$\{\vec{b}_1, \dots, \vec{b}_n\}$ die Lösungsmenge des linearen Gleichungssystems

$$x_{r+1} = d_{r+1}, \quad x_{r+2} = d_{r+2}, \quad \dots, \quad x_n = d_n \,.$$

Durch Rücktransformation auf die Standardbasis des k^n kann man dieses lineare Gleichungssystem in ein lineares Gleichungssystem bezüglich der ursprünglichen Basis umformen.

Spätestens die Interpretation als Lösungsmenge eines linearen Gleichungssystems sollte jedem klargemacht haben, daß der Durchschnitt zweier affiner Unterräume entweder leer ist oder wieder ein affiner Unterraum; in beiden Fällen kann er durch Lösen eines linearen Gleichungssystems ermittelt werden. Seine Dimension hängt offensichtlich von einer ganzen Reihe von Faktoren ab: Schon in der Ebene kann der Durchschnitt zweier Geraden entweder leer sein (parallele Geraden) oder nulldimensional (ein Schnittpunkt) oder eindimensional (zusammenfallende Geraden).

Definition: Für zwei affine Unterräume (A, U) und (B, V) eines affinen Raums (C, W) ist der von A und B aufgespannte affine Unterraum von (C, W) der kleinste affine Unterraum (S, T) von (C, W), der beide enthält.

Bei dieser Definition stellt sich als erstes die Frage, ob sie überhaupt sinnvoll ist, ob es einen solchen kleinsten Raum also auch wirklicht gibt. Dazu bilden wir den Durschnitt aller affiner Unterräume, die (A, U) und (B, V) als Unterräume enthalten. Der Durchschnitt einer beliebigen Menge affiner Unteräume ist entweder leer oder wieder ein affiner Unterraum. Leer kann er hier nicht sein, da wir nur Unterräume bertrachten, die die beiden gegebenen Räume enthalten, also ist er ein affiner Unterraum und offensichtlich der kleinste unter denen, die (A, U) und (B, V) als Unterräume enthalten.

Seine Dimension kann allerdings größer werden als die Summe der Dimensionen der beiden Ausgangsräume; ein wahrscheinlich aus der Schule bekanntes Beispiel dafür sind *windschiefe Geraden* im $\mathbb{R}^3$, d.h zwei Geraden, die zwar nicht parallel zueinander sind, die aber in zwei voneinander verschiedenen parallelen Ebenen liegen.

Ein Beispiel hierfür sind etwa die x-Achse eines kartesischen Koordiantensystems im $\mathbb{R}^3$ und die um eins in z-Richtung verschobene y-Achse. Diese beiden Geraden schneiden sich nicht, denn alle Punkte der ersten haben z-Koordinate null, während auf der zweiten $z = 1$ ist, und sie sind auch offensichtlich nicht parallel.

Wir wollen uns überlegen, daß es keine Ebene gibt, die beide Geraden enthält. Dazu gehen wir aus von der allgemeinsten Gleichung

$$ax + by + cz = d$$

für eine Ebene im $\mathbb{R}^3$. Wenn diese Ebene die x-Achse enthalten soll, also die Menge aller Punkte der Form $(x, 0, 0)$, muß $a = d = 0$ sein. Die um eins in z-Richtung verschobene y-Achse besteht aus allen Punkten der Form $(0, y, 1)$; sie liegt in der Ebenen, wenn $b = 0$ und $c = d$ ist. Für eine Ebene, die beide Geraden enthält, müßte also $a = b = c = d = 0$ sein, und dann haben wir keine Ebene mehr. Da es zwischen einer Ebenen und dem gesamten $\mathbb{R}^3$ keine weiteren affinen Unterräume mehr gibt, ist der kleinste affine Unterraum, der die beiden Geraden enthält, der gesamte $\mathbb{R}^3$.

Damit drängt sich die Frage auf, ob es möglicherweise auch im $\mathbb{R}^4$ oder in Räumen noch höherer Dimension zwei Geraden geben kann, die den gesamten Raum aufspannen. Da die Anschauung hier leider nicht weiterhilft, müssen wir abstrakt mathematisch argumentieren. Das wird am einfachsten, wenn wir rechnen können, und dazu führen wir Koordinatensysteme ein:

Definition: Ein Koordinatensystem des affinen Raums (A, V) besteht aus einem Punkt $O \in A$, genannt der Ursprung oder auch Nullpunkt des Koordinatensystems, und einer Basis von V. Die Geraden durch O mit einem Basisvektor als Richtungsvektor heißen *Koordinatenachsen.*

Bezüglich eines solchen Koordinatensystems hat dann in der Tat jeder Punkt $\boldsymbol{x}$ eines n-dimensionalen affinen Raums A ein n-tupel von Koordinaten, nämlich die Komponenten des Vektors $\overrightarrow{O\boldsymbol{x}}$. Bevor wir aber zeigen können, daß der affine Raum dadurch isomorph wird zu k^n, müssen wir zunächst definieren, was strukturerhaltende Abbildungen zwischen affinen Räumen überhaupt sein sollen:

Definition: Eine affine Abbildung zwischen den affinen Räumen (A, V) und (B, W) besteht aus einer Abbildung $\varphi\colon A \to B$ und einer linearen Abbildung $\psi\colon V \to W$ derart, daß für alle Punkte $\boldsymbol{x}, \boldsymbol{y} \in A$ gilt:

$$\overrightarrow{\varphi(\boldsymbol{x})\,\varphi(\boldsymbol{y})} = \psi\left(\overrightarrow{\boldsymbol{x}\boldsymbol{y}}\right) .$$

Die Abbildung heißt *Isomorphismus* oder *Affinität,* falls φ bijektiv ist.

Offensichtlich ist φ genau dann bijektiv, wenn ψ ein Isomorphismus von Vektorräumen ist, so daß wir dies nicht zusätzlich fordern müssen. Es ist auch klar, daß nach Wahl eines Koordinatensystems eines n-dimensionalen affinen Raums die Koordinaten einen Isomorphismus mit dem affinen Raum k^n definieren.

