x} - x))$ von einer Variablen $t \in [0, 1]$. Es gilt nämlich $\varphi'(t) = \partial_{\tilde{x}-x} f(x + t(\tilde{x} - x)) = |\tilde{x} - x| \partial_u f(x + t(\tilde{x} - x))$, wenn $u = \frac{1}{|\tilde{x}-x|}(\tilde{x} - x)$ der Einheitsvektor in Richtung der Strecke von x nach \tilde{x} ist. (Im Fall $x = \tilde{x}$ ist nichts zu zeigen.) Die Voraussetzung $|\partial_u f| \leq L$ überall auf dieser Strecke hat daher $|\varphi'(t)| \leq L|\tilde{x} - x|$ für $0 \leq t \leq 1$ zur Folge, also liefert der Schrankensatz aus 4.6: $|f(\tilde{x}) - f(x)| = |\varphi(1) - \varphi(0)| \leq L|\tilde{x} - x|$. Im vektoriellen Fall betrachten wir den Einheitsvektor $v \in \mathbb{R}^m$ in Richtung von $F(\tilde{x}) - F(x)$ (im Fall $F(\tilde{x}) = F(x)$ ist wieder nichts zu zeigen) und die Funktion $\psi(t) = v \cdot F(x + t(\tilde{x} - x))$. Wie eben findet man $\psi'(t) = v \cdot \partial_{\tilde{x}-x} F(x + t(\tilde{x} - x))$ und damit $|\psi'(t)| \leq L|\tilde{x} - x|$ für $0 \leq t \leq 1$ (Cauchy-Schwarz-Ungleichung; beachte $|v| = 1$). Gemäß Wahl von v folgt $|F(\tilde{x}) - F(x)| = v \cdot [F(\tilde{x}) - F(x)] = \psi(1) - \psi(0) \leq L|\tilde{x} - x|$.

KOROLLAR (Konstanz-Satz): *Ist die Ableitungsmatrix (bzw. das Differential) von $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ überall Null auf der offenen Menge D und sind je zwei Punkte aus D durch einen in D gelegenen Streckenzug verbindbar, so ist die Funktion F konstant auf D .*

Nach dem Schrankensatz ist nämlich F auf jeder Strecke in D konstant, weil alle Richtungsableitungen Null sind überall auf D .

BEMERKUNGEN: 1) Die **optimale Schranke** L , die für alle Strecken in D funktioniert, ist das Supremum der Normen der Richtungsableitungen $\partial_u F(x)$ über alle Einheitsvektoren $u \in \mathbb{R}^n$ und alle Punkte $x \in D$ (wenn dieses Supremum endlich ist). Ist D eine *konvexe Menge*, d.h. mit zwei Punkten x, \tilde{x} liegt auch ihre Verbindungsstrecke $[x, \tilde{x}] = \{x + t(\tilde{x} - x) : 0 \leq t \leq 1\}$ in D , so ist diese Schranke die *kleinstmögliche Dehnungsschranke* (auch *Lipschitz-Konstante* genannt) für F auf D , d.h. es gilt $|F(\tilde{x}) - F(x)| \leq L|\tilde{x} - x|$ für alle $x, \tilde{x} \in D$, und das ist mit keiner kleineren Konstanten als L für alle $x, \tilde{x} \in D$ richtig.

2) Für $F = (f_1, \dots, f_m)$ und $|u| = 1$ ist das Norm-Quadrat $|\partial_u F(x)|^2 = \sum_{i=1}^m (\partial_u f_i(x))^2 = \sum_{i=1}^m (u \cdot \nabla f_i(x))^2 \leq \sum_{i=1}^m |\nabla f_i(x)|^2 = \sum_{i=1}^m \sum_{j=1}^n \left(\frac{\partial f_i}{\partial x_j}(x)\right)^2$ abgeschätzt durch die Quadratsumme aller Einträge der Ableitungsmatrix. Daher gilt:

- *Als (im Allgemeinen nichtoptimale) Konstante L im Schranken-Satz kann man stets das Supremum der Euklidischen Norm der Ableitungsmatrix auf der Strecke von x nach \tilde{x} nehmen (also der Wurzel aus der Quadratsumme aller Einträge).*

3) Man kann zeigen, dass die partiellen Ableitungen aller Komponentenfunktionen von F auf jeder abgeschlossenen und beschränkten Menge im Definitionsbereich D beschränkt sind, wenn sie darauf existieren und stetig sind. Daher folgt aus 2):

- *Jede stetig partiell differenzierbare Funktion, insbesondere also jede elementare Funktion, von mehreren Veränderlichen ist dehnungsbeschränkt auf abgeschlossenen konvexen Mengen (Kugeln, Würfeln, Quadern, ...) in ihrem offenen Definitionsbereich.*

4) Der Schrankensatz gilt auch, wenn man die Größe der Vektoren mit anderen Normen $\|\cdot\|$ auf \mathbb{R}^n und $\|\cdot\|$ auf \mathbb{R}^m misst als mit den Euklidischen Normen, z.B. mit der Summe oder dem Maximum der Beträge ihrer Komponenten. Es gilt dann $\|F(\tilde{x}) - F(x)\| \leq L\|\tilde{x} - x\|$, wenn die Richtungsableitung $\partial_u F(\xi)$ für den Vektor $u \in \mathbb{R}^n$ der Norm $\|u\| = 1$ in Richtung von $\tilde{x} - x$ an allen Stellen ξ der Strecke von x nach \tilde{x} eine Norm $\|\partial_u F(\xi)\| \leq L$ hat. Nehmen wir z.B. in \mathbb{R}^n und im Zielraum \mathbb{R}^m jeweils das Maximum $|v|_\infty$ der Beträge der Einträge eines Vektors v als Norm, so gilt $|F(\tilde{x}) - F(x)|_\infty \leq L|\tilde{x} - x|_\infty$, wenn das Maximum der Zeilenbetragssummen der Ableitungsmatrix von F an jeder Stelle der Strecke $[x, \tilde{x}]$ höchstens so groß ist wie L .

5) Um den Fehler bei Linearisierung $F(\tilde{x}) - F(x) \approx d_x F(\tilde{x} - x)$ genau abzuschätzen betrachten wir die Abbildung $G(\tilde{x}) = F(\tilde{x}) - d_x F(\tilde{x} - x)$ mit $G(\tilde{x}) - G(x) = F(\tilde{x}) - F(x) - d_x F(\tilde{x} - x)$. Da die lineare Funktion $\tilde{x} \mapsto d_x F(\tilde{x} - x)$ an allen Stellen das Differential $d_x F$ hat, gilt $d_\xi G = d_\xi F - d_x F$. Der Schrankensatz besagt nun $|G(\tilde{x}) - G(x)| \leq \varepsilon|\tilde{x} - x|$, wenn die Richtungsableitung $\partial_u G(\xi) = \partial_u F(\xi) - \partial_u F(x)$ in Richtung von $u = \frac{1}{|\tilde{x} - x|}(\tilde{x} - x)$ an allen Stellen ξ der Strecke $[x, \tilde{x}]$ (Euklidische) Norm $\leq \varepsilon$ hat. Gemäß 2) ist das insbesondere der Fall, wenn die Differenz der Ableitungsmatrizen von F an den Stellen ξ und an der Stelle x stets Euklidische Norm $\leq \varepsilon$ hat. Damit haben wir folgende **Abschätzung des Linearisierungsfehlers**:

- *Für den Linearisierungsfehler bei Linearisierung von F an der Stelle x gilt*

$$|F(\tilde{x}) - F(x) - d_x F(\tilde{x} - x)| \leq \varepsilon|\tilde{x} - x|,$$

wenn die Richtungsableitung $\partial_u F(\xi)$ in Richtung von $u = \frac{1}{|\tilde{x} - x|}(\tilde{x} - x)$ sich an allen Stellen ξ der Strecke von x nach \tilde{x} von derselben Richtungsableitung $\partial_u F(x)$ im Punkt x höchstens um einen Vektor der Euklidischen Norm $\leq \varepsilon$ unterscheidet.

- *Das ist insbesondere dann der Fall, wenn sich die Ableitungsmatrix von F an all diesen Stellen ξ von der Ableitungsmatrix an der Stelle x nur um eine Matrix der Euklidischen Norm $\leq \varepsilon$ unterscheidet.*

Wir erinnern daran, dass man die lineare Funktion $d_x F(\tilde{x} - x)$ konkret als Produkt $(\frac{\partial f_i}{\partial x_j}(x))(\tilde{x} - x)$ der Ableitungsmatrix mit dem (Spalten-)Vektor $\tilde{x} - x$ berechnet. Grob gesagt hat der Linearisierungsfehler die Größe $\lesssim \varepsilon|\tilde{x} - x|$, wenn die partiellen Ableitungen von F auf der Strecke von x nach \tilde{x} um nicht mehr als ε variieren. Bei Stetigkeit der partiellen Ableitungen ist diese Bedingung jedenfalls erfüllt, wenn \tilde{x} nahe genug bei x liegt. Kennt man quantitative Stetigkeitseigenschaften der partiellen Ableitungen z.B. im Sinne einer Dehnungsschranke, so kann man zu gegebenem $\varepsilon > 0$ nun auch quantifizieren, wie nahe \tilde{x} bei x liegen muss, damit die gewünschte Kleinheit $\leq \varepsilon|\tilde{x} - x|$ des Linearisierungsfehlers mit Sicherheit eingehalten ist. ■

5.4 Extremstellenbestimmung, innere Extremstellen

In diesem und im nächsten Abschnitt geht es um die Bestimmung der Maximum- und Minimumstellen einer gegebenen Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ von mehreren Variablen auf einem gegebenen Bereich $D \subset \mathbb{R}^n$. (Diese Problemstellung ist natürlich nur bei *reellwertigen* Funktionen sinnvoll.) Die Terminologie ist dabei ganz wie bei der entsprechenden Aufgabenstellung für Funktionen von einer Veränderlichen in 4.5 und wird zunächst der Vollständigkeit halber wiederholt:

TERMINOLOGIE und DISKUSSION (*Extremstellen und Extremwerte*):

1) Für eine gegebene reelle Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ heißt eine Stelle $x_* \in D$ **(absolute) Minimumstelle** von f auf D , wenn f an keiner anderen Stelle in D einen kleineren Funktionswert hat, wenn also gilt $f(x_*) \leq f(x)$ für alle $x \in D$. Der Funktionswert an einer solchen Stelle heißt dann das **(absolute) Minimum** der Funktion f auf dem Definitionsbereich D , und man sagt, dass f auf D an der Stelle x_* ein **Minimum annimmt**. Wird das Minimum allein an der Stelle x_* angenommen, gilt also $f(x) > f(x_*)$ für alle $x \in D$ mit $x \neq x_*$, so heißt x_* die **eindeutige Minimumstelle** zu f auf D , und wir sagen, dass f ein **einziges Minimum** auf D an der Stelle x_* hat. Entsprechend heißt $x^* \in D$ eine **(absolute) Maximumstelle** von f auf D , und f **nimmt ein Maximum an** im Punkt x^* auf D , wenn $f(x^*) \geq f(x)$ ist für alle $x \in D$. Der Wert $f(x^*)$ heißt dann das **(absolute) Maximum** der Funktion f auf D , und f hat auf D ein **einziges Maximum** im Punkt x^* , wenn sogar $f(x) < f(x^*)$ gilt für alle $x \in D$ mit $x \neq x^*$, in welchem Falle x^* die **eindeutige Maximumstelle** von f auf D genannt wird. Der Sammelbegriff für Maximum- und Minimumstellen ist **(absolute) Extremstellen**, und das Maximum wie das Minimum nennt man einen **(absoluten) Extremwert** oder **Extremum** oder **Optimum** von f auf D . Als Notation für das Minimum und Maximum von f auf D verwendet man, wenn sie existieren:

$$\max_{x \in D} f(x), \quad \min_{x \in D} f(x), \quad \text{oder kurz} \quad \max_D f, \quad \min_D f.$$

2) Man sollte

- *die Extremstellen von den Extremwerten (Extrema) unterscheiden!*

Normalerweise interessieren bei einer Extremwertaufgabe nicht nur die Extremwerte der untersuchten Funktion f , sondern auch und vor allem die Stellen im Definitionsbereich, an denen diese Werte angenommen werden. Man möchte eben z.B. wissen, wie man den Produktions-Outputvektor x wählen muss, damit der Gewinn $G(x)$ maximal ausfällt, und nicht nur, wie groß der maximale Gewinn ist. Kennt man die Extremstellen, so kann man die Extremwerte einfach durch Auswerten der Funktion in diesen Punkten (Einsetzen in die Funktion) ausrechnen.

3) Wenn von Maximum und Minimum die Rede ist, so sollte man sich immer zunächst vor Augen halten:

- *Extremstellen brauchen nicht zu existieren; und wenn sie existieren, so brauchen sie nicht eindeutig zu sein!*

Das Letzte ist klar; eine Funktion kann natürlich ihr Maximum und/oder Minimum an mehreren verschiedenen Stellen annehmen (Extremes Beispiel: Konstante Funktion).

Zum Nachweis der Existenz von Extremstellen ist der *Extremstellensatz* das entscheidende Hilfsmittel, den wir weiter unten für Funktionen von mehreren Veränderlichen beweisen.

Was *immer* existiert, wenn wir dafür auch nötigenfalls die Werte ∞ oder $-\infty$ zulassen, ist das *Infimum* $\inf_D f$ und das *Supremum* $\sup_D f$, also die größte untere Schranke und die kleinste obere Schranke für die Funktionswerte von f auf D . Wenn $\sup_D f = \infty$ ist, so nimmt die Funktion beliebig große Werte auf der Menge D an; dann kann es natürlich kein Maximum, keinen größten Funktionswert, geben; denn ∞ ist kein zugelassener Funktionswert. *Maximum und Minimum sind immer endliche Werte, soweit sie existieren.* Ebenso kann kein Minimum von f auf D existieren, wenn $\inf_D f = -\infty$ ist, wenn also die Funktion negative Werte beliebig großen Betrags annimmt. Aber auch wenn Supremum und Infimum endlich sind, gibt es im allgemeinen keine Extrema, weil Supremum bzw. Infimum nicht als Funktionswerte auf D angenommen werden müssen. (Genau wenn es ein Funktionswert auf D ist, ist das Supremum gleich dem Maximum bzw. das Infimum gleich dem Minimum.) Es kann z.B. sein, dass die Funktionswerte nur bei Annäherung der Argumente an nicht zu D gehörende Randpunkte oder an den “unendlich fernen Punkt” gegen Supremum oder Infimum streben, aber diese Werte auf D nie erreichen. Oder die Funktion ist vielleicht an den Stellen in D , “wo eigentlich die Extremstellen sein müssten”, unstetig und hat dort nicht die “richtigen” Werte. Simple Beispiele für diese Situation haben wir schon in 4.2 bei Funktionen einer Veränderlichen gesehen.

4) In der reellen Analysis ist man oft nicht sehr präzise bei der Angabe von Definitionsbereichen; bei Funktionen, die durch Rechterme gegeben sind, denkt man sich z.B. oft den maximalen Definitionsbereich des Terms in \mathbb{R}^n . Bei Extremwertaufgaben muss man aber genauer sein; denn

- *absolute Extremstellen und Extremwerte einer Funktion hängen davon ab, welcher Definitionsbereich D , betrachtet wird; daher muss man den Optimierungsbereich genau spezifizieren, wenn von Extremstellen oder Extremwerten die Rede ist.*

Eine Maximumstelle zu f auf D ist im Allgemeinen keine mehr zu f auf einem größeren Definitionsbereich $\tilde{D} \supset D$; denn bei Vergrößerung von D können ja neue Funktionswerte hinzukommen, die größer sein können als das Maximum von f auf D . Aus demselben Grund ist z.B. auch das absolute Maximum zu f auf D im Allgemeinen größer als das Maximum von f auf einem kleineren Definitionsbereich $D' \subset D$, vorausgesetzt die beiden Maxima existieren überhaupt. Nur wenn D' schon eine der Maximumstellen zu f auf D enthält, ist das Maximum von f auf D' dasselbe wie auf D (und dann sind die Maximumstellen zu f auf D' genau diejenigen Maximumstellen von f auf D , die schon in D' liegen).

