vom eindimensionalen Fall leiten: Für eine zweimal differenzierbare Funktion $f\colon \mathbb{R} \to \mathbb{R}$ ist

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + o(h^2)\,;$$

mit Hilfe der zweiten Ableitung läßt sich die Funktion also nicht nur linear, sondern auch durch eine quadratische Funktion approximieren; der Approximationsfehler wird dadurch naturgemäß kleiner und geht jetzt für $h \to 0$ sogar schneller gegen Null als h^2.

Im Mehrdimensionalen mußten wir den Term $hf'(x)$ ersetzen durch eine lineare Funktion

$$L(\vec{h}) = \vec{a} \cdot \vec{h} \quad \text{mit} \quad \vec{a} = \operatorname{grad} f\,;$$

entsprechend müssen wir nun den Term $\frac{h^2}{2} f''(x)$ ersetzen durch eine quadratische Funktion

$$\frac{1}{2} Q(\vec{h}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} h_i h_j = \frac{1}{2} {}^t\vec{h} A \vec{h} \quad \text{mit} \quad A = (a_{ij})\,.$$

Die Matrix A können wir dabei als symmetrisch ansetzen, denn da $h_i h_j = h_j h_i$ ist, ändert sich nichts an der Summe, wenn wir a_{ij} und a_{ji} beide durch ihr arithmetisches Mittel ersetzen:

$$a_{ij} h_i h_j + a_{ji} h_j h_i = \frac{a_{ij} + a_{ji}}{2} h_i h_j + \frac{a_{ij} + a_{ji}}{2} h_j h_i\,.$$

Definition: Die Funktion $f\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ heißt zweifach differenzierbar im Punkt $\boldsymbol{x} \in D$, wenn es eine Linearform $L(\vec{h}) = \vec{a} \cdot \vec{h}$ und eine quadratische Form $Q(\vec{h}) = {}^t\vec{h} A \vec{h}$ gibt, so daß

$$f(\boldsymbol{x} + \vec{h}) = f(\boldsymbol{x}) + L(\vec{h}) + \frac{1}{2} Q(\vec{h}) + o\bigl(|\vec{h}|^2\bigr)$$

ist. Die Matrix A, so sie als symmetrische Matrix gewählt wird, heißt dann HESSE-Matrix von f im Punkt $\boldsymbol{x}$ und wird mit $H_f(\boldsymbol{x})$ bezeichnet.

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

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

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

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

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

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

also

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

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

Insbesondere ist also zumindest in diesem Fall $f_{xy} = f_{yx}$.

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 präzieser 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 28 dargestellte Funktion

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

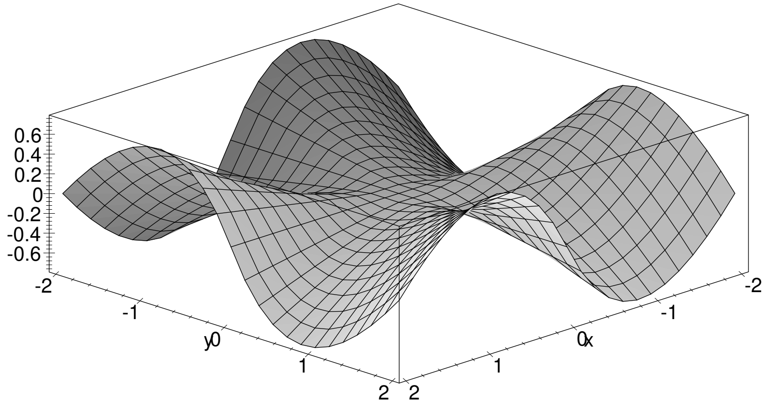


Abb. 28: Ein Gegenbeispiel zum Vertauschungssatz

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

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

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

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 $\tilde{\xi}$ zwischen x und $x+h$, so daß dies gleich $\tilde{g}'(\tilde{\xi})$ ist. Also ist

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

für eine Zahl $\tilde{\eta}$ zwischen y und $y+h$. Somit haben wir gezeigt, daß

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

ist.

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

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 dies 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, §4*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, und die können auch noch als benachbart angenommen werden, denn eine beliebige Transposition läßt sich immer als Produkt von Transpositionen benachbarter Elemente schreiben. 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_1 bis x_{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 $o\big(|\vec{h}|^k\big)$ durch ein Polynom k-ten Grades approximiert werden kann. (Wie das 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 recht schnell unübersichtlich wird.)

