Analysis 2

Johannes Ebert[1]

[1]Vorlesung gehalten im Sommersemester 2015, gesetzt von Henrik Graßhoff

Inhaltsverzeichnis

Teil I

Mehrdimensionale Differentialrechnung

In dieser Vorlesung wollen wir die Erkenntnisse, die wir in Analysis 1 über Folgen und Funktionen in $\mathbb{R}$ gewonnen haben, auf den euklidischen Raum beliebiger endlicher Dimensionen erweitern. Im groben Überblick werden wir in folgender Reihenfolge abstrahieren: Länge, Abstand, Konvergenz, Stetigkeit, Differenzierbarkeit. Auf diesem Weg werden wir uns eingangs mit Grundlagen der Topologie befassen. Alle topologischen Überlegungen beruhen auf einem sehr allgemeinen Begriff der Nähe. Diesen Nähebegriff werden wir im ersten Abschnitt insbesondere mit der Definition offener und abgeschlossener Mengen entwickeln. Zunächst aber befassen wir uns mit der Abstraktion von Länge und Abstand.

1 Metrische Räume

Für $n \in \mathbb{N}$ ist die Menge

$$\mathbb{R}^n := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid x_i \in \mathbb{R} \right\}$$

ist mit komponentenweiser Addition und Multiplikation ein $\mathbb{R}$-Vektorraum der Dimension $\dim_{\mathbb{R}} \mathbb{R}^n = n$. Eine Basis von $\mathbb{R}^n$ ist die Menge der kanonischen Basisvektoren $\{e_1, \ldots, e_n\}$, auch Standardbasis genannt, mit

$$e_i := \begin{pmatrix} \delta_{1i} \\ \vdots \\ \delta_{ni} \end{pmatrix}$$

für $1 \leq i \leq n$, wobei wir mit

$$\delta_{ij} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

das so genannte *Kronecker-Delta* ist. Dieser Vektorraum wird durchgehend unser Standardbeispiel sein, die mehrdimensionale Differentialrechnung werden wir sogar ausschließlich für diesen Vektorraum erarbeiten. Aus diesem Grund sollten wir diesen Vektorraum stets im Hinterkopf behalten.

1.1 Normierte Räume

Definition 1.1.1 (Skalarprodukt). Sei V ein $\mathbb{R}$-Vektorraum. Eine Abbildung $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}, (x, y) \mapsto \langle x, y \rangle$ heißt *Skalarprodukt* auf V, wenn für alle $x, y, z \in V$, $a \in \mathbb{R}$ die folgenden Axiome erfüllt sind:

1. $\langle x, x \rangle \geq 0$ und $\langle x, x \rangle = 0 \Leftrightarrow x = 0$ (Positive Definitheit).
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ und $\langle ax, y \rangle = a\langle x, y \rangle = \langle x, ay \rangle$ (Bilinearität).
3. $\langle x, y \rangle = \langle y, x \rangle$ (Symmetrie).

Das Paar $(V, \langle \cdot, \cdot \rangle)$ heißt euklidischer Vektorraum.

Beispiel (Standardskalarprodukt). Das Standardskalarprodukt auf $\mathbb{R}^n$ ist die Abbildung $\langle \cdot, \cdot \rangle \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ mit

$$\left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle := \sum_{i=1}^{n} x_i y_i.$$

Beispielsweise ist $\left\langle \left(\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} -1 \\ 3 \end{smallmatrix}\right) \right\rangle = 2 \cdot (-1) + 4 \cdot 3 = -2 + 12 = 10$.

Definition 1.1.2 (Norm). Sei V ein $\mathbb{R}$-Vektorraum. Eine *Norm* auf V ist eine Abbildung $\|\cdot\| \colon V \to \mathbb{R}, x \mapsto \|x\|$, welche für $x, y \in V$, $a \in \mathbb{R}$ die folgenden Axiome erfüllt:

1. $\|x\| = 0 \Rightarrow x = 0$ (Definitheit).
2. $\|ax\| = |a| \cdot \|x\|$ (Absolute Homogenität).
3. $\|x + y\| \leq \|x\| + \|y\|$ (Subadditivität bzw. Dreiecksungleichung).

Das Paar $(V, \|\cdot\|)$ heißt normierter Vektorraum.

Korollar 1.1.3. *Sei $(V, \|\cdot\|)$ ein normierter Vektorraum. Für $x, y \in V$ gilt:*

1. $\|x\| \geq 0$.
2. $\|0\| = 0$.
3. $\|-x\| = \|x\|$.
4. $\|x - y\| = \|y - x\|$.

Beweis. Zu 3. Es gilt $\|-x\| = \|(-1) \cdot x\| = |-1| \cdot \|x\| = \|x\|$.

Zu 2. Es ist $\|0\| = \|0 \cdot 0\| = |0| \cdot \|0\| = 0$.

Zu 1. Dies folgt aus $0 = \|0\| = \|x - x\| \leq \|x\| + \|-x\| = 2\|x\|$.

Zu 4. Diese Aussage folgt aus $\|x - y\| = \|-(y - x)\|$ und Aussage 3. □

Die Norm ist also anschaulicherweise eine Abbildung, welche den Begriff der Länge verallgemeinert. All ihre Eigenschaften stimmen mit unserer Vorstellung einer Länge überein, beispielsweise die Definitheit oder die absolute Homogenität. Wir klären nun den Zusammenhang zwischen Skalarprodukt und Norm und geben einige Beispiele für Normen an.

Definition 1.1.4 (Induzierte Norm). Ist $(V, \langle\cdot,\cdot\rangle)$ ein euklidischer Vektorraum, so ist die durch $\langle\cdot,\cdot\rangle$ *induzierte Norm* für $x \in V$ definiert durch

$$\|x\| := \sqrt{\langle x, x\rangle}.$$

Dieser Wert ist wegen der Skalarproduktaxiome wohldefiniert.

Satz 1.1.5. *Sei $(V, \langle\cdot,\cdot\rangle)$ ein euklidischer Vektorraum. Dann genügt die von $\langle\cdot,\cdot\rangle$ induzierte Norm $\|\cdot\|$ den Normaxiomen und für $x, y \in V$ gilt zudem die Cauchy-Schwarz-Ungleichung*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|.$$

Beweis. Aus $\|x\| = \sqrt{\langle x, x\rangle} = 0$ folgt $\langle x, x\rangle = 0$, also $x = 0$ und damit die Definitheit von $\|\cdot\|$. Für $a \in \mathbb{R}$ erhalten wir $\|ax\| = \sqrt{\langle ax, ax\rangle} = \sqrt{a^2\langle x, x\rangle} = \sqrt{a^2}\sqrt{\langle x, x\rangle} = |a| \cdot \|x\|$ und $\|\cdot\|$ ist absolut homogen. Für den Beweis der Dreiecksungleichung benutzen wir die Cauchy-Schwarz-Ungleichung, die wir für den nichttrivialen Spezialfall $y \neq 0$ nun mithilfe einer Hilfsfunktion beweisen: Sei

$$f : \mathbb{R} \to \mathbb{R}, f(t) := \langle x + ty, x + ty\rangle = \|x\|^2 + 2t\langle x, y\rangle + t^2\|y\|^2.$$

Da $\langle\cdot,\cdot\rangle$ positiv definit ist, ist $f \geq 0$. Ferner erkennen wir, dass f eine nach oben geöffnete Parabel ist, also ein globales Minimum besitzt. Wegen

$$f'(t) = 0 \Leftrightarrow 2\langle x, y\rangle + 2t\|y\|^2 = 0 \Leftrightarrow t = -\frac{\langle x, y\rangle}{\|y\|^2} =: T$$

sowie $f''(t) = 2\|y\|^2 > 0$ wird dieses Minimum bei T angenommen. Dort gilt

$$0 \leq f(T) = \|x\|^2 - 2\frac{\langle x, y\rangle^2}{\|y\|^2} + \frac{\langle x, y\rangle^2}{\|y\|^2} = \|x\|^2 - \frac{\langle x, y\rangle^2}{\|y\|^2}$$
$$\Rightarrow \frac{\langle x, y\rangle^2}{\|y\|^2} \leq \|x\|^2 \Rightarrow \langle x, y\rangle^2 \leq \|x\|^2 \cdot \|y\|^2 \Rightarrow |\langle x, y\rangle| \leq \|x\| \cdot \|y\|.$$

Nun beweisen wir die Dreiecksungleichung. Es ist

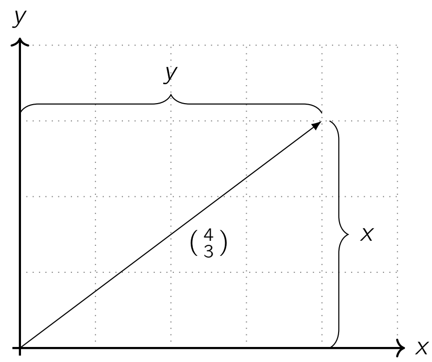
$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2\end{aligned}$$

und mit Radizieren folgt die zu zeigende Dreiecksungleichung. □

Bemerkung. Man notiert die Ungleichung auch $|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$. In den Komponenten von x ausgeschrieben lautet die Ungleichung übrigens

$$|\sum_{i=1}^n x_i y_i| \leq \sqrt{\sum_{i=1}^n x_i^2}\sqrt{\sum_{i=1}^n y_i^2}.$$

Wir wollen nun die Länge eines Vektors so berechnen, wie es unserer intuitiven Anschauung entspricht: Wir beginnen dazu im $\mathbb{R}^2$. Ein Vektor $v = \binom{x}{y}$ stellen wir uns dort als Pfeil vor:

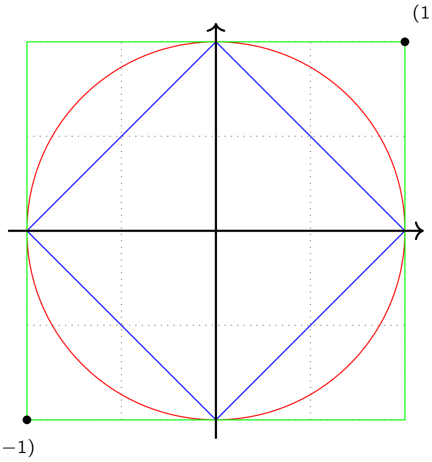


Dessen Länge berechnen wir als Hypotenuse des Dreiecks, welches zwischen den Punkten $(0,0)$, $(4,0)$, $(4,3)$ liegt. Mit dem Satz des Pythagoras folgt Länge $\binom{x}{y} = \sqrt{x^2+y^2}$. Vollzieht man dieselbe Überlegung im $\mathbb{R}^3$, so erhält man wieder durch Anwenden des Satzes des Pythagoras, dass Länge $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \sqrt{x^2+y^2+z^2}$. Gewissermaßen also entspricht diese Längenberechnung unserer natürlichen Vorstellung der Länge. Aus diesem Grund wird diese Norm die für unsere Arbeit wichtigste Norm werden; in der Tat wird sie sogar durch das Standardskalarprodukt induziert. Wir wollen diese Norm jedoch verallgemeinern und gelangen zu folgendem Begriff.

Definition 1.1.6 (p-Norm). Sei $x = \sum_{i=1}^n x_i e_i \in \mathbb{R}^n$. Für $p \in [1, \infty)$ ist die *p-Norm* definiert durch

$$\|x\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}.$$

Man setzt die ∞- bzw. *Maximumsnorm* als $\|x\|_\infty := \max_{1\leq i\leq n}|x_i|$. Für $p = 1$ bzw. $p = 2$ heißen die p-Normen auch *Betrags-* bzw. *euklidische Norm*.



Diese Abbildung zeigt, wie der Einheitskreis $\{x \in \mathbb{R}^2 \mid \|x\|_p = 1\}$ unter verschiedenen p-Normen aussieht, nämlich für $p = 1$, $p = 2$ und $p = \infty$. Wir erkennen hier, dass erstens ein Kreis unter verschiedenen Normen sehr verschieden aussehen kann und dass zweitens (wieder) lediglich die 2-Norm unserer natürlichen Vorstellung eines Kreises gerecht wird.

Satz 1.1.7. *Für jedes $p \in [1, \infty]$ ist die p-Norm eine Norm auf $\mathbb{R}^n$.*

Beweis. Seien $x = (x_1, \dots, x_n)^T, y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$.

Wir zeigen zunächst, dass die 1-Norm eine Norm ist. Dazu müssen wir Definitheit, Betragshomogenität und Subadditivität nachweisen. Aus

$$\|x\|_1 = \sum_{i=1}^n |x_i| = 0 \Rightarrow |x_i| = 0 \text{ für alle } i \Rightarrow x = 0$$

folgt die Definitheit. Die Homogenität ergibt sich für $a \in \mathbb{R}$ aus

$$\|ax\|_1 = \sum_{i=1}^n |ax_i| = |a| \sum_{i=1}^n |x_i| = |a| \cdot \|x\|_1,$$

die Subadditivität folgt für $y \in \mathbb{R}^n$ aus

$$\|x+y\|_1 = \sum_{i=1}^n |x_i + y_i| \le \sum_{i=1}^n |x_i| + |y_i| = \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|x\|_1 + \|y\|_1.$$

Nun betrachten wir die ∞-Norm. Diese ist wegen

$$\|x\| = \max_{1 \le i \le n} |x_i| = 0 \Rightarrow x_i = 0 \text{ für alle } i \Rightarrow x = 0$$

definit und wegen

$$\|ax\|_\infty = \max_{1 \le i \le n} |ax_i| = \max_{1 \le i \le n} \{|a| \cdot |x_i|\} = |a| \max_{1 \le i \le n} |x_i| = |a| \cdot \|x\|_\infty$$

für alle $a \in \mathbb{R}$ betragshomogen. Wir zeigen nun die Dreiecksungleichung. Für $x, y \in \mathbb{R}^n$ gilt wegen der Dreiecksungleichung des Absolutbetrags auf $\mathbb{R}$ die Abschätzung

$$\begin{aligned} \|x+y\|_\infty &= \max_{1 \le i \le n} |x_i + y_i| \le \max_{1 \le i \le n} \{|x_i| + |y_i|\} \\ &\le \max_{1 \le i \le n} |x_i| + \max_{1 \le i \le n} |y_i| = \|x\|_\infty + \|y\|_\infty \end{aligned}$$

und die ∞-Norm genügt den Normaxiomen.

Sei nun $p \in (1, \infty)$. Wir beweisen zunächst Definitheit und Betragshomogenität. Diese Eigenschaften folgen aus

$$\|x\| = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} = 0 \Rightarrow \sum_{i=1}^{n} |x_i|^p = 0 \Rightarrow x_i = 0 \text{ für alle } i \Rightarrow x = 0$$

bzw. aus

$$\|ax\|_p = \left(\sum_{i=1}^{n} |ax_i|^p\right)^{\frac{1}{p}} = (|a|^p)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} = |a| \cdot \|x\|_p.$$

Es bleibt die Dreiecksungleichung (welche in diesem Zusammenhang auch *Minkowskische Ungleichung* genannt wird) zu zeigen. Für deren Beweis werden wir die *Höldersche Ungleichung* verwenden: Sind $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$, so ist

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \|y\|_q.$$

Den Beweis der Hölderungleichung geben wir weiter unten. Sei $z = \left(|x_1 + y_1|^{p-1}, \dots, |x_n + y_n|^{p-1}\right)^T \in \mathbb{R}^n$. Sei $q \in (1, \infty)$ derart, dass $\frac{1}{p} + \frac{1}{q} = 1$. Zwangsweise ist dann $q = \frac{p}{p-1}$. Wir erhalten folgende Identität:

$$\begin{aligned} \|z\|_q &= \left(\sum_{i=1}^{n} \left(|x_i + y_i|^{p-1}\right)^q\right)^{\frac{1}{q}} = \left(\sum_{i=1}^{n} \left(|x_i + y_i|^{p-1}\right)^{\frac{p}{p-1}}\right)^{\frac{1}{q}} \\ &= \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{\frac{1}{q}} = \left(\left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{\frac{1}{p}}\right)^{\frac{p}{q}} = \|x + y\|_p^{p/q}. \end{aligned}$$

Diese werden wir später benutzen. Wegen der Dreiecksungleichung in $\mathbb{R}$ gilt

$$|x_i + y_i|^p = |x_i + y_i| \cdot |x_i + y_i|^{p-1} \leq (|x_i| + |y_i|)(|x_i + y_i|)^{p-1} = |x_i| z_i + |y_i| z_i.$$

Durch Aufsummieren erhalten wir die Ungleichung

$$\|x + y\|_p^p = \sum_{i=1}^{n} |x_i + y_i|^p \leq \sum_{i=1}^{n} |x_i| z_i + \sum_{i=1}^{n} |y_i| z_i. \tag{$*$}$$

Auf beide Summanden der rechten Seite wenden wir die Höldersche Ungleichung an und erhalten

$$\begin{aligned} \sum_{i=1}^{n} |x_i| z_i + \sum_{i=1}^{n} |y_i| z_i &\leq \|x\|_p \|z\|_q + \|y\|_p \|z\|_q = \|z\|_q (\|x\|_p + \|y\|_p) \\ &= \|x + y\|_p^{p/q} (\|x\|_p + \|y\|_p). \end{aligned}$$

In Verbindung mit Ungleichung $(*)$ erhalten wir also

$$\|x + y\|_p^p \leq \|x + y\|_p^{p/q} (\|x\|_p + \|y\|_p).$$

Wegen $\frac{1}{p} + \frac{1}{q} = 1$ ist $\frac{p}{q} = p - 1$. Division durch $\|x + y\|_p^{p-1}$ ergibt

$$\frac{\|x + y\|_p^p}{\|x + y\|_p^{p-1}} = \|x + y\|_p \leq \|x\|_p + \|y\|_p. \qquad \square$$

Satz 1.1.8 (Hölder-Ungleichung). *Sind $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$, so gilt für alle $x, y \in \mathbb{R}^n$ die Ungleichung*

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \|y\|_q.$$

Bevor wir den Beweis geben können, brauchen wir zwei Lemmas.

Lemma 1.1.9. *Sei $I \subset \mathbb{R}$ ein offenes Intervall und $f : I \to \mathbb{R}$ zweimal stetig differenzierbar. Es gelte $f''(x) \geq 0$ für alle $x \in I$. Dann ist f* konvex, *das heißt, für $x, y \in I$ und $t \in [0, 1]$ gilt $f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$.*

Beweis. Ohne Beschränkung der Allgemeinheit dürfen wir $x < y$ und $0 < t < 1$ annehmen. Weil $f'' \geq 0$ ist die Ableitung monoton steigend. Sei $x < z < y$. Nach dem Mittelwertsatz der Differentialrechnung gibt es $\xi \in (x, z)$ und $\eta \in (z, y)$ mit

$$\frac{f(z) - f(x)}{z - x} = f'(\xi) \leq f'(\eta) = \frac{f(y) - f(z)}{y - z}.$$

Umstellung dieser Ungleichung zeigt

$$(f(z) - f(x))(y - z) \leq (f(y) - f(z))(z - x) \Rightarrow f(z) \leq \frac{y - z}{y - x} f(x) + \frac{z - x}{y - x} f(y).$$

Substitution $t := \frac{y-z}{y-x}$ und $1 - t = \frac{z-x}{y-x}$, $tx + (1 - t)y = z$ zeigt das gewünschte Ergebnis, zumindest für $x < y$. □

Lemma 1.1.10 (Young-Ungleichung). *Für $a, b \in [0, \infty)$ und $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q}$ gilt*

$$ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q.$$

Beweis. Die Exponentialfunktion $f(x) = e^x$ erfüllt $f'' \geq 0$ und ist daher nach Lemma 1.1.9 konvex. Setze $x := p \log(a)$, $y := q \log(b)$, $t := \frac{1}{p}$. Dann liefert die Konvexitätsungleichung

$$ab = e^{\frac{1}{p} p \log(a) + \frac{1}{q} q \log(b)} \leq \frac{1}{p} e^{p \log(a)} + \frac{1}{q} e^{q \log(b)} \frac{1}{p} a^p + \frac{1}{q} b^q.$$

□

Beweis der Hölder-Ungleichung. Die Ungleichung ist trivial falls $x = 0$ oder $y = 0$. Falls $\|x\|_p = \|y\|_q = 1$, so sehen wir mit der Young-Ungleichung

$$\sum_{i=1}^{n} |x_i||y_i| \leq \sum_{i=1}^{n} \frac{1}{p} |x_i|^p + \sum_{i=1}^{n} \frac{1}{q} |y_i|^q = \frac{1}{p} \|x\|_p^p + \frac{1}{y} \|y\|_q^q = 1 = \|x\|_p \|y\|_q.$$

Im allgemeinen Fall setze $z := \frac{1}{\|x\|_p} x$, $w := \frac{1}{\|y\|_q} y$ und erhalte

$$\frac{1}{\|x\|_p \|y\|_q} \sum_{i=1}^{n} |x_i||y_i| = \sum_{i=1}^{n} |z_i||w_i| \leq \|z\|_p \|w\|_q = 1.$$

□

Definition 1.1.11. Sei V ein $\mathbb{R}$-Vektorraum. Zwei Normen $\|\cdot\|_1, \|\cdot\|_2$ auf V heißen *äquivalent*, falls $c_1, c_2 > 0$ existierten, sodass für alle $x \in V$ gilt:

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

Korollar 1.1.12. *Die Normenäquivalenz ist eine Äquivalenzrelation.*

Beweis. Wegen $\|x\| \leq \|x\| \leq \|x\|$ ist die Normenäquivalenz reflexiv. Die Symmetrie folgt aus $c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1 \Rightarrow \frac{1}{c_2}\|x\|_2 \leq \|x\|_1 \leq \frac{1}{c_1}\|x\|_2$. Für den Beweis der Transitivität gelte $c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$ sowie $c_3\|x\|_2 \leq \|x\|_3 \leq c_4\|x\|_2$. Dann folgt $c_3c_1\|x\|_1 \leq \|x\|_3 \leq c_4c_2\|x\|_1$. □

Satz 1.1.13. *Die* 1-,2- *und* ∞*-Norm auf* $\mathbb{R}^n$ *sind äquivalent und es gilt*

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty.$$

Beweis. Für $x \in \mathbb{R}^n$ existiert ein $1 \leq j \leq n$ mit $\|x\|_\infty = |x_j|$. Dann ist

$$\|x\|_\infty = |x_j| = \sqrt{|x_j|^2} \leq \sqrt{\sum_{i=1}^n |x_i|^2} = \|x\|_2,$$

da die Wurzelfunktion auf $\mathbb{R}$ streng monoton steigt. Aus der Subadditivität der 2-Norm folgt

$$\|x\|_2 = \left\|\sum_{i=1}^n x_i e_i\right\|_2 \leq \sum_{i=1}^n \|x_i e_i\|_2 = \sum_{i=1}^n |x_i| = \|x\|_1.$$

Zuletzt gilt offensichtlich die Ungleichung

$$\|x\|_1 = \sum_{i=1}^n |x_i| \leq \sum_{i=1}^n \max_{1\leq i\leq n} |x_i| = n\|x\|_\infty.$$ □

Später werden wir sehen, dass allgemein auf dem $\mathbb{R}^n$ alle Normen äquivalent sind.

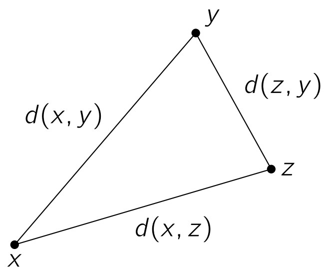
1.2 Metrische Räume

Definition 1.2.1 (Metrik). Sei X eine Menge. Eine *Metrik* auf X ist eine Abbildung $d\colon X\times X \to [0,\infty)$, $(x,y) \mapsto d(x,y)$, welche für alle $x, y, z \in X$ die folgenden Axiome erfüllt:

1. $d(x,y) = 0 \Leftrightarrow x = y$ (Definitheit).
2. $d(x,y) = d(y,x)$ (Symmetrie).
3. $d(x,y) \leq d(x,z) + d(z,y)$ (Dreiecksungleichung).