5) Eine Stelle $x_* \in D \subset \mathbb{R}^n$ heißt **lokale Minimumstelle** oder **relative Minimumstelle** zu der reellen Funktion f auf D , wenn $f(x) \geq f(x_*)$ gilt für alle hinreichend nahe bei x_* gelegenen $x \in D$, d.h. genauer, wenn es $\delta > 0$ gibt, so dass x_* absolute Minimumstelle zu f auf der δ -Umgebung $D \cap U_\delta(x_*) = \{x \in D : |x - x_*| < \delta\}$ von x_* in D ist. Der Funktionswert $f(x_*)$ heißt dann ein **relatives Minimum** oder **lokales Minimum** zu f auf D . Kann dabei $\delta > 0$ so gewählt werden, dass x_* die einzige Minimumstelle zu f in der Umgebung ist, dass also gilt $f(x) > f(x_*)$ für alle von x_* verschiedenen Stellen $x \in D$ nahe genug bei x_* , so heißt x_* **lokale strikte Minimumstelle** oder **lokale isolierte Minimumstelle** von f auf D , und man sagt von der Funktion, dass sie an der Stelle x_* ein **striktes relatives Minimum** oder ein **striktes lokales Minimum** bzgl. D hat. Entsprechend erklärt man **lokale (strikte) Maximumstellen** x^* zu f und **lokale (strikte) Maxima** $f(x^*)$, wobei auch “relativ” statt “lokal” und “isoliert” statt “strikt” gesagt wird. Der Sammelbegriff ist **lokale Extremstellen** und **lokale Extremwerte**.

6) Die Aufgabe, zu einer gegebenen sog. **Zielfunktion** f mit Werten in \mathbb{R} auf irgendeiner Definitionsmenge D die Maximum- oder Minimumstellen zu bestimmen, bezeichnet man als ein allgemeines **Optimierungsproblem**. Es liegt auf der Hand, dass dies für ökonomische Anwendungen die wichtigste mathematische Problemstellung überhaupt ist! Für so allgemeine Optimierungsprobleme gibt es aber keine Lösungsmethoden, sondern man braucht zusätzliche Informationen über den Bereich D und die Funktion f . Dementsprechend unterscheidet man verschiedene Teilgebiete der Optimierungstheorie:

Bei der **diskreten Optimierung** ist D eine Menge von endlich vielen Daten, man braucht dann also nur unter den endlich vielen Werten von f den größten und den kleinsten herauszusuchen — doch bei evtl. vielen Millionen von Werten kann es ein echtes Problem sein, das mit akzeptablem Aufwand zu erledigen, und die Aufgabe ist unter Umständen nur unter zusätzlichen strukturellen Annahmen an die Daten und die Zielfunktion konkret zu lösen. Bei der **linearen Optimierung** ist der Bereich D ein Polyeder in einem Zahlenraum \mathbb{R}^n , das durch endlich viele lineare Gleichungen und Ungleichungen definiert wird, und die Zielfunktion f ist linear; hier setzt man massiv Methoden aus der Linearen Algebra zur Lösung ein, und wir haben in Abschnitt 3.6 eine Einführung in diese Theorie gegeben. Bei der **konvexen Optimierung** ist allgemeiner der Bereich D eine konvexe Menge in \mathbb{R}^n und die Zielfunktion f eine konvexe Funktion.

In diesem und dem nächsten Abschnitt geht es um die **differenzierbare Optimierung**, d.h. die Zielfunktion f ist differenzierbar und der Bereich $D \subset \mathbb{R}^n$, auf dem sie optimiert werden soll, wird durch endlich viele Ungleichungen oder Gleichungen mit differenzierbaren Funktionen definiert. Man kann dann, wie wir sehen werden, mit Hilfe der Differentialrechnung für Funktionen von mehreren Veränderlichen Lösungsverfahren entwickeln, die solche Probleme in vielen konkreten Fällen explizit lösen, die aber auch zu theoretischen Einsichten führen, welche sich als “ökonomische Gesetze” über optimierte ökonomische Situationen formulieren lassen.

7) Grundsätzlich kann eine Funktion $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ ihr Minimum und Maximum — wenn überhaupt — im Inneren oder am Rand des Bereichs D annehmen. Dementsprechend unterscheidet man **innere Extremstellen** von f auf D und **Randextremstellen**. Das kennen wir ja schon von Funktionen einer Veränderlichen auf einem Intervall in \mathbb{R} , wo man auch stets in Betracht ziehen muss, dass die Extremstellen Randpunkte sein können. So ist es auch bei mehreren Veränderlichen mit dem Unterschied, dass dort der Rand ∂D im Allgemeinen aus unendlich vielen Punkten besteht, während ein Intervall in \mathbb{R} höchstens zwei Randpunkte haben kann. Dementsprechend teilen sich die Methoden zur Extremstellenbestimmung bei Funktionen von mehreren Veränderlichen auf in die Methode zum Auffinden innerer Extremstellen, die wir in diesem Abschnitt behandeln, und (kompliziertere) Methoden zur Bestimmung von Randextremstellen im nächsten Abschnitt 5.5. Bei der Randextremstellenbestimmung hat man typischerweise nach Extrema von f auf einer Fläche kleinerer Dimension als die des umgebenden Raumes \mathbb{R}^n zu suchen, etwa auf dem Rand einer Kugel oder eines Quaders in \mathbb{R}^3 . Manchmal gibt es auch noch weitere, die Dimension reduzierende Bedingungen, so dass man z.B. auf einer Kurve in \mathbb{R}^3 oder einer zweidimensionalen Fläche in \mathbb{R}^4 zu optimieren hat. Derartige Probleme bezeichnet man als **Extremwertbestimmung bei Nebenbedingungen**, wobei die Nebenbedingungen Gleichungen sind, welche den niederdimensionalen Optimierungsbereich beschreiben.

8) Ziel aller im Folgenden behandelten Verfahren zur Extremstellenbestimmung ist die Aufstellung einer **vollständigen Kandidatenliste** für ein gegebenes differenzierbares Optimierungsproblem $f \rightsquigarrow \text{opt}$ auf $D \subset \mathbb{R}^n$. Darunter versteht man eine Liste von Punkten aus D , in der mit Sicherheit alle gesuchten Maximum- und / oder Minimumstellen von f vorkommen (das bedeutet "Vollständigkeit" der Liste). Hat man eine vollständige Kandidatenliste, so kann man die Extremstellen einfach durch **Wertevergleich** herausfinden, indem man die Punkte mit dem größten bzw. mit dem kleinsten Funktionswert in der Liste bestimmt. Voraussetzung dabei ist natürlich, dass die gesuchten Extremstellen überhaupt existieren; denn was nicht existiert, kann man auch nicht durch Suchen in einer noch so langen Liste finden! In ökonomischem Kontext ist die Existenz des gesuchten Optimums oft von vorneherein klar; mathematisch bewiesen wird sie, wie gesagt, meistens mit dem Extremstellensatz. Die Kandidatenliste wird oft auch Punkte enthalten, die keine absoluten Extremstellen sind, z.B. (nur) lokale Extremstellen oder Sattelpunkte. Das alles kennen wir im Prinzip schon von der Optimierung bei differenzierbaren Funktionen von einer Veränderlichen (siehe 4.5). Mit **Ableitungskriterien höherer Ordnung** (siehe 5.6) kann man unter Umständen einige dieser "überflüssigen" Kandidaten aus der Kandidatenliste streichen, ohne die Vollständigkeit der Liste zu verlieren. Solche Kriterien höherer Ordnung erlauben auch manchmal, einen Kandidaten wenigstens als lokale strikte Extremstelle oder als Sattelpunkt zu identifizieren, wenn man die absolute Extremalität bzgl. des gegebenen Definitionsbereichs nicht klären kann. ■

Wir beginnen die Diskussion der Methoden zur Extremstellenbestimmung mit dem folgenden sehr einfachen Kriterium für innere Extremstellen. Wir erinnern daran (siehe 5.1), dass ein Punkt $x_0 \in D$ **innerer Punkt** der Teilmenge $D \subset \mathbb{R}^n$ heißt, wenn eine ganze ε -Umgebung, also eine offene Kugel $U_\varepsilon(x) = \{x - x_0 \mid |x - x_0| < \varepsilon\}$ mit Mittelpunkt x_0 und (kleinem) positivem Radius ε , ganz in D enthalten ist.

SATZ (Notwendiges Kriterium erster Ordnung für innere Extremstellen):

Ist x_0 innerer Punkt von $D \subset \mathbb{R}^n$ und hat $f: D \rightarrow \mathbb{R}$ in x_0 eine lokale Extremstelle, so verschwinden dort simultan alle partiellen Ableitungen

$$\frac{\partial f}{\partial x_1}(x_0) = 0, \quad \frac{\partial f}{\partial x_2}(x_0) = 0, \quad \dots, \quad \frac{\partial f}{\partial x_n}(x_0) = 0,$$

sofern sie existieren. Mit anderen Worten: x_0 ist dann eine Nullstelle des Gradientenfelds,

$$\nabla f(x_0) = 0.$$

Das ergibt sich durch Anwendung des Kriteriums für Extremstellen bei Funktionen von einer Variablen auf die partiellen Funktionen $\varphi_j(x_j) = f(x_{01}, \dots, x_{0j-1}, x_j, x_{0j+1}, \dots, x_{0n})$ bei festgehaltenen Komponenten x_{0i} des Punktes $x_0 \in \mathbb{R}^n$. Da x_0 innerer Punkt von D ist, enthält der Definitionsbereich von φ_j die Stelle x_{0j} als inneren Punkt, und diese Stelle ist (lokale) Extremstelle der Funktion φ_j von einer Variablen, weil x_0 (lokale) Extremstelle von f auf D ist. Das Ableitungskriterium für innere Extremstellen bei Funktionen einer Veränderlichen in 4.5 sagt uns, dass dann $\varphi_j'(x_{0j}) = 0$ ist, wenn diese Ableitung existiert. Diese Ableitung der partiellen Funktion φ_j ist aber nichts anderes als die partielle Ableitung $\frac{\partial f}{\partial x_j}(x_0)$.

Weil die Nullstellen des Gradientenfeldes aufgrund dieses Satzes eine wichtige Rolle bei der Extremstellenbestimmung haben, bekommen sie einen besonderen (in 5.2 schon eingeführten) Namen:

DEFINITION: Die Nullstellen des Gradientenfeldes ∇f einer reellen differenzierbaren Funktion f von mehreren Variablen, also die simultanen Nullstellen aller partiellen Ableitungen von f , heißen die **kritischen Punkte** oder **stationären Stellen** der Funktion. ■

Letzteres, weil sich der Wert $f(x_0)$ im Fall $\nabla f(x_0) = 0$ bei kleinen Variationen von x_0 praktisch nicht ändert, also “stationär” bleibt; denn es ist ja $f(x) \approx f(x_0) + \nabla f(x_0) \cdot (x - x_0)$. Der Vollständigkeit halber erwähnen wir noch, obwohl das mit Extremstellenbestimmung nichts zu tun hat, dass man unter einem *kritischen Punkt einer Vektorfunktion* $F: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ eine Stelle $x_0 \in D$ versteht, an der die Ableitungsmatrix $\left(\frac{\partial F}{\partial x}(x_0)\right)$ nicht den maximal möglichen Rang hat, also einen Rang kleiner als $\min\{m, n\}$. (Nur im Fall $m = 1$ oder $n = 1$ bedeutet dies, dass alle Einträge der Ableitungsmatrix an dieser Stelle verschwinden.)

DISKUSSION: 1) Die Bedeutung des Satzes für die Extremstellenbestimmung ist klar:

- Die kritischen Punkte von f im Inneren von D ergeben eine vollständige Kandidatenliste für innere Extremstellen von f auf D , wenn f an allen inneren Stellen von D differenzierbar ist.

Alle inneren Maximum- und Minimumstellen kommen also in dieser Liste vor. Das sagt aber nicht, dass es solche Stellen gibt, auch wenn die Liste nicht leer ist! (Siehe 3) unten.) Sollte f in einzelnen inneren Punkten nicht differenzierbar sein, wie z.B. die Euklidische Normfunktion $f(x) = \sqrt{x_1^2 + \dots + x_n^2}$ im Nullpunkt von \mathbb{R}^n , so macht das nichts, vorausgesetzt

- alle inneren Nichtdifferenzierbarkeitsstellen werden in die Kandidatenliste für innere Extremstellen mit aufgenommen.

2) **Warnung:** Nicht jeder (innere) kritische Punkt ist unbedingt (lokale) Extremstelle!

Der Satz gibt eben nur ein *notwendiges* und kein hinreichendes Kriterium, und es gibt im Allgemeinen auch **Sattelpunkte**. Darunter versteht man kritische Punkte x_0 , die keine lokalen Extremstellen sind, so dass die Funktion in jeder noch so kleinen Umgebung $U_\varepsilon(x_0)$ sowohl größere Werte als auch kleinere Werte annimmt als an der Stelle x_0 . Der Name rührt daher, dass ein Bergwanderer in einem Sattel genau diese Situation für die Höhenfunktion über Nullniveau vorfindet: In allen Richtungen ist die Steigung zwar Null, aber es gibt beliebig nahe beim Sattelpunkt Punkte größerer und Punkte kleinerer Höhe.

Typisches Beispiel ist der Ursprung $(0, 0)$ bei der Funktion $f(x, y) = x^2 - y^2$ von zwei Variablen. Es ist $\nabla f(x, y) = (2x, 2y)$, also $\nabla f(0, 0) = (0, 0)$, aber für $y = 0$ und $0 < |x| \ll 1$ ist $f(x, y) = x^2 > 0 = f(0, 0)$, und für $x = 0$, $0 < |y| \ll 1$ ist $f(x, y) = -y^2 < 0 = f(0, 0)$. Der Graph dieser Funktion ist nahe dem Nullpunkt ein typischer “Sattel”; er enthält die nach oben geöffnete Parabel $z = x^2$ in der (x, z) -Ebene und die nach unten geöffnete Parabel $z = -y^2$ in der dazu senkrechten (y, z) -Ebene.

3) **Warnung:** Unter den inneren kritischen Punkten von f in D braucht überhaupt keine (lokale) Extremstelle von f vorzukommen,

selbst wenn die Liste viele Kandidaten enthält. Es kann eben sein, dass die Funktion Extremwerte nur am Rand von D oder überhaupt nicht annimmt!

4) Aber umgekehrt kann man schließen: Wenn es keine Kandidaten für innere Extremstellen gibt, so existieren auch keine inneren Extremstellen. Mit anderen Worten:

- Ist $\nabla f \neq 0$ an allen inneren Stellen von D , so hat f keine (lokalen) Extremstellen im Inneren von D .

Das lässt sich insbesondere auf lineare Funktionen $f(x) = \mathbf{a} \cdot x = a_1x_1 + \dots + a_nx_n + b$ anwenden. Hier ist $\nabla f \equiv \mathbf{a}$ konstant und überall $\neq 0$, wenn f nicht konstant ist.

- Lineare Funktionen nehmen also (lokale) Extrema auf $D \subset \mathbb{R}^n$ höchstens in Randpunkten von D an.

