

Kapitel 5

Funktionen mehrerer Veränderlicher

Bislang hatten wir nur Funktionen betrachtet, die von einer einzigen Variablen abhängen. Für die meisten Anwendungen ist das zu viel zu speziell: Egal ob wir ein wirtschaftliches, soziales oder technisches System beschreiben wollen, können wir sicher sein, daß es von einer Vielzahl verschiedener Größen abhängt.

Wir könnten versuchen, dieses Problem zu umgehen, indem wir alle diese Größen mit einer Ausnahme konstant halten und die so entstehende Funktion einer Veränderlichen mit den uns bekannten Methoden untersuchen – dieser *ceteris paribus* Ansatz (*das Übrige ist gleich*) wird in der Tat bei manchen volkswirtschaftlichen Problemen gerne verwendet. Er ist aber sicherlich nicht allgemein einsetzbar, denn die eine variable Größe kann je nach Werten der festgehaltenen Variablen sehr unterschiedliche Effekte haben: Mehrproduktion kann beispielsweise ja nach Zustand des Marktes mal zu Gewinnen, mal zu Ladenhütern führen. Hinzu kommt, daß die Kenngrößen eines Systems selten unabhängig voneinander sind und daher Veränderungen einer Größe oft zwangsläufig zu Veränderungen weiterer Größen führen.

Trotz dieser Schwierigkeiten wollen wir die bewährten Methoden aus der Analysis einer Veränderlichen soweit wie möglich weiter benutzen; wie sich zeigen wird, ist das auch möglich, allerdings müssen sie durch zusätzliche Hilfsmittel ergänzt werden.

Ein wesentliches solches Werkzeug sind, wie schon bei Funktionen einer Veränderlichen, *lineare* Funktionen; teilweise werden wir hier im Mehrdimensionalen daher auch Methoden aus der Linearen Algebra brauchen.

§1: Grundlegende Eigenschaften

Bevor wir zur Differential- und Integralrechnung im \mathbb{R}^n kommen, brauchen wir – wie schon im eindimensionalen Fall – einige Vorbereitungen über Konvergenz, Stetigkeit und ähnliche Grundbegriffe. Diese sollen hier zusammengestellt werden.

a) Visualisierung in höheren Dimensionen

Um einen ersten Eindruck von einer Funktion $f: D \rightarrow \mathbb{R}$ auf einer offenen Teilmenge $D \subseteq \mathbb{R}^n$ zu gewinnen, startet man am besten mit einem Bild. Dieses kann zwar nur einen endlichen Ausschnitt des Definitionsbereichs zeigen und ist auch in seiner Genauigkeit begrenzt, bietet aber doch Information, die man anders nur mit erheblich größerem Aufwand darstellen könnte. Als Bild von f betrachten wir üblicherweise den Graphen

$$\Gamma_f \stackrel{\text{def}}{=} \{(x, y) \in \mathbb{R}^2 \mid y = f(x)\},$$

also eine Teilmenge der Ebenen \mathbb{R}^2 , die wir uns sehr gut vorstellen können.

Auch für eine Funktion $f: D \rightarrow \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ können wir deren Graphen

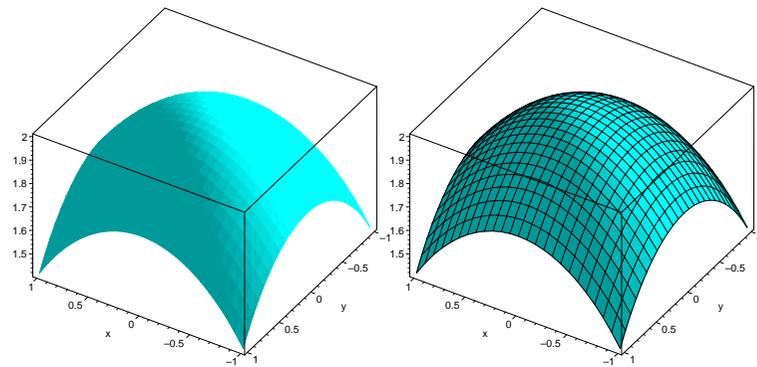
$$\Gamma_f \stackrel{\text{def}}{=} \{(x, y) \in D \times \mathbb{R}^m \mid y = f(x)\}$$

definieren; da dieser in $\mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$ liegt, ist er allerdings nur für $n + m \leq 3$ wirklich anschaulich, wobei es im Fall $n + m = 3$ bei komplizierteren Funktionen stark von der gewählten Perspektive abhängen kann, wieviel man wirklich sieht. Für einfache reellwertige Funktionen zweier Veränderlicher jedoch ist der Graph sicherlich die beste Methode zur Veranschaulichung, wobei man gegebenenfalls zur besseren Übersicht noch Hilfslinien für die Funktionswerte zu ausgewählten Werten von x und/oder y einzeichnen kann.

Beim Graphen der Funktion

$$f: \begin{cases} [-1, 1] \times [-1, 1] & \rightarrow \mathbb{R} \\ (x, y) & \mapsto \sqrt{4 - x^2 - y^2} \end{cases}$$

in der Abbildung unten etwa sieht man bei beiden Darstellungen recht gut, daß Γ_f Teil einer Kugeloberfläche ist.

Graph der Funktion $f(x, y) = \sqrt{4 - x^2 - y^2}$

Eine andere Möglichkeit zur Veranschaulichung von Funktionen zweier Veränderlicher ist von topographischen Karten her bekannt: Dort wird die Höhe über dem Meeresspiegel, eine Funktion der beiden Ebenenkoordinaten, dargestellt durch *Höhenlinien*. Entsprechend können wir für eine beliebige Funktion $f: D \rightarrow \mathbb{R}$ mit $D \subseteq \mathbb{R}^2$ und jeden Wert $c \in \mathbb{R}$ die *Niveaulinie*

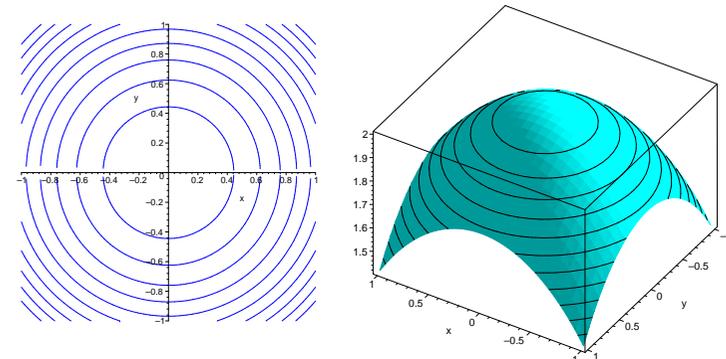
$$N_c(f) \stackrel{\text{def}}{=} \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$$

definieren; sie muß natürlich keine „Linie“ sein, sondern kann auch nur aus einigen Punkten bestehen, leer sein oder – im Falle einer konstanten Funktion – für einen bestimmten Wert c aus dem gesamten Definitionsbereich D bestehen.

Im Falle des obigen Beispiels etwa ist $N_c(f)$ für $c > 2$ und für $c < \sqrt{2}$ die leere Menge; für $c = 2$ besteht sie nur aus dem Nullpunkt, und für $c = \sqrt{2}$ aus den vier Punkten $(0, \pm 1)$ und $(\pm 1, 0)$. Für $\sqrt{2} < c < 2$ erhalten wir die in der nächsten Abbildung für $c = 1,5$ bis $c = 2$ in Schritten von 0,05 dargestellten Kreislinien

$$\sqrt{4 - x^2 - y^2} = c \quad \text{oder} \quad x^2 + y^2 = 4 - c^2,$$

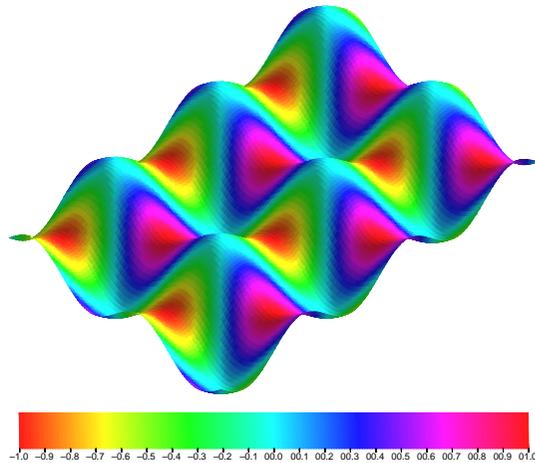
eingeschränkt natürlich auf das Einheitsquadrat als dem Definitionsbereich von f . Die Darstellung von Niveaulinien kann auch kombiniert werden mit der des Graphen.

Niveaulinien der Funktion $f(x, y) = \sqrt{4 - x^2 - y^2}$

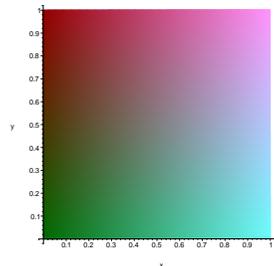
Für Funktionen von mehr als zwei Veränderlichen ist die Visualisierung naturgemäß schwieriger; wir können Graphen und auch Niveaulinien, -flächen usw. zwar problemlos definieren, aber nicht mehr zeichnen – es sei denn, es handelt sich um sehr einfache Niveaulinien im \mathbb{R}^3 . Bei Funktionen mit Werten in einem mehrdimensionalen Raum kommt hinzu, daß die Niveaumengen dann nicht mehr nur von einem, sondern von mehreren Parametern abhängen.

Eine weitere Möglichkeit besteht darin, auf einem zwei- oder dreidimensionalen Graphen durch Farbe, Textur usw. weitere Dimensionen darzustellen; allgemein bekannt ist die Kodierung der Höhe durch von Grün nach Braun laufende Farben in Atlanten oder auch die Darstellung der Temperatur durch Farbverläufe von Blau über Rot nach Weiß, usw.

Wollen wir beispielsweise für $0 \leq x \leq 3\pi$ und $0 \leq y \leq 2\pi$ jene Funktion von \mathbb{R}^2 nach \mathbb{R}^2 veranschaulichen, die einem Punkt (x, y) die beiden Werte $f(x, y) = \sin(x+y) \cos(x-y)$ und $g(x, y) = \cos(x+y) \sin(x-y)$ zuordnet, so können wir den Graphen von f zeichnen und darauf nach einem Farbschema die Funktion g kodieren. Da diese nur Werte zwischen -1 und 1 annimmt, müssen wir dazu einfach eine Funktion festlegen, die jeder dieser Zahlen eine Farbe zuordnet; der entsprechende Farbverlauf ist unter dem Bild abgedruckt. Wenn man genau hinschaut, gewinnt man so einen recht guten Eindruck vom relativen Verlauf der beiden Funktionen.

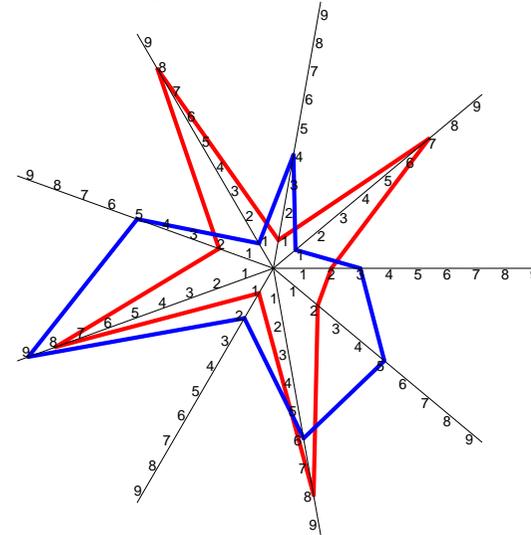


Grundsätzlich kann man mit Farben auch mehr als eine Dimension darstellen: Da wir drei Arten von Sehzäpfchen haben, könnten wir theoretisch bis zu drei Dimensionen darstellen. Tatsächlich sind jedoch nur die wenigsten Menschen in der Lage, einer Farbe deren Rot- Grün- und Blauwerte anzusehen, so daß die Darstellung von drei Dimensionen durch Farbwerte selten nützlich ist. Zwei Dimensionen sind aber durchaus realistisch, vor allem, wenn wir die eine Dimension auf die Helligkeit abbilden und die andere auf einen der beiden Farbwerte in einem Luminanz/ Chromanz-Modell wie etwa dem fürs jpeg-Format verwendeten YCbCr-Modell. Beim Quadrat unten etwa sind die x -Werte durch die Helligkeit kodiert und die y -Werte durch die Chromanz für Rot.



Als letzte Alternative seien noch Stern- oder Netzdiagramme erwähnt: Hier geht es darum, einzelne Punkte in einem höherdimensionalen Raum zu veranschaulichen, z.B. den Vektor der Klausurnoten eines Studenten oder Kompetenzeinschätzungen für Politiker. Um einen Punkt $(x_1, \dots, x_n) \in \mathbb{R}^n$ darzustellen, zeichnet man n vom Nullpunkt ausgehende Strahlen im \mathbb{R}^2 , markiert auf dem i -ten dieser Strahlen den

Punkt x_i , und verbindet die so markierten Punkte zu einem n -Eck. Damit dies übersichtlich bleibt, sollte n hier nicht wesentlich größer als zehn sein und auch die Anzahl der in einem Bild darstellbaren Punkte sollte definitiv einstellig sein.



Die Punkte $(3, 1, 4, 1, 5, 9, 2, 6, 5)$ und $(2, 7, 1, 8, 2, 8, 1, 8, 2)$ aus \mathbb{R}^9

Ein eigenes Forschungsgebiet der Mathematik und Informatik, die Visualisierung, beschäftigt sich mit diesen und weiteren Methoden, die für eine vorgegebene Fragestellung interessanten Aspekte einer (analytisch oder empirisch gegebenen) Funktion mehrerer Veränderlichen graphisch herauszuarbeiten; für uns sollen aber zumindest vorerst Graphen und Niveaumengen genügen.

b) Normierte Vektorräume

Bei der Definition von Konvergenz und Stetigkeit im Eindimensionalen spielte die Betragsfunktion eine große Rolle; die in diesem Abschnitt eingeführten Normen sollen diese Funktion aufs Höherdimensionale verallgemeinern.