Das Paar (X, d) (bei Eindeutigkeit der Metrik auch nur kurz X) heißt *metrischer Raum*; der Wert $d(x,y)$ wird *Abstand* von x und y genannt.

Eine Metrik verallgemeintert den naiven Abstandsbegriff. Auch hier gilt, dass sich eine Metrik so verhält, wie wir es von einem Abstand erwarten würden, ganz eindrücklich zeigen das die Symmetrie oder die Subadditivität.



Die grafische Veranschaulichung der Dreiecksungleichung.

Proposition 1.2.2. *Ist $(V, \|\cdot\|)$ ein normierter Raum, so ist*

$$d(x, y) := \|x - y\|$$

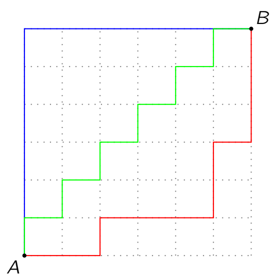
eine Metrik auf V.

Beweis. Wegen $d(x, y) = \|x - y\| = 0 \Leftrightarrow x - y = 0 \Leftrightarrow x = y$ ist d definit. Die Symmetrie ergibt sich aus Korollar 1.1.3, die Dreiecksungleichung folgt aus $d(x, y) = \|x - y\| = \|x - z + z - y\| \leq \|x - z\| + \|z - y\| = d(x, z) + d(z, y)$. □

Diese Metrik heißt (analog zu Definiton 1.1.4) die von $\|\cdot\|$ *induzierte* Metrik. Insgesamt gilt also: Ein Skalarprodukt auf einem Vektorraum V induziert eine Norm, welche wiederum eine Metrik induziert. Deswegen ist jeder euklidische Vektorraum auch ein normierter Raum; und jeder normierte Raum ist auch ein metrischer Raum. Wir vereinbaren für uns: Ist ab sofort X eine Teilmenge von $\mathbb{R}^n$ und nichts weiter angegeben, so ist X immer mit der durch das Standardskalarprodukt induzierten Metrik (dies ist also genau die Metrik, welche durch die euklidische Norm (= 2-Norm) induziert wird) versehen. Diese Metrik hat also die Gestalt

$$d(x, y) = \|x - y\|_2 = \sqrt{\langle x - y, x - y \rangle}.$$

Nebenbei: Die Metrik, welche durch die 1-Norm erzeugt wird, wird auch gelegentlich *Manhattan-* oder *Taxi-Metrik* genannt aus folgendem Grund: Das Straßennetzwerk des New Yorker Stadtteils Manhattan ist ähnlich einem Schachbrettmuster angeordnet. Die Strecke, die nun ein Taxi zwischen zwei Punkten A und B zurücklegt, ist genau die Summe der absoluten Koordinatendifferenzen, also genau $\|A - B\|_1$, wie die folgende Grafik verdeutlicht:



All die eingezeichneten Distanzen haben dieselbe Länge. Man kann sich vorstellen, dass unter verschiedenen Metriken auch sehr verschiedene Geometrien entstehen können. Dies soll uns aber nicht weiter beschäftigen.

1.3 Konvergenz

Da wir nun in allgemeinen metrischen Räumen von Abständen sprechen können, wollen wir den Begriff der Konvergenz, welcher ja maßgeblich auf dem Abstandsbegriff fußt, ebenfalls verallgemeinern.

Definition 1.3.1 (Konvergenz). Sei (X, d) ein metrischer Raum. Eine Abbildung $x\colon \mathbb{N} \to X$, $n \mapsto x(n) =: x_n$ fassen wir wieder als Folge ihrer Funktionswerte $(x_n)_{n\in\mathbb{N}}$ auf. Dann *konvergiert* die Folge $(x_n)_n$ gegen $y \in X$, wenn für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ existiert, sodass für alle $n \geq N$ gilt: $d(x_n, y) < \varepsilon$. Analog zu reellen Folgen übertragen sich die Begriffe *konvergente Folge*, *Divergenz*, *divergente Folge* auf metrische Räume.

Die Definition der Konvergenz überträgt sich also komplett analog zu den reellen Zahlen, bloß dass der Differenzbetrag $|x_n - y|$ durch die Metrik $d(x_n, y)$ ersetzt wird.

Satz 1.3.2. *Sei $X = \mathbb{R}^n$ mit der Metrik, die durch 1-, 2- oder ∞-Norm induziert wird. Sei $(x(k))_{k\in\mathbb{N}} = ((x_1(k), \dots, x_n(k))_k$ eine Folge in X. Sei $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Dann sind folgende Aussagen äquivalent:*

1. *Für alle $1 \leq i \leq n$ gilt $\lim_{k\to\infty} x_i(k) = x_i$.*
2. $\lim_{k\to\infty} \|x(k) - x\|_1 = 0$.
3. $\lim_{k\to\infty} \|x(k) - x\|_2 = 0$.
4. $\lim_{k\to\infty} \|x(k) - x\|_\infty = 0$.

Die Aussage dieses Satzes ist ziemlich intuitiv. Er besagt, dass eine Folge $(x(k))_k$ von Vektoren im $\mathbb{R}^n$ genau dann gegen einen Vektor $x \in \mathbb{R}^n$ (bezüglich der von der 1-, 2- oder ∞-Norm induzierten Metrik) konvergiert, wenn für alle $1 \leq i \leq n$ die Folge der i-ten Komponente von $x(k)$ gegen die i-te Komponente von x konvergiert. Dies ist sehr nützlich, weil damit die Konvergenz im $\mathbb{R}^n$ in n Konvergenzen in $\mathbb{R}$ zerfällt. Es folgt der Beweis:

Beweis. Der Beweis wird als Ringschluss $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$ geführt.

$1 \Rightarrow 2$. Es ist

$$0 \leq \|x(k) - x\|_1 = \sum_{i=1}^{n} \underbrace{|x_i(k) - x_i|}_{\to 0} \to 0.$$

$2 \Rightarrow 3 \Rightarrow 4$. Wir benutzen die Äquivalenz der betrachteten Normen. Wegen $0 \leq \|y\|_\infty \leq \|y\|_2 \leq \|y\|_1 \leq n\|y\|_\infty$ erhalten wir

$$\lim_{k\to\infty} \|x(k) - x\|_1 = 0 \Rightarrow \lim_{k\to\infty} \|x(k) - x\|_2 = 0 \Rightarrow \lim_{k\to\infty} \|x(k) - x\|_\infty = 0.$$

$4 \Rightarrow 1$. Nach Definition ist $\|x(k) - x\|_\infty = \max_{1\leq i\leq n} |x_i(k) - x_i|$. Also ist $|x_i(k) - x_i| \leq \|x(k) - x\|_\infty$ für alle i, damit gilt $|x_i(k) - x_i| \to 0$. □

Satz 1.3.3. *Seien (X, d) ein metrischer Raum. Dann gilt:*

1. *Für $v, w, x, y \in X$ gilt die* Vierecksungleichung $|d(v, w) - d(x, y)| \leq d(v, x) + d(w, y)$.
2. *Falls $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ gegen x, y konvergente Folgen in X sind, so gilt $\lim_{n\to\infty} d(x_n, y_n) = d(x, y)$.*
3. *(Eindeutigkeit des Grenzwertes) Konvergiert $(x_n)_n$ auch gegen $x' \in X$, so ist $x = x'$.*

Beweis. Zu 1. Es gilt $d(v, w) \leq d(v, x) + d(x, y) + d(y, w)$. Subtraktion von $d(x, y)$ ergibt die Ungleichung $d(v, w) - d(x, y) \leq d(v, x) + d(y, w) = d(v, x) + d(w, y)$. Aus der Ungleichung $d(x, y) \leq d(x, v) + d(v, w) + d(w, y)$ erhalten wir durch Subtraktion von $d(v, w)$ die Ungleichung $d(x, y) - d(v, w) \leq d(x, v) + d(w, y)$, welche mit der Symmetrie der Metrik und Multiplikation mit (-1) in die Ungleichung $-d(v, x) - d(w, y) \leq d(v, w) - d(x, y)$ übergeht. Insgesamt erhalten wir also $-(d(v, x) + d(w, y)) \leq d(v, w) - d(x, y) \leq d(v, x) + d(w, y)$, also $|d(v, w) - d(x, y)| \leq d(v, x) + d(w, y)$.

Zu 2. Es ist $0 \leq |d(x_n, y_n) - d(x, y)| \leq d(x_n, x) + d(y_n, y)$. Wegen $x_n \to x$, $y_n \to y$ ist $d(x_n, x) \to 0$, $d(y_n, y) \to 0$ und damit $d(x_n, x) + d(y_n, y) \to 0$. Mit dem Sandwich-Lemma folgt das Korollar.

Zu 3. Wie gezeigt ist $d(x, x') = \lim_{n\to\infty} d(x_n, x_n) = 0$, also $x = x'$. □

Definition 1.3.4 (Cauchy-Folge). Eine Folge $(x_n)_{n\in\mathbb{N}}$ in einem metrischen Raum (X, d) heißt *Cauchy-Folge*, wenn für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ existiert, sodass $d(x_n, x_m) < \varepsilon$ für alle $n, m \geq N$.

Korollar 1.3.5. *Ist (X, d) ein metrischer Raum und $(x_n)_{n\in\mathbb{N}}$ eine Folge in X, die gegen $x \in X$ konvergiert. Dann ist $(x_n)_n$ eine Cauchy-Folge.*

Beweis. Sei $\varepsilon > 0$. Wähle $N \in \mathbb{N}$ derart, dass $d(x, x_n) < \varepsilon/2$ für alle $n \geq N$. Für $n, m \geq N$ gilt dann $d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \varepsilon$. □

Definition 1.3.6 (Vollständigkeit). Ein metrischer Raum X heißt vollständig, wenn jede Cauchy-Folge in X gegen ein $x \in X$ konvergiert.

Beispiel. Der metrische Raum $\mathbb{R}$ mit der Abstandsfunktion ist vollständig. Dies ist gerade die Aussage des Cauchy'schen Konvergenzkriteriums aus der Analysis I. Das Intervall $(0, 1] \subset \mathbb{R}$ ist kein vollständiger metrischer Raum, denn die Cauchy-Folge $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$ besitzt keinen Grenzwert in $(0, 1]$.

Der euklidische Vektorraum $\mathbb{R}^n$ ist vollständig. Denn ist $(x(k))$ eine Cauchyfolge in $\mathbb{R}^n$, so ist jede Komponentenfolge $x_i(k)$ eine Cauchyfolge in $\mathbb{R}$ (denn $|x_i(k) - x_i(l)| \leq \|x(k) - x(l)\|_2$) und besitzt daher einen Grenzwert x_i in $\mathbb{R}$. Nach Satz 1.3.2 konvergiert dann $x(k)$ gegen x.

All diese Definitionen übertragen sich analog aus den eindimensionalen reellen Zahlen und bedürfen wohl keiner ausgiebigeren Erklärung.

1.4 Topologische Grundbegriffe

Wir geben nun eine Einführung in die Grundlagen der Topologie. Die Erkenntnisse, welche wir in diesem Unterkapitel gewinnen, werden wir für alle weiteren Kapitel benutzen. Die Topologie (gr. *topos* = Ort) verallgemeinert das Konzept der Nähe, sodass man darauf beruhende Begriffe der Analysis wie Konvergenz und Stetigkeit, in allgemeinerem Rahmen untersuchen kann. Dafür klären wir zuerst einige Basistermini.

Definition 1.4.1 (ε-Ball). Sei (X, d) ein metrischer Raum, $x \in X$ und $\varepsilon > 0$. Der *ε-Ball* in X um x ist definiert durch

$$B^X_\varepsilon(x) := B_\varepsilon(x) = \{y \in X \mid d(x, y) < \varepsilon\}.$$

Der ε-Ball ist die Verallgemeinerung der ε-Umgebung einer reellen Zahl. In Analysis 1 hatten wir zudem für Punktmengen die Terme *Berührpunkt* und *innerer Punkt* definiert, welche wir auch in unseren folgenden Arbeiten in allgemeinen metrischen Räumen benötigen.

Definition 1.4.2. Sei (X, d) ein metrischer Raum und $Y \subseteq X$. Ein Punkt $x \in X$ heißt

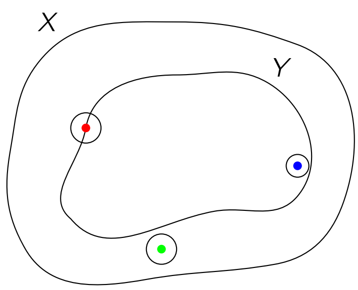
1. *Berührpunkt* von Y, wenn für alle $\varepsilon > 0$ ein $y \in Y$ existiert mit $d(x, y) < \varepsilon$. Äquivalent: Für alle $\varepsilon > 0$ ist $Y \cap B_\varepsilon^X(x) \neq \emptyset$. Die Menge der Berührpunkte von Y nennt man ihren *Abschluss* und schreibt $\overline{Y}$.
2. *innerer Punkt* von Y, falls ein $\varepsilon > 0$ existiert, sodass $B_\varepsilon^X(x) \subseteq Y$. In diesem Fall heißt Y eine *Umgebung* von x in X. Die Menge der inneren Punkte von Y nennt man ihr *Inneres* und schreibt Y°.

Beispiele.

1. Wir betrachten $X = (0, 1) \subset \mathbb{R}$. Dann ist $\overline{X} = X \cup \{0, 1\}$, denn für jedes $\varepsilon > 0$ existieren $a, b \in (0, 1)$ mit $a < \varepsilon$ und $b > 1 - \varepsilon$. Innere Punkte von X sind alle Punkte selbst, also $X^\circ = X$.
2. Wir betrachten $\mathbb{N}$ als Teilmenge von $\mathbb{R}$. Dann besitzt $\mathbb{N}$ keine inneren Punkte in $\mathbb{R}$, denn egal wie klein die Umgebung ist, welche wir um eine natürliche Zahl legen, sie enthält immer eine (sogar überabzählbar viele) reelle Zahl. Fassen wir jedoch $\mathbb{N}$ als Teilmenge von $\mathbb{Z}$ auf, so ist jede Zahl in $\mathbb{N}$ ein innerer Punkt.
3. In Analysis 1 haben wir die Erkenntnis gewonnen, dass $\mathbb{Q}$ dicht in $\mathbb{R}$ liegt. Diese Tatsache können wir topologisch durch $\overline{\mathbb{Q}} = \mathbb{R}$ ausdrücken, denn das heißt, dass in jeder Umgebung einer rellen Zahl liegt eine (sogar unendlich viele) rationale Zahl. Umgekehrt gilt aber auch, dass in der Umgebung jeder rationalen Zahl eine (sogar überabzählbar viele) irrationale Zahlen liegen. Also besitzt $\mathbb{Q}$ aufgefasst als Teilmenge von $\mathbb{R}$ keine inneren Punkte.

Bemerkung. Für eine Teilmenge $Y \subset X$ eines metrischen Raumes gilt offenbar

$$Y^\circ \subset Y \subset \overline{Y}.$$



Diese Abbildung zeigt einen metrischen Raum X und eine Teilmenge $Y \subset X$. Der rote Punkt ist ein Berührpunkt von Y, jedoch kein innerer Punkt, denn egal, wie klein wir eine Umgebung (dargestellt durch einen Kreis) des roten Punktes wählen, diese Umgebung liegt nie ganz in Y. Der blaue Punkt ist sowohl Berührpunkt als auch innerer Punkt von Y, der grüne weder noch. Man erkennt gut, dass es wichtig ist, anzugeben, in welchem Raum man einen Punkt betrachtet: So ist der rote Punkt ein innerer Punkt von X, der grüne ein innerer Punkt von X.

Definition 1.4.3. Eine Teilmenge Y eines metrischen Raumes (X, d) heißt

1. *offen in* X, wenn jeder Punkt in Y ein innerer Punkt ist, also: $Y = Y^\circ$.
2. *abgeschlossen in* X, wenn ihr Komplement $X \setminus Y$ offen in X ist.

Offene Teilmengen sind demnach genau jene, in welchen also jeder Punkt aussschließlich von anderen Punkten dieser Menge umgeben ist. Offenheit und Abgeschlossenheit hängen maßgeblich von der zugrunde liegenden Metrik ab. Man beachte an dieser Stelle übrigens, dass alle bisherigen Definitionen für den metrischen Raum $(\mathbb{R}, |\cdot|)$ exakt den Definitionen aus Analysis 1 entsprechen.

Warnung. Die Begriffe "offen" und "abgeschlossen" sind nur in Bezug auf einen umgebenden metrischen Raum definiert. Die Aussage "Y ist offen" ist nur dann sinnvoll, wenn ein metrischer Raum X angegeben wird, in dem Y offen sein soll (gleiches gilt für "abgeschlossen"). Wir gebrauchen die Sprechweisen "$Y \subset X$ offen", "Y ist eine offene Teilmenge von X".

Beispiele.

1. Offene Intervalle $(a, b) \subset \mathbb{R}$ sind offen in $\mathbb{R}$, denn: Für $x \in (a, b)$ wähle $\delta = \min\{x - a, b - x\} > 0$. Dann ist $(x - \delta, x + \delta) \subseteq (a, b)$. Ebenso sind abgeschlossene Intervalle $[a, b]$ abgeschlossen in $\mathbb{R}$, denn $\mathbb{R} \setminus [a, b] = (-\infty, a) \cup (b, \infty)$ ist offen in $\mathbb{R}$.
2. Das Intervall $(0, 1]$ ist weder offen noch abgeschlossen in $\mathbb{R}$, denn weder $(0, 1]$ noch das Komplement $(-\infty, 0] \cup (1, \infty)$ ist offen in $\mathbb{R}$.
3. Die Menge $\mathbb{N}$ ist nicht offen in $\mathbb{R}$, jedoch offen in $\mathbb{Z}$.

Korollar 1.4.4. *Sei (X, d) ein metrischer Raum, $x \in X, \varepsilon > 0$. Dann ist jeder ε-Ball $B_\varepsilon(x)$ offen in X.*

Beweis. Für $y \in B_\varepsilon(x)$ setze $\delta := \varepsilon - d(x, y) > 0$. Ist dann $z \in B_\delta(y)$, so ist $d(z, x) \leq d(z, y) + d(y, x) < \varepsilon - d(x, y) + d(y, x) = \varepsilon$, also $B_\delta(y) \subseteq B_\varepsilon(x)$. □

Satz 1.4.5. *Sei (X, d) ein metrischer Raum. Dann gilt:*

1. *Die leere Menge und X sind sowohl offen als auch abgeschlossen in X.*
2. *Endliche Durchschnitte offener Mengen sind offen.*
3. *Beliebige Vereinigungen offener Mengen sind offen.*
4. *Beliebige Durchschnitte abgeschlossener Mengen sind abgeschlossen.*
5. *Endliche Vereinigungen abgeschlossener Mengen sind abgeschlossen.*

Beweis. Zu 1. Die leere Menge und X selbst sind trivialerweise offen. Wegen $\emptyset = X \setminus X$ und $X = X \setminus \emptyset$ sind $\emptyset$ und X auch abgeschlossen.

Zu 2. Seien A_1, A_2 offen in X, $x \in A_1 \cap A_2$. Dann existieren $\varepsilon_1, \varepsilon_2 > 0$ mit $B_{\varepsilon_1}(x) \subseteq A_1$ und $B_{\varepsilon_2}(x) \subseteq A_2$. Für $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\} > 0$ ist dann $B_\varepsilon(x) \subseteq A_1 \cap A_2$ und der Durchschnitt zweier offener Mengen ist offen. Induktiv folgt die Aussage für beliebige endliche Durchschnitte.

Zu 3. Sei $(A_i)_{i \in I}$ eine durch I indizierte Familie offener Teilmengen von X. Sei ferner $x \in \bigcup_{i \in I} A_i$. Dann existiert ein Index $i \in I$ mit $x \in A_i$. Da A_i offen ist in X, ist x ein innerer Punkt von A_i, also auch von $\bigcup_{i \in I} A_i$.

Zu 4. Sei $(A_i)_{i \in I}$ eine durch I indizierte Familie abgeschlossener Teilmengen von X. Dann ist

$$X \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (X \setminus A_i).$$

Nun ist $\bigcup_{i \in I} X \setminus A_i$ nach Aussage 3 als Vereinigung offener Mengen offen und $\bigcap_{i \in I} A_i$ damit abgeschlossen in X.

Zu 5. Seien A_1, A_2 abgeschlossen in X. Dann ist

$$X \setminus (A_1 \cup A_2) = X \setminus A_1 \cap X \setminus A_2$$

als Durchschnitt zweier offener Mengen nach Aussage 2 offen in X und damit $A_1 \cap A_2$ abgeschlossen in X. Induktiv folgt die Aussage für beliebige endliche Durchschnitte. □

Bemerkung. Beliebige Durchschnitte offener Mengen sind nicht allgemein offen, betrachte als Gegenbeispiel das Intervall $I_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$, welches für alle n offen in $\mathbb{R}$ ist; jedoch ist $\bigcap_{n \in \mathbb{N}} I_n = \{0\}$ nicht offen in $\mathbb{R}$. Ebenso sind beliebige Vereinigungen abgeschlossener Mengen nicht allgemein abgeschlossen. Ein Gegenbeispiel ist $I_n = \left[-1 + \frac{1}{n}, 1 - \frac{1}{n}\right]$, welches für alle n abgeschlossen in $\mathbb{R}$ ist; jedoch ist $\bigcup_{n \in \mathbb{N}} I_n = (-1, 1)$ nicht abgeschlossen in $\mathbb{R}$.

Eine wichtige Folgerung dieses Satzes ist – wie man in Aussage 1 sieht –, dass Offenheit und Abgeschlossenheit keine Gegensätze sind! Mengen können weder offen noch abgeschlossen sein, genau eine der beiden Eigenschaften besitzen oder auch sowohl offen als auch abgeschlossen sein (solche Mengen heißen selten auch *abgeschloffen*). Wir geben nun ein alternatives (ebenfalls aus Analysis 1 bekanntes) Kriterium für Abgeschlossenheit an und klären kurz den Zusammenhang zwischen Abgeschlossenheit und Vollständigkeit.

Satz 1.4.6. *Sei (X, d) ein metrischer Raum und $Y \subseteq X$. Dann sind folgende Aussagen äquivalent:*

1. *Die Menge Y ist abgeschlossen in X.*
2. *Für jede Folge $(y_n)_{n\in\mathbb{N}}$ in Y, die in X konvergiert, ist $\lim_{n\to\infty} y_n \in Y$.*

Lemma 1.4.7. *Ein Punkt $y \in X$ ist genau dann ein Berührpunkt von Y, wenn es eine Folge $(y_n)_{n\in\mathbb{N}}$ in Y gibt, welche gegen y konvergiert.*

Beweis. »⇒«. Ist y ein Berührpunkt von Y, so existiert für jedes $\varepsilon > 0$ ein $y' \in B_\varepsilon(y)$. Wähle als n-tes Folgenglied y_n ein $y' \in B_{1/n}(y)$. Dann ist $(y_n)_{n\in\mathbb{N}}$ eine Folge in Y, die gegen y konvergiert.