Da wir wissen, wie lineare Abbildungen zwischen Vektorräumen bezüglich zweier Basen aussehen, können wir auch sofort hinschreiben, wie die Abbildung $\varphi\colon A \to B$ bezüglich zweier Koordinatensysteme aussieht:

$$(x_1, \ldots, x_n) \mapsto \Big(\sum_{j=1}^{n} a_{1j}x_j + b_1, \ldots, \sum_{j=1}^{n} a_{mj}x_j + b_m\Big) ,$$

im Gegensatz zu den homogenen linearen Abbildungen zwischen Vektorräumen sind hier also auch inhomogene lineare Abbildungen möglich. In Matrixschreibweise betrachtet man am besten die Spaltenvektoren: Ist

$$\varphi\big((x_1, \ldots, x_n)\big) = (y_1, \ldots, y_m) ,$$

so ist

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \vec{b} \quad \text{mit} \quad \vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} .$$

Geometrisch betrachtet beschreibt der konstante Vektor $\vec{b}$ also eine Verschiebung.

Vor allem in der Computergraphik faßt man die Matrix A und den Vektor $\vec{b}$ oft auch zusammen zu einer Matrix A^* mit $m+1$ Zeilen und $n+1$ Spalten, indem man zunächst den Vektor $\vec{b}$ als $(n+1)$-te Spalte an

die Matrix anhängt und dann als $(m+1)$-te Zeile lauter Nullen einsetzt mit Ausnahme einer Eins an der letzten Stelle:

$$A^* = \begin{pmatrix} A & \vec{b} \\ \mathbf{0} & 1 \end{pmatrix} ,$$

wobei die fette Null für n gewöhnliche Nullen steht. Mit dieser Bezeichnung ist dann

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \\ 1 \end{pmatrix} = A^* \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{pmatrix} ,$$

eine Formel, die zwar rechnerisch aufwendiger ist als die obige, dafür aber etwas kompakter. Sie hat auch den Vorteil, daß sich die Hintereinanderausführung affiner Abbildungen so durch ein Matrixprodukt beschreiben läßt.

Als Anwendung von Koordinatensystemem wollen wir, wie bereits erwähnt, die Dimension des Erzeugnisses zweier affiner Unterräume berechnen.

Betrachten wir zunächst die Dimension des Erzeugnisses zweier Untervektorräume:

Definition: Die Summe $U + V$ zweier Untervektorräume eines festen Vektorraums ist der kleinste Untervektorraum, der beide enthält.

Mit der Notation aus Kapitel I können wir dies auch als $U + V = [U \cup V]$ schreiben, denn für jede Teilmenge M eines Vektorraums hatten wir $[M]$ als kleinsten Untervektorraum definiert, der M enthält.

Lemma: Sind U und V endlichdimensional, so ist

$$\dim(U + V) = \dim U + \dim V - \dim(U \cap V) .$$

Beweis: Wir starten mit einer Basis $\{\vec{b}_1, \ldots, \vec{b}_r\}$ des Durchschnitts $U \cap V$ und ergänzen diese zu einer Basis $\{\vec{b}_1, \ldots, \vec{b}_n\}$ von U und zu einer Basis $\{\vec{b}_1, \ldots, \vec{b}_r, \vec{c}_{r+1}, \ldots, \vec{c}_m\}$ von V. Dann ist $U + V$ der von den Vektoren $\vec{b}_i$ und $\vec{c}_j$ erzeugte Untervektorraum, denn einen kleineren

Untervektorraum, der U und V enthält, kann es nicht geben. Außerdem sind diese Vektoren linear unabhängig, denn aus

$$\lambda_1\vec{b}_1 + \cdots + \lambda_n\vec{b}_n + \mu_{r+1}\vec{c}_{r+1} + \cdots + \mu_m\vec{c}_m = \vec{0}\,,$$

folgt

$$\vec{v} \underset{\text{def}}{=} \lambda_1\vec{b}_1 + \cdots + \lambda_n\vec{b}_n = -\left(\mu_{r+1}\vec{c}_{r+1} + \cdots + \mu_m\vec{c}_m\right) \in U \cap V\,.$$

Damit müssen λ_{r+1} bis λ_n verschwinden, denn Basis des Untervektorraums $U \cap V$ von U sind $\vec{b}_1$ bis $\vec{b}_r$. Dann bleibt aber eine Gleichung übrig, die nur noch $\vec{b}_1$ bis $\vec{b}_r$ und die $\vec{c}_j$ enthält, also die Basisvektoren von V. Da diese linear unabhängig sind, müssen auch alle noch übrigen Koeffizienten verschwinden.

Somit ist $\{\vec{b}_1, \ldots, \vec{b}_n, \vec{c}_{r+1}, \ldots, \vec{c}_m\}$ eine Basis von $U + V$, d.h.

$$\dim(U + V) = n + (m - r) = \dim U + \dim V - \dim(U \cap V)\,. \qquad \blacksquare$$

Bei Untervektorräumen ist damit alles klar; kehren wir zurück zur den affinen Unterräumen; wir betrachten also zwei affinen Unterräume (A, U) und (B, V) eines affinen Raums (C, W) sowie den davon erzeugten affinen Unterraum (S, T).

Falls A und B nichtleeren Durchschnitt haben, können wir Koordinatensysteme mit einem gemeinsamen Nullpunkt finden und W ist der Vektorraum $U + V$, d.h. in diesem Fall ist

$$\begin{aligned}\dim S = \dim T = \dim(U + V) &= \dim U + \dim V - \dim(U \cap V)\\ &= \dim A + \dim B - \dim(A \cap B)\,,\end{aligned}$$

denn $A \cap B$ hat ein Koordinatensystem mit demselben Nullpunkt und daher ist der zugehörige Untervektorraum $U \cap V$.

