Analysis II *

Martin Brokate †

Inhaltsverzeichnis

*Vorlesungsskript, SS 2000

†Zentrum Mathematik, TU München

1 Potenzreihen, Taylorreihen

Für diesen Abschnitt vereinbaren wir wieder: $\mathbb{K}$ steht für den Körper $\mathbb{R}$ oder $\mathbb{C}$, d.h. eine Aussage, in der $\mathbb{K}$ auftaucht, steht als Abkürzung für die beiden entsprechenden Aussagen mit $\mathbb{K} = \mathbb{R}$ und $\mathbb{K} = \mathbb{C}$.

Satz 1.1 (Weierstraß-Kriterium)
Sei $D \subset \mathbb{K}$, sei $(f_k)_{k\geq 0}$ eine Folge von Funktionen in $B(D;\mathbb{K})$, es gelte

$$\sum_{k=0}^{\infty} \|f_k\|_\infty < \infty\,. \tag{1.1}$$

Dann ist für alle $x \in D$ die Reihe

$$\sum_{k=0}^{\infty} f_k(x) \tag{1.2}$$

absolut konvergent, und die Partialsummen $s_n : D \to \mathbb{K}$,

$$s_n = \sum_{k=0}^{n} f_k\,, \tag{1.3}$$

konvergieren gleichmäßig gegen $f : D \to \mathbb{K}$,

$$f(x) = \sum_{k=0}^{\infty} f_k(x)\,. \tag{1.4}$$

Beweis: Für alle $x \in D$ gilt $|f_k(x)| \leq \|f_k\|_\infty$, also auch

$$\sum_{k=0}^{n} |f_k(x)| \leq \sum_{k=0}^{n} \|f_k\|_\infty \leq \sum_{k=0}^{\infty} \|f_k\|_\infty < \infty\,,$$

also ist (1.2) absolut konvergent für alle $x \in D$. Sei nun $\varepsilon > 0$. Aus (1.1) folgt, daß es ein $n_0 \in \mathbb{N}$ gibt mit

$$\sum_{k=n_0+1}^{\infty} \|f_k\|_\infty < \varepsilon\,.$$

Dann gilt für alle $n \geq n_0$ und alle $x \in D$

$$|s_n(x) - f(x)| = \left|\sum_{k=n+1}^{\infty} f_k(x)\right| \leq \sum_{k=n+1}^{\infty} |f_k(x)| \leq \sum_{k=n+1}^{\infty} \|f_k\|_\infty < \varepsilon\,.$$

Also konvergiert s_n gleichmäßig gegen f. □

Beispiel: Für $f_k : \mathbb{R} \to \mathbb{R}$,

$$f_k(x) = \frac{\cos kx}{k^2}\,,$$

gilt

$$\|f_k\|_\infty \leq \frac{1}{k^2}\,,$$

also
$$\sum_{k=1}^{\infty} \|f_k\|_\infty \le \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty\,,$$
also ist Satz 1.1 anwendbar und damit die Reihe
$$\sum_{k=1}^{\infty} \frac{\cos kx}{k^2}$$
absolut und in $\mathbb{R}$ gleichmäßig konvergent.

Potenzreihen. Wir kennen bereits $\exp : \mathbb{C} \to \mathbb{C}$,
$$\exp(z) = e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}\,. \tag{1.5}$$
(1.5) ist ein Beispiel einer Potenzreihe. Die allgemeine Form einer Potenzreihe mit Entwicklungspunkt $a \in \mathbb{C}$ ist
$$\sum_{k=0}^{\infty} c_k (z-a)^k\,, \tag{1.6}$$
wobei $c_k \in \mathbb{C}$ gegebene Koeffizienten sind. In (1.5) ist
$$a = 0\,, \quad c_k = \frac{1}{k!}\,.$$
Es stellt sich die Frage: Für welche $z \in \mathbb{C}$ ist (1.6) konvergent, oder anders formuliert, was ist der Definitionsbereich der durch
$$f(z) = \sum_{k=0}^{\infty} c_k (z-a)^k \tag{1.7}$$
definierten Funktion?

Definition 1.2 *Sei $a \in \mathbb{C}$, $r > 0$. Dann heißt*
$$B(a,r) = \{z : z \in \mathbb{C},\ |z-a| < r\}$$
die offene Kugel um a mit Radius r, und
$$K(a,r) = \{z : z \in \mathbb{C},\ |z-a| \le r\}$$
die abgeschlossene Kugel um a mit Radius r.

Man sieht leicht, daß
$$K(a,r) = \overline{B(a,r)}\,.$$

Satz 1.3 *Sei (c_k) Folge in $\mathbb{C}$, sei $a \in \mathbb{C}$, sei $z_1 \in \mathbb{C}$ und*
$$\sum_{k=0}^{\infty} c_k (z_1 - a)^k$$

sei konvergent. Dann wird durch

$$f(z) = \sum_{k=0}^{\infty} c_k (z-a)^k \tag{1.8}$$

eine Funktion $f : B(a,r) \to \mathbb{C}$, $r = |z_1 - a|$, *definiert, die Konvergenz in (1.8) ist absolut, und für jedes* $\rho < r$ *konvergieren die Partialsummen*

$$s_n(z) = \sum_{k=0}^{n} c_k (z-a)^k \tag{1.9}$$

gleichmäßig gegen f *auf* $K(a,\rho)$. *Für*

$$g(z) = \sum_{k=1}^{\infty} k c_k (z-a)^{k-1} \tag{1.10}$$

gilt ebenfalls, daß die Konvergenz auf $B(a,r)$ *absolut ist und daß für alle* $\rho < r$ *die entsprechenden Partialsummen gleichmäßig gegen* g *auf* $K(a,\rho)$ *konvergieren.*

Beweis: Wir setzen $f_k(z) = c_k(z-a)^k$, dann ist $(f_k(z_1))$ beschränkt, da die Reihe in z_1 konvergiert. Sei

$$|f_k(z_1)| \le M$$

für alle k. Sei $\rho < r$. Dann gilt für alle k und alle $z \in K(a,\rho)$

$$|f_k(z)| = |c_k|\,|z-a|^k = |c_k(z_1-a)^k| \frac{|z-a|^k}{|z_1-a|^k} \le M \left(\frac{\rho}{|z_1-a|} \right)^k ,$$

also

$$\sup_{z \in K(a,\rho)} |f_k(z)| \le M\theta^k , \quad \theta = \frac{\rho}{|z_1-a|} < 1 ,$$

und damit (Supremumsnorm auf $K(a,\rho)$)

$$\sum_{k=0}^{\infty} \|f_k\|_\infty \le M \frac{\theta}{1-\theta} < \infty .$$

Aus Satz 1.1 folgt die gleichmäßige Konvergenz von s_n gegen f auf $K(a,\rho)$. Da $\rho < r$ beliebig ist, folgt aus Satz 1.1 die absolute Konvergenz der Reihe auf

$$B(a,r) = \bigcup_{\rho < r} K(a,\rho) .$$

Wir definieren nun

$$g_k(z) = k c_k (z-a)^{k-1} .$$

Für $z \in K(a,\rho)$, $\rho < r$, folgt

$$|g_k(z)| = k|c_k(z_1-a)^k| \frac{1}{|z_1-a|} \frac{|z-a|^{k-1}}{|z_1-a|^{k-1}} \le k \frac{M}{|z_1-a|} \theta^{k-1} .$$

Die Reihe

$$\sum_{k=1}^{\infty} k \frac{M}{|z_1 - a|} \theta^{k-1}$$

ist also eine (nach dem Quotientenkriterium) konvergente Majorante von

$$\sum_{k=1}^{\infty} \|g_k\|_{\infty} ,$$

also folgen die Behauptungen für g ebenfalls aus Satz 1.1. □

Definition 1.4 (Konvergenzradius)
Sei (c_k) Folge in $\mathbb{C}$, $a \in \mathbb{C}$. Dann heißt

$$r = \sup\{|z - a| : z \in \mathbb{C}, \sum_{k=0}^{\infty} c_k (z-a)^k \text{ ist konvergent}\} \tag{1.11}$$

der Konvergenzradius der durch (c_k) und a definierten Potenzreihe

$$\sum_{k=0}^{\infty} c_k (z-a)^k . \tag{1.12}$$

Folgerung 1.5 *Sei r der Konvergenzradius der Potenzreihe*

$$\sum_{k=0}^{\infty} c_k (z-a)^k . \tag{1.13}$$

Dann konvergiert (1.13) absolut auf $B(a, r)$, und durch

$$f(z) = \sum_{k=0}^{\infty} c_k (z-a)^k \tag{1.14}$$

wird eine stetige Funktion $f : B(a, r) \to \mathbb{C}$ definiert. Die Partialsummen konvergieren gleichmäßig gegen f auf $K(a, \rho)$ für alle $\rho < r$.

Beweis: Für jedes $\rho < r$ gibt es ein $z_1 \in \mathbb{C}$ mit $\rho < |z_1 - a| \le r$, so daß (1.13) für $z = z_1$ konvergiert. Die Behauptung folgt nun aus Satz 1.3; f ist stetig als gleichmäßiger Limes der Folge der (stetigen) Partialsummen. □

Bemerkung 1.6

1. Es gilt außerdem, daß die Potenzreihe (1.13) für alle $z \in \mathbb{C}$ mit $|z - a| > r$ divergiert. Diese Aussage beweist man, indem man analog zum Beweis von Satz 1.3 eine divergente Minorante der Form $\sum_k Mq^k$ mit $q > 1$ konstruiert.

2. Der Konvergenzradius einer Potenzreihe kann 0 sein. Beispiel:

$$\sum_{k=0}^{\infty} k! z^k .$$

Satz 1.7 (Wurzelkriterium)
Sei $\sum_{k=n_0}^{\infty} a_k$ Reihe in $\mathbb{C}$, sei

$$w = \limsup_{k\to\infty} \sqrt[k]{|a_k|}\,. \tag{1.15}$$

Ist $w < 1$, so konvergiert die Reihe; ist $w > 1$, so divergiert sie.

Beweis: Ist $w > 1$, so gibt es unendlich viele k mit $|a_k| > 1$, und (a_k) ist keine Nullfolge. Sei nun $w < 1$. Wähle q mit $w < q < 1$, dann gibt es N mit $|a_k| \le q^k$ für alle $k \ge N$, also konvergiert die Reihe. □

Folgerung 1.8 *Sei (c_k) Folge in $\mathbb{C}$, $a \in \mathbb{C}$. Der Konvergenzradius r der Potenzreihe $\sum_{k=0}^{\infty} c_k(z-a)^k$ ist*

$$r = \frac{1}{\limsup_{k\to\infty} \sqrt[k]{|c_k|}}\,. \tag{1.16}$$

Beweis: Wir wenden Satz 1.7 an mit $a_k = c_k(z-a)^k$. Es gilt

$$w < 1 \quad\Leftrightarrow\quad |z-a| < \frac{1}{\limsup_{k\to\infty} \sqrt[k]{|c_k|}}\,, \tag{1.17}$$

$$w > 1 \quad\Leftrightarrow\quad |z-a| > \frac{1}{\limsup_{k\to\infty} \sqrt[k]{|c_k|}}\,, \tag{1.18}$$

also folgen die Behauptungen. □

Satz 1.9 *Sei (c_k) Folge in $\mathbb{R}$, $a \in \mathbb{R}$,*

$$\sum_{k=0}^{\infty} c_k(x-a)^k \tag{1.19}$$

Potenzreihe mit Konvergenzradius $r > 0$. Dann ist die durch

$$f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k \tag{1.20}$$

definierte Funktion $f : (a-r, a+r) \to \mathbb{R}$ stetig differenzierbar in $(a-r, a+r)$, und es gilt

$$f'(x) = \sum_{k=1}^{\infty} kc_k(x-a)^{k-1}\,. \tag{1.21}$$

Beweis: Wir wenden Satz 1.3 und Satz 14.32 (Analysis I) an. □

Satz 1.9 bedeutet, daß wir die Ableitung einer durch eine Potenzreihe definierte Funktion durch gliedweises Differenzieren erhalten können.

Beispiel 1.10
Die die Funktion

$$f(x) = \sum_{k=0}^{\infty} x^k$$

definierende Potenzreihe ist konvergent für $|x| < 1$, divergent für $|x| > 1$, hat also Konvergenzradius $r = 1$. Für $|x| < 1$ gilt

$$f(x) = \frac{1}{1-x},$$

also nach Satz 1.9

$$f'(x) = \sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

Hieraus erhalten wir die Formel

$$\sum_{k=1}^{\infty} kx^k = x \sum_{k=1}^{\infty} kx^{k-1} = \frac{x}{(1-x)^2}.$$

□

Satz 1.11 *Sei (c_k) Folge in $\mathbb{R}$, $a \in \mathbb{R}$,*

$$\sum_{k=0}^{\infty} c_k(x-a)^k \tag{1.22}$$

Potenzreihe mit Konvergenzradius $r > 0$. Dann ist die durch

$$f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k \tag{1.23}$$

definierte Funktion $f : (a-r, a+r) \to \mathbb{R}$ unendlich oft differenzierbar in $(a-r, a+r)$, und es gilt

$$c_k = \frac{1}{k!} f^{(k)}(a) \tag{1.24}$$

für alle $k \in \mathbb{N} \cup \{0\}$.

Beweis: Wir zeigen mit Induktion über k: f ist k-mal differenzierbar in $(a-r, a+r)$, und $f^{(k)}$ ist gegeben durch die in $(a-r, a+r)$ konvergente Potenzreihe

$$f^{(k)}(x) = \sum_{m=k}^{\infty} m(m-1)\cdots(m-k+1)c_m(x-a)^{m-k}. \tag{1.25}$$

Für $k = 1$ gilt (1.25) nach Satz 1.9. Gilt die Behauptung für k, so können wir Satz 1.9 auf $f^{(k)}$ anwenden und erhalten die Formel (1.25) für $k+1$ durch gliedweises Differenzieren. Die Formel (1.24) folgt, indem wir $x = a$ in (1.25) setzen. □

Für Potenzreihen

$$f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k \tag{1.26}$$

mit Konvergenzradius $r > 0$ gilt also

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(a)(x-a)^k, \quad x \in (a-r, a+r). \tag{1.27}$$

Definition 1.12 (Taylorreihe)
Sei $a \in \mathbb{R}$, sei $I \subset \mathbb{R}$ offenes Intervall mit $a \in I$, sei $f \in C^\infty(I)$. Ist $r > 0$ der Konvergenzradius der Potenzreihe

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \,, \tag{1.28}$$

so heißt die durch

$$T_f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \tag{1.29}$$

definierte Funktion $T_f : (a-r, a+r) \to \mathbb{R}$ die Taylorreihe von f mit Entwicklungspunkt a.

