Kapitel 6 Differenzierbare Funktionen

§ 1 Topologische Strukturen

Inhalt:

Umgebungen, innere Punkte, offene Mengen, abgeschlossene Mengen, Häufungspunkte, offener Kern und abgeschlossene Hülle, Rand einer Menge.

Der Konvergenzbegriff im $\mathbb{R}^n$, kompakte Mengen und der Satz von Heine-Borel, Stetigkeit, Ungleichungen, stetiges Bild einer kompakten Menge, Satz vom globalen Maximum und Minimum, stetige Wege, Gebiete, Konvexität.

Zur Erinnerung: Der Abstand zweier Punkte $\mathbf{x}$ und $\mathbf{y}$ im $\mathbb{R}^n$ ist gegeben durch die Zahl

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

Man nennt die Funktion $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ die *euklidische Metrik* auf dem $\mathbb{R}^n$. Sie hat folgende Eigenschaften:

1. $d(x, y) \geq 0$ für alle $x, y \in X$.
2. $d(x, y) = 0 \iff x = y$.
3. $d(x, y) = d(y, x)$ für alle $x, y \in X$ (Symmetrie).
4. $d(x, y) \leq d(x, z) + d(z, y)$ für $x, y, z \in X$ (Dreiecks-Ungleichung).

Definition:

Sei $\mathbf{x}_0 \in \mathbb{R}^n$ und $\varepsilon > 0$ eine reelle Zahl. Dann heißt

$$U_\varepsilon(\mathbf{x}_0) := \{\mathbf{x} \in \mathbb{R}^n \,:\, d(\mathbf{x}, \mathbf{x}_0) < \varepsilon\}$$

die *ε-Umgebung* von $\mathbf{x}_0$.

In $\mathbb{R}^2$ ist $U_\varepsilon(\mathbf{x}_0)$ eine Kreisscheibe, im $\mathbb{R}^3$ eine Kugel. Wir schreiben auch $B_\varepsilon(\mathbf{x}_0)$ („B“ für „ball“). Der Rand gehört jeweils nicht dazu.

Eine beliebige Menge $M \subset \mathbb{R}^n$ heißt eine *Umgebung* von $\mathbf{x}_0$, falls es ein $\varepsilon > 0$ mit $U_\varepsilon(\mathbf{x}_0) \subset M$ gibt. Der Punkt $\mathbf{x}_0$ hat dann einen „Sicherheitsabstand“ zum Rand der Umgebung. M seinerseits kann aber beliebige Gestalt haben. Natürlich ist jede ε-Umgebung von $\mathbf{x}_0$ auch eine Umgebung von $\mathbf{x}_0$ im obigen Sinne.

Ein Punkt $\mathbf{x}_0 \in M$ heißt *innerer Punkt* von M, falls M noch eine ganze Umgebung von $\mathbf{x}_0$ enthält.

Hausdorffscher Trennungssatz

Sind $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ zwei Punkte mit $\mathbf{x} \neq \mathbf{y}$, so gibt es Umgebungen U von $\mathbf{x}$ und V von $\mathbf{y}$, so daß $U \cap V = \varnothing$ ist.

BEWEIS: Wegen $\mathbf{x} \neq \mathbf{y}$ ist $r := d(\mathbf{x}, \mathbf{y}) > 0$. Nun sei $0 < \varepsilon < r/2$, $U = U_\varepsilon(\mathbf{x})$ und $V = U_\varepsilon(\mathbf{y})$. Wäre $\mathbf{z}$ ein Punkt in $U \cap V$, so wäre $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) < 2\varepsilon < r$. Das wäre ein Widerspruch. ∎

Definition:

Eine Menge $M \subset \mathbb{R}^n$ heißt *offen*, falls es zu jedem $\mathbf{x} \in M$ ein $\varepsilon > 0$ gibt, so daß $U_\varepsilon(\mathbf{x}) \subset M$ ist.

Eine Menge M ist also genau dann offen, wenn sie eine Umgebung von jedem ihrer Punkte ist. Dann ist jeder Punkt von M ein innerer Punkt von M.

Behauptung: Jede ε-Umgebung ist eine offene Menge.

BEWEIS: Sei $\mathbf{y} \in U_\varepsilon(\mathbf{x}_0)$. Wir suchen eine δ-Umgebung von $\mathbf{y}$, die noch ganz in $U_\varepsilon(\mathbf{x}_0)$ enthalten ist. Dazu sei $r := d(\mathbf{y}, \mathbf{x}_0)$. Dann ist $0 \leq r < \varepsilon$. Man kann eine positive reelle Zahl $\delta < \varepsilon - r$ finden. Ist $\mathbf{x} \in U_\delta(\mathbf{y})$, also $d(\mathbf{x}, \mathbf{y}) < \delta$, so ist $d(\mathbf{x}, \mathbf{x}_0) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{x}_0) < \delta + r < (\varepsilon - r) + r = \varepsilon$. Das zeigt, daß $U_\delta(\mathbf{y}) \subset U_\varepsilon(\mathbf{x}_0)$ ist. ∎

Satz (Eigenschaften offener Mengen)

Die offenen Mengen im $\mathbb{R}^n$ besitzen folgende Eigenschaften:

1. *Der $\mathbb{R}^n$ und die leere Menge sind offen.*
2. *Der Durchschnitt endlich vieler offener Mengen ist wieder offen.*
3. *Die Vereinigung beliebig vieler offener Mengen ist wieder offen.*

BEWEIS: 1) Für den $\mathbb{R}^n$ und die leere Menge ist der Beweis trivial.

2) Seien $M_1, \ldots, M_n$ offen und $M := M_1 \cap \ldots \cap M_n$. Ist $\mathbf{x} \in M$, so gibt es Zahlen $\varepsilon_i > 0$ mit $U_{\varepsilon_i}(\mathbf{x}) \subset M_i$ für $i = 1, \ldots, n$. Setzt man $\varepsilon := \min(\varepsilon_1, \ldots, \varepsilon_n)$, so liegt $U_\varepsilon(\mathbf{x})$ in M.

3) Es sei $\mathcal{M} = \{M_\iota \,:\, \iota \in I\}$ eine Familie von offenen Mengen,

$$M = \bigcup_{\iota \in I} M_\iota = \{\mathbf{x} \in \mathbb{R}^n \,:\, \exists\, \iota \in I \text{ mit } \mathbf{x} \in M_\iota\}$$

deren Vereinigung, $\mathbf{x}$ ein Element von M. Ist $\mathbf{x} \in M_\iota$, so gibt es ein $\varepsilon > 0$, so daß $U_\varepsilon(\mathbf{x}) \subset M_\iota$ ist. Aber dann ist erst recht $U_\varepsilon(\mathbf{x}) \subset M$. ∎

Die Menge $\overset{\circ}{M}$ aller inneren Punkte von M nennt man auch den *offenen Kern* von M. Diese Menge ist immer offen. Eine Menge ist genau dann offen, wenn sie mit ihrem offenen Kern übereinstimmt.

Definition:

Eine Menge $M \subset \mathbb{R}^n$ heißt *abgeschlossen*, falls $\mathbb{R}^n \setminus M$ offen ist.

Satz (Eigenschaften abgeschlossener Mengen)

Die abgeschlossenen Mengen in einem metrischen Raum besitzen folgende Eigenschaften:

1. *Der $\mathbb{R}^n$ und die leere Menge sind abgeschlossen.*
2. *Die Vereinigung endlich vieler abgeschlossener Mengen ist wieder abgeschlossen.*
3. *Der Durchschnitt beliebig vieler abgeschlossener Mengen ist wieder abgeschlossen.*

Der Beweis ergibt sich unmittelbar aus den Eigenschaften der offenen Mengen durch Komplement-Bildung.

Definition:

Sei $M \subset \mathbb{R}^n$ eine beliebige Teilmenge. Ein Punkt $\mathbf{x}_0 \in \mathbb{R}^n$ heißt ein *Häufungspunkt* der Menge M, falls in jeder Umgebung von $\mathbf{x}_0$ ein Punkt $\mathbf{x} \neq \mathbf{x}_0$ liegt, der zu M gehört.

Ist $\mathbf{x}_0$ **nicht** Häufungspunkt von M, so gibt es eine Umgebung $U = U(\mathbf{x}_0)$, so daß $U \cap M = \{\mathbf{x}_0\}$ ist. In diesem Falle würde man $\mathbf{x}_0$ einen *isolierten Punkt* von M nennen.

Eine endliche Menge besitzt keine Häufungspunkte. Auch $\mathbb{Z}$ hat keinen Häufungspunkt in $\mathbb{R}$. Aber jede reelle Zahl ist ein Häufungspunkt der Teilmenge $\mathbb{Q} \subset \mathbb{R}$.

Satz

Eine Menge $M \subset \mathbb{R}^n$ ist genau dann abgeschlossen, wenn sie alle ihre Häufungspunkte enthält.

Beweis: 1) Sei M abgeschlossen und $\mathbf{x}_0$ ein Häufungspunkt von M. Würde $\mathbf{x}_0$ nicht zu M gehören, so wäre $\mathbf{x}_0$ ein Element der offenen Menge $\mathbb{R}^n \setminus M$. Dann gäbe es ein $\varepsilon > 0$, so daß auch noch $U := U_\varepsilon(\mathbf{x}_0)$ in $\mathbb{R}^n \setminus M$ enthalten ist. Das wäre ein Widerspruch.

2) Es sei M eine Menge, die alle ihre Häufungspunkte enthält. Wir betrachten einen beliebigen Punkt $\mathbf{x}_0 \in \mathbb{R}^n \setminus M$. Da $\mathbf{x}_0$ kein Häufungspunkt von M ist, gibt es eine Umgebung $V = V(\mathbf{x}_0) \subset \mathbb{R}^n$, die keinen Punkt von M enthält. Weil so etwas mit jedem Punkt $\mathbf{x}_0 \in \mathbb{R}^n \setminus M$ geht, ist $\mathbb{R}^n \setminus M$ offen und M selbst abgeschlossen. ∎

Definition:

Ist $M \subset \mathbb{R}^n$ eine beliebige Teilmenge und $H(M)$ die Menge aller Häufungspunkte von M, so nennt man $\overline{M} := M \cup H(M)$ die *abgeschlossene Hülle* oder den *Abschluß* von M.

Satz

Sei M eine beliebige Teilmenge eines metrischen Raumes X. Dann gilt:

1. *$\overline{M}$ ist abgeschlossen.*
2. *M ist genau dann abgeschlossen, wenn $M = \overline{M}$ ist.*

Beweis: 1) Da $\mathbb{R}^n \setminus \overline{M}$ offen ist, ist $\overline{M}$ abgeschlossen.

2) Es ist $M \subset \overline{M}$. Ist M abgeschlossen, so ist $H(M) \subset M$, also sogar $M = \overline{M}$. Ist umgekehrt diese Gleichheit gegeben, so ist M abgeschlossen, nach (1). ∎

Es ist z.B. $\overline{(a,b)} = [a,b]$, und im $\mathbb{R}^n$ ist $\overline{U_\varepsilon(x_0)} = \{x \in X \, : \, d(x, x_0) \le \varepsilon\}$.

Definition:

Ist $M \subset \mathbb{R}^n$ eine beliebige Menge, so nennt man

$$\partial M := \overline{M} \setminus \mathring{M}$$

den *Rand* von M.

