nur positive Werte an, $f(\boldsymbol{x})$ vergrößert sich also unabhängig davon, in welcher Richtung man sich von $\boldsymbol{x}$ entfernt, und damit muß $f$ in $\boldsymbol{x}$ ein Minimum haben. Entsprechend muß $f$ für negativ definite HESSE-Matrix ein Maximum haben. Im indefiniten Fall gibt es Richtungen, in denen sich $f(\boldsymbol{x})$ vergrößert und Richtungen, in denen es sich verkleinert; daher kann in diesem Fall kein Extremum vorliegen. ■

Klassisches Beispiel eines Punkts, in dem der Gradient verschwindet, ohne daß ein Extremwert vorliegt, ist der in Abbildung 41 gezeigte Sattelpunkt, hier dargestellt als Funktionswert über dem Punkt $(0,0)$ für die Funktion $f(x,y) = x^2 - y^2$.

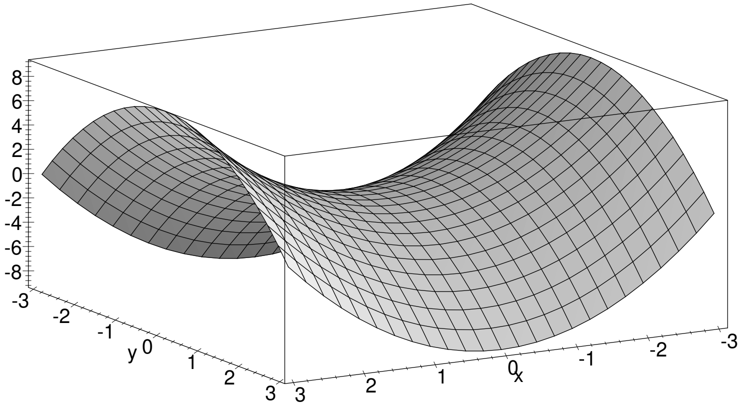


Abb. 41: Graph der Funktion $f(x,y) = x^2 - y^2$

## §2: Grundlegende Sätze

Einige der fundamentalen Rechenregeln für Ableitungen von Funktionen einer reellen Veränderlichen lassen sich problemlos auf Funktionen mehrerer Veränderlicher übertragen; ein Beispiel hierfür ist etwa die Linearität der Ableitung: Für zwei Funktionen $f, g\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ und $\lambda, \mu \in \mathbb{R}$ ist

$$\operatorname{grad}(\lambda f + \mu g) = \lambda \operatorname{grad} f + \mu \operatorname{grad} g\,.$$



Auch die LEIBNIZ-Regel verallgemeinert sich, wie auf dem Übungsblatt nachgerechnet wurde, zu

$$\operatorname{grad}(f \cdot g) = f \operatorname{grad} g + g \operatorname{grad} f\,.$$

Die nächste wichtige Grundregel für die Ableitung von Funktionen einer Veränderlichen ist die Kettenregel; hier müssen wir uns erst überlegen, welche Art von Funktionen wir hier miteinander verknüpfen wollen.

### a) Eine Kettenregel

Wir begnügen uns hier mit der Verknüpfung zwischen skalarwertigen Funktionen mehrerer Veränderlicher mit Funktionen einer Variablen nach $\mathbb{R}^n$. Eine solche Funktion $\gamma\colon (a,b) \to \mathbb{R}^n$ hat als Bild ein *Kurvenstück* im $\mathbb{R}^n$ und wird deshalb auch so bezeichnet. Sie ist gegeben durch $n$ Funktionen $\gamma_i\colon (a,b) \to \mathbb{R}$, den Koordinaten des Punktes $\gamma(t)$, und als ihre Ableitung bezeichnen wir jenen Vektor $\dot\gamma(t)$, dessen Komponenten die Ableitungen $\dot\gamma_i(t)$ sind.

Anschaulich kann man sich ein Kurvenstück durch sein Bild im $\mathbb{R}^n$ vorstellen, allerdings muß man beachten, daß dasselbe Bild durch ganz verschiedene Funktionen parametrisiert werden kann. Als Beispiel für verschiedene Parametrisierungen einer und derselben Kurve betrachten wir den Einheitskreis $x^2 + y^2 = 1$. Seine bekannteste Darstellung als Kurvenstück ist die Parametrisierung

$$\gamma_1\colon [0, 2\pi] \to \mathbb{R}^2; \quad t \mapsto (\cos t, \sin t)\,,$$

aber natürlich wäre auch

$$\gamma_2\colon [0,1] \to \mathbb{R}^2; \quad t \mapsto (\cos 2\pi t, \sin 2\pi t)$$

eine Möglichkeit. Es geht aber auch ganz anders, denn auch bei

$$\gamma_3\colon \mathbb{R} \to \mathbb{R}^2; \quad t \mapsto \left(\frac{t^2-1}{t^2+1}, \frac{2t}{t^2+1}\right)$$

liegen sämtliche Bildpunkte auf dem Einheitskreis, und zumindest für jedes endliche Teilintervall von $\mathbb{R}$ definiert auch $\gamma_3$ ein Kurvenstück. Wie eine kurze Kurvendiskussion (oder ein Blick auf Abbildung 42)