Bemerkung 1.13

1. Der Konvergenzradius der Potenzreihe (1.28) kann 0 sein. (Dann hat T_f ein leeres Definitionsgebiet, ist also nicht definiert.)

2. Ist f durch eine Potenzreihe definiert, so gilt nach Satz 1.11

   $$f = T_f \,.$$

   So ist etwa

   $$\exp(x) = T_{\exp}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \,.$$

3. Für $f \in C^\infty(I)$ ist es möglich, daß $f \neq T_f$. Beispiel:

   $$f(x) = \begin{cases} \exp\left(-\frac{1}{x^2}\right) \,, & x \neq 0 \,, \\ 0 \,, & x = 0 \,. \end{cases}$$

   Es ist $f^{(k)}(0) = 0$ für alle $k \in \mathbb{N}$ (siehe Übungsaufgabe), also ist $T_f = 0$.

□

Die Approximation einer Funktion f durch die Partialsummen ihrer Taylorreihe

$$T_f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_{n+1}(x) \tag{1.30}$$

mit einem Restglied R_{n+1} ist ein wichtiges Werkzeug der Analysis. Sie ist bereits dann sinnvoll, wenn nur die ersten n bzw. $n+1$ Ableitungen von f existieren.

Definition 1.14 (Taylorpolynom)
Sei $I \subset \mathbb{R}$ Intervall, $f \in C^n(I)$ und $a \in I$. Dann heißt das durch

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \tag{1.31}$$

definierte Polynom das n-te Taylorpolynom von f in a.

Satz 1.15 (Taylorentwicklung)
Sei $I \subset \mathbb{R}$ Intervall, $f \in C^{n+1}(I)$ und $a \in I$. Dann gilt für alle $x \in I$

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_{n+1}(x) \tag{1.32}$$

mit

$$R_{n+1}(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t)\, dt\,. \tag{1.33}$$

Beweis: Induktion über n. Für $n = 0$ wird die Behauptung zu

$$f(x) = f(a) + \int_a^x f'(t)\, dt$$

was nach dem Hauptsatz richtig ist. Zum Beweis von $n \to n+1$ sei

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_n(x)\,, \tag{1.34}$$

$$R_n(x) = \frac{1}{(n-1)!} \int_a^x (x-t)^{n-1} f^{(n)}(t)\, dt\,. \tag{1.35}$$

Es gilt dann

$$\begin{aligned} R_n(x) &= \frac{1}{(n-1)!}\left[-\frac{1}{n}(x-t)^n f^{(n)}(t)\Big|_{t=a}^{t=x} - \int_a^x -\frac{1}{n}(x-t)^n f^{(n+1)}(t)\, dt \right] \\ &= \frac{1}{n!}(x-a)^n f^{(n)}(a) + \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\, dt\,. \end{aligned}$$

Einsetzen in (1.34) liefert die Behauptung. □

Satz 1.16 *Seien die Voraussetzungen von Satz 1.15 erfüllt. Dann gibt es zu jedem $x \in I$ ein ξ zwischen a und x mit*

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}\,. \tag{1.36}$$

Beweis: Da $f^{(n+1)}$ stetig ist und die durch

$$g(t) = \frac{(x-t)^n}{n!}$$

definierte Funktion zwischen a und x nicht das Vorzeichen wechselt, können wir den Mittelwertsatz der Integralrechnung (Satz 14.25, Analysis I) anwenden und erhalten, daß es ein ξ zwischen a und x gibt mit

$$\begin{aligned} R_{n+1}(x) &= \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\, dt = \frac{1}{n!} f^{(n+1)}(\xi) \int_a^x (x-t)^n\, dt \\ &= \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}\,. \end{aligned}$$

□

Folgerung 1.17 *Seien die Voraussetzungen von Satz 1.15 erfüllt. Gilt dann zusätzlich $f^{(n+1)} = 0$ in I, so ist f ein Polynom vom Grad $\leq n$.*

Beweis: In diesem Fall ist $R_{n+1} = 0$ in I. □

Folgerung 1.18 *Seien die Voraussetzungen von Satz 1.15 erfüllt. Dann gibt es eine Funktion $\eta : I \to \mathbb{R}$ mit*

$$f(x) = \sum_{k=0}^{n+1} \frac{f^{(k)}(a)}{k!}(x-a)^k + \eta(x)(x-a)^{n+1} \tag{1.37}$$

und

$$\lim_{x \to a} \eta(x) = 0\,. \tag{1.38}$$

Beweis: Nach Satz 1.16 finden wir zu jedem $x \in I$ ein $\xi(x)$ mit $|\xi(x) - a| \leq |x - a|$ und

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

Wir definieren

$$\eta(x) = \frac{f^{(n+1)}(\xi(x)) - f^{(n+1)}(a)}{(n+1)!}\,,$$

dann gilt (1.38), da $f^{(n+1)}$ stetig ist und $\lim_{x\to a} \xi(x) = a$. □

Ist also $f \in C^n(I)$ und T_n das n-te Taylorpolynom von f in a, so geht der Fehler $|f(x) - T_n(x)|$ für $x \to a$ schneller gegen 0 als $(x-a)^n$.

Definition 1.19 (Klein-O und Groß-O)
Sei $\delta > 0$, $I = (-\delta, \delta)$ und $f, g : I \to \mathbb{R}$ Funktionen. Wir sagen

$$f(h) = O(g(h))\,, \tag{1.39}$$

falls es ein $C > 0$ gibt mit

$$|f(h)| \leq C|g(h)|\,, \quad \text{für alle } h \in I. \tag{1.40}$$

Wir sagen

$$f(h) = o(g(h))\,, \tag{1.41}$$

falls $g(h) \neq 0$ für alle $h \neq 0$ und

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{f(h)}{g(h)} = 0\,. \tag{1.42}$$

Beispiel 1.20
Sei T_n das n-te Taylorpolynom von f in $a \in I$, I Intervall, sei

$$r(h) = f(a+h) - T_n(a+h)\,.$$

Dann gilt nach Folgerung 1.18

$$r(h) = o(h^n)\,, \quad \text{falls } f \in C^n(I),$$

und

$$r(h) = O(h^{n+1})\,, \quad \text{falls } f \in C^{n+1}(I),$$

da für eine geeignete Zwischenstelle ξ

$$r(h) = R_{n+1}(a+h) = \frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1}\,,$$

also

$$|r(h)| \le C|h|^{n+1}\,, \quad C = \frac{1}{(N+1)!} \sup_{|\xi - a| \le |h|} |f^{(n+1)}(\xi)|\,.$$

Beispiel 1.21

1. Die Taylorreihe des Sinus mit Entwicklungspunkt $a = 0$ ist

$$T_{\sin}(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}\,, \tag{1.43}$$

und für das Restglied folgt, da alle Ableitungen von sin durch 1 beschränkt sind, aus Satz 1.16

$$|R_n(x)| \le \frac{|x|^n}{n!}\,, \tag{1.44}$$

also gilt für alle $x \in \mathbb{R}$

$$\lim_{n\to\infty} R_n(x) = 0\,,$$

und damit auch

$$\sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \tag{1.45}$$

für alle $x \in \mathbb{R}$.

2. Analog folgt für den Cosinus mit Entwicklungspunkt $a = 0$

$$\cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} \tag{1.46}$$

für alle $x \in \mathbb{R}$.

2 Konvexe Funktionen

Definition 2.1 (Konvexe Menge)
Sei X Vektorraum, $D \subset X$. D heißt konvex, wenn

$$\lambda x + (1-\lambda)y \in D \tag{2.1}$$

für alle $x, y \in D$ und alle $\lambda \in [0,1]$.

Definition 2.2 (Konvexe Funktion)
Sei X Vektorraum, $D \subset X$ konvex, $f : D \to \mathbb{R}$. f heißt konvex, wenn

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \tag{2.2}$$

für alle $x, y \in D$ und alle $\lambda \in [0,1]$. f heißt konkav, wenn $-f$ konvex ist.

Lemma 2.3 *Sei X Vektorraum, $D \subset X$ konvex, $f : D \to \mathbb{R}$ konvex. Seien $x_i \in D$, $\lambda_i \in [0,1]$ für $1 \leq i \leq n$ mit*

$$\sum_{i=1}^{n} \lambda_i = 1\,. \tag{2.3}$$

Dann gilt

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i)\,. \tag{2.4}$$

Beweis: Übungsaufgabe. □

Satz 2.4 *Sei $D \subset \mathbb{R}$ offenes Intervall, $f : D \to \mathbb{R}$ differenzierbar. Dann ist f monoton wachsend genau dann, wenn $f'(x) \geq 0$ für alle $x \in D$.*

Beweis: "$\Leftarrow$": Sei $f'(x) \geq 0$ für alle $x \in D$. Sind $x_1, x_2 \in D$ mit $x_1 < x_2$, so gibt es nach dem Mittelwertsatz der Differentialrechnung ein $\xi \in (x_1, x_2)$ mit

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) \geq 0\,,$$

also ist $f(x_1) \leq f(x_2)$.
"$\Rightarrow$": Sei $x \in D$ mit $f'(x) < 0$. Dann ist für hinreichend kleines $h > 0$ auch

$$\frac{f(x+h) - f(x)}{h} < 0\,,$$

also $f(x+h) < f(x)$, also f nicht monoton wachsend. □

Satz 2.5 *Sei $D \subset \mathbb{R}$ offenes Intervall, $f : D \to \mathbb{R}$ zweimal differenzierbar. Dann ist f konvex genau dann, wenn $f''(x) \geq 0$ für alle $x \in D$.*

Beweis: "$\Leftarrow$": Sei $f''(x) \geq 0$ für alle $x \in D$. Dann ist nach Satz 2.4 die Funktion $f' : D \to \mathbb{R}$ monoton wachsend. Seien nun $x_1, x_2 \in D$ und $\lambda \in (0,1)$ mit $x_1 < x_2$, sei

$$x = \lambda x_1 + (1-\lambda)x_2 \,.$$

Dann gibt es (Mittelwertsatz) $\xi_1 \in (x_1, x)$ und $\xi_2 \in (x, x_2)$ mit

$$\frac{f(x) - f(x_1)}{x - x_1} = f'(\xi_1) \leq f'(\xi_2) = \frac{f(x_2) - f(x)}{x_2 - x} \,.$$

Es ist

$$x - x_1 = (1-\lambda)(x_2 - x_1) \,, \quad x_2 - x = \lambda(x_2 - x_1) \,,$$

also

$$\frac{f(x) - f(x_1)}{1-\lambda} \leq \frac{f(x_2) - f(x)}{\lambda} \,.$$

Multiplikation mit $\lambda(1-\lambda)$ und Umsortieren liefert

$$f(x) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \,.$$

"$\Rightarrow$": Sei $x \in D$. Seien außerdem $y \in D$, $y > x$ und $\lambda \in (0,1)$ beliebig, dann gilt

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x) \,,$$

also

$$f(x + \lambda(y-x)) - f(x) \leq \lambda(f(y) - f(x)) \,,$$

und weiter

$$\frac{f(x + \lambda(y-x)) - f(x)}{\lambda(y-x)} \leq \frac{f(y) - f(x)}{y - x} \,.$$

Grenzübergang $\lambda \downarrow 0$ liefert

$$f'(x) \leq \frac{f(y) - f(x)}{y - x}$$

für alle $y > x$. Sei nun $(x_n)_{n\in\mathbb{N}}$ Folge in D mit $x_n > x$ für alle n und $x_n \to x$. Dann gibt es $\xi_n \in (x, x_n)$ mit

$$f'(x) \leq \frac{f(x_n) - f(x)}{x_n - x} = f'(\xi_n) \,,$$

also

$$0 \leq \frac{f'(\xi_n) - f'(x)}{\xi_n - x} \,,$$

Grenzübergang $n \to \infty$ liefert $\xi_n \to x$, also $f''(x) \geq 0$. □

Satz 2.6 *Seien $x_i > 0$, $\lambda_i \in [0,1]$ für $1 \leq i \leq n$ mit $\sum_i \lambda_i = 1$. Dann gilt*

$$\prod_{i=1}^{n} x_i^{\lambda_i} \leq \sum_{i=1}^{n} \lambda_i x_i \,. \tag{2.5}$$

Beweis: Die Funktion $-\ln : (0,\infty) \to \mathbb{R}$ ist nach Satz 2.5 konvex. Wir wenden Lemma 2.3 an mit $f = -\ln$, dann gilt

$$\ln\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i \ln(x_i)\,,$$

also (da die Exponentialfunktion monoton ist)

$$\sum_{i=1}^{n} \lambda_i x_i \geq \exp\left(\sum_{i=1}^{n} \lambda_i \ln(x_i)\right) = \prod_{i=1}^{n} \exp(\lambda_i \ln x_i) = \prod_{i=1}^{n} x_i^{\lambda_i}\,.$$

□

Satz 2.7 (Youngsche Ungleichung)
Seien $p, q \in (1,\infty)$ mit

$$\frac{1}{p} + \frac{1}{q} = 1\,. \tag{2.6}$$

Dann gilt für alle $x, y \geq 0$

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}\,. \tag{2.7}$$

Beweis: Anwendung von Satz 2.6 mit

$$\lambda_1 = \frac{1}{p}\,, \quad \lambda_2 = \frac{1}{q}\,, \quad x_1 = x^p\,, \quad x_2 = y^q\,.$$

□

Definition 2.8 (p-Norm)
Sei $1 \leq p < \infty$, $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$. Dann definieren wir

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}\,. \tag{2.8}$$

Wir erinnern an die Definition der Maximumsnorm für $x \in \mathbb{C}^n$,

$$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|\,.$$

Satz 2.9 (Höldersche Ungleichung)
Seien $p, q \in (1,\infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$, oder sei $p = 1$, $q = \infty$, oder sei $p = \infty$, $q = 1$. Dann gilt für alle $x, y \in \mathbb{C}^n$

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \|y\|_q\,. \tag{2.9}$$

Beweis: Für $p = 1$, $q = \infty$ (analog $p = \infty$, $q = 1$) gilt

$$\sum_{i=1}^{n} |x_i y_i| \leq \sum_{i=1}^{n} |x_i| \max_{1\leq j\leq n} |y_j| = \|y\|_\infty \sum_{i=1}^{n} |x_i|\,.$$

Seien $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Seien $x, y \in \mathbb{C}^n$, beide von Null verschieden. (Andernfalls ist (2.9) trivialerweise erfüllt.) Dann ist $\|x\|_p \neq 0$, $\|y\|_q \neq 0$. Aus der Youngschen Ungleichung folgt für alle i, $1 \leq i \leq n$,

$$\frac{|x_i|}{\|x\|_p} \cdot \frac{|y_i|}{\|y\|_q} \leq \frac{|x_i|^p}{p\|x\|_p^p} + \frac{|y_i|^p}{q\|y\|_q^q} \,,$$

also (Summation über i)

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

□

Wir betrachten nun das Standardskalarprodukt im $\mathbb{C}^n$,

$$\langle x, y \rangle = \sum_{i=1}^n x_i \overline{y_i} \,. \tag{2.10}$$

Aus der Hölderschen Ungleichung, angewendet mit $p = q = 2$, erhalten wir

$$| \langle x, y \rangle | = \left| \sum_{i=1}^n x_i \overline{y_i} \right| \leq \sum_{i=1}^n |x_i y_i| \leq \|x\|_2 \, \|y\|_2 \,.$$

Die resultierende Ungleichung

$$| \langle x, y \rangle | \leq \|x\|_2 \, \|y\|_2 \,, \quad x, y \in \mathbb{C}^n \,, \tag{2.11}$$

heißt die Cauchy-Schwarzsche Ungleichung.