Falls A und B aber disjunkt sind, müssen wir verschiedene Punkte O_A und O_B als Nullpunkte der entsprechenden Koordinatensysteme wählen; jetzt ist der Vektorraum T des von A und B aufgespannten affinen Raums (S, T) ganz sicher nicht mehr $U + V$, denn er muß nun auch den Verbindungsvektor von O_A und O_B enthalten. Wäre dieser

als Summe $\vec{u} + \vec{v}$ von Vektoren $\vec{u} \in U$ und $\vec{v} \in V$ darstellbar, könnten wir den Vektor $\vec{u}$ an O_A abtragen und $-\vec{v}$ an O_B; beide Vektoren hätten denselben Endpunkt, der somit im (leeren) Durchschnitt von A und B liegen müßte. Also ist T mindestens gleich $U + V$ plus dem vom Verbindungsvektor aufgespannten eindimensionalen Untervektorraum. Das reicht aber auch schon, denn nimmt man etwa O_A als Ursprung des Koordinatensystems, kommt man nun durch Abtragen der Vektoren aus diesem Untervektorraum zu jedem Punkt von A oder B. Also ist in diesem Fall

$$\dim S = \dim T = 1 + \dim(U + V) = \dim U + \dim V - \dim(U \cap V) + 1\,.$$

Insgesamt haben wir also gezeigt

Lemma: (A, U) und (B, V) seien affine Unterräume eines affinen Raum (C, W), und (S, T) sei der von beiden erzeugte affine Unterraum. Dann ist

$$\dim S = \begin{cases} \dim A + \dim B - \dim(A \cap B) & \text{falls } A \cap B \neq \emptyset \\ \dim A + \dim B - \dim(U \cap V) + 1 & \text{falls } A \cap B = \emptyset\,. \end{cases} \qquad \blacksquare$$

Betrachten wir als Beispiel die beiden Unterräume

$$A = \left\{ (1,0,1,0,1) + \lambda \begin{pmatrix} 1\\2\\3\\4\\5 \end{pmatrix} + \mu \begin{pmatrix} 5\\4\\3\\2\\1 \end{pmatrix} \;\middle|\; \lambda, \mu \in \mathbb{R} \right\}$$

und

$$B = \left\{ (0,1,0,1,0) + \lambda \begin{pmatrix} 0\\1\\2\\3\\4 \end{pmatrix} + \mu \begin{pmatrix} 0\\1\\4\\9\\16 \end{pmatrix} \;\middle|\; \lambda, \mu \in \mathbb{R} \right\}$$

von $\mathbb{R}^5$. Der Untervektorraum zu A ist

$$U = \left[\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}, \begin{pmatrix}5\\4\\3\\2\\1\end{pmatrix}\right] = \left[\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}, \begin{pmatrix}6\\6\\6\\6\\6\end{pmatrix}\right] = \left[\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}\right]$$

$$= \left[\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}\right],$$

der zu B ist

$$V = \left[\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}0\\1\\4\\9\\16\end{pmatrix}\right] = \left[\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}0\\0\\2\\6\\12\end{pmatrix}\right].$$

Der Durchschnitt der beiden Untervektorräume ist also eindimensional und wird vom gemeinsamen Basisvektor aufgespannt.

Zur Berechnung des Durchschnitts von A und B müssen wir untersuchen, ob es reelle Zahlen $\lambda_1, \mu_1, \lambda_2, \mu_2$ gibt, so daß

$$(1,0,1,0,1) + \lambda_1 \begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix} + \mu_1 \begin{pmatrix}5\\4\\3\\2\\1\end{pmatrix} = (0,1,0,1,0) + \lambda_2 \begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix} + \mu_2 \begin{pmatrix}0\\1\\4\\9\\16\end{pmatrix}$$

ist. Mit den gerade berechneten neuen Basen von U und V können wir stattdessen auch das etwas einfachere Gleichungssystem

$$(1,0,1,0,1) + \lambda_3 \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} + \mu_3 \begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix} = (0,1,0,1,0) + \lambda_4 \begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix} + \mu_4 \begin{pmatrix}0\\0\\2\\6\\12\end{pmatrix}$$

betrachten mit neuen Variablen $\lambda_3, \mu_3, \lambda_4$ und μ_4, denn der Durchschnitt zweier affiner Unterräume hängt natürlich nicht vom Koordinatensystem

ab. Hier können wir alle Basisvektoren auf die linke Seite bringen und die beiden Basispunkte in Gestalt ihres Verbindungsvektors auf die rechte; mit den neuen Variablen $\lambda = \lambda_3$, $\mu = \mu_3 - \lambda_4$ und $\nu = -\mu_4$ wird das Gleichungssystem dann zu

$$\lambda \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} + \mu \begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix} + \nu \begin{pmatrix}0\\0\\2\\6\\12\end{pmatrix} = \begin{pmatrix}-1\\1\\-1\\1\\-1\end{pmatrix}.$$

Bei diesem System kann das für die erste Komponente nur dann gutgehen, wenn $\lambda = -1$ ist; setzen wir diesen Wert ein und betrachten dann die zweite Komponente, folgt, daß $\mu = 2$ sein muß. Für ν bleibt noch das Gleichungssystem

$$\nu \begin{pmatrix}0\\0\\2\\6\\12\end{pmatrix} = \begin{pmatrix}-1\\1\\-1\\1\\-1\end{pmatrix} + \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} - 2\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix} = \begin{pmatrix}0\\0\\-4\\-4\\-8\end{pmatrix},$$

das offensichtlich unlösbar ist. Damit war auch das ursprüngliche Gleichungssystem unlösbar, wir sind hier also im Fall $A \cap B = \emptyset$.

Nach obiger Formel hat das Erzeugnis von A und B daher die Dimension

$$\dim A + \dim B - \dim(U \cap V) + 1 = 2 + 2 - 1 + 1 = 4;$$

wie wir oben gesehen haben, enthält der zugehörige Untervektorraum außer $U + V$ auch noch den Verbindungsvektor

$$\begin{pmatrix}-1\\1\\-1\\1\\-1\end{pmatrix}$$

der Nullpunkte $(0,1,0,1,0)$ von A und $(1,0,1,0,1)$ von B.