5) Die **Berechnung der kritischen Punkte** läuft auf die Lösung des folgenden Systems von n (im Allgemeinen nichtlinearen) Gleichungen für n Unbekannte hinaus:

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1, x_2, \dots, x_n) &= 0 \\ \frac{\partial f}{\partial x_2}(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ \frac{\partial f}{\partial x_n}(x_1, x_2, \dots, x_n) &= 0. \end{aligned}$$

Die im Inneren von D liegenden Lösungen $x = (x_1, \dots, x_n)$ sind die gesuchten Kandidaten für innere Extremstellen von f . Da es sich hier um ebenso viele Gleichungen wie Unbekannte handelt, wird es "im Normalfall" nur einzelne "isolierte" Lösungen geben. Nur in Ausnahmefällen wird man ein Kontinuum von kritischen Punkten finden, wie z.B. bei der Höhenfunktion eines Gebirges, das einen Bergkamm oder eine Talsohle von exakt konstanter Höhe über einen ausgedehnten Bereich aufweist.

6) Man könnte denken, dass man zusätzliche Gleichungen für die inneren Extremstellen bekommt, wenn man nicht nur das Verschwinden der partiellen Ableitungen dort, sondern auch das Verschwinden aller Richtungsableitungen ins Spiel bringt, was natürlich genauso gezeigt werden kann wie im Beweis des obigen Satzes. Aber Richtungsableitungen $\partial_u f(x)$ lassen sich ja normalerweise (bei Funktionen mit stetigen partiellen Ableitungen, siehe 5.2) durch partielle Ableitungen ausdrücken, $\partial_u f(x) = u \cdot \nabla f(x) = u_1 \frac{\partial f}{\partial x_1}(x) + \dots + u_n \frac{\partial f}{\partial x_n}(x)$, so dass das Verschwinden der Richtungsableitungen keine zusätzliche Information über die Extremstellen ist. ■

Bevor wir einige Beispiele betrachten, bei denen es innere Extremstellen gibt, müssen wir noch den Satz zur Verfügung stellen, der uns überhaupt die Existenz von Extremstellen sichert. Dabei ist der Definitionsbereich D eine sog. **kompakte Menge** in \mathbb{R}^n , d.h. **abgeschlossen und beschränkt**. Abgeschlossenheit heißt (siehe 5.1), dass D sämtliche Randpunkte enthält, also $\partial D \subset D$, und Beschränktheit bedeutet natürlich, dass D in eine große Kugel passt, dass also in D nicht Punkte x mit beliebig großem Abstand $|x|$ zum Ursprung vorkommen. Äquivalent mit Beschränktheit ist, dass es eine endliche Schranke S gibt mit $|x_j| \leq S$ für alle Punkte $x = (x_1, \dots, x_n) \in D$ und alle $j = 1 \dots n$.

SATZ (vom Maximum und Minimum – Extremstellensatz): Ist D kompakt in \mathbb{R}^n , also abgeschlossen und beschränkt, und $f: D \rightarrow \mathbb{R}$ stetig, so hat f (mindestens) eine absolute Maximumstelle und (mindestens) eine absolute Minimumstelle auf D .

Bei Anwendungen ist der Bereich D oft nicht kompakt, man kann aber mit den Argumenten, die zum Extremstellensatz führen, dennoch die Existenz von Extremstellen sichern. Deshalb gehen wir nun auf den *Beweis* ein: Man betrachtet, um z.B. eine Minimumstelle zu finden, zunächst eine sog. **Minimalfolge** von Punkten $x_k = (x_{k1}, \dots, x_{kn}) \in D$, d.h. die Werte $f(x_k)$ konvergieren bei $k \rightarrow \infty$ gegen das Infimum $\inf_D f \geq -\infty$. (Eine solche Folge gibt es nach Definition von $\inf_D f$ als größte untere Schranke für die Werte von f auf D immer.) Wegen der Beschränktheit von D sind dann auch alle Komponentenfolgen x_{1j}, x_{2j}, \dots beschränkt in \mathbb{R} . Nach dem fundamentalen Satz von Bolzano–Weierstraß über beschränkte Zahlenfolgen, können wir durch Streichen von Gliedern zu einer Teilfolge übergehen, derart dass die erste Komponentenfolge konvergiert, dann zu einer weiteren Teilfolge, derart dass auch die zweite Komponentenfolge konvergiert, und so fortfahrend schließlich zu einer Teilfolge, für die alle n Komponentenfolgen gegen reelle Zahlen konvergieren, die erste gegen x_{*1} , die zweite gegen x_{*2}, \dots , die n -te gegen x_{*n} . Dies bedeutet, dass die Teilfolge in \mathbb{R}^n gegen den Punkt $x_* = (x_{*1}, \dots, x_{*n})$ konvergiert. Der Einfachheit halber bezeichnen wir diese konvergente Teilfolge, die auch eine Minimalfolge für f auf D ist, wieder mit x_1, x_2, \dots . Weil D abgeschlossen ist, liegt ihr Limes x_* wieder in D . (x_* kann nämlich kein äußerer Punkt von D sein, weil ja beliebig nahe bei x_* die Punkte $x_k \in D$ mit genügend großem Index k liegen, also ist x_* innerer Punkt oder Randpunkt von D und gehört wegen $\partial D \subset D$ jedenfalls zu D ; siehe 5.1.) Nun kommt die Stetigkeit von f an der Stelle x_* ins Spiel, die besagt, dass sich $f(x_*)$ und $f(x_k)$ beliebig wenig unterscheiden, wenn k groß genug ist, also x_k nahe genug bei x_* . Andererseits aber bilden die x_k eine Minimalfolge für f , d.h. die Werte $f(x_k)$ liegen auch beliebig nahe bei $\inf_D f$ für große k . Dann muss aber $f(x_*) = \inf_D f$ sein, und das bedeutet, dass x_* eine absolute Minimumstelle von f auf D ist.

Wir sehen, dass diese Argumentation auch zu einer Minimumstelle x_* von f auf D führt, wenn D nicht abgeschlossen bzw. nicht beschränkt ist, sofern wir für jede Minimalfolge x_1, x_2, \dots sicherstellen können, dass sie nicht gegen einen nicht zu D gehörenden Randpunkt konvergiert und auch nicht gegen Unendlich im Sinne von $|x_k| \rightarrow \infty$ bei $k \rightarrow \infty$. Daher gilt der

ZUSATZ: *Ist D nicht kompakt, f stetig auf D und kann keine Minimalfolge [bzw. Maximalfolge] gegen Unendlich oder gegen einen nicht zu D gehörenden Randpunkt konvergieren, so hat f (mindestens) eine absolute Minimumstelle [bzw. Maximumstelle] auf D .*

DISKUSSION: 1) Auf keine der Voraussetzungen des Extremstellensatzes kann verzichtet werden. Ist z.B. D nicht abgeschlossen, gibt es also einen Randpunkt a von D mit $a \notin D$, so nimmt die dehnungsbeschränkte Funktion $f(x) = |x - a|$ kein Minimum auf D an, weil sie auf D überall positiv ist und beliebig kleine Werte an Stellen aus D nahe a annimmt, aber nicht den Wert 0. Ist D unbeschränkt, so hat die dehnungsbeschränkte Funktion $f(x) = |x|$ beliebig große Werte auf D und kann deshalb kein Maximum haben. Und $g(x) = \frac{1}{1+|x|}$ hat überall positive Werte und auf der unbeschränkten Menge D auch beliebig kleine Werte, nimmt aber den Wert 0 nicht an und hat daher kein Minimum auf D . Ist schließlich f unstetig, so kann es passieren, dass der Funktionswert $f(x_*)$ im Limes x_* einer Minimalfolge größer ist als $\inf_D f$, und dann ist x_* keine Minimumstelle und es gibt evtl. gar keine. Ein Beispiel ist $f(x) = |x|$ für $x \neq 0$ und $f(0) = 1$. Dann müssen alle Minimalfolgen auf einer Kugel D um den Ursprung gegen 0 konvergieren, aber f ist überall positiv und hat deshalb kein Minimum auf D .

2) *Stetigkeit bei Funktionen von mehreren Veränderlichen* haben wir in 5.2 kurz diskutiert. Alle elementaren Funktionen sind auf ihren offenen Definitionsbereichen stetig. Daher sind auch ökonomische Funktionen, sofern sie nicht aus verschiedenen elementaren Stücken unstetig zusammengesetzt sind, stetig. Deshalb ist die Stetigkeitsvoraussetzung an die zu optimierende Funktion bei ökonomischen Aufgabenstellungen unproblematisch.

3) Das gilt nicht für die Kompaktheitsvoraussetzung im Extremstellensatz, die bei konkreten Problemen oft verletzt ist. Typische Optimierungsbereiche bei ökonomischen Problemen sind z.B. der abgeschlossene, aber unbeschränkte Bereich $\mathbb{R}_{\geq 0}^n$ der (komponentenweise) nichtnegativen Vektoren in \mathbb{R}^n und der Bereich $\mathbb{R}_{> 0}^n$ der (komponentenweise) positiven Vektoren, der weder beschränkt noch abgeschlossen ist. ($\mathbb{R}_{> 0}^n$ ist offen, enthält also keinen seiner Randpunkte.) Hier kann man die Existenz von Extremstellen stetiger Funktionen f unter Umständen mit dem Zusatz zum Extremstellensatz zeigen.

Ist z.B. eine Minimumstelle gesucht und gilt $f(x) \rightarrow \infty$ bei $|x| \rightarrow \infty$ (d.h. $f(x)$ ist größer als eine beliebig groß gegebene Zahl, wenn nur die Norm von x groß genug ist), so kann offensichtlich keine Minimalfolge gegen Unendlich konvergieren, also garantiert der Zusatz die Existenz einer Minimumstelle in $\mathbb{R}_{\geq 0}^n$. (Nicht zu D gehörende Randpunkte gibt es hier ja nicht.) Ähnlich kann man schließen, wenn z.B. $f(x) \rightarrow 0$ strebt bei $|x| \rightarrow \infty$ und wenn f irgendwo auf $\mathbb{R}_{\geq 0}^n$ einen positiven Wert annimmt. Dann kann keine Maximalfolge gegen Unendlich konvergieren (die Funktionswerte müssen darauf ja gegen einen Wert $\geq f(x) > 0$ streben), also hat f eine Maximumstelle auf $\mathbb{R}_{\geq 0}^n$. Derartige Argumentationen sind gemeint mit folgendem Prinzip:

- *Grenzbedingungen an $f(x)$ bei $|x| \rightarrow \infty$ ($x \in D$) können unter Umständen fehlende Beschränktheit des Bereichs D im Extremstellensatz kompensieren.*

Ähnlich ist es bei fehlender Abgeschlossenheit. Gilt z.B. immer $f(x) \rightarrow \infty$ bei Annäherung $D \ni x \rightarrow a$ an einen nicht zu D gehörenden Randpunkt $a \in \partial D$, so kann keine Minimalfolge gegen einen solchen Punkt a konvergieren. Und gilt $f(x) \rightarrow 0$ bei Konvergenz $D \ni x \rightarrow a$ gegen nicht zu D gehörende Randpunkte, so kann keine Maximalfolge dagegen konvergieren, sofern f irgendwo positiv ist auf D . Man sieht:

- *Grenzbedingungen an $f(x)$ bei Konvergenz $D \ni x \rightarrow a$ gegen nicht zu D gehörende Randpunkte können unter Umständen fehlende Abgeschlossenheit des Bereichs D im Extremstellensatz kompensieren.*

Für einen Bereich wie $\mathbb{R}_{> 0}^n$, der weder beschränkt, noch abgeschlossen ist, muss man beide Arten von Grenzbedingungen für die auf $\mathbb{R}_{> 0}^n$ stetige Funktion f haben. Wenn z.B. $f(x) \rightarrow \infty$ gilt, wenn immer eine Komponente von x gegen ∞ oder gegen 0 strebt, so hat f eine Minimumstelle in $\mathbb{R}_{> 0}^n$. Im Fall $|x| \rightarrow \infty$ strebt nämlich mindestens eine Komponente von x gegen ∞ und Randpunkte a von $\mathbb{R}_{> 0}^n$ haben mindestens eine Nullkomponente, so dass aus $x \rightarrow a$ folgt, dass mindestens eine Komponente von x gegen 0 strebt. Da in beiden Fällen nach Voraussetzung $f(x) \rightarrow \infty$ gilt, kann keine Minimalfolge gegen Unendlich oder gegen einen Randpunkt streben, und der Zusatz greift.

4) Der Extremstellensatz für stetige Funktionen f auf kompakten Mengen D garantiert, dass es *endliche* Werte $\min_D f$ und $\max_D f$ gibt, zwischen denen alle Funktionswerte von f auf D liegen. Daher folgt insbesondere:

- *Stetige reelle Funktionen sind auf kompakten Mengen beschränkt,*

d.h. alle Funktionewerte liegen in einem kompakten Intervall $[\alpha, \beta]$, $-\infty < \alpha \leq \beta < \infty$. Auf unbeschränkten Mengen D gilt das nicht allgemein wie $f(x) = |x|$ demonstriert, auf nicht abgeschlossenen Mengen D auch nicht, wie $g(x) = \frac{1}{|x-a|}$ mit $a \in (\partial D) \setminus D$ zeigt. ■

BEISPIELE (innere Extremstellen):

(1) Lineare Funktionen $\ell(x) = \mathbf{a} \cdot x + b = a_1x_1 + \dots + a_nx_n + b$ haben, wie schon gesagt, niemals innere Extremstellen, wenn sie nicht konstant sind; denn $\nabla\ell \equiv \mathbf{a}$ ist dann konstant $\neq 0$. Der Extremstellensatz garantiert andererseits für kompakte Bereiche D , dass die lineare (und insbesondere stetige) Funktion ℓ ein Maximum und ein Minimum auf D annimmt. Diese Extremstellen müssen also im Rand von D liegen.

(2) Da lineare Funktionen kein Beispiel für innere Extremstellen sein können, betrachten wir nun eine einfache quadratische Funktion von zwei Veränderlichen:

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

Das Verhalten im Unendlichen wird durch die rein quadratischen Terme bestimmt, und ein Blick auf die Koeffizienten legt nahe, dass hier $f(x, y) \rightarrow \infty$ gilt bei $|(x, y)| \rightarrow \infty$. Mit quadratischer Ergänzung $f(x, y) = (x + \frac{1}{2})^2 + 2y^2 + \frac{3}{4}$ kann man sich davon überzeugen, dass tatsächlich $f(x, y) \rightarrow \infty$ geht bei $|x| \rightarrow \infty$ und auch bei $|y| \rightarrow \infty$, d.h. wenn (x, y) eine Folge (x_k, y_k) durchläuft, bei der die ersten Komponenten $|x_k|$ beliebig groß werden und die y_k ganz beliebig sind oder umgekehrt. (Es genügt hier nicht zu prüfen, dass $f(x, y) \rightarrow \infty$ strebt bei $|x| \rightarrow \infty$ und festgehaltenem Wert von y bzw. bei $|x| \rightarrow \infty$ und festgehaltenem Wert von x .) Der Zusatz zum Extremstellensatz greift also und garantiert die Existenz mindestens einer Minimumstelle von f auf \mathbb{R}^2 , während es natürlich keine Maximumstelle geben kann, weil ja f von oben unbeschränkt ist. Um die Minimumstellen zu finden, berechnen wir die kritischen Punkte. Weil f quadratisch ist, führt dies hier auf ein lineares Gleichungssystem

$$\frac{\partial f}{\partial x}(x, y) = 2x + 1 = 0, \quad \frac{\partial f}{\partial y}(x, y) = 4y = 0,$$

das genau eine Lösung $x = -\frac{1}{2}$, $y = 0$ hat. Also ist $(-\frac{1}{2}, 0)$ einzige absolute Minimumstelle von f auf \mathbb{R}^2 , das Minimum ist $f(-\frac{1}{2}, 0) = \frac{3}{4}$ und f hat weiter keine lokalen Extremstellen oder Sattelpunkte auf \mathbb{R}^2 . Die eindeutige Minimumstelle kann man bei dieser einfachen Funktion natürlich auch direkt sehen: Zunächst ist klar, dass ein Minimum nur für $y = 0$ erreicht werden kann, dann hat man nur noch die Funktion $x^2 + x + 1 = (x + \frac{1}{2})^2 + \frac{3}{4}$ von einer Variablen zu minimieren.