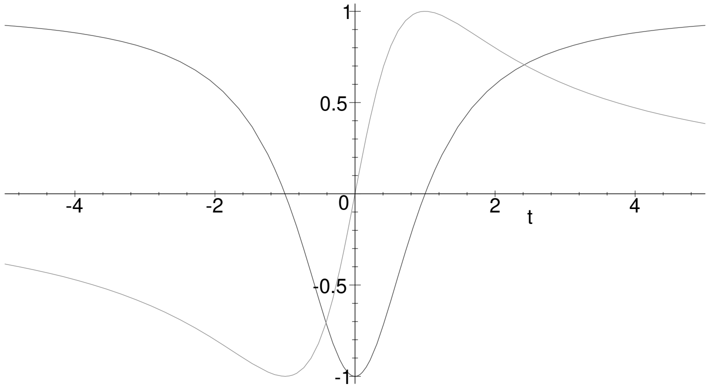


Abb. 42: Graph der Funktionen $x(t) = \frac{t^2-1}{t^2+1}$ und $y(t) = \frac{2t}{t^2+1}$

zeigt, besteht das Bild von $\gamma_3$ aus allen Punkten des Einheitskreises außer $(1,0)$; letzterer Punkt ist der Grenzwert von $\gamma_3(t)$ für $t \to \pm\infty$.

Jede dieser Parametrisierungen führt zu anderen Tangentenvektoren: Für einen festen Kurvenpunkt liegen zwar alle drei Vektoren auf ein und der selben Geraden, der Tangenten an den Kreis, aber sie haben verschiedene Länge: Für $(x,y) = \gamma_1(t)$ ist

$$\dot\gamma_1(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix},$$

für $(x,y) = \gamma_2(t)$ ist

$$\dot\gamma_2(t) = \begin{pmatrix} -2\pi \sin 2\pi t \\ 2\pi \cos 2\pi t \end{pmatrix} = \begin{pmatrix} -2\pi y \\ 2\pi x \end{pmatrix},$$

und für $(x,y) = \gamma_3(t)$ schließlich zeigt eine kurze Rechnung, daß

$$\dot\gamma_3(t) = \frac{2}{(t^2+1)^2} \cdot \begin{pmatrix} 2t \\ 1-t^2 \end{pmatrix} = \frac{2}{t^2+1} \begin{pmatrix} y \\ -x \end{pmatrix}$$

ist. Bei der Parametrisierung mit $\gamma_2$ sind die Tangentenvektoren also jeweils $2\pi$ mal so lang wie bei der mit $\gamma_1$, was man anschaulich so interpretieren kann, daß der Kreis bei dieser Parametrisierung $2\pi$ mal so schnell durchlaufen wird wie bei der mit $\gamma_1$. Bei der Parametrisierung



mit $\gamma_3$ ist die Länge der Vektoren variabel, und sie zeigen auch in die Gegenrichtung zu den Tangentenvektoren an $\gamma_1$ und $\gamma_2$; bei dieser Parametrisierung wird der Kreis also im Uhrzeigersinn durchlaufen.

Für die Verknüpfungen eines Kurvenstücks mit einer Funktion auf dem $\mathbb{R}^n$ haben wir die folgende Kettenregel:

**Satz:** $D \subseteq \mathbb{R}^n$ sei eine offene Teilmenge des $\mathbb{R}^n$ und $f \in \mathcal{C}^1(D,\mathbb{R})$. Für eine stetig differenzierbare Funktion $\gamma\colon (a,b) \to D$ gilt dann

$$\frac{d}{dt} f\bigl(\gamma(t)\bigr) = (\operatorname{grad} f)\bigl(\gamma(t)\bigr) \cdot \dot\gamma(t)\,.$$

*Beweis:* Nach Definition der Ableitung ist

$$\frac{d}{dt} f\bigl(\gamma(t)\bigr) = \lim_{h\to 0} \frac{f\bigl(\gamma(t+h)\bigr) - f\bigl(\gamma(t)\bigr)}{h}\,.$$

Da für jede der Komponenten $\gamma_i$

$$\gamma_i(t+h) = \gamma_i(t) + h\dot\gamma_i(t) + o(h)$$

ist, ist auch

$$\gamma(t+h) = \gamma(t) + h\dot\gamma(t) + o(h)\,,$$

wobei ich hier, einer üblichen Untugend folgend, einen Vektor, dessen sämtliche Komponenten $o(h)$ sind, kurz selbst mit $o(h)$ bezeichnet habe.