Genau wie die klassische Betragsfunktion jeder reellen (oder komplexen) Zahl eine nichtnegative reelle Zahl zuordnet, soll eine Norm jedem

Punkt des \mathbb{R}^n (oder \mathbb{C}^n) eine nichtnegative reelle Zahl zuordnen. Wenn wir den Punkt $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ mit dem Vektor mit denselben Komponenten identifizieren, können wir beispielsweise dessen Länge

$$\sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \dots + x_n^2}$$

nehmen, was oft in der Tat die natürlichste Wahl ist. Andererseits ist das Rechnen mit dieser Länge wegen der Wurzelfunktion gelegentlich recht unangenehm; deshalb wollen wir auch Alternativen betrachten.

Obwohl wir uns in diesem Semester praktisch nur für \mathbb{R}^n interessieren, möchte ich zumindest die Definition für einen beliebigen reellen oder komplexen Vektorraum angeben:

Definition: Ein normierter Vektorraum ist ein \mathbb{R} - oder \mathbb{C} -Vektorraum V zusammen mit einer Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ mit den Eigenschaften

- a) $\|\lambda v\| = |\lambda| \|v\|$ für alle $\lambda \in \mathbb{R}$ bzw. \mathbb{C} und $v \in V$
- b) $\|v\| \geq 0$ für alle $v \in V$, und $\|v\| = 0$ genau dann, wenn $v = 0$
- c) $\|v + w\| \leq \|v\| + \|w\|$ (*Dreiecksungleichung*)

Die Zahl $\|v\|$ wird als *Norm* des Vektors $v \in V$ bezeichnet.

Im Falle $V = \mathbb{R}$ oder auch $V = \mathbb{C}$ ist klar, daß $\|v\| = |v|$ alle drei Forderungen erfüllt; der Begriff der Norm verallgemeinert also in der Tat den des Betrags.

Ist $\langle \cdot, \cdot \rangle$ ein Skalarprodukt auf dem reellen Vektorraum V , kann durch

$$\|v\| \stackrel{\text{def}}{=} \sqrt{\langle v, v \rangle}$$

eine Norm auf V definiert werden: Abgesehen von c) sind alle Forderungen aus der obigen Definition klar; zum Beweis von c) müssen wir beachten, daß wegen der Bilinearität eines Skalarprodukts gilt

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle = \|v\|^2 + \|w\|^2 + 2 \langle v, w \rangle \quad \text{und} \\ (\|v\| + \|w\|)^2 &= \|v\|^2 + \|w\|^2 + 2 \|v\| \|w\|, \end{aligned}$$

so daß die Behauptung im Falle $\langle v, w \rangle \geq 0$ sofort aus der CAUCHY-SCHWARZschen Ungleichung folgt und für $\langle v, w \rangle < 0$ aus der Nichtnegativität der Norm.

Ausgehend vom Standardskalarprodukt

$$\langle v, w \rangle = \langle (v_1, \dots, v_n), (w_1, \dots, w_n) \rangle \stackrel{\text{def}}{=} v_1 w_1 + \dots + v_n w_n$$

auf \mathbb{R}^n erhalten wir so die EUKLIDISCHE Norm

$$\|v\| = \|(v_1, \dots, v_n)\| = \sqrt{v_1^2 + \dots + v_n^2},$$

mit der wir in dieser Vorlesung häufig arbeiten werden.

Der Hauptgrund dafür, daß wir auch noch andere Normen betrachten, liegt in der Unhandlichkeit des Umgangs mit der Wurzel. Die zweite für uns wichtige Norm auf \mathbb{R}^n , die *Maximumsnorm*, vermeidet dies. Sie ist definiert als

$$\|v\|_\infty = \|(v_1, \dots, v_n)\|_\infty \stackrel{\text{def}}{=} \max\{|v_1|, \dots, |v_n|\}.$$

Bedingung a) ist erfüllt, da $|\lambda v_i| = |\lambda| \cdot |v_i|$ für alle i , und auch mit b) gibt es keine Probleme. Für c) betrachten wir zwei Punkte

$$v = (v_1, \dots, v_n) \quad \text{und} \quad w = (w_1, \dots, w_n)$$

aus \mathbb{R}^n . Die Maximumsnorm der Summe

$$v + w = (v_1 + w_1, \dots, v_n + w_n)$$

ist nach Definition der größte Wert eines Betrags $|v_i + w_i|$; dieser werde etwa für den Index j angenommen, d.h. $\|v + w\|_\infty = |v_j + w_j|$.

Die Norm $\|v\|_\infty$ von v ist der größte Betrag eines v_i und damit mindestens gleich $|v_j|$; genauso ist $\|w\|_\infty \geq |w_j|$. Also ist

$$\|v + w\|_\infty = |v_j + w_j| \leq |v_j| + |w_j| \leq \|v\|_\infty + \|w\|_\infty,$$

wie verlangt.

Damit ist auch die Maximumsnorm tatsächlich eine Norm. Sie kann allerdings nicht über ein Skalarprodukt auf \mathbb{R}^n definiert werden: Gäbe es nämlich ein Skalarprodukt $\langle \cdot, \cdot \rangle$, für das $\|v\|_\infty = \sqrt{\langle v, v \rangle}$ wäre; so wäre etwa im \mathbb{R}^2 insbesondere

$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle = \left\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_\infty^2 = 1 \quad \text{und} \quad \left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = \left\| \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\|_\infty^2 = 1,$$

also

$$1 = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_\infty^2 = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 1 + 1 + 2 \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle$$

und somit $\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = -\frac{1}{2}$ und $\left\langle \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = -1$. Mithin wäre

$$4 = \left\| \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\|_{\infty}^2 = \left\langle \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 4 + 1 - 2 = 3,$$

ein offensichtlicher Widerspruch.

Maximumsnormen lassen sich nicht nur für \mathbb{R}^n (und analog \mathbb{C}^n) definieren, sondern auch für Funktionenräume: Wie wir aus Kapitel 2, §5 wissen, nimmt eine stetige Funktion $f: D \rightarrow \mathbb{R}$ auf einem Definitionsbereich, der das abgeschlossene Intervall $[a, b]$ enthält, ihr Maximum wirklich an, d.h die Abbildung

$$\|\cdot\|_{\infty}: \mathcal{C}^0([a, b], \mathbb{R}) \rightarrow \mathbb{R}; \quad f \mapsto \max_{x \in [a, b]} |f(x)|$$

ist wohldefiniert. Sie hat auch die Eigenschaften a) bis c): a) und b) sind, wie in den meisten Fällen, trivial, und sind $f, g \in \mathcal{C}^0([a, b], \mathbb{R})$ zwei Funktionen, so ist auch deren Summe $f + g$ stetig. Wenn sie ihr Betragsmaximum im Punkt $x^* \in [a, b]$ annimmt, haben wir die Abschätzung

$$\begin{aligned} \|f + g\|_{\infty} &= |(f + g)(x^*)| = |f(x^*) + g(x^*)| \leq |f(x^*)| + |g(x^*)| \\ &\leq \max_{x \in [a, b]} |f(x)| + \max_{x \in [a, b]} |g(x)| = \|f\|_{\infty} + \|g\|_{\infty}. \end{aligned}$$

Maximumsnormen spielen unter anderem in der Numerik eine wichtige Rolle, denn sie liefern Fehlerschranken für numerische Rechnungen:

Führen wir beispielsweise auf dem Vektorraum $\mathbb{R}^{n \times m}$ aller $n \times m$ -Matrizen die Maximumsnorm ein, so setzen wir natürlich

$$\|A\|_{\infty} = \max_{i=1}^n \max_{j=1}^m |a_{ij}|.$$

Betrachten wir auch \mathbb{R}^m und \mathbb{R}^n mit der Maximumsnorm, so erhalten wir für einen Vektor $v \in \mathbb{R}^m$ und dessen Produkt $b = Av$ mit einer $n \times m$ -Matrix A die Abschätzung

$$\|b\|_{\infty} \leq m \|A\|_{\infty} \cdot \|v\|_{\infty},$$

denn nach der Multiplikationsregel für Matrizen ist $b_i = \sum_{j=1}^m a_{ij} v_j$ und

damit

$$|b_i| \leq \sum_{j=1}^m |a_{ij}| \cdot |v_j| \leq \sum_{j=1}^m \|A\|_{\infty} \cdot \|v\|_{\infty} = m \|A\|_{\infty} \|v\|_{\infty}.$$

Wird also der Vektor v durch einen Fehlervektor ϵ gestört, so ist

$$A(v + \epsilon) = Av + A\epsilon = b + A\epsilon$$

mit einem Fehler behaftet, dessen Komponenten *mit Sicherheit* kleiner sind als $m \|A\|_{\infty} \cdot \|\epsilon\|_{\infty}$. Entsprechend ändert sich bei einem eindeutig lösbar linearen Gleichungssystem $Ax = b$ die Lösung $x = A^{-1}b$ höchstens um $n \|A^{-1}\|_{\infty} \cdot \|\epsilon\|_{\infty}$, wenn die rechte Seite durch einen Vektor ϵ gestört wird. (Tatsächlich wird diese Schranke in den meisten Fällen viel zu pessimistisch sein, aber realistische Schranken sind in der Numerik oft – wenn überhaupt – nur mit sehr großem Aufwand zu finden.)

Wir haben Normen eingeführt als Verallgemeinerungen der Betragsfunktion, um damit auch Begriffe wie Konvergenz und Stetigkeit auf Mehrdimensionale zu übertragen. Im Falle der Konvergenz führt dies zur

Definition: $(V, \|\cdot\|)$ sei ein normierter \mathbb{R} -Vektorraum. Eine Folge v_1, v_2, \dots von Vektoren aus V konvergiert gegen den Vektor $v \in V$, wenn es zu jedem $\epsilon > 0$ eine natürliche Zahl $N \in \mathbb{N}$ gibt, so daß $\|v - v_n\| < \epsilon$ für alle $n \geq N$.

Da es bei der Konvergenz nur darum geht, daß die Folgenglieder dem Grenzwert beliebig nahe kommen, ist klar, daß sich an der Konvergenz einer Folge nichts ändert, wenn wir die verwendete Norm durch ein positives Vielfaches ersetzen. Allgemeiner definieren wir

Definition: Zwei Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ auf einen Vektorraum V heißen äquivalent, wenn es reelle Konstanten $c_1, c_2 > 0$ gibt, so daß

$$c_1 \|v\|_1 \leq \|v\|_2 \leq c_2 \|v\|_1.$$

Offensichtlich ist dann auch

$$\frac{1}{c_2} \|v\|_2 \leq \|v\|_1 \leq \frac{1}{c_1} \|v\|_2,$$

die Äquivalenz ist also, wie es sein muß, symmetrisch.

Beispielsweise sind auf jedem \mathbb{R}^n die EUKLIDISCHE Norm $\|\cdot\|$ und die Maximumsnorm $\|\cdot\|_\infty$ äquivalent, denn

$$\|v\| = \sqrt{v_1^2 + \dots + v_n^2} \leq \sqrt{n} \|v\|_\infty = \sqrt{n} \|v\|_\infty$$

und

$$\|v\|_\infty = \sqrt{\|v\|_\infty^2} \leq \sqrt{v_1^2 + \dots + v_n^2} = \|v\|,$$

d.h.

$$\|v\|_\infty \leq \|v\| \leq \sqrt{n} \|v\|_\infty.$$

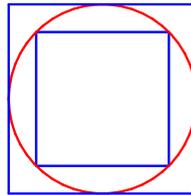
Anschaulich bedeutet dies, daß jeder Würfel in eine Kugel eingebettet werden kann und umgekehrt, denn

$$\{x \in \mathbb{R}^n \mid \|x\|_\infty \leq a\}$$

ist ein Würfel mit Kantenlänge $2a$ und

$$\{x \in \mathbb{R}^n \mid \|x\| \leq r\}$$

eine Kugel mit Radius r .



Für den Nachweis der Konvergenz einer Folge kann mal die eine, mal die andere dieser Normen besser geeignet sein. Zum Glück haben wir die freie Auswahl:

Lemma: $\|\cdot\|_1$ und $\|\cdot\|_2$ seien zwei äquivalente Normen auf dem Vektorraum V . Eine Folge $(v_n)_{n \in \mathbb{N}}$ von Vektoren aus V konvergiert genau dann bezüglich $\|\cdot\|_1$ gegen $v \in V$, wenn sie bezüglich $\|\cdot\|_2$ gegen v konvergiert.

Beweis: Da die Normen äquivalent sind, gibt es positive reelle Zahlen c_1, c_2 , so daß $c_1 \|v\|_1 \leq \|v\|_2 \leq c_2 \|v\|_1$ ist für alle $v \in V$. Falls die Folge $(v_n)_{n \in \mathbb{N}}$ bezüglich $\|\cdot\|_1$ gegen $v \in V$ konvergiert, gibt es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$, so daß $\|v - v_n\|_1 < \varepsilon$ ist für alle $n \geq N$. Also gibt es bei vorgegebenem $\varepsilon > 0$ auch ein $M \in \mathbb{N}$, so daß $\|v - v_n\|_1 < \varepsilon/c_2$ für alle $n \geq M$. Für $n \geq M$ ist dann

$$\|v - v_n\|_2 \leq c_2 \|v - v_n\|_1 \leq c_2 \frac{\varepsilon}{c_2} = \varepsilon,$$

die Folge konvergiert also auch bezüglich $\|\cdot\|_2$ gegen v .

Da $\|v\|_1 \leq \|v\|_2/c_1$ ist für alle $v \in V$, folgt ganz entsprechend, daß jede bezüglich $\|\cdot\|_2$ konvergente Folge auch bezüglich $\|\cdot\|_1$ gegen denselben Grenzwert konvergiert. ■

Man kann zeigen, daß in \mathbb{R}^n alle Normen äquivalent sind, so daß es dort also nur einen Konvergenzbegriff gibt. Wir werden in \mathbb{R}^n praktisch immer mit der EUKLIDISCHE Norm oder der Maximumsnorm arbeiten, deren Äquivalenz wir gezeigt haben; daher können wir immer, wenn von Konvergenz die Rede ist, frei wählen, mit welcher der beiden Normen wir arbeiten möchten. Je nach Anwendung kann

b) Stetigkeit

Stetigkeit bedeutet anschaulich, daß kleine Änderungen der Argumente einer Funktion auch nur zu kleinen Änderungen der Funktionswerte führen. Um Spielraum für die Variation der Argumente zu haben, definierten wir Stetigkeit für Funktionen einer Veränderlichen nur für Funktionen, die auf offenen Intervallen definiert sind. Im Falle mehrerer Veränderlicher wollen wir ähnlich vorgehen; wir brauchen daher als erstes eine mehrdimensionale Entsprechung für offene Intervalle.