(3) Eine solche direkte Argumentation ist bei der folgenden quadratischen Funktion von drei Veränderlichen wegen des Auftretens der gemischten quadratischen Terme xy und yz , die kein eindeutiges Vorzeichen haben, nicht mehr ohne Weiteres möglich:

$$f(x, y, z) = -x^2 - 2y^2 - 3z^2 - xy - yz + 40x + 50y + 80z - 100.$$

Die negativen Vorzeichen der Koeffizienten vor den Quadraten der Variablen suggerieren $f(x, y, z) \rightarrow -\infty$ bei $|(x, y, z)| \rightarrow \infty$, aber wegen der subtrahierten gemischten quadratischen Terme ist das nicht so klar. Hier zeigt eine geeignete quadratische Ergänzung

$$f(x, y, z) = -(x + \frac{1}{2}y)^2 - (\frac{1}{2}y + z)^2 - \frac{3}{2}y^2 - 2z^2 + 40x + 50y + 80z - 100,$$

wegen der Dominanz der negativen quadratischen Terme über die linearen Terme, dass tatsächlich $f(x, y, z) \rightarrow -\infty$ gilt bei $|(x, y, z)| \rightarrow \infty$ (d.h. wenn mindestens eine der Variablen x, y, z eine Folge von Zahlen mit über alle Schranken wachsendem Betrag durchläuft und die anderen Variablen ganz beliebige Zahlenfolgen).

Nach dem Zusatz zum Extremstellensatz gibt es also mindestens eine absolute Maximumstelle auf \mathbb{R}^3 und diese muss eine Lösung des Gleichungssystems für die kritischen Punkte sein, das hier wiederum linear ist:

$$\frac{\partial f}{\partial x} = -2x - y + 40 = 0, \quad \frac{\partial f}{\partial y} = -4y - x - z + 50 = 0, \quad \frac{\partial f}{\partial z} = -6z - y + 80 = 0.$$

Die eindeutige Lösung ist $x = \frac{35}{2}$, $y = 5$, $z = \frac{25}{2}$, also ist $\frac{1}{2}(35, 10, 25)$ die eindeutige Minimumstelle von f auf \mathbb{R}^3 , $f(\frac{35}{2}, 5, \frac{25}{2}) = 875$ das absolute Minimum, und weiter gibt es keine lokalen Extremstellen oder Sattelpunkte von f auf \mathbb{R}^3 .

(4) Auch für allgemeine *quadratische Funktionen* von n Variablen

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

mit symmetrischer $n \times n$ -Matrix $A = (a_{ij})$, $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^n$ und $c \in \mathbb{R}$ ist eine vollständige Diskussion der Extremstellen auf \mathbb{R}^n möglich. In 5.2 haben wir $\nabla q(x) = Ax + \mathbf{b}$ berechnet. Das ist eine lineare Vektorfunktion von $x \in \mathbb{R}^n$, daher führt die Bestimmung der Nullstellen von ∇q auf ein lineares Gleichungssystem

$$Ax = -\mathbf{b}.$$

Gibt es keine Lösung, so hat q keine kritischen Punkte in \mathbb{R}^n , also existieren keine inneren Extremstellen auf irgendeinem Bereich $D \subset \mathbb{R}^n$. Ist aber x_0 eine Lösung, $Ax_0 + \mathbf{b} = 0$, so können wir schreiben

$$q(x) = \frac{1}{2}x \cdot Ax - x \cdot Ax_0 + c = \frac{1}{2}(x-x_0) \cdot A(x-x_0) + c - \frac{1}{2}x_0 \cdot Ax_0 = \frac{1}{2}(x-x_0) \cdot A(x-x_0) + q(x_0).$$

Daraus lesen wir ab, dass x_0 absolute Minimumstelle von q auf \mathbb{R}^n ist, genau wenn $u \cdot Au \geq 0$ gilt für alle $u \in \mathbb{R}^n$ (also wenn A positiv semidefinit ist), und sogar einzige absolute Minimumstelle, wenn $u \cdot Au > 0$ für alle $u \neq 0$ (also wenn A positiv definit ist). Im positiv definiten Fall ist A auch invertierbar, also die Gleichung $Ax + \mathbf{b} = 0$ eindeutig lösbar. Ist aber $v \cdot Av < 0$ für ein $v \in \mathbb{R}^n$, so gilt $q(x_0 + tv) = \frac{1}{2}t^2 v \cdot Av + q(x_0) < q(x_0)$ für alle $t \in \mathbb{R}_{\neq 0}$, also ist x_0 keine lokale Minimumstelle und q von unten unbeschränkt auf \mathbb{R}^n . Hat schließlich $Ax + \mathbf{b} = 0$ keine Lösung, also $\pm \mathbf{b} \notin A\mathbb{R}^n$, so gibt es Vektoren $w \in (A\mathbb{R}^n)^\perp$ mit $w \cdot \mathbf{b} \neq 0$ (sonst wäre $\mathbf{b} \in (A\mathbb{R}^n)^{\perp\perp} = A\mathbb{R}^n$), und dann folgt $q(tw) = \frac{1}{2}t^2 w \cdot Aw + \mathbf{b} \cdot w + c = tw \cdot \mathbf{b} + c$, so dass q von beiden Seiten unbeschränkt ist auf \mathbb{R}^n . Wir fassen zusammen:

- Ist A positiv [negativ] definit, so ist $x_0 = -A^{-1}\mathbf{b}$ einziger kritischer Punkt und absolute Minimumstelle [Maximumstelle] von q auf \mathbb{R}^n ; q ist dann von oben [von unten] unbeschränkt auf \mathbb{R}^n .
- Ist $A \neq 0$ positiv [negativ] semidefinit und nicht definit, so sind die kritischen Punkte von q genau die Lösungen zu $Ax + \mathbf{b} = 0$; gibt es keine, so hat q keine lokalen Extrema und ist beidseitig unbeschränkt; andernfalls ist die Lösungsmenge ein affiner Unterraum der Dimension ≥ 1 und gleich der Menge der absoluten Minimumstellen [Maximumstellen] von q auf \mathbb{R}^n , und q ist von oben [unten] unbeschränkt.
- Ist A indefinit, so hat q keine absoluten Extremstellen auf \mathbb{R}^n und ist von unten und oben unbeschränkt; die kritischen Punkte sind die Lösungen von $Ax + \mathbf{b} = 0$, und wenn es Lösungen gibt, so bilden sie einen affinen Unterraum von Sattelpunkten zu q mit demselben Funktionswert.

Die letzte Aussage gilt, weil für zwei verschiedene Lösungen x_0, \tilde{x}_0 ja $A(\tilde{x}_0 - x_0) = 0$ ist, also $q(\tilde{x}_0) = \frac{1}{2}(\tilde{x}_0 - x_0) \cdot A(\tilde{x}_0 - x_0) + q(x_0) = q(x_0)$. Indefinitheit von A bedeutet, dass es Vektoren $v, w \in \mathbb{R}^n$ gibt mit $v \cdot Av > 0$ und $w \cdot Aw < 0$.

Für die Definitheitsbegriffe und algebraischen *Definitheitskriterien* verweisen wir auf 3.5. Eine Charakterisierung positiver Definitheit [positiver Semidefinitheit] ist, dass alle Eigenwerte λ_i der symmetrischen Matrix A positiv [nichtnegativ] sind; Indefinitheit bedeutet also, dass es einen positiven und einen negativen Eigenwert gibt. Zu jeder symmetrischen Matrix A existiert eine Orthonormalbasis des \mathbb{R}^n aus Eigenvektoren u_1, \dots, u_n . Bezüglich dieser Basis ist A diagonalisiert und $x \cdot Ax = \lambda_1(u_1 \cdot x)^2 + \dots + \lambda_n(u_n \cdot x)^2$ (Hauptachsentransformation; siehe 3.5). Aus dieser Beschreibung des homogenen quadratischen Anteils von $q(x)$ kann man die obigen Aussagen über die absoluten Extremstellen von allgemeinen quadratischen Funktionen auch ohne Differentialrechnung herleiten.

(5) Für Polynomfunktionen $f(x)$ vom Grad $m \geq 3$ auf \mathbb{R}^n ist eine so vollständige Diskussion der Extremstellen auf \mathbb{R}^n wie bei quadratischen Funktionen nicht mehr möglich. Die Bedingung $\nabla f(x) = 0$ für die kritischen Punkte ist nun ein System von n polynomialen Gleichungen des maximalen Grades $m-1 \geq 2$ für x_1, x_2, \dots, x_n , und solche nichtlinearen Gleichungssysteme kann man nur in Ausnahmefällen explizit lösen. Bei einer Polynomfunktion vom Grad 3 wie etwa

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

auf \mathbb{R}^2 ergibt sich für die kritischen Stellen ein System von zwei Gleichungen des Grads ≤ 2 (wobei mindestens eine Gleichung quadratisch ist), hier konkret

$$\frac{\partial f}{\partial x} = 3x^2 - 4xy + y^2 + 2x = 0, \quad \frac{\partial f}{\partial y} = -2x^2 + 2xy - 3y^2 + 2y = 2.$$

Das lässt sich nicht mehr ohne Weiteres auflösen. (Man kann y durch Lösen einer quadratischen Gleichung eliminieren, erhält dann aber eine algebraische Gleichung vom Grad 4 für x .) Geht es allerdings nur um die Bestimmung absoluter Extremstellen auf \mathbb{R}^n , so können wir uns bei einer Polynomfunktion von ungeradem Grad wie hier alle Rechnungen sparen; denn solche Funktionen nehmen ja positive und negative Werte beliebig großen Betrags auf \mathbb{R}^n an, sie haben also bestimmt keine absolute Extremstellen auf \mathbb{R}^n . (Die kritischen Punkte können allenfalls lokale Extremstellen sein oder Sattelpunkte.)

(6) Anders sieht es unter Umständen aus, wenn der Grad der Polynomfunktion eine gerade Zahl ist wie etwa bei

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

auf \mathbb{R}^2 . Hier kann man sehen, dass $f(x, y)$ beliebig groß wird, wenn $|(x, y)|$ hinreichend groß ist. Denn für $|x| \geq 1$ gilt z.B. $f(x, y) \geq (\frac{3}{4}x + \frac{2}{3}y)^2 + \frac{7}{16}x^2 + \frac{1}{18}y^2 \geq \frac{1}{18}(x^2 + y^2)$ und für $|x| \leq 1$ und $x^2 + y^2 \geq 10$ ist $f(x, y) \geq \frac{1}{2}y^2 - |y| = \frac{1}{6}y^2 + |y|(\frac{1}{3}|y| - 1) \geq \frac{1}{6}y^2 \geq \frac{1}{8}(x^2 + y^2)$. Nach dem Zusatz zum Extremstellensatz nimmt daher f ein absolutes Minimum auf \mathbb{R}^2 an, und wir können die absoluten Minimumstellen unter den kritischen Punkten finden. Das Gleichungssystem für die kritischen Punkte lautet hier

$$\frac{\partial f}{\partial x} = 4x^3 + y = 0, \quad \frac{\partial f}{\partial y} = y + x = 0$$

und kann explizit gelöst werden: Einsetzen von $y = -x$ aus der zweiten Gleichung in die erste gibt $4x^3 - x = 0$, also $x = 0$ oder $= \frac{1}{2}$ oder $= -\frac{1}{2}$ und entsprechend $y = 0$ bzw. $= -\frac{1}{2}$ bzw. $= \frac{1}{2}$. Es gibt also genau drei kritische Stellen $(0, 0)$, $(\frac{1}{2}, -\frac{1}{2})$, $(-\frac{1}{2}, \frac{1}{2})$. Die Funktionswerte an diesen Stellen sind der Reihe nach 0 , $-\frac{1}{16}$, $-\frac{1}{16}$. Da wir uns bereits davon überzeugt haben, dass absolute Minimumstellen auf \mathbb{R}^2 existieren, können wir nun schließen, dass genau die beiden kritischen Punkte $\pm(\frac{1}{2}, -\frac{1}{2})$ mit dem kleinsten Funktionswert solche absoluten Minimumstellen sind.

Der kritische Punkt $(0, 0)$ ist dagegen ein Sattelpunkt; denn für $0 < x \ll 1$, $y = 0$ ist $f(x, y) = x^4 > 0$, aber für $0 < x = -y \ll 1$ ist $f(x, y) = y^4 - y^2 = y^2(y^2 - 1) < 0$. Außer den beiden absoluten Minimumstellen hat f also keine weiteren lokalen Extremstellen (das müssten ja kritische Punkte sein). Dass die Diskussion hier ein vollständiges Ergebnis liefert, liegt natürlich daran, dass die zweite Gleichung $x + y = 0$ für die kritischen Punkte so einfach ist und die erste $4x^3 + y = 0$ auch nicht übermäßig kompliziert. Außerdem war wesentlich, dass wir das Grenzverhalten $f(x, y) \rightarrow \infty$ bei $|(x, y)| \rightarrow \infty$ feststellen konnten — sonst hätten wir ohne weitere Analyse nicht sicher sein können, dass die kritischen Punkte $\pm(\frac{1}{2}, -\frac{1}{2})$ tatsächlich absolute Minimumstellen sind und nicht etwa nur lokale Minimumstellen oder lokale Maximumstellen oder ebenfalls Sattelpunkte.

- *Bei unbeschränktem Optimierungsbereich ist die Diskussion der Funktion im Unendlichen unerlässlich!*

(7) Als Beispiel für die Extremstellenbestimmung einer Funktion auf einem beschränkten Optimierungsbereich allein durch Berechnung der inneren kritischen Stellen betrachten wir

$$f(x, y, z) = (1 - x^2 - y^2 - z^2)xy^2z^3 \quad \text{auf der Einheitskugel } \mathbb{B}^3 \subset \mathbb{R}^3.$$

Auf der kompakten Einheitskugel $\mathbb{B}^3 = \{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$ als Optimierungsbereich nimmt f gemäß dem Extremstellensatz ein absolutes Minimum und ein absolutes Maximum an. Die Funktion hat offenbar sowohl positive als auch negative Werte nahe dem Ursprung, also ist das Minimum auf \mathbb{B}^3 negativ und das Maximum positiv. Für alle Randpunkte (x, y, z) der Kugel gilt aber $x^2 + y^2 + z^2 = 1$, also $f(x, y, z) = 0$, daher nimmt die Funktion ihre Extrema nicht in einem Randpunkt an und wir können die Extremstellen *im Inneren*, also unter den kritischen Punkten finden. Berechnung und Nullsetzen der partiellen Ableitungen führt auf folgendes Gleichungssystem für die kritischen Punkte:

$$\begin{aligned} (1 - x^2 - y^2 - z^2)y^2z^3 - 2x^2y^2z^3 &= 0 \\ (1 - x^2 - y^2 - z^2)2xyz^3 - 2xy^3z^3 &= 0 \\ (1 - x^2 - y^2 - z^2)3xy^2z^2 - 2xy^2z^4 &= 0. \end{aligned}$$

Wir sehen, dass alle Punkte mit $y = 0$ oder $z = 0$ Lösungen sind und dass für Lösungen mit $yz \neq 0$ auch $x \neq 0$ sein muss, wenn sie im Inneren der Einheitskugel liegen also $x^2 + y^2 + z^2 < 1$ erfüllen. Für Lösungen ohne Nullkomponente sind die Gleichungen dann äquivalent zu folgendem System linearer Gleichungen für x^2 , y^2 und z^2 :

$$\begin{aligned} 3x^2 + y^2 + z^2 &= 1 \\ x^2 + 2y^2 + z^2 &= 1 \\ 3x^2 + 3y^2 + 5z^2 &= 3. \end{aligned}$$

Die eindeutige Lösung ist $x^2 = \frac{1}{8}$, $y^2 = \frac{1}{4}$, $z^2 = \frac{3}{8}$, und das gibt weitere 8 kritische Punkte $(\pm\frac{\sqrt{2}}{4}, \pm\frac{1}{2}, \pm\frac{\sqrt{6}}{4})$, alle mit Abstand $\sqrt{x^2 + y^2 + z^2} = \frac{1}{4}\sqrt{3} < 1$ zum Ursprung, also im Inneren von \mathbb{B}^3 gelegen. Da der Funktionswert in den kritischen Punkten mit $yz = 0$ Null ist, kommen nur diese 8 Punkte als Extremstellen in Frage, und der Wertevergleich zeigt, dass davon die vier Punkte mit gleichen Vorzeichen der ersten und dritten Komponente die absoluten Maximumstellen von f auf \mathbb{B}^3 sind und die vier anderen Punkte die absoluten Minimumstellen. Für die weiteren kritischen Punkte zeigt direkte Inspektion der Funktion, dass die mit $y = 0 < xz$ lokale Minimumstellen sind, die mit $y = 0 > xz$ lokale Maximumstellen und die mit $z = 0$ oder $x = y = 0$ Sattelpunkte.

