Kapitel 2
Mehrdimensionale Analysis

Im letzten Kapitel hatten wir *lineare* Funktionen zwischen Vektorräumen betrachtet, insbesondere also auch Funktionen von $\mathbb{R}^n$ nach $\mathbb{R}^m$. Um solche Funktionen soll es auch in diesem Kapitel gehen, allerdings lassen wir nun die Forderung der Linearität fallen und verlangen nur noch deutlich schwächere Eigenschaften wie etwa Stetigkeit und/oder Differenzierbarkeit.

§1: Funktionen und ihre Eigenschaften

a) Darstellungsmöglichkeiten für Funktionen

Bei den linearen Funktionen im vorigen Kapitel hatten wir sowohl die Argumente als auch die Bilder als *Vektoren* aufgefaßt. Zumindest bei den Argumenten entspricht dies definitiv nicht der Betrachtungsweise der Analysis oder auch der Geometrie: Wir wollen Funktionen in *Punkten* auswerten, nicht in Vektoren. Wir fassen daher bei einer Funktion $f\colon \mathbb{R}^n \to \mathbb{R}^m$ den $\mathbb{R}^n$ nicht als Vektorraum auf, sondern als Punktmenge, wie wir es schon aus Kapitel I, §3*g)* von den affinen Räumen her gewohnt sind. Wie dort wollen wir aber Punkte mit Vektoren verknüpfen, wobei der Punkt $\boldsymbol{x} + \vec{h}$ der Endpunkt des im Punkt $\boldsymbol{x}$ abgetragenen Vektors $\vec{h}$ sein soll. Seine i-te Koordinate ist also $x_i + h_i$, wobei x_i die i-te Koordinate von $\boldsymbol{x}$ und h_i die i-te Komponente von $\vec{h}$ bezeichnet.

Für Funktionen $f\colon \mathbb{R} \to \mathbb{R}$ sind wir gewohnt, deren Graphen

$$\Gamma_f \underset{\text{def}}{=} \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; y = f(x)\right\}$$

zu zeichnen; entsprechend können wir natürlich auch für eine Funktion $f\colon D \to \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ den Graphen

$$\Gamma_f \underset{\text{def}}{=} \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^m \;\middle|\; \boldsymbol{y} = f(\boldsymbol{x})\right\}$$

definieren; da dieser in $\mathbb{R}^{n+m}$ liegt, ist er allerdings nur für $n + m \leq 3$ wirklich anschaulich, und auch da kann es bei komplizierten Funktionen stark von der gewählten Perspektive abhängen, was man sieht. Für einfache reellwertige Funktionen zweier Veränderlicher jedoch ist der Graph sicherlich die einfachste Methode der Veranschaulichung. Beim Graphen der Funktionen

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

in Abbildung 25 etwa sieht man recht gut, daß Γ_f Teil einer Kugeloberfläche ist.

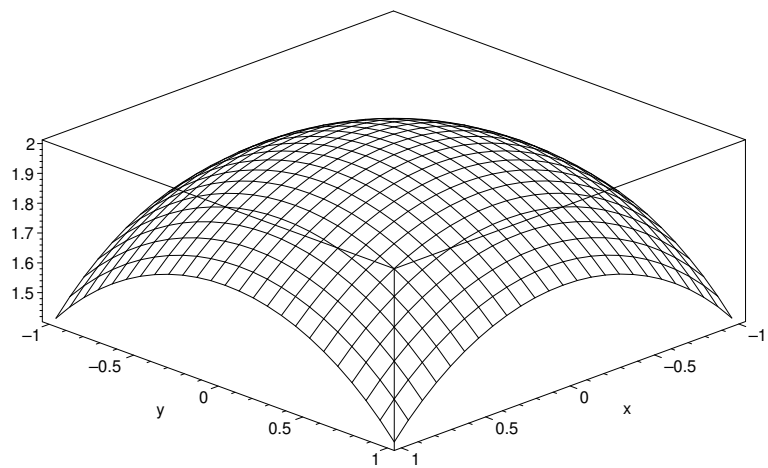


Abb. 25: Graph der Funktion $f(x, y) = \sqrt{4 - x^2 - y^2}$

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 $f\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^2$ und jeden Wert $c \in \mathbb{R}$ die *Niveaulinie*

$$N_c(f) \underset{\text{def}}{=} \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; f(x,y) = c\right\}$$

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 Wert c aus dem gesamten Definitionsbereich D der Funktion 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 Abbildung 26 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.

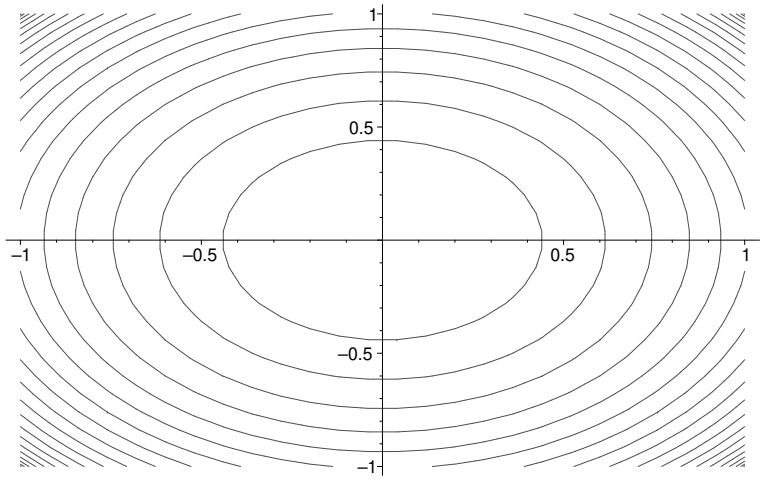


Abb. 26: Niveaulinien der Funktion $f(x,y) = \sqrt{4 - x^2 - y^2}$

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

Ansonsten gibt es für Funktionen von mehr als zwei Veränderlichen beispielsweise die Möglichkeit, einen Teil der Variablen auf interessanten Werten festzuhalten und die so eingeschränkte Funktion darzustellen. Dies gibt natürlich kein vollständiges Bild der Funktion, aber mehrere geschickt ausgewählte solche Bilder können doch einen recht guten Eindruck von der Funktion vermitteln.

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. Grundsätzlich kann man mit Farben auch mehr als eine Dimension darstellen, da wir ja in Kapitel I, §4*d)* gesehen haben, daß Farben durch Punkte eines dreidimensionalen Raums beschrieben werden können. Zwar wird eine RGB-Darstellung von drei Dimensionen die meisten Betrachter überfordern, aber die Farbdarstellung zweier Dimensionen etwa durch eine Luminanz- und eine Chrominanzkoordinate ist durchaus anschaulich.

Ein eigenes Forschungsgebiet der Mathematik und Informatik, die Visualisierung, beschäftigt sich mit den Problem, die für eine bestimmte Fragestellung interessanten Aspekte einer (analytisch oder empirisch gegebenen) Funktion mehrerer Veränderlichen graphisch herauszuarbeiten.

b) Normierte Vektorräume

Mit Hilfe von klassischen wie auch HERMITEschen Skalarprodukten konnten wir die Länge $|\vec{v}| = \sqrt{\vec{v} \cdot \vec{v}}$ eines Vektors definieren und damit

beispielsweise auch Abstände zwischen Punkten eines affinen Raums –sofern der zugehörige Vektorraum EUKLIDisch oder HERMITEsch ist. Manchmal konnt es nur auf diese Längen an, nicht auf die Produkte; daher wollen wir diese hier für sich betrachten:

Sie haben folgende Eigenschaften:

a) $|\lambda\vec{v}| = |\lambda|\,|\vec{v}|$ für alle $\lambda \in \mathbb{R}$ und $\vec{v} \in V$

b) $|\vec{v}| \geq 0$ für alle $\vec{v} \in V$, und $|\vec{v}| = 0$ genau dann, wenn $\vec{v} = \vec{0}$

c) $|\vec{v} + \vec{w}| \leq |\vec{v}| + |\vec{w}|$.

Abgesehen von *c),* der *Dreiecksungleichung,* sind diese Eigenschaften klar; wegen

$$|\vec{v} + \vec{w}|^2 = |\vec{v}|^2 + |\vec{w}|^2 + 2\vec{v}\vec{w} \qquad \text{und}$$
$$(|\vec{v}| + |\vec{w}|)^2 = |\vec{v}|^2 + |\vec{w}|^2 + 2\,|\vec{v}|\,|\vec{w}|$$

folgt letztere im Falle $\vec{v} \cdot \vec{w} \geq 0$ sofort aus der CAUCHY-SCHWARZschen Ungleichung und für $\vec{v} \cdot \vec{w} < 0$ aus der Nichtnegativität der Norm.

Definition: Ein normierter Vektorraum ist ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum V zusammen mit einer Abbildung $\|\cdot\| : V \to \mathbb{R}$ mit den Eigenschaften *a)* bis *c).*

Damit ist also jeder EUKLIDische oder HERMITEsche Vektorraum insbesondere auch ein normierter Vektorraum, es gibt aber auch normierte Vektorräume, deren Norm nichts mit einem Skalarprodukt zu tun hat:

Beispielsweise erfüllt die *Maximumsnorm*

$$\|\cdot\|_\infty : \mathbb{R}^n \to \mathbb{R}; \qquad \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \mapsto \max_{i=1}^{n} |v_i|$$

offensichtlich die Bedingungen *a)* bis *c).* Sie kommt aber nicht von einem Skalarprodukt, denn gäbe es ein Skalarprodukt $\star$ mit $\|\vec{v}\|_\infty = \sqrt{\vec{v} \star \vec{v}}$; so wäre etwa im $\mathbb{R}^2$

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

also

$$1 = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_\infty^2 = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \star \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = 1+1+2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \star \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

und somit $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \star \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{1}{2}$ und $\begin{pmatrix} 2 \\ 0 \end{pmatrix} \star \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -1$. Mithin wäre

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

ein offensichtlicher Widerspruch.

Maximumsnormen lassen sich nicht nur für $\mathbb{R}^n$ oder $\mathbb{C}^n$ definieren, sondern auch für Funktionenräume: Eine stetige Funktion $f : [a,\, b] \to \mathbb{R}$ auf einem *abgeschlossenen* Intervall $[a,\, b]$ nimmt ihr Maximum wirklich an, d.h die Abbildung

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

ist wohldefiniert (für eine stetige Funktion f ist auch $|f|$ eine stetige Funktion) und sie hat die Eigenschaften *a)* bis *c): a)* und *b)* sind, wie in den meisten Fällen, trivial, und sind $f, g \in \mathcal{C}^0\big([a,\, b],\ \mathbb{R}\big)$ zwei Funktionen, von denen f ihr Maximum in x_1 annimmt, g in x_2 und $f+g$ in x^*, so ist

$$\begin{aligned} \|f+g\|_\infty &= \big|(f+g)(x^*)\big| = \big|f(x^*) + g(x^*)\big| \leq \big|f(x^*)\big| + \big|g(x^*)\big| \\ &\leq \max_{x \in [a,\, b]} |f(x)| + \max_{x \in [a,\, b]} |g(x)| = f(x_1) + g(x_2) \\ &= \|f\|_\infty + \|g\|_\infty\ . \end{aligned}$$

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

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

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

Betrachten wir auch $\mathbb{R}^m$ mit der Maximumsnorm, so folgt für ein Produkt $A\vec{v} = \vec{b}$ sofort aus der Multiplikationsregel

$$b_i = \sum_{j=1}^{m} a_{ij} v_j$$

und der gewöhnlichen Dreiecksungleichung aus $\mathbb{R}$

$$|b_i| \leq \sum_{j=1}^{m} |a_{ij}| \cdot |v_j| \ ,$$

daß $\|\vec{b}\|_\infty \leq \|A\|_\infty \cdot \|\vec{v}\|_\infty$ ist. Wird also der Vektor $\vec{v}$ durch einen Fehlervektor $\vec{\epsilon}$ gestört, so ist

$$A(\vec{v} + \vec{\epsilon}) = A\vec{v} + A\vec{\epsilon} = \vec{b} + A\vec{\epsilon}$$

mit einem Fehler behaftet, dessen Komponenten *mit Sicherheit* kleiner sind als $\|A\|_\infty \cdot \|\vec{\epsilon}\|_\infty$. Entsprechend ändert sich bei einem eindeutig lösbaren linearen Gleichungssystem $A\vec{x} = \vec{b}$ die Lösung $\vec{x} = A^{-1}\vec{b}$ *höchstens* um $\|A^{-1}\|_\infty \cdot \|\vec{\epsilon}\|_\infty$, wenn die rechte Seite durch einen Vektor $\vec{\epsilon}$ 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.)

Ebenfalls eine sehr wichtige Rolle spielen Normen bei Funktionenräumen; dafür werden wir im nächsten Semester zahlreiche Beispiele sehen. Der wesentliche Punkt ist, daß man bezüglich einer Norm in offensichtlicher Verallgemeinerung der klassischen Definitionen Begriffe wie Konvergenz und Stetigkeit definieren kann:

Definition: *a)* $(V, \|\cdot\|)$ sei ein normierter $\mathbb{R}$-Vektorraum. Eine Folge $\vec{v}_1, \vec{v}_2, \ldots$ von Vektoren aus V konvergiert gegen den Vektor $\vec{v} \in V$, wenn es zu jedem $\varepsilon > 0$ eine natürliche Zahl $N \in \mathbb{N}$ gibt, so daß $\|\vec{v} - \vec{v}_n\| < \varepsilon$ für alle $n > N$.
b) Eine Abbildung $f\colon D \to W$ von der Teilmenge $D \subseteq V$ eines normierten Vektorraums $(V, \|\cdot\|_1)$ in einen normierten Vektorraum $(W, \|\cdot\|_2)$ heißt stetig in $\vec{v}_0 \in D$, wenn es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß

für alle $\vec{v} \in D$ gilt: Ist $\|\vec{v} - \vec{v}_0\| < \delta$, so ist $\|f(\vec{v}) - f(\vec{v}_0)\| < \varepsilon$.
c) f heißt stetig, wenn f in jedem Punkt $\vec{v}_0 \in D$ stetig ist.

Anschaulich betrachtet bedeutet Konvergenz, daß ich die Punkte immer näher an einen Grenzwert annähern, und Stetigkeit bedeutet, daß die Funktion keine Sprünge macht. Warum brauchen wir für so einfache und anschauliche Dingen so komplizierte Definitionen wie die obigen?

Ursprünglich, bei NEWTON und bei LEIBNIZ geht tatsächlich noch alles ohne δ und ϵ mit Definitionen, die der anschaulichen Vorstellung entsprechen, und beide hatten damit großen Erfolg: Insbesondere wurde die Physik auf eine völlig neue Grundlage gestellt und die naturwissenschaftliche Erklärung der Welt machte rasante Fortschritte.

SIR ISAAC NEWTON (1643–1727) war ab 1661 Student, ab 1669 Professor an der Universität Cambridge. Dort entwickelte er die Infinitesimalrechnung, die er 1671 in seinem Buch *De Methodis Serierum et Fluxionum* beschrieb, arbeitete über Optik, wo er unter anderem dünne Schichten und Beugungsphänomene untersuchte (NEWTONsche Ringe), entdeckte seine Bewegungsgesetze und das Gravitationsgesetz, veröffentlicht 1687 in seinem Buch *Philosophiae naturalis principia mathematica,* das von vielen als bedeutendstes wissenschaftliches Buch aller Zeiten angesehen wird. Nach zwei Nervenzusammenbrüchen ging er 1693 nach London, wo er die königliche Münze leitete.

Die neue Weltsicht einiger Naturwissenschaftler führte zu Spannungen mit der Theologie; mehrere Theologen veröffentlichten ihrerseits mehr oder weniger fundierte Kritiken an den Naturwissenschaften. Heute noch bekannt sind beispielsweise *Gulliver's travels* von JONATHAN SWIFT (1667–1745), wo im dritten Buch vor allem die Mathematiker und Naturwissenschaftler aufs Korn genommen werden.

Am berechtigtsten war die Kritik in BERKELEYs Buch „The Analyst: or a discourse addressed to an infidel mathematician" (gemeint war wahrscheinlich EDMOND HALLEY (1656?–1743), heute vor allem bekannt durch den nach ihm benannten Kometen), in dem er zeigte, wie unsicher die Grundlagen der Infinitesimalrechnung sind und wie leicht man im Umgang damit zu unsinnigen Ergebnissen kommen kann – ganz im Gegensatz, so seine Meinung, zur wissenschaftlich erheblich fundierteren und logischer aufgebauten Theologie. Kleiner Auszug: *And what are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them ghosts of departed quantities?*

Es dauerte geraume Zeit, bis die Mathematik dieser Kritik etwas Substantielles entgegensetzen konnte: Erst um 1800 gelang es CAUCHY die Analysis logisch zweifelsfrei zu

verankern in der Theorie der algebraischen Ungleichungen; seit dieser Zeit arbeiten wir mit ε und δ. Mittlerweile gibt es mit der sogenannten *nonstandard analysis* auch eine Alternative, die in der Prädikatenlogik erster Stufe verankert ist; da die Prädikatenlogik erster Stufe aber zum Beispiel keine vollständige Induktion gestattet, ist diese Alternative jedoch keineswegs einfacher, sondern hängt an recht diffizilen logischen Sätzen.

GEORGE BERKELEY (1685–1753) studierte und lehrte Theologie am Trinity College in Dublin. In der Philosophie gilt er als einer der Begründer des Empirizismus, den er als Gegenposition zum mechanistischen Materialismus der Naturwissenschaften seiner Zeit aufbaute. Als sehr streitbarer Gelehrter war er bald mit fast allen Kollegen am Trinity College verkracht; daher wurde er 1734 zum Bischof von Cloyne ernannt, weit weg von Dublin. Dort kämpfte er sowohl praktisch als auch durch eine dreibändige Streitschrift für die Rechte der Landbevölkerung. Wegen einer Ruhr-Epedemie beschäftigte er sich ab 1741 auch mit Pharmazie.

Mit den obigen Definitionen haben wir somit ein zwar solides, aber doch gelegentlich unhandliches Werkzeug, mit dem wir in jedem normierten Vektorraum Analysis betreiben können. Wir müssen aber damit rechnen, daß die Ergebnisse im allgemeinen stark von der Norm abhängen werden, und in der Tat werden wir dies im nächsten Semester eindrucksvoll sehen können. In diesem Semester allerdings interessieren wir uns vor allem für die Vektorräume $\mathbb{R}^n$, und dort ist die Situation deutlich harmloser.

Offensichtlich definieren zwei Normen denselben Konvergenzbegriff und denselben Stetigkeitsbegriff, wenn sie äquivalent sind im Sinne der folgenden Definition:

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

$$c_1 \|\vec{v}\|_1 \le \|\vec{v}\|_2 \le c_2 \|\vec{v}\|_1 \ .$$

Offensichtlich ist dann auch

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

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

Beispielsweise sind auf jedem $\mathbb{R}^n$ die EUKLIDische Norm $\|\cdot\|_2$ und die Maximumsnorm $\|\cdot\|_\infty$ äquivalent, denn

$$\|\vec{v}\|_2 = \sqrt{v_1^2 + \cdots + v_n^2} \le \sqrt{n \, \|\vec{v}\|_\infty^2} = \sqrt{n} \, \|\vec{v}\|_\infty$$

und

$$\|\vec{v}\|_\infty = \sqrt{\|\vec{v}\|_\infty^2} \le \sqrt{v_1^2 + \cdots + v_n^2} = \|\vec{v}\|_2 \ ,$$

d.h.

$$\|\vec{v}\|_\infty \le \|\vec{v}\|_2 \le \sqrt{n} \, \|\vec{v}\|_\infty \ .$$

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

$$\left\{ \boldsymbol{x} \in \mathbb{R}_n \;\middle|\; \|\boldsymbol{x}\|_\infty \le a \right\}$$

ist ein Würfel mit Kantenlänge $2a$ und

$$\left\{ \boldsymbol{x} \in \mathbb{R}_n \;\middle|\; \|\boldsymbol{x}\|_2 \le r \right\}$$

eine Kugel mit Radius r.