Ein Punkt $\mathbf{x}_0 \in \mathbb{R}^n$ gehört genau dann zum Rand von M, wenn $\mathbf{x}_0$ ein Häufungspunkt, aber kein innerer Punkt von M ist. Dann enthält jede Umgebung von $\mathbf{x}_0$ sowohl Punkte von M als auch Punkte von $\mathbb{R}^n \setminus M$.

Definition:

Eine Folge $(\mathbf{x}_\nu)$ von Punkten im $\mathbb{R}^n$ *konvergiert* gegen einen Punkt $\mathbf{x}_0$, falls folgende Bedingung erfüllt ist:

$$\forall\, \varepsilon > 0 \; \exists\, \nu_0, \text{ so daß } \forall\, \nu \ge \nu_0 \text{ gilt: } \|\mathbf{x}_\nu - \mathbf{x}_0\| < \varepsilon.$$

Man schreibt dann: $\lim\limits_{\nu \to \infty} \mathbf{x}_\nu = \mathbf{x}_0$.

Man kann auch sagen: $(\mathbf{x}_\nu)$ konvergiert im $\mathbb{R}^n$ gegen $\mathbf{x}_0$, falls $d(\mathbf{x}_\nu, \mathbf{x}_0)$ in $\mathbb{R}$ gegen 0 konvergiert.

In $\mathbb{R}$ ergibt das den bereits bekannten Konvergenzbegriff. Der Grenzwert ist eindeutig bestimmt.

Ist $\mathbf{x}_\nu = (x_1^{(\nu)}, \ldots, x_n^{(\nu)})$ eine Punktfolge und $\mathbf{x}_0 = (x_1^{(0)}, \ldots, x_n^{(0)})$ ein fester Punkt, so ist

$$\|\mathbf{x}_\nu - \mathbf{x}_0\| = \sqrt{(x_1^{(\nu)} - x_1^{(0)})^2 + \cdots + (x_n^{(\nu)} - x_n^{(0)})^2}.$$

Die Folge $(\mathbf{x}_\nu)$ konvergiert also genau dann, wenn alle Komponentenfolgen $(x_i^{(\nu)})$ konvergieren.

Satz (Charakterisierung abgeschlossener Mengen)

Eine Menge M ist genau dann abgeschlossen, wenn gilt: Ist $(\mathbf{x}_\nu)$ eine Folge in M, die im $\mathbb{R}^n$ konvergiert, so liegt der Grenzwert ebenfalls in M.

BEWEIS: 1) Sei M abgeschlossen, $(\mathbf{x}_\nu)$ eine Folge in M und $\mathbf{x}_0 = \lim_{\nu\to\infty} \mathbf{x}_\nu$. Ist die Menge der Folgeglieder endlich, so muß $\mathbf{x}_0$ eines dieser Folgeglieder sein und daher in M liegen. Ist sie unendlich, so ist $\mathbf{x}_0$ ein Häufungspunkt von M und es folgt ebenfalls, daß $\mathbf{x}_0$ in M liegt.

2) M erfülle das Kriterium und $\mathbf{x}_0$ sei ein Häufungspunkt von M. Dann liegt in jeder $(1/\nu)$-Umgebung von $\mathbf{x}_0$ ein Punkt $\mathbf{x}_\nu \in M$. Offensichtlich konvergiert $(\mathbf{x}_\nu)$ gegen $\mathbf{x}_0$. Also liegt $\mathbf{x}_0$ schon in M. Damit ist M abgeschlossen. ∎

Eine Menge $M \subset \mathbb{R}^n$ heißt beschränkt, falls es ein $R > 0$ gibt, so daß M in der Kugel $B_R(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{0}) < R\}$ enthalten ist. Eine Folge im $\mathbb{R}^n$ heißt beschränkt, wenn die Menge der Folgeglieder beschränkt ist. Es gilt folgende Verallgemeinerung des Satzes von Bolzano-Weierstraß:

Satz (Bolzano-Weierstraß)

Sei $\mathbf{x}_\nu = (x_1^{(\nu)}, \ldots, x_n^{(\nu)})$ eine beschränkte Folge im $\mathbb{R}^n$. Dann besitzt $(\mathbf{x}_\nu)$ eine konvergente Teilfolge.

BEWEIS: Es gibt ein $R > 0$, so daß alle $\mathbf{x}_\nu$ in $B_R(0)$ liegen. Aber dann liegen sie erst recht in $I^n = I \times \ldots \times I$, mit $I := [-R, R]$.

$(x_1^{(\nu)})$ besitzt eine konvergente Teilfolge $(x_1^{(\nu(i_1))})$ mit einem Grenzwert $x_1^{(0)} \in I$.

$(x_2^{(\nu(i_1))})$ besitzt eine konvergente Teilfolge $(x_2^{(\nu(i_2))})$ mit einem Grenzwert $x_2^{(0)} \in I$, usw.

Schließlich erhält man eine konvergente Teilfolge $(x_{\nu(i_n)})$ von (x_ν). ∎

Definition:

Eine Menge $K \subset \mathbb{R}^n$ heißt *kompakt*, falls jede Punktfolge in K eine konvergente Teilfolge besitzt, deren Grenzwert ebenfalls in K liegt.

Satz von Heine-Borel

Eine Teilmenge K des $\mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.

BEWEIS: 1) Sei K kompakt. Ist K nicht beschränkt, so gibt es eine Punktfolge $(\mathbf{x}_\nu)$ in K mit $\|\mathbf{x}_\nu\| > \nu$. Dann ist auch jede Teilfolge von $(\mathbf{x}_\nu)$ unbeschränkt. Das ist ein Widerspruch.

Sei nun $\mathbf{x}_0$ ein Häufungspunkt von K. Dann gibt es für jedes ν einen Punkt $\mathbf{x}_\nu \in K \cap B_{1/\nu}(\mathbf{x}_0)$. Die Folge $(\mathbf{x}_\nu)$ konvergiert gegen $\mathbf{x}_0$, und nach Voraussetzung konvergiert eine Teilfolge gegen ein Element von K. Das muß dann aber $\mathbf{x}_0$ sein. Also ist K abgeschlossen.

2) Sei jetzt K als abgeschlossen und beschränkt vorausgesetzt. Eine Punktfolge in K ist dann ebenfalls beschränkt, und nach Bolzano-Weierstraß gibt es eine Teilfolge, die gegen ein $\mathbf{x}_0 \in \mathbb{R}^n$ konvergiert. Aber weil K abgeschlossen ist, liegt $\mathbf{x}_0$ in K. ■

Beispiele.

1. In $\mathbb{R}$ ist jedes abgeschlossene Intervall kompakt. Im $\mathbb{R}^n$ ist jede *abgeschlossene Kugel*
$$\overline{B_r(\mathbf{x}_0)} = \{\mathbf{x} \in \mathbb{R}^n \, : \, \|\mathbf{x} - \mathbf{x}_0\| \le r\}$$
kompakt.

2. Jede endliche Teilmenge des $\mathbb{R}^n$ ist kompakt.

3. Sei $(\mathbf{x}_\nu)$ eine konvergente Punktfolge im $\mathbb{R}^n$, mit Grenzwert $\mathbf{x}_0$. Dann ist $M := \{\mathbf{x}_0\} \cup \{\mathbf{x}_\nu \, : \, \nu \in \mathbb{N}\}$ kompakt. Man sieht das so: Jede Folge in M ist eine Teilfolge von $(\mathbf{x}_\nu)$, oder die Folgeglieder nehmen nur endlich viele Werte an. In beiden Fällen gibt es eine Teilfolge, die in M konvergiert.

4. Ist $X \subset \mathbb{R}^n$ kompakt und $M \subset X$ eine abgeschlossene Teilmenge, so ist auch M kompakt. Der Beweis ist trivial.

Definition:

Sei M eine Teilmenge des $\mathbb{R}^m$ und $f : M \to \mathbb{R}^n$ eine Abbildung. f heißt *stetig* in $\mathbf{x}_0 \in M$, falls gilt:

$$\forall \varepsilon > 0 \; \exists \delta > 0 \text{ s.d. } \forall \mathbf{x} \in M \text{ mit } \|\mathbf{x} - \mathbf{x}_0\| < \delta \text{ gilt: } \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon.$$

f heißt stetig auf M, falls f in jedem Punkt von M stetig ist.

Anschaulich bedeutet dies: Zu jeder noch so kleinen Fehlerschranke ε kann man eine davon abhängige Schranke δ finden, so daß gilt: Ist eine Approximation $\mathbf{x}$ von $\mathbf{x}_0$ gegeben und der Fehler $< \delta$, so ist der Bildpunkt $f(\mathbf{x})$ um weniger als ε von $f(\mathbf{x}_0)$ entfernt.

Satz (Gleichwertige Beschreibungen der Stetigkeit)

Folgende Aussagen über $f : M \to \mathbb{R}^m$ und $\mathbf{x}_0 \in M$ sind äquivalent:

1. *f ist stetig in $\mathbf{x}_0$.*
2. *Zu jeder Umgebung $V = V(f(\mathbf{x}_0)) \subset \mathbb{R}^n$ gibt es eine Umgebung $U = U(\mathbf{x}_0) \subset \mathbb{R}^m$ mit $f(U \cap M) \subset V$.*
3. *Für jede Folge $(\mathbf{x}_\nu)$ in M mit $\lim\limits_{\nu\to\infty} \mathbf{x}_\nu = \mathbf{x}_0$ gilt auch $\lim\limits_{\nu\to\infty} f(\mathbf{x}_\nu) = f(\mathbf{x}_0)$.*

BEWEIS: (1) $\Longrightarrow$ (2):

Ist V eine Umgebung von $f(\mathbf{x}_0)$, so enthält V eine ε-Umgebung von $f(\mathbf{x}_0)$. Nach Definition der Stetigkeit gibt es ein $\delta > 0$ mit $f(U_\delta(\mathbf{x}_0) \cap M) \subset U_\varepsilon(f(\mathbf{x}_0))$. Wir setzen $U := U_\delta(\mathbf{x}_0)$.

(2) $\Longrightarrow$ (3):

Sei $(\mathbf{x}_\nu)$ eine Folge in M, die gegen $\mathbf{x}_0$ konvergiert. Außerdem sei ein $\varepsilon > 0$ vorgegeben. Es gibt eine Umgebung $U = U(\mathbf{x}_0)$ mit $f(U \cap M) \subset U_\varepsilon(f(\mathbf{x}_0))$. Für ein geeignetes ν_0 liegen alle Folgeglieder $\mathbf{x}_\nu$ mit $\nu \geq \nu_0$ in U. Dann ist $\|f(\mathbf{x}_\nu) - f(\mathbf{x}_0)\| < \varepsilon$ für $\nu \geq \nu_0$. Das bedeutet, daß $(f(\mathbf{x}_\nu))$ gegen $f(\mathbf{x}_0)$ konvergiert.