(8) Bei dem Optimierungsproblem für

$$f(x, y) = xy + \frac{1}{x} + \frac{1}{y} \quad \text{auf } \mathbb{R}_{>0}^2 = \mathbb{R}_{>0} \times \mathbb{R}_{>0}$$

haben wir einen Optimierungsbereich, den positiven Quadranten, der weder abgeschlossen noch beschränkt ist. Aber man kann sehen, dass $f(x, y) \rightarrow \infty$ geht, wenn $(x, y) \in \mathbb{R}_{>0}^2$ sich einem Randpunkt des Quadranten nähert oder gegen Unendlich strebt. (Bei Annäherung an Randpunkte geht $\frac{1}{x} \rightarrow \infty$ oder $\frac{1}{y} \rightarrow \infty$; wenn aber $\frac{1}{x} \leq c$, $\frac{1}{y} \leq c$ beschränkt bleiben, so gilt $xy \geq \frac{x}{2c} + \frac{y}{2c} \rightarrow \infty$, bei $|(x, y)| \rightarrow \infty$, $(x, y) \in \mathbb{R}_{>0}^2$.) Der Zusatz zum Extremalsatz garantiert also die Existenz einer absoluten Minimumstelle von f auf $\mathbb{R}_{>0}^2$. Die Gleichungen für die kritischen Stellen lauten

$$\frac{\partial f}{\partial x} = y - \frac{1}{x^2} = 0, \quad \frac{\partial f}{\partial y} = x - \frac{1}{y^2} = 0,$$

und haben die eindeutige Lösung $x = 1$, $y = 1$ in $\mathbb{R}_{>0}^2$. Also ist $(1, 1)$ Minimumstelle und $f(1, 1) = 3$ das Minimum von f auf $\mathbb{R}_{>0}^2$. Weitere (lokale) Extremstellen oder Sattelpunkte gibt es nicht.

(9) Die stetige Funktion

$$f(x, y) = y^4 - 3y^2 + \frac{1}{8}x^2 + |x|(y^2 - 1)$$

auf \mathbb{R}^2 hat Nichtdifferenzierbarkeitsstellen bei $x = 0$ (und $y \neq \pm 1$). Die kritischen Punkte mit $x > 0$ berechnen sich aus

$$\frac{1}{4}x + y^2 - 1 = 0, \quad 4y^3 - 6y + 2xy = 0,$$

die mit $x < 0$ aus

$$\frac{1}{4}x - y^2 + 1 = 0, \quad 4y^3 - 6y - 2xy = 0,$$

was beides auf $4y^3 - 6y - 8y(y^2 - 1) = 0$, also $4y^3 = 2y$ und $x = \pm 4|y^2 - 1|$ führt mit den Lösungen $y = 0$ oder $\pm\sqrt{2}$ und $x = \pm 4$. Die Werte von f in diesen insgesamt 6 kritischen Punkten sind $f(\pm 4, 0) = -2$, $f(\pm 4, \pm\sqrt{2}) = 4$. Da wieder $f(x, y) \rightarrow \infty$ gilt bei $|(x, y)| \rightarrow \infty$ (was man etwa aus $f(x, y) \geq y^4 - 3y^2 + \frac{1}{8}x^2 \geq y^2 + \frac{1}{8}x^2$ für $|y| \geq 2$ und $f(x, y) \geq -\frac{9}{4} + \frac{1}{8}|x|^2 - |x| \geq -\frac{9}{4} + |x|$ für $|y| \leq 2$ und $|x| \geq 16$ abliest), hat f mindestens eine absolute Minimumstelle auf \mathbb{R}^2 .

Also sind die beiden kritischen Punkte $(\pm 4, 0)$ mit dem kleineren Funktionswert -2 die absoluten Minimumstellen ???

Nein! Es müssen nämlich auch die Nichtdifferenzierbarkeitsstellen $(0, y)$ als mögliche Extremstellen in Betracht gezogen werden, und $f(0, y) = y^4 - 3y^2$ hat bei $y = \pm\sqrt{3/2}$ (und nur dort) ein Minimum mit Wert $-\frac{9}{4} < -2$. Also sind hier tatsächlich die beiden Nichtdifferenzierbarkeitsstellen $(0, \pm\sqrt{3/2})$ die absoluten Minimumstellen auf \mathbb{R}^2 und nicht etwa die kritischen Punkte mit dem kleinsten Funktionswert.

- *Man darf also Nichtdifferenzierbarkeitsstellen bei der Aufstellung der Kandidatenliste für Extremstellen nicht außer Acht lassen.*

Sonst ist die Kandidatenliste eben unter Umständen nicht vollständig und das Heraussuchen der Kandidaten mit dem kleinsten Funktionswert führt zum schlimmsten Fehler, den man bei der Extremwertbestimmung machen kann: Punkte als absolute Minimumstellen bzw. absolute Maximumstellen zu deklarieren, in denen gar nicht der kleinste bzw. größte Funktionswert vorliegt!

(10) In der Ökonomie kommen häufig *homogene Funktionen* $f(x)$ vor. Dafür gilt nach 5.2 die Eulersche Relation

$$x \cdot \nabla f(x) = s f(x),$$

wenn $s \in \mathbb{R}$ der Homogenitätsgrad ist, also $f(rx) = r^s f(x)$ für $r > 0$. Ist $f(x) > 0$ für $x \neq 0$ und $s \neq 0$, so folgt $\nabla f(x) \neq 0$ für $x \neq 0$.

- Also hat eine positive homogene Funktion f mit Homogenitätsgrad $s \neq 0$ keine kritischen Stellen $x \neq 0$.

Eine innere (lokale) Extremstelle einer solchen Funktion kann daher höchstens im Nullpunkt liegen (wo die Funktion evtl. nicht differenzierbar ist). Das gilt übrigens auch ohne Differenzierbarkeit von f , weil $\mathbb{R}_{>0} \ni r \mapsto f(rx) = r^s f(x)$ streng monoton ist für $s \neq 0$ und $f(x) \neq 0$. Bei positivem Homogenitätsgrad $s > 0$ und $f(0) = 0$ ist $x = 0$ tatsächlich die einzige Minimumstelle der ansonsten positiven homogenen Funktion. Ein Beispiel ist die Euklidische Norm $f(x) = |x|$ mit Homogenitätsgrad 1 und einziger absoluter Minimumstelle 0 auf \mathbb{R}^n , wo die Funktion nicht differenzierbar ist. Ist der Homogenitätsgrad $s < 0$, so gilt $f(rx) = r^s f(x) \rightarrow \infty$ bei $r \searrow 0$ und $x \neq 0$ mit $f(x) > 0$, also ist die Funktion dann im Ursprung nicht stetig ergänzbar und hat keine (lokalen) Extremstellen. Ein Beispiel ist $f(x) = \frac{1}{|x|}$ auf $\mathbb{R}_{\neq 0}^n$. Bei Homogenitätsgrad 0 schließlich ist die Funktion auf jedem Ursprungsstrahl konstant, und daher besteht mit jeder Minimum- bzw. Maximumstelle $x_* \neq 0$ auch der ganze Ursprungsstrahl $\mathbb{R}_{>0}x_*$ durch diesen Punkt aus lauter Minimum- bzw. Maximumstellen. Ist f stetig auf $\mathbb{R}_{\neq 0}^n$, so existieren es solche Extremstellen auch; denn auf der kompakten Einheitssphäre $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ nimmt die Funktion f gemäß Extremstellensatz ihr Minimum und ihr Maximum an und wegen der Konstanz auf Strahlen sind diese Werte dann auch das Minimum und Maximum von f auf $\mathbb{R}_{\neq 0}^n$.

Als konkretes Beispiel für den zuletzt betrachteten Fall des Homogenitätsgrades $s = 0$ nehmen wir

$$f(x, y) = \frac{x^2 - y^2}{x^2 + y^2} \quad \text{auf } \mathbb{R}_{\neq(0,0)}^2.$$

Hier ist

$$\nabla f(x, y) = \frac{4xy}{(x^2 + y^2)^2} \begin{pmatrix} y \\ -x \end{pmatrix}$$

mit Nullstellen genau in den Punkten $(x, 0)$ und $(0, y)$ (verschieden von $(0, 0)$). In den kritischen Punkten $(x, 0)$ hat f den Wert 1, in den anderen kritischen Punkten den Wert -1 , also sind die Stellen $(x, 0)$ mit $x \neq 0$ die absoluten Maximumstellen und die Stellen $(0, y)$ mit $y \neq 0$ die absoluten Minimumstellen von f auf $\mathbb{R}_{\neq(0,0)}^2$. (Das können wir schließen, weil wir schon überlegt haben, dass absolute Minimumstellen und Maximumstellen existieren.) Man sieht, dass die Menge der Extremstellen aus (hier insgesamt 4) Strahlen zusammengesetzt ist — wie allgemein bei vom Grad 0 homogenen Funktionen.

Das Ergebnis hätte man auch direkt aus $-x^2 - y^2 \leq x^2 - y^2 \leq x^2 + y^2$ mit Gleichheit nur für $x = 0$ bzw. für $y = 0$ sehen können, ohne Differentialrechnung zu bemühen. Es ging uns aber bei diesem wie bei allen obigen Beispielen darum, die Methode vorzuführen und nicht ad-hoc-Überlegungen zur Extremstellenbestimmung bei einer spezieller Funktion. ■

BEMERKUNG: Die explizite Auflösung von Systemen nichtlinearer Gleichungen, wie sie sich in den vorangehenden Beispielen für die kritischen Punkte ergeben haben, ist oft nicht möglich. In solchen Fällen ist man auf **numerische Verfahren der Extremstellenbestimmung** angewiesen.

Ein recht einleuchtendes Verfahren zum Auffinden der Maximumstellen ist z.B. die sogenannte **Methode des steilsten Anstiegs**, die auch ein Bergsteiger in der “diretissima” anwendet. Die Idee dabei ist, den Gradientenlinien zu folgen, d.h. den Kurven im Optimierungsbereich $D \subset \mathbb{R}^n$, deren Tangentenvektor stets in Gradientenrichtung, also in die Richtung der größten Funktionssteigung, zeigt. In der Praxis beginnt man mit einem beliebigen (bzw. schon möglichst gut geschätzten) Anfangspunkt $x_1 \in D$, berechnet $\nabla f(x_1)$, definiert dann $x_2 = h_1 \nabla f(x_1)$ mit einer geeignet erscheinenden Schrittgröße $h_1 > 0$, d.h. man bewegt sich vom Punkt x_1 aus ein Stück in Richtung des Gradienten zum nächsten Punkt x_2 . Dort wiederholt man den Verfahrensschritt, geht also weiter zu $x_3 = x_2 + h_2 \nabla f(x_2)$, und hofft, dass die so erhaltene Punktfolge x_1, x_2, x_3, \dots gegen eine Maximumstelle x^* von f auf D konvergiert. Für die Minimumbestimmung wendet man entsprechend die **Methode des steilsten Abstiegs** an, d.h. man folgt der Gegenrichtung des Gradienten $-\nabla f(x)$.

Dabei können freilich vielerlei Probleme auftreten. Beispielsweise braucht die Folge überhaupt keinen Grenzwert zu haben, oder sie kann gegen einen Sattelpunkt konvergieren oder auch gegen eine nichtabsolute lokale Maximumstelle. (Der Bergsteiger wäre dann der “diretissima” folgend nur auf einem Nebengipfel angekommen.) Daher ergänzt man das Verfahren in der Praxis durch Zufallsschritte, wobei z.B. mit sog. Schießverfahren Punkte zufällig ausgewählt und die Funktionswerte verglichen werden, damit man gegebenenfalls erkennt, dass die absolute Maximumstelle noch nicht erreicht ist. ■

BEISPIELE (Anwendung des notwendigen Kriteriums erster Ordnung für innere Extremstellen auf ökonomische Probleme): Die notwendige Bedingung $\frac{\partial f}{\partial x_1}(x_0) = 0, \dots, \frac{\partial f}{\partial x_n}(x_0) = 0$, bzw. kurz $\nabla f(x_0) = 0$, für innere Extremstellen x_0 einer Funktion von n Variablen ist der einfache Kern mancher, oft recht kompliziert und “eindrucksvoll” formulierten, “ökonomischen Gesetze”, wofür wir nun einige Beispiele geben:

1) Optimaler Faktoreinsatz bei der Herstellung eines Produkts:

Hier ist eine reelle *Produktionsfunktion* $x(r_1, \dots, r_n)$ gegeben, die in Abhängigkeit von eingesetzten Produktionsfaktoren r_j (in Faktoreinheiten) den Produktionsoutput x (in Outputeinheiten) angibt, sowie eine *Kostenfunktion* $K(r_1, \dots, r_n)$, welche die Kosten (in Geldeinheiten) für die eingesetzten Faktormengen erfasst, und eine *Preisfunktion* $p(x)$, welche den erzielbaren Marktpreis angibt (in Geldeinheiten pro Outputeinheit), wenn x produzierte Einheiten am Markt angeboten und — wie unterstellt wird — auch abgesetzt werden. Bei *vollständiger Konkurrenz* auf dem Markt ist der Preis p konstant, also unabhängig von der durch den einzelnen Produzenten angebotenen Menge (*Polypol*). Bei einem *monopolistischen Anbieter* ist andererseits $p(x)$ typischerweise eine fallende Funktion von x . Die Kostenfunktion hat oft eine lineare Struktur $K(r_1, \dots, r_n) = k_1 r_1 + \dots + k_n r_n + k_0$, wo k_0 die *Fixkosten* des Produzenten sind und k_j für $j = 1 \dots n$ der konstante *Faktorkostensatz* für eine Einheit des j -ten Produktionsfaktors. (Verwirrenderweise wird bei den Kosten $k_j r_j$ für die eingesetzten Einheiten des j -ten Produktionsfaktors vom j -ten *Faktoreinkommen* gesprochen, ebenso bei $k_1 r_1 + \dots + k_n r_n$ vom *Faktorgesamteinkommen*, obwohl es sich hier — von dem eingenommenen Standpunkt des Produzenten aus gesehen — um Ausgaben handelt, nicht um Einnahmen.)