Tatsächlich kann man zeigen, daß in $\mathbb{R}^n$ alle Normen äquivalent sind, so daß es dort nur einen Konvergenz- und Stetigkeitsbegriff gibt.

Wir werden in $\mathbb{R}^n$ praktisch immer mit der EUKLIDischen Norm oder der Maximumsnorm arbeiten, wobei die erstere anschaulicher ist, die letztere aber meist einfacher für konkretes Nachrechnen.

Als Beispiel sei noch einmal angegeben, wie die obige Stetigkeitsdefinition aussieht für eine Funktion von $\mathbb{R}^n$ nach $\mathbb{R}$, ausgedrückt in der Maximumsnorm:

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

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

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

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

Auf $\mathbb{R}^2 \setminus \{(0,0)\}$ ist die Funktion offensichtlich stetig, denn dort ist sie nur durch 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.

Der Graph von f in einer kleinen Umgebung von $(0,0)$ ist in Abbildung 27 zu sehen. Er 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.

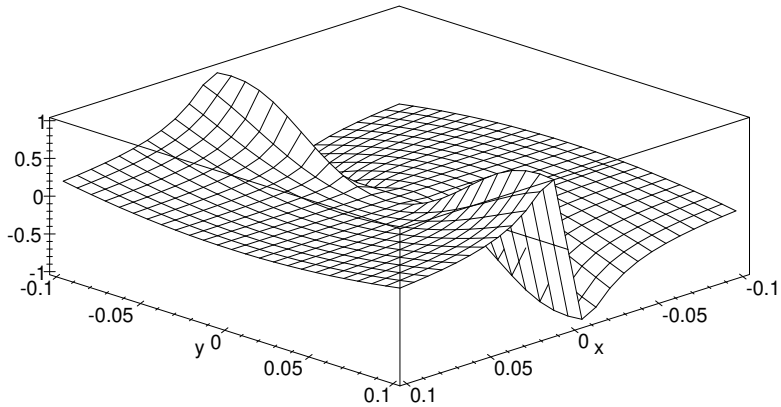


Abb. 27: Ist diese Funktion stetig?

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

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

für $x \neq 0$. Für $x \to 0$ 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 ebenfalls 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, denn für eine stetige Funktion muß schließlich

$$\lim_{(x,y)\to(0,0)} f(x,y) = f(0,0)$$

unabhängig davon sein, auf welchem Weg der Punkt (x,y) gegen $(0,0)$ wandert.

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 28. 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 $N_c(f) = \emptyset$.

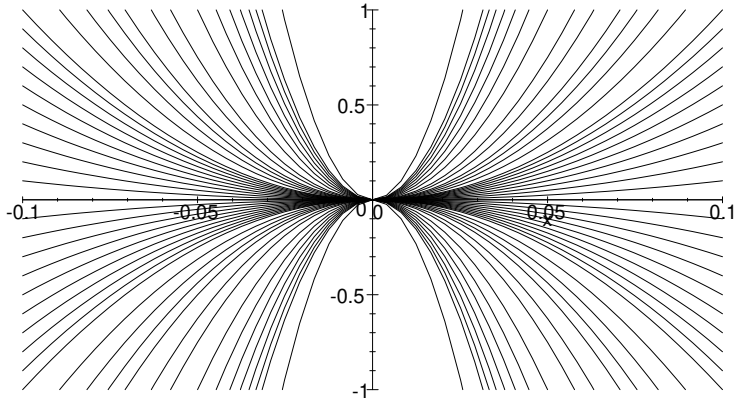


Abb. 28: Die Niveaulinien illustrieren die Unstetigkeit im Nullpunkt

Abbildung 28 zeigt, 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 Graph in Abbildung 27 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. In Abbildung 27 ist dies nicht zu sehen, da Maple, wie die meisten Computergraphikprogramme, zum Zeichnen nur die Linien $x =$ konstant und $y =$ konstant berücksichtigt.

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.

c) Die Ableitung einer Funktion

Als nächstes möchte ich auf die *Differenzierbarkeit* von Funktionen eingehen, als erstes im wohlbekannten Fall einer Funktion auf einem Intervall $(a,b) \subset \mathbb{R}$.

Rein formal betrachtet ist $f\colon (a,b) \to \mathbb{R}$ differenzierbar in einem Punkt $x \in (a,b)$, wenn der Grenzwert

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

existiert; sein Wert wird dann als Ableitung von f im Punkt x bezeichnet.

Hier und im folgenden wird oft die LANDAUsche o-Notation nützlich sein: Wir schreiben $o(h)$, sobald wir *irgendeine* uns nicht weiter interessierende Funktion von h haben, die für $h \to 0$ schneller gegen null geht als h selbst, d.h.

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

$o(h)$ 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

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

Entsprechend schreiben wir auch

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

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 boykotierten 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 das bedeutet gerade, daß

$$\lim_{h\to 0} \frac{f(x+h) - \big(f(x) + hf'(x)\big)}{h} = 0\,.$$

Anschaulich können wir das auch so interpretieren, daß für kleine h

$$f(x+h) \approx f(x) + hf'(x)\,,$$

ist, d.h. die Funktion f sieht in einer hinreichend kleinen Umgebung von x aus wie eine lineare Funktion. Die Abbildungen 29 bis 32 zeigen dies für $f(x) = \sin x$ in der Umgebung von $x = 1$, wobei diese Umgebung von einer Abbildung zur nächsten jeweils um den Faktor fünf vergrößert wird.

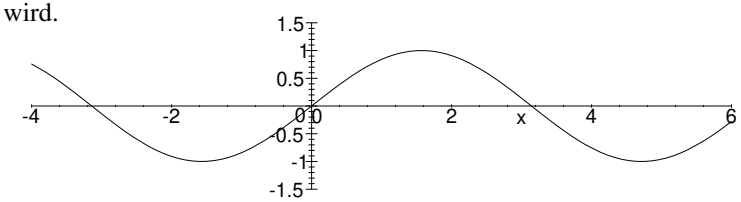


Abb. 29: Graph der Funktion $y = \sin x$

Dies läßt sich durchaus auch praktisch ausnützen, um Funktionen in der Nachbarschaft bekannter Werte näherungsweise auszurechnen: Für

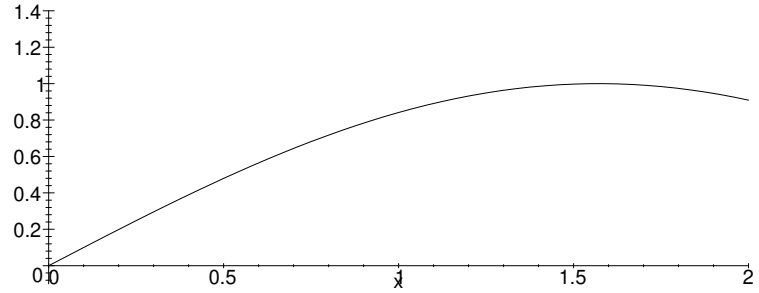


Abb. 30: Vergrößerung um den Faktor fünf um $\big(1, \sin(1)\big)$

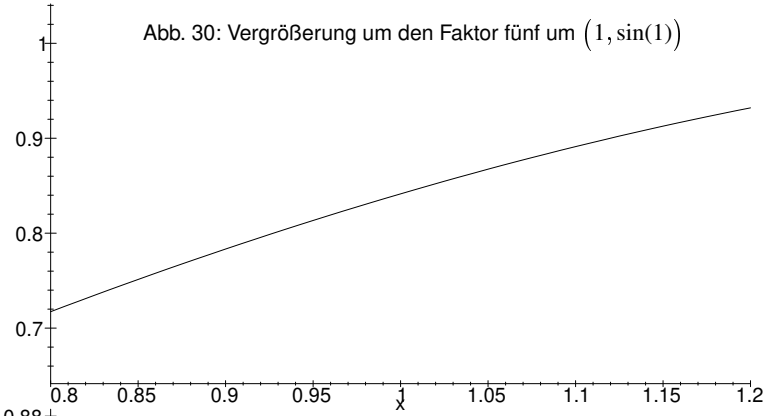


Abb. 31: Nochmalige Vergrößerung um den Faktor fünf um $\big(1, \sin(1)\big)$

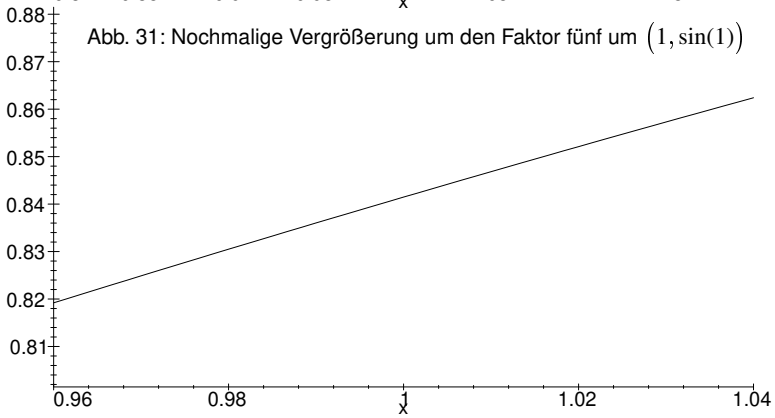


Abb. 32: Nach noch einer Vergrößerung sieht die Funktion praktisch linear aus

$f(x) = x^n$ ist $f'(x) = nx^{n-1}$, also gilt für kleine Werte von h die Näherungsformel $f(x+h) \approx f(x) + hnx^{n-1}$. Damit ist beispielsweise

$$1{,}05^2 \approx 1 + 2 \cdot 0{,}05 = 1{,}1 \quad \text{und} \quad 1{,}05^5 \approx 1 + 5 \cdot 0{,}05 = 1{,}25\,,$$

was sich nicht allzu sehr von den korrekten Werten 1,1025 und 1,276281562 unterscheidet. Für $n = \frac{1}{2}$ enthalten wir entsprechend

$$\sqrt{4{,}1} \approx \sqrt{4} + 0{,}1 \cdot \frac{1}{2\sqrt{4}} = 2{,}025\,,$$

verglichen mit dem korrekten Wert 2,0248 . . . .

Einige weitere populäre Linearisierungen um $x = 0$ oder $x = 1$ sind

$$\sin h \approx h, \quad e^h \approx 1 + h, \quad \frac{1}{1+h} \approx 1 - h\,.$$

Im Falle des Kosinus erhalten wir für kleine Werte von h wegen des Verschwindens der Sinusfunktion bei Null die Formel $\cos h \approx 1$, was nicht von h abhängt aber trotzdem eine recht gute Näherung ist: Für $h \le 0{,}14$ ist $\cos h \ge 0{,}99$, der Fehler also kleiner als ein Prozent, und erst bei $h = 0{,}451$ ist $\cos h \approx 0{,}9$, so daß wir einen Fehler von zehn Prozent bekommen.

Der wohl wichtigste Satz zum theoretischen Umgang mit differenzierbaren Funktionen ist der

Mittelwertsatz der Differentialrechnung: Für eine differenzierbare Funktion $f\colon (a,b) \to \mathbb{R}$ gibt es zu jedem $x \in (a,b)$ und jedem reelen $h \neq 0$, für das auch noch $x+h$ in (a,b) liegt, eine reelle Zahl $0 < \eta < 1$, so daß gilt:

$$\frac{f(x+h) - f(x)}{h} = f'(x + \eta h)\,.$$

Der Differenzenquotient ist also gleich dem Differentialquotient an einem Zwischenpunkt.

Zum *Beweis* betrachten die Funktion

$$\varphi\colon \begin{cases} [0,\ 1] \to \mathbb{R} \\ \quad \eta \mapsto f(x+\eta h) - \eta\big(f(x+h) - f(x)\big) \end{cases}$$

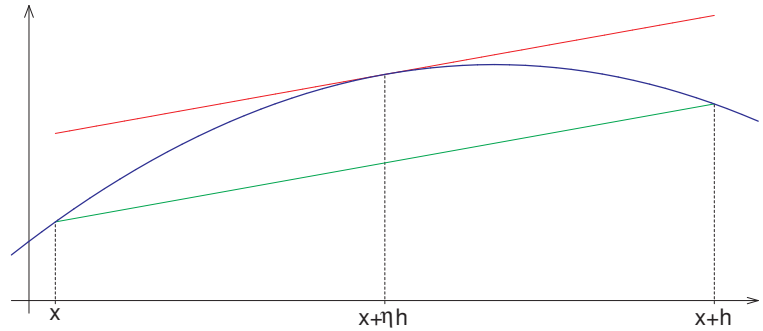


Abb. 33: An einem Zwischenpunkt ist die Tangente parallel zur Sehne

Wegen der Stetigkeit von f ist auch φ eine stetige Funktion; sie muß also im Intervall $[0,\ 1]$ mindestens ein Extremum annehmen. Da $\varphi(0) = \varphi(1) = f(x)$ ist, wird dieses Extremum in einem inneren Punkt η von $[0,\ 1]$ angenommen, und dort ist

$$\varphi'(\eta) = f'(x+\eta h) \cdot h - \big(f(x+h) - f(x)\big) = 0\,,$$

also

$$f'(x+\eta h) = \frac{f(x+h) - f(x)}{h}\,,$$

wie behauptet. ∎

Für spätere Anwendungen sei gleich noch eine Verallgemeinerung dieses Satzes formuliert:

Verallgemeinerter Mittelwertsatz der Differentialrechnung: Die beiden Funktionen $f, g\colon (a,b) \to \mathbb{R}$ seien differenzierbar auf (a,b), und g' habe dort keine Nullstelle. Dann gibt es zu jedem $x \in (a,b)$ und jedem $h \in \mathbb{R} \setminus \{0\}$, für das auch noch $x+h$ in (a,b) liegt, eine reelle Zahl $0 < \eta < 1$, so daß gilt:

$$\frac{f(x+h) - f(x)}{g(x+h) - g(x)} = \frac{f'(x+\eta h)}{g'(x+\eta h)}\,.$$

Der *Beweis* geht analog zum gewöhnlichen Mittelwertsatz: Wir betrachten die Funktion $\varphi\colon [0,\ 1] \to \mathbb{R}$ mit

$$\varphi(\eta) = f(x+\eta h)\big(g(x+h) - g(x)\big) - g(x+\eta h)\big(f(x+h) - f(x)\big)\,.$$

Einsetzen zeigt, daß

$$\varphi(0) = \varphi(1) = f(a)g(b) - f(b)g(a)$$

ist; es gibt also ein $\eta \in (0,\ 1)$, so daß

$$\varphi'(\eta) = h\Big(f'(x+\eta h)\big(g(x+h) - g(x)\big) - g'(x+\eta h)\big(f(x+h) - f(x)\big)\Big)$$

verschwindet. h ist nach Voraussetzung von Null verschieden und dasselbe gilt für $g(x+h) - g(x)$, denn sonst müßte – wie wir gerade im Beweis des Mittelwertsatzes gesehen haben – g' zwischen x und $x+h$ eine Nullstelle haben, was im Satz ausgeschlossen wurde. Somit können wir die Gleichung $\varphi'(\eta) = 0$ umformen zur gewünschten Formel

$$\frac{f(x+h) - f(x)}{g(x+h) - g(x)} = \frac{f'(x+\eta h)}{g'(x+\eta h)}\,.$$

■

Um Differenzierbarkeit für Funktionen mehrerer Veränderlicher zu definieren, können wir ähnlich vorgehen wie im eindimensionalen Fall. Wir betrachten zunächst eine Funktion $f\colon \mathbb{R}^2 \to \mathbb{R}$, 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 Abbildungen 34 bis 37, die zeigen, daß sich dieser 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 hat die Form

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

also ist

$$L(x+h, y+k) = a + b(x+h) + c(y+k) = (a + bx + cy) + ah + bk$$

eine Approximation für f in der Umgebung des betrachteten Punkts (x,y), hier gleich $(1,1)$. Im Punkt (x,y) sollte

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

natürlich mit $f(x,y)$ übereinstimmen, und für $(h,k) \neq (0,0)$ sollte der 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

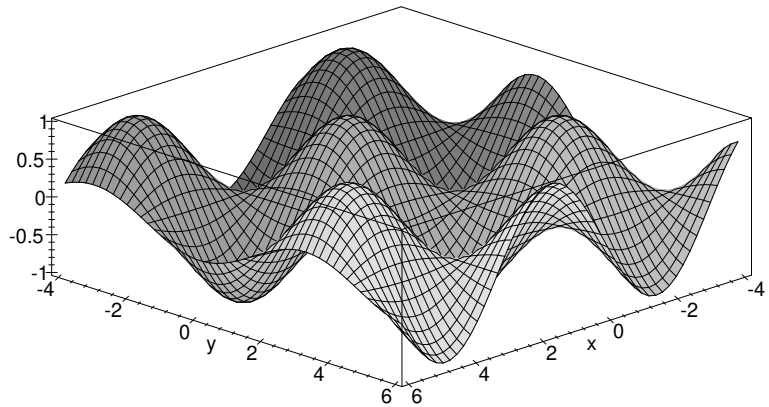


Abb. 34: Graph der Funktion $z = \sin x \cos y$

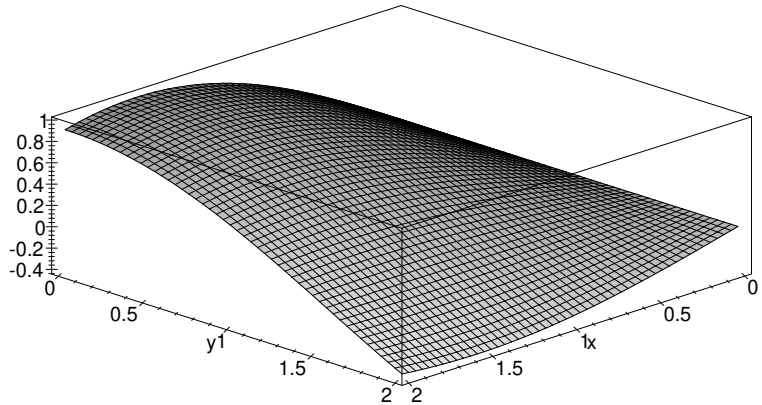


Abb. 35: Vergrößerung um den Faktor fünf um $\big(1, 1, \sin(1)\cos(1)\big)$

erwarten also, daß

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

ist, und genau so läßt sich Differenzierbarkeit allgemein definieren:

Definition: *a)* Die Funktion $f\colon D \to \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ heißt *differenzierbar* im Punkt $\boldsymbol{x} \in D$, wenn es eine lineare Funktion $L\colon \mathbb{R}^n \to \mathbb{R}^m$

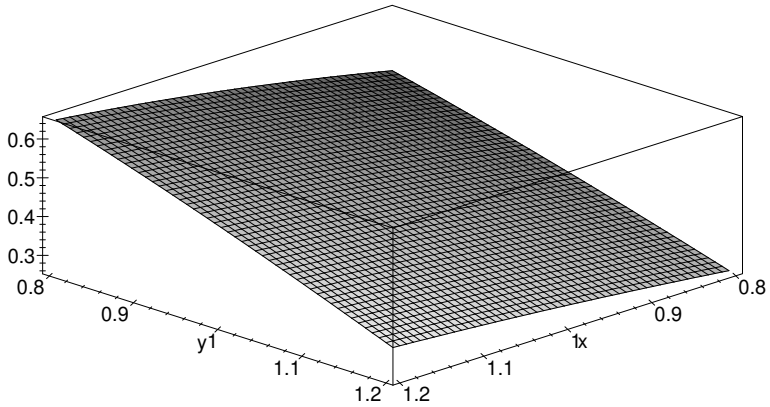


Abb. 36: Nochmalige Vergrößerung um den Faktor fünf um $\big(1, 1, \sin(1)\cos(1)\big)$

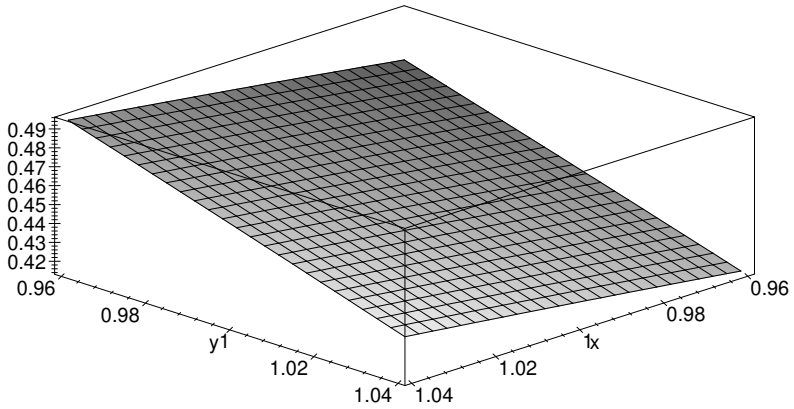


Abb. 37: Nach noch einer Vergrößerung sieht die Funktion praktisch linear aus

gibt, so daß für $\vec{h} \to 0$ gilt

$$f(\boldsymbol{x} + \vec{h}) = f(\boldsymbol{x}) + L\big(\vec{h}\big) + o(\|\vec{h}\|)\,.$$

b) Die Abbildungsmatrix $J_f(\boldsymbol{x})$ von L bezüglich der Standardbasen von $\mathbb{R}^n$ und $\mathbb{R}^m$ heißt dann JACOBI-Matrix von f im Punkt $\boldsymbol{x}$.