Die Aussage, daß eine Funktion $f \in \mathcal{C}^k(D, \mathbb{R})$ bis auf Fehler der Größenordnung $o\big(|\vec{h}|^k\big)$ durch ein Polynom approximiert werden kann, ist natürlich das mehrdimensionale Analogon der TAYLOR-Entwicklung im Eindimensionalen; wir können TAYLOR-Polynome einigermaßen übersichtlich hinschreiben, wenn wir Richtungsableitungen benutzen: Die Richtungsableitung von f 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.*

Zumindest für $n = 2$ ist alles noch einigermaßen handhabbar; hier sehen wir sofort, daß für $f \in \mathcal{C}^2(D,\mathbb{R})$ gilt

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

und nach dem Lemma von SCHWARZ ist dies eine symmetrische Matrix.

§3: Grundlegende Sätze

a) Linearität und Produktregel

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

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

Auch die LEIBNIZ-Regel verallgemeinert sich, wie diese Woche auf dem Übungsblatt nachgerechnet wird, zu

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

b) Kurven und Tangentenvektoren

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

Wir begnügen uns hier mit der Verknüpfung zwischen skalarwertigen Funktionen mehrerer Veränderlicher mit Funktionen einer Variablen nach $\mathbb{R}^n$. Eine solche Funktion $\gamma\colon (a,b) \to \mathbb{R}^n$ hat als Bild ein *Kurvenstück* im $\mathbb{R}^n$ und wird deshalb auch so bezeichnet:

Wir hatten schon in §2 Kurvenstücke als differenzierbare Abbildungen eines Intervalls in den $\mathbb{R}^n$ eingeführt; hier wollen wir allgemeiner auch noch ganze *Kurven* einführen:

Definition: *a)* Ein *Kurvenstück* γ ist eine stetig differenzierbare Abbildung

$$\gamma\colon [a,b] \to \mathbb{R}^n$$

eines Intervalls in den $\mathbb{R}^n$.
b) Eine *Kurve* γ ist eine endliche Folge von Kurvenstücken

$$\gamma_1, \ldots, \gamma_r\colon [a_i, b_i] \to \mathbb{R}^n$$

von Kurvenstücken mit der Eigenschaft, daß

$$\gamma_i(b_i) = \gamma_{i+1}(a_{i+1}) \quad \text{für } i = 1, \ldots, n-1.$$

c) Eine Kurve γ heißt geschlossen, wenn $\gamma_r(b_r) = \gamma_1(a_1)$ ist.

Die Bedingung im Teil *b)* der obigen Definition stellt sicher, daß Kurven, anschaulich betrachtet, zusammenhängend sind; es hat allein praktische Gründe, daß man nicht auch von den Intervallen verlangt, daß sie unmittelbar aneinander anschließen: Oft werden die Formeln für ein Kurvenstück einfacher, wenn man einen bestimmten Anfangswert wie etwa die Null für das Intervall nehmen kann.

Ein wesentlicher Unterschied zwischen einer Kurve und einem Kurvenstück liegt in der Differenzierbarkeit: Durch Verschiebung der Parameterintervalle könnte man jede Kurve durch eine stetige Abbildung $\gamma\colon [a,b] \to \mathbb{R}^n$ beschreiben, aber dort, wo zwei Kurvenstücke aneinanderstoßen, muß diese Abbildung nicht differenzierbar sein; anschaulich gesprochen kann die Kurve dort einen „Knick“ haben.

Die Ableitung $\dot\gamma(t)$ ist ein Vektor, dessen Komponenten die Ableitungen der Koordinatenfunktionen von $\gamma(t)$ sind; wir bezeichnen ihn als

Tangentenvektor der Kurve im Punkt $\gamma(t)$. Gelegentlich wird es wichtig sein, daß dieser Vektor ungleich dem Nullvektor ist; wir definieren daher

Definition: *a)* Ein Kurvenstück $\gamma\colon [a,b] \to \mathbb{R}^n$ heißt *regulär,* wenn $\dot\gamma(t) \neq 0$ für alle $t \in (a,b)$.
b) Eine Kurve γ heißt *stückweise regulär,* wenn sie aus regulären Kurvenstücken zusammengesetzt werden kann.

Man beachte, daß für die Intervallendpunkte nichts gefordert wird: Selbst wenn die Ableitung dort existiert, muß sie nicht von Null verschieden kann. Auf diese Weise läßt sich ein Kurvenstück, bei dem $\dot\gamma(t)$ an endlich vielen Stellen gleich dem Nullvektor ist, immer noch als stückweise reguläre Kurve auffassen.