(3) $\Longrightarrow$ (1):

Es sei das Folgenkriterium erfüllt. Wir nehmen an, f sei nicht stetig in $\mathbf{x}_0$. Dann gibt es ein $\varepsilon > 0$, so daß zu jedem $\nu \in \mathbb{N}$ ein $\mathbf{x}_\nu$ mit $\|\mathbf{x}_\nu - \mathbf{x}_0\| < 1/\nu$ und $\|f(\mathbf{x}_\nu) - f(\mathbf{x}_0)\| \geq \varepsilon$ existiert. Aber das kann nicht sein. ∎

Satz

Es seien $M \subset \mathbb{R}^m$ und $N \subset \mathbb{R}^n$ Teilmengen, $f : M \to \mathbb{R}^n$ und $g : N \to \mathbb{R}^k$ Abbildungen mit $f(M) \subset N$. Ist f stetig in $\mathbf{x}_0 \in M$ und g stetig in $\mathbf{y}_0 := f(\mathbf{x}_0) \in N$, so ist auch $g \circ f : M \to \mathbb{R}^k$ stetig in $\mathbf{x}_0$.

BEWEIS: Sei $\mathbf{z}_0 := g(\mathbf{y}_0) = (g \circ f)(\mathbf{x}_0)$ und $W = W(\mathbf{z}_0) \subset \mathbb{R}^k$ eine Umgebung. Dann gibt es eine Umgebung $V = V(\mathbf{y}_0) \subset \mathbb{R}^n$ mit $g(V \cap N) \subset W$, sowie eine Umgebung $U = U(\mathbf{x}_0) \subset \mathbb{R}^n$ mit $f(U \cap M) \subset V$. Es folgt, daß $(g \circ f)(U \cap M) \subset W$ ist, also $g \circ f$ stetig in $\mathbf{x}_0$. ∎

Beispiele.

1. Jede konstante Abbildung $k : \mathbb{R}^m \to \mathbb{R}^n$ ist stetig, denn die Bildmenge besteht nur aus einem einzigen Punkt.

2. Ist $M \subset \mathbb{R}^n$ eine beliebige Teilmenge, so ist die identische Abbildung $\mathrm{id}_M : M \to \mathbb{R}^n$ stetig, denn für jede offene Teilmenge $U \subset \mathbb{R}^n$ ist $\mathrm{id}_M(U \cap M) \subset U$.

3. Sei $f : \mathbb{R}^m \to \mathbb{R}^n$ linear. Dann ist f bereits durch die Werte $f(\mathbf{e}_i)$, $i = 1, \dots, m$, festgelegt. Wir setzen
$$C := \sum_{i=1}^{m} \|f(\mathbf{e}_i)\|.$$
Dann erhalten wir für $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_m\mathbf{e}_m$ die Abschätzung
$$\begin{aligned} \|f(\mathbf{x})\| &= \|\sum_{i=1}^{m} x_i \cdot f(\mathbf{e}_i)\| \\ &\leq \sum_{i=1}^{m} |x_i| \cdot \|f(\mathbf{e}_i)\| \\ &\leq C \cdot \max_i |x_i| \\ &\leq C \cdot \|\mathbf{x}\|, \end{aligned}$$
denn es ist $\max_i|x_i| = \sqrt{(\max_i|x_i|)^2} \leq \sqrt{(x_1)^2 + \cdots + (x_m)^2}$.

   Aus der gewonnenen Ungleichung leitet man sofort ab, daß f im Nullpunkt stetig ist: Ist $\varepsilon > 0$ gegeben, so wählen wir $\delta := \varepsilon / C$. Für $\|\mathbf{x}\| < \delta$ ist $\|f(\mathbf{x})\| \leq C \cdot \|\mathbf{x}\| < \varepsilon$.

   Ist $\mathbf{x}_0 \in \mathbb{R}^m$ ein beliebiger Punkt, so ist
$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| = \|f(\mathbf{x} - \mathbf{x}_0)\| \leq C \cdot \|\mathbf{x} - \mathbf{x}_0\|.$$
Jetzt folgt wie oben, daß f auch in $\mathbf{x}_0$ (und damit überall) stetig ist.

4. Es sei $M \subset \mathbb{R}^m$ eine beliebige Teilmenge. Sind $f, g : M \to \mathbb{R}^n$ stetige Abbildungen, so sind auch die Abbildungen $f + g : M \to \mathbb{R}^n$ und $f \bullet g : M \to \mathbb{R}$ (mit $(f \bullet g)(\mathbf{x}) := f(\mathbf{x}) \bullet g(\mathbf{x})$) stetig. Auf den BEWEIS verzichten wir hier.

5. Eine Abbildung $f = (f_1, \dots, f_n) : M \to \mathbb{R}^n$ ist genau dann stetig, wenn alle Komponenten-Funktionen $f_i : M \to \mathbb{R}$ stetig sind. Eine komplexe Funktion f ist deshalb genau dann stetig, wenn Realteil und Imaginärteil stetig sind, und dann folgt, daß auch $\overline{f}$ stetig ist.

Satz

Sei $B \subset \mathbb{R}^n$ eine offene Teilmenge und $f : B \to \mathbb{R}$ eine stetige Funktion. Dann ist auch die Menge $M := \{\mathbf{x} \in B \,:\, f(\mathbf{x}) > 0\}$ offen.

BEWEIS: Sei $\mathbf{x}_0 \in M$, also $r_0 := f(\mathbf{x}_0) > 0$. Ist $0 < \varepsilon < r_0$, so gibt es ein $\delta > 0$, so daß $U_\delta(\mathbf{x}_0) \subset B$ und $|f(\mathbf{x}) - f(\mathbf{x}_0)| < \varepsilon$ für $\mathbf{x} \in U_\delta(\mathbf{x}_0)$ ist. Für jedes $\mathbf{x} \in U_\delta(\mathbf{x}_0)$ ist dann $0 < r_0 - \varepsilon = f(\mathbf{x}_0) - \varepsilon < f(\mathbf{x})$, also $\mathbf{x} \in M$. Also ist M offen. ∎

Folgerung

Sei $B \subset \mathbb{R}^n$ offen. Sind $f, g : B \to \mathbb{R}$ stetig, so gilt:

1. *$\{\mathbf{x} \in B \,:\, f(\mathbf{x}) < g(\mathbf{x})\}$ ist offen.*
2. *$\{\mathbf{x} \in B \,:\, f(\mathbf{x}) \neq g(\mathbf{x})\}$ ist offen.*

BEWEIS: 1) $\{f < g\} = \{g - f > 0\}$ ist offen, wegen des Satzes.

2) Da auch $\{f > g\} = \{g < f\}$ offen ist, muß $\{f \neq g\} = \{f < g\} \cup \{f > g\}$ offen sein. ∎

Satz (über das stetige Bild einer kompakten Menge)

Sei $K \subset \mathbb{R}^m$ kompakt und $f : K \to \mathbb{R}^n$ eine stetige Abbildung. Dann ist auch $f(K)$ kompakt.

BEWEIS: Sei $(\mathbf{y}_\nu)$ eine Folge von Punkten in $f(K)$. Dann gibt es zu jedem ν einen Punkt $\mathbf{x}_\nu \in K$ mit $f(\mathbf{x}_\nu) = \mathbf{y}_\nu$. Weil K kompakt ist, besitzt die Folge $(\mathbf{x}_\nu)$ eine in K konvergente Teilfolge $(\mathbf{x}_{\nu_i})$, ihr Grenzwert in K sei mit $\mathbf{x}_0$ bezeichnet. Wegen der Stetigkeit von f konvergiert $(\mathbf{y}_{\nu_i})$ gegen $\mathbf{y}_0 := f(\mathbf{x}_0)$, und dieser Punkt liegt in $f(K)$. ∎

Satz (vom globalen Minimum und Maximum)

Auf einer kompakten Teilmenge $K \subset \mathbb{R}^n$ nimmt jede stetige Funktion ihr Maximum und ihr Minimum an.

Beweis: $f(K) \subset \mathbb{R}$ ist kompakt, also abgeschlossen und beschränkt. Demnach existieren $y_- := \inf f(K)$ und $y_+ := \sup f(K)$, und sie sind in $f(K)$ enthalten. Also gibt es Punkte $\mathbf{x}_-$ und $\mathbf{x}_+$ in K mit $f(\mathbf{x}_-) = y_-$ und $f(\mathbf{x}_+) = y_+$. ∎

Speziell nimmt also eine stetige Funktion $f : [a, b] \to \mathbb{R}$ immer Maximum und Minimum an und ist demnach beschränkt.

Zur Erinnerung: Ein stetiger (parametrisierter) Weg im $\mathbb{R}^n$ ist eine stetige Abbildung $\alpha : I \to \mathbb{R}^n$, wobei I ein endliches oder unendliches Intervall ist.

Beispiele.

1. Sind $\mathbf{x}_0, \mathbf{y}_0$ zwei Punkte im $\mathbb{R}^n$, so wird die *Verbindungsstrecke* von $\mathbf{x}_0$ und $\mathbf{y}_0$ durch
$$\alpha(t) := \mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0) = (1-t)\mathbf{x}_0 + t\mathbf{y}_0$$
parametrisiert, $0 \le t \le 1$. Wir verstehen unter der Verbindungsstrecke aber auch die Bildmenge
$$S(\mathbf{x}_0, \mathbf{y}_0) := \alpha([0,1]) = \{\mathbf{x} = (1-t)\mathbf{x}_0 + t\mathbf{y}_0 \,:\, 0 \le t \le 1\}.$$

2. Im $\mathbb{R}^2$ ist der *Kreis* um $\mathbf{a} = (a_1, a_2)$ mit Radius $r > 0$ gegeben durch
$$\alpha(t) = (a_1 + r\cos(t), a_2 + r\sin(t)), \quad 0 \le t \le 2\pi.$$

Definition:

Eine offene Menge $G \subset \mathbb{R}^n$ heißt *zusammenhängend* oder *ein Gebiet*, falls gilt:

Zu je zwei beliebigen Punkten $\mathbf{x}, \mathbf{y} \in G$ gibt es einen stetigen Weg $\alpha : [0,1] \to G$ mit $\alpha(0) = \mathbf{x}$ und $\alpha(1) = \mathbf{y}$.

Ein Gebiet kann nicht in zwei offene Mengen zerlegt werden.

Satz (von der Unzerlegbarkeit von Gebieten)

Sei $G \subset \mathbb{R}^n$ ein Gebiet und $B \subset G$ eine offene nicht-leere Teilmenge. Ist auch $G \setminus B$ offen, so muß $B = G$ sein.

Beweis: Sei $\mathbf{x}_0 \in B$ und $\mathbf{y}_0$ ein beliebiger Punkt von G. Weil G ein Gebiet ist, gibt es einen stetigen Weg $\alpha : [0,1] \to G$ mit $\alpha(0) = \mathbf{x}_0$ und $\alpha(1) = \mathbf{y}_0$. Für kleines t liegt $\alpha(t)$ noch in der offenen Menge B.

Sei $t_0 := \sup\{t \in [0,1] \,:\, \alpha(s) \in B \text{ für } 0 \le s \le t\}$. Wir wollen zeigen, daß $t_0 = 1$ und damit $\mathbf{y}_0 \in B$ ist. Also nehmen wir an, es sei $t_0 < 1$. Wegen der Offenheit von

B kann $\alpha(t_0)$ nicht in B liegen. Wegen der Offenheit von $G \setminus B$ kann es aber auch nicht in $G \setminus B$ liegen. Das ist ein Widerspruch, die Annahme ist falsch. ∎

Definition:

Eine Teilmenge $M \subset \mathbb{R}^n$ heißt *konvex*, falls mit je zwei Punkten von M auch deren Verbindungsstrecke in M enthalten ist.