»⇐«. Ist $\lim_{n\to\infty} y_n = y$, so existiert für alle $\varepsilon > 0$ ein $y_n \in Y$ mit $d(y, y_n) < \varepsilon$. Also ist y ein Berührpunkt von Y. □

Beweis von Satz 1.4.6. Wir zeigen »$\neg 1 \Leftrightarrow \neg 2$«. Sei Y nicht abgeschlossen. Dies ist äquivalent zu $X \setminus Y$ nicht offen. Dies ist nach Definition genau dann der Fall, wenn ein $y \in X \setminus Y$ existiert, welches kein innererer Punkt von $X \setminus Y$ ist. Genau dann gilt für alle $\varepsilon > 0$, dass $B_\varepsilon(y) \nsubseteq X \setminus Y$, also $B_\varepsilon(y) \cap Y \neq \emptyset$ für alle ε. Dies ist gleichbedeutend mit der Definition eines Berührpunktes, also ist $y \in \overline{Y}$ und nach Lemma 1.4.7 existiert genau in diesem Fall eine Folge $(y_n)_{n\in\mathbb{N}}$ in Y, welche gegen $y \notin Y$ konvergiert. Da wir nur über Äquivalenzen argumentiert haben, ist der Satz gezeigt. □

Korollar 1.4.8. *Sei X ein metrischer Raum. Dann ist $Y \subseteq X$ genau dann abgeschlossen in X, wenn $\overline{Y} = Y$.*

Beweis. Folgt unmittelbar aus Satz 1.4.6 und Lemma 1.4.7. □

Satz 1.4.9. *Sei X ein metrischer Raum. Ist $Y \subseteq X$ vollständig, so ist Y abgeschlossen in X.*

Beweis. Sei $(y_n)_{n\in\mathbb{N}}$ eine Folge in Y, die gegen ein $x \in X$ konvergiert. Wir zeigen die Abgeschlossenheit von Y, indem wir $x \in Y$ zeigen. Da $(y_n)_n$ konvergiert, ist sie eine Cauchy-Folge. Aufgrund der Vollständigkeit von Y konvergiert $(y_n)_n$ gegen ein $y \in Y$. Wegen der Eindeutigkeit des Grenzwertes ist $x = y \in Y$ und Y abgeschlossen in X. □

1.5 Kompaktheit

Wir betrachten nun eine weitere Eigenschaft metrischer Räume, welche folgendermaßen motiviert ist: In Analysis 1 haben wir den Satz von Bolzano-Weierstraß kennengelernt, welcher besagt, dass jede beschränkte Folge in $\mathbb{R}$ eine konvergente Teilfolge besitzt. Dieser Satz war in der Analysis 1 von fundamentaler Bedeutung, und es kann daher nicht überraschen, das wir zu einem Analogon in beliebigen metrischen Räumen gelangen wollen.

Satz 1.5.1 (Satz von Bolzano-Weierstraß in $\mathbb{R}^n$). *Jede beschränkte Folge in $\mathbb{R}^n$ (wobei $(x_n)_{n\in\mathbb{N}}$ beschränkt heißt, wenn $\|x_n\|_2 \leq C$ für ein $C \in \mathbb{R}$ und alle $n \in \mathbb{N}$) besitzt eine konvergente Teilfolge.*

Beweis. Wir führen den Beweis auf den Satz von Bolzano-Weierstraß in $\mathbb{R}$ zurück. Sei $(x_k)_{k\in\mathbb{N}}$ eine Folge von Vektoren in $\mathbb{R}^n$ mit $\|x_k\|_2 \leq C$ für alle k. Dann ist die Folge der i-ten Komponente für alle $1 \leq i \leq n$ beschränkt, wie man sich leicht überlegt. Nach Bolzano-Weierstraß existiert eine Teilfolge von $(x_k)_k$, sodass die Folge der ersten Komponente konvergiert. Wähle aus dieser Teilfolge eine Teilfolge, sodass auch die Folge der zweiten Komponente konvergiert. So erhalten wir schrittweise eine Teilfolge von $(x_k)_k$, in der alle Komponentenfolgen konvergieren und damit auch $(x_{k_m})_m$. □

Definition 1.5.2 (Folgenkompaktheit). Ein metrischer Raum X heißt folgenkompakt, falls jede Folge in X eine in X konvergente Teilfolge besitzt.

Anhand der Definition kann man bereits erkennen: Im Unterschied zu Offenheit und Abgeschlossenheit hängt die Kompaktheit eines metrischen Raumes nicht davon ab, ob der Raum in einen umgebenden (und gegebenenfalls in welchen) metrischen Raum eingebettet ist.

Beispiele.

1. Abgeschlossene reelle Intervalle sind folgenkompakt.
2. Die *Einheitssphäre* $S^{n-1} := \{x \in \mathbb{R}^n \mid \|x\|_2 = 1\}$ ist folgenkompakt.
3. $\mathbb{R}$ ist nicht folgenkompakt, betrachte die Folge $(n)_{n\in\mathbb{N}}$.
4. Die Menge $\mathbb{Q}\cap[0,1] \subseteq \mathbb{R}$ ist nicht folgenkompakt, denn für jede irrationale Zahl $x \in [0,1]\setminus\mathbb{Q}$ existiert eine Folge in $\mathbb{Q}\cap[0,1]$, die gegen x konvergiert und damit keine in $\mathbb{Q}\cap[0,1]$ konvergente Teilfolge besitzt.

Wir werden nun die so genannte *Überdeckungskompaktheit* definieren. Dies ist eine andere Art von Kompaktheit, welche tendenziell abstrakter und eher topologischer Natur ist als die Folgenkompaktheit. Wir werden aber sehen, dass diese Definition äquivalent zur Folgenkompaktheit ist. Ist dies gezeigt, so werden wir ab diesem Zeitpunkt nur noch von »kompakten« Mengen sprechen. Die wichtigste Erkenntnis dieses Unterkapitels wird jedoch der Satz von Heine-Borel sein, welcher uns ein einfaches Kriterium für die Kompaktheit einer Teilmenge des $\mathbb{R}^n$ an die Hand gibt. Zunächst aber wollen wir Überdeckungskompaktheit definieren; dazu bedarf es einiger Vorarbeit.

Definition 1.5.3 (Beschränktheit). Sei X ein metrischer Raum. Eine Teilmenge $Y \subseteq X$ heißt *beschränkt*, wenn $x \in X, \varepsilon > 0$ existieren mit $X \subseteq B_\varepsilon(x)$.

Beschränkte metrische Räume sind also jene, welche sich in einem endlich großen Ball einschließen lassen.

Definition 1.5.4 (Offene Überdeckung). Sei X ein metrischer Raum. Eine *offene Überdeckung* von X ist eine Familie $(U_i)_{i\in I}$ offener Teilmengen von X, sodass

$$\bigcup_{i\in I} U_i = X.$$

Eine offene Überdeckung heißt endlich, falls die Indexmenge endlich ist.

Beispiele.

1. Sei $X = [1,10] \subset \mathbb{R}$. Dann ist $((n-\frac{3}{2}, n+\frac{1}{2}))_{1\leq n\leq 11}$ eine endliche offene Überdeckung von X.
2. Sei $X = \mathbb{R}$. Dann ist $((-n,n))_{n\in\mathbb{N}}$ eine offene Überdeckung von X.

Definition 1.5.5 (Teilüberdeckung). Sei X ein metrischer Raum und $U = (U_i)_{i\in I}$ eine offene Überdeckung von X. Eine *Teilüberdeckung* von U ist eine offene Überdeckung $(V_j)_{j\in J}$ von X, sodass für jedes $j \in J$ ein $i \in I$ existiert mit $V_j = U_i$.

Definition 1.5.6 (Überdeckungskompaktheit). Ein metrischer Raum X heißt *überdeckungskompakt*, falls jede offene Überdeckung von X eine endliche Teilüberdeckung besitzt.

Wir zeigen nun die Äquivalenz der beiden Kompaktheitsbegriffe.

Satz 1.5.7. *Ein metrischer Raum X ist genau dann folgenkompakt, wenn er überdeckungskompakt ist.*

Für den Beweis benutzen wir zwei Lemmata.

Lemma 1.5.8 (Lebesguesches Lemma). *Sei X ein folgenkompakter metrischer Raum und $(U_i)_{i\in I}$ eine offene Überdeckung von X. Dann existiert ein $\delta > 0$, sodass für alle $x \in X$ ein $i \in I$ existiert mit $B_\delta(x) \subseteq U_i$.*

Beweis. Der Beweis wird als Widerspruchsbeweis geführt. Sei also das Lebesguesche Lemma falsch. Dann existiert ein folgenkompakter metrischer Raum X mit einer offenen Überdeckung $(U_i)_{i\in I}$, welche keine Lebesgue-Zahl besitzt. Dann existiert für alle $\delta > 0$ ein $x \in X$, sodass $B_\delta(x)$ in keiner der Mengen U_i enthalten ist. Insbesondere können wir für δ die Folge $(1/n)_{n\in\mathbb{N}}$ betrachten und erhalten die Aussage: Für alle $n \in \mathbb{N}$ existiert ein $x_n \in X$, sodass $B_{1/n}(x_n)$ in keinem U_i enthalten ist. Da X folgenkompakt ist, gibt es eine konvergente Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$, sodass $B_{1/n_k}(x_{n_k})$ in keinem U_i enthalten ist. Sei $x := \lim_{k\to\infty} x_{n_k}$. Da $\bigcup_{i\in I} U_i = X$, existiert ein i' mit $x \in U_{i'}$. Diese Menge $U_{i'}$ ist jedoch offen in X. Damit existiert ein $\varepsilon > 0$, sodass $B_\varepsilon(x) \subseteq U_{i'}$. Dies ist ein Widerspruch, denn: Wähle ein k, sodass $1/n_k < \varepsilon/2$ und $d(x, x_{n_k}) < \varepsilon/2$. Daraus folgt, dass $B_{1/n_k}(x_{n_k}) \subseteq B_\varepsilon(x) \subseteq U_{i'}$, denn: Ist $d(y, x_{n_k}) < 1/n_k$, so ist $d(x, x_{n_k}) \leq d(y, x_{n_k}) + d(x_{n_k}, x) < \frac{1}{n_k} + \frac{\varepsilon}{2} < \varepsilon$ und insgesamt also das Lebesguesche Lemma richtig. ☐

Bemerkung. Ein solches $\delta > 0$ heißt *Lebesgue-Zahl* der Überdeckung $(U_i)_i$.

Lemma 1.5.9. *Sei X ein folgenkompakter metrischer Raum. Dann ist X* total beschränkt, *das heißt für alle $\varepsilon > 0$ existieren endlich viele Punkte $x_1, \ldots, x_n \in X$, sodass $\bigcup_{i=1}^n B_\varepsilon(x_i) = X$.*

Beweis. Wir zeigen, dass ein nicht total beschränkter Raum X nicht folgenkompakt ist. Dann existiert ein $\varepsilon > 0$, sodass für beliebig (aber endlich) viele Punkte $x_1, \ldots, x_n$ gilt: $\bigcup_{i=1}^n B_\varepsilon(x_i) \neq X$. Wir konstruieren nun eine Folge, welche keine konvergente Teilfolge in X besitzt. Wähle $x_1 \in X$ derart, dass $X \setminus B_\varepsilon(x_1) \neq \emptyset$. Wähle $x_2 \in X \setminus B_\varepsilon(x_1)$, sodass $X \setminus (B_\varepsilon(x_1) \cup B_\varepsilon(x_2)) \neq \emptyset$. Allgemein wähle $x_n \in X \setminus \bigcup_{i=1}^{n-1} B_\varepsilon(x_i)$. Für jedes $n \in \mathbb{N}$ finden wir ein $x_n \in X$, da X nicht total beschränkt ist. Dann ist $(x_n)_n$ eine Folge in X mit $d(x_n, x_m) \geq \varepsilon$, denn ist $n > m$, so ist nach Konstruktion $x_n \notin B_\varepsilon(x_m)$. Keine Teilfolge von $(x_n)_n$ kann damit eine Cauchy-Folge sein, also nicht konvergieren. Es folgt also, dass X nicht folgenkompakt ist. ☐

Beweis von Satz 1.5.7. »⇒«. Sei X folgenkompakt und $(U_i)_{i\in I}$ eine offene Überdeckung von X. Zu zeigen ist: Es gibt eine endliche Teilüberdeckung $(V_j)_{j\in J}$. Sei $\delta > 0$ eine Lebesgue-Zahl von $(U_i)_i$, diese existiert nach dem Lebesgueschen Lemma. Da X als folgenkompakter Raum total beschränkt ist, existieren $x_1, \ldots, x_r \in X$, sodass $\bigcup_{j=1}^r B_\delta(x_j) = X$. Für alle $1 \leq j \leq n$ existiert damit ein $i_j \in I$, sodass $B_\delta(x_j) \subseteq U_{i_j}$. Mit $B_\delta(x_1) \cup \cdots \cup B_\delta(x_r) = X$ folgt $U_{i_1} \cup \cdots \cup U_{i_r} = X$.

»⇐«. Der Beweis wird als Widerspruchsbeweis geführt. Sei X überdeckungskompakt, jedoch nicht folgenkompakt und $(x_n)_{n\in\mathbb{N}}$ eine Folge in X, welche keine konvergente Teilfolge besitzt. Dann ist kein Punkt

$x \in X$ Grenzwert einer Teilfolge von $(x_n)_n$, das heißt für jedes x existiert ein $\varepsilon_x > 0$, sodass nur endlich viele Folgenglieder $x_n \in B_{\varepsilon_x}(x)$ liegen. Dann ist $U = (B_{\varepsilon_x}(x))_{x \in X}$ eine offene Überdeckung von X. Da X nach Voraussetzung überdeckungskompakt ist, gibt es eine endliche Teilüberdeckung $\left(B_{\varepsilon_{y_i}}(y_i)\right)_{1 \leq i \leq n}$ mit $X = \bigcup_{i=1}^{n} B_{\varepsilon_{y_i}}(y_i)$. Da die Folge $(x_n)_n$ unendlich viele Folgenglieder besitzt, wir X aber in endlich viele Teilmengen zerlegt haben, müssen in mindestens einem der ε-Bälle unendlich viele Folgenglieder x_n liegen im Widerspruch zur Konstruktion dieser ε-Bälle. Also muss X folgenkompakt sein. □

Da also wie gezeigt Folgenkompaktheit und Überdeckungskompaktheit äquivalent sind, werden wir ab diesem Zeitpunkt lediglich von *kompakten* Mengen sprechen. Für unsere späteren Untersuchungen, insbesondere die mehrdimensionale Differentialrechnung, ist der $\mathbb{R}^n$ von besonderer Bedeutung. Aus diesem Grund wollen wir ein einfaches Kriterium für die Kompaktheit einer Teilmenge des $\mathbb{R}^n$. Dieses liefert der Satz von Heine-Borel:

Satz 1.5.10 (Satz von Heine-Borel). *Eine Teilmenge $X \subseteq \mathbb{R}^n$ ist genau dann kompakt, wenn sie beschränkt und abgeschlossen in $\mathbb{R}^n$ ist.*

Für den Beweis nutzen wir als Kompaktheit die Folgenkompaktheit.

Beweis. »$\Rightarrow$«. Sei X folgenkompakt. Nach Satz 1.5.11 und Satz 1.4.9 ist X abgeschlossen in $\mathbb{R}^n$. Wäre X unbeschränkt, so gäbe es zu jedem $n \in \mathbb{N}$ ein $x_n \in X$ mit $\|x_n\|_2 \geq n$. Die Folge $(x_n)_n$ kann dann keine konvergente Teilfolge besitzen, womit X nicht folgenkompakt wäre. Also ist X beschränkt.

»$\Leftarrow$«. Sei $X \subseteq \mathbb{R}^n$ beschränkt und abgeschlossen in $\mathbb{R}^n$. Eine Folge $(x_n)_{n \in \mathbb{N}}$ in X ist damit beschränkt (dies sieht man leicht, da jede Komponentenfolge beschränkt sein muss). Nach dem Satz von Bolzano-Weierstraß existiert damit eine konvergente Teilfolge $(x_{n_k})_{k \in \mathbb{N}}$, die gegen ein $x \in \mathbb{R}^n$ konvergiert. Da X abgeschlossen in $\mathbb{R}^n$ ist, folgt mit $x \in X$ die Folgenkompaktheit. □

Mit diesem Kriterium sieht man nun sofort, dass beispielsweise die Einheitsspähre S^{n-1} kompakt ist und $\mathbb{R}$ wegen der Unbeschränktheit nicht.

Satz 1.5.11. *Ein kompakter metrischer Raum ist vollständig.*

Beweis. Sei $(x_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge in X. Da X folgenkompakt ist, existiert eine konvergente Teilfolge $x_{n_k} \to x \in X$. Wir zeigen $x_n \to x$. Sei $\varepsilon > 0$. Wähle $N \in \mathbb{N}$ derart, dass $d(x_n, d_m) < \varepsilon/2$ für $n, m \geq N$. Wähle $k \in \mathbb{N}$ mit $n_k \geq N$. Für $n \geq n_k$ gilt dann $d(x, x_n) \leq d(x, x_{n_k}) + d(x_{n_k}, x_n) < \varepsilon$. □

Dieser Satz hat Ähnlichkeit zu Satz 1.4.9. Hier müssen wir im Unterschied nicht von einer Teilmenge Y eines metrischen Raumes sprechen, da Folgenkompaktheit und Vollständigkeit nicht von umgebenden metrischen Räumen abhängen. Mit diesen beiden Sätzen erhalten wir die folgenden Abhängigkeiten:

$$\boxed{\text{Abgeschlossenheit}} \xrightarrow{\text{nötig für}} \boxed{\text{Vollständigkeit}} \xrightarrow{\text{nötig für}} \boxed{\text{Kompaktheit}}$$

Wir wollen an dieser Stelle kurz die Inhalte des ersten Unterkapitels Revue passieren lassen. In diesem Unterkapitel haben wir Skalarprodukt, Norm und Metrik kennengelernt, mit welchen wir den Begriff der Länge und des Abstandes verallgemeinern können. Mit diesen Verallgemeinerungen konnten wir in metrischen Räumen von Offenheit und Abgeschlossenheit sprechen. Eine Vorstellung dieser Begriffe zu bekommen, war wesentliches Ziel dieses Unterkapitels, denn mit all diesen Begriffe werden wir in den nächsten Unterkapiteln hantieren müssen. Bislang haben wir jedoch noch keine Erkenntnisse über Funktionen gewonnen, dies beginnt mit dem nun folgenden Unterkapitel.

2 Stetigkeit in metrischen Räumen

Wir beginnen nun, über Funktionen zu sprechen und unsere Ergebnisse aus Analysis 1 zu verallgemeinern. Erstes Ziel wird es sein, eine adäquate Definition der Stetigkeit zu finden und wesentliche Ergebnisse aus den eindimensionalen reellen Zahlen in metrische Räume zu übertragen.

2.1 Gewöhnliche Stetigkeit

Definition 2.1.1 (Stetigkeit). Seien $(X, d_X), (Y, d_Y)$ metrische Räume und $x \in X$. Eine Abbildung $f\colon X \to Y$ heißt

1. *ε-δ-stetig* in x, wenn für alle $\varepsilon > 0$ ein $\delta > 0$ existiert, sodass für alle $x' \in X$ mit $d_X(x, x') < \delta$ gilt: $d_Y(f(x), f(x')) < \varepsilon$.
2. *folgenstetig* in x, wenn für jede gegen x konvergente Folge $(x_n)_{n\in\mathbb{N}}$ in X die Folge $(f(x_n))_n$ gegen $f(x)$ konvergiert.

Wir erinnern an dieser Stelle an die Konvergenzdefinition. Diese hatte auffallende Ähnlichkeit zu der Konvergenzdefinitionen reeller Folgen, bloß wurde der Differenzbetrag durch die Metrik ersetzt. Ähnliches finden wir hier vor: Die beiden Stetigkeitsdefinitionen reeller Funktionen übertragen sich ebenfalls analog, bloß dass auch hier der Abstand von Punkten bzw. Funktionswerten durch die Metrik des entsprechenden Raumes angegeben wird. In Analysis 1 haben wir gezeigt, dass die beiden Definitionen äquivalent sind. Der folgende Satz zeigt, dass dies auch in metrischen Räumen gilt. Darüber hinaus liefert er noch zwei äquivalente Kriterien für die Stetigkeit.

Satz 2.1.2. *Seien $(X, d_X), (Y, d_Y)$ metrische Räume und $f\colon X \to Y$ eine Abbildung. Dann sind folgende Aussagen äquivalent:*

1. *Die Abbildung f ist folgenstetig.*
2. *Ist $A \subseteq Y$ abgeschlossen in Y, so ist $f^{-1}(A)$ abgeschlossen in X.*
3. *Ist $O \subseteq Y$ offen in Y, so ist $f^{-1}(O)$ offen in X.*
4. *Die Abbildung f ist ε-δ-stetig.*

Beweis. Der Beweis wird als Ringschluss $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$ geführt.

»$1 \Rightarrow 2$«. Sei f folgenstetig und $A \subseteq Y$ abgeschlossen in Y. Wir zeigen: Ist $(x_n)_{n\in\mathbb{N}}$ eine Folge in $f^{-1}(A)$, die gegen $x \in X$ konvergiert, dann ist $x \in f^{-1}(A)$. Da $(x_n)_n$ gegen x konvergiert, konvergiert die Folge $(f(x_n))_n$ in A gegen $f(x) \in Y$. Da A abgeschlossen ist, ist $f(x) \in A$, also $x \in f^{-1}(A)$.

»$2 \Rightarrow 3$«. Ist $O \subseteq Y$ offen, dann ist $Y \setminus O \subseteq Y$ abgeschlossen in Y. Nach Bedingung 2 ist dann $f^{-1}(Y \setminus O) = X \setminus f^{-1}(O)$ abgeschlossen und damit $f^{-1}(O)$ offen in X.

»$3 \Rightarrow 4$«. Sei $x \in X, \varepsilon > 0$. Es ist $B_\varepsilon^Y(f(x)) \subseteq Y$ offen in Y. Nach Bedingung 3 ist dann $f^{-1}\left(B_\varepsilon^Y(f(x))\right) \subseteq X$ offen in X. Wegen $x \in f^{-1}\left(B_\varepsilon^Y(f(x))\right)$ existiert ein $\delta > 0$ mit $B_\delta^X(x) \subseteq f^{-1}\left(B_\varepsilon^Y(f(x))\right)$. Ist also $x' \in B_\delta^X(x)$, so ist $f(x') \in B_\varepsilon^Y(f(x))$, das heißt aus $d(x, x') < \delta$ folgt $d(f(x), f(x')) < \varepsilon$.

»$4 \Rightarrow 1$«. Sei f nun ε-δ-stetig und $(x_n)_{n\in\mathbb{N}}$ eine Folge in X, die gegen $x \in X$ konvergiert. Sei $\varepsilon > 0$. Dann existiert ein $\delta > 0$, sodass $d(f(x), f(x')) < \varepsilon$ für $x' \in X$ mit $d(x, x') < \delta$. Wegen $x_n \to x$ existiert ein $N \in \mathbb{N}$

mit $d(x, x_n) < \delta$ für $n \geq N$. Dann ist $d(f(x), f(x_n)) < \varepsilon$ für $n \geq N$ und $f(x_n) \to f(x)$. □

Aufgrund der Äquivalenz von ε-δ- und Folgenstetigkeit werden wir ab nun lediglich von *stetigen* Funktionen sprechen. Wichtig bei diesem Satz ist, dass die *Urbilder* offener bzw. abgeschlossener Mengen wieder offen bzw. abgeschlossen sind; für die Bilder gilt dies nicht allgemein! Genau umgekehrt verhält es sich mit der Kompaktheit, wie der folgende Satz zeigt.

Satz 2.1.3. *Seien $(X, d_X), (Y, d_Y)$ metrische Räume, $f: X \to Y$ stetig. Ist X kompakt, dann ist auch $f(X) \subseteq Y$ kompakt.*

Beweis. Sei $(y_n)_{n\in\mathbb{N}}$ eine Folge in $f(X)$. Schreibe $y_n = f(x_n)$ für alle n. Die so konstruierte Folge $(x_n)_n$ in X besitzt nach Voraussetzung eine konvergente Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$, deren Grenzwert $x \in X$ ist. Da f stetig, insbesondere folgenstetig, ist, konvergiert die Folge $(f(x_{n_k}))_k$ gegen $f(x) \in f(X)$. Nun ist $f(x_{n_k}) = y_{n_k}$. Dann ist (y_{n_k}) eine konvergente Teilfolge von $(y_n)_n$ mit $\lim_{k\to\infty} y_{n_k} \in f(X)$. □

Wir werden an dieser Stelle keine Beispiele diskutieren, da wir nach der gewöhnlichen Stetigkeit noch zwei verstärkte Stetigkeitsbegriffe definieren und erst danach die betrachteten Beispiele direkt auf die verstärkten Stetigkeiten prüfen. Wir fahren nun fort mit der Übertragung grundlegender Ergebnisse über Stetigkeit aus Analysis 1.