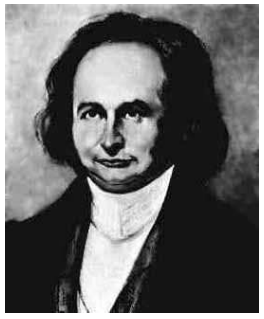
c) Für eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ kann die dann einzeilige JACOBI-Matrix mit einem Vektor aus $\mathbb{R}^n$ identifiziert werden; dieser Vektor

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

heißt *Gradient* von f im Punkt $\boldsymbol{x} \in D$.

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 $\nu\alpha\beta\lambda\alpha$ = 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 rauhe 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

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

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 $f\colon D \to \mathbb{R}^m$ 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 $f\colon \mathbb{R}^n \to \mathbb{R}^m$ ist zusammengesetzt aus m Komponentenfunktionen $f_i\colon \mathbb{R}^n \to \mathbb{R}$; wir beschränken uns daher zunächst auf Funktionen $f\colon \mathbb{R}^n \to \mathbb{R}$.

Schwieriger ist die Reduktion von n auf eins; hier bietet sich trotz der schlechten Erfahrungen beim obigen Beispiel einer unstetigen Funktion an, die Funktion auf eine Gerade einzuschränken.

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

Wenn wir die Einschränkung von f auf eine solche Gerade mit Richtungsvektor $\vec{e}$ betrachten, betrachten wir konkret die Funktion

$$g(t) = f(\boldsymbol{x} + t\vec{e}),$$

die überall dort definiert ist, wo $\boldsymbol{x} + t\vec{e}$ 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

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = \lim_{h\to 0} \frac{f(\boldsymbol{x} + h\vec{e}) - f(\boldsymbol{x})}{h}$$

als *Richtungsableitung* von f in Richtung $\vec{e}$. 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 $\vec{e}$ abhängt, sondern auch von dessen *Länge:* Beispielsweise ist für $h(t) = f(\boldsymbol{x} + 2t\vec{e})$

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

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

$$\vec{e}_i = {}^t(0, \ldots, 1, \ldots, 0),$$

bei dem in der i-ten Zeile eine Eins steht und sonst lauter Nullen.

Alsdann ist für $g_i(t) = f(\boldsymbol{x} + t\vec{e}_i)$

$$\begin{aligned} g_i'(0) &= \lim_{h\to 0} \frac{f(\boldsymbol{x} + h\vec{e}_i) - f(\boldsymbol{x})}{h} \\ &= \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h} \end{aligned}$$

die Ableitung jener Funktion, die nur von x_i abhängt, während alle anderen Koordinaten x_j festgehalten werden. Diese Ableitung, so sie existiert, bezeichnen wir als *partielle Ableitung*

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

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 $\boldsymbol{x} \in D$ differenzierbar ist und D eine Umgebung von $\vec{x}$ enthält (was für offenes D immer der Fall ist), existiert auch jede Richtungsableitung, denn da dann für jeden Vektor $\vec{h}$ gilt

$$f\left(\boldsymbol{x} + \vec{h}\right) = f(\boldsymbol{x}) + L\left(\vec{h}\right) + o(\|\vec{h}\|),$$

ist insbesondere auch

$$f\left(\boldsymbol{x}+t\vec{e}\right) = f(\boldsymbol{x}) + L\left(t\vec{e}\right) + o\left(\|t\vec{e}\|\right) = f(\boldsymbol{x}) + tL\left(\vec{e}\right) + o(t)\,;$$

denn $L\left(\vec{e}\right)$ und $\left|\vec{h}\right|$ sind schließlich Konstanten. Damit existiert

$$\frac{d}{dt} f\left(\boldsymbol{x}+t\vec{e}\right)_{|t=0} = L\left(\vec{e}\right)$$

für jeden Richtungsvektor $\vec{e}$; insbesondere existieren natürlich alle partiellen Ableitungen.

Die Umkehrung gilt leider nicht immer: Bei der (offensichtlich in $(0,0)$ unstetigen) Funktion

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

ist Null auf der x-Achse und der y-Achse, und eins überall sonst. 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 Linearform $L(h,k) = bh+ck$ geben, so daß

$$f(h,k) = f(0,0) + L(h,k) + o\left(\sqrt{h^2+k^2}\right) = o\left(\sqrt{h^2+k^2}\right)$$

ist, denn $L(0,0) = 0$ 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 $f\colon D \to \mathbb{R}$ 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

$$\operatorname{grad} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{pmatrix}.$$

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

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

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

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

und so weiter. Insgesamt erhalten wir

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

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

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 $f(\boldsymbol{x}+\vec{h})$ und $f(\boldsymbol{x})$ nicht durch eine lineare Funktion angenähert ist.

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

$$\lim_{\vec{h}\to 0}\left(f_{x_1}(x_1,x_2+h_2,\ldots,x_n+h_n) - f_{x_1}(x_1,x_2,\ldots,x_n)\right) = 0$$

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

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

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

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

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.

Nächstes Ziel dieses Abschnitts sind höhere Ableitungen von Funktionen mehrerer Veränderlicher. Wir lassen uns bei der Definition wieder vom eindimensionalen Fall leiten: Die zweite Ableitung ist die Ableitung der Ableitung.

Die Ableitung einer differenzierbaren Funktion $f\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ ist der Zeilenvektor zum Gradienten, also die Abbildung

$$D \to \mathbb{R}^n; \quad \boldsymbol{x} \mapsto \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right) ;$$

die Ableitung oder JACOBI-Matrix davon ordnet jedem Punkt aus D eine $n \times n$-Matrix zu, die wir als HESSE-Matrix $H_f(\boldsymbol{x})$ von f im Punkt $\boldsymbol{x}$ bezeichnen.

LUDWIG OTTO HESSE (1811–1874) wurde in Königsberg geboren und unterrichtete zunächst Physik und Chemie am dortigen Gymnasium. 1840 bekam er eine Stelle als Mathematiker an der dortigen Universität, von 1856 bis 1868 war er Professor in Heidelberg, danach in München. Aus der Schule ist er wohl vor allem durch die HESSEsche Normalenform der Ebenengleichung bekannt; der Schwerpunkt seiner Forschungen lag allerdings auf dem Gebiet der Invariantentheorie und der algebraischen Funktionen. Auch die HESSE-Matrix führte er 1842 in einer Arbeit über Invarianten von kubischen und biquadratischen Kurven ein.

Genau wie der Gradient bei gutartigen Funktionen durch die partiellen Ableitungen berechnet werden kann, sollte auch die HESSE-Matrix durch Differentiationsverfahren aus der Analysis einer Veränderlichen berechenbar sein. Das Hilfsmittel dazu sind die zweiten partiellen Ableitungen:

Für eine in ganz D partiell differenzierbare Funktion $f\colon D \to \mathbb{R}$ ist auch jede partielle Ableitung f_{x_i} wieder eine Funktion von D nach $\mathbb{R}$, und

auch diese kann wieder partiell differenzierbar sein. Falls ja, bezeichnen wir die partielle Ableitung von f_{x_i} nach x_j als zweite partielle Ableitung

$$f_{x_i x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \underset{\text{def}}{=} \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

von f nach x_i und x_j. Im Fall $i = j$ schreiben wir kurz

$$f_{x_i x_i} = \frac{\partial^2 f}{\partial x_i^2} .$$

Analog lassen sich auch höhere partielle Ableitungen einführen durch die Definition

$$f_{x_{i_1} x_{i_2} \dots x_{i_k}} \underset{\text{def}}{=} \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_2} \partial x_{i_1}} \underset{\text{def}}{=} \frac{\partial}{\partial x_{i_k}} \frac{\partial}{\partial x_{i_{k-1}}} \cdots \frac{\partial}{\partial x_{i_2}} \frac{\partial f}{\partial x_{i_1}} .$$

Wie wir oben gesehen haben, ist Gradient einer Funktion f gleich dem Vektor der partiellen Ableitungen, falls diese allesamt existieren und stetig sind; die JACOBI-Matrix ist der entsprechende Zeilenvektor. Falls auch die zweiten partiellen Ableitungen allesamt existieren und stetig sind, zeigt dasselbe Lemma, daß deren Ableitungen die Zeilenvektoren

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_1}, \dots, \frac{\partial^2 f}{\partial x_i \partial x_n}\right)$$

sind, d.h.

Lemma: Falls alle ersten und zweiten partiellen Ableitungen von $f\colon D \to \mathbb{R}$ existieren und stetig sind, ist die HESSE-Matrix von f gleich der $n \times n$-Matrix mit Einträgen $\frac{\partial^2 f}{\partial x_i \partial x_j}$. ■

Als erstes Beispiel können wir etwa die zweiten partiellen Ableitungen der Funktion

$$f(x,y) = x^4 + 2x^3 y + 3x^2 y^2 + 4xy^3 + 5y^4$$

berechnen: Die partielle Ableitung nach x ist

$$f_x(x,y) = 4x^3 + 6x^2 y + 6xy^2 + 4y^3 ,$$

also ist

$$f_{xx}(x,y) = \frac{\partial^2 f}{\partial x^2}(x,y) = 12x^2 + 12xy + 6y^2 \quad \text{und}$$

$$f_{xy}(x,y) = \frac{\partial^2 f}{\partial y \partial x}(x,y) = 6x^2 + 12xy + 12y^2 \,.$$

Entsprechend ist

$$f_y(x,y) = 2x^3 + 6x^2 y + 12xy^2 + 20y^3 \,,$$

also

$$f_{yx}(x,y) = \frac{\partial^2 f}{\partial x \partial y}(x,y) = 6x^2 + 12xy + 12y^2 \quad \text{und}$$

$$f_{yy}(x,y) = \frac{\partial^2 f}{\partial y^2}(x,y) = 6x^2 + 24xy + 60y^2 \,.$$

Damit ist

$$H_f(x,y) = \begin{pmatrix} 12x^2 + 12xy + 6y^2 & 6x^2 + 12xy + 12y^2 \\ 6x^2 + 12xy + 12y^2 & 6x^2 + 24xy + 60y^2 \end{pmatrix} \,.$$

Zumindest in diesem Fall ist dies eine symmetrische Matrix, d.h.

$$f_{xy} = f_{yx} \,.$$

Diese Formel gilt, wie wir gleich sehen werden, *fast* immer; in der Tat galt sie für die Mathematiker des 18. Jahrhunderts wie NICOLAUS I. BERNOULLI, der 1719 darüber schrieb, LEONARD EULER (1730), JOSEPH-LOUIS LAGRANGE (1772) und viele andere als selbstverständlich. Erst im 19. Jahrhundert, als sich ein präziser Funktionsbegriff durchzusetzen begann, wurde erkannt, daß Voraussetzungen notwendig sind. Diese waren zu Beginn des Jahrhunderts zunächst unnötig stark; 1873 untersuchte dann HERMANN AMANDUS SCHWARZ, den wir bereits von der CAUCHY-SCHWARZschen Ungleichung kennen, das Problem genauer in seiner Arbeit *Über ein System voneinander unabhängiger Voraussetzungen zum Beweis des Satzes* $\frac{\partial}{\partial y}\left(\frac{\partial f(x,y)}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f(x,y)}{\partial y}\right)$.

Als Gegenbeispiel betrachtet er die in Abbildung 38 dargestellte Funktion

$$f(x,y) = \begin{cases} y^2 \arctan \frac{x}{y} - x^2 \arctan \frac{y}{x} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases} \,.$$

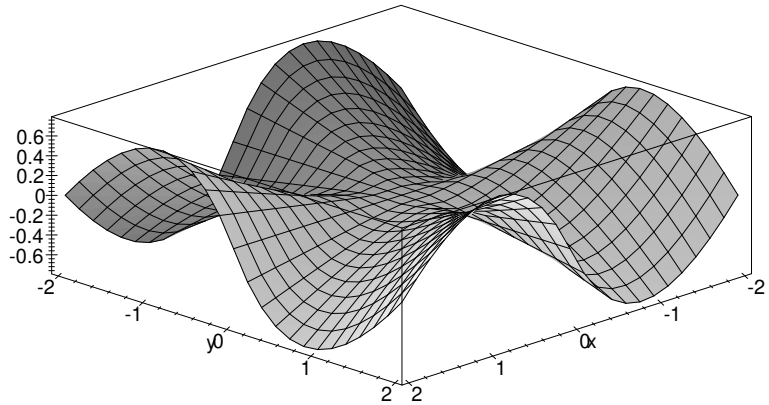


Abb. 38: Ein Gegenbeispiel zum Vertauschungssatz

Ihre ersten partiellen Ableitungen sind

$$f_x(x,y) = \begin{cases} y - 2x \arctan \frac{y}{x} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases}$$

und

$$f_y(x,y) = \begin{cases} -x + 2y \arctan \frac{x}{y} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases} \,.$$

Alle diese Funktionen sind stetig auch im Nullpunkt, denn da der Arkustangens nur Werte zwischen -1 und $+1$ annimmt, sorgt der jeweilige Vorfaktor dafür, daß das Produkt bei jeder Annäherung an den Nullpunkt gegen null geht. Daher können wir trotz der Nenner im Argument des Arkustangens $x = 0$ und $y = 0$ einsetzen und erhalten, daß

$$f_x(0,y) = y \quad \text{und} \quad f_y(x,0) = -x$$

ist, also

$$f_{xy}(0,y) = +1 \quad \text{und} \quad f_{yx}(x,0) = -1 \,.$$

Im Nullpunkt ist somit $f_{xy}(0,0) = +1 \neq -1 = f_{yx}(0,0)$.

Für Punkte $(x,y) \neq (0,0)$ rechnet man leicht nach, daß

$$f_{xy}(x,y) = f_{yx}(x,y) = \frac{y^2 - x^2}{y^2 + x^2}$$

ist. Insbesondere ist

$$f_{xy}(x,0) = -1, \quad f_{xy}(0,y) = +1 \quad \text{und} \quad f_{xy}(x,x) = 0\,;$$

f_{xy} nimmt also in jeder noch so kleinen Umgebung des Nullpunkts jeden der drei Werte $0, 1$ und -1 (und viele andere) an. Damit kann f_{xy} in $(0,0)$ nicht stetig sein, genausowenig wie f_{yx}. Wie SCHWARZ erkannte, ist genau das die fehlende Voraussetzung für die Vertauschbarkeit der partiellen Ableitungen:

Schwarzsches Lemma: $f\colon D \to \mathbb{R}$ sei auf der offenen Menge $D \subseteq \mathbb{R}^n$ erklärt und sowohl die ersten partiellen Ableitungen f_{x_i} als auch die gemischten partiiellen Ableitungen $f_{x_i x_j}$ seien stetig auf D. Dann ist

$$f_{x_i x_j}(\boldsymbol{x}) = f_{x_j x_i}(\boldsymbol{x})$$

für alle $\boldsymbol{x} \in D$ und alle i, j mit $1 \le i, j \le n$.

Beweis: Da bei der partiellen Differentiation alle Variablen außer einer als konstant betrachtet werden, können wir uns auf den Fall $n = 2$ beschränken: Wir interessieren uns nur für die beiden Variablen x_i und x_j, die wir als x und y bezeichnen (wenn sie verschieden sind – andernfalls gibt es aber ohnehin nichts zu beweisen), und betrachten alle sonstigen x_k als konstant.

Für den Punkt (x, y) aus D wählen wir dann $h, k \in \mathbb{R}$ so, daß das Quadrat mit den vier Ecken

$$(x,y),\ (x+h,y),\ (x,y+k) \quad \text{und} \quad (x+h,y+k)$$

vollständig in D liegt; dies ist möglich, da wir D als offene Menge vorausgesetzt haben. Nach Voraussetzung existieren die partiellen Ableitungen f_x, f_y, f_{xy} sowie f_{yx} und sind stetig.

Nach Definition ist

$$f_{xy}(x,y) = \lim_{k\to 0} \frac{f_x(x,y+k) - f_x(x,y)}{k}$$

$$= \lim_{k\to 0} \frac{\lim\limits_{h\to 0} \dfrac{f(x+h,y+k) - f(x,y+k)}{h} - \lim\limits_{h\to 0} \dfrac{f(x+h,y) - f(x,y)}{h}}{k}\,.$$

Falls alle Grenzübergänge miteinander vertauschbar sind (was, wie wir im obigen Gegenbeispiel gesehen haben, keineswegs selbstverständlich ist), ist das ein Limes über den Ausdruck

$$\frac{f(x+h,y+k) - f(x,y+k) - f(x+h,y) + f(x,y)}{hk}$$

für $h, k \to 0$; es liegt also nahe, sich diesen Ausdruck genauer anzuschauen.

Für den Beweis wird es genügen, wenn wir uns auf den Fall $h = k$ beschränken; wir werden den Ausdruck

$$D(h) = \frac{f(x+h,y+h) - f(x,y+h) - f(x+h,y) + f(x,y)}{h^2}$$

auf zwei Arten ausrechnen:

Zunächst fassen wir, wie oben, die beiden ersten und die beiden letzten Summanden zusammen: Mit der Abkürzung

$$g(y) = \frac{f(x+h,y) - f(x,y)}{h}$$

ist dann

$$D(h) = \frac{g(y+h) - g(y)}{h}\,.$$

Nach dem Mittelwertsatz der Differentialrechnung ist dieser Differenzenquotient gleich dem Differentialquotient $g'(\eta)$ für eine (von h abhängige) Zahl η zwischen y und $y+h$. Somit ist

$$D(h) = g'(\eta) = \frac{f_y(x+h,\eta) - f_y(x,\eta)}{h} = f_{yx}(\xi,\eta)$$

für ein η zwischen x und $x+h$, denn natürlich können wir auch auf diesen Differenzenquotienten den Mittelwertsatz anwenden.