Beispiele.

1. Jedes Intervall ist eine konvexe Teilmenge von $\mathbb{R}$.
2. Offene und abgeschlossene Kugeln im $\mathbb{R}^n$ sind konvex.
3. Jede offene konvexe Menge ist ein Gebiet. Umgekehrt braucht ein Gebiet nicht unbedingt konvex zu sein. So ist z.B. das Gebiet

$$G = \{(x, y) \in \mathbb{R}^2 \,:\, -1 < x < 1 \text{ und } 1 - x^2 < y < 2\}$$

nicht konvex.

§ 2 Partielle Differenzierbarkeit

Inhalt:

Richtungsableitungen, partielle Ableitungen, höhere partielle Ableitungen, der Satz von Schwarz, Vektorfelder, der Nabla-Operator, spezielle Kettenregel, Eigenschaften des Gradienten.

Sei nun $G \subset \mathbb{R}^n$ ein Gebiet und $f : G \to \mathbb{R}$ eine Funktion. Wie kann man sich eine solche Funktion veranschaulichen? Ist $n = 2$, so ist der Graph

$$G_f := \{(x_1, x_2, z) \in G \times \mathbb{R} \mid z = f(x_1, x_2)\}$$

eine Fläche im $\mathbb{R}^3$. Jede „vertikale Gerade“ $\{(a, b, z) \mid z \in \mathbb{R}\}$ durch einen festen Punkt $(a, b) \in G$ trifft den Graphen in genau einem Punkt.

Eine andere Möglichkeit der Darstellung ist die Benutzung von „Höhenlinien“. In G liegen die *Niveaumengen*

$$N_c(f) := \{\mathbf{x} \in G \mid f(\mathbf{x}) = c\},$$

im Falle $n = 2$ sind das Linien. Man kennt diese Darstellung von den Landkarten her.

Ist allerdings $n > 2$, so ist eine anschauliche Darstellung von f durch den Graphen oder durch Niveaumengen kaum noch praktikabel.

Definition:

Sei $G \subset \mathbb{R}^n$ ein Gebiet, $\mathbf{a} \in G$ und $f : G \to \mathbb{R}$ eine Funktion. Für $\mathbf{v} \in \mathbb{R}^n$ bezeichnet man

$$D_{\mathbf{v}} f(\mathbf{a}) := \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}$$

als *Richtungsableitung von* f *in* $\mathbf{a}$ *in Richtung* $\mathbf{v}$ (sofern der Grenzwert existiert).

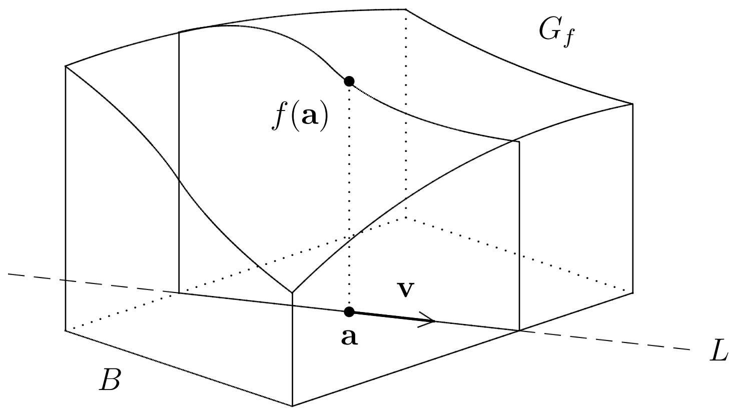
Was bedeutet das anschaulich?

Durch $\alpha(t) := \mathbf{a} + t\mathbf{v}$ wird eine Gerade $L \subset \mathbb{R}^n$ durch den Punkt $\mathbf{a}$ mit Richtungsvektor $\mathbf{v}$ parametrisiert. Die Funktion

$$f_L(t) := f \circ \alpha(t) = f(\mathbf{a} + t\mathbf{v})$$

ist eine gewöhnliche Funktion einer Veränderlichen, und die Richtungsableitung von f in $\mathbf{a}$ mit Richtung $\mathbf{v}$ ist nichts anderes als die gewöhnliche Ableitung $(f_L)'(0)$.

Den Graphen von f_L erhält man, indem man den Graphen von f mit der über der Geraden L gelegenen „senkrechten“ Ebene $\{(\mathbf{x}, z) \in \mathbb{R}^n \times \mathbb{R} \mid \mathbf{x} \in L\}$ schneidet.



Beispiel.

Sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x, y) := 1 - x^2 - y^2$, vektoriell geschrieben also

$$f(\mathbf{x}) = 1 - \mathbf{x} \bullet \mathbf{x}.$$

Ist $\mathbf{a} = (a_1, b_1)$ und $\mathbf{v} = (v_1, v_2)$, so ist

$$\begin{aligned} f_L(t) &= f(\mathbf{a} + t\mathbf{v}) = 1 - (\mathbf{a} + t\mathbf{v}) \bullet (\mathbf{a} + t\mathbf{v}) \\ &= 1 - \mathbf{a} \bullet \mathbf{a} - 2t\mathbf{v} \bullet \mathbf{a} - t^2 \mathbf{v} \bullet \mathbf{v}, \end{aligned}$$

also

$$\begin{aligned} D_{\mathbf{v}} f(\mathbf{a}) &= (f_L)'(0) \\ &= -2\mathbf{v} \bullet \mathbf{a}. \end{aligned}$$

Ist $\mathbf{a} \neq \mathbf{0}$, so verschwindet die Richtungsableitung $D_{\mathbf{v}} f(\mathbf{a}) = 0$ genau dann, wenn der Richtungsvektor $\mathbf{v}$ auf dem Ortsvektor $\mathbf{a}$ senkrecht steht. In $\mathbf{a} = \mathbf{0}$ verschwindet **jede** Richtungsableitung.

Eigenschaften der Richtungsableitung

f und g seien in $\mathbf{a}$ in Richtung $\mathbf{v}$ differenzierbar, c sei eine Konstante. Dann sind auch $c \cdot f$, $f + g$ und $f \cdot g$ in $\mathbf{a}$ in Richtung $\mathbf{v}$ differenzierbar, und es gilt:

1. $D_{\mathbf{v}}(c \cdot f)(\mathbf{a}) = c \cdot D_{\mathbf{v}} f(\mathbf{a})$.
2. $D_{\mathbf{v}}(f + g)(\mathbf{a}) = D_{\mathbf{v}} f(\mathbf{a}) + D_{\mathbf{v}} g(\mathbf{a})$.
3. $D_{\mathbf{v}}(f \cdot g)(\mathbf{a}) = f(a) \cdot D_{\mathbf{v}} g(\mathbf{a}) + D_{\mathbf{v}} f(\mathbf{a}) \cdot g(a)$.

Die Beweise funktionieren wie bei den Funktionen von einer Veränderlichen.

Eine besondere Rolle spielen die Richtungsableitungen in Richtung der Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$:

Definition:

Die Funktion f sei in $\mathbf{a}$ in Richtung des i–ten Einheits–Vektors $\mathbf{e}_i$ differenzierbar. Dann heißt

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) := D_{\mathbf{e}_i} f(\mathbf{a})$$

die *i–te partielle Ableitung von f in* $\mathbf{a}$. Man schreibt auch $f_{x_i}(\mathbf{a})$ dafür.

Wenn alle partiellen Ableitungen von f in $\mathbf{a}$ existieren, dann heißt f in $\mathbf{a}$ *partiell differenzierbar*.

Wie führt man die partielle Differentiation praktisch durch?

Sei $\mathbf{a} = (a_1, \ldots, a_n)$. Dann gilt:

$$\begin{aligned}
\frac{\partial f}{\partial x_i}(\mathbf{a}) &= D_{\mathbf{e}_i} f(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{e}_i) - f(\mathbf{a})}{t} \\
&= \lim_{t \to 0} \frac{1}{t}(f(a_1, \ldots, a_i + t, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)) \\
&= \lim_{s \to a_i} \frac{f(a_1, \ldots, a_{i-1}, s, a_{i+1}, \ldots, a_n) - f(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n)}{s - a_i} \\
&= \frac{d}{ds}\Big|_{s=a_i} f(a_1, \ldots, a_{i-1}, s, a_{i+1}, \ldots, a_n).
\end{aligned}$$

Um also die i–te partielle Ableitung von f in $\mathbf{a}$ auszurechnen, muß man in $f(x_1, \ldots, x_n)$ die Variablen x_j, $j \neq i$, durch die Konstanten a_j (also die Komponenten von $\mathbf{a}$) ersetzen. Danach hängt die Funktion nur noch von der einen verbliebenen Variablen x_i ab und kann im gewöhnlichen Sinne nach dieser Variablen an der Stelle a_i differenziert werden.

Beispiel.

Sei $f(x, y, z) := x^2 \cdot \cos(yz)$.

Um partiell nach x zu differenzieren, muß man die Variablen y und z festhalten und nur die Funktion $x \mapsto x^2 \cdot \cos(yz)$ betrachten. Also ist

$$\frac{\partial f}{\partial x}(x, y, z) = 2x \cdot \cos(yz).$$

Um partiell nach y zu differenzieren, muß man die Variablen x und z festhalten und nur die Funktion $y \mapsto x^2 \cdot \cos(yz)$ betrachten. So erhält man

$$\frac{\partial f}{\partial y}(x, y, z) = x^2 \cdot (-\sin(yz) \cdot z) = -x^2 z \sin(yz)$$

und analog

$$\frac{\partial f}{\partial z}(x, y, z) = -x^2 y \sin(yz).$$

Es sieht so aus, als hätte man die Verallgemeinerung der Differenzierbarkeit auf mehrere Veränderliche gefunden. Aber leider ist die partielle Differenzierbarkeit eine zu schwache Eigenschaft. Sie hat noch nicht einmal die Stetigkeit der Funktion selbst zur Folge:

Beispiel.

Wir betrachten die Funktion

$$f(x, y) := \begin{cases} \dfrac{xy^2}{x^2 + y^4} & \text{für } (x, y) \neq (0, 0) \\ 0 & \text{für } (x, y) = (0, 0). \end{cases}$$

Die Funktionen $x \mapsto f(x, 0) \equiv 0$ und $y \mapsto f(0, y) \equiv 0$ sind sicherlich im Nullpunkt differenzierbar. Also ist f in $\mathbf{0} = (0, 0)$ partiell differenzierbar. Andererseits ist f dort nicht stetig:

Wenn man $\mathbf{y}_\nu := ((a_\nu)^2, a_\nu)$ setzt, mit einer Nullfolge (a_ν), so konvergiert diese Folge gegen $(0, 0)$, aber es ist

$$\lim_{\nu \to \infty} f(\mathbf{y}_\nu) = \lim_{\nu \to \infty} \frac{(a_\nu)^4}{2(a_\nu)^4} = \frac{1}{2}.$$

Das dürfte nicht passieren, wenn f im Nullpunkt stetig wäre.