Die für uns wesentliche Eigenschaft eines offenen Intervalls (a, b) war, daß es dort für jeden Punkt $x \in (a, b)$ sowohl links als auch rechts von x weitere Punkte aus dem Intervall gibt: Ist ε das Minimum der beiden (positiven) Werte $x - a$ und $b - x$, so liegt das Intervall $(x - \varepsilon, x + \varepsilon)$ ganz in (a, b) ; wir können uns also sowohl nach links als auch nach rechts um einen beliebigen Betrag kleiner ε bewegen, ohne das Intervall zu verlassen.

Im \mathbb{R}^2 und erst recht in höherdimensionalen Räumen können wir uns nicht nur nach links und rechts bewegen, sondern in unendlich viele Richtungen; wir wollen verlangen, daß es wieder eine positive Zahl ε gibt, so daß wir uns in jeder Richtung um jeden Betrag echt kleiner ε bewegen können ohne die Menge zu verlassen. Im Zweidimensionalen bedeutet dies, wenn wir mit der EUKLIDISCHE Norm arbeiten, daß es zu jedem Punkt x eine Kreisscheibe um diesen Punkt geben soll, die ganz in der Menge drin liegt; in höheren Dimensionen haben wir entsprechend Kugeln und deren höherdimensionale Verallgemeinerungen.

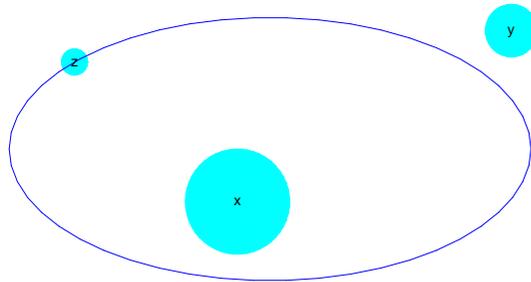
Definition: Eine Teilmenge $M \subseteq V$ eines normierten Vektorraums V mit Norm $\|\cdot\|$ heißt *offen*, wenn es zu jedem Punkt $x \in M$ ein $\varepsilon > 0$ gibt, so daß jedes $y \in V$ mit $\|y - x\| < \varepsilon$ in M liegt. M heißt *abgeschlossen*, wenn die Komplementärmenge $V \setminus M$ offen ist.

Wir können dies auch anders formulieren, indem wir zunächst für eine beliebige Teilmenge $M \subseteq V$ und einen beliebigen Punkt $x \in V$ dessen Lage in Bezug auf M klassifizieren:

Definition: a) Ein Punkt $x \in V$ heißt *innerer Punkt* der Teilmenge $M \subseteq V$, wenn es ein $\varepsilon > 0$ gibt, so daß alle $y \in V$ mit $\|x - y\| < \varepsilon$ in M liegen.

b) x heißt *äußerer Punkt* von M , wenn es ein $\varepsilon > 0$ gibt, so daß alle $y \in V$ mit $\|x - y\| < \varepsilon$ *nicht* in M liegen.

c) x heißt *Randpunkt* von M , wenn es für jedes $\varepsilon > 0$ einen Punkt $y \in M$ gibt mit $\|x - y\| < \varepsilon$ sowie einen Punkt $z \in V \setminus M$, so daß $\|x - z\| < \varepsilon$.



x ist innerer, y äußerer Punkt und z Randpunkt der blau umrandeten Menge

Ein innerer Punkt x von M muß also insbesondere in der Menge M drin liegen, während ein äußerer Punkt von M *nicht* in M liegen darf. In beiden Fällen müssen auch noch alle hinreichend nahe bei x liegenden Punkte dieselbe Eigenschaft haben.

Ein Randpunkt muß nicht in der Menge liegen, kann es aber. Wichtig ist, daß es beliebig nahe von x sowohl Punkte gibt, die in der Menge liegen, als auch solche, die dies nicht tun.

Wenn wir im \mathbb{R}^n mit der EUKLIDischen Norm arbeiten, ist ein Punkt $x \in M$ genau dann ein innerer Punkt von M , wenn es eine Kreisscheibe

bzw. n -dimensionale Kugel um x gibt, die ganz in M liegt; ein Punkt x aus $\mathbb{R}^n \setminus M$ ist genau dann äußerer Punkt, wenn auch noch eine Kreisscheibe bzw. n -dimensionale Kugel um x ganz außerhalb von M liegt. Von einem Randpunkt schließlich verlangen wir, daß jede noch so kleine Kreisscheibe bzw. n -dimensionale Kugel um x sowohl Punkte aus M enthält als auch solche, die nicht in M liegen.

Wenn wir stattdessen mit der Maximumsnorm arbeiten, gilt fast dasselbe; nur haben wir jetzt an Stelle der Kreisscheiben Quadrate und an Stelle von n -dimensionalen Kugeln n -dimensionale Würfel. Die inneren, äußeren und Randpunkte sind offensichtlich bezüglich beider Normen dieselben, und wie oben im Falle der Konvergenz zeigt man leicht

Lemma: Sind $\|\cdot\|_1$ und $\|\cdot\|_2$ äquivalente Normen auf dem Vektorraum V , so ist ein Punkt $x \in V$ genau dann innerer, äußerer bzw. Randpunkt der Teilmenge $M \subseteq V$ bezüglich $\|\cdot\|_1$, wenn er diese Eigenschaft bezüglich $\|\cdot\|_2$ hat. ■

Ebenfalls ziemlich klar ist

Lemma: a) Eine Teilmenge $M \subseteq V$ eines normierten Vektorraums V ist genau dann offen, wenn jeder Punkt $x \in M$ ein innerer Punkt ist.

b) M ist genau dann abgeschlossen, wenn jeder Randpunkt von M in M liegt.

Beweis: a) folgt sofort aus dem Vergleich der Definition offener Mengen und innerer Punkte. Für b) betrachten wir zunächst eine abgeschlossene Teilmenge $M \subseteq V$ und einen Randpunkt x von M . Läge x nicht in M , müßte x in $V \setminus M$ liegen. Da $V \setminus M$ eine offene Menge ist, gäbe es also ein $\varepsilon > 0$, so daß auch alle $y \in V$ mit $\|x - y\| < \varepsilon$ in $V \setminus M$ liegen müßten. Dies widerspricht aber der Definition eines Randpunkts, für den es mindestens ein $y \in M$ geben muß mit $\|x - y\| < \varepsilon$.

Ist umgekehrt $M \subseteq V$ eine Menge, die jeden ihrer Randpunkte enthält, ist keiner der Punkte $x \in V \setminus M$ Randpunkt von M , es gibt also zu jedem $x \in V \setminus M$ ein $\varepsilon > 0$, so daß entweder alle $y \in V$ mit $\|x - y\| < \varepsilon$ in M liegen oder aber alle solche y in $V \setminus M$ liegen. Ersteres ist nicht möglich, da x selbst nicht in M liegt, obwohl $\|x - x\| = 0 < \varepsilon$ ist;

daher müssen alle diese y in $V \setminus M$ liegen, so daß $V \setminus M$ eine offene Menge ist. Damit ist M selbst abgeschlossen, wie behauptet. ■

Nach diesen Vorbereitungen können wir nun stetige Funktionen definieren:

Definition: a) Eine Abbildung $f: D \rightarrow W$ von der offenen Teilmenge $D \subseteq V$ eines normierten Vektorraums $(V, \|\cdot\|_1)$ in einen normierten Vektorraum $(W, \|\cdot\|_2)$ heißt *stetig* in $x \in D$, wenn es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle $y \in D$ gilt: Ist $\|x - y\|_1 < \delta$, so ist $\|f(x) - f(y)\|_2 < \varepsilon$.

b) f heißt *stetig*, wenn f in jedem Punkt $x \in D$ stetig ist.

Auch hier überzeugt man sich leicht, daß sich nichts ändert, wenn wir $\|\cdot\|_1$ und/oder $\|\cdot\|_2$ durch eine äquivalente Norm ersetzen.

Konkret für eine Funktion von n reellen Variablen mit Werten in \mathbb{R} besagt diese Definition, ausgedrückt für die Maximumsnorm

Definition: Eine Funktion $f: D \rightarrow \mathbb{R}$ auf einer offenen Teilmenge D des \mathbb{R}^n heißt *stetig* im Punkt $x = (x_1, \dots, x_n) \in D$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für jeden Punkt $y = (y_1, \dots, y_n) \in D$ gilt: Falls $|y_i - x_i| < \delta$ ist für alle i , dann ist $|f(y) - f(x)| < \varepsilon$.

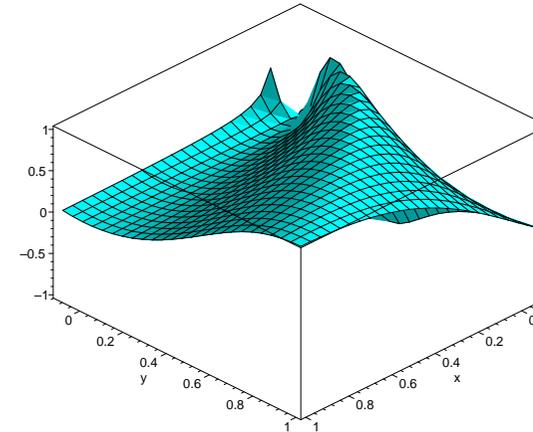
Im Eindimensionalen kann man der graphischen Darstellung einer Funktion leicht ansehen, ob sie stetig ist oder nicht; wir wollen schauen, wie dies im Mehrdimensionalen aussieht.

Der Einfachheit halber beschränken wir uns auf Funktionen zweier Veränderlicher, z.B. die Funktion

$$f: \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto \begin{cases} \frac{2xy^2}{x^2 + y^4} & \text{falls } (x, y) \neq (0, 0) \\ 0 & \text{falls } (x, y) = (0, 0) \end{cases} \end{cases} .$$

Wir sollten erwarten (und werden auch gleich sehen), daß diese Funktion auf $\mathbb{R}^2 \setminus \{(0, 0)\}$ stetig ist, denn dort ist sie nur über Grundrechenarten definiert, wobei Division durch Null ausgeschlossen ist, da $x^2 + y^4$ nur im Nullpunkt verschwindet. Bleibt also der Punkt $(0, 0)$ zu untersuchen.

Die Abbildung unten zeigt den Graphen von f in einer kleinen Umgebung des Nullpunkts. Sie zeigt zwar einen relativ steilen Sprung entlang der Geraden $x = 0$, aber nur für die etwas weiter vom Nullpunkt entfernten x -Werte; bei $y = 0$ sieht alles harmlos und ziemlich eben aus.



Ist diese Funktion stetig?

Nun ist der abgebildete Graph natürlich von einem Computer anhand von nur endlich vielen Stützpunkten konstruiert; um wirklich zu entscheiden, ob f im Nullpunkt stetig ist, können wir uns nicht auf diese Approximation verlassen, sondern müssen die Funktion etwas genauer untersuchen.

Da wir uns im Eindimensionalen recht gut auskennen, können wir beispielsweise die Einschränkungen von f auf die verschiedenen Geraden durch den Nullpunkt betrachten. Abgesehen von der y -Achse haben diese alle die Form $y = ax$ mit $a \in \mathbb{R}$, und

$$f(x, ax) = \frac{2x \cdot (ax)^2}{x^2 + (ax)^4} = \frac{2a^2 x^3}{x^2(1 + a^4 x^2)} = \frac{2a^2 x}{1 + a^4 x^2}$$

für $x \neq 0$. Für $x \rightarrow 0$ geht beim rechtsstehenden Ausdruck der Zähler gegen Null und der Nenner gegen eins; der Grenzwert existiert also und ist gleich $f(0, 0) = 0$, d.h. die Einschränkung von f ist stetig auf der Geraden $y = ax$.

Auf der y -Achse verschwindet f für jeden Wert von x , ist also eben-

falls stetig, d.h. die Einschränkung von f auf jede Gerade durch den Nullpunkt ist stetig.

Betrachten wir zur Vorsicht auch noch die Einschränkung von f auf die Parabel $x = ay^2$! Dort ist

$$f(ay^2, y) = \frac{2ay^2 \cdot y^2}{a^2y^4 + y^4} = \frac{2ay^4}{(1+a^2)y^4} = \frac{2a}{1+a^2}$$

für alle $y \neq 0$, wohingegen $f(0,0) = 0$ ist. Für $a \neq 0$ ist die Einschränkung von f auf diese Parabel also nicht stetig, und damit kann auch f nicht stetig sein: Setzen wir $a = 1$, so ist $2a/(1+a^2) = 1$, die Funktion nimmt also beliebig nahe beim Nullpunkt den Wert eins an. Formal können wir die Unstetigkeit bezüglich der Maximumsnorm folgendermaßen beweisen:

Wenn f im Nullpunkt stetig wäre, gäbe es zu jedem $\varepsilon > 0$, insbesondere also zu $\varepsilon = \frac{1}{2}$, ein $\delta > 0$, so daß für alle Punkte (x, y) mit $\|(x, y)\|_\infty = \max\{|x|, |y|\} < \delta$ der Betrag von $f(x, y)$ kleiner als ε wäre. Für eine reelle Zahl y mit $|y| < \delta$ und $|y| < 1$ ist aber $|y^2| < |y| < \delta$, also $\|(y^2, y)\|_\infty < \delta$, und trotzdem hat

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

einen Betrag größer als $\varepsilon = \frac{1}{2}$.

Dem Graphen konnten wir das nicht ansehen, was bei näherer Überlegung eigentlich auch nicht verwundert: Der Graph wurde von einem Computer gezeichnet, und der kann natürlich nur eine endliche Auswahl x_1, \dots, x_n bzw. y_1, \dots, y_m von Werten berücksichtigen. Dazu berechnet er die Funktionswerte $f(x_i, y_j)$ und verbindet dann die Punkte zu gleichem x_i bzw. gleichem y_j durch Kurven. Entlang dieser Kurven ist f aber, wie wir gesehen haben, stetig.