Satz 2.10 (Minkowskische Ungleichung)
Sei $1 \leq p \leq \infty$. Dann gilt für alle $x, y \in \mathbb{C}^n$

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p \,. \tag{2.12}$$

Beweis: Für $p = \infty$: Siehe Analysis I. Für $p = 1$:

$$\|x + y\|_1 = \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n (|x_i| + |y_i|) = \|x\|_1 + \|y\|_1 \,.$$

Sei nun $1 < p < \infty$, $q = \frac{p}{p-1}$. Dann ist $1 < q < \infty$,

$$\frac{1}{q} + \frac{1}{p} = 1 \,.$$

Seien $x, y \in \mathbb{C}^n$. Wir definieren $z \in \mathbb{C}^n$ durch

$$z_i = |x_i + y_i|^{p-1} \,, \quad 1 \leq i \leq n \,.$$

Es gilt

$$\begin{aligned} \|x+y\|_p^p &= \sum_{i=1}^n |x_i+y_i|^p = \sum_{i=1}^n |x_i+y_i|\,|z_i| \le \sum_{i=1}^n |x_i z_i| + \sum_{i=1}^n |y_i z_i| \\ &\le \|x\|_p \|z\|_q + \|y\|_p \|z\|_q \end{aligned}$$

wegen der Hölderschen Ungleichung, und

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

also

$$\|x+y\|_p^p \le (\|x\|_p + \|y\|_p)\|x+y\|_p^{p-1}.$$

Division durch $\|x+y\|_p^{p-1}$ liefert die Behauptung. □

Folgerung 2.11 *Sei $p \in [1,\infty)$. Mit*

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \tag{2.13}$$

sind $(\mathbb{R}^n, \|\cdot\|_p)$ und $(\mathbb{C}^n, \|\cdot\|_p)$ normierte Räume.

Beweis: Die Minkowskische Ungleichung liefert gerade die Dreiecksungleichung. Die anderen verlangten Eigenschaften einer Norm folgen unmittelbar aus der Definition. □

Lemma 2.12 *Sei $(X, \|\cdot\|)$ normierter Raum. Dann ist die durch*

$$f(x) = \|x\| \tag{2.14}$$

definierte Funktion $f : X \to \mathbb{R}$ konvex.

Beweis: Es gilt für alle $x, y \in X$ und alle $\lambda \in [0,1]$

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \|\lambda x + (1-\lambda)y\| \le \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\| \\ &= \lambda f(x) + (1-\lambda)f(y). \end{aligned} \tag{2.15}$$

□

Lemma 2.13 *Sei X Vektorraum, $D \subset X$ konvex, $f : D \to \mathbb{R}$ konvex. Dann sind die Mengen*

$$D_\alpha = \{x : x \in D,\ f(x) \le \alpha\} \tag{2.16}$$

konvex für alle $\alpha \in \mathbb{R}$.

Beweis: Es gilt für alle $x, y \in D_\alpha$ und alle $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \le \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

□

Folgerung 2.14 *Sei* $(X, \|\cdot\|)$ *normierter Raum. Dann sind die Einheitskugeln*

$$\{x : x \in X,\ \|x\| \leq 1\} \tag{2.17}$$

konvex.

Beweis: Folgt aus Lemma 2.12 und Lemma 2.13. □

(Bilder für die Einheitskugeln in $(\mathbb{R}^2, \|\cdot\|_p)$ siehe Vorlesung.)

3 Metrische Räume

In einem normierten Raum X wird der Abstand $d(x,y)$ zweier Vektoren $x, y \in X$ definiert durch

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

Wir führen nun einen Abstandsbegriff für allgemeine Mengen X ein.

Definition 3.1 (Metrik)
Sei X Menge. Eine Abbildung $d : X \times X \to \mathbb{R}_+$ heißt Metrik auf X, falls gilt

$$d(x,y) = 0 \qquad \Leftrightarrow \qquad x = y \qquad \textit{(Definitheit)} \tag{3.1}$$

$$d(x,y) = d(y,x) \quad \textit{für alle } x, y \in X \qquad \textit{(Symmetrie)} \tag{3.2}$$

$$d(x,z) \le d(x,y) + d(y,z) \quad \textit{für alle } x, y, z \in X \qquad \textit{(Dreiecksungleichung)} \tag{3.3}$$

(X, d) heißt metrischer Raum. □

Daß $d(x,y) \geq 0$ gelten muß, folgt bereits aus (3.1) – (3.3) wegen

$$0 = d(x,x) \le d(x,y) + d(y,x) = 2d(x,y) \,.$$

Satz 3.2 *Sei $(X, \|\cdot\|)$ normierter Raum. Dann definiert*

$$d(x,y) = \|x - y\| \tag{3.4}$$

eine Metrik d auf X. Sie heißt die durch $\|\cdot\|$ erzeugte Metrik.

Beweis: Definitheit:

$$d(x,y) = 0 \;\Leftrightarrow\; \|x-y\| = 0 \;\Leftrightarrow\; x - y = 0 \;\Leftrightarrow\; x = y \,, \quad \text{für alle } x, y \in X.$$

Symmetrie:

$$d(y,x) = \|y - x\| = \|-(y-x)\| = \|x - y\| = d(x,y) \,, \quad \text{für alle } x, y \in X.$$

Dreiecksungleichung:

$$d(x,z) = \|x - z\| \le \|x - y\| + \|y - z\| = d(x,y) + d(y,z) \,, \quad \text{für alle } x, y, z \in X.$$

□

Lemma 3.3 *Sei (X, d) metrischer Raum, sei $A \subset X$. Schränken wir die Metrik auf A ein, $d_A = d|(A \times A)$, so ist (A, d_A) metrischer Raum.*

Beweis: Klar. □

Ist $(X, \|\cdot\|)$ normierter Raum, so können wir auf diese Weise jedes $A \subset X$ zu einem metrischen Raum machen. (Um durch Restriktion der Norm auf A wieder einen normierten Raum zu erhalten, muß A Unterraum sein.)

Beispiel 3.4
Sei S die Oberfläche der Erdkugel. Wir können eine Metrik auf S definieren, indem wir die euklidische Norm im $\mathbb{R}^3$ als Metrik auffassen und auf S einschränken. Eine andere Metrik erhalten wir, wenn wir als Abstand zweier Punkte auf S die Länge der kürzesten auf S verlaufenden Verbindungsstrecke definieren. Sei nun X die Menge aller Städte in Europa, die mit München durch einen Straßenzug verbunden sind. Wir erhalten eine dritte Metrik, indem wir als Abstand zweier Städte die Länge des kürzesten verbindenden Straßenzugs definieren. Keine Metrik ergibt sich, wenn wir als Abstand die kürzeste Fahrzeit laut Bahnfahrplan definieren (dabei betrachten wir nur die Städte, zu denen eine Bahnverbindung existiert).

Definition 3.5 (offene Kugel, Umgebung)
Sei (X, d) metrischer Raum. Für $a \in X$ und $r > 0$ heißt

$$B(a, r) = \{x \in X : d(a, x) < r\} \tag{3.5}$$

die offene Kugel mit Mittelpunkt a und Radius r. Sei nun $x \in X$. Ein $U \subset X$ heißt Umgebung von x, falls es ein $\varepsilon > 0$ gibt mit

$$x \in B(x, \varepsilon) \subset U\,. \tag{3.6}$$

Die offene Kugel $B(x, \varepsilon)$ heißt auch die ε-Umgebung von x. □

In einem normierten Raum gilt für die Kugeln (bezüglich der erzeugten Metrik)

$$B(a, r) = a + rB(0, 1)\,,$$

"alle Kugeln sehen gleich aus".

In $(\mathbb{R}, d)$ mit $d(x, y) = |x-y|$ (Betragsmetrik) ist $B(a, r)$ das offene Intervall $(a-r, a+r)$. In $\mathbb{C}$ mit der Betragsmetrik ist $B(a, r)$ das "Innere" des Kreises mit Mittelpunkt a und Radius r.

Satz 3.6 (Hausdorffsche Trennungseigenschaft)
Sei (X, d) metrischer Raum. Dann haben je zwei verschiedene Punkte disjunkte Umgebungen, das heißt: Für alle $x, y \in X$ mit $x \neq y$ gibt es Umgebungen U von x und V von y mit

$$U \cap V = \emptyset\,. \tag{3.7}$$

Beweis: Seien $x, y \in X$ mit $x \neq y$. Wir setzen

$$\varepsilon = \frac{1}{2}d(x, y)\,, \quad U = B(x, \varepsilon)\,, \quad V = B(y, \varepsilon)\,.$$

Für beliebiges $z \in U$ gilt

$$2\varepsilon = d(x, y) \leq d(x, z) + d(z, y) < \varepsilon + d(z, y)\,,$$

also ist $d(z, y) > \varepsilon$ und damit $z \notin V$. □

Definition 3.7 (Offene Menge)
Sei (X,d) metrischer Raum. Ein $U \subset X$ heißt offen, falls U Umgebung jedes Punktes $x \in U$ ist, d.h. falls gilt: Für alle $x \in U$ existiert ein $\varepsilon > 0$ mit

$$x \in B(x,\varepsilon) \subset U \tag{3.8}$$

□

Jedes offene Intervall (a,b) in $\mathbb{R}$ ist offen (bezüglich der Betragsmetrik). Jede offene Kugel $B(a,r)$ eines metrischen Raums ist offen: Ist $x \in B(a,r)$, so ist $B(x,\varepsilon) \subset B(a,r)$, falls $\varepsilon < r - d(a,x)$.

Ob eine Teilmenge eines metrischen Raums (X,d) offen ist oder nicht, hängt von der Wahl der Metrik d ab. Andererseits wird es sich noch herausstellen, daß im $\mathbb{R}^n$ alle Metriken, die von Normen erzeugt werden, zu denselben offenen Mengen führen.

Man beachte, daß der Begriff "offen" für Teilmengen eines metrischen Raumes X relativ zu X definiert ist. So ist etwa $[0,1]$ nicht offen in $\mathbb{R}$ (d.h. aufgefaßt als Teilmenge von $\mathbb{R}$), wohl aber offen in sich selbst (d.h. aufgefaßt als Teilmenge des metrischen Raums $([0,1], d_{[0,1]})$.

Satz 3.8 *Sei (X,d) metrischer Raum. Dann gilt:*

(i) $\emptyset$ und X sind offen.

(ii) Sind U und V offene Teilmengen, so ist auch $U \cap V$ offen.

(iii) Ist $(U_i)_{i \in I}$ eine Familie offener Teilmengen, so ist auch

$$\bigcup_{i \in I} U_i$$

offen.

Beweis: (i) ist klar. (ii): Seien U, V offen, sei $x \in U \cap V$ beliebig. Wir wählen $\varepsilon_1, \varepsilon_2$ mit $B(x,\varepsilon_1) \subset U$ und $B(x,\varepsilon_2) \subset V$. Für $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ ist dann $B(x,\varepsilon) \subset U \cap V$.
(iii): Ist $x \in \cup_{i \in I} U_i$, so gibt es ein $i \in I$ mit $x \in U_i$ und weiter ein $\varepsilon > 0$ mit $B(x,\varepsilon) \subset U_i$, also $x \in B(x,\varepsilon) \subset \cup_{i \in I} U_i$. □

Folgerung 3.9 *Sind U_i, $1 \le i \le n$, offene Teilmengen eines metrischen Raums (X,d), so ist auch*

$$\bigcap_{i=1}^{n} U_i$$

offen.

Beweis: Folgt direkt aus Teil (ii) des Satzes. □

Ein unendlicher Durchschnitt offener Mengen braucht nicht offen zu sein, so ist etwa

$$\bigcap_{n \in \mathbb{N}} B(x, \frac{1}{n}) = \{x\}\,,$$

und einpunktige Mengen sind "normalerweise" nicht offen.