Satz 2.1.4. *Seien X ein metrischer Raum und $f, g: X \to \mathbb{R}$ Abbildungen. Sind f, g stetig, dann sind auch $f + g, f - g, fg$ und $\frac{f}{g}$ (falls $0 \notin g(X)$) stetig.*

Für den Beweis nutzen wir zwei Lemmata.

Lemma 2.1.5. *Die Funktionen $a, s, m: \mathbb{R}^2 \to \mathbb{R}, d: \mathbb{R} \times \mathbb{R} \setminus \{0\} \to \mathbb{R}$ definiert durch $a(x, y) := x + y, m(x, y) := xy, s(x, y) := x - y, d(x, y) := \frac{x}{y}$ sind stetig.*

Beweis. Wir beweisen die Aussage gleichzeitig für alle Abbildungen. Wir schreiben $D_K \subseteq \mathbb{R}^2$ für Definitionsbereich der Abbildung $K = a, s, m, d$. Sei nun $(x_n, y_n)_{n\in\mathbb{N}}$ eine Folge in D_K, die gegen $(x, y) \in D_K$ konvergiert. Dann ist $\lim_{n\to\infty} x_n = x, \lim_{n\to\infty} y_n = y$ und nach den Grenzwertsätzen für reelle Folgen ist $\lim_{n\to\infty} K(x_n, y_n) = K(x, y)$. Damit ist K stetig auf D_K. □

Lemma 2.1.6. *Seien X, Y, Z metrische Räume und $f: X \to Y, g: Y \to Z$ Abbildungen. Ist f stetig in X und g stetig in $f(x)$, so ist $g \circ f$ stetig in x.*

Beweis. Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in X, welche gegen $x \in X$ konvergiert. Da f stetig in x ist, konvergiert die Folge $(f(x_n))_n$ gegen $f(x)$. Wegen der Stetigkeit von g in $f(x)$ konvergiert dann auch $(g(f(x_n)))_n = ((g \circ f)(x_n))_n$ gegen $g(f(x))$ und folglich ist $g \circ f$ stetig in x. □

Beweis von Satz 2.1.4. Wir betrachten $(f \mid g): X \to \mathbb{R}^2, x \mapsto (f(x), g(x))$. Diese Abbildung ist stetig, denn: Seien $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ Folgen in X, die gegen $x, y \in X$ konvergieren. Wegen der Stetigkeit von f, g konvergieren $(f(x_n))_n, (g(x_n))_n$ gegen $f(x), g(x)$. Damit konvergiert auch $(f(x_n), g(x_n))_n$ gegen $(f(x), g(x))$ und $f \mid g$ ist stetig auf X. Mit Lemma 2.1.5 und Lemma 2.1.6 folgt, dass die Komposition $K \circ (f \mid g) = (fKg)$ stetig auf X ist. □

Satz 2.1.7 (Satz vom Minimum und Maximum). *Sei X ein kompakter metrischer Raum, $f: X \to \mathbb{R}$ stetig.*

Dann existieren $x_{\min}, x_{\max} \in X$, sodass $f(x_{\min}) \leq f(x) \leq f(x_{\max})$ für alle $x \in X$.

Beweis. Da X kompakt und f stetig ist, ist $f(X) \subseteq \mathbb{R}$ kompakt, also abgeschlossen und beschränkt ist. Definiere

$$c_{\inf} := \inf\{f(x) \mid x \in X\} = \inf f(X) \in \mathbb{R},$$
$$c_{\sup} := \sup\{f(x) \mid x \in X\} = \sup f(X) \in \mathbb{R}.$$

Aus Analysis 1 wissen wir, dass $c_{\inf}, c_{\sup}$ Berührpunkte von $f(X)$ sind. Da $f(X)$ abgeschlossen ist, sind $c_{\inf}, c_{\sup} \in f(X)$. Wähle $x_{\min}, x_{\max} \in X$ mit $f(x_{\min}) = c_{\inf}, f(x_{\max}) = c_{\sup}$. Dann ist $f(x_{\min}) \leq f(x) \leq f(x_{\max})$. □

2.2 Gleichmäßige und Lipschitz-Stetigkeit

Wir betrachten nun noch zwei Verschärfungen der Stetigkeit.

Definition 2.2.1. Seien $(X, d_X), (Y, d_Y)$ metrische Räume. Eine Funktion $f\colon X \to Y$ heißt *gleichmäßig stetig*, wenn für alle $\varepsilon > 0$ ein $\delta > 0$ existiert, sodass $d_Y(f(x), f(x')) < \varepsilon$ für alle $x, x' \in X$ mit $d_X(x, x') < \delta$.

Diese Definition ähnelt der Definition der gewöhnlichen Stetigkeit sehr. Der Unterschied ist, dass das gesuchte δ nur vom vorgegebenen ε, nicht aber vom betrachteten Punkt x abhängen darf. Zu einem ε muss man demnach eine Art *Universal-*δ finden, sodass $d_Y(f(x) - f(x')) < \varepsilon$ für alle $x, x' \in X$ mit $d_X(x, x') < \delta$ gilt. Ein Standardgegenbeispiel ist die Normalparabel $x \mapsto x^2$. Diese Funktion ist auf ganz $\mathbb{R}$ stetig, jedoch nicht gleichmäßig stetig; dies kann man sich geometrisch überlegen: Seien $\delta > 0$ fest und $x, x' \in \mathbb{R}$ mit $|x - x'| < \delta$. Werden nun x, x' sehr groß, so strebt die Differenz der Funktionswerte gegen ∞, damit kann kein δ ein gefordertes Universal-δ für ein vorgegebenes ε sein und x^2 ist nicht gleichmäßig stetig. Es gilt aber:

Proposition 2.2.2. *Seien $(X, d_X), (Y, d_Y)$ metrische Räume und $f\colon X \to Y$ stetig. Ist X kompakt, dann ist f gleichmäßig stetig.*

Beweis. Sei $\varepsilon > 0$. Für einen Punkt $x \in X$ betrachten wir die Menge

$$U_x = \{x' \in X \mid d_Y(f(x) - f(x')) < \varepsilon/2\} = f^{-1}(B_{\varepsilon/2}(f(x))).$$

Da f stetig ist, ist diese Menge offen in X und damit $(U_x)_{x \in X}$ eine offene Überdeckung von X. Nach dem Lebesgueschen Lemma existiert damit ein $\delta > 0$, sodass für alle $z \in X$ ein x existiert mit $B_\delta(z) \subseteq U_x$. Für $z' \in X$ gilt dann: Ist $d(z, z') < \delta$, so sind $z, z' \in B_\delta(z) \subseteq U_x$ und es folgt mit

$$|f(z) - f(z')| \leq |f(z) - f(x)| + |f(x) - f(z')| < \varepsilon$$

die gleichmäßige Stetigkeit von f. □

Die folgende Stetigkeit ist eine Veschärfung der gleichmäßigen Stetigkeit.

Definition 2.2.3 (Lipschitz-Stetigkeit). Eine Abbildung $f\colon X \to Y$ zwischen metrischen Räumen $(X, d_X), (Y, d_Y)$ heißt *Lipschitz-stetig*, wenn ein $L \geq 0$ existiert, sodass für alle $x, x' \in X$ gilt: $d_Y(f(x), f(x')) \leq L d_X(x, x')$. Die Zahl L heißt *Lipschitz-Konstante* der Funktion f.

Wir überlegen zunächst, dass jede Lipschitz-stetige Funktion gleichmäßig stetig ist. Für ein $\varepsilon > 0$ wählen wir $\delta > 0$ derart, dass $\delta L \leq \varepsilon$. Ist dann $d_X(x, x') < \delta$, so ist $d_Y(f(x), f(x')) \leq L d_X(x, x') < L\delta \leq \varepsilon$ und f folglich gleichmäßig stetig. Nun interpretieren wir die Lipschitz-Stetigkeit geometrisch anhand einer reellen Funktion f. In diesem Fall gilt die Ungleichung $|f(x) - f(x')| \leq L|x - x'|$. Stellen wir diese um, so erhalten wir $\frac{|f(x)-f(x')|}{|x-x'|} \leq L$. Die linke Seite der Gleichung ist, wie wir erkennen, die Steigung der Sekante zwischen den Punkten x, x'. Eine Lipschitz-stetige reelle Funktion ist also genau eine Funktion, deren Steigung überall durch L beschränkt ist. Aus diesem Grund ist beispielsweise die Wurzelfunktion $x \mapsto \sqrt{x}$ nicht Lipschitz-stetig, denn für x, x' sehr nahe bei 0 strebt die Sekantensteigung gegen ∞. Die Wurzelfunktion ist jedoch gleichmäßig stetig, wie man sich geometrisch überlegen kann.

2.3 Beispiele stetiger Funktionen

Wir diskutieren nun einige Beispiele. Zunächst prüfen wir die drei Abbildungen, welche wir ganz am Anfang dieses Kapitels kennengelernt haben: Skalarprodukt, Norm und Metrik.

Wir beginnen mit der Norm, da wir diese Erkenntnis für die Prüfung des Skalarprodukts verwenden. Sei $(V, \|\cdot\|)$ ein normierter Vektorraum. Dann ist $\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|$, also $\|x\| - \|y\| \leq \|x - y\|$. Andererseits ist $-(\|x\| - \|y\|) = \|y\| - \|x\| \leq \|y - x\| = \|x - y\|$, also $-\|x - y\| \leq \|x\| - \|y\|$. Insgesamt ist damit $|\|x\| - \|y\|| \leq \|x - y\|$ und aus diesem Grund die Norm Lipschitz-stetig mit Lipschitz-Konstante 1.

Nun prüfen wir das Skalarprodukt. Seien $(V, \langle\cdot,\cdot\rangle)$ ein euklidischer Vektorraum und $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ zwei Folgen in V, welche gegen $x, y \in V$ konvergieren. Dann ist

$$\begin{aligned}
|\langle x_n, y_n\rangle - \langle x, y\rangle| &= |\langle x_n, y_n\rangle - \langle x_n, y\rangle + \langle x_n, y\rangle - \langle x, y\rangle| \\
&= |\langle x_n, y_n - y\rangle + \langle x_n - x, y\rangle| \\
&\leq |\langle x_n, y_n - y\rangle| + |\langle x_n - x, y\rangle| \\
&\leq \|x_n\| \cdot \|y_n - y\| + \|x_n - x\| \cdot \|y\|,
\end{aligned}$$

wobei $\|\cdot\|$ die durch das Skalarprodukt induzierte Norm ist. Nun ist diese Norm wie gesehen stetig. Für $n \to \infty$ konvergieren also $\|x_n\|, \|y\|$ gegen $\|x\|, \|y\|$ und insbesondere $\|y_n - y\|, \|x_n - x\|$ gegen 0, der gesamte Ausdruck konvergiert damit gegen 0, woraus die Stetigkeit des Skalarproduktes folgt.

Es bleibt die Metrik. Und auch hier gilt: Jede Metrik ist stetig, wir haben dies in Satz 1.3.3 gesehen, wiederholen dies aber noch kurz: Es sei (X, d) ein metrischer Raum und $((x_n, y_n))_{n\in\mathbb{N}}$ eine Folge in X, welche gegen $(x, y) \in X$ konvergiert. Nach Definition der Konvergenz heißt dies aber genau, dass $\lim_{n\to\infty} d(x_n, x) = \lim_{n\to\infty} d(y_n, y) = 0$. Mit der Dreiecksungleichung folgt

$$\begin{aligned}
|d(x_n, y_n) - d(x, y)| &\leq |d(x_n, x) + d(x, y) + d(y, y_n) - d(x, y)| \\
&= |d(x_n, x) + d(y_n, y)| \to 0
\end{aligned}$$

und damit die Stetigkeit der Metrik. Man könnte vermuten, dass die durch $d(x, x) = 0$ und $d(x, y) = 1$ für $x \neq y$ definierte, so genannte *diskrete Metrik* ein Beispiel einer unstetigen Metrik ist. Dies ist jedoch ein falscher Gedanke aus folgendem Grund: Stetigkeit bedeutet, dass für jede gegen $(x, y) \in X$ konvergente Folge $((x_n, y_n))_n$ in X die Folge $(d(x_n, y_n))_n$ gegen $d(x, y)$ konvergiert. Wenn (x_n, y_n) aber gegen (x, y) konvergiert, heißt dies, dass der *Abstand* von (x_n, y_n) und (x, y) beliebig klein wird. Dieser Abstand, und das ist der Knackpunkt dieser Argumentation, wird aber durch die Metrik angegeben; das heißt in jedem Fall greift also unsere oben getroffene Abschätzung, auch wenn die Metrik auf den ersten Blick unstetig zu sein scheint.

Die Stetigkeit der Norm führt uns zu folgendem Korollar.

Korollar 2.3.1. *Je zwei Normen auf $\mathbb{R}^n$ sind äquivalent.*

Beweis. Da die Normenäquivalenz eine Äquivalenzrelation ist, reicht es zu zeigen, dass jede Norm auf $\mathbb{R}^n$ äquivalent zur 2-Norm ist. Sei $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ eine Norm auf $\mathbb{R}^n$. Zu zeigen ist: Es existieren $c_1, c_2 > 0$, sodass für alle $x \in \mathbb{R}^n$ gilt: $c_1\|x\|_2 \leq \|x\| \leq c_2\|x\|_2$. Für $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ ist dann

$$\begin{aligned}\|x\| = \left\|\sum_{i=1}^n x_i e_i\right\| \leq \sum_{i=1}^n |x_i|\|e_i\| &= \left\langle \begin{pmatrix}|x_1|\\ \vdots \\ |x_n|\end{pmatrix}, \begin{pmatrix}\|e_1\|\\ \vdots \\ \|e_n\|\end{pmatrix}\right\rangle \\ &\leq \left\|\begin{pmatrix}|x_1|\\ \vdots \\ |x_n|\end{pmatrix}\right\|_2 \cdot \left\|\begin{pmatrix}\|e_1\|\\ \vdots \\ \|e_n\|\end{pmatrix}\right\|_2 = \|x\|_2 \cdot \underbrace{\left\|\begin{pmatrix}\|e_1\|\\ \vdots \\ \|e_n\|\end{pmatrix}\right\|_2}_{=:c_2>0}\end{aligned}$$

Da $S := \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ kompakt und die Norm stetig ist, existiert nach dem Satz vom Minimum und Maximum ein $x_{\min} \in S$, sodass $\|x\| \geq \|x_{\min}\|$ für alle $x \in S$. Setze $c_1 := \|x_{\min}\| > 0$ (c_1 ist positiv, weil $x_{\min} \neq 0$). Sei $y \in \mathbb{R}^n \setminus \{0\}$ beliebig. Dann ist

$$\|y\| = \left\|\|y\|_2 \frac{y}{\|y\|_2}\right\| = \|y\|_2 \cdot \left\|\frac{y}{\|y\|_2}\right\| \geq \|y\|_2 \cdot \|x_{\min}\| = c_1\|y\|_2. \qquad \square$$

Satz 2.3.2. *Jede lineare Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ ist Lipschitz-stetig.*

Beweis. Aus der linearen Algebra wissen wir, dass es genau eine Matrix $A = (a_{ij}) \in \mathrm{Mat}_{m,n}(\mathbb{R})$ gibt, welche $f(x) = Ax$ für alle $x \in \mathbb{R}^n$ erfüllt, wobei die Matrizenmultiplikation für $x = (x_1, \ldots, x_n)^T$ definiert ist durch

$$Ax = \begin{pmatrix}\sum_{i=1}^n a_{1n}x_n \\ \vdots \\ \sum_{i=1}^n a_{mi}x_i\end{pmatrix} \in \mathbb{R}^m.$$

Behauptung: In Abhängigkeit von A existiert ein L, sodass für $x \in \mathbb{R}^n$ die Ungleichung $\|Ax\|_2 \leq L\|x\|_2$ gilt. Schreiben wir $Ax = ((Ax)_1, \ldots, (Ax)_m)^T$ und wenden die Cauchy-Schwarz-Ungleichung an, erhalten wir

$$\begin{aligned}\|Ax\|_2 \leq \|Ax\|_1 = \sum_{i=1}^m |(Ax)_i| &= \sum_{j=1}^m \left|\sum_{i=1}^n a_{ji}x_i\right| \\ &\leq \sum_{j=1}^m \left(\left(\sum_{i=1}^n a_{ji}^2\right)^{1/2} \cdot \|x\|_2\right) = \underbrace{\left(\sum_{j=1}^m \left(\sum_{i=1}^n a_{ji}^2\right)^{1/2}\right)}_{=:L} \cdot \|x\|_2. \qquad \square\end{aligned}$$

Definition 2.3.3 (Operatornorm). Für $A \in \mathrm{Mat}_{m,n}(\mathbb{R})$ ist die *Operatornorm* von A definiert durch $\|A\| := \sup_{x \in S^{n-1}} \|Ax\|_2$.

Aus Satz 2.3.2 ist dieser Ausdruck stets endlich, denn ist $L > 0$ mit $\|Ax\|_2 \leq L\|x\|_2$ für alle $x \in \mathbb{R}^n$, so gilt $\|Ax\|_2 \leq L$ für alle $x \in S^{n-1}$, also $\|A\| \leq L < \infty$. Wir verifizieren nun kurz, dass die Operatornorm eine Norm auf $\mathrm{Mat}_{m,n}(\mathbb{R})$ ist und halten einige ihrer wesentlichen Eigenschaften fest.

Proposition 2.3.4. *Die Operatornorm ist eine Norm auf* $\mathrm{Mat}_{m,n}(\mathbb{R})$ *und erfüllt:*

1. $\|Ax\| \leq \|A\|\|x\|_2$.
2. $\|AB\| \leq \|A\|\|B\|$ *(Submultiplikativität).*

Beweis. Wir zeigen zunächst, dass die Operatornorm eine Norm ist. Zur Definitheit. Ist $\|A\| = 0$, so ist insbesondere $Ae_i = 0$ für alle Standardbasisvektoren $e_1, \dots, e_m$, also folgt $A = 0$. Die Betragshomogenität ergibt sich für $\lambda \in \mathbb{R}$ aus

$$\|\lambda A\| = \sup_{x \in S^{n-1}} \|\lambda Ax\|_2 = \sup_x |\lambda| \|Ax\|_2 = |\lambda| \sup_x \|Ax\|_= |\lambda| \|A\|.$$

Schließlich folgt die Subadditivität aus

$$\begin{aligned}\|A + B\| &= \sup_{x \in S^{n-1}} \|(A+B)x\|_2 = \sup_x \|Ax + Bx\|_2 \leq \sup_x \|Ax\|_2 + \|Bx\|_2 \\ &\leq \sup_x \|Ax\|_2 + \sup_x \|Bx\|_2 = \|A\| + \|B\|\end{aligned}$$

und die Operatornorm ist folglich eine Norm.

Zu 1. Für $x = 0$ ist $0 \leq 0$ trivial. Für $x \neq 0, x' := \frac{x}{\|x\|_2}$ ist

$$\|Ax\|_2 = \left\| A \frac{\|x\|_2 \cdot x}{\|x\|_2} \right\|_2 = \|x\|_2 \cdot \left\| A \frac{x}{\|x\|_2} \right\|_2 = \|x\|_2 \cdot \|Ax'\|_2 \leq \|x\|_2 \cdot \|A\|,$$

da $\|x'\|_2 = 1$.

Zu 2. Es ist

$$\begin{aligned}\|AB\| &= \sup_{x \in S^{n-1}} \|(AB)x\|_2 = \sup_x \|A(Bx)\|_2 \\ &\leq \sup_x \|A\| \|Bx\|_2 = \|A\| \sup_x \|Bx\|_2 \leq \|A\| \|B\|.\end{aligned}$$

□

Zwar ist die Operatornorm wegen $\mathrm{Mat}_{m,n}(\mathbb{R}) \cong \mathbb{R}^{mn}$ äquivalent zu jeder anderen Norm auf $\mathrm{Mat}_{m,n}(\mathbb{R})$, dennoch werden wir die Operatornorm an einigen Stellen gebrauchen.

3 Mehrdimensionale Differentialrechnung

Nachdem wir nun die Stetigkeit auf metrische Räume verallgemeinert haben, wollen wir uns nun der aus Analysis 1 bekannten Differentialrechnung widmen. Dabei werden wir uns aber auf den $\mathbb{R}^n$ beschränken. Am Ende dieses Unterkapitels werden wir einen Differenzierbarkeitsbegriff erarbeitet haben, welcher uns die Übertragung vieler Ergebnisse aus Analysis 1 ermöglicht. Wir werden uns aber zunächst einigen intuitiveren Konstruktionen widmen.

Konvention. Ab sofort sei $U \subseteq \mathbb{R}^n$ stets eine in $\mathbb{R}^n$ offene Teilmenge.

3.1 Partielle und Richtungsableitungen

,

Die Differenzierbarkeit einer Funktion $f \colon U \to \mathbb{R}^m$ im Punkt $x \in U$ gibt einen Überblick darüber, wie sich f in einer (gegebenenfalls sehr kleinen) offenen Umgebung von x verhält. Nun erstreckt sich diese Umgebung aber in n Dimensionen. Aus Analysis 1 sind uns keine Methoden bekannt, diesen n-dimensionalen Fall zu untersuchen. Wir können jedoch folgende Überlegung verfolgen. Wir betrachten nicht, wie sich f in einer Umgebung von x verhält, sondern schränken uns auf das Verhalten von f in Richtung der Koordinatenachsen ein. Dies heißt, wir ignorieren alle Funktionswerte außerhalb der Geraden $x + te_i$ für $t \in \mathbb{R}, 1 \leq i \leq n$ und schauen lediglich, wie f auf dieser Geraden aussieht. Da wir damit f als eindimensionale Funktion auffassen können, können wir sehr ähnlich wie in Analysis 1 differenzieren.

Definition 3.1.1 (Partielle Ableitung)**.** Für eine Funktion $f \colon U \to \mathbb{R}^m$ ist die *partielle Ableitung nach der i-ten Variable* im Punkt $x = (x_1, \dots, x_n) \in U$ ist definiert durch

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial}{\partial x_i} f(x) = \partial_i f(x) := \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h} \in \mathbb{R}^m,$$

falls dieser Grenzwert existiert (hierbei ist e_i der i-te Standardbasisvektor). Existieren alle $\partial_i f$ in jedem Punkt $x \in U$, so heißt f *partiell differenzierbar* in x.

Diese Definition ähnelt stark der Differenzierbarkeitsdefinition aus Analysis 1. Dort hieß eine Funktion $f \colon U \to \mathbb{R}$, wobei U ein offenes Intervall war, differenzierbar in $x \in U$, wenn der Grenzwert

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

existierte. Wir überlegen uns nun, wie wir diese partiellen Ableitungen berechnen. Da wir die Funktion f auf die Gerade $x + te_i$ einschränken, bleiben alle anderen Funktionswerte fest. Schreiben wir $f(x) = f(x_1, \dots, x_n)$, so können wir die Funktion $\widetilde{f_i} \colon U \to \mathbb{R}^m$ betrachten, welche bis auf x_i alle Variablen festhält und nur noch von der i-ten Variable abhängt, also $\widetilde{f_i}(\xi) = f(x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_n)$. Dann ist die partielle Ableitung von f nach x_i im Punkt x die gewöhnliche Ableitung von $\widetilde{f_i}$ an der Stelle x_i, es gilt also

$$\frac{\partial f}{\partial x_i}(x) = \frac{\mathrm{d}\widetilde{f_i}}{\mathrm{d}\xi}(x_i).$$

Dies liefert uns eine einfache Methode zur Berechnung partieller Ableitungen. Dabei werden wir jetzt lediglich reellwertige Funktionen betrachten (also Funktionen, welche in die reellen Zahlen abbilden). Vektorwertige Funktionen werden wir nach Satz 3.1.3 diskutieren, welcher uns die Berechnung derer partiellen Ableitungen vereinfacht.