Anders sieht es aus, wenn wir stattdessen die Niveaulinien von f betrachten: Zusammen mit den Koordinatenachsen überdecken die Parabeln $x = ay^2$ mit $a \in \mathbb{R} \setminus \{0\}$ den gesamten \mathbb{R}^2 , die Niveaulinie $N_0(f)$ von f besteht also aus den beiden Koordinatenachsen, während die anderen Niveaulinien aus Parabeln $x = ay^2$ jeweils ohne den Nullpunkt

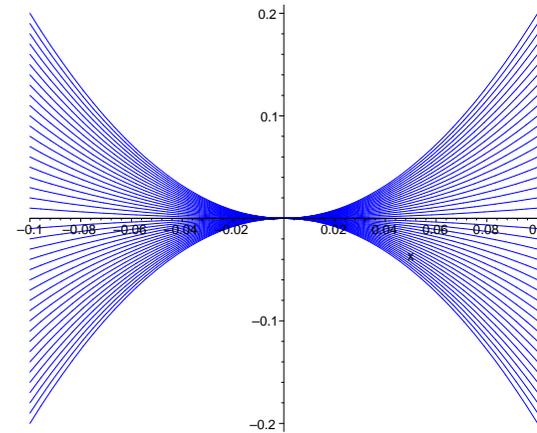
bestehen; siehe Abbildung. Da die Gleichung

$$\frac{2a}{1+a^2} = c$$

für $c \neq 0$ die beiden Lösungen

$$a = \frac{1 \pm \sqrt{1-c^2}}{c}$$

hat, besteht jede Niveaulinie für $0 < |c| < 1$ aus zwei dieser Parabeln, für $|c| = 1$ aus einer, und für $|c| > 1$ ist $N_c(f) = \emptyset$.



Die Niveaulinien illustrieren die Unstetigkeit im Nullpunkt

Dies zeigt auch anschaulich, daß f im Nullpunkt unstetig ist, denn alle Niveaulinien kommen dem Nullpunkt beliebig nahe, obwohl dieser nur auf $N_0(f)$ liegt. Daran sehen wir übrigens auch, daß der oben abgebildete Graph falsch ist: In jeder Umgebung von $(0,0)$ wird jeder Wert c zwischen -1 und 1 angenommen, der Abschluß des Graphen enthält also die Strecke $[-1, 1]$ auf der z -Achse. Wegen der oben beschriebenen Vorgehensweise von Maple (und auch fast aller anderer Computergraphikprogramme) ist das aber in der Abbildung nicht zu sehen.

Bei Funktionen, in deren Definition Fallunterscheidungen eingehen, ist also größere Vorsicht geboten als im Eindimensionalen; bei in der Praxis auftretenden Funktionen wird es allerdings wohl meist so sein, daß die Unstetigkeitsstellen genau dort auftreten, wo man Sprünge definiert

hat. Schwierig wird es nur, wenn man wie im obigen Beispiel in einem Punkt eine Situation der Art „0/0“ hat und entscheiden muß, ob man stetig ergänzen kann: Hier hilft im Mehrdimensionalen keine DE L'HOSPITALSche Regel, es hilft auch nicht, die Annäherung der Funktion an den problematischen Punkt aus allen Richtungen zu untersuchen, sondern man muß wirklich auf die Definition der Stetigkeit zurückgehen.

Zum Glück sind Funktionen wie die obige allerdings nicht der Regelfall, mit dem wir es in Anwendungen zu tun haben; für die meisten gängigen Funktionen ist wie im Eindimensionalen ziemlich klar, daß sie stetig sind.

Beginnen wir mit linearen Funktionen! Da ein konstanter Summand an der Form des Graphen nicht ändert, wollen dabei im Gegensatz zur Linearen Algebra auch inhomogene Funktionen betrachten:

Lemma: Jede lineare Funktion

$$f: \begin{cases} \mathbb{R}^n \rightarrow \mathbb{R} \\ x = (x_1, \dots, x_n) \mapsto \sum_{j=1}^n a_j x_j + b \end{cases}$$

ist stetig.

Beweis: $x = (x_1, \dots, x_n)$ sei ein fester Punkt aus \mathbb{R}^n und $\varepsilon > 0$ sei eine positive reelle Zahl. Für einen weiteren Punkt $u = (u_1, \dots, u_n) \in \mathbb{R}^n$ ist dann

$$\begin{aligned} |f(u) - f(x)| &= \left| \sum_{j=1}^n a_j (u_j - x_j) \right| \leq \sum_{j=1}^n |a_j| \cdot |u_j - x_j| \\ &\leq \sum_{j=1}^n |a_j| \cdot \|u - x\|_\infty. \end{aligned}$$

Falls alle a_j verschwinden, ist dies gleich Null, also insbesondere kleiner als ε ; andernfalls ist es kleiner als ε , falls

$$\|u - x\|_\infty < \delta = \frac{\varepsilon}{\sum_{j=1}^n |a_j|}.$$

In beiden Fällen folgt, daß f stetig im Punkt x ist und damit stetig auf ganz \mathbb{R}^n , denn $x \in \mathbb{R}^n$ war ein beliebiger Punkt. ■

Damit haben wir zwar nur lineare Funktionen mit Werten in \mathbb{R} , aber das reicht: Eine Funktion $f: D \rightarrow \mathbb{R}^m$ ordnet jedes $x \in D$ einen Punkt $f(x) \in \mathbb{R}^m$ zu, und dieser ist gegeben durch m reelle Zahlen $f_1(x), \dots, f_m(x)$. Bezeichnen wir die so definierten Funktionen $f_i: D \rightarrow \mathbb{R}$ als die *Komponenten* von f , so gilt:

Lemma: Eine Funktion $f: D \rightarrow \mathbb{R}^m$ auf einer offenen Teilmenge $D \subseteq \mathbb{R}^n$ ist genau dann stetig in einem Punkt $x \in D$, wenn jede ihrer Komponenten $f_i: D \rightarrow \mathbb{R}$ stetig in x ist.

Beweis: Wir arbeiten mit den Maximumnormen auf \mathbb{R}^n und \mathbb{R}^m . Falls f in $x \in D$ stetig ist, gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß

$$\|f(u) - f(x)\|_\infty = \max\{|f_i(u) - f_i(x)| \mid i = 1, \dots, n\} < \varepsilon$$

falls $\|u - x\|_\infty < \delta$. In diesen Fall ist erst recht $|f_i(u) - f_i(x)| < \varepsilon$ für alle i , also sind alle f_i stetig in x .

Umgekehrt seien alle f_i stetig in x . Dann gibt es für alle $\varepsilon > 0$ positive reelle Zahlen $\delta_1, \dots, \delta_n > 0$, so daß

$$|f_i(u) - f_i(x)| < \varepsilon \quad \text{falls} \quad \|u - x\|_\infty < \delta_i.$$

Bezeichnet δ die kleinste unter den n Zahlen δ_i , ist daher

$$\|f(u) - f(x)\|_\infty = \max\{|f_i(u) - f_i(x)| \mid i = 1, \dots, n\} < \varepsilon$$

falls $\|u - x\|_\infty < \delta$. Somit ist auch f stetig in x . ■

Da jede Komponente einer linearen Funktion linear ist, folgt daraus sofort

Lemma: Jede lineare Funktion

$$f: \begin{cases} \mathbb{R}^n \rightarrow \mathbb{R}^m \\ x = (x_1, \dots, x_n) \mapsto (f_1(x), \dots, f_m(x)) \end{cases}$$

mit $f_i(x) = \sum_{j=1}^n a_{ij} x_j + b_i$ ist stetig. ■

Die folgende Charakterisierung stetiger Funktionen ist gelegentlich einfacher anzuwenden als die Definition:

Lemma: Eine Funktion $f: D \rightarrow \mathbb{R}^m$ ist genau dann stetig auf der offenen Menge $D \subseteq \mathbb{R}^n$, wenn für jede offene Menge $U \subseteq \mathbb{R}^m$ auch deren Urbild $f^{-1}(U) = \{x \in D \mid f(x) \in U\}$ offen ist.

Beweis: Sei zunächst f stetig auf D und $U \subseteq \mathbb{R}^m$ eine offene Menge. Wir müssen zeigen, daß $f^{-1}(U)$ eine offene Menge ist, daß also jeder Punkt $x \in D$ mit $f(x) \in U$ ein innerer Punkt von $f^{-1}(U)$ ist.

Wir betrachten einen festen solchen Punkt x und sein Bild $f(x) \in U$. Wegen der Offenheit von U gibt es dann ein $\varepsilon > 0$, so daß alle $y \in \mathbb{R}^m$ mit $\|y - f(x)\| < \varepsilon$ in U liegen. Zu diesem ε wiederum gibt es wegen der Stetigkeit von f in x ein $\delta > 0$, so daß für alle $u \in D$ mit $\|x - u\| < \delta$ gilt: $\|f(x) - f(u)\| < \varepsilon$. Somit liegt $f(u)$ in U und u selbst liegt in $f^{-1}(U)$. Somit liegen alle $u \in D$ mit $\|x - u\| < \delta$ in $f^{-1}(U)$, so daß x ein innerer Punkt von $f^{-1}(U)$ ist. Da x beliebig war, folgt daraus die Offenheit von $f^{-1}(U)$.

Umgekehrt habe f die Eigenschaft, daß das Urbild einer jeden offenen Menge wieder offen ist. Wir müssen zeigen, daß f in jedem Punkt $x \in D$ stetig ist.

Dazu sei $\varepsilon > 0$ eine beliebige positive Zahl und

$$U = \{y \in \mathbb{R}^m \mid \|f(x) - y\| < \varepsilon\}.$$

Dies ist eine offene Menge, also ist nach Voraussetzung auch ihr Urbild $f^{-1}(U)$ offen. Da x in diesem Urbild liegt, ist x ein innerer Punkt; es gibt also ein $\delta > 0$, so daß alle $u \in \mathbb{R}^n$ mit $\|u - x\| < \delta$ in $f^{-1}(U)$ liegen. Somit liegt für jedes solche u der Bildpunkt $f(u)$ in U , d.h. $\|f(x) - f(u)\| < \varepsilon$. Dies zeigt die Stetigkeit von f in x . ■

Lemma: $f: D \rightarrow \mathbb{R}^m$ und $g: E \rightarrow \mathbb{R}^p$ seien stetige Funktionen auf den offenen Teilmengen $D \subseteq \mathbb{R}^n$ und $E \subseteq \mathbb{R}^m$, und $f(D)$ liege in E . Dann ist auch die Hintereinanderausführung

$$h = g \circ f: \begin{cases} D \rightarrow \mathbb{R}^p \\ x \mapsto g(f(x)) \end{cases}$$

stetig.

Beweis: Dies folgt am einfachsten aus dem vorigen Lemma: Ist $U \subseteq \mathbb{R}^p$ eine offene Menge, so ist wegen der Stetigkeit von g auch $g^{-1}(U) \subseteq \mathbb{R}^m$ offen, also, wegen der Stetigkeit von f , auch $f^{-1}(g^{-1}(U))$. Das ist aber gerade das Urbild von U unter der zusammengesetzten Abbildung $g \circ f$. ■

Lemma: a) Die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $f(x, y) = x \pm y$ ist stetig.

b) Die Funktion $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $g(x, y) = xy$ ist stetig.

c) Die Funktion $h: \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \rightarrow \mathbb{R}$ mit $h(x, y) = x/y$ ist stetig.

Beweis: a) ist klar, da es sich hier um lineare Funktionen handelt, und deren Stetigkeit haben wir bereits allgemein bewiesen.

Der Beweis von b) beruht im wesentlichen auf demselben Trick, mit dem wir in Kapitel 2 bewiesen haben, daß für zwei konvergente Folgen die Produktfolge gegen das Produkt aus deren Grenzwerten konvergiert. Wie dort müssen wir die Fälle, daß einer oder gar beide Faktoren verschwinden, gesondert behandeln.

Sei also zunächst $(x, y) = (0, 0)$ und $\varepsilon > 0$. Dann ist auch $\delta = \sqrt{\varepsilon} > 0$ und für alle Punkte $(u, v) \in \mathbb{R}^2$ mit $\|(u, v)\|_\infty < \delta$ sind $|u|$ und $|v|$ kleiner als δ , also ist

$$|uv - xy| = |uv| = |u| \cdot |v| < \delta^2 = \varepsilon.$$

Damit ist die Stetigkeit im Punkt $(0, 0)$ gezeigt.

Nun sei $x \neq 0$, aber $y = 0$. Für $(u, v) \in \mathbb{R}^2$ ist dann

$$|uv - xy| = |uv - 0| = |u(v - y)| = |u| \cdot |v - y|.$$

Wir müssen zu einem vorgegebenen $\varepsilon > 0$ ein $\delta > 0$ finden, so daß dies kleiner ist als ε falls

$$\|(x, y) - (u, v)\|_\infty = \max\{|x - u|, |y - v|\} = \max\{|x - u|, |v|\} < \delta.$$

Wählen wir $\delta \leq \frac{1}{2}|x|$, so erfüllt jedes u mit $|x - u| < \delta$ die Ungleichung

$$|u| < |x| + \frac{1}{2}|x| = \frac{3}{2}|x|.$$

Wählen wir δ so, daß zusätzlich auch noch $\delta \leq 2\varepsilon/3|x|$ ist, also beispielsweise

$$\delta = \min \left\{ \frac{|x|}{2}, \frac{\varepsilon}{2|x|} \right\},$$

so ist

$$|uv - xy| = |u| \cdot |v - y| < \frac{3|x|}{2} \cdot \frac{2\varepsilon}{3|x|} = \varepsilon$$

für alle $(u, v) \in \mathbb{R}^2$ mit $\|(x, y) - (u, v)\|_\infty < \delta$.