Das Ziel des Produzenten ist natürlich, den Einsatz der Faktoren so zu steuern, dass sein Gewinn maximiert wird. Dazu liefert die sog. **Grenzproduktivitätstheorie** folgendes ökonomische Gesetz:

- *Im Gewinnmaximum stimmen für jeden Produktionsfaktor die Grenzkosten überein mit seiner durch den Grenzerlös bewerteten Grenzproduktivität.*

Was steckt dahinter? Zunächst ist die Gewinnfunktion in Abhängigkeit von den Faktoreinsätzen r_1, \dots, r_n aufzustellen. Dazu hat man in die Erlösfunktion $E(x) = p(x)x$ die Produktionsfunktion $x(r_1, \dots, r_n)$ einzusetzen und die Kostenfunktion davon zu subtrahieren. Man erhält so die *Gewinnfunktion*

$$G(r_1, \dots, r_n) = E(x(r_1, \dots, r_n)) - K(r_1, \dots, r_n).$$

Die notwendige Bedingung für einen nicht durch Kapazitätsgrenzen eingeschränkten (also im Inneren des Optimierungsbereiches gelegenen) optimalen Faktoreinsatz lautet nun $\frac{\partial}{\partial r_j} G(r_1, \dots, r_n) = 0$ für $j = 1 \dots n$. Mit Differenzregel und Kettenregel $\frac{\partial}{\partial r_j} (E \circ x) = (E' \circ x) \frac{\partial x}{\partial r_j}$ ist das äquivalent mit

$$\underbrace{E'(x(r_1, \dots, r_n))}_{\text{Grenzerlös}} \cdot \underbrace{\frac{\partial x}{\partial r_j}(r_1, \dots, r_n)}_{\text{Grenzproduktivität}} - \underbrace{\frac{\partial K}{\partial r_j}(r_1, \dots, r_n)}_{\text{Grenzkosten}} = 0 \quad \text{für } j = 1 \dots n,$$

und das oben ausgesprochene Gesetz ist nur eine Verbalisierung dieser Gleichung, die wir in 5.3 als ein Beispiel für die Anwendung der Kettenregel schon einmal hergeleitet haben. Die in Outputeinheiten pro Faktoreinheit angegebene Grenzproduktivität, also die Mehrproduktion bei zusätzlichem Einsatz einer Einheit des j -ten Produktionsfaktors, wird hier bewertet (multipliziert) mit dem in Geldeinheiten pro Outputeinheit gemessenen Grenzerlös, also dem Preis für eine zusätzlich angebotene und abgesetzte Outputeinheit, um so einen Wert in Geldeinheiten für die durch die Faktorerhöhung entstehende Outputzunahme festzulegen. Das Gesetz besagt, dass dieser Wert im Gewinnmaximum mit den Grenzkosten bzgl. desselben Faktors übereinstimmt, also mit den Kosten für Einbringung einer zusätzlichen Einheit des j -ten Faktors. Das ist mathematisch gesehen nichts anderes als das notwendige Kriterium erster Ordnung für innere Extremstellen in Verbindung mit der Kettenregel. Und es ist auch ökonomisch unmittelbar einleuchtend: Solange der Wert der Mehrproduktion bei Einsatz einer zusätzlichen Faktoreinheit die Kosten dafür übersteigt, ist das Gewinnmaximum noch nicht erreicht, und der Produzent wird den Faktoreinsatz erhöhen (wenn das nicht wegen Kapazitätsgrenzen unmöglich ist).

Im Fall eines Polypols ist p konstant, $E(x) = px$, $E'(x) \equiv p$, und dann ist die Grenzproduktivität mit dem konstanten Marktpreis p zu bewerten. Bei linearer Kostenfunktion sind die Grenzkosten $\frac{\partial K}{\partial x_j} \equiv k_j$ gleich dem konstanten Faktorkostensatz. Man sagt, wenn beides gegeben ist, dass

- *im Gewinnmaximum jeder Produktionsfaktor mit dem Marktwert seiner Grenzproduktivität entlohnt wird.*

In Formeln heißt das $p \cdot \frac{\partial x}{\partial r_j}(r_1, \dots, r_n) = k_j$, wozu sich in diesem Fall die notwendige Bedingung für innere Extrema der Gewinnfunktion spezialisiert. Die aufzubringenden Faktorkosten k_j pro Einheit werden auch als "Lohn" bezeichnet (insbesondere wenn man an den Produktionsfaktor Arbeit denkt). Das Gesetz drückt also aus, dass zur Gewinnmaximierenden Produktion der Faktoreinsatz solange gesteigert wird, bis für jeden Faktor der Marktwert der durch höheren Faktoreinsatz bewirkten Mehrproduktion kompensiert wird durch den vom Produzenten aufzubringenden Mehrlohn für die Faktoreinsatz-erhöhung. (Immer unterstellt, dass keine Kapazitätsgrenzen erreicht werden.)

Bei variablem Marktpreis $p(x)$ (Monopol) ist der Grenzerlös $E'(x)$ nicht gleich dem Marktpreis, sondern es gilt $E'(x) = \frac{d}{dx}(xp(x)) = p(x) + xp'(x)$, was wegen $xp'(x) = p(x) \cdot \frac{xp'(x)}{p(x)} = p(x)\varepsilon_p(x) = p(x)\frac{1}{\varepsilon_x(p(x))}$ multiplikativ so geschrieben werden kann:

$$E'(x) = p(x)[1 + \varepsilon_p(x)] = p(x) \left[1 + \frac{1}{\varepsilon_x(p(x))} \right].$$

Das ist die sog. *Amoroso–Robinson–Beziehung* zwischen Grenzerlös und Elastizität des Preises bzgl. der Nachfrage bzw. der Nachfrage bzgl. des Preises (siehe 4.4; mathematisch handelt es sich dabei nur um die Produktregel $\frac{d}{dx}(xp(x)) = p(x) + xp'(x)$.) Die notwendige Bedingung für Gewinn–maximierenden Faktoreinsatz lässt sich damit so formulieren:

$$p(x(r)) [1 + \varepsilon_p(x(r))] \frac{\partial x}{\partial r_j}(r) = \frac{\partial K}{\partial r_j}(r) \quad \text{für } j = 1 \dots n,$$

wobei wir $r = (r_1, \dots, r_n)$ für den Faktoreinsatzvektor abgekürzt haben.

Von ökonomischer Bedeutung ist es, den Kostenanteil für die einzelnen Faktoren am gesamten Produktionswert (hier der Erlös, da gemäß Annahme der ganze Output abgesetzt wird) zu kennen. Bei einer linearen Kostenfunktion $K(r) = k_1 r_1 + \dots + k_n r_n + k_0$ ist $r_j k_j = r_j \frac{\partial K}{\partial r_j}(r)$ der Kostenanteil für den j -ten Faktor, und bei optimaler Produktion lässt sich dies gemäß der notwendigen Bedingungen ausdrücken durch die Grenzproduktivität, also sind

$$r_j k_j = r_j E'(x(r)) \frac{\partial x}{\partial r_j}(r) = r_j p(x(r)) [1 + \varepsilon_p(x(r))] \frac{\partial x}{\partial r_j}(r)$$

die *Kosten für den Einsatz des j -ten Faktors* (das j -te “Faktoreinkommen”). Division durch $E(x) = xp(x)$ gibt mit der Definition der Output–Elastizität $\varepsilon_{x,r_j}(r) = r_j \frac{x}{r_j}(r)/x(r)$ dann den *Anteil der j -ten Faktorkosten am Gesamtproduktionswert* (“Einkommensanteil” des j -ten Faktors):

$$\frac{r_j k_j}{x(r)p(x(r))} = [1 + \varepsilon_p(x(r))] \varepsilon_{x,r_j}(r) \quad \text{für } j = 1 \dots n,$$

und diese n Gleichungen sind (im Fall einer linearen Kostenfunktion) äquivalent zu den notwendigen Bedingungen für ein Gewinnoptimum, da man alles auch zurückrechnen kann.

Bei konstantem Marktpreis p ist $\varepsilon_p(x) = 0$, also reduzieren sich die rechten Seiten zu $\varepsilon_{x,r_j}(r)$. Die Produktionselastizitäten geben also in dieser speziellen Situation, d.h. bei linearer Kostenfunktion und konstantem Marktpreis, gerade die Anteile der Faktorkosten am Produktionswert an, wenn mit maximalem Gewinn produziert wird. Der allgemeiner für lineare Kostenfunktionen gültige “ökonomische Lehrsatz”

- *Im Gewinnmaximum ist das Kostenverhältnis für je zwei Faktoren gleich dem Verhältnis ihrer Produktionselastizitäten.*

entpuppt sich als Kürzungsregel für das Bruchrechnen:

$$\frac{r_i k_i}{r_j k_j} = \frac{x(r)p(x(r))[1 + \varepsilon_p(x(r))]\varepsilon_{x,r_i}(r)}{x(r)p(x(r))[1 + \varepsilon_p(x(r))]\varepsilon_{x,r_j}(r)} = \frac{\varepsilon_{x,r_i}(r)}{\varepsilon_{x,r_j}(r)}.$$

Wichtig ist auch der Kostenanteil aller Faktoren zusammen am Produktionswert (das ist 1 minus der Fixkostenanteil). Dies ist — im Gewinnmaximum und bei linearer Kostenfunktion — gleich $[1 + \varepsilon_p(x(r))] \sum_{j=1}^n \varepsilon_{x,r_j}(r)$ und kann mit der *Skalenelastizität* der Produktionsfunktion $\sigma_x = \varepsilon_{x,r_1} + \dots + \varepsilon_{x,r_n}$ ausgedrückt werden (zu σ_x siehe 5.2). Der *Kostenanteil aller Faktoren am Gesamtproduktionswert* ist demnach

$$[1 + \varepsilon_p(x(r))] \sigma_x(r) = [1 + \varepsilon_p(x(r))] (\varepsilon_{x,r_1}(r) + \dots + \varepsilon_{x,r_n}(r)) .$$

Nehmen wir (realistisch) an, dass die Produktionsfunktion $x(r_1, \dots, r_n)$ streng wachsend von jedem Faktor r_j abhängt, so dass die Produktionselastizitäten ε_{x,r_j} und σ_x positiv sind, und nehmen wir (realistisch für Monopole) weiter an, dass der Preis $p(x)$ streng fallend vom Marktangebot (gleich Output) abhängt, so dass $\varepsilon_p(x) < 0$ ist, so folgt aus der im Gewinnmaximum gültigen Darstellung $[1 + \varepsilon_p(x(r))] \sigma_x(r)$ für den (positiven) Kostenanteil aller Faktoren, dass $-1 < \varepsilon_p(x(r)) < 0$ sein muss und somit $\varepsilon_x(p(x(r))) = 1/\varepsilon_p(x(r)) < -1$. Das ist ein “weiterer ökonomischer Lehrsatz”, der bei linearer Kostenfunktion gilt:

- *Im Gewinnmaximum ist der Faktoreinsatz eines monopolistischen Anbieters so gestaltet, dass sich der Produktionsoutput im Bereich elastischer Nachfrage befindet,*

also in dem Bereich, in dem Output-Zuwächse prozentual größere Nachfrageeinbrüche hervorrufen

Da bei Produktion mit Gewinn der Gesamtanteil aller Faktorkosten am Erlös kleiner als 1 ist, muss in diesem Fall $[1 + \varepsilon_p(x(r))] \sigma_x(r) < 1$ sein. Wenn also $[1 + \varepsilon_p(x(r))] \sigma_x(r) \geq 1$ ausfällt für jede Wahl des Faktoreinsatzvektors r , so kann nicht mit Gewinn produziert werden. Bei vollständiger Konkurrenz, also p konstant und $\varepsilon_p = 0$, ist daher die Skalenproduktionselastizität $\sigma_x(r) < 1$ im Gewinnmaximum, falls mit (positivem) Gewinn produziert wird. Mit anderen Worten: Wenn $\sigma_x(r) \geq 1$ ist für alle Wahlen des Faktoreinsatzvektors r , so ist Produktion mit Gewinn unmöglich. Da bei variablem Marktpreis die Bedingung $[1 + \varepsilon_p(x(r))] \sigma_x(r) < 1$ wegen $-1 < \varepsilon_p(x(r)) < 0$ eine geringere Einschränkung an die Skalenelastizität ist als $\sigma_x < 1$, lässt sich generell sagen:

- *In einer Monopolsituation kann noch bei höherer Skalenelastizität des Outputs mit Gewinn produziert werden als bei vollkommener Konkurrenz am Markt.*

2) Wir konkretisieren die Überlegungen aus 1) am Beispiel einer *Cobb–Douglas–Produktionsfunktion*

$$x(r_1, \dots, r_n) = c \cdot r_1^{s_1} r_2^{s_2} \cdot \dots \cdot r_n^{s_n} \quad (c, s, \dots, s_n \in \mathbb{R}_{>0}),$$

wobei $s = s_1 + s_2 + \dots + s_n < 1$ sei (d.h. bei Einsatz von großen Faktormengen wächst der Output prozentual langsamer als die Faktoreinsätze — eine ökonomisch sinnvolle Annahme). Bei konstantem Marktpreis p und linearer Kostenfunktion $K(r) = k_1 r_1 + \dots + k_n r_n + k_0$ (mit Fixkosten $k_0 \geq 0$ und Faktorkostensätzen $k_j > 0$ für $j = 1 \dots n$) lautet die Bedingung für ein Gewinnmaximum wegen $\frac{\partial x}{\partial r_j}(r) = \frac{s_j}{r_j} x(r)$ nun konkret

$$p \frac{s_j}{r_j} x(r) = k_j \quad \text{für } j = 1 \dots n .$$

Diese Gleichungen lassen sich eindeutig nach den r_j auflösen: Die Gleichungen zeigen zunächst, dass $r = (r_1, \dots, r_n)$ ein Vielfaches des Vektors $(\frac{s_1}{k_1}, \dots, \frac{s_n}{k_n})$ ist, und wenn

man $r = t(\frac{s_1}{k_1}, \dots, \frac{s_n}{k_n})$ einsetzt, so ergibt sich

$$p \frac{s_j}{t s_j / k_j} t^s x\left(\frac{s_1}{k_1}, \dots, \frac{s_n}{k_n}\right) = k_j, \quad \text{also} \quad t = p^{\frac{1}{1-s}} x\left(\frac{s_1}{k_1}, \dots, \frac{s_n}{k_n}\right)^{\frac{1}{1-s}}.$$

Somit gibt es in dem hier relevanten Bereich $\mathbb{R}_{>0}^n$ der (komponentenweise) positiven Faktoreinsatzvektoren genau einen kritischen Punkt r , und zwar

$$r_j = (cp)^{\frac{1}{1-s}} \left(\frac{s_1}{k_1}\right)^{\frac{s_1}{1-s}} \left(\frac{s_2}{k_2}\right)^{\frac{s_2}{1-s}} \cdots \left(\frac{s_n}{k_n}\right)^{\frac{s_n}{1-s}} \frac{s_j}{k_j} \quad \text{für } j = 1 \dots n.$$

Diese r_j heißen die *optimalen Cobb-Douglas-Faktoreinsätze*.