Da wir $U + V$ und auch den Verbindungsvektor kennen, läßt sich das Erzeugnis von A und B nun leicht explizit angeben: Es ist der affine

Unterraum

$$\left\{ (1,0,1,0,1) + \lambda \begin{pmatrix} 1\\1\\1\\1\\1 \end{pmatrix} + \mu \begin{pmatrix} 0\\1\\2\\3\\4 \end{pmatrix} + \nu \begin{pmatrix} 0\\1\\4\\9\\16 \end{pmatrix} + \rho \begin{pmatrix} -1\\1\\-1\\1\\-1 \end{pmatrix} \;\middle|\; \begin{matrix} \lambda, \mu, \nu, \\ \rho \in \mathbb{R} \end{matrix} \right\}.$$

Damit ist dieser Abschnitt fast zu Ende; das einzige, was noch fehlt, ist das Hauptthema, um das es hier geht, die EUKLIDischen Räume. Die Definition sollte fast selbstverständlich sein:

Definition: Ein EUKLIDischer Raum ist gegeben durch eine Menge E zusammen mit einem EUKLIDischen Vektorraum V derart, daß (E, V) ein affiner Raum ist. Der Abstand zwischen zwei Punkten $\boldsymbol{x}, \boldsymbol{y} \in E$ ist definiert als die Länge ihres Verbindungsvektors $\overrightarrow{\boldsymbol{xy}}$, der Winkel zwischen zwei sich schneidenden Geraden als der Winkel zwischen zwei (beliebigen von $\vec{0}$ verschiedenen) Vektoren auf diesen Geraden.

In EUKLIDischen Räumen lassen sich daher alle geometrischen Konstruktionen durchführen, die aus der EUKLIDischen Geometrie von der Schule her bekannt sind; lediglich Orientierungen können nicht behandelt werden, da wir Winkel nur modulo 180° definiert haben.

d) Normen, Stetigkeit und Konvergenz

Mit Hilfe des Skalarprodukts haben wir in Abschnitt *b)* eine Norm auf einem EUKLIDischen Vektorraum definiert durch

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}\,.$$

Diese hat 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 $\mathbb{R}$-Vektorraum V zusammen mit einer Abbildung $\|\cdot\|: V \to \mathbb{R}$ mit den Eigenschaften *a)* bis *c).*

Damit ist also jeder EUKLIDische Vektorraum insbesondere auch ein normierter Vektorraum; man kann Normen aber auch noch anders definieren als über Skalarprodukte: Auch die *Maximumsnorm*

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

erfüllt 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 $\begin{pmatrix}1\\0\end{pmatrix} \star \begin{pmatrix}0\\1\end{pmatrix} = -\dfrac{1}{2}$ 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 den $\mathbb{R}^n$ 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

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

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

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

Wir betrachten Normen, um Begriffe, wie Konvergenz und Stetigkeit zu definieren; dafür haben wir nun genügend Vorbereitungen getroffen:

Definition: *a)* $\big(V, \|\cdot\|\big)$ sei ein normierter $\mathbb{R}$-Vektorraum. Eine Folge $\vec v_1, \vec v_2, \dots$ 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 $f\colon V \to W$ zwischen zwei normierten $\mathbb{R}$-Vektorräumen $\big(V, \|\cdot\|_1\big)$ und $\big(W, \|\cdot\|_2\big)$ heißt stetig in $\vec v_0 \in V$, wenn es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle $\vec v \in V$ gilt: Ist $\|\vec v - \vec v_0\| < \delta$, so ist $\|f(\vec v) - f(\vec v_0)\| < \varepsilon$.

Um zu sehen, daß wir hiermit Stetigkeit in ganz neuen Zusammenhängen betrachten können, betrachten wir die lineare Abbildung

$$\varphi\colon \mathcal{C}^0([0\,,1]\,,\mathbb{R}) \to \mathbb{R}^2; \quad f \mapsto \begin{pmatrix} f(0) \\ f(1) \end{pmatrix}.$$

Wenn wir auf beiden Vektorräumen die Maximumsnorm einführen, ist diese Abbildung stetig in jedem $f \in \mathcal{C}^0([0,\,1],\ \mathbb{R})$, denn für jedes $\varepsilon > 0$ können wir hier einfach $\delta = \varepsilon$ setzen: Ist

$$\|f-g\|_\infty = \max_{x\in[0,1]} |f(x)-g(x)| < \varepsilon\,,$$

so ist erst recht

$$|f(0)-g(0)| < \varepsilon \quad \text{und} \quad |f(0)-g(0)| < \varepsilon\,,$$

also

$$\left\| \begin{pmatrix} f(0) \\ f(1) \end{pmatrix} - \begin{pmatrix} g(0) \\ g(1) \end{pmatrix} \right\|_\infty = \max\big\{|f(0)-g(0)|\,,|f(1)-g(1)|\big\} < \varepsilon\,.$$

Die so definierte Stetigkeit ist aber noch nicht ganz das, was wir wollen: Uns geht es schließlich in erster Linie um Funktionen mit *Punkten* als Argumenten, nicht um solche von Vektoren. Deshalb müssen wir zunächst noch über die Norm von Vektoren einen Abstand von Punkten definieren:

Definition: (A, V) sei ein affiner Raum mit einem normierten $\mathbb{R}$-Vektorraum V. Für zwei Punkte $\boldsymbol{x}, \boldsymbol{y} \in A$ heißt

$$d(\boldsymbol{x}, \boldsymbol{y}) \underset{\text{def}}{=} \left\|\overrightarrow{\boldsymbol{xy}}\right\|$$

der *Abstand* von $\boldsymbol{x}$ und $\boldsymbol{y}$.

Der Abstand ist also gerade gleich der Länge des Verbindungsvektors. Aus den Eigenschaften *a)* bis *c)* einer Norm folgen sofort die drei Eigenschaften
(1) $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$
(2) $d(\boldsymbol{x}, \boldsymbol{y}) \ge 0$ und $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ genau dann, wenn $\boldsymbol{x} = \boldsymbol{y}$
(3) Die *Dreiecksungleichung* $d(\boldsymbol{x}, \boldsymbol{z}) \le d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$

(Man beachte, daß (1) nur ein sehr schwacher Abglanz von *a)* ist; allerdings wird jeder, der schon einmal einen Stadtplan mit variablem Maßtab gesehen hat, verstehen, daß man die strenge Homogenität, die der Bedingung *a)* entspricht, nicht für *jeden* Abstand fordern will.)