Für die zweite Berechnung fassen wir in $D(h)$ den ersten und den dritten sowie den zweiten und den vierten Term zusammen. Mit der Abkürzung

$$\widetilde{g}(x) = \frac{f(x,y+h) - f(x,y)}{h}$$

ist dann dieses Mal

$$D(h) = \frac{\widetilde{g}(x+h) - \widetilde{g}(y)}{h}\,,$$

und nach dem Mittelwertsatz der Differentialrechnung gibt es dazu ein $\widetilde{\xi}$ zwischen x und $x+h$, so daß dies gleich $\widetilde{g}'(\widetilde{\xi})$ ist. Also ist

$$D(h) = \widetilde{g}'(\widetilde{\xi}) = \frac{f_x(\widetilde{\xi}, y+h) - f_x(\widetilde{\xi}, y)}{h} = f_{xy}(\widetilde{\xi}, \widetilde{\eta})$$

für eine Zahl $\widetilde{\eta}$ zwischen y und $y+h$. Somit ist

$$D(h) = f_{yx}(\xi, \eta) = f_{xy}(\widetilde{\xi}, \widetilde{\eta})\,.$$

Lassen wir nun h gegen Null gehen, konvergieren ξ und $\widetilde{\xi}$ gegen x und η wie auch $\widetilde{\eta}$ gegen y. Wegen der vorausgesetzten Stetigkeit der zweiten partiellen Ableitungen konvergiert daher $f_{yx}(\xi, \eta)$ gegen $f_{yx}(x, y)$ und $f_{xy}(\widetilde{\xi}, \widetilde{\eta})$ gegen $f_{xy}(x, y)$, d.h. der Grenzwert existiert und

$$D(0) = f_{yx}(x, y) = f_{xy}(x, y)\,.$$

Damit ist das Lemma bewiesen. ■

Tatsächlich bewies SCHWARZ diesen Satz (für $n = 2$) unter einer etwas schwächeren Voraussetzung: Es reicht, wenn *eine* der partiellen Ableitungen f_{xy} oder f_{yx} existiert und stetig ist. Am Beweis ändert sich wenig; falls etwa über die Ableitung f_{yx} nichts vorausgesetzt ist, muß man die Existenz aller damit zusammenhängenden Grenzwerte explizit durch Abschätzungen nachweisen und daraus nachträglich die Existenz und Stetigkeit von f_{yx} folgern. Ein Leser, der seine *Analysis I* noch nicht ganz vergessen hat, sollte dies auf etwa einer Seite tun können. Für Anwendungen ist die SCHWARZsche Formulierung etwas nützlicher als die obige, denn wenn man beispielsweise f_{xy} berechnet und seine Stetigkeit nachgewiesen hat, folgt automatisch, daß auch f_{yx} existiert und gleich f_{xy} ist. Für uns wird das keine sehr große Rolle spielen, denn bei den meisten uns interessierenden Funktionen wird die *Existenz* und Stetigkeit der partiellen Ableitungen klar sein; lediglich ihre Berechnung wird im allgemeinen mit Arbeit verbunden sein.

Ein analoger Satz zum SCHWARZschen Lemma gilt auch für höhere partielle Ableitungen; für k-fache Ableitungen müssen wir entsprechend voraussetzen, daß die partiellen Ableitungen bis zur k-fachen existieren und stetig sind (wobei diese Voraussetzung wieder strenggenommen nicht für alle k-fachen wirklich notwendig ist).

Definition: Für eine offene Teilmenge $D \subseteq \mathbb{R}^n$ bezeichne $\mathcal{C}^k(D, \mathbb{R})$ die Menge aller Funktionen $f\colon D \to \mathbb{R}$, deren sämtliche partielle Ableitungen bis zu den k-ten existieren und stetig sind. Für $k = 0$ sei $\mathcal{C}^0(D, \mathbb{R})$ einfach die Menge aller stetiger Funktionen $D \to \mathbb{R}$.

Man überlegt sich sofort, daß $\mathcal{C}^k(D, \mathbb{R})$ ein $\mathbb{R}$-Vektorraum ist, und es ist auch nicht schwer einzusehen, daß die Funktionen aus $\mathcal{C}^k(D, \mathbb{R})$ alle die Eigenschaften haben, die man bei der Betrachtung von k-ten Ableitungen wünscht:

Erstens ist die Berechnung einer k-ten partiellen Ableitung von der Reihenfolge der partiellen Differentiationen unabhängig: Wie wir aus Kapitel I, §5*d)* wissen, kann jede Permutation als Produkt von Transpositionen geschrieben werden; es genügt also zu zeigen, daß man die Reihenfolge zweiter partieller Differentiationen vertauschen kann. Eine Transposition $(i_r\ i_{r+k})$ wiederumg läßt sich gemäß

$$(i_r\ i_{r+k}) = (i_r\ i_{r+1}) \cdots (i_{r+k-1}\ i_{r+k})(i_{r+k-2}\ i_{r+k-1}) \cdots (i_r\ i_{r+1})$$

als Produkt von Transpositionen benachbarter Elemente schreiben, und für eine solche Transposition ist

$$\frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_r}} \frac{\partial}{\partial x_{i_{r+1}}} \cdots \frac{\partial f}{\partial x_{i_k}} = \frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_{r+1}}} \frac{\partial}{\partial x_{i_r}} \cdots \frac{\partial f}{\partial x_{i_k}}\,.$$

Für $f \in \mathcal{C}^k(D, \mathbb{R})$ ist sichergestellt, daß

$$\frac{\partial}{\partial x_{i_{r+2}}} \cdots \frac{\partial f}{\partial x_{i_k}} \in \mathcal{C}^2(D, \mathbb{R})\,;$$

nach obigem Lemma ist daher

$$\frac{\partial}{\partial x_{i_r}} \frac{\partial}{\partial x_{i_{r+1}}} \left(\frac{\partial}{\partial x_{i_{r+2}}} \cdots \frac{\partial f}{\partial x_{i_k}} \right) = \frac{\partial}{\partial x_{i_{r+1}}} \frac{\partial}{\partial x_{i_r}} \left(\frac{\partial}{\partial x_{i_{r+2}}} \cdots \frac{\partial f}{\partial x_{i_k}} \right)\,.$$

Differenziert man hier beide Seiten noch partiell nach x_{i_1} bis $x_{i_{r-1}}$, ändert dies natürlich nichts an der Gleichheit.

Zweitens besagt das vorletzte Lemma, daß eine Funktion $f \in \mathcal{C}^1(D, \mathbb{R})$ differenzierbar ist. Eine naheliegende Verallgemeinerung des dortigen Beweises, bei der man anstelle von linearen Approximationen solche höherer Ordnung betrachtet, zeigt, daß eine Funktion $f \in \mathcal{C}^2(D, \mathbb{R})$ zweifach differenzierbar ist, und daß entsprechend eine Funktion f aus $\mathcal{C}^k(D, \mathbb{R})$ bis auf einen Fehler der Größenordnung $o\left(\left|\vec{h}\right|^k\right)$ durch ein Polynom k-ten Grades approximiert werden kann. Wie das im einzelnen aussieht, wollen wir uns im nächsten Abschnitt genauer anschauen.

d) Taylor-Reihen

Beginnen wir auch hier wieder mit Funktionen einer Veränderlichen. Hier gilt

Satz: $f\colon [a,\, b] \to \mathbb{R}$ sei stetig und mindestens $(n+1)$-fach stetig differenzierbar auf dem Intervall $(a,\, b) \subseteq \mathbb{R}$. Dann gilt für jedes x aus $(a,\, b)$ und jedes $h \in \mathbb{R}$ mit $x + h \in (a,\, b)$ die Formel

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \cdots + \frac{h^n}{n!}f^{(n)}(x) + R_{n+1}(x,h) \\ &= \sum_{i=0}^{n} \frac{h^i}{i!}f^{(i)}(x) + R_{n+1}(x,h) \end{aligned}$$

mit einem Restglied $R_{n+1} = O(h^{n+1})$. Dieses kann beispielsweise dargestellt werden als

$$R_{n+1}(x,h) = \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(x+\eta h)$$

mit einer reellen Zahl η zwischen 0 und 1.

Definition: *a)* $T_{f,x,n}(h) = \sum_{i=0}^{n} \frac{h^i}{i!}f^{(i)}(x)$ heißt TAYLOR-Polynom n-ten Grades von f um den Punkt x.
b) $R_{n+1}(x,h) = \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(x+\eta h)$ ist die LAGRANGEsche Form des Restglieds.
c) Die Reihe $T_{f,x}(h) = \sum_{i=0}^{\infty} \frac{1}{i!}f^{(i)}(x)h^i$ heißt TAYLOR-Reihe von f um x.
d) Die Funktion f heißt *analytisch* an der Stelle x, wenn es ein $\delta > 0$ gibt, so daß die TAYLOR-Reihe $T_{f,x}(h)$ existiert und für alle h mit $|h| < \delta$ gegen $f(x+h)$ konvergiert.

Beweis des Satzes: Wir betrachten für festgehaltenes x und h die Hilfsfunktion

$$F\colon \begin{cases} (a,\, b) \to \mathbb{R} \\ \qquad y \mapsto \displaystyle\sum_{i=0}^{n} \frac{f^{(i)}(x)}{i!}(x+h-y)^i \end{cases} .$$

Für $y = x$ stimmt diese mit dem TAYLOR-Polynom $T_{f,x,n}(h)$ überein; für $y = x+h$ ist $x+h-y = 0$, so daß alle Terme mit positivem i verschwinden und nur der konstante Term $F(x+h) = f(x+h)$ übrigbleibt.

Die Ableitung von F ist nach der Produkt- und Kettenregel

$$\begin{aligned} F'(y) &= \sum_{i=0}^{n} \frac{f^{(i+1)}(x)}{i!}(x+h-y)^i - \sum_{i=0}^{n} \frac{f^{(i)}(x)}{i!} \cdot i \cdot (x+h-y)^{i-1} \\ &= \sum_{i=1}^{n+1} \frac{f^{(i)}(x)}{(i-1)!}(x+h-y)^{i-1} - \sum_{i=1}^{n} \frac{f^{(i)}(x)}{(i-1)!} \cdot (x+h-y)^{i-1} \\ &= \frac{f^{(n+1)}(x)}{n!}(x+h-y)^n . \end{aligned}$$

Nach dem verallgemeinerten Mittelwertsatz gibt es für jede zwischen x und $x+h$ differenzierbare Funktion g eine reelle Zahl η zwischen 0 und 1, so daß

$$\frac{F'(x+\eta h)}{g'(x+\eta h)} = \frac{F(x+h) - F(x)}{g(x+h) - g(x)} = \frac{f(x+h) - T_{f,x,n}(x)}{g(x+h) - g(x)}$$

oder

$$\begin{aligned} f(x+h) &= T_{f,x,n}(h) + \frac{g(x+h) - g(x)}{g(x+\eta h)}F'(x+\eta h) \\ &= T_{f,x,n}(h) + \frac{g(x+h) - g(x)}{g(x+\eta h)}\frac{f^{(n+1)}(x+\eta h)}{n!}(h-\eta h)^n \end{aligned}$$

ist. Speziell für die Funktion $g(y) = (x+h-y)^{n+1}$ ist

$$g(x+h) = 0, \quad g(x) = h^{n+1} \quad \text{und} \quad g'(x+\eta h) = -(n+1)(h-\eta h)^n ,$$

insgesamt also

$$\begin{aligned} f(x+h) &= T_{f,x,n}(h) + \frac{h^{n+1}}{(n+1)(h-\eta h)^n}\frac{f^{(n+1)}(x+\eta h)}{n!}(h-\eta h)^n \\ &= T_{f,x,n}(h) + \frac{f^{(n+1)}(x+\eta h)}{(n+1)!}h^{n+1} , \end{aligned}$$

wie behauptet. ∎

BROOK TAYLOR (1685–1731) war Sohn wohlhabender Eltern und wurde daher, bevor er 1703 an die Universität Cambridge ging, nur von privaten Hauslehrern ausgebildet. In Cambridge beschäftigte er sich hauptsächlich mit Mathematik, woran sich auch nach seinem Studienabschluß nichts änderte. Sein 1715 erschienenes Buch *Methodus incrementorum directa et inversa* enthält unter anderem TAYLOR-Polynome (die in Spezialfällen bereits LEIBNIZ, NEWTON und anderen bekannt waren), sowie die Methode der partiellen Integration. Weitere Bücher und Arbeiten beschäftigen sich unter anderem mit der Perspektive sowie mit Fragen aus der Physik.

Als Beispiel betrachten wir die Funktion $f(x) = \sinh x$. Da die Ableitung des *Sinus hyperbolicus* der *Cosinus hyperbolicus* ist und umgekehrt, ist

$$f^{(n)}(x) = \begin{cases} \sinh x & \text{für gerade } n \\ \cosh x & \text{für ungerade } n \end{cases} ;$$

speziell für $x = 0$ verschwinden also alle geraden Ableitungen und die ungeaden haben den Wert eins. Das TAYLOR-Polynom vom Grad $2n$ ist daher

$$T_{\sinh,0,2n}(h) = \sum_{i=0}^{n-1} \frac{h^{2n+1}}{(2n+1)!} ,$$

und das Restglied ist

$$R_{2n+1}(0,h) = \frac{h^{2n+1}}{(2n+1)!} \cosh \eta h$$

mit einer reellen Zahl η zwischen 0 und 1. Die Zahl η hängt dabei natürlich von n ab, aber da der *Cosinus hyperbolicus* im Positiven monoton steigend und im Negativen monoton fallend ist, ist auf jeden Fall

$$\cosh(\eta h) \le \cosh h .$$

Damit ist

$$|R_{2n+1}(0,h)| = \left| \frac{h^{2n+1}}{(2n+1)!} \cosh(\eta h \right| \le \frac{|h|^{2n+1}}{(2n+1)!} \cosh h ,$$

und rechts steht für festes h und wachsendes n eine Nullfolge. Daher ist nicht nur für jedes n

$$\sinh x = \sum_{i=0}^{n-1} \frac{x^{2i+1}}{(2i+1)!} + R_{2n+1}(0,x) ,$$

sondern auch im Limes

$$\sinh x = \sum_{i=0}^{\infty} \frac{x^{2i+1}}{(2i+1)!} .$$

Im Sinne der obigen Definitionen ist also der *Cosinus hyperbolicus* analytisch im Nullpunkt; tatsächlich überzeugt man sich leicht, daß er sogar analytisch auf ganz $\mathbb{R}$ ist.

Eine Funktion kann natürlich nur dann analytisch in x sein, wenn sie dort beliebig oft differenzierbar ist – andernfalls läßt sich die TAYLOR-Reihe nicht einmal definieren. Analytizität ist aber eine noch stärkere Eigenschaft als beliebige Differenzierbarkeit, wie das folgende Beispiel zeigt: Wir setzen

$$f(x) = \begin{cases} e^{-1/x^2} & \text{für } x \neq 0 \\ 0 & \text{für } x = 0 \end{cases} .$$

Diese Funktion ist stetig, denn für $x \to 0$ geht $-1/x^2$ gegen $-\infty$, e^{-1/x^2} also gegen null.

Auch mit der Differenzierbarkeit gibt es keine Probleme, denn für $x \neq 0$ ist

$$f'(x) = \frac{2}{x^3} e^{-1/x^2} ,$$

und auch alle weiteren Ableitungen haben die Form

$$f^{(n)}(x) = \textit{rationale Funktion} \times e^{-1/x^2} .$$

Die rationale Funktion geht zwar gegen unendlich für $x \to 0$, aber da e^{-1/x^2} viel schneller gegen null geht als irgendeine rationale Funktion gegen unendlich, ist der Grenzwert des Produkts für $x \to 0$ gleich null. Also ist f auch im Nullpunkt beliebig oft differenzierbar, und alle Ableitungen verschwinden. Die TAYLOR-Reihe von f um den Nullpunkt ist daher die Nullfunktion, d.h. die Reihe konvergiert zwar überall, aber außerhalb des Nullpunkts nicht gegen e^{-1/x^2}.

Damit kennen wir TAYLOR-Reihen im Eindimensionalen; zur Verallgemeinerung auf das Mehrdimensionale verwenden wir Richtungsableitungen. Da wir eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}^m$ problemlos komponentenweise behandeln können, genügt es, den Fall $m = 1$ zu betrachten.

Die Richtungsableitung einer Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ in Richtung $\vec{v}$ ist nach Definition die Ableitung der Funktion einer Veränderlichen

$$g\colon \begin{cases} (-a, a) & \to \mathbb{R} \\ t & \mapsto f(\boldsymbol{x} + t\vec{v}) \end{cases}$$

mit geeignet gewähltem $a \in \mathbb{R}_+$ nach t für $t = 0$; wir können sie berechnen als Skalarprodukt des Gradienten mit dem Vektor $\vec{v}$.

Natürlich können wir g nicht nur einmal ableiten; für $f \in \mathcal{C}^k(D, \mathbb{R})$ existiert das TAYLOR-Polynom k-ten Grades

$$g(h) = g(0) + h \cdot g'(0) + \frac{h^2}{2} g''(0) + \cdots + \frac{h^k}{k!} g^{(k)}(0) + o(h^k)\,.$$

Führen wir nun für die Richtungsableitung in Richtung $\vec{v}$ einen Operator $\partial_{\vec{v}}$ ein durch

$$\partial_{\vec{v}} f(x) = \operatorname{grad} f(x) \cdot \vec{v} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}) \cdot v_i\,,$$

so wird dies zu

$$f(\boldsymbol{x} + h\vec{v}) = f(\boldsymbol{x}) + h\partial_{\vec{v}} f(\boldsymbol{x}) + \cdots + \frac{h^k}{k!} \partial_{\vec{v}}^k f(\boldsymbol{x}) + o(h^k)\,.$$

Speziell für $h = 1$ erhalten wir die kompakte Schreibweise der TAYLOR-Formel, nämlich

$$f(\boldsymbol{x} + \vec{v}) = f(\boldsymbol{x}) + \partial_{\vec{v}} f(\boldsymbol{x}) + \cdots + \frac{1}{k!} \partial_{\vec{v}}^k f(\boldsymbol{x}) + o\left(\left|\vec{v}\right|^k\right)\,.$$

$\partial_{\vec{v}}^k$ steht hierbei natürlich für die k-fache Anwendung des Operators $\partial_{\vec{v}}$; was das explizit bedeutet, sollte man sich anhand obiger Summendarstellung von $\partial_{\vec{v}}$ klarmachen. Insbesondere beachte man, daß alle Potenzprodukte der v_i in den Potenzen des Operatiors $\partial_{\vec{v}}$ stecken und daß diese nach Ausmultiplizieren schnell ziemlich unübersichtlich werden. Einige Leser werden dieses Phänomen wohl aus der Physik kennen:

Die *Multipolentwicklung* eines elektrischen Feldes um einen Punkt ist nichts anderes als die TAYLOR-Entwicklung; der konstante Term hängt ab von der elektrischen Ladung, der lineare Term vom Dipolmoment, der quadratische vom Quadrupolmoment *usw.*

Wie das für beliebiges n im einzelnen aussieht, wollen wir uns lieber nicht so genau anschauen: Eine Form vom Grad k ist gegeben durch eine k-fache Summation von 1 bis n, hat also n^k Summanden, was sehr schnell sehr unübersichtlich wird.