Da hier $x(r) = 0$ und $G(r) = px(r) - K(r) \leq -k_0 \leq 0$ gilt, wenn ein $r_j = 0$ ist (was bedeutet, dass r im Rande des Bereichs $\mathbb{R}_{>0}^n$ der positiven Faktoreinsatzvektoren liegt), und da $G(r) \rightarrow -\infty$ strebt bei $|r| \rightarrow \infty$, $r \in \mathbb{R}_{>0}^n$, (weil die Kosten linear anwachsen, der Erlös aber nur sublinear), während für kleine $r_j > 0$ der Gewinn $G(r)$ jedenfalls $> -k_0$ ist, kann man auch einsehen, dass die Gewinnfunktion $G(r)$ tatsächlich ein Maximum im offenen Bereich $\mathbb{R}_{>0}^n$ annimmt. (Siehe die Diskussion zum Zusatz des Extremstellensatzes oben.) Dieses Maximum kann nur in dem einzigen kritischen Punkt liegen, also wird bei Wahl der oben angegebenen optimalen Faktoreinsätze tatsächlich das Gewinnmaximum erreicht. Aber *das heißt noch nicht, dass mit Gewinn produziert werden kann*; denn der maximale Gewinn könnte negativ sein! Sicher ist nur, dass er größer ist als $-k_0$, so dass der Erlös die variablen Faktorkosten abdeckt, möglicherweise aber nicht die Fixkosten. Die Berechnung des Gewinns bei optimalem Faktoreinsatz durch Einsetzen in die Gewinnfunktion gibt als maximal erzielbaren Gewinn

$$G_{\max} = (1-s) c^{\frac{1}{1-s}} p^{\frac{s}{1-s}} \left(\frac{s_1}{k_1}\right)^{\frac{s_1}{1-s}} \left(\frac{s_2}{k_2}\right)^{\frac{s_2}{1-s}} \cdots \left(\frac{s_n}{k_n}\right)^{\frac{s_n}{1-s}} - k_0,$$

und dies muss positiv sein, damit überhaupt mit Gewinn produziert werden kann.

In der Praxis sind für die Faktoreinsatzmengen r_j meistens Kapazitätsgrenzen gegeben. Man kann dann an den Gleichungen für den optimalen Faktoreinsatz sehen, wie hoch die Kostensätze k_j sein dürfen, damit die zur Verfügung stehenden Faktoreinsatzmengen alle voll ausgenutzt werden. Dazu setzt man in

$$k_j = p \frac{s_j}{r_j} x(r_1, \dots, r_n) \quad \text{für } j = 1 \dots n$$

rechts die maximalen Faktormengen ein. Da $x(r)/r_j$ bei Verkleinerung von r_j zunimmt (wegen $s_j < 1$), kann bei solchermaßen festgelegten größten akzeptablen Faktorkostensätzen k_j für keinen Faktor r_j unterhalb des maximal möglichen Wertes die entsprechende Gleichung für optimalen Faktoreinsatz schon erfüllt sein.

Ist der Homogenitätsgrad $s = s_1 + \dots + s_n$ der Cobb-Douglas-Produktionsfunktion gleich 1, so versagt die obige Rechnung. Aber dann ist ja auch $G(r) + k_0 = px(r) - k_1 r_1 - \dots - k_n r_n$ vom Grad 1 positiv homogen auf $\mathbb{R}_{>0}^n$, so dass $G(r)$ in diesem Bereich sicher kein positives Maximum annimmt. (Entweder ist dort überall $G \leq -k_0$, oder es gilt $G(r) > -k_0$ für gewisse $r \in \mathbb{R}_{>0}^n$ und dann $\lim_{t \rightarrow \infty} G(tr) = \lim_{t \rightarrow \infty} t(G(r) + k_0) - k_0 = \infty$.) Vom Grad 1 homogene Produktionsfunktionen sind also nicht sinnvoll, jedenfalls ohne Faktor-Kapazitätsgrenzen bei konstantem Preis und linearer Kostenfunktion. Anders ist das, wenn $p(x)$ fallend in x ist oder $K(r)$ überproportional wachsend in den r_j . Dann sind auch homogen lineare Produktionsfunktionen $x(r)$ unter Umständen ökonomisch sinnvoll, d.h. die Gewinnfunktion $G(r) = p(x(r))x(r) - K(r)$ nimmt ein positives Maximum im Bereich $\mathbb{R}_{>0}^n$ an.

3) Optimaler Faktoreinsatz bei mehreren Produkten:

Hier hat man $m \geq 2$ Produktionsfunktionen $x_1(r_{11}, \dots, r_{1n}), \dots, x_m(r_{m1}, \dots, r_{mn})$, wobei r_{ij} die Einsatzmenge des j -ten Produktionsfaktors bei der Herstellung des i -ten Produkts ist. Der Gesamteinsatz des j -ten Faktors ist also $r_j = r_{1j} + \dots + r_{mj}$. Sind Preis-Angebots-Funktionen $p_i(x_1, \dots, x_m)$ gegeben (die Produkte können konkurrierend oder substitutiv sein, so dass der Marktpreis für jedes von ihnen auch vom Angebot der anderen Produkte abhängt) und eine Kostenfunktion $K(r_1, \dots, r_n)$, so lautet die Gewinnfunktion:

$$\begin{aligned} G(r) &= E(x_1(r_{11}, \dots, r_{1n}), \dots, x_m(r_{m1}, \dots, r_{mn})) - K(r_1, \dots, r_n) \\ &= \sum_{i=1}^m x_i(r_{i1}, \dots, r_{in}) p_i(x_1(r_{11}, \dots, r_{1n}), \dots, x_m(r_{m1}, \dots, r_{mn})) \\ &\quad - K(r_{11} + \dots + r_{m1}, r_{12} + \dots + r_{m2}, \dots, r_{1n} + \dots + r_{mn}), \end{aligned}$$

wo $r = (r_{ij})_{i=1 \dots m, j=1 \dots n}$ die Faktoreinsatzmatrix ist.

Man könnte die $m \cdot n$ Variablen r_{11}, \dots, r_{mn} auch mit einem einzigen Index durchnummerieren, aber das macht die Struktur weniger deutlich. Die notwendige Bedingung für eine Extremstelle von G auf dem Raum $\mathbb{R}_{\geq 0}^{m \times n}$ der Faktoreinsatzmatrizen mit nichtnegativen Einträgen lautet nun, wenn es sich um eine innere Stelle handelt (d.h. alle r_{ij} sind positiv): $\frac{\partial G}{\partial r_{ij}}(r) = 0$ für alle i, j , also gemäß Kettenregel:

$$\frac{\partial E}{\partial x_i}(x(r)) \frac{\partial x_i}{\partial r_{ij}}(r) = \frac{\partial K}{\partial r_j}(r_1, \dots, r_n) \quad \begin{array}{l} \text{für } i = 1 \dots m \\ \text{und } j = 1 \dots n. \end{array}$$

Dabei haben wir die m Produktionsfunktionen zu einer \mathbb{R}^m -wertigen Vektorfunktion $x(r)$ zusammengefasst. Die Voraussetzung, dass hierbei alle $r_{ij} > 0$ sind, ist unrealistisch, weil sie ja bedeutet, dass bei jedem Produkt alle n Produktionsfaktoren eingesetzt werden. Die Gleichung für optimalen Faktoreinsatz gilt aber auch schon für i, j , wenn nur $r_{ij} > 0$ ist (und evtl. andere $r_{hk} = 0$); denn dann hat die partielle Gewinnfunktion der Variablen r_{ij} eine innere Minimumstelle, also verschwindet ihre Ableitung dort. Die Gleichungen für Gewinn-maximierenden Faktoreinsatz kann man so verbalisieren:

- *Im Gewinnmaximum sind für jedes Produkt und jeden dafür eingesetzten Produktionsfaktor die mit dem Grenzerlös des Produkts bewertete Grenzproduktivität bzgl. des Faktors gleich den bzgl. dieses Faktors gebildeten Grenzkosten.*

Bei vollständiger Konkurrenz auf dem Markt (Polypol) erfolgt die Bewertung mit den konstanten Marktpreisen $\frac{\partial E}{\partial x_i} \equiv p_i$. Man sieht insbesondere, dass das Verhältnis zweier Grenzproduktivitäten $\frac{\partial x_i}{\partial r_{ij}}(r) : \frac{\partial x_i}{\partial r_{ik}}(r)$ desselben Produkts gleich dem Grenzkostenverhältnis $\frac{\partial K}{\partial r_j}(r_1, \dots, r_n) : \frac{\partial K}{\partial r_k}(r_1, \dots, r_n)$ und unabhängig vom Produkt ist. Dieses von vorneherein vielleicht nicht unmittelbar zu erwartende Ergebnis liegt an der speziellen Struktur der Kostenfunktion, die nur von den Variablensummen $r_j = r_{1j} + \dots + r_{mj}$ abhängt. Eine weitere Diskussion wie in 1) ist möglich, aber recht kompliziert (vor allem hinsichtlich der Notation); wir verzichten darauf.

4) Gewinn-optimales Angebot bei mehreren Produkten:

Hier sind Preis-Angebots-Funktionen $p_i(x_1, \dots, x_m)$ für m Produkte gegeben und eine von den angebotenen Outputmengen x_1, \dots, x_m dieser Produkte abhängige Gesamtkostenfunktion $K(x_1, \dots, x_m)$. Die angebotenen Mengen sollen Gewinn-maximierend be-

stimmt werden. (Das ist *nicht* dieselbe Problemstellung wie in 3), weil eine vom Faktoreinsatz abhängige Kostenfunktion im Allgemeinen nicht als Funktion der Produktions-Outputs aufgefasst werden kann. Nur wenn für alle gegebenen Output-Werte x_1, \dots, x_m jeweils *eindeutige* kostenoptimale Faktoreinsätze bestimmt werden können, lässt sich die Kostenfunktion — unterstellt der Faktoreinsatz erfolgt immer möglichst kostengünstig — als Funktion der Outputs auffassen.) Die Gewinnfunktion lautet jetzt

$$G(x_1, \dots, x_m) = E(x_1, \dots, x_m) - K(x_1, \dots, x_m),$$

mit der Erlösfunktion

$$E(x_1, \dots, x_m) = \sum_{i=1}^m x_i p_i(x_1, \dots, x_m),$$

und die Optimalitätsbedingung (notwendige Bedingung für Extremstellen in der offenen Menge $\mathbb{R}_{>0}^m$) ist $\frac{\partial G}{\partial x_i} = 0$ für $i = 1 \dots m$, also

$$\frac{\partial E}{\partial x_i}(x_1, \dots, x_m) = \frac{\partial K}{\partial x_i}(x_1, \dots, x_m).$$

- Bei Gewinn-optimal angebotenen Outputmengen stimmen für jedes Produkt Grenzkosten und Grenzerlös überein;
- im Fall vollständiger Konkurrenz am Markt für ein Produkt stimmen dann also seine Grenzkosten mit seinem Marktpreis überein;

denn bei konstantem Marktpreis p_i ist ja $\frac{\partial E}{\partial x_i}(x_1, \dots, x_m) = p_i$. Dieses ökonomische Gesetz ist nicht überraschend, da wohlbekannt (aus 4.5) im Falle der Produktion eines einzigen Produkts. Zu betonen ist wieder, dass das Gesetz nur gilt, wenn keiner der Outputs durch Erreichen einer Kapazitätsgrenze restringiert ist (sonst handelt es sich nicht um eine *innere* Gewinn-Maximumstelle). Wir erwähnen noch, dass die Grenzerlöse gemäß Produktregel gegeben sind durch

$$\frac{\partial E}{\partial x_h}(x_1, \dots, x_m) = p_h(x_1, \dots, x_m) + \sum_{i=1}^m x_i \frac{\partial p_i}{\partial x_h}(x_1, \dots, x_m),$$

das ist also der Preis $p_h(x_1, \dots, x_m)$ zuzüglich der mit ihren Grenzpreisen bzgl. x_h bewerteten Outputs aller Produkte.

Liegen Absatz-Preis-Funktionen $x_i(p_1, \dots, p_m)$ vor, die beschreiben, welche Mengen des i -ten Produkts auf den Markt gebracht und abgesetzt werden können, wenn die Preise für alle m Produkte gegeben sind, so kann man analoge Optimalitätskriterien für den Gewinn aufstellen durch den Ansatz

mit $G(p_1, \dots, p_m) = E(p_1, \dots, p_m) - K(x_1(p_1, \dots, p_m), \dots, x_m(p_1, \dots, p_m))$

$$E(p_1, \dots, p_m) = \sum_{i=1}^m p_i x_i(p_1, \dots, p_m).$$

Die Bedingung $\frac{\partial G}{\partial p_i}$ für **Gewinn-optimale Preisgestaltung** lautet jetzt unter Anwendung der Kettenregel (mit der Abkürzung $x_i(p)$ für $x_i(p_1, \dots, p_m)$):

$$\begin{aligned} \frac{\partial E}{\partial p_i}(p_1, \dots, p_m) &= \frac{\partial}{\partial p_i} K(x_1(p_1, \dots, p_m), \dots, x_m(p_1, \dots, p_m)) \\ &= \sum_{h=1}^m \frac{\partial K}{\partial x_h}(x_1(p), \dots, x_m(p)) \frac{\partial x_h}{\partial p_i}(p_1, \dots, p_m) \quad \text{für } i = 1 \dots m. \end{aligned}$$

5) Optimales Angebot von einem Produkt für mehrere Märkte:

Hier bezeichnen x_1, \dots, x_n die von einem Produkt auf n verschiedenen Märkten, in denen der Anbieter ein Monopol hat, angebotenen und abgesetzten Mengen. Die Märkte sind durch *Preis-Angebots-Funktionen* $p_j(x_j)$ gekennzeichnet, die Produktionskosten durch eine nur von der Gesamtproduktion abhängige *Kostenfunktion* $K(x_1 + \dots + x_n)$. Dann ist

$$G(x_1, \dots, x_n) = \sum_{j=1}^n E_j(x_j) - K(x_1 + \dots + x_n) = \sum_{j=1}^n x_j p_j(x_j) - K(x_1 + \dots + x_n)$$

die *Gewinnfunktion*, und die notwendige Bedingung $\frac{\partial G}{\partial x_j} = 0$ für eine innere Gewinn-Maximumstelle (auf $\mathbb{R}_{>0}^n$) lautet hier:

$$E'_j(x_j) = K'(x_1 + \dots + x_n) \quad \text{für } j = 1 \dots n.$$

- *Bei Gewinn-optimalen Angebot stimmt der Grenzerlös für jeden Teilmarkt mit den Grenzkosten überein.*

Wegen der Produktregel

$$E'_j(x_j) = \frac{d}{dx_j}(x_j p_j(x_j)) = p_j(x_j) + x_j p'_j(x_j) = p_j(x_j) [1 + \varepsilon_{p_j}(x_j)]$$

und weil die Nachfrageelastizität des Preises $\varepsilon_{p_j}(x_j)$ das Reziproke der Preiselastizität der Nachfrage $\varepsilon_{x_j}(p_j)$ ist (für $p_j = p_j(x_j)$), folgt aus der notwendigen Optimalitätsbedingung:

$$\frac{p_j(x_j)}{p_k(x_k)} = \frac{1 + \varepsilon_{p_k}(x_k)}{1 + \varepsilon_{p_j}(x_j)} = \frac{1 + 1/\varepsilon_{x_k}(p_k(x_k))}{1 + 1/\varepsilon_{x_j}(p_j(x_j))}.$$

Daraus liest man zum Beispiel ab:

- *Bei Gewinn-optimalen Angebot ist die Preiselastizität der Nachfrage auf allen Teilmärkten elastisch,*
- *auf einem Teilmarkt mit höherer Preiselastizität der Nachfrage wird ein niedrigerer Preis erzielt;*
- *bei gleichem Preis auf Märkten mit unterschiedlicher Preiselastizität der Nachfrage kann das Angebot also nicht Gewinn-optimierend sein.*

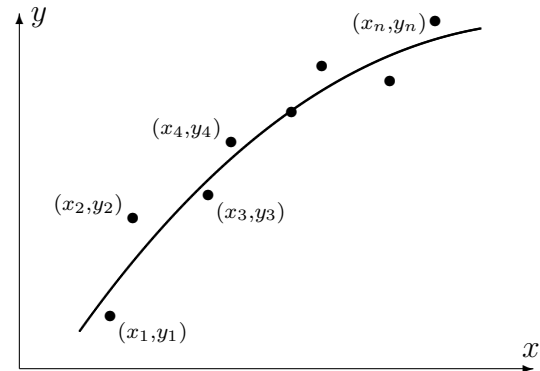
Wegen $p_j(x_j)[1 + \varepsilon_{p_j}(x_j)] = E'_j(x_j) = K'(x_1 + \dots + x_n) > 0$ im Gewinnmaximum ist nämlich $1 + \varepsilon_{p_j}(x_j) > 0$, also $\varepsilon_{p_j}(x_j) > -1$ und $\varepsilon_{x_j}(p_j(x_j)) < -1$ (da natürlich $\varepsilon_{p_j} < 0$ ist), und größere Elastizität bei x_k bedeutet $\varepsilon_{x_k}(p_k(x_k)) < \varepsilon_{x_j}(p_j(x_j)) < -1$, also $-1 < \varepsilon_{p_j}(x_j) < \varepsilon_{p_k}(x_k) < 0$ und somit $p_k(x_k) < p_j(x_j)$. ■

Die Liste solcher Beispiele könnte man fortsetzen — mehrere Produkte auf verschiedenen Teilmärkten etc. —, aber der mathematische Kern ist immer derselbe: In inneren Extremstellen einer differenzierbaren Funktion verschwinden alle partiellen Ableitungen. Alles Weitere ist nur die Anwendung von Differentiationsrechenregeln und ökonomischer Fachjargon. Deshalb brechen wir die Liste der Beispiele für ökonomische Anwendungen des notwendigen Kriteriums erster Ordnung für innere Extremstellen hier ab und wenden uns einer anders gelagerten Anwendung dieses Kriteriums zu, die auch für die Wirtschaftswissenschaft relevant ist.