Bemerkung 3.10 (Topologie)
Satz 3.8 ist ein Ausgangspunkt, um den Begriff des metrischen Raumes weiter zu verallgemeinern, und zwar zum Begriff des *topologischen Raumes*. Statt des Abstandsbegriffs wird hier der Begriff der offenen Menge axiomatisiert: Sei X Menge. Eine Familie $\mathcal{T} \subset \mathcal{P}(X)$ von Teilmengen von X heißt Topologie auf X, falls gilt

(i) $\emptyset \in \mathcal{T}$, $X \in \mathcal{T}$,

(ii) $U, V \in \mathcal{T} \;\Rightarrow\; U \cap V \in \mathcal{T}$,

(iii) $(U_i)_{i\in I} \subset \mathcal{T} \;\Rightarrow\; \cup_{i\in I} U_i \in \mathcal{T}$.

$(X, \mathcal{T})$ heißt topologischer Raum, die Elemente von $\mathcal{T}$ heißen offene Mengen in $(X, \mathcal{T})$.

□

Definition 3.11 (Abgeschlossene Menge)
Sei (X, d) metrischer Raum. Ein $A \subset X$ heißt abgeschlossen, wenn das Komplement $X \setminus A$ offen ist.

□

Definition 3.12 (Abgeschlossene Kugel)
Sei (X, d) metrischer Raum, $a \in X$, $r > 0$. Dann heißt

$$K(a, r) = \{x \in X : d(a, x) \leq r\} \tag{3.9}$$

die abgeschlossene Kugel mit Mittelpunkt a und Radius r.

□

Jede abgeschlossene Kugel $K(a, r)$ ist abgeschlossen (siehe Übung). Jedes abgeschlossene Intervall in $\mathbb{R}$ ist abgeschlossen.

Satz 3.13 *Sei (X, d) metrischer Raum. Dann gilt*

(i) $\emptyset$ und X sind abgeschlossen.

(ii) Sind $(A_i)_{1\leq i\leq n}$ abgeschlossen, so ist auch

$$\bigcup_{i=1}^{n} A_i$$

abgeschlossen.

(iii) Ist $(A_i)_{i\in I}$ eine Familie abgeschlossener Teilmengen, so ist auch

$$\bigcap_{i\in I} A_i$$

abgeschlossen.

Beweis: Folgt direkt aus Satz 3.8 und Folgerung 3.9. $\square$

Beispiel 3.14 (Cantorsches Diskontinuum)
Sei $C_0 = [0,1]$. Wir setzen $C_1 = C_0 \setminus (1/3, 2/3)$. C_1 besteht aus den beiden abgeschlossenen Intervalle $[0,1/3]$ und $[2/3,1]$. Wir entfernen nun jeweils deren offenes mittleres Drittel $(1/9, 2/9)$ bzw. $(7/9, 8/9)$, um C_2 zu erhalten. Entsprechend erhalten wir C_{n+1} aus C_n. Die Menge

$$C = \bigcap_{n \in \mathbb{N}} C_n \tag{3.10}$$

heißt *Cantorsches Diskontinuum*. C ist abgeschlossen. Es ist

$$C = \left\{ \sum_{i=1}^{\infty} a_i 3^{-i} : a_i \in \{0,2\} \text{ für alle } i \right\} . \tag{3.11}$$

Definition 3.15 (Rand, Inneres, Abschluß)
Sei (X,d) metrischer Raum, sei $Y \subset X$. Ein $x \in X$ heißt Randpunkt von Y, falls jede ε-Umgebung von x sowohl einen Punkt von Y als auch einen Punkt von $X \setminus Y$ enthält. Die Menge

$$\partial Y = \{x : x \in X,\ x \text{ ist Randpunkt von } Y\} \tag{3.12}$$

heißt Rand von Y (in X). Die Menge

$$\operatorname{int}(Y) = Y \setminus \partial Y \tag{3.13}$$

heißt das Innere von Y. Die Menge

$$\overline{Y} = Y \cup \partial Y \tag{3.14}$$

heißt der Abschluß von Y.

$\square$

Lemma 3.16 *Sei (X,d) metrischer Raum, $Y \subset X$. Dann gilt*

$$\operatorname{int}(Y) \subset Y \subset \overline{Y}\,, \quad \partial Y = \overline{Y} \setminus \operatorname{int}(Y)\,. \tag{3.15}$$

Beweis: Folgt direkt aus der Definition. $\square$

In $\mathbb{R}$ mit der Standardmetrik gilt: Ist Y eine Teilmenge mit $(a,b) \subset Y \subset [a,b]$, so ist

$$\partial Y = \{a,b\}\,, \quad \operatorname{int}(Y) = (a,b)\,, \quad \overline{Y} = [a,b]\,.$$

Ebenso gilt in $\mathbb{R}$: $\partial\mathbb{Q} = \overline{\mathbb{Q}} = \mathbb{R}$, $\operatorname{int}(\mathbb{Q}) = \emptyset$.

Definieren wir in einem normierten Raum $(X, \|\cdot\|)$ die Sphäre durch

$$S(a,r) = \{x : x \in X,\ \|x - a\| = r\}\,, \tag{3.16}$$

so ist

$$S(a,r) = \partial K(a,r) = \partial B(a,r) = K(a,r) \setminus B(a,r)\,.$$

Man beachte: Für beliebige Teilmengen Y eines metrischen Raumes gilt im allgemeinen **nicht**

$$\overline{Y} = \overline{\operatorname{int}(Y)}\,. \tag{3.17}$$

In einem normierten Raum gilt (3.17), falls Y konvex ist und $\operatorname{int}(Y) \neq \emptyset$.

Satz 3.17 *Sei (X,d) metrischer Raum, sei $Y \subset X$. Dann gilt*

(i) $\mathrm{int}\,(Y)$ *ist offen.*

(ii) $\overline{Y}$ *ist abgeschlossen.*

(iii) ∂Y *ist abgeschlossen.*

Beweis: (i): Sei $a \in \mathrm{int}\,(Y) = Y \setminus \partial Y$. Da $a \notin \partial Y$, gibt es ein $\varepsilon > 0$, so daß $B(a,\varepsilon)$ keinen Punkt von $X \setminus Y$ enthält, also $B(a,\varepsilon) \subset Y$. Sei nun $y \in B(a,\varepsilon)$ beliebig. Für hinreichend kleines $\delta > 0$ gilt $B(y,\delta) \subset B(a,\varepsilon) \subset Y$, also ist $y \notin \partial Y$. Da y beliebig war, folgt $B(a,\varepsilon) \subset Y \setminus \partial Y = \mathrm{int}\,(Y)$.
(ii): Für $Z = X \setminus Y$ gilt $\partial Z = \partial Y$ nach Definition. Es folgt

$$\overline{Y} = Y \cup \partial Y = (X \setminus Z) \cup \partial Z = X \setminus (Z \setminus \partial Z) = X \setminus \mathrm{int}\,(Z)\,.$$

Nach (i) ist $\mathrm{int}\,(Z)$ offen, also ist $\overline{Y}$ abgeschlossen.
(iii) Es ist

$$\partial Y = \overline{Y} \setminus \mathrm{int}\,(Y) = \overline{Y} \cap (X \setminus \mathrm{int}\,(Y))\,,$$

und damit abgeschlossen als Durchschnitt der abgeschlossenen Mengen $\overline{Y}$ und $X \setminus \mathrm{int}\,(Y)$.
□

Satz 3.18 (Produktmetrik)
Seien (X_1,d_1) und (X_2,d_2) metrische Räume. Dann wird durch

$$d((x_1,x_2),(y_1,y_2)) = \max\{d_1(x_1,y_1), d_2(x_2,y_2)\} \tag{3.18}$$

eine Metrik auf $X_1 \times X_2$ definiert, die sogenannte Produktmetrik. Es gilt

$$B((x_1,x_2),\varepsilon) = B(x_1,\varepsilon) \times B(x_2,\varepsilon)\,, \tag{3.19}$$
$$K((x_1,x_2),\varepsilon) = K(x_1,\varepsilon) \times K(x_2,\varepsilon)\,, \tag{3.20}$$

für alle $x_i \in X_i$, $\varepsilon > 0$.

Beweis: Übungsaufgabe. □

Satz 3.19 (Offene und abgeschlossene Mengen im Produktraum)
Seien (X_1,d_1) und (X_2,d_2) metrische Räume. Sind $U_i \subset X_i$ offen in X_i, $i = 1,2$, so ist $U_1 \times U_2$ offen in $X_1 \times X_2$, versehen mit der Produktmetrik. Ebenso ist $A_1 \times A_2$ abgeschlossen in $X_1 \times X_2$, falls $A_i \subset X_i$ abgeschlossen sind für $i = 1,2$.

Beweis: Seien U_i offen, sei $(x_1,x_2) \in U_1 \times U_2$. Wähle $\varepsilon_1, \varepsilon_2$ mit $B(x_i,\varepsilon_i) \subset U_i$ $(i = 1,2)$, dann ist $B((x_1,x_2),\varepsilon) \subset U_1 \times U_2$ für $\varepsilon = \min\{\varepsilon_1,\varepsilon_2\}$. Seien $A_i \subset X_i$ abgeschlossen. Es ist

$$(X_1 \times X_2) \setminus (A_1 \times A_2) = ((X_1 \setminus A_1) \times X_2) \cup (X_1 \times (X_2 \setminus A_2))\,,$$

und die rechte Seite ist offen als Vereinigung zweier offener Mengen. □

Folgerung 3.20 *Seien (X_i, d_i) metrische Räume, $1 \le i \le n$. Dann wird auf*

$$X = \prod_{i=1}^{n} X_i$$

eine Metrik d definiert durch

$$d(x, y) = \max_{1 \le i \le n} d_i(x_i, y_i) \tag{3.21}$$

für $x, y \in X$, $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$. Sind U_i offen in X_i bzw. A_i abgeschlossen in X_i, so sind auch

$$\prod_{i=1}^{n} U_i\,, \quad \prod_{i=1}^{n} A_i\,,$$

offen bzw. abgeschlossen in (X, d).

Beweis: Folgt mit vollständiger Induktion aus den Sätzen 3.18 und 3.19. □

Setzen wir $X_i = \mathbb{K}$ und wählen wir für d_i die Betragsmetrik, so ist $X = \mathbb{K}^n$, und die Produktmetrik

$$d(x, y) = \max_{1 \le i \le n} |x_i - y_i|$$

entspricht der von der Maximumnorm auf $\mathbb{K}^n$ erzeugten Metrik. Aus Satz 3.20 ergibt sich dann für $\mathbb{K} = \mathbb{R}$, daß alle achsenparallelen Quader

$$\prod_{i=1}^{n}(a_i, b_i)\,, \quad \prod_{i=1}^{n}[a_i, b_i]$$

offen bzw. abgeschlossen sind in $(\mathbb{R}^n, \|\cdot\|_\infty)$.

Betrachten wir ein Produkt

$$X_1 \times X_2 \times X_3$$

dreier metrischer Räume (X_i, d_i), und konstruieren wir eine Metrik durch zweimalige Produktbildung gemäß

$$(X_1 \times X_2) \times X_3 \quad \text{, oder} \quad X_1 \times (X_2 \times X_3)\,,$$

so erhalten wir in jedem Fall dieselbe Metrik, nämlich die in (3.21) definierte.

4 Metrische Räume: Konvergenz, Stetigkeit

Hier steht $\mathbb{K}$ für $\mathbb{R}$ oder $\mathbb{C}$ wie gehabt.

Definition 4.1 (Konvergenz)
Sei (X,d) metrischer Raum. Eine Folge $(x_k)_{k\in\mathbb{N}}$ in X heißt konvergent gegen den Grenzwert $a \in X$, falls

$$\lim_{k\to\infty} d(x_k, a) = 0\,. \tag{4.1}$$

Wir schreiben wieder

$$\lim_{k\to\infty} x_k = a \qquad \textit{(in } (X,d)\textit{)}\,,$$

oder einfach "$x_k \to a$".

□

Falls d von einer Norm erzeugt wird, $d(x,y) = \|x-y\|$, so erhalten wir die Definition der Konvergenz im normierten Raum (Analysis I).

Satz 4.2 *Seien (X_i, d_i) metrische Räume, $1 \le i \le n$, sei $X = \prod_{i=1}^n X_i$ versehen mit der Produktmetrik d. Sei $(x_k)_{k\in\mathbb{N}}$ Folge in X mit $x_k = (x_{k,1}, \ldots, x_{k,n})$, sei $a = (a_1, \ldots, a_n) \in X$. Dann gilt*

$$\lim_{k\to\infty} x_k = a \qquad \Leftrightarrow \qquad \lim_{k\to\infty} x_{k,i} = a_i \quad \textit{für alle } i,\ 1 \le i \le n. \tag{4.2}$$

Beweis: Nach Definition der Produktmetrik d gilt

$$d(x_k, a) \to 0 \quad \Leftrightarrow \quad \max_{1\le i\le n} d_i(x_{k,i}, a_i) \to 0 \quad \Leftrightarrow \quad d_i(x_{k,i}, a_i) \to 0 \quad \text{für alle } i.$$

□

Folgerung 4.3 *Sei $(x_k)_{k\in\mathbb{N}}$ eine Folge in $\mathbb{K}^n$, sei $a \in \mathbb{K}^n$. Dann gilt*

$$\lim_{k\to\infty} x_k = a$$

in $(\mathbb{K}^n, \|\cdot\|_\infty)$ genau dann, wenn

$$\lim_{k\to\infty} x_{k,i} = a_i$$

für alle i, $1 \le i \le n$.

Folgt aus Satz 4.2 mit $(X_i, d_i) = (\mathbb{R}, |\cdot|)$. □

Lemma 4.4 (Konvergenz und Umgebung)
Seien (X,d) metrischer Raum, $(x_k)_{k\in\mathbb{N}}$ Folge in X, $a \in X$. Dann gilt $x_k \to a$ genau dann, wenn es zu jeder Umgebung U von a ein $N \in \mathbb{N}$ gibt mit $x_k \in U$ für alle $k \ge N$.