Definition: Eine Menge M zusammen mit einer Abbildung

$$d\colon M \times M \to \mathbb{R}$$

heißt *metrischer Raum* und d heißt *Metrik,* wenn die Bedingungen (1) bis (3) erfüllt sind.

Der Abstand definiert also eine Metrik, und wir können Konvergenz und Stetigkeit nun für beliebige metrische Räume definieren:

Definition: *a)* (M,d) sei ein metrischer Raum. Eine Folge $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ von Punkten aus M konvergiert gegen $\boldsymbol{x} \in M$, wenn es zu jedem $\varepsilon > 0$ eine natürliche Zahl $N \in \mathbb{N}$ gibt, so daß $d(\boldsymbol{x}, \boldsymbol{x}_n) < \varepsilon$ für alle $n > N$.
b) Eine Abbildung $f\colon M \to L$ zwischen zwei metrischen Räumen (M,d) und (L,e) heißt stetig in $\boldsymbol{x}_0 \in M$, wenn es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle $\boldsymbol{x} \in M$ gilt: Ist $d(\boldsymbol{x}, \boldsymbol{x}_0) < \delta$, so ist $d\big(f(\boldsymbol{x}), f(\boldsymbol{x}_0)\big) < \varepsilon$. *c)* Eine Teilmenge $U \subseteq M$ eines metrischen Raums heißt *offen,* wenn es zu jedem $\boldsymbol{x} \in U$ ein $\varepsilon > 0$ gibt, so daß jeder Punkt $\boldsymbol{y} \in M$ mit $d(\boldsymbol{x}, \boldsymbol{y}) < \varepsilon$ in U liegt.

Uns interessiert in diesem Semester vor allem der Spezialfall $M = \mathbb{R}^n$; hier kennen wir bislang zwei Normen auf dem Vektorraum $\mathbb{R}^n$ und damit auch zwei Metriken auf der Menge $\mathbb{R}^n$: Die EUKLIDische Norm

$$\|\vec{v}\| = \sqrt{v_1^2 + \cdots + v_n^2}$$

führt auf die Metrik

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2}\,,$$

und die Maximumsnorm

$$\|\vec{v}\|_\infty = \max_{i=1}^{n} |v_i|$$

führt auf die Metrix

$$d_\infty(\boldsymbol{x}, \boldsymbol{y}) = \max_{i=1}^{n} |y_i - x_i|\ .$$

Um den Unterschied zwischen den beiden Metriken zu verstehen, betrachten wir im $\mathbb{R}^2$ jeweils Menge aller Punkte, deren Abstand vom Punkt $(a,b) \in \mathbb{R}^2$ höchstens gleich ε ist.

Für die EUKLIDische Metrik ist das

$$\begin{aligned} U_\varepsilon(a,b) &= \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; \sqrt{(x-a)^2 + (y-b)^2} < \varepsilon\right\} \\ &= \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; (x-a)^2 + (y-b)^2 < \varepsilon^2\right\}, \end{aligned}$$

also eine (offene) Kreisscheibe mit Radius ε um (a,b).

Für die Maximumsnorm erhalten wir

$$\begin{aligned} V_\varepsilon(a,b) &= \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; \max\{|x-a|\,, |y-b|\} < \varepsilon\right\} \\ &= \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; |x-a| < \varepsilon \text{ und } |y-b| < \varepsilon\right\}, \end{aligned}$$

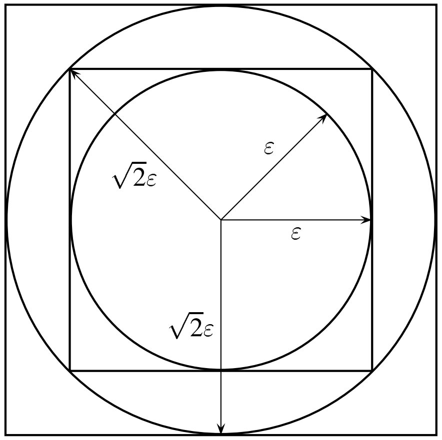


Abb. 15: Die Umgebungen U_ε und V_ε

also ein (offenes) Quadrat mit Seitenlänge 2ε und Mittelpunkt (a,b).

Wie Abbildung 15 zeigt, ist

$$U_\varepsilon(a,b) \subset V_\varepsilon(a,b) \subset U_{\sqrt{2}\varepsilon}(a,b) \subset V_{\sqrt{2}\varepsilon}(a,b)\,,$$

jede der beiden Mengen $U_\varepsilon(a,b)$ und $V_\varepsilon(a,b)$ liegt also zwischen zwei entsprechenden Mengen bezüglich der jeweils anderen Norm. Somit sind die offenen Mengen bezüglich beider Abstände dieselben, und auch Konvergenz und Stetigkeit hängen nicht davon ab, welche der beiden Metriken man verwendet: Man kann also für jede Anwendung die jeweils angenehmste verwenden.

Tatsächlich führen sogar *alle* Normen auf $\mathbb{R}^n$ zu Metriken, die auf dieselben offenen Mengen, konvergenten Folgen und stetigen Funktionen führen wie die beiden hier betrachteten; auf einen Beweis sei hier aber verzichtet.

§2: Skalarwertige Funktionen mehrer Veränderlicher

Eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}^m$ läßt sich in der Form

$$(x_1, \dots, x_n) \mapsto \big(f_1(x_1, \dots, x_n),\ \dots\ , f_m(x_1, \dots, x_n)\big)$$

schreiben mit m Funktionen $f_j\colon \mathbb{R}^n \to \mathbb{R}$. Somit spielt der Fall $m = 1$ eine fundamentale Rolle und verdient es, vorab behandelt zu werden.

Wir bezeichnen Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ als *skalarwertig,* da wir gewohnt sind, Elemente von $\mathbb{R}^n$ als Vektoren, solche von $\mathbb{R}$ aber als Skalare zu bezeichnen.