Der quadratische Term der TAYLOR-Reihe ist allerdings noch einigermaßen gut handhabbar: Dafür brauchen wir nur die HESSE-Matrix

$$H_f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}\,,$$

und nach dem Lemma von SCHWARZ ist diese für $f \in \mathcal{C}^2(D, \mathbb{R})$ eine symmetrische Matrix, was den Rechenaufwand noch einmal fast um die Hälfte reduziert.

e) 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 $F(\boldsymbol{x}) = 0$ 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. Der Satz über implizite Funktionen besagt, das dieses Nichtverschwinden der partiellen Ableitung bereits ausreicht um die Existenz einer eindeutigen Auflösung zu zeigen.

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

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

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

Die Funktion f ist stetig und differenzierbar; ihre Ableitung ist

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

Beweis: Wir beginnen mit einer Reduktion zwecks Vereinfachung der Schreibarbeit: Offensichtlich genügt es, wenn wir den Fall $x_0 = y_0 = 0$ 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

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

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

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

Damit ist f als Funktion festgelegt, und nach Konstruktion ist

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

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

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

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

ist.

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 $\big|F(x',y')\big| < \varepsilon$ 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

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

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

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

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

$$F\big(x+h, f(x+h)\big) = F\big(x+h, f(x)+k\big) = F\left(\big(x, f(x)\big) + \tbinom{h}{k}\right),$$

und setzen wir

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

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

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

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

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

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

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

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

also folgt

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

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

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

d.h.

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

Für Extremwertbetrachtungen bei implizit definierten Funktionen braucht man, wie wir im nächsten Semester sehen werden, die zweite Ableitung vor allem in den Punkten, in denen die erste verschwindet; dort vereinfacht sich die Formel zu

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

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

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

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

Für alle $i \leq n-1$ ist

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

■

§2: Vektorfelder

Eine Funktion $f\colon D \to \mathbb{R}^m$ auf einer Teilmenge $D \subseteq \mathbb{R}^n$ läßt sich für $n, m \geq 2$ nicht mehr durch einen Graphen veranschaulichen, denn der Graph

$$\Gamma_f = \big\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^m \;\big|\; \boldsymbol{y} = f(\boldsymbol{x})\big\}$$

liegt dann im mindestens vierdimensionalen $\mathbb{R}^{n+m}$.

Im für uns wichtigsten Fall $n = m = 2$ oder 3 gibt es aber eine andere Möglichkeit der Veranschaulichung: Für $\boldsymbol{x} \in D$ läßt sich $f(\boldsymbol{x})$ als Vektor auffassen, und f kann veranschaulicht werden, indem man für hinreichend viele Punkte $\boldsymbol{x} \in D$ diesen Vektor (oder ein geeignetes konstantes Vielfaches davon) im Punkt $\boldsymbol{x}$ anträgt; siehe dazu die Abbildungen 39 und 40.

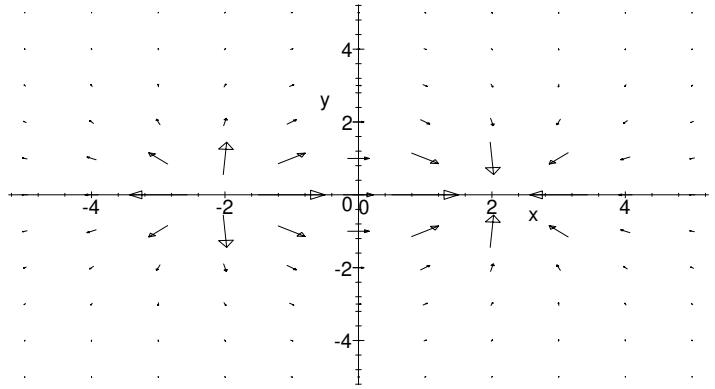


Abb. 39: Das elektrische Feld zweier entgegengesetzt gleicher Punktladungen

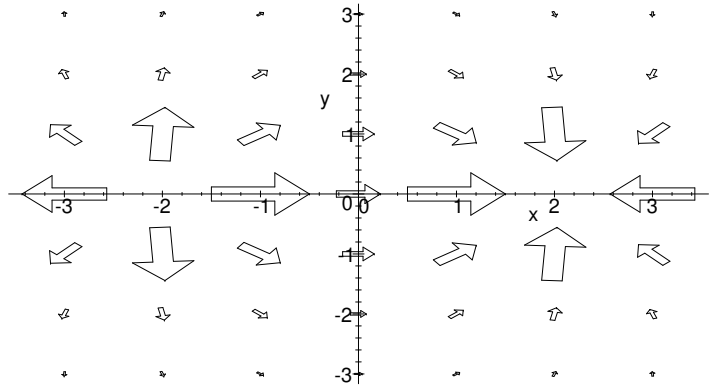


Abb. 40: Dasselbe Feld in einem kleineren Bereich mit „flächigen“ Vektoren

Wie Abbildung 39 zeigt, kann es dabei bei betragsmäßig stark variierenden Funktionswerten zu Problemen kommen: Wie man die Konstante auch wählt, mit der alle Vektoren multipliziert werden, hat man immer entweder einige extrem lange Vektoren oder aber Vektoren, die so kurz sind, daß man nichts mehr erkennen kann: Ab einer gewissen Entfernung vom Zentrum sind die Richtungen des Felds praktisch nicht mehr wahrnehmbar. Eine leichte Verbesserung ergibt sich, wenn man wie in Abbildung 40 die Vektoren „flächig“ zeichnet, wobei die Länge des Vektors proportional zur *Fläche* des eingezeichneten Pfeils ist, aber auch das hilft nicht sehr viel.

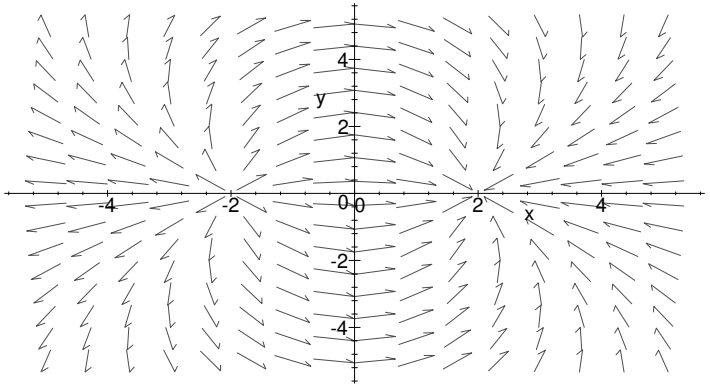


Abb. 41: Das Richtungfeld zum obigen Feld

Da gerade die Richtung sehr wesentlich für die Konstruktion von Feldlinien ist, begnügt man sich oft damit, nur die Richtungen darzustellen, ansonsten aber sämtlichen Vektoren dieselbe Länge zu geben; siehe dazu etwa Abbildung 41. Dadurch geht zwar ein Teil der Information verloren, aber das entsprechende Bild kann trotzdem oft mehr aussagen, als eines, bei dem fast alle Pfeile visuell zu Punkten degeneriert sind. Auch bei der Darstellung als Richtungsfeld müssen allerdings natürlich alle Punkte $\boldsymbol{x}$, für die $f(\boldsymbol{x}) = (0, \ldots, 0)$ ist, durch Punkte oder überhaupt nicht dargestellt werden, denn dort gibt es keine Richtung.

a) Der Begriff des Vektorfelds

Nicht nur zur Veranschaulichung kann es nützlich sein, die Werte einer Funktion als Vektoren zu betrachten: Die für uns wichtigsten Beispiele von Funktionen mehrerer Veränderlicher, deren Wert nicht in $\mathbb{R}$ sondern in einem $\mathbb{R}^n$ liegt, sind die aus der Physik bekannten Felder wie etwa ein Kraftfeld, ein elektrisches Feld oder ein Magnetfeld. Da solche Felder typischerweise als Vektoren geschrieben werden, schreiben wir deshalb in diesem Paragraphen die *Funktionswerte* vektoriell, wohingegen die Argumente weiterhin Punkte sind, die als Tupel geschrieben werden. Wir betrachten somit Funktionen

$$\vec{V}\colon D \to \mathbb{R}^m; \quad \boldsymbol{x} \mapsto \vec{V}(\boldsymbol{x}) = \begin{pmatrix} V_1(\boldsymbol{x}) \\ \vdots \\ V_n(\boldsymbol{x}) \end{pmatrix},$$

wobei die Komponentenfunktionen $V_i\colon D \to \mathbb{R}$ genau die im vorigen Paragraphen betrachteten skalarwertigen Funktionen mehrerers Veränderlicher sind.

Der mathematische Begriff, der physikalische Felder beschreibt, ist der des Vektorfelds:

Definition: Ein Vektorfeld auf der Teilmenge $D \subseteq \mathbb{R}^n$ ist eine Funktion $\vec{V}\colon D \to \mathbb{R}^n$.

(Da die typischen elektromagnetischen Felder wie $\vec{E}, \vec{D}, \vec{B}, \vec{H}$ oder auch das Kraftfeld $\vec{F}$ üblicherweise durch Großbuchstaben bezeichnet werden, verwenden wir auch für Vektorfelder meist Großbuchstaben wie $\vec{V}$ oder $\vec{W}$.)

b) Die Jacobi-Matrix

Sowohl bezüglich der Maximumsnorm als auch bezüglich der EUKLIDischen Norm (und in der Tat jeder möglichen Norm auf $\mathbb{R}^n$ *bzw.* $\mathbb{R}^m$) ist klar, daß eine Funktion $\vec{f}\colon D \to \mathbb{R}^m$ auf $D \subseteq \mathbb{R}^n$ genau dann stetig *bzw.* differenzierbar ist, wenn alle Komponentenfunktionen f_i stetig *bzw.* differenzierbar sind. Entsprechend definieren wir

auch die Menge $\mathcal{C}^k(D,\mathbb{R}^m)$ aller k-fach stetig differenzierbarer Funktionen von D nach $\mathbb{R}^m$ als Menge aller Funktionen, deren Komponenten $f_i \in \mathcal{C}^k(D,\mathbb{R})$ k-mal stetig differenzierbar sind.

Wie bei skalarwertigen Funktionen überlegt sich leicht, daß $\mathcal{C}^k(D,\mathbb{R}^m)$ ein $\mathbb{R}$-Vektorraum ist.

Bei einer differenzierbare Funktion $\vec{f}$ gilt für jede ihrer Komponenten

$$f_i(\boldsymbol{x}+\vec{h}) = f_i(\boldsymbol{x}) + \operatorname{grad} f_i(\boldsymbol{x}) \cdot \vec{h} + o(\|\vec{h}\|)\,,$$

wobei der Punkt hier für das Skalarprodukt von Vektoren steht. Im Hinblick auf vektorwertige Funktionen ist es aber besser, anstelle des Skalarprodukt ein Matrixprodukt zu verwenden; dazu müssen wir den Gradienten als Zeilenvektor schreiben, d.h. wenn $\cdot$ das Matrixprodukt bezeichnet, ist

$$f_i(\boldsymbol{x}+\vec{h}) = f_i(\boldsymbol{x}) + {}^t\operatorname{grad} f_i(\boldsymbol{x}) \cdot \vec{h} + o(\|\vec{h}\|)\,.$$

In dieser Notation können wir die als Zeilenvektoren aufgefaßten Gradienten zu einer $m \times n$-Matrix zusammenfassen, der JACOBI-Matrix

$$J_{\vec{f}}(\boldsymbol{x}) = \begin{pmatrix} {}^t\operatorname{grad} f_1(\boldsymbol{x}) \\ \vdots \\ {}^t\operatorname{grad} f_m(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}\,.$$

Mit dieser Matrix ist dann

$$\vec{f}(\boldsymbol{x}+\vec{h}) = \vec{f}(\boldsymbol{x}) + J_{\vec{f}}(\boldsymbol{x}) \cdot \vec{h} + o(\|\vec{h}\|)\,,$$

wobei hier das Produkt als Matrixprodukt zu verstehen wird und das Symbol $o(\|\vec{h}\|)$ schlampigerweise auch für einen Vektor benutzt wird, dessen sämtliche Komponenten $o(\|\vec{h}\|)$ sind.

c) Die Divergenz eines Vektorfelds

Die JACOBI-Matrix eines Vektorfelds hängt stark ab vom gewählten Koordinatensystem. Wir wollen aus dieser Matrix Größen ableiten, von denen wir später sehen werden, daß sie *intrinsische,* d.h. also vom Koordinatensystem unabhängige Eigenschaften des Vektorfelds beschreiben.

Definition: *a)* Die *Spur* einer $n \times n$-Matrix $A = (a_{ij})$ ist die Summe

$$\operatorname{Spur} A = a_{11} + a_{22} + \dots + a_{nn}$$

der Diagonalelemente von A.
b) Die Spur der JACOBI-Matrix $J_{\vec{V}}(\boldsymbol{x})$ eines Vektorfelds heißt *Divergenz* oder *Quellendichte* von $\vec{V}$:

$$\operatorname{div} \vec{V}(\boldsymbol{x}) = \operatorname{Spur} J_{\vec{V}}(\boldsymbol{x})\,.$$

Man überlegt sich leicht, daß sich die Spur einer Matrix bei einem Basiswechsel nicht ändert: Aus der definierenden Formel für die Determinante folgt leicht, daß Spur A der Koeffizient von λ^{n-1} in $\det(A+\lambda E)$ ist, und diese Determinante bleibt natürlich invariant bei Basiswechel. Insbesondere ist die Spur im Falle einer diagonalisierbaren Matrix stets gleich der Summe der Eigenwerte.

Um ein anschauliches Verständnis von der Divergenz zu bekommen, betrachten wir den (sehr speziellen) Fall, daß die JACOBI-Matrix von $\vec{V}$ eine Diagonalmatrix ist:

$$J_{\vec{V}}(\boldsymbol{x}) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

Dann ist für jeden Einheitsvektor $\vec{e}_i$

$$\vec{V}(\boldsymbol{x}+h\vec{e}_i) = \vec{V}(\boldsymbol{x}) + h \cdot J_{\vec{V}}(\boldsymbol{x})\vec{e}_i + o(h) = \vec{V}(\boldsymbol{x}) + h\lambda_i\vec{e}_i + o(h)\,,$$

und für einen beliebigen Vektor $\vec{h} = h_1\vec{e}_1 + \dots + h_n\vec{e}_n$ ist

$$\vec{V}(\boldsymbol{x}+\vec{h}) = \vec{V}(\boldsymbol{x}) + J_{\vec{V}}(\boldsymbol{x})\vec{h} + o(\|\vec{h}\|) = \vec{V}(\boldsymbol{x}) + \begin{pmatrix} \lambda_1 h_1 \\ \vdots \\ \lambda_n h_n \end{pmatrix} + o(\|\vec{h}\|)\,.$$

Falls alle λ_i positiv sind, hat hier jede Komponente des Vektors $\vec{V}(\boldsymbol{x})+J_{\vec{V}}(\boldsymbol{x})\vec{h}$ dasselbe Vorzeichen wie die entsprechende Komponente von $\vec{h}$; entfernt man sich also vom Punkt $\boldsymbol{x}$, so zeigt auch die Veränderung des Vektorfelds $\vec{V}$ weg vom Punkt x und umgekehrt; die Situation

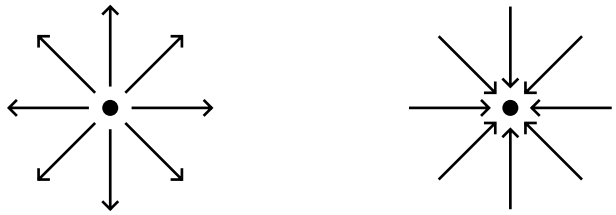


Abb. 42: Quellen und Senken

ist also ungefähr so wie auf der linken Seite von Abbildung 42. Man bezeichnet $\boldsymbol{x}$ in einem solchen Fall als eine *Quelle* des Vektorfelds $\vec{V}$.

Falls alle λ_i negativ sind, gehen die Änderungen jeweils in Gegenrichtung, es sieht also ungefähr so aus wie im rechten Teil von Abbildung 42, und man spricht von einer *Senke.*

Im allgemeinen werden allerdings weder *alle* λ_i positiv noch *alle* λ_i negativ sein; die Divergenz als Summe der λ_i sagt uns dann, ob sich der Punkt eher wie eine Quelle oder eher wie eine Senke verhält.

Nun wird es natürlich nur selten vorkommen, daß die JACOBI-Matrix eine Diagonalmatrix ist; wie wir am Ende dieses Kapitels sehen werden, ist die Divergenz trotzdem auch im allgemeinen Fall eine Maßzahl dafür, ob ein Punkt Quelle oder Senke ist, und zwar in einer viel präziseren Weise, als es die obige Diskussion für Diagonalmatrizen zeigte.

Im übrigen sind Diagonalmatrizen doch nicht so speziell, wie es zunächst den Anschein hat: Beispielsweise läßt sich immer dann, wenn der $\mathbb{R}^n$ eine Basis aus Eigenvektoren der betrachteten Matrix hat, ein Koordinatensystem finden, bezüglich dessen sie Diagonalform hat. Im nächsten Semester werden wir sehen, daß dies bei symmetrischen Matrizen immer der Fall ist; zumindest für den symmetrischen Anteil der JACOBI-Matrix ist die obige Diskussion also richtig. Mit dem antisymmetrischen Anteil beschäftigen wir uns im nächsten Abschnitt.

In diesem Abschnitt wollen wir uns nur noch überlegen, wie man die Divergenz eines Vektorfelds tatsächlich ausrechnet. Die Divergenz ist

nach Definition gleich der Spur der JACOBI-Matrix, also ist

$$\operatorname{div}\vec{V} = \operatorname{Spur} J_{\vec{V}}(x) = \operatorname{Spur}\begin{pmatrix} \frac{\partial V_1}{\partial x_1} & \frac{\partial V_1}{\partial x_2} & \cdots & \frac{\partial V_1}{\partial x_n} \\ \frac{\partial V_2}{\partial x_1} & \frac{\partial V_2}{\partial x_2} & \cdots & \frac{\partial V_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial V_n}{\partial x_1} & \frac{\partial V_n}{\partial x_2} & \cdots & \frac{\partial V_n}{\partial x_n} \end{pmatrix}$$
$$= \frac{\partial V_1}{\partial x_1} + \frac{\partial V_2}{\partial x_2} + \cdots + \frac{\partial V_n}{\partial x_n}\,.$$

Genauso wie wir den Gradienten einer skalarwertigen Funktion f auch als ∇f schreiben, schreiben wir die Divergenz eines Vektorfelds $\vec{V}$ auch als $\nabla\vec{V}$. Beides kann formal auch als Produkt *bzw.* Skalarprodukt mit einem vektorwertigen Operator interpretiert werden:

$$\nabla f = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \cdot f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

und

$$\nabla\vec{V} = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix} = \frac{\partial V_1}{\partial x_1} + \frac{\partial V_2}{\partial x_2} + \cdots + \frac{\partial V_n}{\partial x_n}\,.$$

Diese Interpretation darf allerdings nicht *zu* ernst genommen werden, denn obwohl Skalarprodukt und Produkt mit einem Skalar beide kommutativ sind, ist weder die Gleichung $\nabla\vec{V} = \vec{V}\nabla$ noch die Gleichung $\nabla f = f\nabla$ auch nur annähernd sinnvoll, geschweige denn richtig.

d) Vektorprodukt und Rotation im Dreidimensionalen

Erinnern wir uns zunächst an symmetrische und antisymmetrische Matrizen: Eine Matrix heißt *symmetrisch,* wenn sie gleich ihrer transponierten ist, wenn also $a_{ij} = a_{ji}$ ist für alle Indexpaare (i, j); sie heißt *antisymmetrisch,* wenn ${}^tA = -A$ ist, d.h. $a_{ij} = -a_{ji}$. Jede reelle Matrix A kann

als Summe einer symmetrischen Matrix S und einer antisymmetrischen Matrix T geschrieben werden, denn mit

$$S = \frac{1}{2}\left(A + {}^tA\right) \quad \text{und} \quad T = \frac{1}{2}\left(A - {}^tA\right) \quad \text{ist} \quad S + T = A\,.$$

Wie wir im nächsten Semester sehen werden, ist jede symmetrische reelle Matrix diagonalisierbar, und da bedeutet, daß uns die Divergenz eines Vektorfelds recht viel Information gibt über den symmetrischen Anteil der JACOBI-Matrix.