Eine weitere Schwäche der partiellen Differenzierbarkeit tritt auf, wenn man höhere Ableitungen betrachtet:

Ist $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ in allen Punkten von B partiell differenzierbar, so bilden die partiellen Ableitungen $\dfrac{\partial f}{\partial x_i}(\mathbf{x})$ wieder reellwertige Funktionen auf B. Sind sie alle stetig, so nennt man f *stetig partiell differenzierbar*.

Definition:

Sei $B \subset \mathbb{R}^n$ offen, $\mathbf{a} \in B$ und $f : B \to \mathbb{R}$ überall partiell differenzierbar. Alle partiellen Ableitungen $\dfrac{\partial f}{\partial x_i}$ seien in $\mathbf{a}$ noch einmal partiell differenzierbar. Dann definiert man für $i, j = 1, \ldots, n$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) := \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(\mathbf{a}).$$

Man nennt diesen Ausdruck auch die *2–te partielle Ableitung von f nach x_i und x_j an der Stelle* $\mathbf{a}$, und schreibt dafür auch $f_{x_i x_j}(\mathbf{a})$.

Man beachte die Reihenfolge! Zuerst wird nach der Variablen differenziert, die am weitesten rechts steht!

Beispiel.

Sei $f(x_1, x_2) := e^{k \cdot x_1} \cdot \cos(x_2)$. Dann gilt:

$$\frac{\partial f}{\partial x_1}(\mathbf{x}) = k \cdot e^{k \cdot x_1} \cdot \cos(x_2) \quad \text{und} \quad \frac{\partial f}{\partial x_2}(\mathbf{x}) = -e^{k \cdot x_1} \cdot \sin(x_2),$$

sowie

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) = -ke^{ka_1} \sin(a_2).$$

Man kann sich nun fragen, ob man die 2-ten Ableitungen immer miteinander vertauschen kann, ob es also bei höheren partiellen Ableitungen nicht auf die Reihenfolge ankommt. Leider ist das nicht generell der Fall:

Beispiel.

$$\text{Sei} \quad f(x, y) := \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & \text{für } (x, y) \neq (0, 0), \\ 0 & \text{für } (x, y) = (0, 0). \end{cases}$$

Dann gilt für $(x, y) \neq (0, 0)$:

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= \frac{\partial}{\partial x}\left(\frac{x^3y-y^3x}{x^2+y^2}\right)\\ &= \frac{(3x^2y-y^3)(x^2+y^2)-(x^3y-y^3x)2x}{(x^2+y^2)^2}\\ &= \frac{x^4y+4x^2y^3-y^5}{(x^2+y^2)^2},\end{aligned}$$

also

$$\frac{\partial f}{\partial x}(0,y) = -y \qquad (\text{für } y \neq 0).$$

Weiter ist

$$\frac{\partial f}{\partial x}(0,0) = \lim_{x\to 0}\frac{f(x,0)-f(0,0)}{x} = 0.$$

Also ist sogar $\dfrac{\partial f}{\partial x}(0,y) \equiv -y$ für alle y und $\dfrac{\partial^2 f}{\partial y \partial x}(0,0) = -1$.

Entsprechend erhalten wir für $(x,y) \neq (0,0)$:

$$\begin{aligned}\frac{\partial f}{\partial y}(x,y) &= \frac{\partial}{\partial y}\left(\frac{x^3y-y^3x}{x^2+y^2}\right)\\ &= \frac{(x^3-3y^2x)(x^2+y^2)-(x^3y-y^3x)2y}{(x^2+y^2)^2}\\ &= \frac{x^5-4x^3y^2-xy^4}{(x^2+y^2)^2},\end{aligned}$$

also

$$\frac{\partial f}{\partial y}(x,0) \equiv x \text{ für } x \neq 0,$$

und

$$\frac{\partial f}{\partial y}(0,0) = \lim_{y\to 0}\frac{f(0,y)-f(0,0)}{y} = 0.$$

Somit ist $\dfrac{\partial^2 f}{\partial x \partial y}(0,0) = +1$.

Zum Glück gilt folgendes hinreichende Kriterium für die Gleichheit der gemischten zweiten Ableitungen:

Satz von Schwarz

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ auf ganz B nach allen Variablen partiell differenzierbar, $\mathbf{a} \in B$.

Wenn die gemischten zweiten Ableitungen $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$ und $\dfrac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x})$ auf einer Umgebung von $\mathbf{a}$ in B existieren und in $\mathbf{a}$ stetig sind, so ist

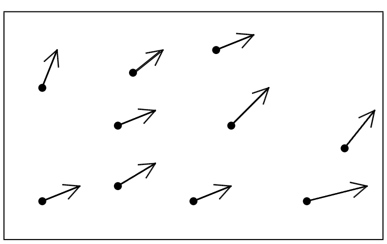
$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}).$$

Auf den etwas technischen Beweis verzichten wir hier.

Definition:

Sei $G \subset \mathbb{R}^n$ ein Gebiet. Ein *Vektorfeld* auf G ist eine Abbildung $F : G \to \mathbb{R}^n$, die jedem $\mathbf{x} \in G$ einen Vektor $F(\mathbf{x}) \in \mathbb{R}^n$ zuordnet.

Graphisch stellt man das Vektorfeld dar, indem man in jedem Punkt $\mathbf{x}$ den zugeordneten Vektorpfeil $F(\mathbf{x})$ zeichnet. Dadurch wird deutlich gemacht, daß es auf die gesamte Abbildung F ankommt, nicht nur auf die einzelnen Werte.



Manchmal versteht man deshalb unter einem Vektorfeld auf G auch die Menge aller Paare $(\mathbf{x}, F(\mathbf{x}))$ mit $\mathbf{x} \in G$.

Die Bildung der partiellen Ableitungen $\dfrac{\partial f}{\partial x_i}$ einer Funktion f kann man auch als Anwendung des „linearen Operators“ $D_i := \dfrac{\partial}{\partial x_i}$ auf die Funktion f auffassen.

Man faßt nun gerne die n Operatoren $D_1, \ldots, D_n$ zu einem vektoriellen Operator zusammen:

$$\nabla := \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right). \quad (\text{„Nabla“})$$

Dieser Operator kann auf verschiedene Weise wirken.

Sei $G \subset \mathbb{R}^n$ offen.

1. Ist $f : G \to \mathbb{R}$ eine stetig partiell differenzierbare Funktion, so heißt das Vektorfeld

$$\mathbf{grad}(f) := \nabla f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$$

das *Gradientenfeld* von f. Der Wert $\mathbf{grad}(f)(\mathbf{a})$ wird als *Gradient von f in* $\mathbf{a}$ bezeichnet.

2. Sei $\mathbf{v} = (v_1, \ldots, v_n) : G \to \mathbb{R}^n$ ein Vektorfeld, dessen sämtliche Komponenten v_i stetig partiell differenzierbar sind. Dann heißt die Funktion

$$\mathbf{div}(\mathbf{v}) := \nabla \bullet \mathbf{v} = \frac{\partial v_1}{\partial x_1} + \cdots + \frac{\partial v_n}{\partial x_n}$$

die *Divergenz* von $\mathbf{v}$.

3. Sei jetzt speziell $n = 3$ und $\mathbf{v} : G \to \mathbb{R}^3$ ein stetig partiell differenzierbares Vektorfeld. Dann heißt das Vektorfeld

$$\mathbf{rot}(\mathbf{v}) := \nabla \times \mathbf{v} = (\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}, \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}, \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2})$$

die *Rotation* von $\mathbf{v}$.

Man beachte, daß bei $\nabla \bullet \mathbf{v}$ und $\nabla \times \mathbf{v}$ nicht einfach nur Multiplikationen zwischen den Komponenten von ∇ und denen von $\mathbf{v}$ durchgeführt werden, sondern daß die partiellen Ableitungen in ∇ als Operatoren auf den Komponenten von $\mathbf{v}$ wirken! Die vereinfachte Schreibweise mit dem ∇ kann daher leicht zu Fehlern führen.

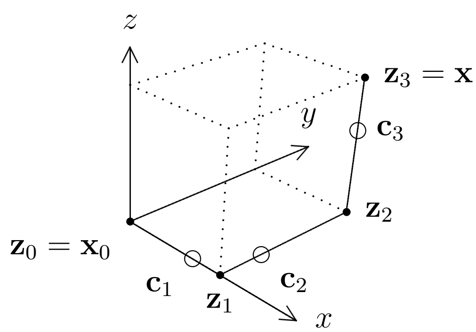
Divergenz und Rotation werden später ausführlicher in einem Kapitel über Vektoranalysis behandelt werden, mit dem Gradienten und seiner Bedeutung beschäftigen wir uns noch einmal weiter unten in diesem Paragraphen.

Lemma (schwacher Mittelwertsatz)

Sei $\mathbf{x}_0 = (x_1^{(0)}, \ldots, x_n^{(0)}) \in \mathbb{R}^n$, $f : U_\varepsilon(\mathbf{x}_0) \to \mathbb{R}$ partiell differenzierbar und $\mathbf{x} \in U_\varepsilon(\mathbf{x}_0)$ beliebig. Die Punkte $\mathbf{z}_0, \ldots, \mathbf{z}_n$ seien definiert durch $\mathbf{z}_0 := \mathbf{x}_0$ und $\mathbf{z}_i := \mathbf{z}_{i-1} + (x_i - x_i^{(0)}) \cdot \mathbf{e}_i$ für $i = 1, \ldots, n$.

Dann liegen alle $\mathbf{z}_i$ und die Verbindungsstrecken von $\mathbf{z}_{i-1}$ nach $\mathbf{z}_i$ in $U_\varepsilon(\mathbf{x}_0)$, und auf jeder dieser Verbindungsstrecken gibt es einen Punkt $\mathbf{c}_i$, so daß gilt:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}_i) \cdot (x_i - x_i^{(0)})\,.$$



Beweis: Es ist $\mathbf{z}_i = (x_1, \ldots, x_i, x_{i+1}^{(0)}, \ldots, x_n^{(0)})$, also $\|\mathbf{z}_i - \mathbf{x}_0\| \le \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon$. Wegen der Konvexität der Kugel liegen auch die Verbindungsstrecken in $U_\varepsilon(\mathbf{x}_0)$.