Damit ist, wegen der stetigen Differenzierbarkeit von $f$,

$$\begin{aligned} \frac{d}{dt} f\bigl(\gamma(t)\bigr) &= \lim_{h\to 0} \frac{f\bigl(\gamma(t)\bigr) + (\operatorname{grad} f)\bigl(\dot\gamma(t) + o(h)\bigr) - f\bigl(\gamma(t)\bigr)}{h} \\ &= \lim_{h\to 0} \frac{h(\operatorname{grad} f)\bigl(\dot\gamma(t)\bigr) + o(h)}{h} \\ &= (\operatorname{grad} f)\bigl(\gamma(t)\bigr) \cdot \dot\gamma(t)\,, \end{aligned}$$

womit die Behauptung bewiesen wäre. ∎

## b) Der Satz über implizite Funktionen

Hier geht es, wie bereits erwähnt, um die Auflösung einer Gleichung $F(\boldsymbol{x}) = 0$ nach einer der Variablen $x_i$. In einfachen Fällen ist dies trivial möglich, beispielsweise läßt sich

$$F(x, y, z) = ax + by + cz = 0$$

für $c \neq 0$ durch

$$z = \frac{-ax - by}{c}$$

nach $z$ auflösen. In etwas komplizierteren Fällen, wie etwa bei

$$F(x, y) = x^2 + y^2 - 1 = 0\,,$$

kann man für die Punkte, die nicht auf der $x$-Achse $y = 0$ liegen, zumindest lokal eindeutig explizit auflösen durch

$$y = \pm\sqrt{1 - x^2}\,,$$

wobei das Vorzeichen gleich dem von $y$ im betrachteten Intervall ist.

Im allgemeinen gibt es jedoch keine Möglichkeit für eine explizite Auflösung mit den „üblichen" mathematischen Funktionen, d.h. man kann höchstens dann auflösen, wenn man neue Funktionen einführt.

Wie das Beispiel der Kreislinie zeigt, ist auch das nicht immer möglich: Für die beiden Punkte auf der $x$-Achse gibt es offensichtlich keine *eindeutige* Auflösung, da sowohl die positive wie auch die negative Wurzel Teilauflösungen sind.

Diese Existenz mehrerer Teilauflösungen hängt mit dem Verschwinden der partiellen Ableitung nach $y$ zusammen: Falls diese partielle Ableitung ungleich Null ist, gibt sie an, wie sich $F$ verändert, wenn man $y$ ändert, und sie gibt damit zumindest in erster Näherung auch an, wie man $y$ verändern muß, um bei einer Änderung von $x$ die Bedingung $F(x, y) = 0$ zu erhalten. Der Satz über implizite Funktionen besagt, das dieses Nichtverschwinden der partiellen Ableitung bereits ausreicht um die Existenz einer eindeutigen Auflösung zu zeigen.

Um den Beweis wenigstens einigermaßen überschaubar zu halten, möchte ich mich zunächst auf Funktionen zweier Veränderlicher beschränken:



**Satz:** $D \subseteq \mathbb{R}^2$ sei offen und $F \in \mathcal{C}^1(D, \mathbb{R})$. Dann gibt es für jeden Punkt $(x_0, y_0) \in D$ mit $F(x_0, y_0) = 0$ und $F_y(x_0, y_0) \neq 0$ Intervallumgebungen $I$ von $x_0$ und $K$ von $y_0$ sowie eine eindeutig bestimmte Funktion $g\colon I \to K$, so daß für alle $x \in I$ gilt:

$$F\big(x, f(x)\big) = 0\,.$$

Die Funktion $f$ ist stetig und differenzierbar; ihre Ableitung ist

$$f'(x) = -\frac{F_x(x, y)}{F_y(x, y)} \quad \text{mit} \quad y = f(x)\,.$$

*Beweis:* Wir beginnen mit einer Reduktion zwecks Vereinfachung der Schreibarbeit: Offensichtlich genügt es, wenn wir den Fall $x_0 = y_0 = 0$ betrachten. Gilt nämlich der Satz für die Funktion

$$G(x, y) = F(x + x_0, y + y_0)$$

im Punkt $(0, 0)$, so folgt er sofort auch für $F$ im Punkt $(x_0, y_0)$. Außerdem können wir o.B.d.A. annehmen, daß $F_y(0, 0)$ positiv ist, denn nach Voraussetzung ist dieser Wert ungleich Null, und falls er negativ sein sollte, ersetzen wir einfach $F$ durch $-F$.