Über den antisymmetrischen Anteil kann sie uns nichts sagen, denn aus $a_{ij} = -a_{ji}$ folgt insbesondere, daß alle Diagonaleinträge a_{ii} verschwinden müssen, d.h. der antisymmetrische Anteil liefert keine Beiträge zur Spur einer Matrix.

Wir wollen uns überlegen, wie der antisymmetrische Anteil im Fall kleiner Dimensionen aussieht.

Für $n = 1$ ist jede $n \times n$-„Matrix“ symmetrisch, es gibt als keinen antisymmetrischen Anteil.

Für $n = 2$ hat jede antisymmetrische Matrix A die Form

$$\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix},$$

für den antisymmetrischen Anteil der JACOBI-Matrix ist

$$a = \frac{1}{2}\left(\frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x}\right).$$

Damit ist der antisymmetrische Anteil durch diese eine Zahl bestimmt.

Für $n > 3$ gibt es keine einfachere Darstellung des antisymmetrischen Anteils als die schiefsymmetrische Matrix selbst, bleibt also noch der für viele Anwendungen besonders wichtige Fall $n = 3$.

Eine antisymmetrische 3×3-Matrix A hat die Form

$$A = \begin{pmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ -\beta & -\gamma & 0 \end{pmatrix},$$

hängt also nur von drei Parametern ab. Da drei auch die Dimension des $\mathbb{R}^3$ ist, können wir diese als Komponenten eines Vektors interpretieren und hoffen, daß sich das Produkt von A mit einem Vektor als ein Produkt zweier Vektoren auffassen läßt. Dieses Produkt muß allerdings – im Gegensatz zum Skalarprodukt – zwei Vektoren wieder einen Vektor zuordnen.

Ein solches Produkt gibt es nur im Dreidimensionalen; dort leistet das (vielleicht aus der Schule bekannte) Vektorprodukt oder Kreuzprodukt das Verlangte. Wie schon der Name sagt, ordnet es je zwei Vektoren $\vec{v}$ und $\vec{w}$ aus $\mathbb{R}^3$ einen *Vektor* zu, und dieser wird mit $\vec{v} \times \vec{w} \in \mathbb{R}^3$ bezeichnet. Er ist festgelegt durch folgende Eigenschaften:

- $\vec{v} \times \vec{w}$ hat die Länge $|\vec{v} \times \vec{w}| = |\vec{v}|\,|\vec{w}|\,|\sin \measuredangle(\vec{v}, \vec{w})|$.
  Insbesondere ist also $\vec{v} \times \vec{w} = \vec{0}$, wenn $\vec{v}$ und $\vec{w}$ auf einer Geraden liegen, denn dann bilden sie einen Winkel von null oder 180 Grad, so daß der Sinus verschwindet.
- $\vec{v} \times \vec{w}$ steht senkrecht sowohl auf $\vec{v}$ als auch auf $\vec{w}$.
  Falls $\vec{v} \times \vec{w} \neq \vec{0}$ ist, spannen $\vec{v}$ und $\vec{w}$ eine Ebene auf, auf der (da wir im $\mathbb{R}^3$ sind) genau ein eindimensionaler Unterraum senkrecht steht. Darin gibt es allerdings für jede vorgegebene positive Länge zwei Vektoren, die sich durch ihr Vorzeichen unterscheiden. Um $\vec{v} \times \vec{w}$ eindeutig festzulegen, brauchen wir daher noch eine weitere Bedingung:
- Die drei Vektoren $\vec{v}, \vec{w}$ und $\vec{v} \times \vec{w}$ bilden ein Rechtssystem, d.h. wenn sich die Finger der *rechten* Hand so ausrichten lassen, daß der Daumen in Richtung von $\vec{v}$ zeigt, der Zeigefinger in Richtung von $\vec{w}$ und der Mittelfinger in Richtung von $\vec{v} \times \vec{w}$.

Alternativ kann man ein Rechtssystem auch so definieren, daß sich, ein von $\vec{v}$ nach $\vec{w}$ gedrehter Korkenzieher in Richtung $\vec{v} \times \vec{w}$ in den Kork bohrt. Ähnlich geht es auch mit Schrauben; da es allerdings neben den (üblichen) Rechtsschrauben auch die (seltenen) Linksschrauben gibt, ist diese Definition eventuell zirkulär: Alles hängt davon ab, wie man Rechtsschrauben definiert.

Aus jeder dieser Regeln folgt sofort die *Antikommutativität* des Vektorprodukts:

$$\vec{v} \times \vec{w} = -\vec{w} \times \vec{v}\,.$$

Weitere Rechenregeln lassen sich leicht geometrisch ableiten: Da der Sinus eines Winkels gleich Gegenkathete durch Hypothenuse ist, ist in der von $\vec{v}$ und $\vec{w}$ aufgespannten Ebenen $|\vec{w}|\,|\sin\measuredangle(\vec{v},\vec{w})|$ gleich der Länge des auf die senkrecht auf $\vec{v}$ stehenden Geraden projizierten Vektors $\vec{w}$, das heißt also gleich der Höhe des in Abbildung 42 eingezeichneten Rechtecks. Die Länge des Vektors $\vec{v} \times \vec{w}$ ist daher gleich dem Flächeninhalt dieses Rechtecks und damit – wie eine Scherung zeigt – gleich der Fläche des von $\vec{v}$ und $\vec{w}$ aufgespannten Parallelogramms.

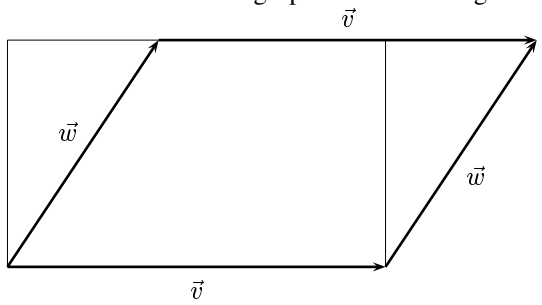


Abb. 42: Geometrische Interpretation des Vektorprodukts

Dies zeigt das Distributivgesetz

$$\vec{v} \times (\vec{w} + \vec{u}) = \vec{v} \times \vec{w} + \vec{v} \times \vec{u}\,,$$

für den zweiten Faktor, und wegen der Antikommutativität folgt daraus auch das für den ersten:

$$(\vec{u} + \vec{v}) \times \vec{w} = \vec{u} \times \vec{w} + \vec{v} \times \vec{w}\,.$$

Um das Vektorprodukt in Koordinaten ausrechnen zu können, müssen wir zunächst die Produkte der Koordinateneinheitsvektoren $\vec{e}_i$ kennen. Da sie allesamt die Länge eins haben und paarweise aufeinander senkrecht stehen, ist klar, daß das Produkt zweier verschiedener dieser Vektoren bis aufs Vorzeichen gleich dem dritten ist; das Vorzeichen hängt

ab von der Orientierung des Koordinatensystems. Das Produkt eines Vektors $\vec{e}_i$ mit sich selbst ist natürlich, wie jedes Produkt eines Vektors mit sich selbst, gleich dem Nullvektor, denn der eingeschlossene Winkel ist null Grad.

Für die folgende Rechnung wollen wir annehmen, daß $\vec{e}_1, \vec{e}_2$ und $\vec{e}_3$ in dieser Reihenfolge ein Rechtssystem bilden; das ist beispielsweise dann der Fall, wenn $\vec{e}_1$ nach rechts, $\vec{e}_2$ nach vorne und $\vec{e}_3$ nach oben zeigt. Dann folgt sofort, daß

$$\vec{e}_1 \times \vec{e}_2 = \vec{e}_3$$

ist, und nach einigen Fingerübungen auch findet man auch die Formeln

$$\vec{e}_2 \times \vec{e}_3 = \vec{e}_1 \quad \text{und} \quad \vec{e}_1 \times \vec{e}_3 = -\vec{e}_2\,.$$

Die Produkte mit vertauschten Faktoren sind natürlich gerade das negative davon, und $\vec{e}_i \times \vec{e}_i = 0$. Für

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \text{und} \quad \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

ist also

$$\vec{v} \times \vec{w} = (v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3) \times (w_1\vec{e}_1 + w_2\vec{e}_2 + w_3\vec{e}_3)\,,$$

und nach obigen Rechenregeln ist dies gleich

$$\begin{aligned} &\sum_{i=1}^{3}\sum_{j=1}^{3} v_i w_j\, \vec{e}_i \times \vec{e}_j \\ &= (v_2w_3 - v_3w_2)\vec{e}_1 + (v_3w_1 - v_1w_3)\vec{e}_2 + (v_1w_2 - v_2w_1)\vec{e}_3\,, \end{aligned}$$

d.h.

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \\ v_1w_2 - v_2w_1 \end{pmatrix}\,.$$

Dies läßt sich dadurch merken, daß man im Schema

$$\begin{array}{ccccccccc} \vec{e}_1 & & \vec{e}_2 & & \vec{e}_3 & & \vec{e}_1 & & \vec{e}_2 \\ & \searrow & & \swarrow\!\!\!\searrow & & \swarrow\!\!\!\searrow & & \swarrow & \\ v_1 & & v_2 & & v_3 & & v_1 & & v_2 \\ & \swarrow & & \swarrow\!\!\!\searrow & & \swarrow\!\!\!\searrow & & \searrow & \\ w_1 & & w_2 & & w_3 & & w_1 & & w_2 \end{array}$$

von $\vec{e}_i$ ausgeht und als dessen Koeffizient das Zweierprodukt entlang der schrägen Linie nach rechts unten *positiv* und das entlang der schrägen Linie nach links unten *negativ* nimmt; man wendet also die SARRUSsche Regel an auf die „Determinante“

$$\begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} .$$

Das Produkt einer schiefsymmetrischen Matrix $A \in \mathbb{R}^{3\times 3}$ mit einem Vektor $\vec{x} \in \mathbb{R}^3$ ist

$$A\vec{x} = \begin{pmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ -\beta & -\gamma & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \alpha y + \beta z \\ -\alpha x + \gamma z \\ -\beta x - \gamma y \end{pmatrix} ,$$

während ein Vektorprodukt $\vec{a} \times \vec{x}$ sich zu

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} bz - cy \\ cx - az \\ ay - bx \end{pmatrix}$$

berechnet, d.h.

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

und damit

$$\begin{pmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ -\beta & -\gamma & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\gamma \\ \beta \\ -\alpha \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

Somit entspricht im Dreidimensionalen die Multiplikation mit einer schiefsymmetrischen Matrix genau dem Vektorprodukt (das im übrigen auch nur im Dreidimensionalen definierbar ist).

Der Vektor $\vec{a}$ läßt sich für das Produkt mit dem schiefsymetrischen Anteil der JACOBI-Matrix leicht ausrechnen: Die JACOBI-Matrix

$$J_{\vec{V}}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial V_1}{\partial x} & \frac{\partial V_1}{\partial y} & \frac{\partial V_1}{\partial z} \\ \frac{\partial V_2}{\partial x} & \frac{\partial V_2}{\partial y} & \frac{\partial V_2}{\partial z} \\ \frac{\partial V_3}{\partial x} & \frac{\partial V_3}{\partial y} & \frac{\partial V_3}{\partial z} \end{pmatrix}$$

eines dreidimensionalen Vektorfelds $\vec{V}$ hat antisymmetrischen Anteil

$$A = \frac{1}{2}\left(A - {}^tA\right) = \frac{1}{2}\begin{pmatrix} 0 & \frac{\partial V_1}{\partial y} - \frac{\partial V_2}{\partial x} & \frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x} \\ \frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} & 0 & \frac{\partial V_2}{\partial z} - \frac{\partial V_3}{\partial y} \\ \frac{\partial V_3}{\partial x} - \frac{\partial V_1}{\partial z} & \frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z} & 0 \end{pmatrix} ,$$

und wie wir uns gerade überlegt haben, ist für jeden Vektor $\vec{v}$

$$A\vec{v} = \vec{a} \times \vec{v} \quad \text{mit} \quad \vec{a} = \frac{1}{2}\begin{pmatrix} \frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z} \\ \frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x} \\ \frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \end{pmatrix} .$$

Da der Faktor $\frac{1}{2}$ in konkreten Rechnungen nur hinderlich wäre, betrachten wir den Vektor $2\vec{a}$ und definieren:

Definition: $\vec{V}\colon D \to \mathbb{R}^3$ sei ein Vektorfeld auf der offenen Teilmenge $D \subseteq \mathbb{R}^3$. Dann bezeichnen wir das Vektorfeld

$$\operatorname{rot} V\colon D \to \mathbb{R}^3; \quad \boldsymbol{x} \mapsto \begin{pmatrix} \frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z} \\ \frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x} \\ \frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \end{pmatrix}$$

als Rotation von $\vec{V}$.

Für ein Vektorfeld $\vec{V} \in \mathcal{C}^1(D, \mathbb{R}^3)$, dessen JACOBI-Matrix symmetrischen Anteil S hat, ist also

$$\vec{V}(\boldsymbol{x} + \vec{h}) = \vec{V}(\boldsymbol{x}) + S\vec{h} + \tfrac{1}{2}(\operatorname{rot}\vec{V}) \times \vec{h} + o(\|\vec{h}\|) .$$

Der Name *Rotation* wird im folgenden (etwas umfangreichen) Beispiel verständlich: Wir gehen aus vom Vektor

$$\vec{a} = 2 \operatorname{rot} \vec{V}(\boldsymbol{x}) .$$

Nach Definition des Vektorprodukts steht $\vec{a} \times \vec{h}$ senkrecht sowohl auf $\vec{a}$ als auch auf $\vec{h}$. Bedenkt man noch, daß Vektoren Äquivalenzklassen von Pfeilen sind, die man sich nicht unbedingt so vorstellen muß, daß sie alle im selben Punkt beginnen, bietet sich folgende Interpretation an: Wenn wir den Vektor $\vec{h}$ um die von $\vec{a}$ aufgespannte Achse drehen, zeigt $\vec{a} \times \vec{h}$ in Richtung der Tangente der Bahn des Endpunkts von $\vec{h}$. (Damit dies stimmt, muß die Achse so orientiert werden, daß sich $\vec{h}$ im mathematisch positiven, d.h. Gegenuhrzeigersinn, um $\vec{a}$ dreht. Von der Ebene der Uhr aus gesehen zeigt dann $\vec{a}$ nach oben, $\vec{h}$ zu dem sich drehenden Punkt, und $\vec{a} \times \vec{h}$ in Richtung des Gegenuhrzeigersinns, wie man sich z.B. mit der Dreifingerregel leicht veranschaulicht.)

Der antisymmetrische Anteil der JACOBI-Matrix sollte also etwas mit einer Drehbewegung zu tun habe, genauer gesagt mit deren Geschwindigkeitsvektor, denn dieser zeigt in Richtung der Tangente der Bahnkurve.

Um dies genauer zu untersuchen, betrachten wir zunächst eine einfache Drehung eines Vektors um die z-Achse mit Drehwinkel φ. Diese ist aufgrund der Definition von Sinus und Cosinus ausdrückbar durch die lineare Abbildung

$$\begin{aligned} x &\mapsto x\cos\varphi - y\sin\varphi \\ y &\mapsto x\sin\varphi + y\cos\varphi \ . \\ z &\mapsto z \end{aligned}$$

Als nächstes wollen wir einen Punkt $\boldsymbol{x} \in \mathbb{R}^3$ um eine beliebige Achse drehen; der Drehwinkel sei weiterhin φ. Dazu sei $\vec{a}$ der Einheitsvektor in Richtung der Achse; er sei so orientiert, daß der Drehsinn (bei positivem Winkel φ) gleich dem Gegenuhrzeigersinn ist. Dann gelten in den auf $\vec{a}$ senkrecht stehenden Ebenen im wesentlichen dieselbe Formeln, falls wir nur die Koordinateneinheitsvektoren der x- und der y-Achse ersetzen durch zwei aufeinander und auf $\vec{a}$ senkrecht stehende gleichlange Vektoren. Als ein solcher Vektor bietet sich der Verbindungsvektor $\vec{r}$ von der Achse zum Punkt $\boldsymbol{x}$ an, d.h. jener Vektor, der den Abstand zwischen $\boldsymbol{x}$ und der Achse beschreibt. Ein sowohl auf $\vec{r}$ als auch auf $\vec{a}$ senkrecht stehender Vektor ist $\vec{s} = \vec{a} \times \vec{r}$; da $\vec{a}$ die Länge eins hat, ist $\vec{s}$ genauso lang wie $\vec{r}$.

In der von $\vec{r}$ und $\vec{s}$ aufgespannten Ebene bildet die Drehung um den Winkel φ somit den Vektor $\alpha\vec{r} + \beta\vec{s}$ ab auf

$$(\alpha\cos\varphi - \beta\sin\varphi)\,\vec{r} + (\alpha\sin\varphi + \beta\cos\varphi)\,\vec{s}\ .$$

Nun zerlegen wir den Ortsvektor $\vec{x}$ eines Punktes $\boldsymbol{x}$ in einen Vektor $\vec{x}_a$ auf der Achse und den radialen, d.h. senkrecht auf der Achse stehenden, Anteil $\vec{r}$.

$\vec{x}_a$ ist die Projektion von $\vec{x}$ auf die Drehachse; da wir $\vec{a}$ als Einheitsvektor vorausgesetzt haben, ist dies nach den allgemeinen Regeln der Vektorrechnung einfach

$$\vec{x}_a = (\vec{a}\cdot\vec{x})\,\vec{a}\ ,$$

denn die Länge des Vektors wird bei der Projektion mit dem Cosinus des Winkels zur Achse multipliziert, und die Richtung des auf die Achse projezierten Vektors ist natürlich gleich der Richtung der Achse.