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

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

aber natürlich wäre auch

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

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

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

liegen sämtliche Bildpunkte auf dem Einheitskreis, und zumindest für jedes endliche Teilintervall von $\mathbb{R}$ definiert auch γ_3 ein Kurvenstück. Wie eine kurze Kurvendiskussion (oder ein Blick auf Abbildung 29) zeigt, besteht das Bild von γ_3 aus allen Punkten des Einheitskreises außer $(1,0)$; letzterer Punkt ist der Grenzwert von $\gamma_3(t)$ für $t \to \pm\infty$.

Jede dieser Parametrisierungen führt zu anderen Tangentenvektoren: Für einen festen Kurvenpunkt liegen zwar alle drei Vektoren auf ein

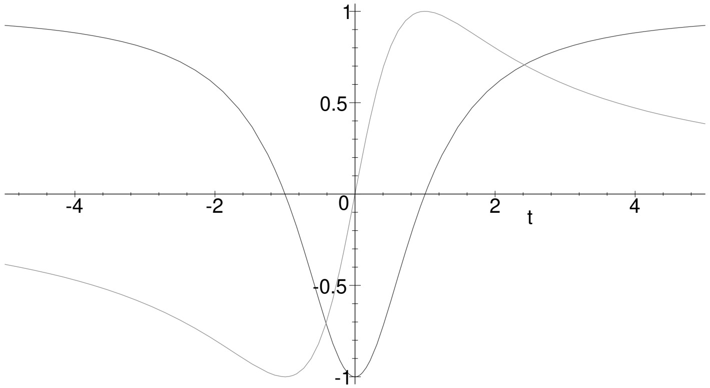


Abb. 29: Graph der Funktionen $x(t) = \frac{t^2-1}{t^2+1}$ und $y(t) = \frac{2t}{t^2+1}$

und derselben Geraden, der Tangenten an den Kreis, aber sie haben verschiedene Länge: Für $(x, y) = \gamma_1(t)$ ist

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

für $(x, y) = \gamma_2(t)$ ist

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

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

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

ist. Bei der Parametrisierung mit γ_2 sind die Tangentenvektoren also jeweils 2π mal so lang wie bei der mit γ_1, was man anschaulich so interpretieren kann, daß der Kreis bei dieser Parametrisierung 2π mal so schnell durchlaufen wird wie bei der mit γ_1. Bei der Parametrisierung mit γ_3 ist die Länge der Vektoren variabel, und sie zeigen auch in die Gegenrichtung zu den Tangentenvektoren an γ_1 und γ_2; bei dieser Parametrisierung wird der Kreis also im Uhrzeigersinn durchlaufen.

c) Eine Kettenregel

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

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

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

Beweis: Nach Definition der Ableitung ist

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

Da für jede der Komponenten γ_i

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

ist, ist auch

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

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

Damit ist, wegen der stetigen Differenzierbarkeit von f,

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

womit die Behauptung bewiesen wäre. ∎

d) Der Satz über implizite Funktionen

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

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

für $c \neq 0$ durch

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

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

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

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

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

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

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

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

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

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

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

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

Die Funktion f ist stetig und differenzierbar; ihre Ableitung ist

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

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

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

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

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

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

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

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

Damit ist f als Funktion festgelegt, und nach Konstruktion ist

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

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

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

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

ist.

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

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

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

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

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

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

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

und setzen wir

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

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 die zweite Ableitung in den Punkten, in denen die erste verschwindet; dort vereinfacht sich die Formel zu

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

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

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

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

Für alle $i \leq n-1$ ist

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

■

§4: Vektorwertige Funktionen mehrer Veränderlicher

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

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

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 30 und 31.