Beispiele.

1. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x,y) \mapsto x^2 + y^2 - y$. Die partiellen Ableitungen sind dann $\partial_1 f(x,y) = 2x$ und $\partial_2 f(x,y) = 2y - 1$.
2. Sei $f\colon \mathbb{R}^3 \to \mathbb{R}, (x,y,z) \mapsto \sin(x+y) + \exp(yz)$. Die partiellen Ableitungen sind dann $\partial_1 f(x,y,z) = \cos(x+y), \partial_2 f(x,y,z) = \cos(x+y) + z\exp(yz)$ und $\partial_3 f(x,y,z) = y\exp(yz)$ und f ist stetig partiell differenzierbar auf $\mathbb{R}^3$.

Wir verallgemeinern nun die partiellen Ableitungen. Statt eine Funktion auf eine Gerade in Richtung einer Koordinatenachse einzuschränken, betrachten wir sie eingeschränkt auf eine beliebige Richtung.

Definition 3.1.2 (Richtungsableitung). Sei $f\colon U \to \mathbb{R}^m$ eine Funktion. Die *Richtungsableitung* von f im Punkt $x \in U$ entlang dem Vektor $v \in \mathbb{R}^n \setminus \{0\}$ (ein solcher Vektor heißt *Richtungsvektor*) ist definiert durch

$$D_v f(x) := \lim_{h\to 0} \frac{f(x+hv) - f(x)}{h} \in \mathbb{R}^m$$

falls dieser Grenzwert existiert.

Für $v = e_i$ erhalten wir also genau $D_{e_i} f(x) = \partial_i f$. Der folgende Satz liefert ein Kriterium für die Richtungs-differenzierbarkeit einer Funktion f und vereinfacht die Berechnung.

Satz 3.1.3. *Sei $x \in U$ ein Punkt und $v \in \mathbb{R}^n$ ein Richtungsvektor. Sei ferner $f\colon U \to \mathbb{R}^m$ eine Funktion in Komponentendarstellung*

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

Dann existiert $D_v f$ genau dann, wenn alle $D_v f_i$ existieren und es gilt

$$D_v f(x) = \begin{pmatrix} D_v f_1(x) \\ \vdots \\ D_v f_m(x) \end{pmatrix}.$$

Beweis. Für die Richtungsableitung gilt

$$D_v f(x) = \lim_{h\to 0} \frac{f(x+hv) - f(x)}{h} = \lim_{h\to 0} \begin{pmatrix} \frac{f_1(x+hv) - f_1(x)}{h} \\ \vdots \\ \frac{f_m(x+hv) - f_m(x)}{h} \end{pmatrix}.$$

Existiert $D_v f$, so existieren nach Korollar 1.3.2 alle Komponentengrenzwerte und es ist

$$D_v f(x) = \begin{pmatrix} \lim_{h\to 0} \frac{f_1(x+hv) - f_1(x)}{h} \\ \vdots \\ \lim_{h\to 0} \frac{f_m(x+hv) - f_m(x)}{h} \end{pmatrix} = \begin{pmatrix} D_v f_1(x) \\ \vdots \\ D_v f_m(x) \end{pmatrix}.$$

Existieren andererseits die Richtungsableitungen aller Komponentenfunktionen, so existiert auch $D_v f$ und ebenfalls gilt die gezeigte Identität. $\square$

Wie bei der Konvergenz reduziert sich die Richtungsableitungen einer Funktion in n-Variablen auf n Ableitungen von Funktionen einer Variablen. Da partielle Ableitungen besondere Richtungsableitungen sind, folgt

$$\partial_i f(x) = \begin{pmatrix} \partial_i f_1(x) \\ \vdots \\ \partial_i f_m(x) \end{pmatrix}.$$

Diese Identität werden wir häufig gebrauchen. Damit sind partielle Ableitungen vektorwertiger Funktionen nicht wesentlich schwerer als partielle Ableitungen reellwertiger Funktionen.

Beispiele.
1. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ definiert durch $f(x,y) \mapsto \big(x^2y, x-3y\big)$. Dann ist $\partial_1 f(x,y) = (2xy, 1)$ und $\partial_2 f(x,y) = \big(x^2, -3\big)$.
2. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ definiert durch $f(x,y) \mapsto \big(\sin(x)y, x^2, \exp(2y)\big)$. Dann ist $\partial_1 f(x,y) = (\cos(x)y, 2x, 0)$ und $\partial_2 f(x,y) = (\sin(x), 0, 2\exp(y))$.

Definition 3.1.4. Seien $x \in U$ und $v_1, \ldots, v_k \in \mathbb{R}^n$ Richtungsvektoren sowie $f\colon U \to \mathbb{R}^m$. Dann heißt eine Ableitung der Form $D_{v_1} D_{v_2} \cdots D_{v_k} f$ *Richtungsableitung k-ter Ordnung* oder *k-te Richtungsableitung*. Insbesondere bezeichnet man partielle Ableitungen der Form $\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} f$, wobei $1 \leq i_1, \ldots, i_k \leq n$, als *$k$-te partielle Ableitung*. Die Funktion f heißt *partielle Ableitung k-ter Ordnung* oder *k-mal (stetig) partiell differenzierbar in x*, wenn alle partiellen Ableitungen von höchstens k-ter Ordnung in x existieren (und stetig sind). Man schreibt auch

$$\frac{\partial^k f}{\partial_{x_{i_1}} \cdots \partial_{x_{i_k}}} := \partial_{i_1} \cdots \partial_{i_k} f, \quad \text{und insbesondere} \quad \frac{\partial^2 f}{\partial x_i^2} = \partial_i^2 f := \partial_i \partial_i f.$$

Beispiel. Wir betrachten die Funktion $f\colon \mathbb{R}^2 \to \mathbb{R}^2, (x,y) \mapsto x^3y + y^2e^x - y$. Die partiellen Ableitungen von f sind dann $\partial_1 f(x,y) = 3x^2y + y^2e^x$ und $\partial_2 f(x,y) = x^3 + 2ye^x - 1$. Leiten wir diese Ausdrücke nochmal partiell ab, so erhalten wir $\partial_1\partial_1 f(x,y) = 6xy + y^2e^x$, $\partial_2\partial_1 f(x,y) = 3x^2 + 2ye^x$ sowie $\partial_1\partial_2 f(x,y) = 3x^2 + 2ye^x$, $\partial_2\partial_2 f(x,y) = 2e^x$.

Wir erkennen, dass $\partial_1\partial_2 f(x,y) = \partial_2\partial_1 f(x,y)$. Berechnet man weitere übliche Beispiele, so stellt man fest, dass unter nicht allzu strengen Voraussetzungen diese Gleichheit gilt:

Satz 3.1.5 (Satz von Schwarz). *Sei $f\colon U \to \mathbb{R}$ eine zweimal stetig partiell differenzierbare Funktion. Dann ist $\partial_i\partial_j f = \partial_j\partial_i f$ für $1 \leq i, j \leq n$.*

Vor der Beweisführung erläutern wir die Grundidee des Beweises: Ohne Beschränkung der Allgemeinheit nehmen wir $n = 2, i = 1, j = 2$ an. Um einen fixen Punkt x legen wir ein Quadrat, welches ganz in U liegt. Mithilfe des Mittelwertsatzes der eindimensionalen Analysis finden wir zwei Punkte, in welchen die beiden zweiten partiellen Ableitungen übereinstimmen. Wird das Quadrat dann beliebig klein, so folgt die Gleichheit der beiden zweiten partiellen Ableitungen in x. Es folgt der Beweis:

Beweis. Sei ohne Beschränkung der Allgemeinheit $n = 2, i = 1, j = 2$. Für einen Punkt $x \in U$ sei $r > 0$ so, dass das Quadrat $\{x + se_1 + te_2 \mid |s|, |t| \leq r\}$ ganz in U enthalten ist. Wir fixieren solche $s, t \neq 0$ mit $|s|, |t| \leq r$ und betrachten die differenzierbare Hilfsfunktion

$$\varphi\colon [0,s] \to \mathbb{R}, h \mapsto f(x + he_1 + te_2) - f(x + he_1),$$

Nach dem Mittelwertsatz der Differentialrechnung existiert ein $\xi_1 \in (0,s)$ mit

$$\begin{aligned} \varphi(s) - \varphi(0) &= f(x + se_1 + te_2) - f(x + se_1) - f(x + te_2) + f(x) \\ &= s\varphi'(\xi_1) = s(\partial_1 f(x + \xi_1 e_1 + te_2) - \partial_1 f(x + \xi_1 e_1)). \end{aligned}$$

Nun betrachten wir die differenzierbare Abbildung

$$\psi\colon [0,t] \to \mathbb{R}, k \mapsto \partial_1 f(x + \xi e_1 + k e_2).$$

Anwenden des Mittelwertsatzes bringt ein $\mu_1 \in (0,t)$ mit

$$\begin{aligned}\psi(t) - \psi(0) &= \partial_1 f(x + \xi_1 e_1 + t e_2) - \partial_1 f(x + \xi_1 e_1) \\ &= t\psi'(\mu_1) = t\partial_2\partial_1 f(x + \xi_1 e_1 + \mu_1 e_2).\end{aligned}$$

Insgesamt erhalten wir also

$$f(x + se_1 + te_2) - f(x + se_1) - f(x + te_2) + f(x) = st\partial_2\partial_1 f(x + \xi_1 e_1 + \mu_1 e_2).$$

Wenden wir diese Rechnung in umgekehrter Reihenfolge an, so erhalten wir $\xi_2 \in (0,s)$ und $\mu_2 \in (0,t)$ mit

$$f(x + se_1 + te_2) - f(x + se_1) - f(x + te_2) + f(x) = st\partial_1\partial_2 f(x + \xi_2 e_1 + \mu_2 e_2),$$

also

$$\partial_2\partial_1 f(x + \xi_1 e_1 + \mu_1 e_2) = \partial_1\partial_2 f(x + \xi_2 e_1 + \mu_2 e_2). \tag{$*$}$$

Für alle vorgegebenen $s, t \neq 0$ mit $|s|, |t| \leq r$ existieren also ξ_1, ξ_2 bzw. μ_1, μ_2, deren Betrag nicht größer ist als der Betrag von s bzw. t und welche $(*)$ erfüllen. Zieht sich das eingangs beschriebene Quadrat nun auf den Punkt x zusammen, streben also $s, t \to 0$, dann konvergieren auch $\xi_1, \xi_2, \mu_1, \mu_2 \to 0$ und aus der vorausgesetzten Stetigkeit der zweiten partiellen Ableitungen folgt die Identität $\partial_1\partial_2 f(x) = \partial_2\partial_1 f(x)$. □

Bemerkung. Der Satz von Schwarz gilt allgemein für Richtungsableitungen in beliebige Richtungen $v, u \in \mathbb{R}^n$. Im Beweis sind dann $e_1, e_2, \partial_1, \partial_2$ durch u, v, D_u, D_v zu ersetzen.

Definition 3.1.6 (Hesse-Matrix). Sei $f : U \to \mathbb{R}$ zweimal stetig differenzierbar. Die *Hesse-Matrix* von f bei x ist definiert durch

$$Hf(x) = \begin{pmatrix} \partial_1\partial_1 f(x) & \cdots & \partial_1\partial_n f(x) \\ \vdots & & \vdots \\ \partial_n\partial_1 f(x) & \cdots & \partial_n\partial_n f(x) \end{pmatrix} \in \mathrm{Mat}_n(\mathbb{R}).$$

Nach dem Satz von Schwarz ist die Hesse-Matrix symmetrisch, das heißt $Hf(x)^T = Hf(x)$. Das folgende Korollar werden wir an einigen Stellen später verwenden.

Korollar 3.1.7. *Sind $v, w \in \mathbb{R}^n$ Richtungsvektoren, so ist*

$$D_v D_w f(x) = \langle Hf(x)v, w\rangle = \langle v, Hf(x)w\rangle.$$

Für den Beweis werden wir die Erkenntnis aus der linearen Algebra benutzen, dass für eine symmetrische Matrix die Identität $\langle Av, w\rangle = \langle v, Aw\rangle$ gilt.

Beweis. Schreibe $v = \sum_{i=1}^n v_i e_i$ und $w = \sum_{i=1}^n w_i e_i$ für $v_i, w_i \in \mathbb{R}$. Dann ist

$$\begin{aligned} D_v D_w f(x) &= \sum_{i=1}^n v_i \partial_i D_w f(x) = \sum_{i,j=1}^n v_i w_j \partial_i \partial_j f(x) = \sum_{i,j=1}^n v_i w_j (Hf(x))_{ji} \\ &= \sum_{j=1}^n w_j (Hf(x)v)_j = \langle w, Hf(x)v\rangle = \langle Hf(x)v, w\rangle. \end{aligned}$$ □

3.2 Totale Differenzierbarkeit

Wir wollen uns nun dem Begriff nähern, dass eine Funktion nicht nur entlang von Richtungsvektoren differenzierbar ist, sondern im gesamten Raum durch eine lineare Funktion approximiert werden kann. Dazu können wir jedoch nicht einfach die übliche Grenzwertdefinition aus Analysis 1 verallgemeinern. Denn der Grenzwert

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \in \mathbb{R} \tag{$*$}$$

würde voraussetzen, dass wir durch einen Vektor $h \in \mathbb{R}^n$ teilen. Also benötigen wir eine anderen Ansatz. In Analysis 1 hatten wir eine zu $(*)$ äquivalente Definition. Dort hieß eine Funktion $f : U \to \mathbb{R}$ (wobei $U \subseteq \mathbb{R}$ ein offenes Intervall war) differenzierbar in $x \in U$, wenn es eine reelle Zahl c gibt, sodass

$$f(x+h) = f(x) + ch + r(h),$$

wobei $r : \mathbb{R} \to \mathbb{R}$ eine Funktion war mit $\lim_{h\to 0} \frac{r(h)}{h} = 0$. Anschaulich bedeutet das: Am fixen Punkt $(x, f(x))$ legen wir die Tangente mit Steigung c an die Funktion an. Dann soll die Funktion in einer gewissen Nähe von x bis auf einen Fehler $r(h)$ mit der Tangente übereinstimmen. Dieser Restterm $r(h)$ muss dabei schnell genug klein werden, sodass $\lim_{h\to 0} \frac{r(h)}{h} = 0$ erfüllt ist. Existieren c und r, so hieß c die Ableitung von f an der Stelle x und war eindeutig bestimmt, denn es war $c = f'(x)$. Wir nutzen diese Definition, um zu formulieren, dass eine Funktion $\mathbb{R}^n \to \mathbb{R}^m$ lokal durch eine lineare Funktion approximiert werden kann.

Definition 3.2.1 (Totale Differenzierbarkeit)**.** Eine Funktion $f : U \to \mathbb{R}^m$ heißt *(total) differenzierbar im Punkt* $x \in U$, falls eine lineare Abbildung $L : \mathbb{R}^n \to \mathbb{R}^m$ und eine Restfunktion $r : U \to \mathbb{R}^m$ existieren, sodass

$$f(x+h) = f(x) + L(h) + r(h) \quad \text{und} \quad \lim_{h\to 0} \frac{r(h)}{\|h\|} = 0.$$

Dabei ist $\|\cdot\|$ irgendeine der (äquivalenten) Normen auf $\mathbb{R}^m$. Man bezeichnet die lineare Abbildung L als *(totales) Differential von f in x* (oder manchmal auch *Ableitung*) und schreibt $Df(x)$. Aufgrund des Vektorraumisomorphismus $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathrm{Mat}_{m,n}(\mathbb{R})$ entspricht jede lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ genau einer $m \times n$-Matrix mit reellen Koeffizienten. Damit geht die Differenzierbarkeitsdefinition über in

$$f(x+h) = f(x) + Ah + r(h),$$

wobei $A \in \mathrm{Mat}_{m,n}(\mathbb{R})$ und $\lim_{h\to 0} \frac{r(h)}{\|h\|} = 0$. Für die meisten Zwecke werden wir auf die Definition mittels Matrizen zurückgreifen. Dies ist aufgrund des oben beschriebenen Isomorphismus und der daraus folgenden Äquivalenz der Definitionen möglich.

Korollar 3.2.2. *Ist f differenzierbar in x, so ist $Df(x)$ eindeutig bestimmt.*

Beweis. Sei $\widetilde{D}f(x) \in \mathrm{Mat}_{m,n}(\mathbb{R})$ eine weitere Matrix mit der Eigenschaft, dass $f(x+h) = f(x) + \widetilde{D}f(x)h + \widetilde{r}(h)$ mit einer Restfunktion $\widetilde{r}$, welche $\lim_{h\to 0} \frac{\widetilde{r}(h)}{\|h\|} = 0$ erfüllt. Dann folgt $Df(x)h + r(h) = \widetilde{D}f(x)h + \widetilde{r}(h)$, also $Df(x)h - \widetilde{D}f(x)h = \widetilde{r}(h) - r(h)$. Division durch $\|h\|$ und Anwenden des Grenzwertes ergibt

$$\lim_{h\to 0} \frac{Df(x)h - \widetilde{D}f(x)h}{\|h\|} = \lim_{h\to 0} \frac{\widetilde{r}(h) - r(h)}{\|h\|} = 0.$$

Schreibe $h = tv$ für $t \in \mathbb{R}$ und einen Einheitsvektor $v \in \mathbb{R}^n$ (das heißt es ist $\|v\| = 1$). Wir erhalten wegen der Linearität von $Df(x)$, $\widetilde{D}f(x)$ die Gleichung

$$\lim_{t\to 0} \frac{Df(x)(tv) - \widetilde{D}f(x)(tv)}{t} = \lim_{t\to 0} \frac{tDf(x)v - t\widetilde{D}f(x)v}{t}$$
$$= \lim_{t\to 0} Df(x)v - \widetilde{D}f(x)v = 0.$$

Also ist stimmen $Df(x)$ und $\widetilde{D}f(x)$ auf allen Einheitsvektoren überein, damit insbesondere auf allen Standardbasisvektoren, also (da lineare Abbildungen eindeutig durch die Bilder von Basisvektoren bestimmt sind) auf $\mathbb{R}^n$. □

Wir wissen nun also, dass der Ausdruck $Df(x)$ wohldefiniert ist. Nun stellen wir uns die Frage, wie dieses totale Differential aussieht.

Satz 3.2.3. *Sei $f : U \to \mathbb{R}^m$ mit Komponenten $f_1, \dots, f_m$ differenzierbar in $x \in U$. Dann sind alle $f_1, \dots, f_m$ partiell differenzierbar und es ist*

$$Df(x) = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix}.$$

Lemma 3.2.4. *Sei $f : U \to \mathbb{R}^m$ eine in x differenzierbare Funktion. Dann ist f in x in jede Richtung $v \in \mathbb{R}^n \setminus \{0\}$ differenzierbar mit $D_v f(x) = Df(x)v$.*

Beweis. Schreibe $f(x+h) = f(x) + Df(x)h + r(h)$ für $Df(x) \in \mathrm{Mat}_{m,n}(\mathbb{R})$ und $\lim_{h\to 0} \frac{r(h)}{\|h\|} = 0$. Dann ist

$$D_v f(x) = \frac{f(x+hv) - f(x)}{h} = \frac{f(x) + Df(x)(hv) + r(hv) - f(x)}{h}$$
$$= \frac{hDf(x)v + r(hv)}{h} = Df(x)v + \frac{r(hv)}{h} \xrightarrow{h\to 0} Df(x)v,$$

also existiert die Richtungsableitung und es gilt die behauptete Identität. □

Bemerkung. Insbesondere gilt: Ist f differenzierbar in x, so hängen die Richtungsableitungen in x linear von der Richtung ab. Ist dies umgekehrt nicht der Fall, so ist f nicht differenzierbar.

Beweis von Satz 3.2.3. Nach Lemma 3.2.4 und Korollar 3.1.3 sind alle Komponentenfunktionen partiell differenzierbar und es folgt

$$i\text{-te Spalte von } Df(x) = Df(x)e_i = \partial_i f(x) = \begin{pmatrix} \partial_i f_1(x) \\ \vdots \\ \partial_i f_m(x) \end{pmatrix}.$$ □

Definition 3.2.5 (Jacobi-Matrix). Sei $f : U \to \mathbb{R}^m$ eine in $x \in U$ differenzierbare Funktion mit Komponentenfunktionen $f_1, \dots, f_m$. Die *Jaboci-Matrix von f im Punkt x* ist definiert durch

$$Jf(x) := \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix} \in \mathrm{Mat}_{m,n}(\mathbb{R}).$$

Da wie gesehen dieser Matrix die Rolle der Ableitung (also der approximierenden linearen Funktion) zukommt, schreibt man für $Jf(x)$ auch $Df(x)$.

Ist f also differenzierbar, so hat das totale Differential die Gestalt der Jacobi-Matrix. Dieser Ausdruck ist jedoch nur dann sinvoll, wenn f auch differenzierbar ist, denn umgekehrt gilt nicht, dass jede Funktion, deren partielle Ableitungen existieren, differenzierbar ist. Wir diskutieren einige Beispiele.

Beispiele.

1. Betrachte $f\colon \mathbb{R}^n \to \mathbb{R}, x \mapsto \langle x, x\rangle = \|x\|_2^2$. Dann ist
$$f(x+h) = \langle x, x\rangle + \langle h, h\rangle + 2\langle x, h\rangle = f(x) + \langle h, h\rangle + 2\langle x, h\rangle.$$
Wir suchen also $A \in \mathrm{Mat}_{1,n}(\mathbb{R})$ und $r\colon \mathbb{R}^n \to \mathbb{R}$, sodass $Ah + r(h) = \langle h, h\rangle + 2\langle x, h\rangle$ und $\lim_{h\to 0} \frac{r(h)}{\|h\|_2} = 0$. Sei $A = (2x_1, \dots, 2x_n)$ und $r(h) := \langle h, h\rangle$. Dann ist
$$Ah = \sum_{i=1}^{n} 2x_i h_i = 2\langle x, h\rangle \quad \text{und} \quad \lim_{h\to 0} \frac{r(h)}{\|h\|_2} = \lim_{h\to 0} \frac{\|h\|_2^2}{\|h\|_2} = 0.$$
Also ist f differenzierbar und es ist $Df(x) = (2x_1, \dots, 2x_n)$.
2. Für $A \in \mathrm{Mat}_{m,n}(\mathbb{R})$ ist die lineare Abbildung $f\colon \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Ax$ differenzierbar und es ist $Df(x) = A$, denn trivialerweise gilt $f(x+h) = A(x+h) = Ax + Ah = f(x) + Ah$.

3.3 Zusammenhänge der Differenzierbarkeitsbegriffe

In Analysis 1 haben wir gesehen, dass jede differenzierbare Funktion stetig ist. Da wir in der mehrdimensionalen Differentialrechnung verschiedene Differenzierbarkeitsbegriffe kennengelernt haben, wollen wir kurz die Zusammenhänge dieser Ableitungen festhalten.

Satz 3.3.1. *Ist $f\colon U \to \mathbb{R}^m$ differenzierbar in $x \in U$, so ist f stetig in x.*

Beweis. Es gelte $f(x+h) = f(x) + Ah + r(h)$ mit $A \in \mathrm{Mat}_{m,n}(\mathbb{R})$ und $\lim_{h\to 0} \frac{r(h)}{\|h\|} = 0$. Dann folgt aus $\|f(x+h) - f(x)\| = \|Ah + r(h)\| \xrightarrow{h\to 0} 0$ die Stetigkeit von f im Punkt x. □

Satz 3.3.2. *Sei $f\colon U \to \mathbb{R}$ eine Funktion, welche im Punkt $x \in U$ alle Richtungsableitungen besitzt. Dann ist f nicht notwendigerweise (1) stetig oder (2) differenzierbar in x.*

Beweis. Zu 1. Wir betrachten die Funktion
$$f\colon \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto \begin{cases} \frac{x^2 y}{x^4 + y^2}, & (x, y) \neq 0 \\ 0, & (x, y) = 0 \end{cases}.$$
Es sei $v = (v_1, v_2) \in \mathbb{R}^2$ ein Richtungsvektor. Dann ist
$$\begin{aligned} D_v f(0,0) &= \lim_{h\to 0} \frac{f(hv_1, hv_2) - f(0,0)}{h} = \lim_{h\to 0} \frac{\frac{h^2 v_1^2 h v_2}{h^4 v_1^4 + h^2 v_2^2}}{h} \\ &= \lim_{h\to 0} \frac{h^3 v_1^2 v_2}{h^5 v_1^4 + h^3 v_2^2} = \lim_{h\to 0} \frac{v_1^2 v_2}{h^2 v_1^4 + v_2^2} = \frac{v_1^2}{v_2}. \end{aligned}$$
Also existieren alle Richtungsableitungen von f im Nullpunkt. Dennoch ist f in diesem Punkt unstetig, denn:
$$\lim_{n\to\infty} f\left(\frac{1}{n}, \frac{1}{n^2}\right) = \lim_{n\to\infty} \frac{\frac{1}{n^3}}{\frac{1}{n^4} + \frac{1}{n^4}} = \lim_{n\to\infty} \frac{n}{2} = \infty \neq 0 = f(0,0),$$

womit die Aussage über die Stetigkeit gezeigt ist.