Damit läßt sich auch die radiale Komponente problemlos berechnen: Da beide Komponenten zusammen den Vektor $\vec{x}$ ergeben müssen, ist

$$\vec{r} = \vec{x} - \vec{x}_a = \vec{x} - (\vec{a}\cdot\vec{x})\,\vec{a}\ .$$

Nun ist $\vec{x}_a$ parallel zu $\vec{a}$, und das Vektorprodukt zweier paralleler Vektoren verschwindet. Deshalb können wir auch schreiben

$$\vec{s} = \vec{a}\times\vec{r} = \vec{a}\times\Big(\vec{x} - (\vec{x}\cdot\vec{a})\vec{a}\Big) = \vec{a}\times\vec{x}\ .$$

Da die Axialkomponente eines Vektors bei Drehungen erhalten bleibt, wird speziell der Vektor $\vec{x} = \vec{x}_a + \vec{r}$ also abgebildet auf

$$\begin{aligned} &\vec{x}_a + \cos\varphi\,\vec{r} + \sin\varphi\,\vec{s} \\ &= (\vec{x}\cdot\vec{a})\,\vec{a} + \cos\varphi\,\Big(\boldsymbol{x} - (\vec{x}\cdot\vec{a})\,\vec{a}\Big) + \sin\varphi\,\vec{a}\times\vec{x} \\ &= (1-\cos\varphi)(\vec{x}\cdot\vec{a})\,\vec{a} + \cos\varphi\,\vec{x} + \sin\varphi\,\vec{a}\times\vec{x}\ . \end{aligned}$$

Um den Geschwindigkeitsvektor einer Drehung auszurechnen, müssen wir diese zunächst zu einem dynamischen Prozeß machen, d.h. anstelle einer Drehung mit konstantem Drehwinkel φ betrachten wir eine Drehung mit Winkelgeschwindigkeit ω. Wir gehen wieder aus von einem festen Punkt $\boldsymbol{x}$ und betrachten dessen Position zum Zeitpunkt t, d.h. wir betrachten die Funktion

$$\vec{f}\colon \begin{cases} \mathbb{R} & \to \mathbb{R}^3 \\ t & \mapsto (1-\cos\omega t)(\vec{x}\cdot\vec{a})\,\vec{a} + \cos\omega t\,\vec{x} + \sin\omega t\,\vec{a}\times\vec{x}\ . \end{cases}$$

Der Geschwindigkeitsvektor im Punkt $\boldsymbol{x}$ ist die Ableitung dieser Funktion im Nullpunkt, also der Vektor

$$\frac{d\vec{f}}{dt}(0) = \Big(\omega\sin\omega t\,(\vec{x}\cdot\vec{a})\vec{a} - \omega\sin\omega t\,\vec{x} + \omega\cos\omega t\,\vec{a}\times\vec{x}\Big)_{|t=0} = \omega\vec{a}\times\vec{x}\ .$$

Mit anderen Worten: Bei der Drehung mit Winkelgeschwindigkeit ω um die von $\vec{a}$ aufgespannte Achse ist der Geschwindigkeitsvektor im Punkt $\boldsymbol{x}$ gleich $\omega\vec{a}\times\vec{x}$.

Der Vektor $\omega\vec{a}$ läßt sich auffassen als eine vektorielle Winkelgeschwindigkeit (die in der Physik in der Tat genau so definiert wird), und für einen beliebigen Vektor $\vec{\omega} \in \mathbb{R}^3$ ist das Vektorfeld

$$\vec{V}\colon \mathbb{R}^3 \to \mathbb{R}^3;\quad \boldsymbol{x} \mapsto \vec{\omega}\times\vec{x}$$

das Feld der Geschwindigkeitsvektoren zu einer Rotation mit vektorieller Winkelgeschwindigkeit $\vec{\omega}$.

Es sei noch darauf hingewiesen, daß $\operatorname{rot}\vec{V}$ in der englischsprachigen Literatur meist als $\operatorname{curl}\vec{V}$ geschrieben wird nach dem englischen Wort *curl* = Locke.

Wie Gradient und Divergenz können wir auch die Rotation formal mit Hilfe des Operators ∇ ausdrücken: Für ein dreidimensionales Vektorfeld $\vec{V}$ ist

$$\nabla\times\vec{V} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \times \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} \frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z} \\ \frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x} \\ \frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \end{pmatrix} = \operatorname{rot}\vec{V}\ .$$

e) Erste Beispiele

Zum besseren Verständnis von Divergenz und Rotation und um den Umgang mit den beiden Begriffen zu üben, wollen wir sie für zwei einfache Felder ausrechnen: Für das elektrische Feld einer Punktladung und für das Magnetfeld eines stromdurchflossenen Leiters.

1) Das elektrische Feld einer Punktladung: Eine Punktladung q im Nullpunkt erzeugt nach dem COULOMBschen Gesetz ein elektrisches Feld $\vec{E}$ der Stärke

$$\vec{E}(\boldsymbol{x}) = \frac{q}{4\pi\varepsilon_0}\,\frac{\vec{x}}{|\vec{x}|^3} = \frac{q}{4\pi\varepsilon_0|\vec{x}|^3}\begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

denn die Feldstärke nimmt mit der Entfernung vom Nullpunkt quadratisch ab und hat die Richtung (oder – je nach Vorzeichen von q – auch Gegenrichtung) des Ortsvektors., d.h. der Einheitsvektor der Richtung ist $\pm\vec{x}/|\vec{x}|$, und er muß noch durch $|\vec{x}|^2$ dividiert werden.

Das Feld $\vec{E}$ ist nur definiert auf $\mathbb{R}^3 \setminus \{(0,0,0)\}$; im Nullpunkt haben wir einen Nenner Null, und auch physikalisch ist es nicht sinnvoll, nach dem Feld *im Innern* einer Punktladung zu fragen: Punktladungen sind nun einmal mathematische Idealisierungen, und wirklich anwendbar ist das COULOMBsche Gesetz ohnehin erst ab einem gewissen Mindestabstand: Bei zu kleinen Abständen spielen andere Wechselwirkungen eine wesentlich größere Rolle.

Zur Berechnung der Divergenz müssen wir die erste Komponente nach x, die zweite nach y und die dritte nach z ableiten und dann die drei Ableitungen addieren. Beachten wir, daß $|\vec{x}|^3 = (x^2+y^2+z^2)^{3/2}$ ist, gibt uns die Quotientenregel sofort das Ergebnis

$$\begin{aligned}\frac{\partial E_1}{\partial x}(\boldsymbol{x}) &= \frac{q}{4\pi\varepsilon_0}\,\frac{\partial}{\partial x}\,\frac{x}{|\vec{x}|^3} = \frac{q}{4\pi\varepsilon_0}\,\frac{\|\vec{x}\|^3 - x\cdot\frac{3}{2}\cdot 2x|\vec{x}|}{|\vec{x}|^6} \\ &= \frac{q}{4\pi\varepsilon_0}\,\frac{|\vec{x}|^2 - 3x^2}{|\vec{x}|^5} = \frac{q}{4\pi\varepsilon_0|\vec{x}|^5}(-2x^2+y^2+z^2)\,.\end{aligned}$$

Entsprechend ist

$$\frac{\partial E_2}{\partial y}(\boldsymbol{x}) = \frac{q}{4\pi\varepsilon_0|\vec{x}|^5}(x^2-2y^2+z^2)$$

und

$$\frac{\partial E_3}{\partial z}(\boldsymbol{x}) = \frac{q}{4\pi\varepsilon_0|\vec{x}|^5}(x^2+y^2-2z^2)\,,$$

also

$$\begin{aligned}\operatorname{div}\vec{E}(\boldsymbol{x}) &= \frac{q}{4\pi\varepsilon_0|\vec{x}|^5}(-2x^2+y^2+z^2+x^2-2y^2+z^2+x^2+y^2-2z^2) \\ &= 0\,.\end{aligned}$$

Zur Berechnung der Rotation benötigen wir die Einträge der JACOBI-Matrix, die nicht in der Diagonale stehen; dabei werden die Komponenten von $\vec{E}$ nach jenen Variablen abgeleitet, die *nicht* im Zähler stehen. Die Situation ist also stets dieselbe, und es genügt, wenn wir beispielsweise die Ableitung von E_1 nach y ausrechnen.

Aus den üblichen Ableitungsregeln folgt sofort, daß

$$\frac{\partial E_1}{\partial y} = \frac{q}{4\pi\varepsilon_0}\left(-\frac{3}{2}\right)\frac{x\cdot 2y}{|\vec{x}|^5} = -\frac{3q}{4\pi\varepsilon_0}\,\frac{xy}{|\vec{x}|^5}$$

ist, und analog ist auch

$$\frac{\partial E_2}{\partial x} = \frac{q}{4\pi\varepsilon_0}\left(-\frac{3}{2}\right)\frac{y\cdot 2x}{|\vec{x}|^5} = -\frac{3q}{4\pi\varepsilon_0}\,\frac{xy}{|\vec{x}|^5}\,.$$

Die x-Komponente von rot $\vec{E}$ ist somit

$$\frac{\partial E_2}{\partial x} - \frac{\partial E_1}{\partial y} = 0\,,$$

und genauso überzeugt man sich auch vom Verschwinden der anderen Komponenten.

Somit ist also das Feld einer Punktladung außerhalb des Punktes selbst sowohl quellen- als auch wirbelfrei, genau wie es nach den MAXWELLschen Gleichungen auch sein muß.

2) Das Magnetfeld eines stromdurchflossenen Leiters: Das Magnetfeld eines von einem Strom der Stärke I durchflossenen geradlinigen Drahts ist nach dem BIOT-SAVARTschen Gesetz

$$\vec{H}(\boldsymbol{x}) = \frac{I}{4\pi\left|\vec{r}\right|^2}\,\vec{a}\times\vec{x}\,,$$

wobei $\vec{a}$ der Einheitsvektor in Richtung des Stroms und $\vec{r}$ der Abstandsvektor zwischen dem Draht und dem Punkt $\vec{x}$ ist. $\vec{H}$ ist überall im $\mathbb{R}^3$ erklärt, außer auf der von $\vec{a}$ aufgespannten Geraden, wo $\vec{r}$ der Nullvektor ist.

Zum einfacherer Rechnen wählen wir das Koordinatensystem so, daß $\vec{a}$ der Einheitsvektor in Richtung der positiven z-Achse ist; für den Punkt mit Koordinaten (x, y, z) ist dann

$$\vec{r} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \quad \text{und} \quad \vec{a}\times\vec{x} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix}\,,$$

d.h.

$$\vec{H}(\boldsymbol{x}) = \frac{I}{4\pi(x^2+y^2)}\begin{pmatrix} -y \\ x \\ 0 \end{pmatrix}\,.$$

Dieses Vektorfeld ist erklärt auf $\mathbb{R}^3$ minus der z-Achse; auf der z-Achse, d.h. im Innern des Leiters, bekommen wir eine Null in den Nenner.

Die drei Komponenten dieses Felds könnten wir nun wie oben nach den drei Variablen x, y, z ableiten, um so Rotation und Divergenz zu berechnen; es gibt allerdings eine einfachere Möglichkeit, die JACOBI-Matrix eines solchen Felds zu bestimmen:

Wie auch das $\vec{E}$-Feld im obigen Beispiel läßt sich $\vec{H}$ schreiben als Produkt einer skalaren Funktion mit einer *linearen* Abbildung $\mathbb{R}^3 \to \mathbb{R}^3$. Solche Funktionen, bei denen alle Nichtlinearität in einer skalaren Funktion steckt, bezeichnen wir als *quasilinear:*

Definition: Eine Funktion $\vec{f}\colon D \to \mathbb{R}^m$ auf einer Teilmenge $D \subseteq \mathbb{R}^n$ heißt *quasilinear,* wenn es eine Funktion $\varphi\colon D \to \mathbb{R}$ und eine Matrix $A \in \mathbb{R}^{n\times m}$ gibt, so daß $\vec{f}(\boldsymbol{x}) = \varphi(\boldsymbol{x})\cdot A\vec{x}$ ist.

In diesem Sinne ist $\vec{H}(x)$ quasilinear mit

$$\varphi(\boldsymbol{x}) = \frac{I}{4\pi(x^2+y^2)} \quad \text{und} \quad A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

und das obige Feld $\vec{E}$ ist quasilinear mit

$$\varphi(\boldsymbol{x}) = \frac{1}{4\pi\varepsilon_0\left|\vec{r}\right|^3} \quad \text{und} \quad A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\,.$$

Lemma: Für eine quasilineare Funktion $\vec{f}\colon D \to \mathbb{R}^m$ auf einer Teilmenge $D \subseteq \mathbb{R}^n$ ist

$$J_{\vec{f}}(\boldsymbol{x}) = \varphi(\boldsymbol{x})\cdot A + (A\vec{x})\cdot {}^t\operatorname{grad}\varphi(\boldsymbol{x})\big)\,.$$

In dieser Formel steht der erste Malpunkt für die Multiplikation eines Skalars mit einer Matrix, der zweite für das Matrixmultiplikation des *Spaltenvektors* $A\vec{x}$ mit dem zum *Zeilenvektor* transponierten Gradienten. Es geht hier also nicht um ein Skalarprodukt, sondern um eine Matrixmultiplikation, deren Ergebnis eine $n\times m$-Matrix ist.

Ausgeschrieben wird die Formel zu

$$\frac{\partial f_i}{\partial x_j} = \varphi(\boldsymbol{x})\cdot a_{ij} + \left(\sum_{\nu=1}^{n} a_{i\nu}x_\nu\right)\cdot\frac{\partial\varphi}{\partial x_j}\,.$$

Zum *Beweis* leiten wir $f_i(\boldsymbol{x}) = \varphi(\boldsymbol{x})\sum_{\nu=1}^{n} a_{i\nu}x_\nu$ nach x_j ab; nach der Produktregel ist, wie gewünscht,

$$\begin{aligned}\frac{\partial f_i}{\partial x_j}(\boldsymbol{x}) &= \varphi(\boldsymbol{x})\frac{\partial}{\partial x_j}\left(\sum_{\nu=1}^{n} a_{i\nu}x_\nu\right) + \frac{\partial\varphi}{\partial x_j}(\boldsymbol{x})\sum_{\nu=1}^{n} a_{i\nu}x_\nu \\ &= \varphi(\boldsymbol{x})\cdot a_{ij} + \frac{\partial\varphi}{\partial x_j}(\boldsymbol{x})\sum_{\nu=1}^{n} a_{i\nu}x_\nu\,.\end{aligned}$$

■

Dieses Lemma wenden wir gleich an auf das Feld $\vec{H}(\boldsymbol{x})$; hier ist

$$\frac{\partial\varphi}{\partial x}(\boldsymbol{x}) = \frac{-2x\cdot I}{4\pi(x^2+y^2)^2} \quad \text{und} \quad \frac{\partial\varphi}{\partial y}(\boldsymbol{x}) = \frac{-2y\cdot I}{4\pi(x^2+y^2)^2}\,,$$

d.h.

$$\operatorname{grad}\varphi(\boldsymbol{x}) = \frac{-2I}{4\pi(x^2+y^2)^2}\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Somit ist

$$\begin{aligned}
J_{\vec{H}}(\boldsymbol{x}) &= \frac{I}{4\pi(x^2+y^2)}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
&\quad + \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}\cdot\frac{-2I}{4\pi(x^2+y^2)^2}\begin{pmatrix} x & y & 0 \end{pmatrix} \\
&= \frac{I}{4\pi(x^2+y^2)^2}\left(\begin{pmatrix} 0 & -x^2-y^2 & 0 \\ x^2+y^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right. \\
&\quad \left. + \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix}\cdot\begin{pmatrix} -2x & -2y & 0 \end{pmatrix}\right) \\
&= \frac{I}{4\pi(x^2+y^2)^2}\left(\begin{pmatrix} 0 & -x^2-y^2 & 0 \\ x^2+y^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right. \\
&\quad \left. + \begin{pmatrix} 2xy & 2y^2 & 0 \\ -2x^2 & -2xy & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) \\
&= \frac{I}{4\pi(x^2+y^2)^2}\begin{pmatrix} 2xy & y^2-x^2 & 0 \\ y^2-x^2 & -2xy & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

Diese Matrix ist symmetrisch und hat Spur Null, d.h auch hier ist

$$\operatorname{div}\vec{H}(\boldsymbol{x}) = 0 \quad \text{und} \quad \operatorname{rot}\vec{H}(\boldsymbol{x}) = \vec{0}\,,$$

und auch dies natürlich wieder in Übereinstimmung mit den MAXWELL-schen Gleichungen.

Da die Feldlinien des hier betrachteten Felds Kreise sind, mag es auf den ersten Blick seltsam erscheinen, daß die Rotation von $\vec{H}$ verschwindet. Der Grund liegt darin, daß alle Kreise ihren Mittelpunkt auf der z-Achse haben, die wir in der obigen Rechnung ausschließen mußten: Die Rotation in einem Punkt $\boldsymbol{x}$ mißt die Rotation um $\boldsymbol{x}$, und beim betrachteten Feld rotiert eben alles um die z-Achse, so daß die Rotation überall sonst verschwindet.

f) Allgemeine Rechenregeln

Nachdem wir gerade mit relativ großem Aufwand für zwei Beispiele Divergenz und Rotation berechnet haben, wollen wir nun einige Rechenregeln herleiten, die uns unter anderem zeigen werden, daß zumindest ein Teil der obigen Rechnungen hätte vermieden werden können.

Wir wollen uns als nächstes überlegen, was passiert, wenn wir zwei der drei Operatoren grad, div und rot hintereinanderausführen – sofern dies möglich ist.

Der Gradient einer skalaren Funktion $f\colon D \to \mathbb{R}$ auf einer offenen Teilmenge $D \subseteq \mathbb{R}^n$ ist ein Vektorfeld; es ist also möglich, dessen Divergenz und Rotation zu berechnen. Um keine Probleme mit der Existenz und der Reihenfolge von Ableitungen zu haben, setzen wir voraus, daß f in $\mathcal{C}^2(D,\mathbb{R})$ liegt; dann ist

$$\operatorname{div}\operatorname{grad} f = \operatorname{div}\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}\,.$$

Die rechte Seite schreiben wir auch kurz als

$$\Delta f \underset{\text{def}}{=} \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}$$

und bezeichnen

$$\Delta \underset{\text{def}}{=} \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$$

als LAPLACE-Operator.

(Der LAPLACE-Operator ist älter als Divergenz, Gradient, Rotation und so weiter; er spielt eine große Rolle sowohl bei Wellen als auch bei Potentialfeldern. HAMILTON wählte das Zeichen ∇ in Anlehnung an diesen Operator; formal kann man $\Delta = \nabla \cdot \nabla$ schreiben.)

Auch die Rotation eines Gradientenfelds läßt sich leicht berechnen, allerdings natürlich nur im $\mathbb{R}^3$, da wir für andere Dimensionen keine Rotation definiert haben.

$$\operatorname{rot}\operatorname{grad} f = \operatorname{rot}\begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f}{\partial z \partial y} - \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial x \partial z} - \frac{\partial^2 f}{\partial z \partial x} \\ \frac{\partial^2 f}{\partial y \partial x} - \frac{\partial^2 f}{\partial x \partial y} \end{pmatrix} = \vec{0}\,.$$

Die Rotation eines stetig differenzierbaren Gradientenfelds ist also stets Null; da das elektrische Feld der negative Gradient des elektrischen Potentials ist, hätten wir oben also für

$$\vec{E}(\boldsymbol{x}) = \frac{1}{4\pi\varepsilon_0}\frac{\vec{x}}{\left|\vec{x}\right|^3} = -\operatorname{grad}\frac{1}{4\pi\varepsilon_0\left|\vec{x}\right|}$$

auf die Berechnung der Rotation verzichten können.

Die Divergenz eines Vektorfelds ist eine skalare Funktion; der einzige Operator, der darauf angewandt werden kann, ist daher der Gradient, und das Ergebnis ist völlig uninteressant:

$$\operatorname{grad}\operatorname{div}\vec{V} = \operatorname{grad}\sum_{i=1}^{n}\frac{\partial V_i}{\partial x_i} = \begin{pmatrix} \sum_{i=1}^{n}\frac{\partial^2 V_i}{\partial x_1 \partial x_i} \\ \vdots \\ \sum_{i=1}^{n}\frac{\partial^2 V_i}{\partial x_n \partial x_i} \end{pmatrix}.$$

Die Rotation eines Vektorfelds im Dreidimensionalen schließlich ist wieder ein Vektorfeld, wir können also Divergenz und Rotation davon berechnen:

$$\begin{aligned}\operatorname{div}\operatorname{rot}\vec{V} &= \operatorname{div}\begin{pmatrix} \frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z} \\ \frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x} \\ \frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \end{pmatrix} \\ &= \frac{\partial^2 V_3}{\partial x \partial y} - \frac{\partial^2 V_2}{\partial x \partial z} + \frac{\partial^2 V_1}{\partial y \partial z} - \frac{\partial^2 V_3}{\partial y \partial x} + \frac{\partial^2 V_2}{\partial z \partial x} - \frac{\partial^2 V_1}{\partial z \partial y} = 0\end{aligned}$$

für ein zweimal stetig differenzierbares Vektorfeld. Da ein Magnetfeld als Rotation eines Vektorpotentials geschrieben werden kann, erklärt dies, warum im obigen Beispiel $\operatorname{div}\vec{H} = 0$ war.