Beweis: Nach Definition 4.1 gilt

$$\lim_{k\to\infty} x_k = a \qquad \Leftrightarrow \qquad \forall\, \varepsilon > 0 \quad \exists\, N \in \mathbb{N} \quad \forall\, k \geq N : \quad x_k \in B(a,\varepsilon)$$

"$\Leftarrow$": Klar, da $B(a,\varepsilon)$ Umgebung von a ist.
"$\Rightarrow$": Sei U Umgebung von a. Wähle $\varepsilon > 0$ mit $a \in B(a,\varepsilon) \subset U$, wähle N mit $x_k \in B(a,\varepsilon)$ für alle $k \geq N$, dann ist $x_k \in U$ für alle $k \geq N$. □

Es kommt also für die Frage, ob eine gegebene Folge konvergiert oder nicht, nur darauf an, wie die Umgebungen (oder äquivalent: die offenen Mengen) in (X,d) aussehen. Sind d_1 und d_2 Metriken auf X, die auf dieselben offenen Mengen führen, so gilt

$$x_k \to a \text{ in } (X,d_1) \qquad \Leftrightarrow \qquad x_k \to a \text{ in } (X,d_2)\,.$$

Lemma 4.5 *Sei (X,d) metrischer Raum, sei $(x_k)_{k\in\mathbb{N}}$ eine konvergente Folge in X. Dann ist der Grenzwert von $(x_k)_{k\in\mathbb{N}}$ eindeutig bestimmt, und jede Teilfolge von $(x_k)_{k\in\mathbb{N}}$ konvergiert gegen ihn.*

Beweis: Sind a,b Grenzwerte von $(x_k)_{k\in\mathbb{N}}$, so gilt

$$0 \leq d(a,b) \leq d(a,x_k) + d(x_k,b) \to 0$$

für $k \to \infty$, also $a = b$. Aus $d(a,x_k) \to 0$ folgt dasselbe für jede Teilfolge. □

Satz 4.6 (Konvergenz und Abschluß)
Sei (X,d) metrischer Raum, sei $A \subset X$. Dann sind äquivalent:

(i) A ist abgeschlossen in X.

(ii) $A = \overline{A}$.

(iii) Für jede konvergente Folge $(x_k)_{k\in\mathbb{N}}$ in X gilt: Ist $x_k \in A$ für alle k, so ist auch

$$\lim_{k\to\infty} x_k \in A\,.$$

Beweis: "(ii)$\Rightarrow$(i)": Satz 3.15.
"(iii)$\Rightarrow$(ii)": Kontraposition. Sei $A \neq \overline{A}$. Wähle $a \in \partial A$ mit $a \notin A$. Nach Definition des Randes gibt es zu jedem $k \in \mathbb{N}$ ein $x_k \in A$ mit

$$x_k \in B(a, \frac{1}{k})$$

also $d(a,x_k) \to 0$, $x_k \to a \notin A$, (iii) gilt also nicht.
"(i)$\Rightarrow$(iii)": Sei A abgeschlossen, sei $(x_k)_{k\in\mathbb{N}}$ Folge in A mit $x_k \to a$. Wäre $a \notin A$, also $a \in X \setminus A$, so gäbe es ein $\varepsilon > 0$ mit $B(a,\varepsilon) \subset X \setminus A$ (da $X \setminus A$ offen), im Widerspruch zu $x_k \in A$ und $x_k \to a$. □

Definition 4.7 (Beschränktheit)
Sei (X,d) metrischer Raum. Eine Teilmenge $Y \subset X$ heißt beschränkt, falls

$$\operatorname{diam}(Y) := \sup_{x,y\in Y} d(x,y) < \infty \tag{4.3}$$

gilt. diam (Y) *heißt der Durchmesser von Y.*

□

Lemma 4.8 *Sei $(X, \|\cdot\|)$ normierter Raum. Eine Teilmenge $Y \subset X$ ist beschränkt genau dann, wenn es ein $C > 0$ gibt mit $\|x\| \leq C$ für alle $x \in Y$.*

Beweis: "$\Leftarrow$": $\operatorname{diam}(Y) \leq 2C$.
"$\Rightarrow$": Kontraposition. Gibt es eine Folge $(x_k)_{k\in\mathbb{N}}$ in Y mit $\|x_k\| \to \infty$, so gilt

$$\|x_k - x_1\| \geq \|x_k\| - \|x_1\| \to \infty\,,$$

also $\operatorname{diam}(Y) = +\infty$. □

Definition 4.9 (Cauchyfolge, Vollständigkeit)
Sei (X, d) metrischer Raum. Eine Folge $(x_k)_{k\in\mathbb{N}}$ in X heißt Cauchyfolge, falls es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$ gibt mit

$$d(x_k, x_m) < \varepsilon\,, \quad \text{für alle } k, m \geq N\,.$$

(X, d) heißt vollständig, falls jede Cauchyfolge in X konvergiert.

Definition 4.10 (Banachraum)
Ein normierter Raum $(X, \|\cdot\|)$ heißt Banachraum, wenn er vollständig ist.

Lemma 4.11 *Sei (X, d) metrischer Raum. Dann gilt:*

(i) Jede konvergente Folge in X ist eine Cauchyfolge.

(ii) Jede Cauchyfolge ist beschränkt.

Beweis: (i): Sei (x_k) Folge mit $x_k \to a$. Zu $\varepsilon > 0$ wähle N mit $d(x_k, a) < \varepsilon$ für alle $k \geq N$. Dann gilt für $k, m \geq N$

$$d(x_k, x_m) \leq d(x_k, a) + d(a, x_m) < \varepsilon + \varepsilon = 2\varepsilon\,.$$

(ii): Sei (x_k) Cauchyfolge. Wähle N so, daß gilt $d(x_k, x_m) < 1$ für alle $k, m \geq N$. Dann gilt für $Y = \{x_k : k \in \mathbb{N}\}$

$$\operatorname{diam}(Y) \leq \max_{1\leq i,j\leq N} d(x_i, x_j) + 2 < \infty\,.$$

□

Satz 4.12 (Vollständigkeit von $\mathbb{K}^n$)
Der Raum $(\mathbb{K}^n, \|\cdot\|_\infty)$ ist vollständig, also ein Banachraum.

Beweis: Sei $(x_k)_{k\in\mathbb{N}}$ Cauchyfolge in $(\mathbb{K}^n, \|\cdot\|_\infty)$. Wegen $|x_{k,i} - x_{m,i}| \leq \|x_k - x_m\|_\infty$ ist auch jede Komponentenfolge eine Cauchyfolge in $\mathbb{K}$. Da $\mathbb{K}$ vollständig ist, gilt $x_{k,i} \to a_i$ für ein $a_i \in \mathbb{K}$. Nach Satz 4.3 gilt $x_k \to a = (a_1, \ldots, a_n)$. □

Wir erinnern an die Definition des Raums $B(D; \mathbb{K})$ der beschränkten Funktionen auf einer beliebigen Menge D mit Werten in $\mathbb{K}$ und daran, daß er mit der Supremumsnorm zu einem normierten Raum wird.

Satz 4.13 (Vollständigkeit des Raums der beschränkten Funktionen) *Sei D Menge. Der Raum $(B(D;\mathbb{K}), \|\cdot\|_\infty)$ ist ein Banachraum.*

Beweis: Sei $(f_k)_{k\in\mathbb{N}}$ Cauchyfolge in $(B(D;\mathbb{K}), \|\cdot\|_\infty)$.
(1) Wegen

$$|f_k(t) - f_m(t)| \leq \|f_k - f_m\|_\infty = \sup_{s\in D} |f_k(s) - f_m(s)|$$

ist $(f_k(t))_{k\in\mathbb{N}}$ für alle $t \in D$ eine Cauchyfolge in $\mathbb{K}$. Da $\mathbb{K}$ vollständig ist, können wir also definieren

$$f(t) = \lim_{k\to\infty} f_k(t)\,.$$

(2) f ist eine beschränkte Funktion: Nach Satz 4.11 ist die Folge $(f_k)_{k\in\mathbb{N}}$ beschränkt in $B(D;\mathbb{K})$, also existiert $C > 0$ mit

$$|f_k(t)| \leq \|f_k\|_\infty \leq C$$

für alle $k \in \mathbb{N}$ und alle $t \in D$, also ist auch $|f(t)| \leq C$ für alle $t \in D$.
(3): Wir zeigen $\|f_k - f\|_\infty \to 0$: Sei $\varepsilon > 0$. Wähle N, so daß

$$\|f_k - f_m\|_\infty < \frac{\varepsilon}{2} \qquad \text{für alle } k, m \geq N.$$

Sei nun $t \in D$. Wähle $m \geq N$ so, daß

$$|f(t) - f_m(t)| < \frac{\varepsilon}{2}\,.$$

Dann gilt für alle $k \geq N$

$$|f(t) - f_k(t)| \leq |f(t) - f_m(t)| + |f_m(t) - f_k(t)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Da $t \in D$ beliebig war, folgt

$$\|f - f_k\|_\infty = \sup_{t\in D} |f(t) - f_k(t)| \leq \varepsilon$$

für alle $k \geq N$. $\square$

Satz 4.14 *Sei (X, d) vollständiger metrischer Raum, sei $Y \subset X$. Dann sind äquivalent:*

(i) Y ist abgeschlossen in X.

(ii) Der metrische Raum (Y, d_Y) ist vollständig.

Beweis: "(i)⇒(ii)": Sei (x_k) Cauchyfolge in Y. Da X vollständig ist, gilt $x_k \to a$, $a \in X$. Da Y abgeschlossen ist, gilt $a \in Y$ nach Satz 4.6.
"(ii)⇒(i)": Sei (x_k) Folge in Y mit $x_k \to a$, $a \in X$. Dann ist (x_k) Cauchyfolge in (X, d), also auch Cauchyfolge in (Y, d_Y). Da (Y, d_Y) vollständig ist, gilt $x_k \to b \in Y$. Da der Grenzwert eindeutig ist, gilt $a = b \in Y$. Also ist Y abgeschlossen nach Satz 4.6. $\square$

Folgerung 4.15 *Der Raum $(C[a, b], \|\cdot\|_\infty)$ ist ein Banachraum.*

Beweis: Nach Satz 4.13 ist $(B([a,b];\mathbb{R}), \|\cdot\|_\infty)$ ein Banachraum. Sei $(f_n)_{n\in\mathbb{N}}$ Folge in $C[a,b]$ mit $f_n \to f$ gleichmäßig in $B([a,b])$. Nach einem Satz der Analysis I ist f stetig. Also ist $C[a,b]$ abgeschlossen und nach Satz 4.14 vollständig. $\square$

Definition 4.16 (Stetigkeit)
Seien (X,d_1), (Y,d_2) metrische Räume, sei $D \subset X$, $f : D \to Y$. Ein $c \in Y$ heißt Grenzwert von f in $a \in \overline{D}$, falls $c = \lim_{k\to\infty} f(x_k)$ gilt für alle Folgen $(x_k)_{k\in\mathbb{N}}$ in D mit $\lim x_k = a$. Wir schreiben

$$c = \lim_{\substack{x\to a\\ x\in D}} f(x)\,, \qquad \textit{oder} \quad f(x) \underset{x\to a}{\longrightarrow} c\,.$$

f heißt stetig in $a \in D$, falls gilt

$$\lim_{x\to a} f(x) = f(a)\,.$$

f heißt stetig auf $M \subset D$, falls f in jedem Punkt $a \in M$ stetig ist. Wir definieren

$$C(X;Y) = \{f | f : X \to Y,\ f \text{ stetig}\}\,, \quad C(X) = C(X;\mathbb{R})\,.$$

$\square$

Satz 4.17 *Seien (X,d_1), (Y,d_2), (Z,d_3) metrische Räume, seien $f : X \to Y$, $g : Y \to Z$. Ist f stetig in $a \in X$ und g stetig in $f(a)$, so ist $g \circ f$ stetig in a.*

Beweis: Sei $(x_k)_{k\in\mathbb{N}}$ Folge in X mit $x_k \to a$, dann folgt $f(x_k) \to f(a)$, da f stetig in a, und weiter $g(f(x_k)) \to g(f(a))$. $\square$

Satz 4.18 *Sei (X,d) metrischer Raum, sei $f : X \to \mathbb{K}^n$, $f(x) = (f_1(x),\ldots,f_n(x))$, $a \in X$. Dann ist f stetig in a genau dann, wenn alle Komponentenfunktionen $f_i : X \to \mathbb{K}$ in a stetig sind.*

Beweis: Aus Satz 4.3 folgt, daß $f(x_k) \to f(a)$ gilt genau dann, wenn $f_i(x_k) \to f_i(a)$ für alle i. $\square$

Satz 4.19 *Addition, Subtraktion, Multiplikation und Division definieren stetige Funktionen von $(\mathbb{K}^2, \|\cdot\|_\infty)$ nach $(\mathbb{K}, |\cdot|)$ (die Division $(x_1,x_2) \mapsto x_1/x_2$ auf $D = \{(x_1,x_2) : x_2 \neq 0\}$).*

Beweis: Sei $f(x) = x_1 + x_2$ für $x = (x_1,x_2) \in \mathbb{K}^2$. Sei $(x_k)_{k\in\mathbb{N}}$ konvergente Folge in $\mathbb{K}^2$, $x_k = (x_{k,1}, x_{k_2})$, dann gilt

$$\begin{aligned} f(\lim x_k) &= f((\lim x_{k,1}, \lim x_{k,2})) = \lim x_{k,1} + \lim x_{k,2} && (4.4)\\ &= \lim(x_{k,1} + x_{k,2}) = \lim f(x_k)\,. && (4.5)\end{aligned}$$

Analog für die anderen Operationen. $\square$

Satz 4.20 *Sei (X,d) metrischer Raum, seien $f,g : X \to \mathbb{K}$ stetig in $a \in X$. Dann sind auch $f+g$, $f-g$, $f\cdot g$ und f/g stetig in a (letzteres, falls $g(a) \neq 0$).*

Beweis: $f+g : X \to \mathbb{K}$ läßt sich schreiben als Komposition

$$X \underset{(f,g)}{\longrightarrow} \mathbb{K}^2 \underset{+}{\longrightarrow} \mathbb{K}\,, \quad x \mapsto (f(x), g(x)) \mapsto f(x)+g(x)\,.$$

Die Behauptung folgt aus den Sätzen 4.18, 4.19 und 4.20. Analog für die anderen Operationen. □

Beispiel 4.21

(i) Jede konstante Abbildung $f : \mathbb{K}^n \to \mathbb{K}$ ist stetig.

(ii) Die Abbildungen $p_i : \mathbb{K}^n \to \mathbb{K}$, $p_i(x_1, \dots, x_n) = x_i$, sind stetig.