Zu 2. Wäre f zwangsweise differenzierbar, so wäre f nach Satz 3.3.1 auch zwangsweise stetig, was nach Aussage 1 nicht der Fall ist. □

Satz 3.3.3 (Differenzierbarkeitskriterium). *Sei $f\colon U \to \mathbb{R}^m$ partiell differenzierbar und alle partiellen Ableitungen seien in einem Punkt $x \in U$ stetig. Dann ist f differenzierbar in x.*

Beweis. Da f genau dann differenzierbar ist, wenn alle Komponentenfunktionen differenzierbar sind, können wir uns auf dem Fall $m = 1$ beschränken. Dann ist $Jf(x) = (\partial_1 f(x), \dots, \partial_n(x)) \in \mathrm{Mat}_{1,n}(\mathbb{R})$. Der Satz ist bewiesen, wenn wir

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - Jf(x)h}{\|h\|} = 0 \tag{$*$}$$

zeigen können. Schreibe $x = (x_1, \dots, x_n)$. Da U offen ist, gibt es ein $\varepsilon > 0$, sodass $B_\varepsilon(x) \subseteq U$. Es sei nun $x + h = x + (h_1, \dots, h_n) \in B_\varepsilon(x)$ ein solcher Punkt. Wir definieren Punkte $a_k := x + \sum_{i=1}^k h_i e_i$ für $0 \le k \le n$. Dann ist $a_0 = x$ und $a_n = x + h$. Wir schreiben die Differenz $f(x+h) - f(x)$ als Teleskopsumme:

$$f(x+h) - f(x) = \sum_{i=1}^n f(a_i) - f(a_{i-1})$$

und werden diese mithilfe des Mittelwertsatzes umformen. Dazu definieren wir für $1 \le i \le n$ die Hilfsfunktion

$$\varphi_i\colon [0, h_i] \to \mathbb{R}, t \mapsto f(a_{i-1} + te_i).$$

Damit ist $f(a_i) - f(a_{i-1}) = \varphi_i(h_i) - \varphi_i(0)$. Die φ_i sind wegen der partiellen Differenzierbarkeit von f differenzierbar mit $\varphi_i'(t) = \partial_i f(a_{i-1} + te_i)$. Aufgrund ihrer vorausgesetzten Stetigkeit erhalten wir mit dem Mittelwert für jedes $1 \le i \le n$ ein $\xi_i \in (0, h_i)$ mit

$$\varphi_i(h_i) - \varphi_i(0) = h_i \varphi_i'(a_{i-1} + \xi_i e_i) = h_i \partial_i f(z_i).$$

mit $z_i := a_{i-1} + \xi_i e_i$. Unsere Teleskopsumme schreibt sich damit als

$$f(x+h) - f(x) = \sum_{i=1}^n h_i \partial_i f(z_i).$$

Mit $Jf(x)h = \sum_{i=1}^n \partial_i f(x) h_i$ erhalten wir

$$f(x+h) - f(x) - Jf(x)h = \sum_{i=1}^n h_i(\partial_i f(z_i) - \partial_i f(x)).$$

Die rechte Seite ist ein Skalarprodukt, auf welches wir die Cauchy-Schwarz-Ungleichung anwenden können. Wir erhalten

$$|f(x+h) - f(x) - Jf(x)h| \le \|h\| \cdot \left(\sum_{i=1}^n |\partial_i f(z_i) - \partial_i f(x)| \right)^{1/2},$$

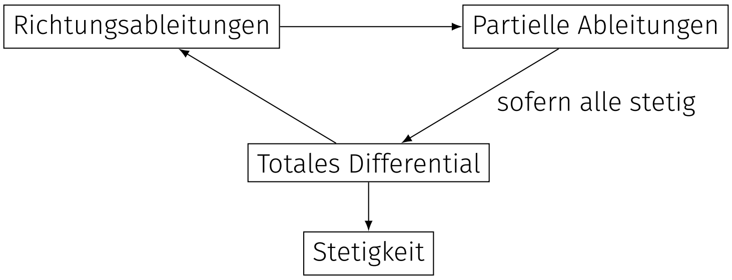
also

$$\frac{f(x+h) - f(x) - Jf(x)h}{\|h\|} \le \left(\sum_{i=1}^n |\partial_i f(z_i) - \partial_i f(x)| \right)^{1/2}.$$

Für $h \to 0$ strebt $z_i = a_{i-1} + \xi_i e_i$ nach Definition von ξ_i gegen a_{i-1}, welches wiederum gegen x strebt. Folglich ist mit

$$0 \le \lim_{h\to 0} \frac{f(x+h) - f(x) - Jf(x)h}{\|h\|} \le \lim_{h\to 0} \left(\sum_{i=1}^n |\partial_i f(z_i) - \partial_i f(x)| \right)^{1/2} = 0$$

die Aussage bewiesen. □



Dieses Diagramm zeigt die Zusammenhänge zwischen den verschiedenen Differentiationen und der Stetigkeit. Existiert das totale Differential, so hängen insbesondere die Richtungsableitungen linear von der Richtung ab. Existieren alle partiellen Ableitungen und sind stetig, so existiert das totale Differential und ist insbesondere ebenfalls stetig.

3.4 Ableitungsregeln

Satz 3.4.1 (Kettenregel). *Sei $V \subseteq \mathbb{R}^m$ offen. Es seien $f\colon U \to \mathbb{R}^m$ (wobei $f(U) \subseteq V$) und $g\colon V \to \mathbb{R}^p$ in x bzw. $f(x)$ differenzierbar. Dann ist $g \circ f$ differenzierbar in x und es ist*

$$D(g \circ f)(x) = (Dg)(f(x)) \cdot Df(x).$$

Beweis. Es seien $A := Df(x)$, $B := (Dg)(f(x))$ und es gelte

$$\begin{aligned} f(x+h) &= f(x) + Ah + \varphi(h) \quad \text{sowie} \\ g(f(x)+k) &= g(f(x)) + Bk + \psi(k), \end{aligned}$$

wobei $\lim_{h\to 0} \frac{\varphi(h)}{\|h\|} = 0$ und $\lim_{k\to 0} \frac{\psi(k)}{\|k\|} = 0$. Dann ist

$$\begin{aligned} (g \circ f)(x+h) = g(f(x+h)) &= g(f(x) + Ah + \varphi(h)) \\ &= g(f(x)) + BAh + B\varphi(h) + \psi(Ah + \varphi(h)) \\ &= g(f(x)) + BAh + \omega(h) \end{aligned}$$

mit $\omega(h) = B\varphi(h) + \psi(Ah + \varphi(h))$. Wir zeigen, dass $\lim_{h\to 0} \frac{\omega(h)}{\|h\|} = 0$. Nach Voraussetzung ist $\lim_{h\to 0} \frac{\varphi(h)}{\|h\|} = 0$, also $\lim_{h\to 0} \frac{B\varphi(h)}{\|h\|} = 0$. Nun gelten folgende Abschätzungen: Wegen $\lim_{h\to 0} \frac{\|\varphi(h)\|}{\|h\|} = 0$ existiert ein $\delta > 0$, sodass $\|\varphi(h)\| < \|h\|$ für alle $\|h\| < \delta$ und wegen $\lim_{k\to 0} \frac{\psi(k)}{\|k\|} = 0$ existiert eine Funktion ψ' mit $\psi(k) \leq \|k\|\psi'(k)$ und $\lim_{k\to 0} \psi'(k) = 0$. Ist also h klein genug, dass $Ah + \varphi(h) < \delta$, so gilt also

$$\begin{aligned} \psi(Ah + \varphi(h)) &\leq \|Ah + \varphi(h)\|\psi'(Ah + \psi(h)) \\ &\leq (\|Ah\| + \|\varphi(h)\|)\psi'(Ah + \psi(h)) \\ &\leq (\|A\|\|h\| + \|h\|)\psi'(Ah + \psi(h)) \\ &= (\|A\| + 1)\|h\|\psi'(Ah + \psi(h)) \end{aligned}$$

und es folgt $\lim_{h\to 0} \frac{\psi(Ah+\varphi(h))}{\|h\|} \leq \lim_{h\to 0}(\|A\| + 1)\psi'(Ah + \psi(h)) = 0$ □

Satz 3.4.2 (Produktregel). *Seien $f, g\colon U \to \mathbb{R}$ differenzierbar. Dann ist fg differenzierbar und es ist $D(fg)(x) = Df(x)g(x) + f(x)Dg(x)$.*

Beweis. Wir definieren Hilfsfunktionen

$$\varphi\colon U \to \mathbb{R}^2, x \mapsto \begin{pmatrix} f(x) \\ g(x) \end{pmatrix} \quad \text{sowie} \quad m\colon \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto xy.$$

Die Funktion m ist differenzierbar, da die partiellen Ableitungen

$$\partial_1 m(x,y) = \lim_{h\to 0} \frac{(x+h)y - xy}{h} = \lim_{h\to 0} \frac{xy + hy - xy}{h} = y \quad \text{und}$$
$$\partial_2 m(x,y) = \lim_{h\to 0} \frac{x(y+h) - xy}{h} = \lim_{h\to 0} \frac{xy + hx - xy}{h} = x$$

auf U existieren und stetig sind. Also ist $Dm(x,y) = (y,x)$. Ebenfalls ist φ differenzierbar, da f und g differenzierbar sind. Wegen $fg = m \circ \varphi$ folgt mit

$$D(fg)(x) = Dm(\varphi(x)) \cdot D\varphi(x) = (g(x), f(x)) \cdot \begin{pmatrix} Df(x) \\ Dg(x) \end{pmatrix}$$

und Matrizenmultiplikation die Behauptung. □

Satz 3.4.3 (Quotientenregel)**.** *Seien $f, g\colon U \to \mathbb{R}$ differenzierbar, wobei $0 \notin g(U)$. Dann ist $\frac{f}{g}$ und es ist*

$$D\left(\frac{f}{g}\right) = \frac{Df(x)g(x) - f(x)Dg(x)}{g(x)^2}.$$

Beweis. Wir definieren Hilfsfunktionen

$$\varphi\colon U \to \mathbb{R}^2, x \mapsto \begin{pmatrix} f(x) \\ g(x) \end{pmatrix} \quad \text{sowie} \quad q\colon \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto \frac{x}{y}.$$

Die Funktion q ist differenzierbar, da die partiellen Ableitungen

$$\partial_1 q(x,y) = \lim_{h\to 0} \frac{(x+h) - x}{hy} = \lim_{h\to 0} \frac{1}{y} = \frac{1}{y} \quad \text{und}$$
$$\partial_2 q(x,y) = \lim_{h\to 0} \frac{1}{h}\left(\frac{x}{y+h} - \frac{x}{y}\right) = \lim_{h\to 0} \frac{1}{h}\left(\frac{-xh}{y^2+hy}\right) = \frac{-x}{y^2}$$

auf U existieren und stetig sind. Also ist $Dq(x,y) = \left(\frac{1}{y}, \frac{-x}{y^2}\right)$. Ebenfalls ist φ differenzierbar, da f und g differenzierbar sind. Wegen $\frac{f}{g} = q \circ \varphi$ folgt mit

$$D\left(\frac{f}{g}\right)(x) = Dq(\varphi(x)) \cdot D\varphi(x) = \left(\frac{1}{g(x)}, \frac{-f(x)}{g(x)^2}\right) \cdot \begin{pmatrix} Df(x) \\ Dg(x) \end{pmatrix}$$

und Matrizenmultiplikation die Behauptung. □

3.5 Eine nützliche Tatsache zum Matrizenraum

Wir wollen zum Abschluss des Kapitels ein Beispiel behandeln, welches eine Brücke zur linearen Algebra schlägt und zu einer nützlichen Vorstellung des Vektorraums der Matrizen $\mathrm{Mat}_n(\mathbb{R})$ verhilft. Darüber hinaus ist dies ein gutes Beispiel dafür, dass Stetigkeits- und Differenzierbarkeitsbeweise auch argumentativ geführt werden können.

Proposition 3.5.1. *Die Menge der invertierbaren Matrizen* $\mathrm{GL}_n(\mathbb{R})$ *ist offen in* $\mathrm{Mat}_n(\mathbb{R})$ *und die Funktion* $F\colon \mathrm{GL}_n(\mathbb{R}) \to \mathrm{GL}_n(\mathbb{R}), A \mapsto A^{-1}$ *ist beliebig oft stetig differenzierbar.*

Beweis. Aus der linearen Algebra wissen wir, dass $\mathrm{Mat}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$. Demnach können wir eine Matrix $A = (a_{ij}) \in \mathrm{Mat}_n(\mathbb{R})$ auffassen als Tupel ihrer transponierten Spaltenvektoren; also ist

$$A = (a_{11}, \dots, a_{n1}, a_{12}, \dots, a_{n2}, \dots, a_{1n}, \dots, a_{nn}) \in \mathbb{R}^{n^2}.$$

Damit können wir die Operatornorm als eine Norm auf $\mathbb{R}^{n^2}$ auffassen. Nun ist die Determinantenabbildung $\det\colon \mathrm{Mat}_n(\mathbb{R}) \to \mathbb{R}$ stetig auf $\mathbb{R}$, denn für $A = (a_{ij}) \in \mathrm{Mat}_n(\mathbb{R})$ ist nach der Leibniz-Formel die Determinante wegen

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn} \sigma \, a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}$$

ein reelles Polynom der Matrixkoeffizienten. Diese Stetigkeit gilt wegen der Normenäquivalenz bezüglich Operator- und euklidischer Norm. Nun wissen wir aus der linearen Algebra, dass $A \in \mathrm{Mat}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$ genau dann invertierbar ist, wenn $\det A \neq 0$. Da die Determinante wie gezeigt stetig ist, existiert ein $\delta > 0$, sodass $\|A - A'\| < \delta \Rightarrow \det A' \neq 0 \Rightarrow A' \in \mathrm{GL}_n(\mathbb{R})$. Folglich ist $\mathrm{GL}_n(\mathbb{R})$ offen in $\mathrm{Mat}_n(\mathbb{R})$. Nun zeigen wir, dass die Inversionsabbildung $F\colon \mathrm{GL}_n(\mathbb{R}) \to \mathrm{GL}_n(\mathbb{R}), A \mapsto A^{-1}$ beliebig oft stetig differenzierbar ist. Sei $B = (b_{ij}) := A^{-1}$ die zu A inverse Matrix mit Spaltenvektoren $b^1, \dots, b^n$. Dann ist b eindeutig bestimmt durch das lineare Gleichungssystem

$$\begin{array}{rcl} Ab^1 & = & e_1 \\ \vdots & & \vdots\,, \\ Ab^n & = & e_n \end{array}$$

wobei $e_i = (\delta_{1i}, \dots, \delta_{ni}) \in \mathbb{R}^n$. Nach der Cramerschen Regel ist

$$b^i = \frac{\det(a^1, \dots, a^{i-1}, e_i, a^{i+1}, \dots, a^n)}{\det A}.$$

Da Zähler und Nenner als reellwertige Polynome beliebig oft stetig differenzierbar sind, ist auch der Quotient beliebig oft stetig differenzierbar. □

Die Benutzung der Determinante ist nicht nötig, und der folgende alternative Zugang liefert sogar bessere Ergebnisse.

Lemma 3.5.2. *Es sei* $A \in \mathrm{GL}_n(\mathbb{R})$ *eine invertierbare Matrix, und* $B \in \mathrm{Mat}_{n,n}(\mathbb{R})$ *mit* $\|B\| < \frac{1}{\|A^{-1}\|}$ *(Operatornorm!). Dann ist* $A + B$ *invertierbar, und es gilt*

$$\|(A+B)^{-1}\| \leq \frac{1}{\frac{1}{\|A^{-1}\|} - \|B\|}$$

Beweis. Schreibe $A + B = A + BA^{-1}A = (1 + BA^{-1})A$. Es folgt dann für $x \in \mathbb{R}^n$ mit der umgekehrten Dreiecksungleichung

$$\|(A+B)x\| = \|(1 + BA^{-1})(Ax)\| \geq \|Ax\| - \|BA^{-1}\|\|Ax\| = (1 - \|BA^{-1}\|)\|Ax\| =$$

$$= (1 - \|BA^{-1}\|)\|A^{-1}\|^{-1}\|A^{-1}\|\|Ax\| \geq \frac{(1 - \|BA^{-1}\|)}{\|A^{-1}\|}\|A^{-1}Ax\| \geq \frac{(1 - \|B\|\|A^{-1}\|)}{\|A^{-1}\|}\|x\| = \left(\frac{1}{\|A^{-1}\|} - \|B\|\right)\|x\|$$

Zusammengenommen folgt $(A+B)x = 0 \Rightarrow x = 0$. Also ist $(A+B)$ injektiv und folglich (LAI!) invertierbar. Ferner erhalten wir für die Operatornorm von $(A+B)^{-1}$:

$$\|(A+B)^{-1}y\| \leq \frac{1}{\frac{1}{\|A^{-1}\|} - \|B\|}\|(A+B)(A+B)^{-1}y\|,$$

also wie behauptet
$$\|(A+B)^{-1}\| \leq \frac{1}{\frac{1}{\|A^{-1}\|} - \|B\|}.$$
□

Lemma 3.5.3. *Die Funktion $F : \mathrm{GL}_n(\mathbb{R}) \to \mathrm{GL}_n(\mathbb{R})$, $A \mapsto A^{-1}$, ist stetig.*

Beweis. Sei $H \in \mathrm{Mat}_{n,n}(\mathbb{R})$ mit $\|H\|$ klein. Dann gilt
$$(A+H)^{-1} - A^{-1} = (A+H)^{-1}AA^{-1} - (A+H)^{-1}(A+H)A^{-1} = -(A+H)^{-1}HA^{-1}$$
und daher
$$\|(A+H)^{-1} - A^{-1}\| \leq \|(A+H)^{-1}\|\|A\|\|H\|.$$
Es gilt ferner $\|(A+H)^{-1}\| \to \|A^{-1}\|$ nach dem vorigen Lemma, und daher $\lim_{H\to 0} \|(A+H)^{-1} - A^{-1}\| = 0$. □

Lemma 3.5.4. *Die Funktion F ist differenzierbar, und es gilt*
$$DF(A)H = -A^{-1}HA^{-1}.$$

Beweis. Nach der Differenzierbarkeitsdefinition ist nur zu zeigen, dass
$$\lim_{H\to 0} \frac{1}{\|H\|}((A+H)^{-1} - A^{-1} + A^{-1}HA^{-1}) = 0,$$
aber es gilt
$$(A+H)^{-1} - A^{-1} + A^{-1}HA^{-1} = -(A+H)^{-1}HA^{-1} + A^{-1}HA^{-1} = (A^{-1} - (A+H)^{-1})HA^{-1}$$
und daher
$$\|\frac{1}{\|H\|}((A+H)^{-1} - A^{-1} + A^{-1}HA^{-1})\| \leq \frac{1}{\|H\|}\|A^{-1} - (A+H)^{-1}\|\|H\|\|A^{-1}\| \leq \|(A+H)^{-1}\|\|A\|\|H\|\|A^{-1}\| \to 0.$$
□

4 Extremwertprobleme

4.1 Mehrdimensionale Taylorentwicklung

Erinnerung. In Analysis 1 haben wir die Methode kennengelernt, um Funktionen in einer Umgebung eines Punktes a durch das so genannte Taylorpolynom n-ter Ordnung zu approximieren, welches für $f \in C^m(I, \mathbb{R})$, wobei $I \subseteq \mathbb{R}$ ein offenes Intervall ist, definiert war durch

$$T_{m,a}f(x) := \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a)(x-a)^k.$$

Dieses Polynom ist das einzige Polynom vom Grad m, für welches die Gleichheit der Ableitungen $T_{m,a}^{(k)}f(a) = f^{(k)}(a)$ für alle $0 \leq k \leq m$ gilt. Ferner gilt die Taylorformel mit Integralrest:

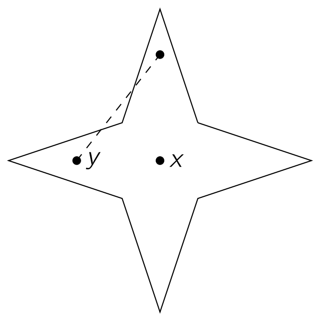
$$f(x) = T_{m,a}f(x) + \frac{1}{m!} \int_a^x (x-t)^m f^{(m+1)}(t)\mathrm{d}t,$$

sofern f sogar $(m+1)$-mal stetig differenzierbar ist. In diesem Fall existiert die Langrangesche Restglieddarstellung: Für ein ξ zwischen a und x gilt

$$f(x) = T_{m,a}(x) + \frac{1}{(m+1)!} f^{(m+1)}(\xi)(x-a)^{m+1}.$$

Diese Erkenntnisse wollen wir auf den höherdimensionalen $\mathbb{R}^n$ übertragen.

Definition 4.1.1 (Sternförmige Menge). Sei $U \subseteq \mathbb{R}^n$, $x \in U$. Dann heißt U *sternförming bezüglich* x, falls für alle $y \in U$, $t \in [0,1]$ gilt: $tx + (1-t)y \in U$.



Diese Menge ist sternförmig bezüglich x, nicht aber bezüglich y.

Satz 4.1.2. *Sei $U \subseteq \mathbb{R}^n$ sternförmig bezüglich $x \in U$, $f \in C^{m+1}(U, \mathbb{R})$ und $v \in \mathbb{R}^n$ mit $x + v \in U$. Dann* *existiert ein ξ zwischen x und $x+v$ (das heißt $\xi = x+tv$ für ein $t \in [0,1]$), sodass*

$$f(x+v) = \sum_{k=0}^{m} \frac{1}{k!} \underbrace{D_v \cdots D_v}_{k\text{-mal}} f(x) + \frac{1}{(m+1)!} \underbrace{D_v \cdots D_v}_{(m+1)\text{-mal}} f(\xi).$$

Beweis. Setze $g\colon I := \{t \in \mathbb{R} \mid x+tv \in U\} \supseteq [0,1] \to \mathbb{R}$, $t \mapsto f(x+tv)$. Da I eine offene Teilmenge von $\mathbb{R}$ ist, können wir g im Punkt x entwickeln und finden ein $t \in [0,1]$, für dass

$$f(x+v) = g(1) = \sum_{k=0}^{m} \frac{1}{k!} g^{(k)}(0)(1-0)^k + \frac{1}{(m+1)!} g^{(m+1)}(t)(1-0)^{m+1} \tag{$*$}$$

gilt. Für die Ableitungen von g gilt dabei

$$g^{(k)}(t) = \underbrace{D_v \cdots D_v}_{k\text{-mal}} f(x+tv),$$

was man leicht durch vollständige Induktion über k oder scharfes Hinsehen beweist. Der Beweis der Aussage erfolgt mit Einsetzen von $g^{(k)}(t)$ in $(*)$. □

Bemerkung. Für den Spezialfall $m = 1$ erhalten wir

$$\begin{aligned} f(x+v) &= f(x) + D_v f(x) + \frac{1}{2} D_v D_v f(\xi) \\ &= f(x) + Df(x)v + \frac{1}{2} \langle Hf(\xi)v, v \rangle . \end{aligned}$$

4.2 Extremwertprobleme

Motivation. In diesem Kapitel wollen wir die aus Analysis 1 bekannte Berechnung von Extremwerten auf Funktionen im $\mathbb{R}^n$ verallgemeinern, wobei wir zwischendurch einen kurzen Exkurs in die lineare Algebra machen.