Falls wir allerdings Elemente von $\mathbb{R}^n$ wirklich als *Vektoren* betrachten wollen, entspricht dies nicht der Sichtweise der Analysis; wir unterscheiden also weiterhin zwischen Punkten und Vektoren, wobei wir Funktionen meist als Funktionen auf Punktmengen betrachten.

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 $f\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ den Graphen

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

definieren; da dieser in $\mathbb{R}^{n+1}$ liegt, ist er allerdings nur für $n = 2$ wirklich anschaulich, und auch da kann es bei komplizierten Funktionen stark von der gewählten Perspektive abhängen, was man sieht. Für einfache 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 16 etwa sieht man recht gut, daß Γ_f Teil einer Kugeloberfläche ist.

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*

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

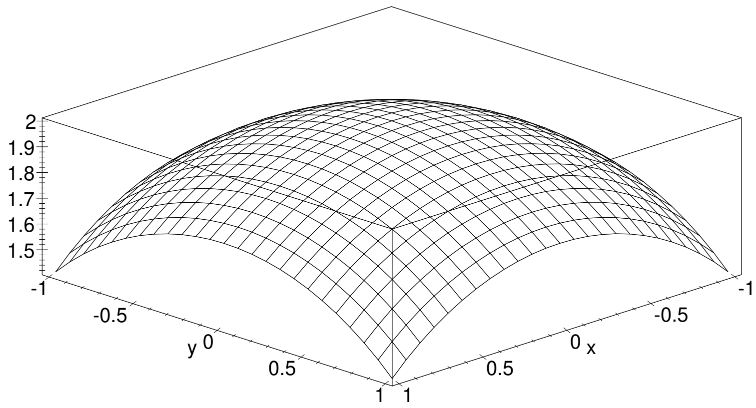


Abb. 16: Graph der Funktion $f(x,y) = \sqrt{4-x^2-y^2}$

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 17 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.

Für Funktionen von mehr als zwei Veränderlichen ist die Visualisierung naturgemäß schwieriger; wir können Graphen und Nivealinien, -flächen *usw.* zwar problemlos definieren, aber nicht mehr zeichnen – es sei denn, es handelt sich um sehr einfache Niveauflächen im $\mathbb{R}^3$.

Ansonsten gibt es für Funktionen von mehr als zwei Veränderlichen im allgemeinen keine andere Möglichkeiten, als einen Teil der Variablen auf

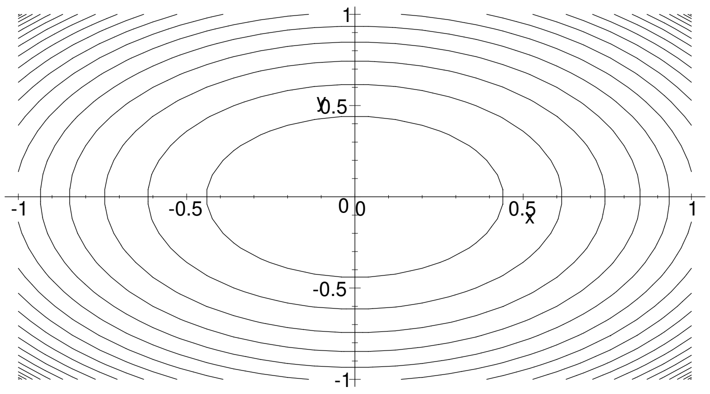


Abb. 17: Niveaulinien der Funktion $f(x, y) = \sqrt{4 - x^2 - y^2}$

interessanten Werten festzuhalten und die so eingeschränkte Funktion darzustellen. Dies gibt natürlich kein vollständiges Bild der Funktion.

Als Beispiel für die Vor- und Nachteile von Graph und Niveaulinien wollen wir die Stetigkeit einer Funktion anhand beider Darstellungen zu erraten versuchen. Zunächst die Definition:

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 $|f(\boldsymbol{y}) - f(\boldsymbol{x})| < \varepsilon$.

Anstelle der Bedingung, daß alle $|y_i - x_i| < \delta$ sein sollen, hätten wir natürlich genauso gut fordern können, daß der EUKLIDische Abstand von $\boldsymbol{x}$ und $\boldsymbol{y}$ kleiner als δ sein soll; die Zahlenwerte für δ hätten sich dadurch zwar eventuell geändert, aber da jeder n-dimensionale Würfel in eine n-dimensionale Kugel eingebettet werden kann und umgekehrt definiert beides genau die gleiche Stetigkeit.

Als Beispiel betrachten wir 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 18 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 eben aus.

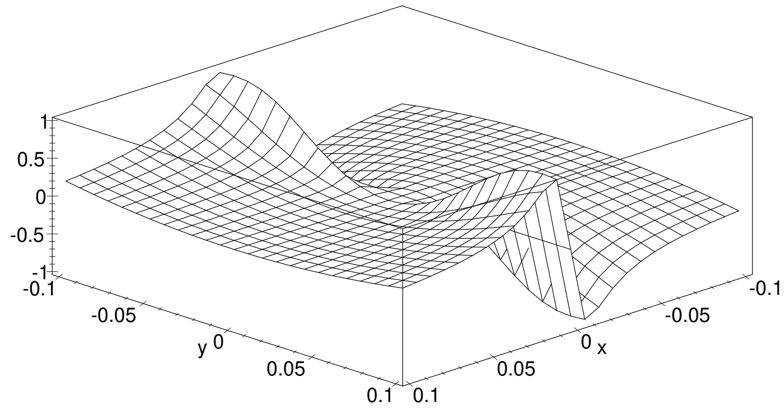


Abb. 18: Ist diese Funktion stetig?

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 $a \in \mathbb{R} \smallsetminus \{0\}$ 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 32. 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 $N_c(f) = \emptyset$.