(iii) Ein $f : \mathbb{K}^n \to \mathbb{K}$ der Form

$$f(x_1, \dots, x_n) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}\,, \quad \alpha_i \in \mathbb{N} \cup \{0\}\,,$$

heißt Monom. Monome sind stetig.

(iv) Sei I eine endliche Teilmenge von $\prod_{i=1}^n (\mathbb{N} \cup \{0\})$. Ein $f : \mathbb{K}^n \to \mathbb{K}$ der Form

$$f(x_1, \dots, x_n) = \sum_{\alpha \in I} c_{\alpha_1\alpha_2\cdots\alpha_n} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

heißt Polynom vom Grad N, wobei

$$N = \max\{\sum_{i=1}^n \alpha_i : \alpha = (\alpha_1, \dots, \alpha_n) \in I\}\,.$$

Polynome sind stetig.

(v) Jede lineare Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$ ist stetig: Die Komponentenfunktionen f_i haben die Form

$$f_i(x) = \sum_{j=1}^n a_{ij} x_j\,,$$

wobei $A = (a_{ij})$ die Matrix ist, welche f in der kanonischen Basis darstellt.

Satz 4.22 *Seien (X, d_1), (Y, d_2) metrische Räume, sei $f : X \to Y$, $a \in X$. Dann ist f stetig in a genau dann, wenn gilt:*

$$\forall\, \varepsilon > 0 \quad \exists\, \delta > 0 \text{ so daß gilt: } d_1(x,a) < \delta \quad \Rightarrow \quad d_2(f(x), f(a)) < \varepsilon\,. \tag{4.6}$$

Beweis: "$\Leftarrow$": Sei (x_k) Folge in X mit $x_k \to a$. Sei $\varepsilon > 0$. Wähle δ gemäß (4.6) und $N \in \mathbb{N}$ so, daß $d_1(x_k, a) < \delta$ für alle $k \geq N$. Dann gilt $d_2(f(x_k), f(a)) < \varepsilon$. Also ist f stetig in a.
"$\Rightarrow$": Kontraposition. Gilt (4.6) nicht, so gibt es ein $\varepsilon > 0$ und eine Folge (x_k) mit

$$d_1(x_k, a) < \frac{1}{k}\,, \quad d_2(f(x_k), f(a)) \geq \varepsilon\,,$$

also $x_k \to a$, aber $f(x_k) \not\to f(a)$. □

Ist (X,d) metrischer Raum, $x_0 \in X$ und $f : X \to \mathbb{R}$ definiert durch

$$f(x) = d(x, x_0)\,,$$

so ist f stetig wegen

$$|f(x) - f(a)| = |d(x, x_0) - d(a, x_0)| \le d(x, a)\,.$$

Insbesondere ist in jedem normierten Raum $(X, \|\cdot\|)$ die Norm

$$f(x) = \|x\|$$

stetig.

Satz 4.23 *Seien $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ normierte Räume, sei $f : X \to Y$ linear. Dann sind äquivalent:*

(i) f ist stetig auf X.

(ii) f ist stetig in 0.

(iii) Es gibt ein $C > 0$ mit

$$\|f(x)\|_Y \le C\|x\|_X\,, \quad \textit{für alle } x \in X. \tag{4.7}$$

Beweis: "(iii)⇒(i)": Sei $a \in X$, (x_k) Folge in X mit $x_k \to a$. Dann gilt

$$0 \le \|f(x_k) - f(a)\|_Y = \|f(x_k - a)\|_Y \le C\|x_k - a\|_X \to 0\,,$$

also $\|f(x_k) - f(a)\|_Y \to 0$, also $f(x_k) \to f(a)$.
"(i)⇒(ii)": klar.
"(ii)⇒(iii)": Wir wenden (4.6) an mit $\varepsilon = 1$ und $a = 0$ (also $f(a) = 0$). Wähle $\delta > 0$ mit $\|f(x)\|_Y < 1$ für alle $x \in X$ mit $\|x\|_X < \delta$. Sei nun $x \in X$ beliebig, $x \neq 0$. Dann ist

$$\left\| \frac{\delta}{2\|x\|_X} x \right\|_X = \frac{\delta}{2} < \delta\,,$$

also

$$\|f(x)\|_Y = \frac{2\|x\|_X}{\delta} \left\| f\left(\frac{\delta}{2\|x\|_X} x \right) \right\|_Y < \frac{2}{\delta}\|x\|_X\,.$$

□

Beispiel 4.24
Wir betrachten $X = C[a,b]$ mit den beiden Normen

$$\|f\|_\infty = \max_{a \le t \le b} |f(t)|\,, \quad \|f\|_1 = \int_a^b |f(t)|\,dt\,.$$

Die lineare Abbildung

$$T : X \to \mathbb{R}\,, \quad T(f) = \int_a^b f(t)\,dt\,, \tag{4.8}$$

ist stetig bezüglich beider Normen, da

$$|T(f)| = \left| \int_a^b f(t)\,dt \right| \le \int_a^b |f(t)|\,dt = \|f\|_1 \le (b-a)\|f\|_\infty\,.$$

Die lineare Abbildung

$$T : X \to \mathbb{R}\,, \quad T(f) = f(a)\,, \tag{4.9}$$

ist stetig bezüglich $\|\cdot\|_\infty$, da $|T(f)| = |f(a)| \le \|f\|_\infty$, aber nicht stetig bezüglich $\|\cdot\|_1$, da für

$$f_n(t) = \max\{1 - n(t-a)\,,\,0\}$$

gilt

$$|T(f_n)| = |f_n(a)| = 1\,,$$

aber (für $n \ge \frac{1}{b-a}$)

$$\|f_n\|_1 = \frac{1}{2n} \to 0\,,$$

und somit (4.7) für kein $C > 0$ erfüllt ist.

Satz 4.25 *Seien (X, d_1), (Y, d_2) metrische Räume, sei $f : X \to Y$. Dann sind äquivalent:*

(i) f ist stetig auf X.

(ii) Das Urbild $f^{-1}(V)$ jeder offenen Menge $V \subset Y$ ist offen in X.

(iii) Das Urbild $f^{-1}(A)$ jeder abgeschlossenen Menge $A \subset Y$ ist abgeschlossen in X.

Beweis: "(i)⇒(iii)": Sei $A \subset Y$ abgeschlossen, sei $(x_k)_{k\in\mathbb{N}}$ Folge in $f^{-1}(A)$ mit $x_k \to x$, $x \in X$. Dann ist $f(x_k) \in A$ für alle $k \in \mathbb{N}$ und $f(x_k) \to f(x)$ (da f stetig), also ist $f(x) \in A$ nach Satz 4.6, also $x \in f^{-1}(A)$. Wiederum nach Satz 4.6 ist $f^{-1}(A)$ abgeschlossen.
"(iii)⇒(ii)": Ist V offen in Y, so ist $Y \setminus V$ abgeschlossen, also auch $f^{-1}(Y \setminus V)$, und damit $f^{-1}(V) = X \setminus f^{-1}(Y \setminus V)$ offen in X.
"(ii)⇒(i)": Sei $(x_k)_{k\in\mathbb{N}}$ Folge in X mit $x_k \to a$. Sei $\varepsilon > 0$. Wir setzen $U = f^{-1}(B(f(a), \varepsilon))$. Dann ist $a \in U$ und U offen in x. Nach Lemma 4.4 gibt es $N \in \mathbb{N}$ mit $x_k \in U$ für alle $k \ge N$, also $f(x_k) \in B(f(a), \varepsilon)$ für alle $k \ge N$. Es folgt $f(x_k) \to f(a)$. □

Folgerung 4.26 *Seien (X, d) metrischer Raum, $f : X \to \mathbb{R}$ stetig, $c \in \mathbb{R}$. Dann sind*

$$\{x : x \in X,\ f(x) < c\}\,, \quad \{x : x \in X,\ f(x) > c\}\,,$$

offen in X, und

$$\{x : x \in X,\ f(x) \le c\}\,, \quad \{x : x \in X,\ f(x) \ge c\}\,,$$

sowie die Niveaumenge von f zum Niveau c,

$$N_c(f) = \{x : x \in X,\ f(x) = c\}\,,$$

abgeschlossen in X.

Beweis: Folgt aus Satz 4.25, da die Mengen $(-\infty, c)$ und (c, ∞) offen in $\mathbb{R}$ und die Mengen $(-\infty, c]$, $[c, \infty)$ und $\{c\}$ abgeschlossen in $\mathbb{R}$ sind. □

5 Metrische Räume: Kompaktheit

Definition 5.1 (Kompakte Menge)
Sei (X, d) metrischer Raum. (X, d) heißt kompakt, falls jede Folge $(x_k)_{k\in\mathbb{N}}$ in X eine konvergente Teilfolge besitzt. □

Ist Y Teilmenge eines metrischen Raums (X, d), so sagen wir abkürzend "Y ist kompakt" statt "(Y, d_Y) ist kompakt". Für eine Teilmenge Y von (X, d) bedeutet also "Y ist kompakt":

> Jede Folge $(x_k)_{k\in\mathbb{N}}$ in Y besitzt eine konvergente Teilfolge, deren Grenzwert ebenfalls ein Y liegt.

Kompaktheit ist also immer eine Eigenschaft eines metrischen Raums und hängt nur von der Metrik ab. Offenheit und Abgeschlossenheit sind Eigenschaften von Teilmengen eines metrischen Raums und hängen nicht nur von der Metrik ab, sondern u.U. auch davon, wie groß der umfassende metrische Raum gewählt wird.

Jedes abgeschlossene beschränkte Intervall $[a, b]$ ist kompakt in der Betragsmetrik (Analysis I).

In allgemeinen topologischen Räumen $(X, \mathcal{T})$ wird Kompaktheit anders definiert, nämlich über die sogenannte "endliche Überdeckungseigenschaft". In metrischen Räumen sind beide Definitionen äquivalent.

Satz 5.2 *Seien (X_1, d_1), (X_2, d_2) kompakte metrische Räume. Dann ist $(X_1 \times X_2, d)$, d Produktmetrik, kompakt.*

Beweis: Sei $(x_k)_{k\in\mathbb{N}}$, $x_k = (x_{k,1}, x_{k,2})$ Folge in $X = X_1 \times X_2$. Wähle eine in X_1 konvergente Teilfolge $(x_{k_j,1})_{j\in\mathbb{N}}$ von $(x_{k,1})$, wähle eine in X_2 konvergente Teilfolge $(x_{k_{j_m},2})_{m\in\mathbb{N}}$ von $(x_{k_j,2})_{j\in\mathbb{N}}$. Dann ist auch $(x_{k_{j_m},1})_{m\in\mathbb{N}}$ in X_1 konvergent, also auch $(x_{k_{j_m}})_{m\in\mathbb{N}}$ konvergent in X. □

Folgerung 5.3 *Seien (X_i, d_i), $1 \le i \le n$, kompakte metrische Räume. Dann ist auch $(\prod_{i=1}^n X_i, d)$, d Produktmetrik, kompakt.*

Beweis: Mit Induktion aus Satz 5.2. □

Folgerung 5.4 *Jeder abgeschlossene achsenparallele Quader*

$$Q = \prod_{i=1}^n [a_i, b_i]$$

mit $a_i, b_i \in \mathbb{R}$, $a_i \le b_i$, ist kompakt in $(\mathbb{R}^n, \|\cdot\|_\infty)$.

Beweis: Folgt aus Folgerung 5.3, da alle $[a_i, b_i]$ kompakt sind. □

Satz 5.5 *Sei (X,d) metrischer Raum, sei $Y \subset X$. Dann gilt:*

$$X \text{ kompakt, } Y \text{ abgeschlossen in } X \quad \Rightarrow \quad Y \text{ kompakt}, \tag{5.1}$$

$$Y \text{ kompakt} \quad \Rightarrow \quad Y \text{ beschränkt und abgeschlossen in } X\,. \tag{5.2}$$

Beweis: Zu (5.1): Sei $(x_k)_{k\in\mathbb{N}}$ Folge in Y, dann gibt es eine in X konvergente Teilfolge, $x_{k_m} \to a$, $a \in X$. Da Y abgeschlossen ist, ist $a \in Y$.
Zu (5.2): Y ist abgeschlossen: Sei $(x_k)_{k\in\mathbb{N}}$ Folge in Y mit $x_k \to a$, $a \in X$. Da Y kompakt, gibt es Teilfolge (x_{k_m}) und $b \in Y$ mit $x_{k_m} \to b$. Es folgt $a = b$, also $a \in Y$.
Y ist beschränkt: Kontraposition. Sei Y unbeschränkt. Wir zeigen

$$\text{es gibt eine Folge } (x_k)_{k\in\mathbb{N}} \text{ in } Y \text{ mit } d(x_k, x_m) > 1 \text{ für alle } k \neq m. \tag{5.3}$$

(Dann sind wir fertig, da keine Teilfolge einer solchen Folge eine Cauchyfolge sein kann.)
Die Folge (x_k) wird rekursiv konstruiert. Wähle $x_1 \in Y$ beliebig. Seien $x_1, \ldots, x_k$ bereits konstruiert. Für alle $y, z \in Y$ und alle $1 \le i, j \le k$ gilt dann

$$d(y,z) \le d(y,x_i) + d(x_i,x_j) + d(x_j,z) \le d(y,x_i) + d(x_j,z) + m\,,$$

wobei $m = \max_{1\le i,j\le k} d(x_i,x_j)$. Es folgt für alle $y, z \in Y$

$$d(y,z) - m \le \min_{1\le i\le k} d(y,x_i) + \min_{1\le j\le k} d(z,x_j)\,.$$

Wähle nun $y, z \in Y$ mit $d(y,z) > m + 2$ (möglich, da Y unbeschränkt), dann ist

$$\min_{1\le i\le k} d(y,x_i) > 1 \qquad \text{oder} \qquad \min_{1\le j\le k} d(z,x_j) > 1\,.$$

Im ersteren Fall setzen wir $x_{k+1} = y$, andernfalls $x_{k+1} = z$. $\square$

Satz 5.6 (Bolzano-Weierstraß)
Sei Y Teilmenge von $(\mathbb{R}^n, \|\cdot\|_\infty)$. Dann gilt

$$Y \text{ kompakt} \quad \Leftrightarrow \quad Y \text{ abgeschlossen in } \mathbb{R}^n \text{ und beschränkt.}$$

Beweis: "$\Rightarrow$": Satz 5.5.
"$\Leftarrow$": Für hinreichend großes $C > 0$ gilt

$$Y \subset Q = \prod_{i=1}^{n} [-C, C]\,.$$

Sei $(x_k)_{k\in\mathbb{N}}$ Folge in Y. Da Q kompakt ist nach Folgerung 5.4 hat $(x_k)_{k\in\mathbb{N}}$ eine bezüglich $\|\cdot\|_\infty$ konvergente Teilfolge, deren Limes auch in Y liegt, da Y abgeschlossen ist. $\square$

Die Implikation "$\Leftarrow$" aus Satz 5.6 gilt in **keinem** unendlichdimensionalen normierten Raum.