Sei $g_i : [0,1] \to \mathbb{R}$ definiert durch $g_i(t) := x_i^{(0)} + t(x_i - x_i^{(0)})$. Dann ist

$$\mathbf{z}_{i-1} + t(\mathbf{z}_i - \mathbf{z}_{i-1}) = (x_1, \ldots, x_{i-1}, g_i(t), x_{i+1}^{(0)}, \ldots, x_n^{(0)})\,.$$

Die Funktion $f_i(s) := f(x_1, \ldots, x_{i-1}, s, x_{i+1}^{(0)}, \ldots, x_n^{(0)})$ ist für jedes $t \in [0,1]$ in $g_i(t)$ differenzierbar, und es gilt:

$$f_i \circ g_i(t) = f(\mathbf{z}_{i-1} + t(\mathbf{z}_i - \mathbf{z}_{i-1})).$$

Weiter ist $f_i'(s) = f_{x_i}(x_1, \ldots, x_{i-1}, s, x_{i+1}^{(0)}, \ldots, x_n^{(0)})$ und daher

$$(f_i \circ g_i)'(t) = f_i'(g_i(t)) \cdot g_i'(t) = \frac{\partial f}{\partial x_i}(\mathbf{z}_{i-1} + t(\mathbf{z}_i - \mathbf{z}_{i-1})) \cdot (x_i - x_i^{(0)}).$$

Nach dem Mittelwertsatz gibt es ein $\xi_i \in (0,1)$ mit

$$(f_i \circ g_i)'(\xi_i) = f_i(g_i(1)) - f_i(g_i(0)) = f(\mathbf{z}_i) - f(\mathbf{z}_{i-1})\,.$$

Setzen wir $\mathbf{c}_i := \mathbf{z}_{i-1} + \xi_i(\mathbf{z}_i - \mathbf{z}_{i-1})$, so ist

$$\begin{aligned}\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{c}_i) \cdot (x_i - x_i^{(0)}) &= \sum_{i=1}^n (f(\mathbf{z}_i) - f(\mathbf{z}_{i-1})) \\ &= f(\mathbf{x}) - f(\mathbf{x}_0)\,. \end{aligned}$$

■

Folgerung (Spezielle Kettenregel)

Ist $B \subset \mathbb{R}^n$ offen, I ein Intervall, $\alpha : I \to B$ in $t_0 \in I$ ein differenzierbarer Weg und $f : B \to \mathbb{R}$ partiell differenzierbar und in $\mathbf{a} := \alpha(t_0)$ sogar stetig partiell differenzierbar, so ist auch $f \circ \alpha : I \to \mathbb{R}$ in t_0 differenzierbar, und es gilt:

$$(f \circ \alpha)'(t_0) = \nabla f(\mathbf{a}) \bullet \alpha'(t_0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\alpha(t_0)) \cdot \alpha_i'(t_0).$$

BEWEIS: Wir wählen ein $\varepsilon > 0$, so daß $U_\varepsilon(\mathbf{a}) \subset B$ ist, und ein $\delta > 0$, so daß $\alpha(t) \in U_\varepsilon$ ist, für $|t - t_0| < \delta$. Nach dem gerade bewiesenen Satz kann man zu jedem t Punkte $\mathbf{c}_i = \mathbf{c}_i(t)$, $i = 1, \ldots, n$, mit $\|\mathbf{c}_i - \mathbf{a}\| \le \|\alpha(t) - \mathbf{a}\|$ finden, so daß gilt:

$$f(\alpha(t)) - f(\alpha(t_0)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}_i)(\alpha_i(t) - \alpha_i(t_0)).$$

Teilt man beide Seiten durch $t - t_0$ und läßt man t gegen t_0 gehen, so streben alle Punkte $\mathbf{c}_i(t)$ gegen $\mathbf{a}$, und man erhält die Behauptung. ∎

Folgerung

Ist $B \subset \mathbb{R}^n$ offen, $f : B \to \mathbb{R}$ partiell differenzierbar und in $\mathbf{a} \in B$ sogar stetig partiell differenzierbar, so existieren in $\mathbf{a}$ alle Richtungsableitungen von f, und es ist $D_\mathbf{v} f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v}$.

BEWEIS: Für einen beliebigen Richtungsvektor $\mathbf{v} \ne \mathbf{0}$ sei $\alpha(t) := \mathbf{a} + t\mathbf{v}$. Dann ist $f \circ \alpha$ in $t = 0$ differenzierbar, und weil $\alpha'(t) \equiv \mathbf{v}$ ist, folgt:

$$(f \circ \alpha)'(0) = \nabla f(\mathbf{a}) \bullet \mathbf{v}.$$

Andererseits ist

$$(f \circ \alpha)'(0) = \lim_{t \to 0} \frac{f \circ \alpha(t) - f \circ \alpha(0)}{t - 0} = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t},$$

und das ist die Richtungsableitung $D_\mathbf{v} f(\mathbf{a})$. ∎

Wir können jetzt das Wesen des Gradienten etwas besser ergründen:

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ eine stetig partiell differenzierbare Funktion. Für $c \in \mathbb{R}$ sei

$$F_c := \{\mathbf{x} \in B \mid f(\mathbf{x}) = c\}$$

die entsprechende Niveaumenge von f.

Satz

Sei $\mathbf{a} \in B$, $f(\mathbf{a}) = c$ und $\nabla f(\mathbf{a}) \ne \mathbf{0}$.

1. *$\nabla f(\mathbf{a})$ zeigt in die Richtung, in der f am schnellsten wächst.*
2. *Ist $\alpha : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ ein differenzierbarer Weg mit $\alpha(0) = \mathbf{a}$, der ganz in F_c verläuft, so steht $\nabla f(\mathbf{a})$ auf $\alpha'(0)$ senkrecht.*

Beweis: 1) Wir betrachten beliebige Vektoren $\mathbf{v}$ mit $\|\mathbf{v}\| = 1$. Zu zeigen ist, daß $D_{\mathbf{v}}f(\mathbf{a})$ genau dann sein Maximum annimmt, wenn $\mathbf{v}$ in die Richtung des Gradienten zeigt. Tatsächlich ist

$$\begin{aligned} D_{\mathbf{v}}f(\mathbf{a}) &= \nabla f(\mathbf{a}) \bullet \mathbf{v} \\ &= \|\nabla f(\mathbf{a})\| \cdot \|\mathbf{v}\| \cdot \cos\theta, \end{aligned}$$

wobei $\theta \in [0, \pi]$ der Winkel zwischen $\mathbf{v}$ und $\nabla f(\mathbf{a})$ ist.

Dieser Ausdruck wird genau dann maximal, wenn $\theta = 0$ ist, also $\mathbf{v} = \dfrac{\nabla f(\mathbf{a})}{\|\nabla f(\mathbf{a})\|}$.

2) Verläuft α ganz in F_c, so ist $f \circ \alpha(t) \equiv c$, also

$$0 = (f \circ \alpha)'(0) = \nabla f(\mathbf{a}) \bullet \alpha'(0).$$

∎

Man sagt dann auch, der Gradient steht auf der Niveaumenge senkrecht.

§ 3 Totale Differenzierbarkeit

Inhalt:

Linearformen und Tangentialebenen, totale Differenzierbarkeit, Differential, Berechnung der totalen Ableitung, Differenzierbarkeitskriterium, Beispiele differenzierbarer Funktionen, Mittelwertsatz.

Wir wollen jetzt den Differenzierbarkeitsbegriff noch einmal überdenken. Bei der partiellen Differenzierbarkeit haben wir folgende Mängel festgestellt:

- Eine partiell differenzierbare Funktion braucht nicht stetig zu sein.
- Ist eine Funktion $2\times$ partiell differenzierbar, so hängen die Werte der zweiten Ableitungen von der Reihenfolge der Differentiation ab.

Erinnern wir uns noch einmal an die Situation in einer Veränderlichen:

Sei $I \subset \mathbb{R}$ ein offenes Intervall, $t_0 \in I$ und $f : I \to \mathbb{R}$ eine Funktion. Ist f in t_0 differenzierbar, so existiert der Grenzwert

$$f'(t_0) := \lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}.$$

Setzen wir

$$\delta(t) := \frac{f(t) - f(t_0)}{t - t_0} - f'(t_0),$$

so gilt:

1. $f(t) = f(t_0) + f'(t_0) \cdot (t - t_0) + \delta(t) \cdot (t - t_0)$ für $t \in I$.

   Hier ist $L(t) := f(t_0) + f'(t_0) \cdot (t - t_0)$ eine affin-lineare Funktion mit $L(t_0) = f(t_0)$, und der Ausdruck $\delta(t) \cdot (t - t_0)$ ist der „Fehler“, den man macht, wenn man f durch L approximiert.

2. $\lim_{t \to t_0} \delta(t) = 0$.

   Das zeigt, daß der Fehler mit $t \to t_0$ quadratisch gegen Null geht. Dadurch wird zum Ausdruck gebracht, daß sich die Graphen von f und L über t_0 nicht nur treffen, sondern sich sogar „tangential“ berühren.

3. Die Tangente an den Graphen von f im Punkte $(t_0, f(t_0))$ ist der Graph der affin-linearen Funktion

$$L(t) = f(t_0) + f'(t_0) \cdot (t - t_0).$$

Allgemein ist eine Gerade im $\mathbb{R}^2$ durch einen Punkt (a, b) gegeben durch eine Gleichung der Form

$$A(x-a)+B(y-b)=0, \text{ mit } A, B \in \mathbb{R} \text{ und } (A, B) \neq (0,0).$$

Soll diese Gerade der Graph einer affin-linearen Funktion sein, so darf sie nicht „senkrecht“ (d.h. nicht parallel zur y-Achse) verlaufen. Es muß also $B \neq 0$ sein, und man kann die Gleichung nach y auflösen:

$$y=b+\alpha(x-a), \text{ mit } \alpha:=-A/B.$$

Soll die Gerade in (a, b) den Graphen von f treffen, so muß $b=f(a)$ sein. Soll sie dort außerdem die gleiche Steigung wie f besitzen, so muß $\alpha=f'(a)$ sein, so daß die Geradengleichung die Form

$$y=f(a)+f'(a)(x-a)$$

annimmt.

Im Falle mehrerer Veränderlicher versuchen wir jetzt genauso vorzugehen. Der Einfachheit halber betrachten wir zunächst $n=2$.

Sei $G \subset \mathbb{R}^2$ ein Gebiet und $z=f(x, y)$ eine stetig partiell differenzierbare Funktion. Wir suchen die Tangentialebene an den Graphen im Punkt (a, b, c) mit $c=f(a, b)$. Eine solche Ebene im $\mathbb{R}^3$ wird durch eine Gleichung der Form

$$A(x-a)+B(y-b)+C(z-c)=0$$

beschrieben, mit $(A, B, C) \neq (0,0,0)$.

Damit die Ebene nicht senkrecht auf der x-y-Ebene steht, muß $C \neq 0$ sein. Also kann man die Gleichung folgendermaßen auflösen:

$$z=c+p(x-a)+q(y-b), \text{ mit } p=-A/C \text{ und } q=-B/C.$$

Die (senkrechte) Ebene $y=b$ trifft den Graphen von f in einem 1-dimensionalen Graphen $z=f(x, b)$. Die Tangente an diesen Graphen im Punkt $(x, b, z)=(a, b, c)$ ist durch die Gleichung $z=c+\dfrac{\partial f}{\partial x}(a, b)(x-a)$ gegeben. Da diese Tangente in der Tangentialebene enthalten sein soll, ist $p=\dfrac{\partial f}{\partial x}(a, b)$. Analog funktioniert es im Falle $x=a$. Die Gleichung der Tangentialebene ist also

$$\boxed{z=f(a, b)+\frac{\partial f}{\partial x}(a, b) \cdot(x-a)+\frac{\partial f}{\partial y}(a, b) \cdot(y-a).}$$

Jetzt müssen wir noch irgendwie zum Ausdruck bringen, daß die Tangentialebene den Graphen berührt. Wir versuchen es wieder mit einem Fehlerterm, der für $(x, y) \rightarrow(a, b)$ quadratisch gegen Null strebt.