Definition 4.2.1. Sei $f: U \to \mathbb{R}$ eine Funktion. Ein Punkt $x \in U$ heißt *lokales Maximum* bzw. *lokales Minimum* von f, falls eine Umgebung $V \subseteq U$ von x existiert, in welcher $f(x) \geq f(v)$ bzw. $f(x) \leq f(v)$ für alle $v \in V$ gilt. Als *lokales Extremum* bezeichnet man einen Punkt, welcher ein lokales Maximum oder Minimum ist.

Lemma 4.2.2. *Ist $f: U \to \mathbb{R}$ differenzierbar und besitzt im Punkt $x \in U$ ein lokales Minimum oder Maximum, so gilt $Df(x) = 0$.*

Beweis. Betrachte die Funtion $g_v(t) := f(x+tv)$ für $v \in \mathbb{R}^n$. Diese Funktion ist definiert auf einer offenen Umgebung von 0, differenzierbar und besitzt ein lokales Extremum im Nullpunkt. Aus Analysis 1 wissen wir, dass dann $g_v'(0) = D_v f(x) = Df(x)v = 0$. Da v beliebig war, folgt $Df(x) = 0$. □

Definition 4.2.3 **(Kritischer Punkt).** Sei $f: U \to \mathbb{R}$ differenzierbar. Ein Punkt $x \in U$ mit $Df(x) = 0$ heißt *kritischer Punkt von f*.

Bemerkung. Sei $f \in C^2(U, \mathbb{R})$ und $x \in U$ ein kritischer Punkt von f. Wir finden einen Radius $r > 0$, sodass $B_r(x) \subseteq U$ eine sternförmige offene Umgebung von x ist. Für alle v mit $\|v\| < r$ gilt dann

$$f(x+v) = f(x) + Df(x)v + \frac{1}{2}\langle Hf(\xi)v, v\rangle = f(x) + \frac{1}{2}\langle Hf(\xi)v, v\rangle$$

für ein ξ zwischen x und $x+v$ nach der Taylorformel.

4.3 Definitheit symmetrischer Matrizen

Definition 4.3.1. Wir schreiben $\mathrm{Sym}_n(\mathbb{R}) := \left\{A \in \mathrm{Mat}_n(\mathbb{R}) \mid A^T = A\right\}$ für den Vektorraum der symmetrischen Matrizen.

Definition 4.3.2 (Definitheit). Eine Matrix $A \in \mathrm{Sym}_n(\mathbb{R})$ heißt *positiv definit* bzw. *positiv semidefinit* bzw. *negativ definit* bzw. *negativ semidefinit*, falls $\langle x, Ax\rangle > 0$ bzw. $\langle x, Ax\rangle \geq 0$ bzw. $\langle x, Ax\rangle < 0$ bzw. $\langle x, Ax\rangle \leq 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$. Gilt keine der Bedingungen, so heißt A *indefinit*.

Definition 4.3.3 (Eigenwert/-vektor/-raum). Sei $A \in \mathrm{Mat}_n(\mathbb{R})$. Ein Skalar $\lambda \in \mathbb{R}$ heißt *Eigenwert* von A, wenn ein $v \in \mathbb{R}^n \setminus \{0\}$ existiert mit $Av = \lambda v$. Der Vektor v heißt dann *Eigenvektor* von A zum Eigenwert λ. Man definiert das *Spektrum* $\sigma(A)$ als die Menge der Eigenwerte und den *Eigenraum* von A zum Eigenwert λ durch $\mathrm{Eig}_A(\lambda) = \{x \in \mathbb{R}^n \mid Ax = \lambda x\}$.

Satz 4.3.4. *Eine symmetrische Matrix ist genau dann positiv definit bzw. positiv semidefinit bzw. negativ definit bzw. negativ semidefinit, falls alle Eigenwerte von A positiv bzw. nicht-negativ bzw. negativ bzw. nicht-positiv sind.*

Lemma 4.3.5 (Spektralsatz). *Sei $A \in \mathrm{Sym}_n(\mathbb{R})$. Dann existiert eine Orthonormalbasis von $\mathbb{R}^n$ (das heißt eine Familie von Vektoren, welche jeweils die Länge 1 besitzen, zueinander senkrecht stehen und eine Basis von $\mathbb{R}^n$ bilden), welche aus Eigenvektoren von A besteht. Äquivalent: Es existieren $\lambda_1 \leq \cdots \leq \lambda_n \in \mathbb{R}$ und $v_1, \ldots, v_n \in \mathbb{R}^n$ mit $\langle v_i, v_j\rangle = \delta_{ij}$ und $Av_i = \lambda v_i$. Für die Basis $b = (v_1, \ldots, v_n)$ ist dann*

$$_b[A]_b = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Beweis. Wird in der linearen Algebra geführt. □

Beweis von Satz 4.3.4. Wähle eine Orthonormalbasis $(v_1, \ldots, v_n)$ von $\mathbb{R}^n$ aus Eigenvektoren von A und notiere λ_i für den Eigenwert zu v_i für $1 \leq i \leq n$. Für $x = \sum_{i=1}^n a_i v_i$ ist dann

$$\langle x, Ax\rangle = \left\langle \sum_{i=1}^n a_i v_i, \sum_{i=1}^n a_i \lambda_i v_i \right\rangle = \sum_{i=1}^n a_i^2 \lambda_i \langle v_i, v_i\rangle = \sum_{i=1}^n a_i^2 \lambda_i,$$

da aufgrund der Normiertheit von v_i die Gleichung $\langle v_i, v_i\rangle = 1$ gilt. Die behauptete Äquivalenz folgt mit $\langle v_i, Av_i\rangle = \lambda_i$. □

4.4 Zurück zum Eigenwertproblem

Satz 4.4.1 (Extremwertkriterium). *Sei $f : U \to \mathbb{R}$ zweimal stetig differenzierbar und $x \in U$ ein kritischer Punkt von f. Im Punkt x gilt dann:*

1. *Ist $Hf(x)$ positiv bzw. negativ definit, so hat f ein lokales Minimum bzw. Maximum.*
2. *Hat $Hf(x)$ einen positiven bzw. negativen Eigenwert, so hat f kein lokales Maximum bzw. Minimum.*
3. *Ist $Hf(x)$ indefinitv, so hat f einen Sattelpunkt.*
4. *Ist $Hf(x)$ semidefinit, so lässt sich keine Aussagen treffen.*

Lemma 4.4.2. *Für $A \in \mathrm{Sym}_n(\mathbb{R})$ mit Eigenwerten $\lambda_1 \leq \cdots \leq \lambda_n$ gilt:*

1. *Für alle $x \in \mathbb{R}^n$ ist $\lambda_1 \|x\|_2^2 \leq \langle Ax, x\rangle \leq \lambda_n \|x\|_2^2$.*
2. $\{\langle Ax, x\rangle \mid \|x\| = 1\} = [\lambda_1, \lambda_n]$

Beweis. Zu 1. Wähle Orthonormalbasis $(v_1, \dots, v_n)$ von $\mathbb{R}^n$, wobei $v_i \in \mathrm{Eig}_A(\lambda_i)$. Für $x = \sum_{i=1}^n x_i v_i \in \mathbb{R}^n$ ist dann

$$\begin{aligned}\langle Ax, x\rangle &= \sum_{i,j=1}^n \langle Ax_i v_i, x_j v_j\rangle = \sum_{i,j=1}^n x_i x_j \langle Av_i, v_j\rangle \\ &= \sum_{i,j=1}^n x_i x_j \lambda_i \langle v_i, v_j\rangle = \sum_{i=1}^n \lambda_i x_i^2\end{aligned}$$

und es folgt

$$\lambda_1 \|x\|_2^2 = \lambda_1 \sum_{i=1}^n x_i^2 \leq \sum_{i=1}^n \lambda_i x_i^2 \leq \lambda_n \sum_{i=1}^n x_i^2 = \lambda_n \|x\|_2^2.$$

Zu 2. Betrachte

$$f(t) := \langle A(\cos(t)v_n + \sin(t)v_1), (\cos(t)v_n + \sin(t)v_1)\rangle.$$

Diese Funktion f ist stetig und es ist $\|\cos(t)v_n + \sin(t)v_1\| = 1$. Dann ist

$$\begin{aligned}f(0) &= \langle Av_n, v_n\rangle = \lambda_n \langle v_n, v_n\rangle = \lambda_n \\ f(\pi/2) &= \langle Av_1, v_1\rangle = \lambda_1 \langle v_1, v_1\rangle = \lambda_1.\end{aligned}$$

Nach dem Zwischenwertsatz gilt $f([0, \pi/2]) \supseteq [\lambda_1, \lambda_n]$ und nach Teil 1 die umgekehrte Inklusion $[\lambda_1, \lambda_n] \subseteq f([0, \pi/2])$, folglich die Behauptung. $\square$

Beweis von Satz 4.4.1. Die Aussagen 2 bzw. 4 ergeben sich aus 1 bzw. 3 durch Betrachtung der Funktion $-f$.

Zu 1. Sei $v \in \mathbb{R}^n$ ein Richtungsvektor derart, dass $x + tv \in U$ für alle $t \in [-1, 1]$. Dann existiert ein $\Theta \in [0, 1]$ mit

$$\begin{aligned}f(x + v) &= f(x) + \frac{1}{2}\langle Hf(x + \Theta v)v, v\rangle \\ &= f(x) + \frac{1}{2}\langle Hf(x)v, v\rangle + \frac{1}{2}\langle (Hf(x + \Theta v) - Hf(x))v, v\rangle.\end{aligned}$$

Seien $0 < \lambda_1 \leq \cdots \leq \lambda_n$ die Eigenwerte von $Hf(x)$. Nach Lemma 4.4.2 ist dann $\langle Hf(x)v, v\rangle \geq \lambda_1\|v\|^2$. Da f zweimal-stetig differenzierbar ist, ist $x \mapsto Hf(x)$ stetig. Also existiert ein $\delta > 0$, sodass $\|Hf(x+\Theta v) - Hf(x)\| < \lambda_1/2$ für $\|v\| < \delta$. In diesem Fall gilt

$$\begin{aligned} f(x+v) &= f(x) + \frac{1}{2}\langle Hf(x)v, v\rangle + \frac{1}{2}\langle (Hf(x+\Theta v) - Hf(x))v, v\rangle \\ &\geq f(x) + \frac{1}{2}\lambda_1\|v\|^2 - \underbrace{\frac{1}{2}\|Hf(x+\Theta v) - Hf(x)\|}_{<\lambda_1/2} \cdot \|v\|^2 \\ &\geq f(x) + \frac{1}{4}\lambda_1\|v\|^2 > f(x), \end{aligned}$$

falls $v \neq 0$. Also folgt $f(x+v) > f(x)$, falls $\|v\| < \delta$ und $v \neq 0$ und damit, dass f ein lokales Minimum besitzt.

Zu 3. Sei $\lambda > 0$ ein Eigenwert von $Hf(x)$ und $w \in \mathrm{Eig}_{Hf(x)}(\lambda)$ mit $\|w\| = 1$. Für hinreichend kleine $t \in \mathbb{R}$ ist $x + tw \in U$ und wir erhalten

$$\begin{aligned} f(x+tw) &= f(x) + \frac{1}{2}\langle Hf(x)tw, tw\rangle + \frac{1}{2}\langle (Hf(x+\Theta tw) - Hf(x))tw, tw\rangle \\ &= f(x) + \frac{t^2}{2}\langle Hf(x)w, w\rangle + \frac{t^2}{2}\langle (Hf(x+\Theta tw) - Hf(x))w, w\rangle \\ &= f(x) + \frac{t^2\lambda}{2} + \frac{t^2}{2}\langle (Hf(x+\Theta tw) - Hf(x))w, w\rangle \\ &\geq f(x) + \frac{t^2\lambda}{2} - \frac{t^2}{2}\|Hf(x+\Theta tw) - Hf(x)\| \cdot \underbrace{\|w\|^2}_{=1} \end{aligned}$$

für ein $\Theta \in [0,1]$. Wähle nun $\delta > 0$ klein genug, dass für alle $\|w\| \leq 1, |t| \leq \delta$ die Abschätzung $\|Hf(x + t\Theta w) - Hf(x)\| < \lambda/2$ gilt. Für $|t| < \delta$ erhalten wir mit $f(x+tw) \geq f(x) + \frac{1}{4}t^2\lambda > f(x)$ für $t \neq 0$ einen Wert, welcher größer als $f(x)$ ist. Damit hat f kein lokales Maximum in x.

Zu 5. Da $Hf(x)$ einen positiven und einen negativen Eigenwert besitzt, existieren in jeder Umgebung von x zwei Funktionswerte, von denen einer größer und einer kleiner als $f(x)$ ist. Damit hat f in x einen Sattelpunkt.

Zu 6. Wir betrachten als Gegenbeispiele die Funktionen $f_1(x,y) := x^2$ und $f_2(x,y) := x^2 + y^3$. Für jede dieser Funktionen ist der Nullpunkt ein kritischer Punkt. In diesem gilt $Hf_k(0,0) = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ für $k = 1, 2$. Jedoch gilt:

1. f_1 besitzt in 0 ein lokales Minimum.
2. f_2 besitzt in 0 weder ein lokales Minimum noch Maximum. □

Wir geben nun eine Methode an, um die Eigenwerte einer Matrix zu bestimmen:

Definition 4.4.3 (Charakteristisches Polynom). Für $A \in \mathrm{Mat}_n(\mathbb{R})$ heißt

$$\chi_A(\lambda) := \det(\lambda I_n - A)$$

das *charakteristische Polynom* von A.

Bemerkung. Nach Definition ist χ_A die Determinante der Matrix $\lambda I_n - A$, welche nach der Leibniz-Formel ein normiertes Polynom vom Grad n ist.

Satz 4.4.4. *Sei $A \in \mathrm{Mat}_n(\mathbb{R})$. Dann ist $\lambda \in \mathbb{R}$ genau dann ein Eigenwert von A, wenn $\chi_A(\lambda) = 0$.*

Beweis. Die Zahl λ ist genau dann ein Eigenwert, wenn ein $v \in \mathbb{R}^n \setminus \{0\}$ existiert mit $Av = \lambda v = (\lambda I_n)v$, also wenn $(\lambda I_n - A)$ nicht injektiv ist (da 0 durch A immer auf ein Vielfaches von sich abgebildet wird). Dies ist bei einer quadratischen Matrix genau dann der Fall, wenn $\det(\lambda I_n - A)$ verschwindet. □

Beispiel. Zu untersuchen sind lokale Extrema der Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto 6xy - 3y^2 - 2x^3.$$

Zunächst berechnen wir die Jacobi-Matrix

$$Jf(x) = Df(x) = \begin{pmatrix} 6y - 6x^2 \\ 6x - 6y \end{pmatrix}.$$

Die kritischen Punkte von f sind genau jene, in welchen $Df(x)$ verschwindet, in welchen also $6y - 6x^2 = 0$ und $6x - 6y = 0$ gilt. Die zweite Gleichung ergibt $x = y$. Einsetzen in die erste Gleichung ergibt

$$-6x^2 + 6x = 0 \Leftrightarrow x^2 - x = 0 \Leftrightarrow x(x - 1) = 0 \Leftrightarrow x = 0 \vee x = 1.$$

Kritische Punkt von f sind also $(0, 0)$ und $(1, 1)$. Wir prüfen nun die Definitheit der Hesse-Matrix $Hf(x)$ in diesen Punkten: Es ist

$$Hf(0, 0) = \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix} \quad \text{und} \quad Hf(1, 1) = \begin{pmatrix} -12 & 6 \\ 6 & -6 \end{pmatrix}.$$

Deren charakteristische Polynome sind

$$\chi_{Hf(0,0)}(\lambda) = \begin{vmatrix} \lambda & -6 \\ -6 & \lambda + 6 \end{vmatrix} = \lambda^2 + 6\lambda - 36 \quad \text{und} \quad \chi_{Hf(1,1)}(\lambda) = \lambda^2 + 18\lambda + 36.$$

Deren Nullstellen sind

$$\chi_{Hf(0,0)}(\lambda) = 0 \Leftrightarrow \lambda = -3 \pm \sqrt{3^2 + 36} = -3 \pm \sqrt{45} \Rightarrow \lambda \approx -10 \vee \lambda \approx 4,$$

womit $Hf(0, 0)$ indefinit ist und f in $(0, 0)$ einen Sattelpunkt besitzt, und

$$\chi_{Hf(1,1)}(\lambda) = 0 \Leftrightarrow \lambda = -9 \pm \sqrt{9^2 - 36} = -9 \pm \sqrt{45} \Rightarrow \lambda < -2,$$

wonach $Hf(1, 1)$ negativ definit ist und f in $(1, 1)$ folglich ein Maximum besitzt.

5 Der Umkehrsatz und implizite Funktionen

Erinnerung. In Analysis 1 hatten wir folgenden Satz: Ist $I \subseteq \mathbb{R}$ ein offenes Intervall, $f\colon I \to \mathbb{R}$ stetig differenzierbar und $f'(x) \neq 0$ für alle $x \in I$. Dann ist $f(I) \subseteq \mathbb{R}$ ein offenes Intervall und $f\colon I \to f(I)$ bijektiv. Ihre Umkehrfunktion $g\colon f(I) \to I$ ist stetig differenzierbar mit $g'(f(x)) = 1/f'(x)$.

5.1 Diffeomorphismen und der Umkehrsatz

Definition 5.1.1 (Diffeomorphismus). Seien $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$ offen. Als *Diffeomorphismus* bezeichnet man eine Bijektion $f \in C^1(U, V)$, deren Umkehrabbildung $f^{-1}\colon V \to U$ stetig differenzierbar ist.

Proposition 5.1.2. *Sei $V \subseteq \mathbb{R}^m$ offen und $f\colon U \to V$ ein Diffeomorphismus mit Umkehrabbildung $g = f^{-1}\colon V \to U$. Dann gilt:*

1. $m = n$.
2. $Df(x) \in \mathrm{GL}_n(\mathbb{R})$ *für alle* $x \in U$.
3. $Df^{-1}(f(x)) = (Df(x))^{-1} \in \mathrm{GL}_n(\mathbb{R})$.

Beweis. Es ist $f^{-1} \circ f = \mathrm{id}_U$ und $f \circ f^{-1} = \mathrm{id}_V$. Mit der Kettenregel folgt

$$Df^{-1}(f(x)) \cdot Df(x) = I_n \quad \text{und} \quad Df(f^{-1}(y)) \cdot Df^{-1}(y) = I_m.$$

In der rechten Gleichung erhalten wir für $y = f(x)$ die Identität $Df(x) \cdot Df^{-1}(f(x)) = I_m$. Folglich sind $Df^{-1}(f(x))$ und $Df(x)$ invers zueinander und es folgen alle behaupteten Aussagen. □

Satz 5.1.3 (Umkehrsatz). *Sei $f \in C^1(U, \mathbb{R}^n)$ derart, dass $Df(x) \in \mathrm{GL}_n(\mathbb{R})$ für alle $x \in U$. Dann gilt:*

1. *Für alle offenen Teilmengen $V \subseteq U$ ist $f(V) \subseteq \mathbb{R}^n$ offen.*
2. *Zu jedem $x \in U$ existiert eine offene Umgebung $V \subseteq U$ von x, sodass $f|_V\colon V \to f(V)$ bijektiv und $f|_V^{-1}$ stetig differenzierbar ist.*

Definition 5.1.4 (Lokaler Diffeomorphismus). Eine Funktion $f \in C^1(U, \mathbb{R}^n)$, deren Differential auf ganz U invertierbar ist, heißt *lokaler Diffeomorphismus*.

Wir wollen nun einen Beweis des Umkehrsatzes erbringen, für welchen wir zwei Lemmata verwenden werden:

Lemma 5.1.5 (Schrankensatz). *Sei U konvex (das heißt sternförmig bezüglich jedes Punktes $x \in U$) und $f \in C^1(U, \mathbb{R}^m)$. Es gebe ein $L \geq 0$, sodass $\|Df(x)\| \leq L$ für alle $x \in U$. Dann ist f Lipschitz-stetig mit Lipschitz-Konstante L, das heißt $\|f(x) - f(y)\| \leq L\|x - y\|$ für alle $x, y \in U$.*

Beweis. Seien $x, y \in U$. Nach dem Hauptsatz der Differential- und Integralrechnung ist dann $f(y) = f(x) + \int_0^1 Df(x + t(y-x))(y-x)\mathrm{d}t$. Also ist

$$\begin{aligned}
\|f(y) - f(x)\| &= \left\| \int_0^1 Df(x + t(y-x))(y-x)\mathrm{d}t \right\| \\
&\leq \int_0^1 \|Df(x + t(y-x))(y-x)\|\mathrm{d}t \\
&\leq \int_0^1 \underbrace{\|Df(x + t(y-x))\|}_{\leq L} \|y - x\|\mathrm{d}t \leq L\|y - x\|.
\end{aligned}$$ □

Lemma 5.1.6 (Banachscher Fixpunktsatz). *Sei $(X \neq \emptyset, d)$ ein vollständiger metrischer Raum und $f\colon X \to X$ eine Funktion, welche q-Lipschitz-stetig ist, wobei $q < 1$. Dann besitzt f genau einen Fixpunkt $x \in X$.*

Beweis. Zur Existenz: Für $n \in \mathbb{N}_0$ definieren wir $f^n\colon X \to X$ induktiv durch $f^0 := \mathrm{id}_X$ und $f^{n+1} := f \circ f^n$. Sei $x \in X$ beliebig und $x_n := f^n(x)$. Behauptung: Die Folge $(x_n)_{n\in\mathbb{N}_0}$ konvergiert gegen einen Fixpunkt $x_\infty \in X$. Wir zeigen, dass $(x_n)_n$ eine Cauchy-Folge ist. Es gilt die Abschätzung

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq q d(x_{n-1}, x_n) \leq \cdots \leq q^n d(x_0, x_1)$$

und nach der Dreiecksungleichung gilt allgemeiner

$$d(x_n, x_{n+m}) \leq \sum_{k=0}^{m-1} d(x_{n+k}, x_{n+k+1}) \leq \sum_{k=0}^{m-1} q^{n+k} d(x_0, x_1) \xrightarrow{n\to\infty} 0.$$

Da X vollständig ist, konvergiert x_n gegen ein $x_\infty \in X$. Nun gilt

$$\begin{aligned}
d(x_\infty, f(x_\infty)) &= d\left(\lim_{n\to\infty} x_n, f\left(\lim_{n\to\infty} x_n\right)\right) = d\left(\lim_{n\to\infty} x_n, \lim_{n\to\infty} f(x_n)\right) \\
&= \lim_{n\to\infty} d(x_n, f(x_n)) = \lim_{n\to\infty} d(x_n, x_{n+1}) \\
&\leq \lim_{n\to\infty} q^n d(x_0, x_1) = 0,
\end{aligned}$$

da f und d stetig sind. Also ist $f(x_\infty) = x_\infty$ und ein Fixpunkt von f gefunden. Nun zeigen wir die Eindeutigkeit. Seien $x, y \in X$ verschiedene Fixpunkte von f. In diesem Fall erhalten wir mit $d(x, y) = d(f(x), f(y)) < d(x, y)$ einen Widerspruch und es folgt $x = y$. □