Bleibt noch der Fall $y \neq 0$. Hier haben wir für einen Punkt $(u, v) \in \mathbb{R}^2$ die Abschätzung

$$|uv - xy| = |v(u - x) + x(v - y)| \leq |v| \cdot |u - x| + |x| \cdot |v - y|.$$

Falls wir nur $v \in \mathbb{R}$ mit $|v - y| < \frac{1}{2}|y|$ betrachten, können wir $|v|$ durch $\frac{3}{2}|y|$ nach oben abschätzen und erhalten die Ungleichung

$$|uv - xy| \leq \frac{3|y|}{2} \cdot |u - x| + |x| \cdot |v - y|.$$

Ist nun ein $\varepsilon > 0$ vorgegeben, setzen wir

$$\delta = \begin{cases} \frac{2\varepsilon}{3|y|} & \text{falls } x = 0 \\ \min \left\{ \frac{\varepsilon}{3|y|}, \frac{\varepsilon}{2|x|} \right\} & \text{falls } x \neq 0 \end{cases};$$

dann ist $|uv - xy| < \varepsilon$ für alle $(u, v) \in \mathbb{R}^2$ mit $\|(u, v) - (x, y)\|_\infty < \delta$. Somit ist die Funktion stetig in jedem Punkt $(x, y) \in \mathbb{R}^2$.

Die in c) behauptete Stetigkeit der Division könnten wir ähnlich beweisen: schneller geht es aber, wenn wir die aus der Analysis I bekannte Stetigkeit der Funktion $\mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ mit $y \mapsto 1/y$ ausnutzen. Da eine Funktion genau dann stetig ist, wenn alle ihre Komponenten stetig sind, ist auch die Funktion

$$\begin{cases} \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \rightarrow \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \\ (x, y) \mapsto \left(x, \frac{1}{y}\right) \end{cases}$$

stetig, nach dem vorigen Lemma also auch ihre Schachtelung mit der Multiplikation, die Division. Damit haben wir die Stetigkeit aller Grundrechenarten bewiesen. ■

Die gerade bewiesenen Lemmata lassen sich zusammenfassen zum folgenden Prinzip:

Wenn eine Funktion $f: D \rightarrow \mathbb{R}^m$ auf einer offenen Teilmenge $D \subseteq \mathbb{R}^n$ so aus Grundrechenarten und stetigen Funktionen zusammengesetzt ist,

daß nie eine Funktion außerhalb ihrer Stetigkeitsbereiche verwendet wird und auch Divisionen durch Null ausgeschlossen sind, dann ist f stetig auf D .

§2: Differenzierbare Funktionen

Nachdem wir wissen, was Konvergenz und Stetigkeit im Mehrdimensionalen bedeuten, können wir uns der Differenzierbarkeit zuwenden. Wir beginnen mit einer kurzen Wiederholung des eindimensionalen Falls:

a) Funktionen einer Veränderlichen

Wir bezeichnen eine Funktion $f: (a, b) \rightarrow \mathbb{R}$ als differenzierbar im Punkt $x \in (a, b)$, wenn sie in dessen Umgebung durch eine lineare Funktion angenähert werden kann. In Kapitel 3 hatten wir dies so formuliert, daß es ein $c \in \mathbb{R}$ sowie eine Funktion \tilde{f} mit $\tilde{f}(0) = 0$ geben muß, so daß für kleine Werte von h gilt

$$f(x+h) = f(x) + ch + h\tilde{f}(h).$$

Da diese Formulierung mit explizit angegebener Fehlerfunktion \tilde{f} bei längeren Betrachtungen recht umständlich ist, wollen wir eine abkürzende Sprechweise einführen, die LANDAUSche o -Notation: Wir schreiben $o(h)$, sobald wir irgendeine uns nicht weiter interessierende Funktion von h haben, die für $h \rightarrow 0$ schneller gegen Null geht als h selbst, d.h.

$$\varphi(h) = o(h) \iff \lim_{h \rightarrow 0} \frac{\varphi(h)}{h} = 0.$$

$o(h)$ ist hier also keine Funktion, sondern steht für eine ganze Klasse von Funktionen; beispielsweise ist

$$h^2 = o(h), \quad h^5 = o(h) \quad \text{und} \quad h \cdot \sin h = o(h),$$

aber $\sin h$ können wir nicht als $o(h)$ schreiben, denn

$$\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1.$$

Entsprechend schreiben wir auch

$$\varphi(h) = o(\psi(h)), \quad \text{wenn} \quad \lim_{h \rightarrow 0} \frac{\varphi(h)}{\psi(h)} = 0$$

ist.



EDMUND GEORG HERMANN LANDAU (1877–1938) wurde in Berlin geboren und studierte an der dortigen Universität, wo er auch von 1899 bis 1909 lehrte. Dann bekam er einen Ruf an die damals führende deutsche Mathematikfakultät in Göttingen. 1933 verlor er seinen dortigen Lehrstuhl, denn die Studenten boykottierten seine Vorlesungen, da sie meinten, sie könnten Mathematik nur von einem Professor ihrer eigenen Rasse lernen. LANDAUS zahlreiche Publikationen beschäftigen sich vor allem mit der Zahlentheorie, über die er auch ein bedeutendes Lehrbuch schrieb; sehr bekannt sind seine Arbeiten über die Verteilung von Primzahlen.

Mit LANDAUS o -Notation können wir kurz sagen, die Funktion f sei genau dann differenzierbar in x mit Ableitung $f'(x)$, wenn

$$f(x+h) = f(x) + hf'(x) + o(h)$$

ist, denn natürlich ist $hf'(h) = o(h)$, denn der Quotient $hf'(h)/h = \tilde{f}(h)$ geht für $h \rightarrow 0$ gegen $\tilde{f}(0) = 0$.

b) Differenzierbarkeit im Mehrdimensionalen

Um Differenzierbarkeit für Funktionen mehrerer Veränderlicher zu definieren, können wir ähnlich vorgehen wie im eindimensionalen Fall. Wir betrachten zunächst eine Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, zum Beispiel $f(x, y) = \sin x \cos y$ in der Umgebung des Punktes $(1, 1)$. Indem wir ihren Graphen sukzessive um den Faktor fünf vergrößern, erhalten wir die unten folgende Abbildungen. Sie zeigen, daß sich der Graph in einer hinreichend kleinen Umgebung von $(1, 1)$ nur wenig von einer Ebenen unterscheidet, d.h. die Funktion ist dort annähernd linear.

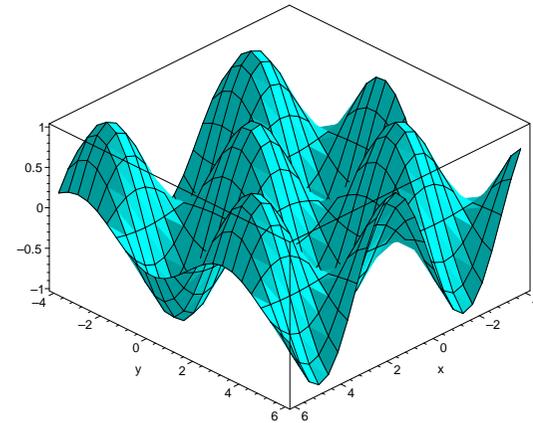
Eine lineare Funktion zweier Veränderlicher, bei der wir auch hier wieder im Gegensatz zur Linearen Algebra einen konstanten Term zulassen, läßt sich schreiben als

$$L(x, y) = a + bx + cy;$$

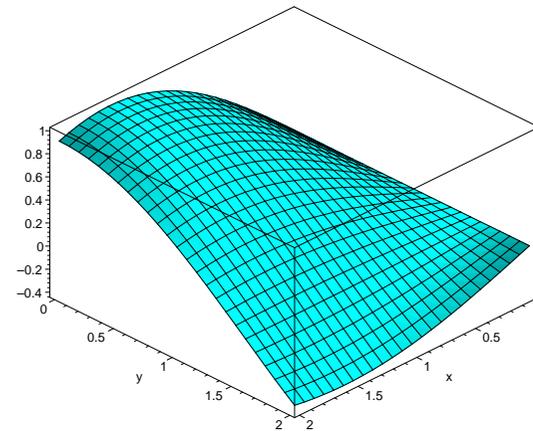
also ist

$$L(x+h, y+k) = a + b(x+h) + c(y+k) = L(x, y) + bh + ck$$

eine Approximation für f in der Umgebung des betrachteten Punkts (x, y) . Im Punkt (x, y) sollte $L(x, y)$ natürlich mit $f(x, y)$ übereinstimmen, so daß $a = f(x, y)$ sein muß, und für $(h, k) \neq (0, 0)$ sollte der



Graph der Funktion $z = \sin x \cos y$

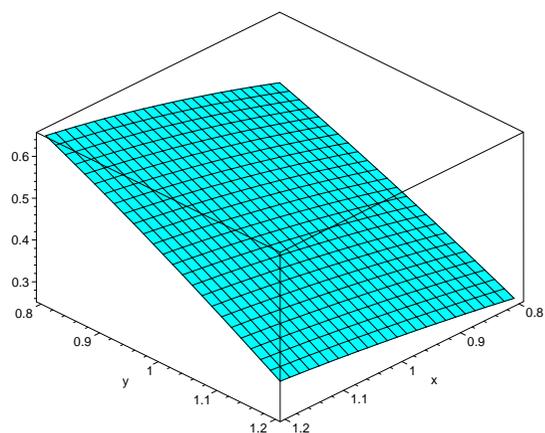


Vergrößerung um den Faktor fünf um $(1, 1, \sin(1) \cos(1))$

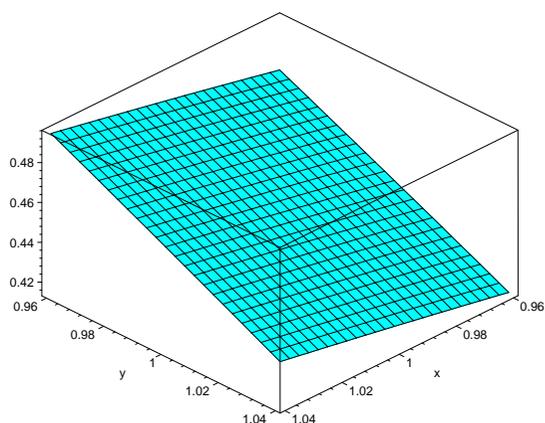
Unterschied zwischen f und L schneller gegen Null gehen als der Abstand zwischen $(x+h, y+k)$ und (x, y) , d.h. schneller als $\sqrt{h^2 + k^2}$. Wir erwarten daher, daß

$$f(x+h, y+k) = f(x, y) + bh + ck + o(\sqrt{h^2 + k^2})$$

ist, und genau so läßt sich Differenzierbarkeit allgemein definieren.



Nochmalige Vergrößerung um den Faktor fünf um $(1, 1, \sin(1) \cos(1))$



Nach noch einer Vergrößerung sieht die Funktion praktisch linear aus

Wenn wir eine Funktion von n Veränderlichen in der Umgebung eines Punkts $x \in \mathbb{R}^n$ betrachten, müssen wir alle n Variablen variieren; die Nachbarpunkte sind also Punkte der Form $x + h$ mit Vektoren $h \in \mathbb{R}^n$. Im Falle einer Funktion mit Werten in \mathbb{R} ist $f(x + h)$ eine reelle Zahl

und eine Linearisierung von f um x hat die Form

$$f(x) + \sum_{i=1}^n \gamma_i h_i,$$

wobei h_i die Komponenten des Vektors h sind. Für eine differenzierbare Funktion erwarten wir, daß der Fehler dieser Linearisierung schneller gegen Null geht als der Vektor h , das heißt also, schneller als dessen Norm. Ob wir dabei die EUKLIDISCHE oder die Maximumsnorm verwenden, bleibt sich gleich, denn die beiden sind ja äquivalent. Wir verlangen daher einfach, daß der Fehler $o(\|h\|)$ sein soll, ohne uns auf eine spezielle Norm festzulegen.

Für eine Funktion mit Werten in \mathbb{R}^m ändert sich nichts wesentliches: $f: D \rightarrow \mathbb{R}^m$ ist gegeben durch m Komponentenfunktionen $f_i: D \rightarrow \mathbb{R}$, und eine Linearisierung von f ist gleichbedeutend mit der Linearisierung der sämtlichen f_i . Im Falle der Differenzierbarkeit soll es also für jedes i reelle Zahlen $\gamma_{i1}, \dots, \gamma_{in}$ geben, so daß

$$f_i(x + h) = f_i(x) + \sum_{j=1}^n \gamma_{ij} h_j + o(\|h\|)$$

ist. Den Vektor aus \mathbb{R}^m , dessen i -te Komponente die Summe der $\gamma_{ij} h_j$ ist, können wir auch kurz schreiben als Produkt

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \gamma_{1j} h_j \\ \sum_{j=1}^n \gamma_{2j} h_j \\ \vdots \\ \sum_{j=1}^n \gamma_{mj} h_j \end{pmatrix}$$

der Matrix mit Einträgen γ_{ij} mit dem Vektor h . Für eine differenzierbare Funktion $f: D \rightarrow \mathbb{R}^m$ erwarten wir, daß es in jeder der m Komponenten einen Fehler der Form $o(\|h\|)$ gibt; wir haben also einen Vektor aus \mathbb{R}^m , dessen sämtliche Komponenten $o(\|h\|)$ sind. Um eine kurze Schreibweise zu haben, bezeichnen wir auch diesen Vektor einfach mit $o(\|h\|)$ und definieren:

Definition: a) Die Funktion $f: D \rightarrow \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ heißt *differenzierbar* im Punkt $x \in D$, wenn es eine $n \times m$ -Matrix $J_f(x)$ gibt, so daß

für Vektoren $h \in \mathbb{R}^n$ mit $h \rightarrow 0$ gilt

$$f(x+h) = f(x) + J_f(x)h + o(\|h\|).$$

Die Matrix $J_f(x)$ heißt *Ableitung* oder *JACOBI-Matrix* von f im Punkt x .

b) Für eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ kann die dann einzeilige JACOBI-Matrix mit einem Vektor aus \mathbb{R}^n identifiziert werden; dieser Vektor

$$\text{grad } f(x) \stackrel{\text{def}}{=} \nabla f(x) \stackrel{\text{def}}{=} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}$$

heißt *Gradient* von f im Punkt $x \in D$. c) Eine differenzierbare Funktion heißt *stetig differenzierbar*, wenn ihre Ableitung stetig ist.

Das Symbol ∇ sieht zwar aus wie ein griechischer Buchstabe, ist aber keiner; es ist ein auf den Kopf gestelltes großes Delta (Δ). ∇f wird „Nabla f“ ausgesprochen nach dem griechischen Wort $\nu\alpha\beta\lambda\alpha = \text{Leier}$; die Bezeichnung wurde eingeführt von dem irischen Mathematiker WILLIAM ROWEN HAMILTON, den die Form von ∇ an eine Leier erinnerte.