Weil wir vorausgesetzt haben, daß $z = f(x,y)$ stetig partiell differenzierbar ist, haben wir die Darstellung

$$\begin{aligned} f(x,y) &= f(a,b) + f_x(\mathbf{c}_1)(x-a) + f_y(\mathbf{c}_2)(y-b) \\ &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &\qquad + \delta_1(x,y)(x-a) + \delta_2(x,y)(y-b), \end{aligned}$$

mit $\delta_1(x,y) := f_x(\mathbf{c}_1) - f_x(a,b)$ und $\delta_2(x,y) := f_y(\mathbf{c}_2) - f_y(a,b)$. Dabei liegen die Punkte $\mathbf{c}_1, \mathbf{c}_2$ jeweils sehr nahe bei (a,b). Für $(x,y) \to (a,b)$ streben sie gegen (a,b), und wegen der Stetigkeit der partiellen Ableitungen strebt dann $\delta(x,y) = (\delta_1(x,y), \delta_2(x,y))$ für $(x,y) \to (a,b)$ gegen Null.

Definition:

Sei $G \subset \mathbb{R}^n$ ein Gebiet, $f : G \to \mathbb{R}$ eine Funktion und $\mathbf{x}_0 \in G$ ein Punkt.

f heißt in $\mathbf{x}_0$ *(total) differenzierbar*, wenn es eine Linearform $\lambda : \mathbb{R}^n \to \mathbb{R}$ und eine auf G definierte vektorwertige Funktion δ gibt, so daß in der Nähe von $\mathbf{x}_0$ gilt:

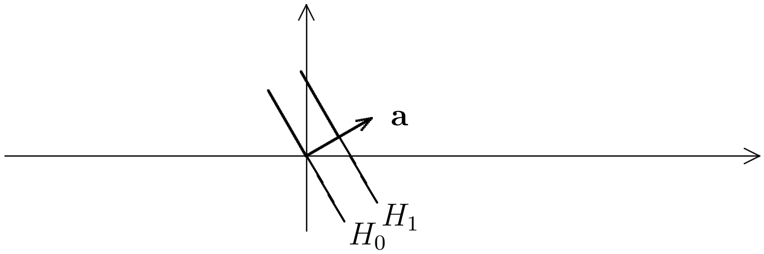
1. $f(\mathbf{x}) = f(\mathbf{x}_0) + \lambda(\mathbf{x} - \mathbf{x}_0) + \delta(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0)$.
2. $\lim\limits_{\mathbf{x} \to \mathbf{x}_0} \delta(\mathbf{x}) = 0$.

Die (dadurch eindeutig bestimmte) Linearform λ heißt die *(totale) Ableitung* oder das *(totale) Differential* von f in $\mathbf{x}_0$. Wir schreiben dafür auch $(df)_{\mathbf{x}_0}$.

Eine Linearform $\lambda \in L(\mathbb{R}^n, \mathbb{R})$ kann immer in der Form

$$\lambda(\mathbf{x}) = \mathbf{a} \bullet \mathbf{x} = \mathbf{a} \cdot \mathbf{x}^t = a_1 x_1 + \cdots + a_n x_n$$

dargestellt werden, mit einem festen Vektor $\mathbf{a}$. Anschaulich wird so eine Linearform durch die Hyperebenen $H_0 = \{\mathbf{x} \,:\, \lambda(\mathbf{x}) = 0\} = \mathrm{Ker}(\lambda)$ und $H_1 = \{\mathbf{x} \,:\, \lambda(\mathbf{x}) = 1\}$ repräsentiert. Dabei ist $\mathrm{Ker}(\lambda)$ das orthogonale Komplement zu der Geraden $\mathbb{R}\mathbf{a}$, also die Menge aller Vektoren, die auf $\mathbf{a}$ senkrecht stehen. Im Falle $n = 2$ sieht das folgendermaßen aus:



Die Hyperebene H_1 ist parallel zu H_0 und geht durch den Punkt $\mathbf{a}_1 := \mathbf{a}/\|\mathbf{a}\|^2$. Ein beliebiger Vektor $\mathbf{x} \in \mathbb{R}^n$ besitzt eine eindeutige Darstellung $\mathbf{x} = x_1\mathbf{a}_1 + \mathbf{a}_0$, mit $a_1 \in \mathbb{R}$ und $\mathbf{a}_0 = \mathrm{pr}_{H_0}(\mathbf{x}) \in \mathrm{Ker}(\lambda)$. Dann ist

$$\lambda(\mathbf{x}) = x_1\lambda(\mathbf{a}_1) + \lambda(\mathbf{a}_0) = x_1.$$

Der Graph G_L der affin-linearen Abbildung $L(\mathbf{x}) = f(\mathbf{x}_0) + \lambda(\mathbf{x} - \mathbf{x}_0)$ ist eine Hyperebene im Raum $\mathbb{R}^{n+1}$ und berührt dort den Graphen $G_f = \{(\mathbf{x}, z) \in \mathbb{R}^n \times \mathbb{R} : z = f(\mathbf{x})\}$ im Punkt $(\mathbf{x}_0, f(\mathbf{x}_0))$. Definieren wir die Linearform $\Lambda \in L(\mathbb{R}^{n+1}, \mathbb{R})$ durch

$$\Lambda(\mathbf{h}, t) := \lambda(\mathbf{h}) - t,$$

so ist

$$\begin{aligned}(\mathbf{x}_0, f(\mathbf{x}_0)) + \mathrm{Ker}(\Lambda) &= \{(\mathbf{x}, z) \,:\, \Lambda(\mathbf{x} - \mathbf{x}_0, z - f(\mathbf{x}_0)) = 0\} \\ &= \{(\mathbf{x}, z) \,:\, f(\mathbf{x}_0) + \lambda(\mathbf{x} - \mathbf{x}_0) - z = 0\} \;=\; G_L.\end{aligned}$$

Bemerkung. Die Bedingungen (1) und (2) für die totale Differenzierbarkeit in einem Punkt $\mathbf{x}_0$ sind zu der folgenden Aussage äquivalent:

Es gibt eine Linearform $\lambda : \mathbb{R}^n \to \mathbb{R}$, so daß gilt:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \lambda(\mathbf{h})}{\|\mathbf{h}\|} = 0.$$

Ist nämlich f differenzierbar, so benutzen wir die Ungleichung von Cauchy-Schwarz:

$$|\delta(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0)| \le \|\delta(\mathbf{x})\| \cdot \|\mathbf{x} - \mathbf{x}_0\|\,.$$

Daraus folgt:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \lambda(\mathbf{h})}{\|\mathbf{h}\|} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{|\delta(\mathbf{x}_0 + \mathbf{h}) \bullet \mathbf{h}|}{\|\mathbf{h}\|} \le \lim_{\mathbf{h}\to\mathbf{0}} \|\delta(\mathbf{x}_0 + \mathbf{h})\| = 0.$$

Ist umgekehrt das Kriterium erfüllt, so setzen wir

$$\delta(\mathbf{x}) := \frac{\mathbf{x} - \mathbf{x}_0}{\|\mathbf{x} - \mathbf{x}_0\|^2} \cdot \big(f(\mathbf{x}) - f(\mathbf{x}_0) - \lambda(\mathbf{x} - \mathbf{x}_0)\big)\,.$$

Dann sind die Bedingungen für die Differenzierbarkeit erfüllt.

Berechnung der totalen Ableitung

Sei $f : B \to \mathbb{R}$ in $\mathbf{x}_0 \in B$ differenzierbar. Dann existieren in $\mathbf{x}_0$ auch sämtliche Richtungsableitungen von f, und es gilt:

$$(df)_{\mathbf{x}_0}(\mathbf{v}) = D_{\mathbf{v}} f(\mathbf{x}_0).$$

Insbesondere ist f in $\mathbf{x}_0$ nach allen Variablen partiell differenzierbar und

$$(df)_{\mathbf{x}_0}(\mathbf{v}) = \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \cdot v_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \cdot v_n = \nabla f(\mathbf{x}_0) \bullet \mathbf{v}.$$

Aus diesem Satz folgt auch sofort die Eindeutigkeit der Ableitung!

BEWEIS: Ist f in $\mathbf{x}_0$ differenzierbar, so haben wir die Darstellung

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \lambda(\mathbf{x} - \mathbf{x}_0) + \delta(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0).$$

Sei jetzt $\mathbf{v} = (v_1, \ldots, v_n)$ ein Richtungsvektor $\neq \mathbf{0}$ und $t \in \mathbb{R}$, $t \neq 0$. Dann gilt:

$$f(\mathbf{x}_0 + t\mathbf{v}) = f(\mathbf{x}_0) + t \cdot \lambda(\mathbf{v}) + t \cdot \delta(\mathbf{x}_0 + t\mathbf{v}) \bullet \mathbf{v},$$

mit

$$\lim_{t \to 0} \delta(\mathbf{x}_0 + t\mathbf{v}) = 0.$$

Also ist

$$\lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0) - t \cdot \lambda(\mathbf{v})}{t} = 0$$

und damit

$$D_{\mathbf{v}} f(\mathbf{x}_0) = \lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = \lambda(\mathbf{v}).$$

Insbesondere existieren die partiellen Ableitungen $\dfrac{\partial f}{\partial x_i}(\mathbf{x}_0) = D_{\mathbf{e}_i} f(\mathbf{x}_0)$ für $i = 1, \ldots, n$, und es gilt:

$$\lambda(\mathbf{v}) = \lambda(\sum_{i=1}^{n} v_i \mathbf{e}_i) = \sum_{i=1}^{n} v_i \lambda(\mathbf{e}_i) = \sum_{i=1}^{n} v_i D_{\mathbf{e}_i} f(\mathbf{x}_0) = \mathbf{v} \bullet \nabla f(\mathbf{x}_0).$$

■

Dieser Satz erlaubt es jetzt, totale Ableitungen mit Hilfe von partiellen Ableitungen auszurechnen, und für die letzteren brauchen wir ja nur den Kalkül aus der Theorie einer Veränderlichen zu übernehmen.

Eine **Warnung** muß allerdings ausgesprochen werden! Es gibt Funktionen, die partiell, aber nicht total differenzierbar sind, während andererseits gilt:

Total differenzierbare Funktionen sind stetig

Ist f in $\mathbf{x}_0$ total differenzierbar, so ist f dort auch stetig.

BEWEIS: Wir haben

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \lambda(\mathbf{x} - \mathbf{x}_0) + \delta(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0),$$

mit einer (stetigen) Linearform λ und $\delta(\mathbf{x}) \to 0$ für $\mathbf{x} \to \mathbf{x}_0$. Also strebt $f(\mathbf{x})$ für $\mathbf{x} \to \mathbf{x}_0$ gegen $f(\mathbf{x}_0)$. ■

Wir haben schon ein Beispiel einer Funktion gesehen, die im Nullpunkt partiell differenzierbar, aber nicht stetig ist. Sie kann dann natürlich erst recht nicht total differenzierbar sein.

Wir stehen damit vor einem Dilemma: Bevor wir die Ableitung einer Funktion f mit Hilfe der partiellen Ableitungen ausrechnen können, müssen wir die totale Differenzierbarkeit beweisen. Zum Glück gibt es folgendes einfache Kriterium:

Differenzierbarkeitskriterium

Sei $B \subset \mathbb{R}^n$ offen, $f : B \to \mathbb{R}$ eine Funktion und $\mathbf{x}_0 \in B$ ein Punkt.

Wenn es eine offene Umgebung U von $\mathbf{x}_0$ in B gibt, so daß alle partiellen Ableitungen von f auf U existieren und in $\mathbf{x}_0$ stetig sind, dann ist f in $\mathbf{x}_0$ total differenzierbar.