Bleibt noch die Rotation der Rotation; für diese ist nach Aufgabe 1 des elften Übungsblatts

$$\operatorname{rot}\operatorname{rot}\vec{V} = \operatorname{grad}\operatorname{div}\vec{V} - \begin{pmatrix} \Delta V_1 \\ \Delta V_2 \\ \Delta V_3 \end{pmatrix}.$$

Ebenfalls dort sind auch die beiden „Produktregeln“

$$\operatorname{div}(f\vec{V}) = f\operatorname{div}\vec{V} + (\operatorname{grad} f)\cdot\vec{V}$$

und

$$\operatorname{rot}(f\vec{V}) = f\operatorname{rot}\vec{V} + (\operatorname{grad} f)\times\vec{V}$$

zu finden. Mit dem Operator ∇ geschrieben werden sie völlig analog zur klassischen LEIBNIZ-Regel:

$$\nabla\cdot(f\vec{V}) = f\nabla\cdot\vec{V} + \nabla f\cdot\vec{V}$$

und

$$\nabla\times(f\vec{V}) = f\nabla\times\vec{V} + \nabla f\times\vec{V}$$

Speziell für ein lineares Vektorfeld

$$\vec{V}\colon\mathbb{R}^n\to\mathbb{R}^n;\quad \boldsymbol{x}\mapsto A\vec{x}\quad\text{mit}\quad A\in\mathbb{R}^{n\times n}$$

ist

$$\operatorname{div}\vec{V} = \sum_{i=1}^{n}\frac{\partial V_i}{\partial x_i} = \sum_{i=1}^{n} a_{ii} = \operatorname{Spur} A$$

und, im Falle $n = 3$,

$$\operatorname{rot}\vec{V} = \begin{pmatrix} a_{32} - a_{23} \\ a_{13} - a_{31} \\ a_{21} - a_{12} \end{pmatrix};$$

für eine quasilineare Funktion $\vec{f}(\boldsymbol{x}) = \varphi(\boldsymbol{x})A\vec{x}$ ist also nach den beiden Produktregeln

$$\operatorname{div}\vec{f} = \varphi\operatorname{Spur} A + (\operatorname{grad}\varphi)\cdot A\vec{x}$$

und

$$\operatorname{rot} \vec{f} = \varphi \cdot \begin{pmatrix} a_{32} - a_{23} \\ a_{13} - a_{31} \\ a_{21} - a_{12} \end{pmatrix} + (\operatorname{grad} \varphi) \times A\vec{x}\,,$$

was uns in den Beispielen des vorigen Abschnitts manche Rechnung erspart hätte.

g) Nichtkartesische Koordinatensysteme

Auch in diesem letzten Abschnitt von §2 soll es um Möglichkeiten gehen, wie wir bei den oben betrachteten und anderen Beispielen effizienter rechnen können. Das Feld einer Punktladung beispielsweise ist kugelsymmetrisch, jedoch haben wir bei der Berechnung von Divergenz und Rotation dieses Feldes die Kugelsymmetrie mutwillig zerstört, indem wir in einem kartesischen Koordinatensystem rechneten. Hier wollen wir uns beschäftigen, wie man Koordinatensysteme finden kann, die einer solchen Symmetrie angepaßt sind.

Entsprechend der Vielzahl möglicher Symmetrien kennt die Mathematik auch eine Vielzahl solcher Koordinatensysteme; wir wollen uns auf die beiden wichtigsten un am häufigsten angewandten beschränken: die Kugel- und die Zylinderkoordinaten. Zum Einstieg in das Thema betrachten wir aber zunächst die etwas einfachere Situation der Polarkoordinaten in der Ebene.

1) Polarkoordinaten in $\mathbb{R}^2$***:*** Ein Punkt $P \in \mathbb{R}^2$ wird charakterisiert durch seien Abstand r vom Nullpunkt O sowie den Winkel φ zwischen dem Vektor $\overrightarrow{OP}$ und dem Einheitsvektor der (positiven) x-Achse, wobei dieser Winkel für $P = O$ nicht definiert ist. Ansonsten messen wir ihn im Bogenmaß und wählen den Drehsinn so, die Richtung von der positiven x-Achse zur positiven y-Achse positiv gerechnet wird.

Für einen Punkt mit kartesischen Koordinaten (x, y) ist dann

$$r = \sqrt{x^2 + y^2}$$

und für $(x, y) \neq (0, 0)$ ist φ so bestimmt, daß gilt

$$x = r \cos \varphi \quad \text{und} \quad y = r \sin \varphi\,.$$

φ ist dabei, wie üblich, natürlich nur modulo 2π bestimmt. Da Sinus und Cosinus nur auf Intervallen der Länge π injektiv sind, reicht eine der beiden Umkehrfunktionen allein nicht aus, um φ direkt anzugeben, sondern man braucht noch mindestens eine Fallunterscheidung. Eine mögliche Formel für φ, auf Basis des Arkuskosinus, ist etwa

$$\varphi = \begin{cases} \arccos \frac{x}{r} & \text{falls } y \geq 0 \\ -\arccos \frac{x}{r} & \text{falls } y < 0 \end{cases}.$$

Natürlich kann man entsprechende Formeln auch mit dem Arkussinus oder Arkustangens aufstellen.

Anstelle des Gitternetzes eines kartesischen Koordinatensystems hat man bei den Poalrkoordinaten ein Netz aus Kreisen um den Nullpunkt, auf denen der Radius r konstant ist, und aus vom Nullpunkt ausgehenden Strahlen, auf denen φ konstant ist.

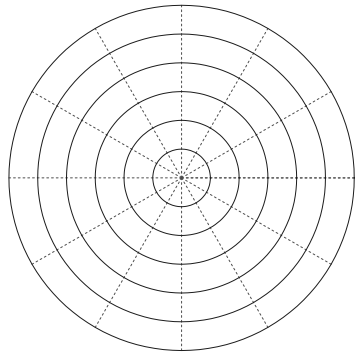


Abb. 43: Das Netz der Polarkoordinaten

Für eine Funktion $f(x, y)$ auf einer Teilmenge des $\mathbb{R}^2$ können wir die Polarkoordinatendarstellung

$$F(r, \varphi) \underset{\text{def}}{=} f(r \cos \varphi, r \sin \varphi)$$

betrachten; wir wollen sehen, wie man mit F allein immer noch den

Gradienten oder beispielsweise auch die LAPLACEsche Δf berechnen kann.

Dazu müssen wir zunächst die partiellen Ableitungen von f nach x und y durch die von F nach r und φ ausdrücken. Dies ist auf direkte Weise etwas schwierig, da wir keine gute Formel für die Berechnung von φ aus x und y haben. Die umgekehrte Aufgabe dagegen ist umso einfacher: Nach der Kettenregel ist

$$\frac{\partial F}{\partial r} = \frac{\partial}{\partial r} f(r\cos\varphi, r\sin\varphi) = \frac{\partial f}{\partial x}\cos\varphi + \frac{\partial f}{\partial y}\sin\varphi$$

und

$$\frac{\partial F}{\partial \varphi} = \frac{\partial}{\partial \varphi} f(r\cos\varphi, r\sin\varphi) = -\frac{\partial f}{\partial x} r\sin\varphi + \frac{\partial f}{\partial y} r\cos\varphi\,.$$

Da $r > 0$ ist, können wir dies auch in der Form

$$\begin{aligned} F_r &= f_x\cos\varphi + f_y\sin\varphi \\ \frac{F_\varphi}{r} &= -f_x\sin\varphi + f_y\cos\varphi \end{aligned}$$

schreiben. Dies ist ein lineares Gleichungssystem für f_x und f_y; wenn wir von der mit $\cos\varphi$ multiplizierten ersten Gleichung die mit $\sin\varphi$ multiplizierte zweite subtrahieren, erhalten wir

$$f_x = F_r\cos\varphi - \frac{F_\varphi}{r}\sin\varphi\,;$$

entsprechend erhalten wir

$$f_y = F_r\sin\varphi + \frac{F_\varphi}{r}\cos\varphi\,,$$

wenn wir die mit $\sin\varphi$ multiplizierte erste Gleichung zur mit $\cos\varphi$ multiplizierten zweiten addieren.

Entsprechend können wir auch die zweiten partiellen Ableitungen von f durch partielle Ableitungen von F ausdrücken, allerdings werden hier die Ausdrücke schon deutlich komplexer, so daß hier auf die detailierte Rechnung verzichtet sei. Wichtig ist vor allem der LAPLACE-Operator,

und wenn es auch sehr umständlich wäre, dessen Darstellung in Polarkoordinaten *herzuleiten* können wir die fertige Formel

$$\Delta f = f_{xx} + f_{yy} = F_{rr} + \frac{F_r}{r} + \frac{F_{\varphi\varphi}}{r^2}\,,$$

doch leicht und relativ schnell beweisen kann, indem wir

$$F_{rr} = \big(f_{xx}\cos\varphi + f_{xy}\sin\varphi\big)\cos\varphi + \big(f_{xy}\cos\varphi + f_{yy}\sin\varphi\big)\sin\varphi$$

und

$$\begin{aligned} F_{\varphi\varphi} = {} & \big(f_{xx} r\sin\varphi - f_{xy} r\cos\varphi\big) r\sin\varphi - f_x r\cos\varphi \\ & - \big(f_{xy} r\sin\varphi - f_{yy} r\cos\varphi\big) r\cos\varphi - f_y r\sin\varphi \end{aligned}$$

nach der Kettenregel berechnen und einsetzen.

2) Zylinderkoordinaten im $\mathbb{R}^3$***:*** Für den $\mathbb{R}^3$ gibt es zwei naheliegende Verallgemeinerung der Polarkoordinaten: Einmal können wir in der (x, y)-Ebene Polarkoordinaten einführen, die z-Koordinate aber unverändert lassen, zum anderen können wir neben dem Abstand vom Nullpunkt zwei Winkelkoordinaten einführen, wie z.B. die geographische Länge und Breite. Hier soll es zunächst um den ersten Fall gehen, wir haben also drei Koordinaten (r, φ, z) mit $r > 0$, die über die Gleichungen

$$x = r\cos\varphi, \quad y = r\sin\varphi \quad \text{und} \quad z = z$$

mit den kartesischen Koordinaten verbunden sind. Diese Koordinaten heißen *Zylinderkoordinaten,* da die Flächen r = *konstant* Zylinder sind. Insbesondere ist die Koordinate r hier nicht mehr gleich dem Abstand eines Punktes vom Nullpunkt, sondern gleich dem Abstand von der z-Achse.

Da die z-Koordinate des kartesischen Koordinatensystems übernommen wird, brauchen wir für diese Koordinaten nicht neu zu rechnen, sondern können die obigen Ergebnisse praktisch wörtlich übernehmen: Mit

$$F(r, \varphi, z) = f(r\cos\varphi, r\sin\varphi, z)$$

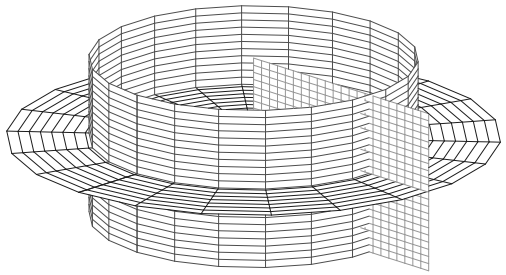


Abb. 44: Flächen mit konstantem r, φ und z für Zylinderkoordinaten

ist

$$f_x = F_r \cos\varphi - \frac{F_\varphi}{r}\sin\varphi$$
$$f_y = F_r \sin\varphi + \frac{F_\varphi}{r}\cos\varphi$$
$$f_z = F_z$$

und

$$\Delta f = f_{xx} + f_{yy} + f_{zz} = F_{rr} + \frac{F_r}{r} + \frac{F_{\varphi\varphi}}{r^2} + F_{zz}\,.$$

3) Kugelkoordinaten: Hier führen wir zusätzlich zu den beiden Polarkoordinaten der Ebenen als dritte Koordinate noch eine *Winkelkoordinate* ϑ ein, wir haben also drei Koordinaten (r, φ, ϑ), wobei r wieder positiv sein muß und den Abstand vom Nulpunkt bezeichnet.

Im Gegensatz zur Konvention der Geographen, wonach der Äquator, d.h. der Schnittkreis einer Kugel um den Nullpunkt mit der (x, y)-Ebenen, die Breite Null hat, ist es in der Mathematik, Physik und Technik üblich, dem Äquator die Koordinate $\vartheta = \pi/2$ zu geben, dem „Nordpol", d.h. dem Schnittpunkt mit der positiven z-Achse, $\vartheta = 0$ und dem Südpol $\vartheta = \pi$. In der (x, z)-Ebene ist also

$$x = r\sin\vartheta \quad \text{und} \quad z = r\cos\vartheta\,,$$

in der (y, z)-Ebene

$$y = r\sin\vartheta \quad \text{und} \quad z = r\cos\vartheta\,,$$

und ganz allgemein ist

$$R = r\sin\vartheta \quad \text{und} \quad z = r\cos\vartheta\,,$$

wobei R den Abstand eines Punktes von der z-Achse bezeichnet.

Für einen Punkt im Abstand R von der z-Achse ist

$$x = R\cos\varphi \quad \text{und} \quad y = R\sin\varphi\,,$$

zusammen mit den obigen Formeln erhalten wir also als Zusammenhang zwischen Kugelkoordinaten und kartesischen Koordinaten

$$x = r\cos\varphi\sin\vartheta$$
$$y = r\sin\varphi\sin\vartheta$$
$$z = r\cos\vartheta\,.$$

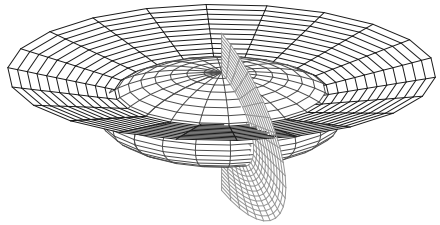


Abb. 45: Flächen mit konstantem r, φ und ϑ für Kugelkoordinaten

Die Flächen mit konstantem r sind hier natürlich Kugeln, konstantes φ führt wie bei den Zylinderkoordinaten auf Halbebenen, und die Flächen mit konstantem ϑ schließlich sind Kegel um die z-Achse.

Wieder können wir einer Funktion $f(x, y, z)$ in kartesischen Koordinaten durch

$$F(r, \varphi, \vartheta) \underset{\text{def}}{=} f(r\cos\varphi\sin\vartheta, r\sin\varphi\sin\vartheta, r\cos\vartheta)$$

eine neue Funktion zuordnen und durch Anwendung der Kettenregel sowie Lösen eines linearen Gleichungssystems die partiellen Ableitungen

zueinander in Beziehung setzen; die Vorgehensweise ist im wesentlichen wie bei den ebenen Polarkoordinaten, nur daß wir jetzt ein System aus drei linearen Gleichungen bekommen:

$$\begin{aligned} F_r &= f_x \cos\varphi \sin\vartheta + f_y \sin\varphi \sin\vartheta + f_z \cos\vartheta \\ \frac{F_\varphi}{r} &= -f_x \sin\varphi \sin\vartheta + f_y \cos\varphi \sin\vartheta \\ \frac{F_\vartheta}{r} &= f_x \cos\varphi \cos\vartheta + f_y \sin\varphi \cos\vartheta - f_z \sin\vartheta\,. \end{aligned}$$

Der GAUSS-Algorithmus führt schnell auf die Lösung

$$\begin{aligned} f_x &= F_r \cos\varphi \sin\vartheta - \frac{F_\varphi \sin\varphi}{r \sin\vartheta} + \frac{F_\vartheta \cos\varphi \cos\vartheta}{r} \\ f_y &= F_r \sin\varphi \sin\vartheta - \frac{F_\varphi \cos\varphi}{r \sin\vartheta} + \frac{F_\vartheta \sin\varphi \cos\vartheta}{r} \\ f_z &= F_r \cos\vartheta - \frac{F_\vartheta \sin\vartheta}{r}\,. \end{aligned}$$

Im Prinzip genauso können auch die zweiten Ableitungen berechnet werden; eine grausame, wenn auch nicht sonderlich schwierige Rechnung führt beispielsweise auf die Formel

$$\begin{aligned} \Delta f &= f_{xx} + f_{yy} + f_{zz} \\ &= F_{rr} + \frac{2F_r}{r} + \frac{F_{\varphi\varphi}}{r^2 \sin^2\vartheta} + \frac{F_{\vartheta\vartheta}}{r^2} + \frac{F_\vartheta}{r^2 \tan\vartheta}\,. \end{aligned}$$

Als Beispiel wollen wir damit noch einmal die Divergenz des elektrischen Felds einer Ladung q im Nullpunkt berechnen: Diese erzeugt ein Potential

$$U(r,\varphi,\vartheta) = \frac{q}{4\pi\varepsilon_0 r}\,,$$

das Feld ist $\vec{E} = -\operatorname{grad} U$ und

$$\operatorname{div}\vec{E} = -\operatorname{div}\operatorname{grad} U = -\Delta U = -\frac{q}{4\pi\varepsilon_0}\left(\frac{2}{r^3} + \frac{2}{r}\cdot\frac{-1}{r^2}\right) = 0\,.$$

§ 3: Integralrechnung

Die Integralrechnung ist neben der Differentialrechnung die zweite wichtige Säule der Analysis. Wir betrachten zunächst den eindimensionalen Fall, der sich auch für die mehrdimensionale Theorie an mehreren Stellen als wichtig erweisen wird.

a) Heuristische Vorüberlegungen

Seiner großen Bedeutung entsprechend, wollen wir uns dem Integralbegriff zunächst heuristisch nähern, bevor wir ihn – mit beträchtlichem technischem Aufwand – im nächsten Paragraphen formal einführen.

Die Integration dient primär drei Zwecken:

- Sie dient zur Umkehrung der Differentiation.
- Sie dient zur Flächenberechnung.
- Sie dient zur Durchschnittsberechnung.

Wie wollen uns alle drei Aspekte kurz ansehen.

1) Integration als Umkehrung der Differentiation: Das klassische Beispiel zur Einführung der Differentiation ist die Geschwindigkeit als Ableitung des Wegs nach der Zeit. Es liegt daher nahe, zur Einführung des Integrals die Frage nach der Berechnung des Wegs aus der Geschwindigkeit zu betrachten.

Wir nehmen also an, ein Fahrzeug sei zur Zeit $t = 0$ an der Stelle $s = 0$ und beschleunige in fünf Sekunden auf die im Stadtverkehr zulässige Hchstgeschwindigkeit von 13,9 m/sec. Welchen Weg hat es bis dahin zurückgelegt?

Die Geschwindigkeit sei durch folgende Tabelle gegeben:

t [sec]	0	1	2	3	4	5
v [m/sec]	2,5	6,1	9,1	11,5	13,4	13,9

Ist $s(t)$ der bis zum Zeitpunkt t zurückgelegte Weg und $v(t)$ die dann erreichte Geschwindigkeit, so ist bekanntlich

$$v(t) = \frac{ds}{dt} \approx \frac{\Delta s}{\Delta t} = \frac{s(t + 1\,sec) - s(t)}{(t + 1\,sec) - t} = \frac{s(t + 1\,sec) - s(t)}{1\,sec}\,.$$