Nach Voraussetzung sind die partiellen Ableitungen von $F$ stetig; daher ist $F_y$ nicht nur im Nullpunkt positiv, sondern auch noch in einer gewissen Umgebung davon. In dieser Umgebung wählen wir ein Rechteck

$$\big\{(x, y) \in \mathbb{R}^2 \;\big|\; -\alpha \leq x \leq \alpha \quad \text{und} \quad -\beta \leq y \leq \beta\big\}\,.$$

Da $F_y$ auf diesem Rechteck überall positiv ist, wächst die Funktion $y \mapsto F(x_0, y)$ für jedes $x_0 \in [-\alpha, \alpha]$ streng monoton; wegen $F(0, 0) = 0$ ist insbesondere $F(0, -\beta) < 0$ und $F(0, \beta) > 0$. Aufgrund der Stetigkeit von $F$ ist damit für $x_1$ aus einer gewissen Umgebung der Null auch $F(x_1, -\beta) < 0$ und $F(x_1, \beta) > 0$. Indem wir nötigenfalls $\alpha$ noch etwas verkleinern, können wir annehmen, daß dies für alle $x_1 \in [-\alpha, \alpha]$ gilt.

Damit gibt es nach dem Zwischenwertsatz für jedes $x_1 \in [-\alpha, \alpha]$ ein $y_1$, so daß $f(x_1, y_1)$ verschwindet; wegen der strengen Monotonie der Funktion $y \mapsto f(x_1, y)$ ist dieser Wert $y_1$ eindeutig bestimmt. Wir setzen daher $I = (-\alpha, \alpha)$, $K = (-\beta, \beta)$ und

$$f\colon \begin{cases} I & \to K \\ x_1 & \mapsto y_1 \end{cases}\,.$$

Damit ist $f$ als Funktion festgelegt, und nach Konstruktion ist

$$F\big(x, f(x)\big) = 0 \quad \text{für alle } x \in I\,.$$

Wir müssen uns noch überlegen, daß die so konstruierte Funktion $f$ stetig und differenzierbar ist.

Die Stetigkeit können wir etwa dadurch nachweisen, daß wir für jede gegen ein $x \in I$ konvergierende Folge $(x_n)$ aus $I$ zeigen, daß

$$f\big(\lim_{n\to\infty} x_n\big) = \lim_{n\to\infty} f(x_n)$$

ist.

Sei $y = \lim\limits_{n\to\infty} f(x_n)$. Wegen der vorausgesetzten Stetigkeit von $F$ ist zunächst

$$\lim_{n\to\infty} F\big(x_n, f(x_n)\big) = F\big(\lim_{n\to\infty} x_n, \lim_{n\to\infty} f(x_n)\big) = F(x,y)\,.$$

Da die Zahlen $F\big(x_n, f(x_n)\big)$ allesamt Null sind, verschwindet auch ihr Grenzwert, also $F(x,y)$, und damit ist $y = \lim_{n\to\infty} f(x_n) = f(x)$ der eindeutig bestimmte Wert, für den $F(x,y)$ verschwindet. Also muß $y$ gleich $f(x)$ sein, wie gewünscht.

Somit ist $f$ stetig; als nächstes müssen wir noch die Differenzierbarkeit zeigen. Für ein $x \in I$ und ein hinreichend kleines $h \in \mathbb{R}$, für das auch noch $x+h$ in $I$ liegt, ist nach Definition von $f$

$$F\big(x+h, f(x+h)\big) = 0\,.$$

Andererseits können wir diesen Funktionswert auch nach dem Mittelwertsatz berechnen: Mit $k = f(x+h) - f(x)$ ist

$$F\big(x+h, f(x+h)\big) = F\big(x+h, f(x)+k\big) = F\Big(\big(x, f(x)\big) + \binom{h}{k}\Big)\,,$$

und setzen wir

$$\varphi(t) \underset{\text{def}}{=} F\Big(\big(x, f(x)\big) + t\binom{h}{k}\Big)\,,$$

so ist nach dem Mittelwertsatz der Differentialrechnung

$$\varphi(1) = \varphi(0) + \dot\varphi(\tau)$$

für ein $\tau$ zwischen null und eins. Mit $\xi = x + \tau h$ und $\eta = f(x) + \tau k$ ist daher

$$F\big(x+h, f(x)+k\big) = F\big(x, f(x)\big) + hF_x(\xi,\eta) + kF_y(\xi,\eta)\,.$$