Satz 5.7 *Seien (X, d_1), (Y, d_2) metrische Räume, sei $f : X \to Y$ stetig und X kompakt. Dann ist auch $f(X)$ kompakt.*

Beweis: Sei (y_k) Folge in $f(X)$. Für alle $k \in \mathbb{N}$ wähle $x_k \in X$ mit $f(x_k) = y_k$. Da X kompakt, gibt es eine Teilfolge (x_{k_m}) und ein $a \in X$ mit $x_{k_m} \to a$, also auch $y_{k_m} = f(x_{k_m}) \to f(a) \in f(X)$, da f stetig. $\square$

Satz 5.8 *Sei (X, d) metrischer Raum, sei $f : X \to \mathbb{R}$ stetig und X kompakt, $X \neq \emptyset$. Dann ist $f(X)$ beschränkt, und f nimmt auf X sein Maximum und Minimum an, d.h. es gibt $p, q \in X$ mit*

$$f(p) = \sup_{x \in X} f(x) \,, \quad f(q) = \inf_{x \in X} f(x) \,.$$

Beweis: Nach Satz 5.7 ist $f(X)$ kompakt, also beschränkt und abgeschlossen in $\mathbb{R}$ nach Satz 5.5. Aus der Beschränktheit folgt $\sup f(X) < \infty$, aus der Abgeschlossenheit folgt $\sup f(X) \in f(X)$, es gibt also ein $p \in X$ mit $f(p) = \sup_{x \in X} f(x)$. Analog für das Minimum. $\square$

Satz 5.9 *Zu jeder Norm $\| \cdot \|$ auf $\mathbb{R}^n$ gibt es Zahlen $c_1, c_2 > 0$ mit*

$$c_1 \|x\|_\infty \leq \|x\| \leq c_2 \|x\|_\infty \,, \quad \textit{für alle } x \in \mathbb{R}^n. \tag{5.4}$$

Beweis: Konstruktion von c_2: Für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ gilt

$$x = \sum_{i=1}^{n} x_i e_i \,,$$

wobei e_i der i-te Einheitsvektor ist. Es folgt

$$\|x\| \leq \sum_{i=1}^{n} |x_i| \|e_i\| \leq \left(\sum_{i=1}^{n} \|e_i\| \right) \max_{1 \leq j \leq n} |x_j| = \underbrace{\left(\sum_{i=1}^{n} \|e_i\| \right)}_{=: c_2} \|x\|_\infty \,.$$

Existenz von c_1: Wir definieren $f : (\mathbb{R}^n, \| \cdot \|_\infty) \to (\mathbb{R}, | \cdot |)$ durch

$$f(x) = \|x\| \,.$$

f ist stetig, da für jede Folge $(x_k)_{k \in \mathbb{N}}$ in $\mathbb{R}^n$ mit $x_k \to a$ gilt

$$0 \leq |f(x_k) - f(a)| = | \, \|x_k\| - \|a\| \, | \leq \|x_k - a\| \leq c_2 \|x_k - a\|_\infty \,,$$

also auch $f(x_k) \to f(a)$. Wir definieren

$$Y = \{x : x \in \mathbb{R}^n, \; \|x\|_\infty = 1\} \,.$$

Y ist abgeschlossene beschränkte Teilmenge von $\mathbb{R}^n$, also kompakt nach Satz 5.6, also gibt es nach Satz 5.8 ein $y \in \mathbb{R}^n$ mit $\|y\|_\infty = 1$ und

$$0 < \|y\| = \min_{x \in Y} \|x\| \,.$$

Wir setzen $c_1 = \|y\|$. Es gilt dann für alle $x \in \mathbb{R}^n$ mit $x \neq 0$, daß

$$c_1 \leq \left\| \frac{x}{\|x\|_\infty} \right\| \,,$$

also gilt die linke Ungleichung in (5.4) für alle $x \in \mathbb{R}^n$. $\square$

Aus Satz 5.9 folgt: Ist $(x_k)_{k \in \mathbb{N}}$ Folge in $\mathbb{R}^n$, so hängt die Richtigkeit der Aussagen

(x_k) ist konvergent in $(\mathbb{R}^n, \|\cdot\|)$
(x_k) ist Cauchyfolge bezüglich $(\mathbb{R}^n, \|\cdot\|)$

nicht von der Wahl der Norm $\|\cdot\|$ ab, “in allen Normen sind die gleichen Folgen konvergent”. Ebenso gilt: Ist $Y \subset \mathbb{R}^n$, so hängen die Mengen $\operatorname{int}(Y)$, $\overline{Y}$, ∂Y sowie die Eigenschaften “offen”, “abgeschlossen”, “kompakt” und “beschränkt” ebenfalls nicht davon ab, aus welcher Norm man die Metrik erzeugt. Siehe Übung.

Definition 5.10 (Äquivalenz von Normen)
Sei X Vektorraum, seien $\|\cdot\|_1$, $\|\cdot\|_2$ Normen auf X. Die beiden Normen heißen äquivalent, falls es Konstanten $c_1, c_2 > 0$ gibt mit

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

für alle $x \in X$. □

Satz 5.9 erhält somit die prägnante Form:

“Auf dem $\mathbb{R}^n$ sind alle Normen äquivalent.”

Definition 5.11 (Gleichmäßige Stetigkeit)
Seien (X, d_1), (Y, d_2) metrische Räume, sei $f : X \to Y$. Dann heißt f gleichmäßig stetig auf X, falls gilt

$$\forall\, \varepsilon > 0 \quad \exists\, \delta > 0 \quad \forall\, x, y \in X : \quad \textit{Aus } d_1(x,y) < \delta \textit{ folgt } d_2(f(x), f(y)) < \varepsilon. \tag{5.5}$$

□

Offensichtlich ist jede auf X gleichmäßig stetige Funktion auch stetig auf X.

Satz 5.12 *Seien (X, d_1), (Y, d_2) metrische Räume, sei $f : X \to Y$. Dann gilt: Ist f stetig auf X und ist X kompakt, so ist f gleichmäßig stetig auf X.*

Beweis: Wir nehmen an, f sei nicht gleichmäßig stetig. Dann gibt es ein $\varepsilon > 0$ und für alle $k \in \mathbb{N}$ Elemente $x_k, y_k \in X$ mit

$$d_1(x_k, y_k) < \frac{1}{k}, \quad \text{aber} \quad d_2(f(x_k), f(y_k)) \geq \varepsilon\,. \tag{5.6}$$

Sei (x_{k_m}) konvergente Teilfolge von (x_k) mit $x_{k_m} \to a \in X$. Aus

$$d_1(y_{k_m}, a) \leq d_1(y_{k_m}, x_{k_m}) + d_1(x_{k_m}, a) < \frac{1}{k_m} + d_1(x_{k_m}, a)$$

folgt auch $y_{k_m} \to a$. Es gilt

$$0 \leq d_2(f(y_{k_m}), f(x_{k_m})) \leq d_2(f(x_{k_m}), f(a)) + d_2(f(y_{k_m}), f(a)) \to 0\,,$$

da f stetig in a, im Widerspruch zu (5.6). □

Definition 5.13 (Abstand von Mengen)
Sei (X,d) metrischer Raum. Wir definieren den Abstand eines Punktes $x \in X$ von einer Teilmenge $A \subset X$ durch

$$\operatorname{dist}(x,A) = \inf_{a \in A} d(x,a)\,, \tag{5.7}$$

und den Abstand zweier Teilmengen $A, B \subset X$ durch

$$\operatorname{dist}(A,B) = \inf_{\substack{a \in A \\ b \in B}} d(a,b)\,. \tag{5.8}$$

Durch 'dist' wird keine Metrik auf der Potenzmenge von X definiert, da $\operatorname{dist}(A,B) = 0$ falls $A \cap B \neq \emptyset$. Es gibt aber andere Abstandsbegriffe für Mengen, die zu einer Metrik (auf einer geeigneten Teilmenge von $\mathcal{P}(X)$) führen, der wichtigste davon ist wohl der Begriff des Hausdorff-Abstands (später).

Unsere Konvention über das Infimum leerer Mengen führt zu

$$\operatorname{dist}(x,\emptyset) = +\infty\,, \quad \operatorname{dist}(A,\emptyset) = +\infty\,.$$

Satz 5.14 *Sei (X,d) metrischer Raum, sei $A \subset X$, $A \neq \emptyset$. Dann gilt*

$$|\operatorname{dist}(x,A) - \operatorname{dist}(y,A)| \le d(x,y)\,, \tag{5.9}$$

für alle $x, y \in X$, und die durch

$$f(x) = \operatorname{dist}(x,A)$$

definierte Funktion $f : X \to \mathbb{R}$ ist gleichmäßig stetig.

Beweis: Für alle $x, y \in X$ und alle $a \in A$ gilt

$$\operatorname{dist}(y,A) \le d(y,a) \le d(y,x) + d(x,a)\,,$$

also gilt für alle $x, y \in X$

$$\operatorname{dist}(y,A) \le d(y,x) + \operatorname{dist}(x,A)\,,$$

und weiter

$$|\operatorname{dist}(y,A) - \operatorname{dist}(x,A)| \le d(y,x)\,.$$

Gleichmäßige Stetigkeit: Wähle $\delta = \varepsilon$ in der Definition. □

Satz 5.15 *Sei (X,d) metrischer Raum, seien $A, K \subset X$ abgeschlossen mit $A \cap K = \emptyset$. Dann gilt: Ist K kompakt, so ist* $\operatorname{dist}(A,K) > 0$.

Beweis: Es ist

$$\operatorname{dist}(A,K) = \inf_{\substack{a \in A \\ x \in K}} d(a,x) = \inf_{x \in K} \inf_{a \in A} d(a,x) = \inf_{x \in K} \operatorname{dist}(x,A)\,.$$

Wegen Satz 5.14 und Satz 5.8 wird rechts das Minimum angenommen, also gibt es $q \in K$ mit

$$\operatorname{dist}(q,A) = \operatorname{dist}(A,K)\,.$$

Da A abgeschlossen und $q \notin A$, gibt es $\varepsilon > 0$ mit $B(q, \varepsilon) \cap A = \emptyset$, also $\operatorname{dist}(q, A) \geq \varepsilon > 0$.
$\square$

Ohne die Voraussetzung der Kompaktheit von K stimmt die Aussage von Satz 5.15 im allgemeinen nicht, ist etwa $X = \mathbb{R}^2$, $A = \{(x, 0) : x \in \mathbb{R}\}$ die x-Achse, und $K = \{(x, \exp(x)) : x \in \mathbb{R}\}$ der Graph der Exponentialfunktion, dann sind A und K abgeschlossen, $A \cap K = \emptyset$, aber $\operatorname{dist}(A, K) = 0$.

6 Kurven im $\mathbb{R}^n$

Definition 6.1 (Kurve)
Sei $I \subset \mathbb{R}$ Intervall, (X, d) metrischer Raum.

(i) Jede stetige Funktion $f : I \to X$ heißt Kurve (in X). Die Bildmenge $f(I)$ heißt Spur der Kurve f. Die Elemente $f(t)$ der Spur von f heißen Kurvenpunkte.

(ii) Sei $X = \mathbb{R}^n$. Ein $f = (f_1, \dots, f_n) : I \to \mathbb{R}^n$ heißt (stetig) differenzierbar, falls alle $f_i : I \to \mathbb{R}$ (stetig) differenzierbar sind. Die durch

$$f'(t) = \lim_{\substack{h \to 0 \\ h \neq 0}} \frac{f(t+h) - f(t)}{h} = (f_1'(t), \dots, f_n'(t)) \in \mathbb{R}^n \tag{6.1}$$

definierte Ableitung von f in $t \in I$ heißt Tangentialvektor an f in t. Falls $f'(t) \neq 0$, so heißt

$$\frac{f'(t)}{\|f'(t)\|_2}$$

der Tangenteneinheitsvektor an f in t.

□

Wir betrachten eine Reihe von Beispielen. Für $a, v \in \mathbb{R}^n$, $v \neq 0$, wird durch

$$f(t) = a + tv$$

eine Kurve $f : \mathbb{R} \to \mathbb{R}^n$ definiert, nämlich die Gerade durch a mit Richtung v. Es gilt

$$f'(t) = v\,, \quad \frac{f'(t)}{\|f'(t)\|_2} = \frac{v}{\|v\|_2}\,.$$

Weiter:

$$f : [0, 2\pi] \to \mathbb{R}^2\,, \quad f(t) = (r\cos t, r\sin t)\,, \quad r > 0\,,$$

definiert eine Kurve, deren Spur der Kreis um 0 mit Radius r ist. Es ist

$$f'(t) = (-r\sin t, r\cos t)\,, \quad \|f'(t)\|_2 = r\,, \frac{f'(t)}{\|f'(t)\|_2} = (-\sin t, \cos t)\,.$$

Eine Schraubenlinie erhalten wir durch

$$f : \mathbb{R} \to \mathbb{R}^3\,, \quad f(t) = (r\cos t, r\sin t, ct)\,, \quad r > 0\,, c > 0\,.$$

Es ist

$$f'(t) = (-r\sin t, r\cos t, c)\,.$$

Für eine beliebige Kurve $\varphi : I \to \mathbb{R}^n$ können wir $\operatorname{graph}(\varphi)$ als Kurve im $\mathbb{R}^{n+1}$ auffassen gemäß $f(t) = (t, \varphi(t))$.