WILLIAM ROWEN HAMILTON (1805–1865) wurde in Dublin geboren; bereits mit fünf Jahren sprach er Latein, Griechisch und Hebräisch. Mit dreizehn begann er, mathematische Literatur zu lesen, mit 21 wurde er, noch als Student, Professor der Astronomie am Trinity College in Dublin. Er verlor allerdings schon bald sein Interesse für Astronomie und arbeitete weiterhin auf dem Gebiet der Mathematik und Physik. Am bekanntesten ist seine Entdeckung der Quaternionen 1843, vorher publizierte er aber auch bedeutende Arbeiten über Optik, Dynamik und Algebra.



CARL GUSTAV JACOB JACOBI (1804–1851) wurde in Potsdam als Sohn eines jüdischen Bankiers geboren und erhielt den Vornamen Jacques Simon. Im Alter von zwölf Jahren bestand er sein Abitur, mußte aber noch vier Jahre in Abschlußklasse des Gymnasiums bleiben, da die Berliner Universität nur Studenten mit mindestens 16 Jahren aufnahm. 1824 beendete er seine Studien mit dem Staatsexamen für Mathematik, Griechisch und Latein und wurde Lehrer. Außerdem promovierte er 1825 und begann mit seiner Habilitation. Etwa gleichzeitig konvertierte er zum Christentum, so daß er ab 1825 an der Universität Berlin und ab 1826 in

Königsberg lehren konnte. 1832 wurde er dort Professor. Zehn Jahre später mußte er aus gesundheitlichen Gründen das raue Klima Königsbergs verlassen und lebte zunächst in Italien, danach für den Rest seines Lebens in Berlin. Er ist vor allem berühmt durch seine Arbeiten zur Zahlentheorie und über elliptische Integrale.

Mit obiger Definition haben wir etwas ähnliches wie die Definition

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

für Funktionen einer Veränderlicher: Auf diese Weise ist die Ableitung zwar *definiert*, aber sie wird – außerhalb von Anfängervorlesungen – praktisch nie so ausgerechnet. Entsprechend sollten wir auch für Funktionen mehrerer Veränderlicher eine effizientere Methode der Differentiation finden.

Am einfachsten wäre es, wenn wir den wohlbekannten Kalkül der Differentialrechnung für Funktionen einer Veränderlicher benutzen könnten, also versuchen wir, aus einer Funktion $f: D \rightarrow \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ Funktionen einer Veränderlichen zu machen.

Dabei können wir uns sofort auf den Fall $m = 1$ beschränken, denn jede Funktion $f: D \rightarrow \mathbb{R}^m$ ist zusammengesetzt aus m Komponentenfunktionen $f_i: D \rightarrow \mathbb{R}$; wir beschränken uns daher zunächst auf Funktionen $f: D \rightarrow \mathbb{R}$.

Schwieriger ist die Reduktion von n auf eins. Trotz der schlechten Erfahrungen im vorigen Paragraphen, wo eine unstetige Funktion nach Einschränkung auf eine beliebige Gerade stets stetig wurde, wollen wir uns exakt diese Einschränkungen genauer anschauen.

Eine Gerade durch einen gegebenen Punkt x ist eindeutig festgelegt durch einen Richtungsvektor e , wobei umgekehrt der Vektor e durch die Gerade natürlich *nicht* eindeutig festgelegt ist: Jedes Vielfache von e (außer dem Nullvektor) definiert dieselbe Gerade.

Wenn wir die Einschränkung von f auf eine solche Gerade mit Richtungsvektor e betrachten, betrachten wir konkret die Funktion

$$g(t) = f(x + te),$$

einer einzigen Variablen $t \in \mathbb{R}$, die überall dort definiert ist, wo $x + te$ im Definitionsbereich D von f liegt; für eine offene Menge D also

zumindest in einem gewissen offenen Intervall um den Nullpunkt der reellen Geraden.

Damit können wir nach der Differenzierbarkeit dieser Funktion für $t = 0$ fragen; falls sie differenzierbar ist, bezeichnen wir die Ableitung

$$g'(0) = \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{f(x + he) - f(x)}{h}$$

als *Richtungsableitung* von f in Richtung e . Eine einfache Anwendung der Kettenregel, die jeder Leser am Rand des Skriptums kurz durchführen sollte, zeigt, daß diese „Richtungsableitung“ nicht nur von der *Richtung* des Vektors e abhängt, sondern auch von dessen *Länge*: Beispielsweise ist für $k(t) = f(x + 2te)$

$$k'(0) = 2g'(0).$$

Speziell können wir diese Richtungsableitungen betrachten für den Fall, daß e ein *Einheitsvektor* ist (genau ist der Grund für die Bezeichnung e), beispielsweise einer der Koordinateneinheitsvektoren

$$e_i = (0, \dots, 1, \dots, 0),$$

bei dem an der i -ten Stelle eine Eins steht und sonst lauter Nullen.

Alsdann ist für $g_i(t) = f(x + te_i)$

$$\begin{aligned} g_i'(0) &= \lim_{h \rightarrow 0} \frac{f(x + he_i) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h} \end{aligned}$$

die Ableitung jener Funktion, die nur von x_i abhängt, während alle anderen Koordinaten x_j festgehalten werden. Wir behandeln also beim Differenzieren alle Variablen x_j mit Ausnahme von x_i als Konstanten und leiten die so entstehende Funktion in der üblichen Weise ab nach x_i . Diese Ableitung, so sie existiert, bezeichnen wir als *partielle Ableitung*

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

von f nach x_i ; das Symbol ∂ wird, wenn überhaupt, als „del“ ausgesprochen, wobei *del* natürlich eine Abkürzung für *delta* ist. Partielle Ableitungen, so sie existieren, lassen sich nach den üblichen Regeln der

Differentialrechnung für Funktionen einer Veränderlichen berechnen, sind also für „gutartige“ Funktionen problemlos.

Falls die Funktion f in $x \in D$ differenzierbar ist, existiert auch jede Richtungsableitung, denn da dann für jeden Vektor h gilt

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + o(\|h\|),$$

ist insbesondere auch

$$f(x + te) = f(x) + \langle \nabla f(x), te \rangle + o(\|te\|) = f(x) + t \langle \nabla f(x), e \rangle + o(t);$$

denn $\langle \nabla f(x), e \rangle$ und $\|e\|$ sind schließlich Konstanten. Damit existiert

$$\left. \frac{d}{dt} f(x + te) \right|_{t=0} = \langle \nabla f(x), e \rangle$$

für jeden Richtungsvektor e ; insbesondere existieren natürlich alle partiellen Ableitungen.

Die Umkehrung gilt leider nicht immer: Die (offensichtlich in $(0, 0)$ unstetige) Funktion

$$f: \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto \begin{cases} 1 & \text{falls } xy \neq 0 \\ 0 & \text{falls } xy = 0 \end{cases} \end{cases}$$

verschwindet auf der x -Achse und der y -Achse; überall sonst hat sie den Wert eins. Damit existieren im Punkt $(0, 0)$ beide partielle Ableitungen und sind identisch Null. Trotzdem ist f natürlich nicht differenzierbar in $(0, 0)$, denn da $f(h, k)$ für jeden Punkt, der nicht auf einer der beiden Koordinatenachsen liegt, gleich eins ist, kann es keine Zahlen $b, c \in \mathbb{R}$ geben, so daß

$$f(h, k) = f(0, 0) + bh + ck + o(\sqrt{h^2 + k^2})$$

ist, denn $f(0, 0) = 0$ und $f(h, k) = 1$ für $hk \neq 0$.

Nun wird natürlich jeder vernünftige Mensch einwenden, daß dieses Beispiel sehr künstlich ist, und in der Tat verhalten sich „gutartige“ Funktionen nicht so:

Lemma: Falls $f: D \rightarrow \mathbb{R}$ auf der offenen Teilmenge $D \subseteq \mathbb{R}^n$ stetig ist und auch alle partiellen Ableitungen von f dort existieren und stetig

sind, ist f in D differenzierbar und

$$\text{grad } f(x) = \nabla f(x) = \begin{pmatrix} f_{x_1}(x) \\ \vdots \\ f_{x_n}(x) \end{pmatrix}.$$

Beweis: Sei $x \in D$. Da D offen ist, gibt es eine Kugel um x , die ganz in D liegt; wir betrachten im folgenden nur Vektoren h , deren Länge höchstens gleich dem Radius dieser Kugel ist, so daß $x + h$ stets in D liegt.

Wir betrachten f zunächst nur als Funktion der ersten Variablen; da deren Ableitung f_{x_1} in ganz D existiert, ist

$$\begin{aligned} f(x+h) &= f(x_1+h_1, \dots, x_n+h_n) \\ &= f(x_1, x_2+h_2, \dots, x_n+h_n) + f_{x_1}(x_1, x_2+h_2, \dots, x_n+h_n)h_1 + o(h_1). \end{aligned}$$

Genauso ist, da die partielle Ableitung nach x_2 in ganz D existiert,

$$\begin{aligned} f(x_1, x_2+h_2, \dots, x_n+h_n) &= \\ f(x_1, x_2, x_3, \dots, x_n+h_n) &+ f_{x_2}(x_1, x_2, x_3+h_3, \dots, x_n+h_n)h_2 + o(h_2) \end{aligned}$$

und so weiter. Insgesamt erhalten wir

$$\begin{aligned} f(x+h) &= f(x) + f_{x_1}(x_1, x_2+h_2, x_3+h_3, \dots, x_n+h_n)h_1 + o(h_1) \\ &\quad + f_{x_2}(x_1, x_2, x_3+h_3, \dots, x_n+h_n)h_2 + o(h_2) \\ &\quad \vdots \\ &\quad + f_{x_{n-1}}(x_1, \dots, x_{n-1}, x_n+h_n)h_{n-1} + o(h_{n-1}) \\ &\quad + f_{x_n}(x_1, \dots, x_n)h_n + o(h_n). \end{aligned}$$

Die LANDAU-Symbole $o(h_1), \dots, o(h_n)$ können wir zu $o(\|h\|)$ zusammenfassen, denn da $|h_i| \leq \|h\|$ für jedes i , kann keines der h_i langsamer gegen Null gehen als $|h|$.

Damit sind wir schon ziemlich nahe an dem, was wir für die Differenzierbarkeit brauchen; allerdings hängen die partiellen Ableitungen noch von den h_i ab, so daß die Differenz zwischen $f(x+h)$ und $f(x)$ nicht durch eine lineare Funktion angenähert ist.

Hier kommt nun die Stetigkeit der partiellen Ableitungen ins Spiel: Diese impliziert, daß

$$\lim_{h \rightarrow 0} (f_{x_1}(x_1, x_2+h_2, \dots, x_n+h_n) - f_{x_1}(x_1, x_2, \dots, x_n)) = 0$$

ist. Damit ist $(f_{x_1}(x_1, x_2+h_2, \dots, x_n+h_n) - f_{x_1}(x_1, x_2, \dots, x_n))h_1 = o(\|h\|)$, denn wenn h_1 mit einem Ausdruck multipliziert wird, der gegen Null geht, strebt das Produkt für $h \rightarrow 0$ schneller gegen Null als h_1 allein, und $o(h_1)$ kann durch $o(\|h\|)$ abgeschätzt werden. Also ist

$$\begin{aligned} f_{x_1}(x_1, x_2+h_2, \dots, x_n+h_n)h_1 \\ = f_{x_1}(x_1, x_2, \dots, x_n)h_1 + o(\|h\|) = f_{x_1}(x)h_1 + o(\|h\|). \end{aligned}$$

Entsprechend können wir auch bei den übrigen partiellen Ableitungen argumentieren und erhalten insgesamt, daß

$$\begin{aligned} f(x+h) &= f(x) + f_{x_1}(x)h_1 + \dots + f_{x_n}(x)h_n + o(\|h\|) \\ &= f(x) + \begin{pmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{pmatrix} \cdot h + o(\|h\|) \end{aligned}$$

ist. Damit ist das Lemma bewiesen. ■

Für Funktionen mit stetigen partiellen Ableitungen ist der Gradient also gerade der Vektor der partiellen Ableitungen; er kann damit über die bekannten Ableitungsregeln für Funktionen einer Veränderlichen berechnet werden.

NB: Häufig wird der Gradient durch diese Formel *definiert*; in diesem Fall folgt natürlich aus der Existenz des Gradienten nicht die Differenzierbarkeit der Funktion; siehe obiges Beispiel einer unstetigen Funktion, für die alle partiellen Ableitungen in $(0, 0)$ existieren.

c) Ableitungsregeln

Da wir Ableitungen von Funktionen mehrerer Veränderlichen auf die mit Methoden der eindimensionalen Analysis bestimmbar partiellen Ableitungen zurückgeführt haben, sollten wir erwarten, daß sich auch die gewohnten Rechenregeln der Differentialrechnung einer Veränderlichen übertragen lassen. Dies ist in der Tat der Fall, wobei selbst die

Beweismethoden fast wörtlich übernommen werden können. Daher sollen hier nur ganz kurz einige einfache Regeln gezeigt werden; für den Rest sei auf die Übungen verwiesen.

Am einfachsten ist die Linearität der Ableitung:

Lemma: Sind $f, g: D \rightarrow \mathbb{R}^m$ differenzierbare Funktionen auf der offenen Teilmenge $D \subseteq \mathbb{R}^n$, so ist auch für alle $a, b \in \mathbb{R}$ die Funktion $af + bg$ differenzierbar und $J_{af+bg}(x) = aJ_f(x) + bJ_g(x)$ für alle $x \in D$. Speziell im Falle $m = 1$ ist also $\nabla(af + bg)(x) = a\nabla f(x) + b\nabla g(x)$.