Auch Abbildung 19 zeigt, daß f im Nullpunkt unstetig ist, denn alle Niveaulinien kommen dem Nullpunkt beliebig nahe, obwohl dieser nur

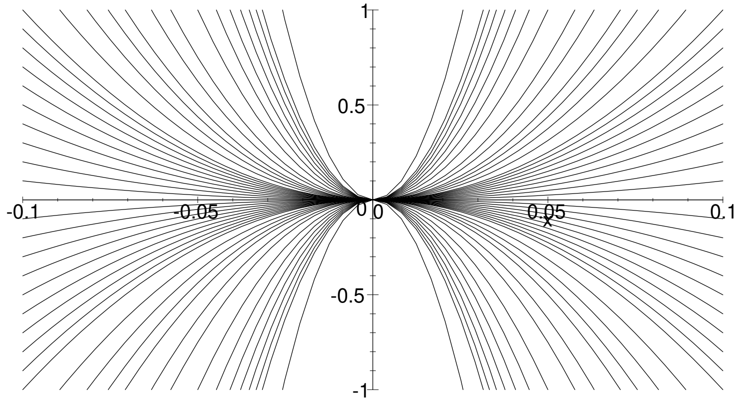


Abb. 19: Die Niveaulinien illustrieren die Unstetigkeit im Nullpunkt

auf $N_0(f)$ liegt. Daran sehen wir übrigens auch, daß der Graph in Abbildung 18 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 18 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.

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 interessierendeFunktion 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 auch 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 20 bis 23 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.

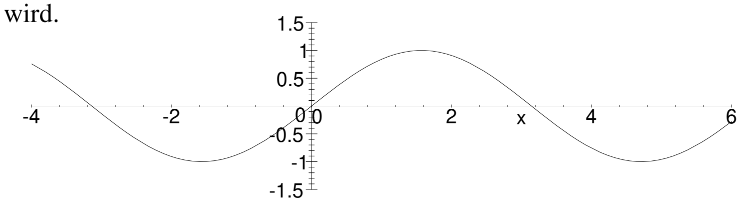


Abb. 20: Graph der Funktion $y = \sin x$

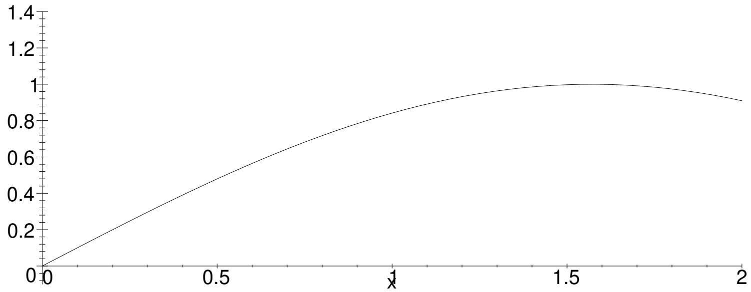


Abb. 21: Vergrößerung um den Faktor fünf um $\big(1, \sin(1)\big)$

Im Zweidimensionalen können wir ganz entsprechend vorgehen: Wenn wir etwa die Funktion $f(x, y) = \sin x \cos y$ auf $\mathbb{R}^2$ in der Umgebung des Punktes $(1, 1)$ betrachten und diese wieder sukzessive um den Faktor fünf vergrößern, erhalten wir die Abbildungen 24 bis 27, die zeigen, daß sich der Graph der Funktion in einer hinreichend kleinen Umgebung des

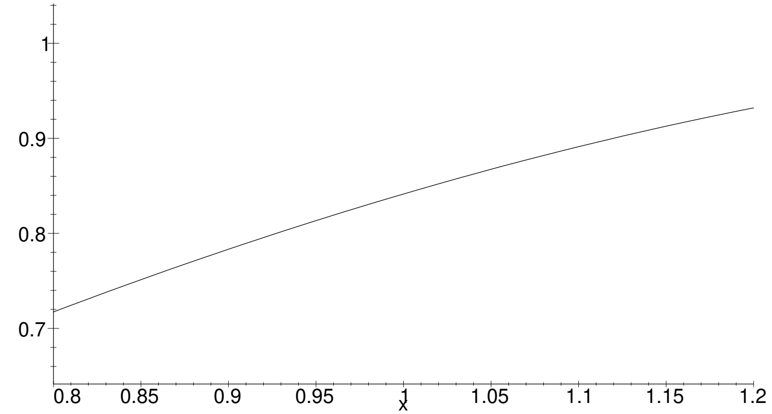


Abb. 22: Nochmalige Vergrößerung um den Faktor fünf um $\big(1, \sin(1)\big)$

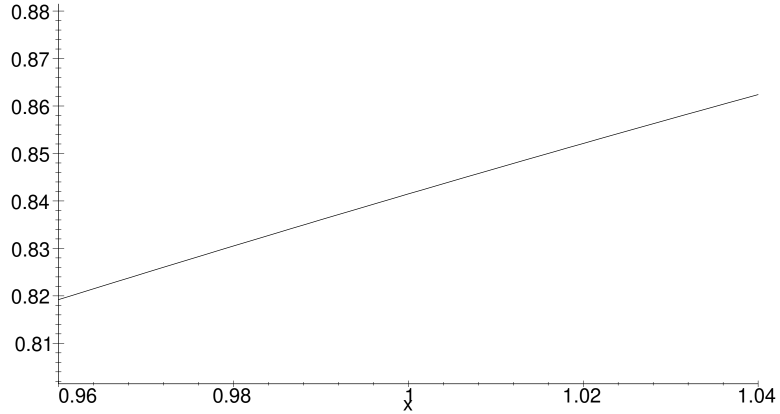


Abb. 23: Nach noch einer Vergrößerung sieht die Funktion praktisch linear aus

betrachteten Punktes nur wenig von einer Ebenen unterscheidet, d.h. die Funktion ist dort annähernd linear.