Da $F\big(x+h, f(x)+k\big)$ und $F\big(x, f(x)\big)$ beide verschwinden, folgt

$$\frac{f(x+h)-f(x)}{h} = \frac{k}{h} = -\frac{F_x(\xi,\eta)}{F_y(\xi,\eta)}\,.$$

Für $h \to 0$ geht die rechte Seite wegen der Stetigkeit der partiellen Ableitungen gegen $-F_x(x,y)/F_y(x,y)$, insbesondere existiert also der Grenzwert. (Man beachte, daß hier nochmals die Voraussetzung $F_y \neq 0$ benötigt wird). Damit existiert auch der Grenzwert des linksstehenden Differenzenquotienten für $h \to 0$, d.h. $f$ ist differenzierbar und hat die behauptete Ableitung.

■

Nachdem wir wissen, daß die Ableitung von $f$ existiert, ist ihre Berechnung, unabhängig vom gerade bewiesenen Satz, eine einfache Übungsaufgabe: Die Funktion $F\big(x, f(x)\big)$ ist gleich der Nullfunktion, und damit verschwindet natürlich auch ihre Ableitung. Andererseits ist diese Ableitung nach der Kettenregel gleich

$$F_x\big(x, f(x)\big) + F_y\big(x, f(x)\big)\cdot f'(x)\,,$$

also folgt

$$f'(x) = -\frac{F_x\big(x, f(x)\big)}{F_y\big(x, f(x)\big)}\,.$$

Genauso kann auch die zweite Ableitung von $f$ berechnet werden –falls sie existiert. Wenn $F$ und $f$ zweimal stetig differenzierbar sind, können wir $F\big(x, f(x)\big)$ zweimal ableiten, was natürlich immer noch null ist. Nach der Kettenregel ist aber die zweite Ableitung von $F\big(x.f(x)\big)$ (der Übersichtlichkeit halber jeweils ohne das Argument $\big(x, f(x)\big)$ geschrieben) gleich

$$\begin{aligned}&\frac{\partial}{\partial x}\big(F_x + F_y\cdot f'(x)\big) + \frac{\partial}{\partial y}\big(F_x + F_y\cdot f'(x)\big)\cdot f'(x)\\ &= F_{xx} + F_{yx}\cdot f'(x) + F_y\cdot f''(x) + \big(F_{xy} + F_{yy}\cdot f'(x)\big)\cdot f'(x)\,,\end{aligned}$$

d.h.

$$\begin{aligned}f''(x) &= -\frac{1}{F_y}\big(F_{xx} + 2F_{xy}\cdot f'(x) + F_{yy}\cdot f'(x)^2\big)\\ &= -\frac{1}{F_y^3}\big(F_{xx}^2F_y^2 - 2F_{xy}F_xF_y + F_{yy}F_x^2\big)\,.\end{aligned}$$

Für Extremwertbetrachtungen bei implizit definierten Funktionen braucht man die zweite Ableitung in den Punkten, in denen die erste verschwindet; dort vereinfacht sich die Formel zu

$$f''(x) = -\frac{F_{xx}\big(x, f(x)\big)}{F_y\big(x, f(x)\big)} \quad \text{falls} \quad f'(x) = 0\,.$$

Genau wie Funktionen einer Veränderlichen können auch Funktionen mehrerer Veränderlicher implizit definiert sein; die entsprechende Verallgemeinerung des Satzes über implizite Funktionen folgt fast vollständig aus der koordinatenweisen Anwendung des obigen Satzes, lediglich für die Stetigkeit der Funktion muß man das entsprechende Argument aus dem gerade beendeten Beweis noch einmal anwenden. Die Aussage ist

**Satz:** $D \subseteq \mathbb{R}^n$ sei eine offene Menge, $F \in \mathcal{C}^1(D, \mathbb{R})$ eine stetig differenzierbare Funktion auf $D$, und $\boldsymbol{a} = (a_1, \ldots, a_n) \in D$ sei ein Punkt mit $f(\boldsymbol{a}) = 0$. Falls $f_{x_n}(\boldsymbol{a}) \neq 0$ ist, gibt es ein eine offene Umgebung $U$ von $(a_1, \ldots, a_{n-1})$ in $\mathbb{R}^{n-1}$ und eine Funktion $f \in \mathcal{C}^1(U, \mathbb{R})$, so daß für alle Punkte $\boldsymbol{y} = (y_1, \ldots, y_{n-1}) \in U$ gilt:

$$F\big(\boldsymbol{y}, f(\boldsymbol{y})\big) = 0\,.$$

Für alle $i \leq n-1$ ist

$$f_{x_i}(\boldsymbol{y}) = -\frac{F_{x_i}\big(\boldsymbol{y}, f(\boldsymbol{y})\big)}{F_{x_n}\big(\boldsymbol{y}, f(\boldsymbol{y})\big)}\,.$$