Beweis: Wegen der Differenzierbarkeit von f und g ist für alle $x \in D$ und $h \in \mathbb{R}^n$ mit $x + h \in D$

$$f(x+h) = f(x) + J_f(x)h + o(\|h\|) \quad \text{und} \quad g(x+h) = g(x) + J_g(x)h + o(\|h\|).$$

Damit ist auch

$$\begin{aligned} (af + bg)(x+h) &= af(x+h) + bg(x+h) \\ &= af(x) + bg(x) + aJ_f(x)h + bJ_g(x)h + o(\|h\|) \\ &= (af + bg)(x) + (aJ_f(x) + bJ_g(x))h + o(\|h\|), \end{aligned}$$

denn auch jede Linearkombination zweier Funktionen der Form $o(\|h\|)$ ist wieder $o(\|h\|)$. Somit ist $af + bg$ differenzierbar mit Ableitung $J_{af+bg}(x) = aJ_f(x) + bJ_g(x)$ für alle $x \in D$. ■

Auch die LEIBNIZ-Regel gilt, d.h.

Lemma: Sind $f, g: D \rightarrow \mathbb{R}$ stetig differenzierbare Funktionen auf der offenen Teilmenge $D \subseteq \mathbb{R}^n$, so ist auch $fg: D \rightarrow \mathbb{R}$ differenzierbar und $\nabla(fg)(x) = f\nabla g(x) + g\nabla f(x)$.

Beweis: Wir können entweder wie oben mit der Definition argumentieren, oder aber mit partiellen Ableitungen: Nach der Produktregel für Funktionen einer Veränderlichen ist die Produktfunktion für jede der Variablen x_i partiell differenzierbar und

$$\frac{\partial(fg)}{\partial x_i}(x) = f(x) \frac{\partial g}{\partial x_i}(x) + g(x) \frac{\partial f}{\partial x_i}(x).$$

Wegen der vorausgesetzten stetigen Differenzierbarkeit sind diese Ableitungen auch stetig, also ist fg differenzierbar mit diesen Ableitungen

als Komponenten des Gradienten. Das sind aber genau die Komponenten von $f\nabla g(x) + g\nabla f(x)$. ■

Schließlich gilt auch eine Kettenregel:

Lemma: $f: D \rightarrow \mathbb{R}^m$ sei differenzierbar auf $D \subseteq \mathbb{R}^n$ und $g: E \rightarrow \mathbb{R}^p$ sei differenzierbar auf $E \subseteq \mathbb{R}^m$, wobei $f(D) \subseteq E$ sei. Dann ist auch $g \circ f$ differenzierbar und $J_{g \circ f}(x) = J_g(f(x))J_f(x)$ für alle $x \in D$.

Beweis: Wegen der Differenzierbarkeit von f und g ist für alle $x \in D$, $y \in E$ und alle $h \in \mathbb{R}^n$, $k \in \mathbb{R}^m$ mit $x+h \in D$ und $y+k \in E$

$$f(x+h) = f(x) + J_f(x)h + o(\|h\|) \quad \text{und} \quad g(y+k) = g(y) + J_g(y)k + o(\|k\|).$$

Ist $f(x+h) \in E$, gilt daher auch

$$\begin{aligned} g(f(x+h)) &= g(f(x) + J_f(x)h + o(\|h\|)) \\ &= g(f(x)) + J_g(f(x))(J_f(x)h + o(\|h\|)) + (J_f(x)h + o(\|h\|)) \\ &= g(f(x)) + J_g(f(x))J_f(x)h + o(\|h\|), \end{aligned}$$

denn bei Multiplikation mit der (bezüglich h) konstanten Matrix $J_f(x)$ bleibt eine Funktion $o(\|h\|)$ von dieser Form. ■

Für weitere Ableitungsregeln sei auf die Übungen verwiesen.

Natürlich gibt es auch im Mehrdimensionalen nicht nur eine Ableitung, sondern wir können Funktionen, entsprechende Differenzierbarkeit vorausgesetzt, auch mehrfach ableiten. Was genau wir unter den höheren Ableitungen einer Funktion mehrerer Veränderlicher verstehen wollen, soll aber erst weiter hinten definiert werden, da wir wegen der bereits zu Beginn dieses Semesters in der Mikroökonomie wichtigen Anwendungen auf Extremwertprobleme mit Nebenbedingungen zunächst solche Probleme betrachten wollen – soweit dies mit der ersten Ableitung allein möglich ist. Auch der nächste Abschnitt ist für uns vor allem wichtig für den Umgang mit Nebenbedingungen.

d) Der Satz über implizite Funktionen

Der Zusammenhang zwischen zwei Größen x und y ist nicht immer explizit in der Form $y = f(x)$ gegeben; gelegentlich hat man auch nur

einen impliziten Zusammenhang $F(x, y) = 0$; entsprechend auch für mehr als zwei Variablen. In diesem Abschnitt soll untersucht werden, wann eine Gleichung der Form $F(x) = 0$ nach einer der Variablen x_i aufgelöst werden kann.

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. Falls sie aber verschwindet, fehlt diese Information.

Der Satz über implizite Funktionen besagt, daß das Nichtverschwinden dieser 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 $F: D \rightarrow \mathbb{R}$ sei stetig differenzierbar. 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 $f: I \rightarrow K$, so daß für alle $x \in I$ gilt:

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

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 $x_0 = y_0 = 0$ zu 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

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

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 α 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

$$f: \begin{cases} I & \rightarrow K \\ x_1 & \mapsto y_1 \end{cases}.$$

Damit ist f als Funktion festgelegt, und nach Konstruktion ist

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

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

$$f\left(\lim_{n \rightarrow \infty} x_n\right) = \lim_{n \rightarrow \infty} f(x_n)$$

ist.

Da x in I liegt, wissen wir, daß es dazu ein eindeutig bestimmtes $y \in K$ gibt, so daß $F(x, y) = 0$ ist, nämlich $y = f(x)$. Für jedes $\varepsilon > 0$ gibt es daher ein $\delta > 0$, so daß $|F(x', y')| < \varepsilon$ ist, falls der Abstand zwischen (x', y') und (x, y) kleiner ist als δ . Zu diesem δ wiederum gibt es ein $N \in \mathbb{N}$, so daß $|x - x_n| < \delta$ für alle $n > N$, so daß für alle solchen n gilt: $|F(x_n, y)| < \varepsilon$. Läßt man hier ε gegen Null gehen, folgt, daß

$$F(x, y) = F\left(\lim_{n \rightarrow \infty} x_n, y\right) = 0$$

ist, d.h. $F(x, y) = 0$ und damit $f(x) = y$, 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

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

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

$$F(x + h, f(x + h)) = F(x + h, f(x) + k) = F\left((x, f(x)) + \begin{pmatrix} h \\ k \end{pmatrix}\right),$$

und setzen wir

$$\varphi(t) \stackrel{\text{def}}{=} F\left((x, f(x)) + t \begin{pmatrix} h \\ k \end{pmatrix}\right),$$

so ist nach dem Mittelwertsatz der Differentialrechnung

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

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

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

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

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

Für $h \rightarrow 0$ 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 $h \rightarrow 0$, 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 $F(x, f(x))$ ist gleich der Nullfunktion, und damit verschwindet natürlich auch ihre Ableitung. Andererseits ist diese Ableitung nach der Kettenregel gleich

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

also folgt

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

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, $f: D \rightarrow \mathbb{R}$ eine stetig differenzierbare Funktion auf D , und $a = (a_1, \dots, a_n) \in D$ sei ein Punkt

mit $F(a) = 0$. Falls $F_{x_n}(a) \neq 0$ ist, gibt es eine offene Umgebung U von (a_1, \dots, a_{n-1}) in \mathbb{R}^{n-1} und eine Funktion $f \in \mathcal{C}^1(U, \mathbb{R})$, so daß für alle Punkte $y = (y_1, \dots, y_{n-1}) \in U$ gilt:

$$F(y, f(y)) = 0.$$

Für alle $i \leq n - 1$ ist

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

■

e) Ableitungen und Extrema

Im Eindimensionalen ist das Verschwinden der Ableitung eine notwendige Bedingung für einen Extremwert. Dies gilt genau so auch im Mehrdimensionalen:

Für eine differenzierbare Funktion $f: D \rightarrow \mathbb{R}$ auf einer offenen Teilmenge $D \subset \mathbb{R}^n$ ist im Punkt $x_0 \in D$

$$f(x_0 + h) = f(x_0) + \text{grad } f(x_0) \cdot h + o(\|h\|).$$

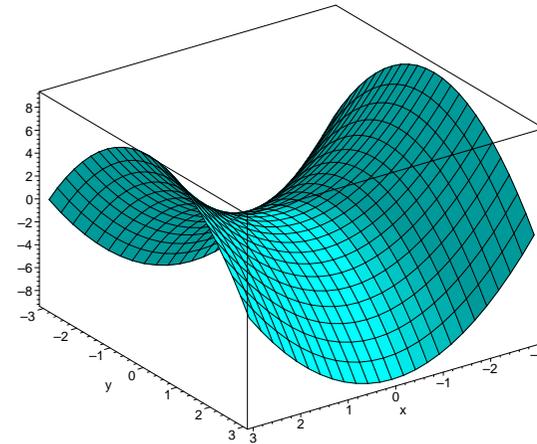
Daher muß für jeden Extremwert $\text{grad } f(x_0)$ gleich dem Nullvektor sein, denn setzt man für h ein kleines Vielfaches $t \cdot \text{grad } f(x_0)$ des Gradienten ein, wäre sonst

$$f(x_0 + h) = f(x_0) + t(\text{grad } f(x_0) \cdot \text{grad } f(x_0)) + o(\|h\|)$$

für kleine positive t größer als $f(x_0)$ und für kleine negative t kleiner.

Die Frage, welche Nullstellen des Gradienten wirklich Extremwerten entsprechen, ist schwieriger; in der Praxis wird es oft am einfachsten sein, sich die Umgebung des betreffenden Punktes mit irgendwelchen *ad hoc*-Methoden genauer anzusehen und dann zu entscheiden.

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



Graph der Funktion $f(x, y) = x^2 - y^2$

Wie wir bald sehen werden, kann man auch hier mit den (noch zu definierenden) höheren Ableitungen auch hinreichende Bedingungen finden; diese sind allerdings schon für Funktionen von zwei oder drei Veränderlichen deutlich aufwendiger als im Falle einer Veränderlichen.

f) Extremwerte unter Nebenbedingungen

Bei einem realen physikalischen, technischen wirtschaftlichen oder sozialen Prozeß können sich die Variablen selten frei im gesamten \mathbb{R}^n bewegen: Sinnvoll und realistisch ist meist nur eine beschränkte Teilmenge. Im Gegensatz zur Dimension eins, wo diese Teilmenge praktisch immer ein Intervall ist, kann sie im Mehrdimensionalen durch Randbedingungen aller Art charakterisiert sein und damit auch beliebig kompliziert aussehen.

Maxima und Minima auf solchen Teilmengen 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 im vorigen Abschnitt diskutiert haben, solche Extremwerte üblicherweise nicht finden. Wir brauchen daher neue Werkzeuge, und die soll dieser Paragraphen bereitstellen.

Die Ausgangslage ist typischerweise die folgende: Gegeben ist eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 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 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ gegeben ist. Falls wir uns für Extremwerte auf einer Kugel vom Radius r um den Nullpunkt interessieren, wäre dies etwa die Funktion

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

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 anwenden können:

Definition: Wir sagen, die Funktion $f: D \rightarrow \mathbb{R}$ auf einer Teilmenge $D \subseteq \mathbb{R}^n$ habe im Punkt $a \in D$ ein lokales $\begin{cases} \text{Maximum} \\ \text{Minimum} \end{cases}$ unter der Nebenbedingung $g = 0$, wobei $g: D \rightarrow \mathbb{R}$ eine weitere Funktion ist, wenn $g(a) = 0$ ist und es eine Umgebung U von a gibt, so daß für alle $x \in U$ gilt: Ist $g(x) = 0$, so ist $f(x) \begin{cases} \leq \\ \geq \end{cases} f(a)$.

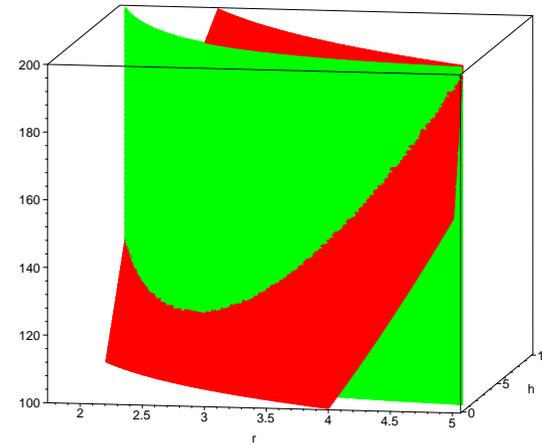
Als Einstiegsbeispiel betrachten wir eine beliebige Schulbuchaufgabe zur Minimumsbestimmung: Eine Konservendose soll bei einem vorgegebenen Volumen von 100 cm^3 möglichst wenig Blech benötigen, d.h. ihre Oberfläche soll minimal sein.

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

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

wegen der Nebenbedingung für das Volumen $V = \pi r^2 h$ muß gelten

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



Oberfläche einer Konservendose mit festem Volumen

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 gilt

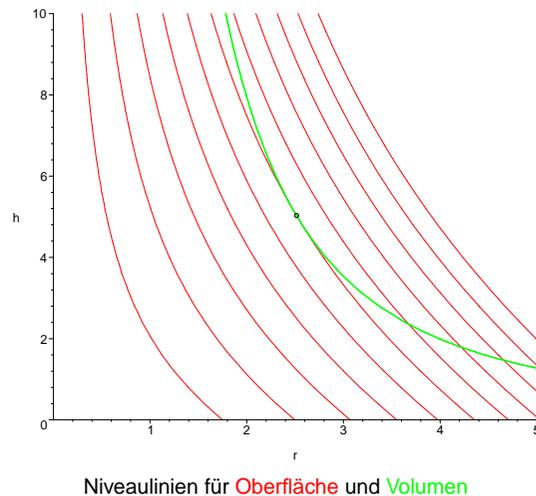
$$4\pi r^3 = 200 \quad \text{oder} \quad r = \sqrt[3]{\frac{50}{\pi}} .$$

In diesem einfachen Fall konnten wir die Optimierung also zurückführen auf gewöhnliche Extremwertaufgaben einer Veränderlichen, indem wir die Nebenbedingung nach einer der Variablen auflösten und diese dann in f einsetzten. Im allgemeinen wird dies aber 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.