Eine lineare Funktion zweier Veränderlicher hat die Form

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

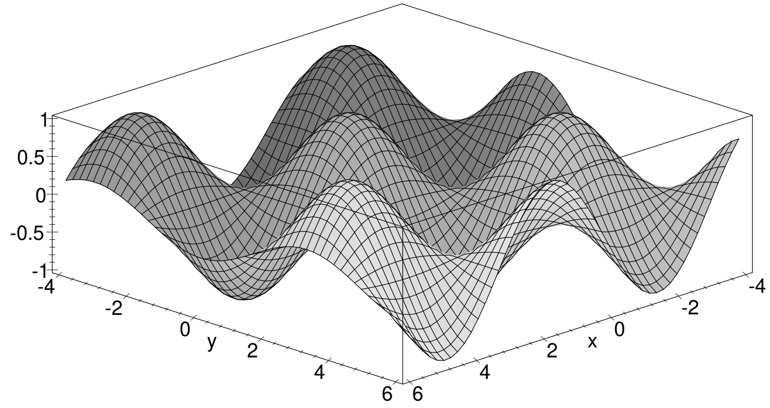


Abb. 24: Graph der Funktion $z = \sin x \cos y$

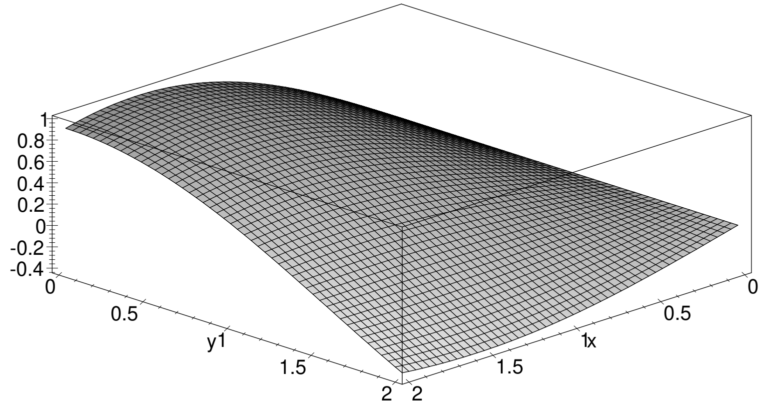


Abb. 25: Vergrößerung um den Faktor fünf um $\big(1, 1, \sin(1)\cos(1)\big)$

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

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

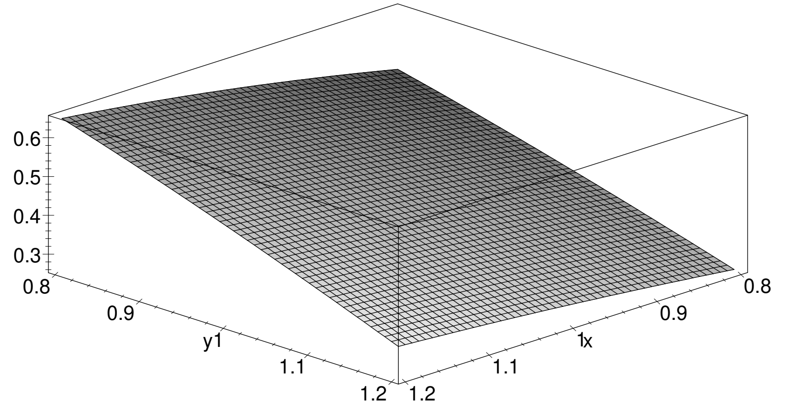


Abb. 26: Nochmalige Vergrößerung um den Faktor fünf um $\big(1, 1, \sin(1)\cos(1)\big)$

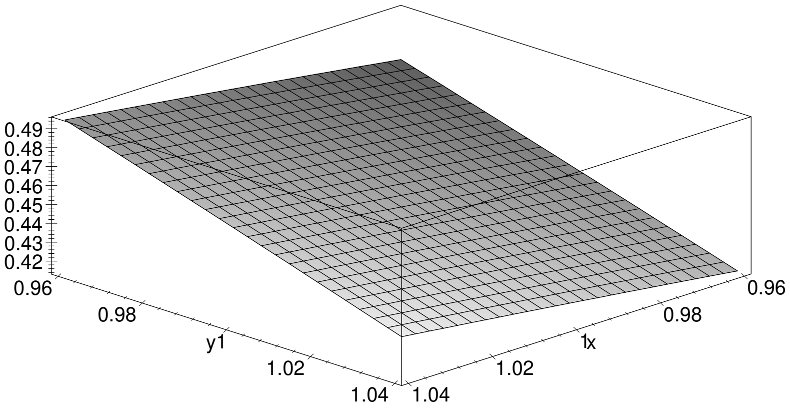


Abb. 27: Nach noch einer Vergrößerung sieht die Funktion praktisch linear aus

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 erwarten also, daß

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

ist und bekanntlich ist Differenzierbarkeit im $\mathbb{R}^n$ genau so definiert:

Definition: *a)* Die Funktion $f\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ heißt *differenzierbar* im Punkt $\boldsymbol{x} \in D$, wenn es eine lineare Funktion $L\colon \mathbb{R}^n \to \mathbb{R}^n$ gibt, so daß für $\vec{h} \to 0$ gilt

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

b) Ist für $\vec{h} = {}^t(h_1, \ldots, h_n)$

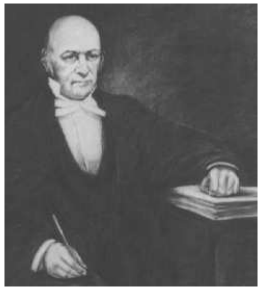
$$L\big(\vec{h}\big) = \gamma_1 h_1 + \cdots + \gamma_n h_n\ ,$$

so bezeichnen wir den Vektor

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

als *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.

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 $f\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ Funktionen einer Veränderlichen zu machen. Trotz der schlechten Erfahrungen beim obigen Beispiel einer unstetigen Funktion bietet sich hier 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

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

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

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

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

die Ableitung jener Funktion, die nur von x_i abhängt, während alle anderen Koordinaten x_j festgehalten werden. 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

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

ist insbesondere auch

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

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

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

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 $o\left(\left|\vec{h}\right|\right)$ zusammenfassen, denn da $\left|h_i\right| \le \left|\vec{h}\right|$ 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

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

ist. Damit ist

$$\Big(f_{x_1}(x_1,x_2+h_2,\ldots,x_n+h_n) - f_{x_1}(x_1,x_2,\ldots,x_n)\Big)h_1 = o\left(\left|\vec{h}\right|\right),$$

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 $o\left(\left|\vec{h}\right|\right)$ abgeschätzt werden. Also ist

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

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

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

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 Paragraphen sind höhere Ableitungen von Funktionen mehrerer Veränderlicher. Wir lassen uns bei der Definition wieder