■

## §3: Maxima und Minima unter Nebenbedingungen

Bei einem realen physikalischen oder technischen Prozeß können sich die Variablen selten frei im gesamten $\mathbb{R}^n$ bewegen: Physikalisch sinnvoll ist meist nur eine beschränkte Teilmenge. Im Gegensatz zur Dimension eins, wo diese Teilmenge praktisch immer ein Intervall ist, gibt es aber im Mehrdimensionalen keinen Grund, warum diese Teilmenge offen oder zumindest der Abschluß einer offenen Teilmenge sein sollte:



Im $\mathbb{R}^3$ kann man sich beispielsweise auch interessieren für das Maximum oder Minimum der Ladungsdichte auf einer Kugeloberfläche oder die elektrische Feldstärke oder Temperaturverteilung auf der Innenhaut eines Reaktordruckbehälters.

Diese Maxima oder Minima sind im allgemeinen keine lokalen Maxima oder Minima der betrachteten Funktion: Wenn man die jeweilige Fläche verläßt, läßt sich der Funktionswert selbst für einen solchen Extremwert meist noch – je nach Richtung – sowohl vergrößern als auch verkleinern. Dementsprechend können die Methoden, die wir am Ende von §1 diskutiert haben, solche Extremwerte üblicherweise nicht finden; wir brauchen weitere Werkzeuge, die in diesem Paragraphen bereitgestellt werden sollen.

Die Situation, um die es hier geht, ist typischerweise die folgende: Gegeben ist eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$, möglicherweise auch nur auf einer Teilmenge $D \subset \mathbb{R}^n$ definiert, deren Extremwerte nicht auf $\mathbb{R}^n$ oder $D$ gesucht werden, sondern nur auf einer Teilmenge, die beispielsweise durch das Verschwinden einer weiteren Funktion $g\colon \mathbb{R}^n \to \mathbb{R}$ gegeben ist. Falls wir uns für Extremwerte auf einer Kugel vom Radius $r$ um den Nullpunkt interessieren, wäre dies etwa die Funktion

$$g\colon \begin{cases} \mathbb{R}^3 & \to \mathbb{R} \\ (x, y, z) & \mapsto x^2 + y^2 + z^2 - r^2 \end{cases}.$$

Eine mögliche Strategie zur Lösung solcher Probleme besteht darin, die Gleichung $g = 0$ nach einer der Variablen aufzulösen, diese dann in $f$ einzusetzen und sodann eine gewöhnliche Extremwertaufgabe zu lösen. Diese Auflösung ist *explizit* nur in sehr einfachen Fällen möglich, aber selbst wenn wir nur wissen, daß eine solche Auflösung *existiert,* können wir doch damit argumentieren und Kriterien ableiten.

Unter Maxima und Minima sollen hier *lokale* Extrema verstanden werden, so daß wir die üblichen Kriterien mit Ableitungen anwenden können:

**Definition:** Wir sagen, die Funktion $f\colon D \to \mathbb{R}$ auf einer Teilmenge $D \subseteq \mathbb{R}^n$ habe im Punkt $\boldsymbol{a} \in D$ ein lokales $\left\{\begin{matrix}\text{Maximum}\\ \text{Minimum}\end{matrix}\right\}$ unter der

Nebenbedinung $g = 0$, wobei $g\colon D \to \mathbb{R}$ eine weitere Funktion ist, wenn $g(\boldsymbol{a}) = 0$ ist und es eine Umgebung $U$ von $\boldsymbol{a}$ gibt, so daß für alle $\boldsymbol{x} \in U$ gilt: Ist $g(\boldsymbol{x}) = 0$, so ist $f(\boldsymbol{x}) \begin{Bmatrix} \le \\ \ge \end{Bmatrix} f(\boldsymbol{a})$.

Als Einstiegsbeispiel betrachten wir eine beliebte Schulbuchaufgabe zur Minimumsbestimmung: Eine Konservendose soll bei einem vorgegebenen Volumen von 100 $cm^3$ möglichst wenig Blech benötigen, d.h. ihre Oberfläche soll minimal sein.

Die Oberfläche eines Zylinders der Höhe $h$ mit einer Grundfläche vom Radius $r$ ist

$$f(r, h) = 2\pi r^2 + 2\pi r \cdot h\,;$$

die Nebenbedingung für das Volumen $V = \pi r^2 h$ besagt, daß

$$g(r, h) = \pi r^2 h - 100 = 0$$

sein soll.