Beweis des Umkehrsatzes. Durch gegebenenfalls Verschiebungen und Streckungen (bzw. Stauchungen) in Definitions- und Zielraum können wir ohne Beschränkung der Allgemeinheit annehmen, dass $x = 0$, $f(x) = 0$ und $Df(0) = I_n$ ist. Wir wollen für kleine y die Gleichung $f(x) = y$ lösen. Wir schreiben die Gleichung als $x = y + x - f(x)$, setzen $h_y(x) := y + x - f(x)$ und suchen also einen Fixpunkt von h_y. Um nun den Banachschen Fixpunktsatz anwenden zu können, benötigen wir einen vollständigen metrischen Raum, auf dem h_y definiert ist. Dazu suchen wir einen abgeschlossenen (denn Abgeschlossenheit ist nach Satz 1.4.9 nötig für Vollständigkeit) Ball $\overline{B_r(0)} = \{x \in \mathbb{R}^n \mid \|x\| \leq r\} \subseteq U$, sodass für genügend kleine y die Abbildung h_y eine Abbildung von $\overline{B_r(0)}$ in sich selbst und Lipschitz-stetig mit Lipschitz-Konstante $1/2$ ist. Diesen Ball finden wir wie folgt: Wähle $r > 0$ derart, dass $overlineB_r(0) \subseteq U$ und für alle x mit $\|x\| < r$ die Ungleichung

$\|D(\mathrm{id}-f)(x)\| = \|I_n - Df(x)\| \le 1/2$ erfüllt ist. Dies können wir fordern, da $Df(0) = I_n$ und $x \mapsto Df(x)$ nach Voraussetzung stetig ist. Sei nun $\|y\| \le r/2$ und $\|x\| < r$. Nach dem Schrankensatz (SchS) ist dann

$$\begin{aligned}\|h_y(x)\| &= \|y + x - f(x)\| \le \|y\| + \|x - f(x)\| \\ &\le \|y\| + \|(x - f(x)) - (0 - f(0))\| \overset{\text{SchS}}{\le} \|y\| + \frac{1}{2}\|x - 0\| < r.\end{aligned}$$

Für $\|y\| \le r/2$ gilt also $h_y\left(\overline{B_r(0)}\right) \subseteq \overline{B_r(0)}$. Wir zeigen nun, dass h_y Lipschitz-stetig ist. Seien dafür $\|y\| \le r/2$ und $x, z \in \overline{B_r(0)}$. Mit demselben Argument wie vorher folgt

$$\begin{aligned}\|h_y(x) - h_y(z)\| &= \|y + x - f(x) - y - z + f(z)\| \\ &= \|(x - f(x)) - (z - f(z))\| \overset{\text{SchS}}{\le} \frac{1}{2}\|x - z\|;\end{aligned}$$

Insgesamt folgt: Für $\|y\| \le r/2$ ist $h_y : \overline{B_r(0)} \to \overline{B_r(0)}$ Lipschitz-stetig mit Lipschitz-Konstante $1/2$. Aus dem Banachschen Fixpunktsatz folgt: Es gibt genau einen Punkt $x \in \overline{B_r(0)}$ mit $h_y(x) = x$, also $f(x) = y$ für $\|y\| \le r/2$. Ist sogar $\|y\| < r/2$, so ist $h_y\left(\overline{B_r(0)}\right) \subseteq B_r(0)$; deshalb muss der Fixpunkt x von h_y sogar in $B_r(0)$ liegen. Nun können wir eine offene Umgebung $V \subseteq U$ von 0 angeben, nämlich $V := \{x \in U \mid \|x\| < r \wedge \|f(x)\| < r/2\}$. Wir haben bereits durch Fixpunktberechnung von h_y gezeigt, dass $f : V \to B_{r/2}(0)$ bijektiv ist. Sei $g\colon B_{r/2}(0) \to V$ die Umkehrabbildung. Der Satz ist bewiesen, wenn wir $g \in C^1(B_{r/2}(0), V)$ zeigen können. Seien $x, y \in B_r(0)$. Nach Proposition 5.1.2 muss dann $Dg(0) = I_n$ gelten. Wir zeigen, dass

$$\lim_{h\to 0} \frac{g(h) - g(0) - \mathrm{id}(h)}{\|h\|} = \lim_{h\to 0} \frac{g(h) - h}{\|h\|} = 0$$

für $h \in B_{r/2}(0)$. Da h_0 Lipschitz-stetig mit Konstante $1/2$ ist, folgt

$$\begin{aligned}\|x - z\| &= \|h_0(x) + f(x) - h_0(z) - f(z)\| \\ &\le \|h_0(x) - h_0(z)\| + \|f(x) - f(z)\| \\ &\le \frac{1}{2}\|x - z\| + \|f(x) - f(z)\|, \\ \Leftrightarrow \|x - z\| &\le 2\|f(x) - f(z)\|.\end{aligned}$$

Da $f(0 + h) = f(0) + \mathrm{id}(h) + r(h)$ mit $\lim_{h\to 0} \frac{r(h)}{\|h\|} = 0$ gilt $\lim_{h\to 0} \frac{f(h)-h}{\|h\|} = 0$ und es folgt

$$\lim_{h\to 0} \frac{\|g(h) - h\|}{\|h\|} \le \lim_{h\to 0} \frac{\|f(g(h)) - f(h)\|}{\|h\|} = \lim_{h\to 0} \frac{\|h - f(h)\|}{\|h\|} = 0.$$

Aufgrund der Definitheit der Norm gilt dann auch $\lim_{h\to 0} \frac{g(h)-h}{\|h\|} = 0$, womit die Differenzierbarkeit von g gezeigt ist. Die Stetigkeit der Ableitung von g ergibt sich aus der Stetigkeit der Inversenbildung. □

Korollar 5.1.7. *Sei $f \in C^1(U, \mathbb{R}^n)$ eine Bijektion, deren Differential $Df(x)$ für alle $x \in U$ invertierbar ist. Dann ist f ein Diffeomorphismus.*

Beweis. Folgt aus dem Umkehrsatz. □

Beispiel. Wir zeigen, dass die komplexe Exponentialfunktion exp ein Diffeomorphismus

$$S_1 := (-\infty, \infty) \times (0, 2\pi) \to \mathbb{C}^- := \mathbb{R}^2 \setminus \{(x, 0) \mid x \ge 0\}$$

ist. Dazu fassen wir eine komplexe Zahl z als $z = (x, y) \in \mathbb{R}^2$ auf. Unter dieser Identifikation ist

$$\exp(z) = e^x \begin{pmatrix} \cos y \\ \sin y \end{pmatrix}.$$

Wir erkennen: Der Wert x stammt aus dem Intervall $(-\infty, \infty)$, wird durch e^x also auf das Intervall $(0, \infty)$ abgebildet; der Wert y stammt aus dem Intervall $(0, 2\pi)$. Die Bijektivität ergibt sich damit aus folgender Überlegung: Jede komplexe Zahl lässt sich schreiben als $z = r(\cos y, \sin y)$, wobei diese Konstruktion informell folgendermaßen funktioniert: »Man läuft mit $(\cos y, \sin y)$ um den Winkel y auf dem Einheitskreis entlang und streckt/staucht den resultierenden Vektor mit dem Faktor r, um den gewünschten Punkt in der komplexen Zahlenebene zu treffen.« Nun ist die Bijektivität von $\exp$ klar: Wegen $e^x \in (0, \infty)$ können wir jeden Einheitskreisvektor beliebig strecken und stauchen. Da $y \in (0, 2\pi)$, ist die Abbildung $(0, 2\pi) \to S^1, y \mapsto (\cos y, \sin y)$ injektiv. Wegen $y \neq 0, 2\pi$ treffen wir also jeden Vektor im $\mathbb{R}^2$ außer den positiven Abschnitt der x-Achse, also genau $\mathbb{C}^-$ und dies, wie beschrieben, auf injektive Weise. Nun berechnen wir die partiellen Ableitungen. Diese sind:

$$\partial_1 \exp(x, y) = e^x \begin{pmatrix} \cos y \\ \sin y \end{pmatrix} \quad \text{und} \quad \partial_2 \exp(x, y) = e^x \begin{pmatrix} -\sin y \\ \cos y \end{pmatrix}.$$

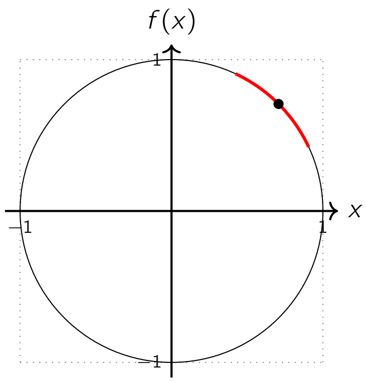
Da die partiellen Abbildungen stetig sind, ist $\exp$ stetig differenzierbar und es ist $D\exp(x, y) = e^x \begin{pmatrix} \cos y & -\sin y \\ \sin y & \cos y \end{pmatrix}$ sowie

$$\det D\exp(x, y) = e^{2x}(\cos^2 \phi + \sin^2 \phi) = e^{2x} \neq 0,$$

also ist $D(x, y)$ invertierbar und damit $\exp$ ein Diffeomorphismus.

5.2 Satz über implizite Funktionen

Motivation. Es sei eine Funktion $f\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k \ni z$ gegeben. Wir stellen uns die Frage, ob man in einer Umgebung von x die die Gleichung $f(x, y) = z$ eindeutig nach y auflösen kann. Anders ausgedrückt: Gibt es eine Funktion $h\colon \mathbb{R}^n \to \mathbb{R}^k$, die in Abhängigkeit von x explizit ein y ausgibt, welches $(x, h(x)) = z$ erfüllt?



Der Einheitskreis ist gegeben durch die Menge $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$, also das Urbild von 1 unter der Funktion $f(x, y) = x^2 + y^2$. Für keinen Punkt $x \in (-1, 1)$ lässt sich y als eindeutige Weise in Abhängigkeit von x schreiben, da sets $\pm\sqrt{1 - x^2}$ zwei Lösungen von $f(x, y) = 1$ sind. Um den markierten Punkt können wir jedoch eine Umgebung anlegen, sodass der Graph von f lokal aussieht wie eine Funktion, welche nur von einer Variablen abhängt. Dort ist eindeutig $y = \sqrt{1 - x^2}$, das heißt, wir können in der Umgebung die Funktion eindeutig nach y auflösen. Der Satz über die implizite Funktion gibt uns ein Kriterium, in welchen Fällen dies allgemein funktioniert.

Satz 5.2.1 (Satz über die implizite Funktion). *Seien $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^k$ offen, $f \in C^1(U \times V, \mathbb{R}^k)$ mit Komponentenfunktionen $f_1, \dots, f_k$, $(x_0, y_0) \in U \times V$, $z := f(x_0, y_0)$. Das totale Differential von f im Punkt*

$a := (x_0, y_0)$ zerfällt in zwei Teilmatrizen

$$\left(\begin{array}{ccc|ccc} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) & \frac{\partial f_1}{\partial y_1}(a) & \cdots & \frac{\partial f_1}{\partial y_m}(a) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) & \frac{\partial f_m}{\partial y_1}(a) & \cdots & \frac{\partial f_m}{\partial y_m}(a) \end{array}\right) =: (\partial_1 f(a) \mid \partial_2 f(a)).$$

Ist $\partial_2 f(a) \in \mathrm{GL}_k(\mathbb{R})$, dann gibt es offene Umgebungen $x_0 \in U_0 \subseteq U$ bzw. $y_o \in V_0 \subseteq V$ und eine stetig differenzierbare Funktion $h\colon U_0 \to V_0$ mit

1. *$h(x_0) = y_0$.*
2. *Für alle $x \in U_0$ gilt $f(x, h(x)) = z$.*

Wenn $f(x, y) = z$ für alle $x \in U_0, y \in V_0$ gilt, dann ist $y = h(x)$.

Differenzieren der Identität $Df(x, h(x)) = z$ nach x zeigt

$$\underbrace{D_1 f(x, h(x))}_{\mathrm{Mat}_{k,n}(\mathbb{R})} + \underbrace{D_2 f(x, h(x))}_{\in \mathrm{Mat}_k(\mathbb{R})} \cdot \underbrace{Dh(x)}_{\in \mathrm{Mat}_{k,n}(\mathbb{R})} = 0.$$

Dies ist äquivalent zu

$$Dh(x) = -(D_2 f(x, h(x)))^{-1} \cdot D_1 f(x, h(x)),$$

insbesondere ist $Dh(x_0) = -(D_2 f(x_0, y_0))^{-1} \cdot D_1 f(x_0, y_0)$.

Beweis. Da $D_2 f$ stetig und im Punkt (x_0, y_0) invertierbar ist, ist allgemein $D_2 f(x, y)$ invertierbar, falls x nahe bei x_0 und y nahe bei y_0. Es gelte also ohne Beschränkung der Allgemeinheit $D_2 f(x, y) \in \mathrm{GL}_k(\mathbb{R})$ für alle $x \in U, y \in V$ (dies erreichen wir durch Verkleinerung von U und V). Wir definieren eine Hilfsfunktion

$$g\colon U \times V \to \mathbb{R}^n \times \mathbb{R}^k, (x, y) \mapsto (x, f(x, y)).$$

Deren Ableitung ist die Blockmatrix

$$Dg(x, y) = \begin{pmatrix} I_n & 0 \\ D_1 f(x, y) & D_2 f(x, y) \end{pmatrix}$$

mit Determinante $\det Dg(x, y) = \det(I_n \cdot D_2 f(x, y)) = \det D_2 f(x, y) \neq 0$. Nun wenden wir den Umkehrsatz an und es gilt: Es existieren offene Umgebungen $x_0 \in U_1 \subseteq U$ und $y_0 \in V_1 \subseteq V$, sodass $g\colon U_1 \times V_1 \to g(U_1 \times V_1)$ ein Diffeomorphismus. Wähle $U_0, x_0 \in U_0 \subseteq U_1$, sodass $U_0 \times \{z\} \subseteq g(U_1 \times V_1)$. Dies geht, weil $g(x_0, y_0) = (x_0, z)$. Seien

$$\begin{aligned} p_k &\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k, (x, y) \mapsto y \quad \text{und} \\ p_n &\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n, (x, y) \mapsto x \end{aligned}$$

Projektionen. Definiere $h\colon U_0 \to V_1$ durch

$$h(x) := p_k \underbrace{\left(g^{-1} \underbrace{(x, z)}_{\in U_0 \times \{z\}} \right)}_{\in U_1 \times V_1} \in V_1$$

Dann ist h stetig differenzierbar, da $g^{-1} \in C^k$ und $p_k \in C^\infty$. Wir reichnen nun nach, dass h die geforderten Bedingungen erfüllt. Es ist

$$h(x_0) = p_k \left(g^{-1} \underbrace{(x_0, z)}_{= g(x_0, z)} \right) = p_k(x_0, y_0) = y_0.$$

Ferner ist

$$\begin{aligned} f(x,h(x)) &= f\left((x,p_k\left(g^{-1}(x,z)\right)\right) = f\left(p_n\left(g^{-1}(x,z)\right), p_k\left(g^{-1}(x,z)\right)\right) \\ &= f\left(g^{-1}(x,z)\right) = (p_k \circ g)\left(g^{-1}(x,z)\right) = p_k(x,z) = z. \end{aligned}$$

Damit besitzt h die geforderten Eigenschaften. Wir zeigen nun die Eindeutigkeit von h. Setze $V_0 := V_1$. Falls $x \in U_0, y \in V_0, f(x,y) = z$. Dann ist $g(x,y) = (x,z) \in U_0 \times \{z\} \subseteq g(U_1 \times V_0)$, also $(x,y) = g^{-1}(x,z)$, also $x = pr_k\left(g^{-1}(x,z)\right) = h(x)$ ☐

Beispiel. Dieses Beispiel behandelt die Nullstellen von Polynomen. Es sei $f\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, (x,y) \mapsto y^n + \sum_{i=0}^{n-1} x_{n-i} y^i$. Für festes x ist $f_x(y) := f(x,y)$ ein Polynom. Sei $(x_0,y_0) \in \mathbb{R}^n \times \mathbb{R}$ derart, dass $f(x_0,y_0) = f_{x_0}(y_0) = 0$. Ist h wie in 5.2.1, so ist $h(x)$ eine Nullstelle von f_x. Nebenbei ist

$$D_2 f(x,y) = n y^{n-1} + \sum_{i=1}^{n} -1 x_{n-i} i y^{i-1} = f_x'(y).$$

Falls $f_{x_0}(y_0) = 0$ und $f_{x_0}'(y_0) \neq 0$, so gibt es eine Funktion $h\colon U \to \mathbb{R}$ (wobei $U \subseteq \mathbb{R}^n$ eine offene Umgebung von x_0 ist) mit der Eigenschaft, dass $h(x_0) = y_0$ und $f(x,h(x)) = f_x(h(x)) = 0$. Also folgt: Ist $f_{x_0}'(y_0) \neq 0$, so hängen die Nullstellen von f_x differenzierbar von x ab. Dabei ist $f_{x_0}'(y_0) = 0$ genau dann, wenn y_0 eine einfache Nullstelle von f_{x_0} ist.

Für $n = 2$ haben wir $f(x_1,x_2,y) = y^2 + x_1 y + x_2$. Deren Nullstellen sind $-x_1/2 \pm \sqrt{x_1^2/4 - x_2}$. Diese Funktion ist differenzierbar, falls $x_1^2/4 - x_2 > 0$.

5.3 Geometrische Version des Satzes über implizite Funktionen

Definition 5.3.1. Sei $U \subseteq \mathbb{R}^m$ offen und $f \in C^1(U,\mathbb{R}^k)$. Ein Punkt $x \in U$ heißt *regulärer Punkt* von f, falls $\operatorname{rk} Df(x) = k$; das heißt $Df(x)$ stellt einen Epimorphismus (eine surjektive lineare Abbildung) $\mathbb{R}^m \to \mathbb{R}^n$ darstellt. Ist x kein regulärer Punkt, so heißt er *singulär* oder *kritisch*. Ein $y \in \mathbb{R}^k$ heißt *regulärer Wert* von f, falls jedes Urbild von y ein regulärer Punkt ist.

Bemerkungen.

1. Für $k = 1$ stimmt die Definition eines kritischen Punktes mit Definition 4.2.3 aus dem Kapitel »Extremwertprobleme« überein.
2. In der Definition eines regulären Wertes ist nicht gefordert, dass y überhaupt im Bild von f liegt. Dies hat zur Folge: Ist $f^{-1}(y) = \emptyset$, so ist y ein regulärer Wert von f.
3. Ist $k > m$, so ist $y \in \mathbb{R}^k$ genau dann ein regulärer Wert von f, wenn $y \notin f(U)$.
4. Ist $k = m$, so ist $x \in U$ genau dann ein regulärer Punkt von f, wenn $Df(x) \in \mathrm{GL}_n(\mathbb{R})$.

Satz 5.3.2. *Sei $f \in C^1(U,\mathbb{R}^k)$ und $y_0 \in \mathbb{R}^k$ ein regulärer Wert von f mit Urbild $M := f^{-1}(y_0) \subseteq U$. Dann existiert zu jedem $x_0 \in M$ eine offene Umgebung $U_0 \subseteq U$ von x_0 und ein Diffeomorphismus $h\colon U_0 \to h(U_0) \subseteq \mathbb{R}^m$, sodass $h(M \cap U_0) = \{0\}^k \times \mathbb{R}^{m-k} \subseteq \mathbb{R}^m$.*

Beispiel. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}, x \mapsto \langle x,x\rangle = \|x\|_2^2$. Deren Ableitung $Df(x)v = 2\langle x,v\rangle$ ist surjektiv genau dann, wenn $x \neq 0$ ist. Damit ist jeder Punkt $x \in \mathbb{R}^n \setminus \{0\}$ ein regulärer Punkt von f, insbesondere auch $x = 1$. Dessen Urbild ist $f^{-1}(1) = S^{n-1}$

Beweis von Satz 5.3.2. Sei $x_0 \in M$ und $Df(x_0)\colon \mathbb{R}^m \to \mathbb{R}^k$ surjektiv. Wähle eine Basis $(v_1,\dots,v_m)$ von $\mathbb{R}^m$ derart, dass

$$Df(x_0)v_i = \begin{cases} e_i, & i \le k \\ 0, & i > k \end{cases}.$$

Eine solche Basis existiert aus folgendem Grund: Nach der Dimensionsformel ist Es ist $\dim \ker Df(x_0) = m-k$. Sei $(v_{k+1}, \dots, v_m)$ eine Basis von $\ker Df(x_0)$. Weil $Df(x_0)$ surjektiv ist, gibt es $v_1, \dots, v_k$ mit $Df(x_0)v_i = e_i$ für $1 \le i \le k$. Dann ist $(v_1, \dots, v_m)$ eine Basis von $\mathbb{R}^m$, da sie linear unabhängig sind. Sei $p\colon \mathbb{R}^m \to \langle e_{k+1}, \dots, e_m \rangle = \{0\}^k \times \mathbb{R}^{m-k} \subseteq \mathbb{R}^m$ die lineare Abbildung

$$p(v_i) = \begin{cases} 0, & i \le k \\ e_i, & i > k \end{cases}.$$

Setze $h\colon U \to \mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^{m-k}, x \mapsto (f(x) - y_0, p(x))$. Diese Abbildung ist stetig differenzierbar. Nun ist $Dh(x_0) \in \mathrm{Mat}_m(\mathbb{R})$ ist invertierbar, denn:

$$Dh(x_0)v_i = (Df(x_0)v_i, p(v_i)) = \begin{cases} (e_i, 0), & i \le k \\ (0, e_i), & i > k \end{cases}.$$

Damit bildet $Dh(x_0)$ eine Basis auf eine Basis ab, ist damit ein Isomorphismus, also invertierbar. Weil $x \mapsto Dh(x)$ stetig ist, gibt es eine Umgebung $U_1 \subseteq U$ von x_0 derart, dass $Dh(x) \in \mathrm{GL}_n(\mathbb{R})$ für alle $x \in U_1$. Nach dem Umkehrsatz existiert eine offene Umgebung $U_0 \subseteq U_1$ von x_0 und $h\colon U_0 \to h(U_0)$ ein Diffeomorphismus ist. Wir zeigen nun $h(M \cap U_0) = \{0\}^k \times \mathbb{R}^{m-k}$. Für $x \in U_0 \cap M$ ist $f(x) = y_0$ und $h(x) = (f(x) - y_0, p(x)) = (0, p(x)) \in \{0\}^k \times \mathbb{R}^{m-k}$. Ist andererseits $h(x) \in \{0\}^k \times \mathbb{R}^{m-k}$, so ist $h(x) - y_0 = 0$, also $x \in M \cap U_0$. □

Bemerkung. Man kann zeigen: Ist $A \subseteq \mathbb{R}^n$ abgeschlossen, so gibt es eine C^∞-Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ mit $f^{-1}(0) = A$.

Definition 5.3.3. Eine Teilmenge $M \subseteq \mathbb{R}^{n+k}$ heißt *n-dimensionale Untermannigfaltigkeit* von $\mathbb{R}^{n+k}$, falls gilt: Zu jedem Punkt $x_0 \in M$ existiert eine offene Umgebung $x_0 \in U \subseteq \mathbb{R}^{n+k}$ und ein Diffeomorphismus $h\colon U \to h(U) \subseteq \mathbb{R}^{n+k}$ mit $h(U \cap M) = \{0\}^k \times \mathbb{R}^n \subseteq \mathbb{R}^{n+k}$. Ein solches Paar (U, h) heißt *Karte* von M. Eine Familie $((U_i, h_i))_{i \in I}$ von Karten heißt *Atlas* von M, falls $\bigcup_{i \in I} U_i \cap M = M$

Beispiele.

1. Sei $U \subseteq \mathbb{R}^m$ und $f \in C^1(U, \mathbb{R}^m)$. Ist $y_0 \in \mathbb{R}^k$ ein regulärer Wert, dann ist $f^{-1}(y_0) \subseteq \mathbb{R}^m$ ist eine Untermannigfaltigkeit der Dimension $m - k$; insbesondere ist S^{n-1} eine $(n-1)$-dimensionale Untermannigfaltigkeit.
2. Eine offene Menge $U \subseteq \mathbb{R}^n$ ist eine n-dimensionale Untermannigfaltigkeit von $\mathbb{R}^n$.