Das notwendige Kriterium für innere Extremstellen hat auch wichtige Anwendungen auf Probleme der Statistik und damit — indirekt über die Wirtschaftsstatistik — auch in der Ökonomie. Mit den folgenden Ausführungen geben wir in diese Richtung nur einen

AUSBLICK (auf Optimierungsprobleme in der Datenanalyse und Statistik):

1) Bei der sog. **Regressionsanalyse** geht es darum, aus einer endlichen Folge von beobachteten reellen Datenpaaren (x_j, y_j) ($j = 1 \dots n$), welche z.B. die Werte zweier in Beziehung zueinander stehenden ökonomischen Größen x, y zu n verschiedenen Zeitpunkten darstellen, “in sinnvoller Weise” eine Funktion $y = f(x)$ zu bestimmen, welche die Gleichungen $y_j = f(x_j)$ für $j = 1 \dots n$ “möglichst gut” erfüllt. Geometrisch gesprochen sind die Paare (x_j, y_j) Punkte in der Ebene \mathbb{R}^2 und man sucht einen Funktionsgraphen, der möglichst nahe bei diesen Punkten verläuft.



Natürlich kann man, wenn alle x_j verschieden sind, immer Funktionen f mit $f(x_j) = y_j$ für $j = 1 \dots n$ angeben; es gibt sogar eine eindeutige Polynomfunktion vom Grad $< n$ mit dieser Eigenschaft. Aber das ist hier nicht der Punkt. In der Regressionsanalyse will man ein möglicherweise bestehendes *einfaches Gesetz* über den quantitativen Zusammenhang zwischen x und y herausfinden, und das ist nur möglich, wenn man *die Regressionsfunktion unter Funktionen eines vorgegebenen einfachen Typs sucht*.

Beispiele für solche in der Praxis verwendete sog. **Ansatzfunktionen** sind:

<i>Lineare Funktionen</i>	$ax + b,$
<i>quadratische Funktionen</i>	$ax^2 + bx + c,$
<i>Potenzfunktionen</i>	$ax^b + c,$
<i>Exponentialfunktionen</i>	$ae^{bx} + c,$
<i>logistische Funktionen</i>	$\frac{a}{1 + be^{cx}}.$

Die Auswahl des Typs ist ein Frage der Vorinformation über das Problem und der Erfahrung bzw. Konvention. Hat man sich für den Typ der Regressionsfunktion entschieden, so geht es meistens nur noch darum, die darin auftretenden endlich vielen Parameter a, b, c, \dots so zu bestimmen, dass die Abweichung des Funktionsgraphen von den beobachteten Wertepaaren möglichst klein ist. Dies ist nun ein Optimierungsproblem in endlich vielen Variablen (den Parametern der Ansatzfunktion), das definiert ist, sobald man ein Maß für die Abweichung eines Funktionsgraphen $\text{Graph}(f)$ von einer endlichen Menge vorgegebener Wertepaare (x_j, y_j) spezifiziert hat. Die gängigste Wahl eines solchen Maßes ist die Summe der Quadrate der vertikalen Abweichungen der Punkte vom Graphen

$$\sum_{j=1}^n |f(x_j) - y_j|^2;$$

man spricht dann von der **Methode der kleinsten Quadrate**. Dieses Maß für die Abweichung hat etwa im Vergleich zu dem naheliegenden Maß $\sum_{j=1}^n |f(x_j) - y_j|$ rechen-technische Vorteile. Es berücksichtigt allerdings einzelne aus dem Rahmen fallende große Abweichungen $|f(x_j) - y_j|$ ziemlich stark. Jedes vernünftige Maß für die Abweichung eines Graphen von endlich vielen Punktpaaren ist natürlich genau dann Null, wenn alle Punktpaare exakt auf dem Graphen liegen.

2) Wir führen die Regressionsanalyse für den Fall einer linearen Ansatzfunktion vor (siehe auch 1.5): Zu gegebenen Paaren $(x_1, y_1), \dots, (x_n, y_n)$ reeller Zahlen soll also eine **lineare Regressionsfunktion** $\ell(x) = ax + b$ bestimmt werden. Der Graph dieser Funktion heißt dann **Regressionsgerade**. Wenn wir zur Messung der Abweichung der Geraden von den Datenpaaren die Methode der kleinsten Quadrate verwenden, so haben wir dann die Funktion

$$\varphi(a, b) = \sum_{j=1}^n (\ell(x_j) - y_j)^2 = \sum_{j=1}^n (ax_j + b - y_j)^2$$

der beiden reellen Variablen a, b zu minimieren. Diese Funktion ist (echt) quadratisch und nichtnegativ auf \mathbb{R}^2 und nimmt daher ein Minimum auf \mathbb{R}^2 an (wie man der früheren Diskussion der Extremstellen quadratischer Funktionen von mehreren Veränderlichen entnehmen kann). Die partiellen Ableitungen sind

$$\begin{aligned} \frac{\partial \varphi}{\partial a}(a, b) &= \sum_{j=1}^n 2(ax_j + b - y_j)x_j = 2a \sum_{j=1}^n x_j^2 + 2b \sum_{j=1}^n x_j - 2 \sum_{j=1}^n x_j y_j, \\ \frac{\partial \varphi}{\partial b}(a, b) &= \sum_{j=1}^n 2(ax_j + b - y_j) = 2a \sum_{j=1}^n x_j + 2bn - 2 \sum_{j=1}^n y_j. \end{aligned}$$

Setzen wir die partiellen Ableitungen von φ Null, so ergeben sich zwei lineare Gleichungen für a, b . Mit den Abkürzungen $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$, $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$ für die *Mittelwerte* und $s_{x^2} = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ für die sog. *Varianz* der x_j sowie $s_{xy} = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})$ für die sog. *Kovarianz* der x_j, y_j findet man

$$\begin{aligned} a(s_{x^2} + \bar{x}^2) + b\bar{x} &= s_{xy} + \bar{x}\bar{y} \\ a\bar{x} + b &= \bar{y} \end{aligned}$$

mit der Lösung

$$a = \frac{s_{xy}}{s_{x^2}} = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}, \quad b = \bar{y} - a\bar{x} = \frac{1}{n} \sum_{j=1}^n (y_j - ax_j),$$

sofern $s_{x^2} \neq 0$ ist, also nicht alle x_j übereinstimmen. Die lineare Regressionsfunktion ist dann

$$\ell(x) = \frac{s_{xy}}{s_{x^2}}(x - \bar{x}) + \bar{y}.$$

(Im Fall $x_j = \bar{x}$ für alle j sind alle linearen Funktionen $\ell(x) = a(x - \bar{x}) + \bar{y}$ mit beliebigem $a \in \mathbb{R}$ Regressionsfunktionen.)

3) Ein anderes Regressionsproblem: Bei einem Vorgang, der zu "Inputvektoren" $x \in \mathbb{R}^n$ durch eine lineare Transformation $x \mapsto Ax \in \mathbb{R}^m$ mit bekannter $m \times n$ -Matrix A "Outputvektoren" $Ax \in \mathbb{R}^m$ liefert, wird zu einem *beobachteten Vektor* $y \in \mathbb{R}^m$ ein *optimaler Parametervektor* $x \in \mathbb{R}^n$ gesucht, derart dass die quadratische Abweichung

$$\varphi(x) = |y - Ax|^2 = \sum_{i=1}^m \left(y_i - \sum_{j=1}^n a_{ij} x_j \right)^2$$

minimal wird. Die Matrix A heißt auch *Design-Matrix* und die ganze Aufgabenstellung **multiple lineare Regression**.

Wieder ist $\varphi(x)$ eine nichtnegative quadratische Funktion und nimmt folglich ein Minimum auf \mathbb{R}^n an. Mit der Produktregel finden wir für alle Richtungsvektoren $u \in \mathbb{R}^n$:

$$\begin{aligned}\partial_u \varphi(x) &= \left. \frac{d}{dt} \right|_{t=0} [(y - A(x + tu)) \cdot (y - A(x + tu))] \\ &= -(Au) \cdot (y - Ax) - (y - Ax) \cdot (Au) = -2y \cdot (Au) + 2(Ax) \cdot (Au) \\ &= 2A^\top (Ax - y) \cdot u.\end{aligned}$$

Die kritischen Punkte x von φ sind gekennzeichnet durch $\partial_u \varphi(x) = 0$ für alle $u \in \mathbb{R}^n$, werden also beschrieben durch das lineare Gleichungssystem

$$A^\top Ax = A^\top y \quad (\text{sog. Normalgleichungen für } x).$$

Diese Gleichungen sind äquivalent mit $(Ax - y) \cdot Au = (A^\top (Ax - y)) \cdot u = 0$ für alle $u \in \mathbb{R}^n$, also mit $Ax - y \perp A\mathbb{R}^n$. Da φ ein Minimum annimmt und als quadratische Funktion dann nur Minimumstellen als kritische Punkte hat, gibt es (mindestens) eine Lösung und jede Lösung ist eine Minimumstelle von φ auf \mathbb{R}^n .

(All dies hätten wir ohne Differentialrechnung mit Linearer Algebra direkt sehen können; denn die Aufgabe, $\varphi(x) = |Ax - y|^2$ zu minimieren ist äquivalent mit der Bestimmung eines zu y nächsten Punktes im Unterraum $A\mathbb{R}^n$ von \mathbb{R}^m , und diesen erhält man nach 3.5 in der Form $Ax = y^\top$ durch orthogonale Zerlegung $y = y^\top + y^\perp$ mit $y^\top \in A\mathbb{R}^m$, $y^\perp \in (A\mathbb{R}^m)^\perp$. Es ging hier aber darum, die Anwendung der notwendigen Bedingung für Extremstellen zu demonstrieren.)

4) Eine verwandte Aufgabenstellung bezieht sich darauf, zu n beobachteten Werten x_1, \dots, x_n einer reellen "Zufallsgröße" eine Verteilungsfunktion auf \mathbb{R} anzugeben, welche mit den Beobachtungen möglichst gut übereinstimmt. Hier nimmt man als Ansatzfunktion z.B. die

$$\text{Gaußsche Normalverteilung } v_{\sigma, \mu}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2},$$

deren Parameter $\mu \in \mathbb{R}$ und $\sigma^2 \in \mathbb{R}_{>0}$ Erwartungswert und Varianz heißen. Die Funktion erfüllt $\int_{-\infty}^{\infty} v_{\sigma, \mu}(x) dx = 1$, und die Integrale $\int_a^b v_{\sigma, \mu}(x) dx$ sind zu interpretieren als Wahrscheinlichkeit dafür, dass eine gemäß $v_{\sigma, \mu}$ verteilte reelle Zufallsgröße ihren Wert im Intervall $[a, b]$ hat. Eine Methode der optimalen Anpassung der Parameter μ, σ an die beobachteten n Werte besteht nun darin, die Parameter so zu wählen, dass die Wahrscheinlichkeit, diese n Werte bei tatsächlicher Verteilung gemäß $v_{\sigma, \mu}$ zu beobachten, möglichst groß wird. Diese sog. **Maximum-Likelihood-Methode** läuft im vorliegenden Fall (aufgrund hier unterschlagener wahrscheinlichkeitstheoretischer Überlegungen) darauf hinaus, die Funktion

$$\varphi(\sigma, \mu) = \prod_{i=1}^n v_{\sigma, \mu}(x_i) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left(\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2 \right)$$

zu maximieren. Die sich so ergebenden Parameter μ und σ^2 heißen *Maximum-Likelihood-Schätzer* des Erwartungswerts und der Varianz zu den Beobachtungswerten x_1, \dots, x_n ("Schätzer", weil man die Parameter der "wahren" Normalverteilung nicht kennt). Die ganze Aufgabenstellung: Schätzen von Parametern in Verteilungsfunktionen aufgrund beobachteter Daten, gehört in den Bereich der **Statistik**.

Die Funktion φ mag furchterregend aussehen, aber wenn wir statt φ äquivalent den Logarithmus $\psi = \ln \varphi$ maximieren, so ist alles schon viel einfacher:

$$\psi(\sigma, \mu) = \ln \varphi(\sigma, \mu) = n \ln \frac{1}{\sqrt{2\pi}} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Bei $\sigma \searrow 0$ strebt $\psi(\sigma, \mu) \rightarrow -\infty$ (wenn nicht alle $x_i = \mu$ sind; in diesem Fall ist der Limes $+\infty$ und ein Maximum nicht existent), ebenso bei $|\mu| \rightarrow \infty$ und bei $\sigma \rightarrow \infty$. Dies suggeriert, dass $\psi(\sigma, \mu)$ in $\mathbb{R}_{>0} \times \mathbb{R}$ eine (innere) Maximumstelle hat, und das kann durch eine etwas genauere Überlegung auch bewiesen werden. Zur Bestimmung der Stelle berechnen wir

$$\begin{aligned} \frac{\partial \psi}{\partial \sigma}(\sigma, \mu) &= \frac{-n}{\sigma} - \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2, \\ \frac{\partial \psi}{\partial \mu}(\sigma, \mu) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{n}{\sigma^2} (\bar{x} - \mu) \end{aligned}$$

mit dem Mittelwert $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Die notwendige Bedingung für eine innere Extremstelle von ψ lautet also hier:

$$\begin{aligned} \mu &= \bar{x} \quad \text{der Mittelwert der } x_i, \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{die Varianz der } x_i. \end{aligned}$$

Da es nur diese eine kritische Stelle gibt (wenn nicht alle x_i gleich sind), muss sie die eindeutige Minimumstelle von ψ auf $\mathbb{R}_{>0} \times \mathbb{R}$ sein.

Die Maximum-Likelyhood-Methode liefert also für den Erwartungswert μ und die Varianz σ^2 der an die Beobachtungen am besten angepassten Normalverteilung $v_{\sigma, \mu}$ den Mittelwert der x_i als Schätzwert für μ und die Varianz der x_i als Schätzwert für σ^2 . ■