Hier läßt sich natürlich die Nebenbedingung sofort nach $h$ auflösen:

$$h = \frac{100}{\pi r^2}\,,$$

und wir müssen nur noch die Funktion

$$F(r) = f\left(r, \frac{100}{\pi r^2}\right) = 2\pi r^2 + \frac{200}{r}$$

minimieren. Für diese ist

$$F'(r) = 4\pi r - \frac{200}{r^2}\,,$$

und dies verschwindet genau dann, wenn

$$4\pi r^3 = 200 \quad \text{oder} \quad r = \sqrt[3]{\frac{50}{\pi}}$$

ist.

In diesem einfachen Fall kann man solche Aufgaben also zurückführen auf gewöhnliche Extremwertaufgaben, indem man die Nebenbedingung nach einer der Variablen auflöst und diese dann in $f$ einsetzt; in anderen



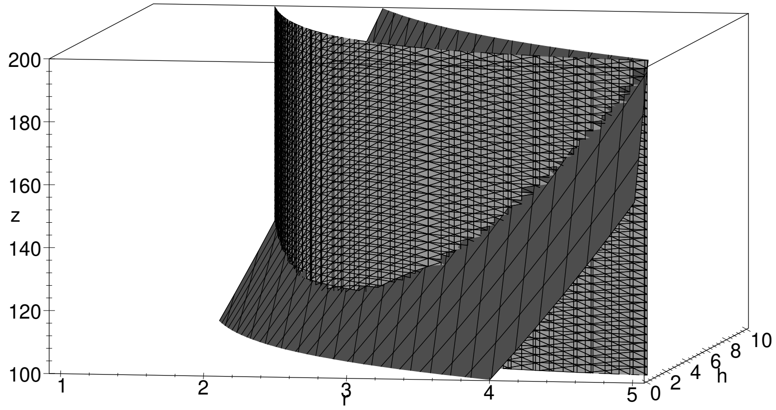


Abb. 43: Oberfläche einer Konservendose mit festem Volumen

Fällen kann man gelegentlich die Nebendedingung durch geeignete Parameterwahl oder Wahl eines angepaßten Koordinatensystems berücksichtigen. Im allgemeinen wird aber beides nicht möglich sein, so daß wir andere Methoden brauchen.

Unser bisherige Theorie für lokale Extrema ist in dieser Situation nicht anwendbar, denn die lokalen Extrema von $f$ werden nur in den seltensten Fällen die Nebenbedingung $g = 0$ erfüllen; im obigen Beispiel zeigt Abbildung 43 die Nebenbedingung als eng schraffierte Fläche dargestellt und der Graph von $f$ als weiter schraffierte; wie man sieht, läßt sich der Wert von $f$ problemlos verkleinern, wenn man nur die Fläche $g = 0$ verläßt, und in der Tat ist auch ohne jede Mathematik sofort klar, daß man mit weniger Blech auskommt, wenn man die Konservendose einfach schmäler oder kürzer macht.

Die Grundidee für ein alternatives Verfahren wird klar bei der Betrachtung der Niveaulinien in Abbildung 44: Die Niveaulinie für $g = 0$ ist gestrichelt eingezeichnet, verschiedene Niveaulinien von $f$ als durchgezogene Kurven.

Wie man sieht, schneiden einige dieser Niveaulinien die gestrichelte Kurve überhaupt nicht: Wenn man zu wenig Blech hat, kann man keine Dose mit 100 $cm^3$ Inhalt zusammenlöten. Wenn es dagegen genug Blech

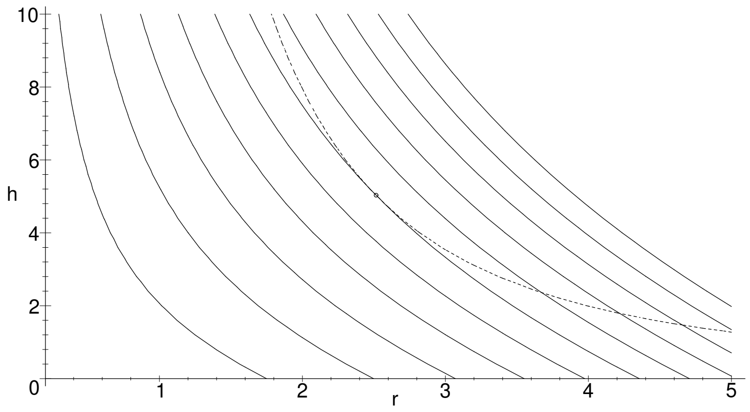


Abb. 44: Niveaulinien für Oberfläche und Volumen

gibt, gibt es gleich zwei Schnittpunkte: Die Dose kann entweder eher höher oder eher breiter gemacht werden. In einem solchen Fall kann man die Niveaulinie durch eine zu einem etwas niedrigeren Niveau ersetzen, die im allgemeinen auch wieder Schnittpunkte haben wird, so daß das Niveau noch nicht minimal sein kann. Erst wenn man im Minimum ist, fallen die beiden Schnittpunkte zusammen; wenn man nun das Niveau noch weiter erniedrigt, gibt es keine Schnittpunkte mehr.