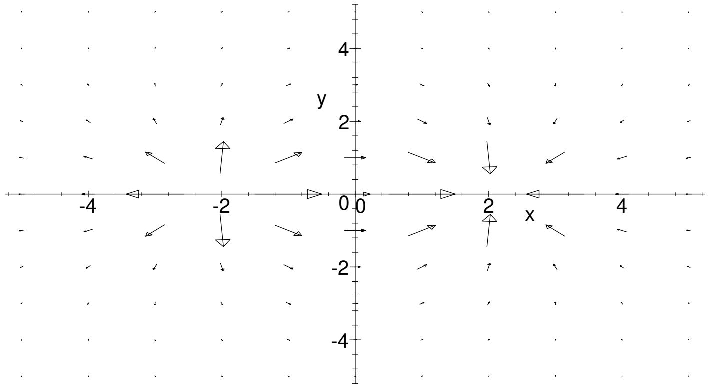


Abb. 30: Das elektrische Feld zweier entgegengesetzt gleicher Punktladungen

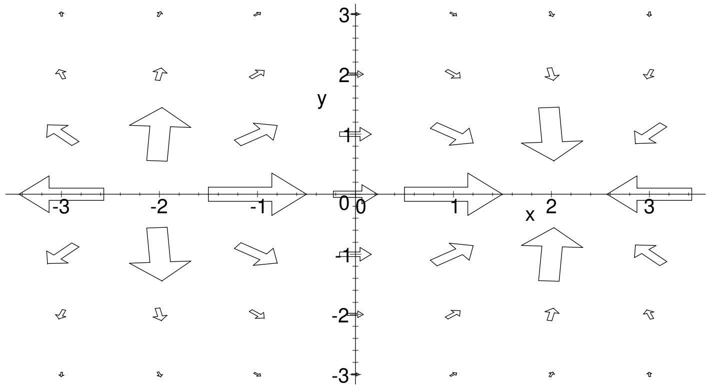


Abb. 31: Dasselbe Feld in einem kleineren Bereich mit „flächigen“ Vektoren

Wie Abbildung 30 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 31 die Vektoren „flächig" zeichnet, wobei die Länge des Vektors proportional zur *Fläche* des eingezeichneten Pfeils ist, aber auch das hilft nicht sehr viel.

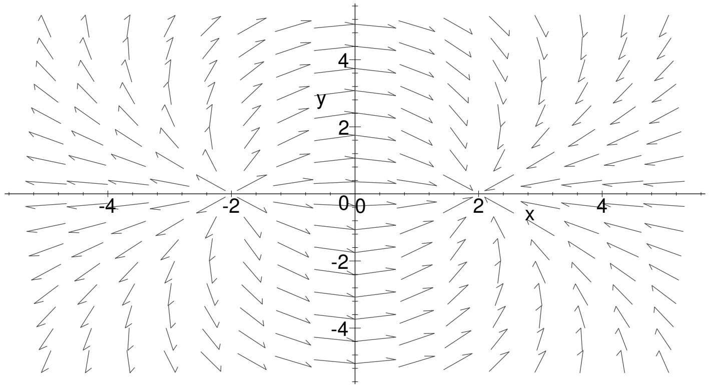


Abb. 32: Das Richtungfeld zum obigen Feld

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 32. 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 $f(\boldsymbol{x}) = (0, \dots, 0)$ 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 mehrerer 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

Wir werden uns in diesem Paragraphen zwar hauptsächlich mit Vektorfeldern im $\mathbb{R}^n$ beschäftigen, und uns selbst da teilweise auf den Fall $n = 3$ beschränken; zumindest die grundlegenden Definitionen sollen aber für beliebige Funktionen auf einer Teilmenge D von $\mathbb{R}^n$ mit Werten in irgendeinem $\mathbb{R}^m$ betrachtet werden.

Fast von selbst versteht sich die

Definition: Eine Funktion $\vec{f}\colon D \to \mathbb{R}^m$ auf $D \subseteq \mathbb{R}^n$ heißt stetig *bzw.* differenzierbar, 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. Damit wissen wir also, was Differenzierbarkeit einer Funktion $\vec{f}\colon D \to \mathbb{R}^n$ bedeutet; was aber ist ihre Ableitung?

Bei skalarwertigen Funktionen einer oder mehrerer Veränderlicher hatten wir die Ableitung jeweils als Linearisierung der Funktion in der Umgebung des betrachteten Punkts definiert; genauso gehen wir auch hier vor: Für jede der Komponentenfunktion f_i ist

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

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

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

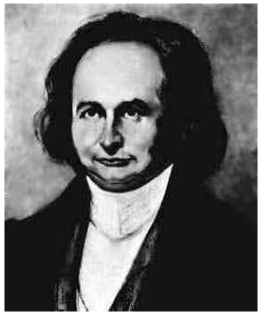
In dieser Notation können wir die als Zeilenvektoren aufgefaßten Gradienten zu einer $n \times n$-Matrix zusammenfassen, der JACOBI-Matrix

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

Mit dieser Matrix ist dann

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

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



CARL GUSTAV JACOB JACOBI (1804–51) 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 war 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.

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

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

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

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

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:

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