Den Beweis haben wir uns oben im Falle $n = 2$ schon überlegt. Bei beliebigem n geht's genauso.

Beispiele.

1. Sei $f(\mathbf{x}) \equiv c$ konstant. Dann verschwinden alle partiellen Ableitungen, und da die Nullfunktion stetig ist, ist f total differenzierbar und $(df)_{\mathbf{x}} = 0$ (die „Null-Form") in jedem Punkt $\mathbf{x}$ des $\mathbb{R}^n$.

2. Sei $f(\mathbf{x}) := \mathbf{u} \bullet \mathbf{x} = u_1x_1 + \cdots + u_nx_n$ selbst schon eine Linearform. Dann ist $f_{x_i}(\mathbf{x}) \equiv u_i$ konstant (und damit stetig) für alle i. Also ist f überall total differenzierbar, und offensichtlich ist $(df)_{\mathbf{x}}(\mathbf{v}) = \mathbf{u} \bullet \mathbf{v} = f(\mathbf{v})$ für jedes $\mathbf{x}$. Die Ableitung einer Linearform f stimmt in jedem Punkt $\mathbf{x}$ des $\mathbb{R}^n$ mit genau dieser Linearform überein.

   Ein Spezialfall ist die Linearform $x_i : \mathbf{v} \mapsto \mathbf{e}_i \bullet \mathbf{v} = v_i$. In jedem Punkt $\mathbf{x}$ ist das Differential $(dx_i)_{\mathbf{x}}$ wieder die Projektion auf die i-te Komponente. Deshalb bezeichnen wir diese Projektion auch einfach mit dem Symbol dx_i (ohne Index $\mathbf{x}$).

3. Nun sei $A = (a_{ij}) \in M_{n,n}(\mathbb{R})$ eine symmetrische Matrix, d.h. $A^t = A$, und

$$\begin{aligned} f(\mathbf{x}) &:= \mathbf{x} \cdot A \cdot \mathbf{x}^t = (x_1, \ldots, x_n) \cdot \Big(\sum_{j=1}^n a_{1j}x_j, \ldots, \sum_{j=1}^n a_{nj}x_j \Big) \\ &= \sum_{i,j=1}^n a_{ij}x_ix_j \end{aligned}$$

   die durch A bestimmte „quadratische Form". Um die Ableitung in einem Punkt $\mathbf{a}$ zu bestimmen, bleiben wir bei der vektoriellen Schreibweise. Es ist

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) &= (\mathbf{a}+\mathbf{h}) \cdot A \cdot (\mathbf{a}+\mathbf{h})^t - \mathbf{a} \cdot A \cdot \mathbf{a}^t \\ &= \mathbf{a} \cdot A \cdot \mathbf{a}^t + \mathbf{h} \cdot A \cdot \mathbf{a}^t + \mathbf{a} \cdot A \cdot \mathbf{h}^t + \mathbf{h} \cdot A \cdot \mathbf{h}^t \\ &\quad -\mathbf{a} \cdot A \cdot \mathbf{a}^t \\ &= 2\mathbf{a} \cdot A \cdot \mathbf{h}^t + \mathbf{h} \cdot A \cdot \mathbf{h}^t \;=\; \lambda(\mathbf{h}) + \delta(\mathbf{a}+\mathbf{h}) \bullet \mathbf{h}, \end{aligned}$$

mit $L(\mathbf{h}) := (2\mathbf{a} \cdot A) \bullet \mathbf{h}$ und $\delta(\mathbf{x}) := (\mathbf{x} - \mathbf{a}) \cdot A$. Offensichtlich ist λ eine Linearform, und

$$\lim_{\mathbf{x} \to \mathbf{a}} \delta(\mathbf{x}) = 0.$$

Somit ist f total differenzierbar und $(df)_{\mathbf{a}}(\mathbf{h}) = 2\mathbf{a} \cdot A \cdot \mathbf{h}^t$.

4. Ist f eine beliebige total differenzierbare Funktion, so ist

$$\begin{aligned} (df)_{\mathbf{x}}(\mathbf{v}) &= f_{x_1}(\mathbf{x}) \cdot v_1 + \cdots + f_{x_n}(\mathbf{x}) \cdot v_n \\ &= f_{x_1}(\mathbf{x}) \cdot (dx_1)_{\mathbf{x}}(\mathbf{v}) + \cdots + f_{x_n}(\mathbf{x}) \cdot (dx_n)_{\mathbf{x}}(\mathbf{v}), \end{aligned}$$

also

$$(df)_{\mathbf{x}} = f_{x_1}(\mathbf{x}) \cdot (dx_1)_{\mathbf{x}} + \cdots + f_{x_n}(\mathbf{x}) \cdot (dx_n)_{\mathbf{x}}.$$

Im Falle $n = 1$ (und einer Funktion $y = f(x)$) wird daraus die Formel

$$(df)_x = f'(x) \cdot (dx)_x.$$

Hier sind $(df)_x$ und $(dx)_x$ Linearformen auf $\mathbb{R}$ und keine reellen Zahlen. Deshalb ist $f'(x) = (df)_x/(dx)_x$ kein echter Quotient (von Differentialen).

5. Sei $f(x,y) := e^{x^2} \cdot \cos(y)$ und $\mathbf{a} := (0, \frac{\pi}{4})$.

Dann ist $f_x = 2xe^{x^2} \cdot \cos(y)$ und $f_y = -e^{x^2} \cdot \sin(y)$, also

$$(df)_{\mathbf{a}} = f_x(\mathbf{a})\,(dx)_{\mathbf{a}} + f_y(\mathbf{a})\,(dy)_{\mathbf{a}} = -\frac{1}{2}\sqrt{2}\,(dy)_{\mathbf{a}}.$$

6. Sei $f(x,y) := \begin{cases} \dfrac{xy^2}{x^2+y^2} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0). \end{cases}$

Wir zeigen zunächst, daß f im Nullpunkt stetig ist: Sei $(\mathbf{x}_\nu)$ eine Nullfolge. Dann können wir schreiben:

$$\mathbf{x}_\nu = (r_\nu \cos(\varphi_\nu), r_\nu \sin(\varphi_\nu)), \text{ für } \nu \in \mathbb{N}.$$

Dabei konvergiert $r_\nu = \|\mathbf{x}_\nu\|$ gegen Null, und unabhängig von φ_ν ist

$$(\cos\varphi_\nu)^2 + (\sin\varphi_\nu)^2 = 1 \text{ und } 0 \le |\cos\varphi_\nu|, |\sin\varphi_\nu| \le 1.$$

Also konvergiert

$$|f(\mathbf{x}_\nu)| = |\frac{r_\nu^3 \cos\varphi_\nu (\sin\varphi_\nu)^2}{r_\nu^2}| \le r_\nu$$

gegen Null.

Weiter ist $f(x,0) \equiv 0$ und $f(0,y) \equiv 0$. Also ist f im Nullpunkt auch partiell differenzierbar, und es gilt:

$$\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0.$$

Es existieren sogar beliebige Richtungsableitungen:

Da $f(tx,ty) = t \cdot f(x,y)$ für alle t und beliebiges (x,y) gilt (man nennt f daher auch eine *homogene* Funktion), ist

$$D_{\mathbf{h}} f(\mathbf{0}) = \lim_{t\to 0} \frac{f(t\mathbf{h}) - f(\mathbf{0})}{t} = f(\mathbf{h}).$$

Man kann also an G_f im Nullpunkt in jeder beliebigen Richtung eine Tangente legen.

Wäre f in $\mathbf{0}$ total differenzierbar, so müßte $(df)_{\mathbf{0}} = 0$ sein. Für $\mathbf{h} := (r,r)$ ist aber

$$\frac{f(\mathbf{h}) - f(\mathbf{0}) - 0}{\|\mathbf{h}\|} = \frac{r^3}{2r^2 \cdot \sqrt{2}|r|} = \pm\frac{1}{2\sqrt{2}},$$

und das kann nicht gegen Null konvergieren.

Also ist f im Nullpunkt nicht total differenzierbar, und der Graph von f besitzt dort keine Tangentialebene. Wie soll man sich das vorstellen?

Da f homogen ist, gehört mit $(\mathbf{x}, z)$ auch jeder Punkt $(t\mathbf{x}, tz)$ zum Graphen von f, also die ganze Gerade durch $(\mathbf{x}, z)$ und den Nullpunkt. Diese Geraden sind dann natürlich auch Tangenten, und sie müßten daher auch in einer etwa existierenden Tangentialebene enthalten sein. Das ist nicht möglich, weil die Geraden gar nicht alle in einer Ebene liegen. Die Punkte $(1,1,\frac{1}{2})$, $(1,-1,\frac{1}{2})$ und $(1,0,0)$ liegen z.B. auf G_f, sind aber linear unabhängig.

Tatsächlich hat G_f im Nullpunkt eine „Spitze“, und dieser Mangel an Glattheit verhindert die totale Differenzierbarkeit.

Der Mittelwertsatz

Sei $B \subset \mathbb{R}^n$ offen und konvex, $f : B \to \mathbb{R}$ eine differenzierbare Funktion, $\mathbf{a}, \mathbf{b} \in B$ zwei Punkte.

Dann gibt es ein $t \in (0,1)$ mit

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) \bullet (\mathbf{b} - \mathbf{a}).$$

BEWEIS: Wir setzen $\alpha(t) := \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ und $h(t) := f \circ \alpha(t)$. Dies ist eine auf $[0, 1]$ stetige und auf $(0, 1)$ differenzierbare Funktion. Nach dem Mittelwertsatz in einer Veränderlichen gibt es ein $t \in (0, 1)$, so daß

$$h(1) - h(0) = h'(t) \cdot (1 - 0) = h'(t)$$

ist. Es ist aber $h(1) - h(0) = f(\mathbf{b}) - f(\mathbf{a})$ und

$$h'(t) = D_{\mathbf{b}-\mathbf{a}} f(\alpha(t)) = \nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) \bullet (\mathbf{b} - \mathbf{a}).$$

■

Folgerung

Sei $G \subset \mathbb{R}^n$ ein Gebiet, $f : G \to \mathbb{R}$ differenzierbar und $\nabla f(\mathbf{x}) = \mathbf{0}$ für alle $\mathbf{x} \in G$. Dann ist f konstant.

BEWEIS: Sei $\mathbf{x}_0 \in G$ und $c := f(\mathbf{x}_0)$. Dann ist die Menge

$$M := \{\mathbf{x} \in G \, : \, f(\mathbf{x}) = c\}$$

nicht leer. Ist $\mathbf{y} \in M$, so gibt es eine kleine Kugel $U = U_\varepsilon(\mathbf{y})$, die noch ganz in G liegt. Für $\mathbf{x} \in U$ ist

$$f(\mathbf{x}) - f(\mathbf{y}) = \nabla f(\mathbf{y} + t(\mathbf{y} - \mathbf{x})) \bullet (\mathbf{y} - \mathbf{x}) = 0,$$

also $f(\mathbf{x}) = f(\mathbf{y}) = c$. Damit gehört U zu M und M ist offen.

Weil f stetig ist, ist auch die Menge $G \setminus M = \{\mathbf{x} \in G \, : \, f(\mathbf{x}) \neq c\}$ offen. Also muß $M = G$ sein. ■