Da somit im Minimum zwei Schnittpunkte zusammenfallen, berühren sich dort die Niveaulinien von $f$ und von $g$, d.h. sie haben eine gemeinsame Tangente. Da der Gradient, wie wir wissen, senkrecht auf der Tangenten der Niveaulinien steht (die Richtungsableitung entlang einer Niveaulinie ist schließlich null), sind somit die Gradienten von $f$ und $g$ im Minimum zueinander parallel, d.h. der eine ist ein Vielfaches des anderen.

Dies gilt nicht nur im vorliegenden Beispiel, sondern allgemein:

**Satz:** $D \subseteq \mathbb{R}^n$ sei eine offene Menge und $f, g \in \mathcal{C}^1(D, \mathbb{R})$ seien stetig differenzierbare Funktionen auf $D$. Falls $f$ im Punkt $\boldsymbol{a} \in D$ ein Extremum hat unter der Nebenbedinung $g(\boldsymbol{x}) = 0$, so sind grad $f(\boldsymbol{a})$ und grad $g(\boldsymbol{a})$ linear abhängig.



*Beweis:* Die Grundidee ist einfach: Auch wenn wir die Nebenbedingung nicht *explizit* nach einer der Variablen auflösen können, sagt uns der Satz über implizite Funktionen in vielen Fällen dennoch, daß zumindest lokal eine Auflösung existiert. Diese Auflösung kennen wir zwar nicht, aber wir können mit ihr argumentieren und, zumindest formal, auch rechnen.

Falls grad $g(\boldsymbol{a})$ der Nullvektor ist, gibt es nichts mehr zu beweisen, denn jede Menge, die den Nullvektor enthält, ist linear abhängig.

Wir können daher annehmen, daß grad $g(\boldsymbol{a})$ mindestens eine von Null verschiedene Komponente hat, und durch Umnummerierung der Koordinaten können wir auch o.B.d.A. annehmen, daß die $n$-te Komponente nicht verschwindet, d.h. $g_{x_n}(\boldsymbol{a}) \neq 0$.

Dann gibt es nach dem Satz über implizite Funktionen eine Umgebung $U$ von $(a_1, \ldots, a_{n-1})$ und eine Funktion $h: U \to \mathbb{R}$ mit $h(a_1, \ldots, a_{n-1}) = a_n$, so daß

$$g\big(x_1, \ldots, x_{n-1}, h(x_1, \ldots, x_{n-1})\big) = 0 \quad \text{für alle } (x_1, \ldots, x_{n-1}) \in U\,.$$

Nachdem $f$ in $\boldsymbol{a}$ ein lokales Extremum unter der Nebenbedinung $g = 0$ hat, nimmt die Funktion

$$F(x_1, \ldots, x_{n-1}) \underset{\text{def}}{=} f\big(x_1, \ldots, x_{n-1}, h(x_1, \ldots, x_{n-1})\big)$$

in $(a_1, \ldots, a_{n-1})$ ein lokales Extremum im üblichen Sinne an, d.h. der Gradient von $F$ verschwindet dort.

Nach der Kettenregel ist für $i = 1, \ldots, n-1$

$$F_{x_i}(a_1, \ldots, a_{n-1}) = f_{x_i}(\boldsymbol{a}) + f_{x_n}(\boldsymbol{a}) \cdot h_{x_i}(a_1, \ldots, a_{n-1})\,,$$

und nach dem Satz über implizite Funktionen ist $h_{x_i} = -g_{x_i}/g_{x_n}$, d.h.

$$F_{x_i}(a_1, \ldots, a_{n-1}) = f_{x_i}(\boldsymbol{a}) - f_{x_n}(\boldsymbol{a}) \frac{g_{x_i}(\boldsymbol{a})}{g_{x_n}(\boldsymbol{a})}\,.$$

Da die linke Seite verschwindet, gilt dasselbe auch für die rechte. Die rechte Seite ist im Gegensatz zur linken auch für $i = n$ definiert und verschwindet aus trivialen Gründen; also ist für alle $i$

$$f_{x_i}(\boldsymbol{a}) - \frac{f_{x_n}(\boldsymbol{a})}{g_{x_n}(\boldsymbol{a})} g_{x_i}(\boldsymbol{a}) = 0$$