In der obigen Abbildung ist die Nebenbedingung als grüne Fläche dargestellt und der Graph von f als rote; 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 schmaler oder kürzer macht.

Die Grundidee für ein alternatives Verfahren wird klar bei der Betrachtung der Niveaulinien: Die Niveaulinie für $g = 0$ ist grün eingezeichnet, verschiedene Niveaulinien von f als rote 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 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 $a \in D$ ein Extremum hat unter der Nebenbedingung $g(x) = 0$, so sind $\text{grad } f(a)$ und $\text{grad } g(a)$ 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 $\text{grad } g(a)$ der Nullvektor ist, gibt es nichts mehr zu beweisen, denn jede Menge, die den Nullvektor enthält, ist linear abhängig.

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

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

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

Nachdem f in a ein lokales Extremum unter der Nebenbedingung $g = 0$ hat, nimmt die Funktion

$$F(x_1, \dots, x_{n-1}) \stackrel{\text{def}}{=} f(x_1, \dots, x_{n-1}, h(x_1, \dots, x_{n-1}))$$

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

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

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

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

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

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

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

oder, anders ausgedrückt,

$$\text{grad } f(a) - \frac{f_{x_n}(a)}{g_{x_n}(a)} \text{grad } g(a) = 0.$$

Damit sind die beiden Gradienten in der Tat linear abhängig. ■

Falls der Gradient von g im Punkt a nicht verschwindet, gibt es somit eine Zahl $\lambda \in \mathbb{R}$, so daß

$$\text{grad } f(a) - \lambda \text{grad } g(a) = 0$$

ist, nämlich

$$\lambda = \frac{f_{x_n}(a)}{g_{x_n}(a)}.$$

Diese Zahl bezeichnet man als LAGRANGESchen Multiplikator; mit seiner inhaltlichen Interpretation werden wir uns in Kürze beschäftigen.



JOSEPH-LOUIS LAGRANGE (1736–1813) wurde als GIUSEPPE LODOVICO LAGRANGIA in Turin geboren und studierte dort zunächst Latein. Erst eine alte Arbeit von HALLEY über algebraische Methoden in der Optik weckte sein Interesse an der Mathematik, woraus ein ausgedehnter Briefwechsel mit EULER entstand. In einem Brief vom 12. August 1755 berichtete er diesem unter anderem über seine Methode zur Berechnung von Maxima und Minima; 1756 wurde er, auf EULERS Vorschlag, Mitglied der Berliner Akademie; zehn Jahre später zog er nach Berlin und wurde dort EULERS Nachfolger als mathematischer Direktor der

Akademie. 1787 wechselte er an die Pariser Académie des Sciences, wo er bis zu seinem Tod blieb und unter anderem an der Einführung des metrischen Systems beteiligt war. Seine Arbeiten umspannen weite Teile der Analysis, Algebra und Geometrie.

Zur praktischen Bestimmung von Extremwerten unter Nebenbedingungen geht man wie folgt vor: Über die Punkte, in denen der Gradient von g verschwindet, macht obiger Satz keine verwertbare Aussage; diese Punkte müssen also vorab berechnet und untersucht werden.

Danach müssen die Punkte gefunden werden, in denen es ein $\lambda \in \mathbb{R}$ gibt, so daß

$$\begin{aligned} f_{x_1}(x) - \lambda g_{x_1}(x) &= 0 \\ &\vdots \\ f_{x_n}(x) - \lambda g_{x_n}(x) &= 0 \\ g(x) &= 0 \end{aligned}$$

ist. Mit der LAGRANGE-Funktion

$$\mathcal{L}(x_1, \dots, x_n, \lambda) = f(x) - \lambda g(x)$$

läßt sich dies auch kurz schreiben als $\nabla \mathcal{L}(x_1, \dots, x_n, \lambda) = 0$, denn die partiellen Ableitungen von \mathcal{L} sind abgesehen vom Vorzeichen der Ableitung nach λ gerade die linken Seiten diesen Gleichungssystems.

Dieses ist ein System von $n + 1$ Gleichungen für $n + 1$ Unbekannte, allerdings ist es nur selten linear und damit oft nicht mit den uns bekannten Methoden lösbar. Manchmal kann man das Gleichungssystem durch geeignete Umformungen und Fallunterscheidungen vollständig lösen, in anderen Fällen helfen nur die aus der Numerik bekannten Näherungsverfahren wie etwa die Methode von NEWTON-RAPHSON.

Falls alle Gleichungen Polynomgleichungen sind (oder durch Einführung geeigneter zusätzlicher Variablen auf Polynomgleichungen zurückgeführt werden können), kann man im Falle einer endlichen Lösungsmenge diese auch exakt bestimmen: Genau wie der GAUSS-Algorithmus zur Lösung eines linearen Gleichungssystems dieses auf eine Treppengestalt bringt, aus der man die Lösungen einfach ermitteln kann, gibt es in der Computeralgebra einen Algorithmus, der dasselbe für beliebige Systeme von Polynomgleichungen versucht; die Gleichungen, die dieser Algorithmus liefert, bezeichnet man als GRÖBNER-Basis oder Standardbasis. Zum Verständnis dieses Algorithmus, den man als eine Art Synthese aus EUKLIDischen Algorithmus und GAUSS-Algorithmus ansehen kann, sind Kenntnisse der kommutativen Algebra erforderlich, für die die Zeit in dieser

Vorlesung nicht ausreicht; bei einigen Implementierungen werden zusätzlich auch noch Algorithmen aus der Informatik eingesetzt, die typischerweise nicht in Grundvorlesungen behandelt werden. Deshalb sei hier nur darauf hingewiesen, daß die gängigen universellen Computeralgebrasysteme wie Maple, Mathematica, MuPad allesamt entsprechende Routinen enthalten, mit denen man auch dann experimentieren kann, wenn man die dahinter stehende Theorie nicht versteht.

Als Beispiel, wie gelegentlich auch ein nichtlineares Gleichungssystem elementar gelöst werden kann, betrachten wir eine Anwendung aus den Wirtschaftswissenschaften: Die Gesamtproduktion eines Unternehmens oder eines Staats in Abhängigkeit von n eingesetzten Ressourcen x_1, \dots, x_n wird oft modelliert durch eine sogenannte COBB-DOUGLAS-Funktion der Form

$$P(x_1, \dots, x_n) = \alpha x_1^{e_1} \dots x_n^{e_n},$$

benannt nach den beiden Wissenschaftlern, die dieses Modell 1928 für die amerikanische Gesamtproduktion in Abhängigkeit von Kapital und Arbeit in den Jahren 1899 bis 1922 entwickelten. (Sie fanden $P \approx 1,01A^{3/4}K^{1/4}$ mit A = Anzahl der Beschäftigten und K = Kapitaleinsatz.)

Betrachten wir stattdessen die Produktion eines Wirtschaftsguts aus zwei Ressourcen x, y gemäß der Funktion

$$f(x, y) = P(x, y) = x^{1/2}y^{1/4}.$$

Falls wir der Einfachheit halber annehmen, daß die Kosten pro Einheit für x und y gleich sind und die Gesamtkosten höchstens gleich zwölf sein dürfen, müssen wir f maximieren unter der Nebenbedingung

$$x + y \leq 12.$$

Nun ist aber f eine monoton wachsende Funktion sowohl von x als auch von y , d.h. die maximale Produktion wird sicherlich erreicht in einem Punkt, für den $x + y = 12$ ist, denn für jeden anderen Punkt (x, y) mit $x + y < 12$ ist $f(x, y) < f(x, 12 - x)$. Daher können wir die Nebenbedingung in der gewohnten Form

$$g(x, y) = x + y - 12 = 0$$

schreiben. Diese Nebenbedingung sowie die zu maximierende Funktion sind in nächsten Abbildung dargestellt.

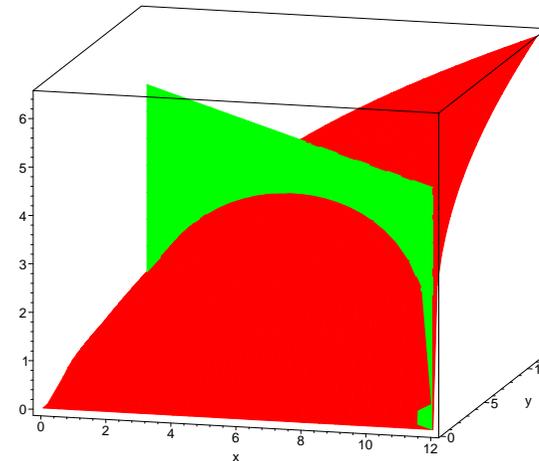
Ableitung beider Funktionen zeigt, daß

$$\text{grad } g = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{und} \quad \text{grad } f = \begin{pmatrix} y^{1/4}/2x^{1/2} \\ x^{1/2}/4y^{3/4} \end{pmatrix}$$

ist; das zu lösende Gleichungssystem wird also zu

$$\begin{aligned} \frac{y^{1/4}}{2x^{1/2}} - \lambda &= 0 \\ \frac{x^{1/2}}{4y^{3/4}} - \lambda &= 0 \\ x + y - 12 &= 0 \end{aligned}$$

(Die Nenner brauchen uns nicht zu stören, denn da $f(0, y) = f(x, 0) = 0$ ist, kommen Lösungen mit $x = 0$ oder $y = 0$ für das Maximum ohnehin nicht in Betracht; wir können sie also getrost ausschließen.)



Maximierung einer Produktionsfunktion bei festem Kapitaleinsatz

Als Ansatz zu einer möglichen Lösung können wir ausnutzen, daß λ in den beiden ersten Gleichungen isoliert steht; wenn wir danach auflösen und gleichsetzen, erhalten wir die Gleichung

$$\frac{y^{1/4}}{2x^{1/2}} = \frac{x^{1/2}}{4y^{3/4}}.$$

Multiplikation mit dem Hauptnenner macht daraus

$$4y^{1/4}y^{3/4} = 2x^{1/2}x^{1/2} \quad \text{oder} \quad 2y = x.$$

Einsetzen in die dritte Gleichung ergibt $3y = 12$, also ist

$$y = 4 \quad \text{und} \quad x = 8;$$

der Maximalwert von f ist

$$f(8, 4) = 8^{1/2} \cdot 4^{1/4} = 2\sqrt{2} \cdot \sqrt{2} = 4.$$

Auch den LAGRANGESchen Multiplikator λ können wir noch ausrechnen:

$$\lambda = \frac{y^{1/4}}{2x^{1/2}} = \frac{4^{1/4}}{2 \cdot 8^{1/2}} = \frac{\sqrt{2}}{2 \cdot 2\sqrt{2}} = \frac{1}{4}.$$

Die Berechnung von λ war für die Bestimmung des Optimums eigentlich überflüssig; λ ist nur eine Hilfsgröße zur Berechnung des Extremums. Wir wollen uns als nächstes überlegen, daß wir λ auch inhaltlich interpretieren können: Dazu betrachten wir eine Nebenbedingung

$$g(x_1, \dots, x_n) = c$$

mit *variabler* rechter Seite c und ein Extremum der Funktion

$$f(x_1, \dots, x_n).$$

Dieses Extremum wird natürlich von c abhängen; wir schreiben es in der Form

$$(x_1(c), \dots, x_n(c))$$

und nehmen an, daß die Funktionen $x_i(c)$ stetig differenzierbar seien. (Ein interessierter Leser kann sich anhand des Satzes über implizite Funktionen überlegen, welche Bedingungen f und g erfüllen müssen, damit dies garantiert ist.) Der Optimalwert von f in Abhängigkeit von c ist dann

$$F(c) \stackrel{\text{def}}{=} f(x_1(c), \dots, x_n(c)).$$

Nach der Kettenregel ist

$$F'(c) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i(c)}{dc}.$$

Genauso folgt für $G(c) \stackrel{\text{def}}{=} g(x_1(c), \dots, x_n(c))$ die Gleichung

$$G'(c) = \sum_{i=1}^n \frac{\partial g}{\partial x_i} \frac{dx_i(c)}{dc}.$$

Da $(x_1(c), \dots, x_n(c))$ ein Optimum ist, sind dort die Gradienten von f und $g - c$ proportional mit Proportionalitätsfaktor λ , und da wir bei der Gradientenbildung nur nach den x_i ableiten, von denen die rechte Seite c nicht abhängt, ist der Gradient von $g - c$ gleich dem von g selbst, d.h.

$$\frac{\partial f}{\partial x_i} = \lambda \frac{\partial g}{\partial x_i} \quad \text{für alle } i.$$

Somit ist $F'(c) = \lambda G'(c)$. Nun erfüllt aber der Punkt $(x_1(c), \dots, x_n(c))$ die Nebenbedingung mit rechter Seite c , d.h. $G(c) = c$ und $G'(c) \equiv 1$. Damit ist $\lambda = F'(c)$ die Wachstumsrate für das Optimum bei Änderung der rechten Seite der Nebenbedingung.

Im obigen Beispiel steigt also die Maximallmenge $f(x, y)$, die mit Kapitaleinsatz 12 produziert werden kann, für kleines h ungefähr um $h/4$, wenn wir den Kapitaleinsatz auf $12 + h$ erhöhen. Die Erhöhung des Kapitaleinsatzes lohnt sich, wenn für das fertige Produkt ein Preis pro Einheit erzielt werden kann, der größer ist als vier.

Als letztes wollen wir uns noch überlegen, was passiert, wenn wir nicht nur eine, sondern mehrere Nebenbedingungen erfüllen müssen. Es geht also wieder darum, eine Funktion $f(x_1, \dots, x_n)$ zu optimieren, jetzt aber unter den Nebenbedingungen

$$g_1(x_1, \dots, x_n) \geq 0, \quad \dots \quad g_r(x_1, \dots, x_n) \geq 0.$$

(Es genügt, Bedingungen mit \geq zu betrachten, denn durch Multiplikation mit minus Eins kann man jede Ungleichung mit \leq in eine mit \geq überführen. Auch Gleichungen $g_i = 0$ kann man zumindest formal durch die beiden Ungleichungen $g_i \geq 0$ und $-g_i \geq 0$ ausdrücken.)

Die wichtigsten Beispiele solcher Optimierungsaufgaben sind die Fälle mit linearen Funktionen f und g_i ; hier redet man von *linearen Programmen*. (Das Wort *Programm* in diesem Zusammenhang hat natürlich