Verschiedene Kurven können dieselbe Spur haben, beispielsweise

$$f : [0, 2\pi] \to \mathbb{R}^2\,, \qquad f(t) = (r\cos t, r\sin t)\,, \tag{6.2}$$
$$f : \mathbb{R} \to \mathbb{R}^2\,, \qquad f(t) = (r\cos t, r\sin t)\,, \tag{6.3}$$
$$f : [0, \pi] \to \mathbb{R}^2\,, \qquad f(t) = (r\cos 2t, r\sin 2t)\,. \tag{6.4}$$

Eine Kurve muß nicht injektiv sein.

Definition 6.2 (Doppelpunkt)
Sei $I \subset \mathbb{R}$ Intervall, (X,d) metrischer Raum, $f : I \to X$ Kurve. Ein $x \in X$ heißt Doppelpunkt von f, falls es $t, s \in I$ gibt mit $t \neq s$ und $x = f(t) = f(s)$. □

Für die Kurve (6.3) gilt etwa, daß jeder Kurvenpunkt ein Doppelpunkt ist. Die Kurve

$$f : \mathbb{R} \to \mathbb{R}^2 , \quad f(t) = (t^2 - 1, t^3 - t) ,$$

hat den Doppelpunkt $0 = f(1) = f(-1)$.

Die Ordnungsrelation "$\le$" auf I definiert eine Orientierung für eine Kurve $f : I \to X$. Sind $t, s \in I$ mit $t < s$, so sagen wir, daß $f(t)$ vor $f(s)$ liegt. Ist $I = [a,b]$, so heißt $f(a)$ der Anfangspunkt und $f(b)$ der Endpunkt der Kurve. Die gleiche Spur kann in verschiedene Richtungen durchlaufen werden, etwa

$$f : [0, 2\pi] \to \mathbb{R}^2 , \qquad f(t) = (r \cos t, r \sin t) , \tag{6.5}$$

$$f : [0, 2\pi] \to \mathbb{R}^2 , \qquad f(t) = (r \cos t, -r \sin t) . \tag{6.6}$$

Definition 6.3 (Reguläre Kurve)
Sei $I \subset \mathbb{R}$ Intervall, $f : I \to \mathbb{R}^n$ Kurve. f heißt regulär, falls f stetig differenzierbar auf I ist und $f'(t) \neq 0$ gilt für alle $t \in I$. Ein $t \in I$ mit $f'(t) = 0$ heißt singuläre Stelle von f. Ist $I = [a,b]$, so heißt f stückweise regulär, falls es eine Zerlegung $a = t_0 < t_1 < \ldots < t_N = b$ gibt, so daß $f|[t_j, t_{j+1}]$ regulär ist für alle j. □

Die Kurve

$$f : \mathbb{R} \to \mathbb{R}^2 , \quad f(t) = (t^2 - 1, t^3 - t) ,$$

ist regulär. Die Kurve

$$f : \mathbb{R} \to \mathbb{R}^2 , \quad f(t) = (t^2, t^3) ,$$

ist nicht regulär; sie hat die singuläre Stelle 0. Die Kurve

$$f : [0, 2\pi] \to \mathbb{R}^2 , \quad f(t) = (r \cos(t^2), r \sin(t^2)) ,$$

hat ebenfalls die singuläre Stelle 0; ihre Spur ist der Kreis um 0 mit Radius r. (Dieselbe Spur kann also durch reguläre und nicht reguläre Kurven erzeugt werden.) Der Rand eines Rechtecks im $\mathbb{R}^2$ ist Spur einer stückweise regulären Kurve $f : [0,1] \to \mathbb{R}^2$.

Definition 6.4 (Schnittpunkt, Schnittwinkel)
Seien $f : I \to \mathbb{R}^n$, $g : J \to \mathbb{R}^n$ reguläre Kurven. Gibt es $t \in I$, $\tau \in J$ mit $x = f(t) = g(\tau)$, so heißt x Schnittpunkt von f und g. Der Winkel θ zwischen den Tangentialvektoren $f'(t)$ und $g'(\tau)$ heißt Schnittwinkel von f und g in x. θ ist definiert durch

$$\cos \theta = \frac{\langle f'(t), g'(\tau) \rangle}{\|f'(t)\|_2 \|g'(\tau)\|_2} . \tag{6.7}$$

□

Ist $f : I \to \mathbb{R}^n$ Kurve, sind $f(t_0), \ldots, f(t_N)$ Kurvenpunkte, so ist die Länge des Polygonzugs, der diese Kurvenpunkte nacheinander verbindet, gegeben durch

$$\sum_{j=1}^{N} \|f(t_j) - f(t_{j-1})\| .$$

Definition 6.5 (Länge einer Kurve)
Sei $(X, \|\cdot\|)$ *normierter Raum, sei* $f : [a, b] \to X$ *Kurve. Für eine Zerlegung* $Z = (t_0, \ldots, t_N)$ *von* $[a, b]$*, also* $a = t_0 < t_1 < \ldots < t_N = b$*, definieren wir deren Feinheit*

$$|Z| = \max_{1 \le j \le N} (t_j - t_{j-1}) , \tag{6.8}$$

sowie

$$L(f, Z) = \sum_{j=1}^{N} \|f(t_j) - f(t_{j-1})\| . \tag{6.9}$$

Dann heißt

$$L(f) = \sup_{Z \in \mathcal{Z}} L(f, Z) , \quad \mathcal{Z} = \{Z : Z \text{ Zerlegung von } [a, b]\} , \tag{6.10}$$

die Länge der Kurve f *in* $(X, \|\cdot\|)$*. Die Kurve* f *heißt rektifizierbar, falls* $L(f) < \infty$*.* □

Für Kurven $f : [a, b] \to \mathbb{R}^n$ hängt die Antwort auf die Frage, ob f rektifizierbar ist, nicht von der Wahl der Norm ab (da im $\mathbb{R}^n$ alle Normen äquivalent sind), wohl aber die Länge $L(f)$. Für $\|\cdot\| = \|\cdot\|_2$ erhalten wir das, was wir uns üblicherweise als Länge vorstellen.

Definition 6.6 *Sei* $f : [a, b] \to \mathbb{R}^n$ *Kurve. Wir definieren*

$$\int_a^b f(t)\, dt = \left(\int_a^b f_1(t)\, dt , \ldots , \int_a^b f_n(t)\, dt \right) . \tag{6.11}$$

□

Satz 6.7 *Sei* $f : [a, b] \to (\mathbb{R}^n, \|\cdot\|)$ *eine stetig differenzierbare Kurve. Dann ist* f *rektifizierbar, und*

$$L(f) = \int_a^b \|f'(t)\|\, dt = \lim_{\substack{\delta \to 0 \\ |Z| \le \delta, Z \in \mathcal{Z}}} L(f, Z) , \tag{6.12}$$

das heißt: Für alle $\varepsilon > 0$ *gibt es ein* $\delta > 0$*, so daß für alle Zerlegungen* Z *von* $[a, b]$ *mit* $|Z| \le \delta$ *gilt*

$$|L(f) - L(f, Z)| < \varepsilon .$$

Beweis: Wir zeigen die Behauptung (B): Für alle $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß für alle Zerlegungen Z von $[a, b]$ mit $|Z| \le \delta$ gilt

$$\left| \int_a^b \|f'(t)\|\, dt - L(f, Z) \right| < \varepsilon . \tag{6.13}$$

Sei $Z = (t_j)$, $a = t_0 < t_1 < \ldots < t_N = b$, Zerlegung von $[a,b]$. Dann ist die Abbildung $t \mapsto \|f'(t)\|$ stetig, und es gilt für alle j, $1 \leq j \leq N$

$$\begin{aligned}
\left| \int_{t_{j-1}}^{t_j} \|f'(t)\|\, dt - \|f(t_j) - f(t_{j-1})\| \right| &= \left| \int_{t_{j-1}}^{t_j} \|f'(t)\| - \frac{\|f(t_j) - f(t_{j-1})\|}{t_j - t_{j-1}}\, dt \right| \\
&\leq \int_{t_{j-1}}^{t_j} \left| \|f'(t)\| - \frac{\|f(t_j) - f(t_{j-1})\|}{t_j - t_{j-1}} \right| dt \\
&\leq \int_{t_{j-1}}^{t_j} \left\| f'(t) - \frac{f(t_j) - f(t_{j-1})}{t_j - t_{j-1}} \right\| dt \\
&= \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} \|f(t_j) - f(t_{j-1}) - (t_j - t_{j-1}) f'(t)\|\, dt \\
&= \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} \left\| \int_{t_{j-1}}^{t_j} f'(s)\, ds - (t_j - t_{j-1}) f'(t) \right\| dt \\
&= \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} \left\| \int_{t_{j-1}}^{t_j} f'(s) - f'(t)\, ds \right\| dt \\
&\leq \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} c \max_{1 \leq i \leq n} \left(\left| \int_{t_{j-1}}^{t_j} f_i'(s) - f_i'(t)\, ds \right| \right) dt\,,
\end{aligned} \tag{6.14}$$

letzteres wegen der Äquivalenz aller Normen im $\mathbb{R}^n$, mit einer von j unabhängigen Konstante c. Wir setzen nun für $\delta > 0$

$$\alpha(\delta) = \max_{1 \leq i \leq n} \max_{\substack{|s-t| \leq \delta \\ a \leq s,t \leq b}} |f_i'(s) - f_i'(t)|\,.$$

Dann ist $\alpha(\delta) < \infty$, da alle f_i' stetig sind und $[a,b]$ kompakt ist, und es gilt

$$\lim_{\delta \to 0} \alpha(\delta) = 0\,, \tag{6.15}$$

da alle f_i' gleichmäßig stetig sind auf $[a,b]$ nach Satz 5.12. Aus (6.14) folgt also

$$\begin{aligned}
\left| \int_{t_{j-1}}^{t_j} \|f'(t)\|\, dt - \|f(t_j) - f(t_{j-1})\| \right| &\leq \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} c \max_{1 \leq i \leq n} \left(\left| \int_{t_{j-1}}^{t_j} f_i'(s) - f_i'(t)\, ds \right| \right) dt \\
&\leq \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} c(t_j - t_{j-1}) \alpha(|Z|)\, dt = c(t_j - t_{j-1}) \alpha(|Z|)\,,
\end{aligned} \tag{6.16}$$

und weiter

$$\begin{aligned}
\left| \int_a^b \|f'(t)\|\, dt - L(f,Z) \right| &= \left| \sum_{j=1}^N \left[\int_{t_{j-1}}^{t_j} \|f'(t)\|\, dt - \|f(t_j) - f(t_{j-1})\| \right] \right| \\
&\leq \sum_{j=1}^N c(t_j - t_{j-1}) \alpha(|Z|) = c(b-a) \alpha(|Z|)\,.
\end{aligned} \tag{6.17}$$

Für $\varepsilon > 0$ wähle $\delta > 0$ so, daß

$$\alpha(\delta) < \frac{\varepsilon}{c(b-a)}\,,$$

das ist möglich wegen (6.15), und hieraus folgt die Behauptung (B) und damit die zweite Gleichung in (6.12). Sind nun Z, Z' Zerlegungen, und ist Z' Verfeinerung von Z, so folgt mit Hilfe der Dreiecksungleichung

$$L(f,Z) \le L(f,Z')\,.$$

Hieraus folgt

$$L(f) = \sup_{Z\in\mathcal{Z}} L(f,Z) = \sup_{\substack{Z\in\mathcal{Z}\\ |Z|\le\delta}} L(f,Z)$$

für alle $\delta > 0$, und nach dem oben Bewiesenen auch die erste Gleichung in (6.12). □

Die Länge eines Kreisbogens (gemessen in der euklidischen Norm)

$$f : [0,\varphi] \to \mathbb{R}^2\,, \quad f(t) = (\cos t, \sin t)\,,$$

errechnet sich also wegen

$$f'(t) = (-\sin t, \cos t)\,, \quad \|f'(t)\|_2 = \sqrt{\sin^2(t) + \cos^2(t)} = 1$$

zu

$$L(f) = \int_0^\varphi \|f'(t)\|_2\, dt = \varphi\,.$$

Die Länge des Graphen einer Funktion können wir ebenfalls berechnen. Ist $f : [a,b] \to \mathbb{R}$ stetig differenzierbar, so setzen wir

$$\tilde{f} : [a,b] \to \mathbb{R}^2\,, \quad \tilde{f}(t) = (t, f(t))\,,$$

also

$$L(\tilde{f}) = \int_a^b \|\tilde{f}'(t)\|_2\, dt = \int_a^b \sqrt{1 + f'(t)^2}\, dt\,.$$

Definition 6.8 (Parametertransformation)
Sei $\varphi : [\alpha,\beta] \to [a,b]$ bijektiv und stetig. Dann heißt φ Parametertransformation von $[a,b]$ auf $[\alpha,\beta]$. Ist $f : [a,b] \to \mathbb{R}^n$ Kurve, so sagen wir, daß

$$g : [\alpha,\beta] \to \mathbb{R}^n\,, \quad g = f \circ \varphi\,,$$

aus f durch die Parametertransformation φ hervorgeht. Falls φ und φ^{-1} stetig differenzierbar sind, heißt φ eine C^1-Parametertransformation; falls zusätzlich $\varphi'(t) > 0$ ($\varphi'(t) < 0$) gilt für alle $t \in [\alpha,\beta]$, so heißt φ orientierungstreu (orientierungsumkehrend). □

Die Parametertransformation

$$\varphi : [a,b] \to [a,b]\,, \quad \varphi(t) = a + b - t\,,$$

kehrt die Orientierung einer Kurve $f : [a,b] \to \mathbb{R}^n$ um,

$$(f\circ\varphi)(t) = f(a+b-t)\,, \quad (f\circ\varphi)(b) = f(a)\,, \quad (f\circ\varphi)(a) = f(b)\,.$$

Ist $f : [a,b] \to \mathbb{R}^n$ eine differenzierbare Kurve, $\varphi : [\alpha,\beta] \to [a,b]$ eine C^1-Parametertransformation, so gilt für $g = f \circ \varphi$

$$g'(\tau) = f'(\varphi(\tau))\varphi'(\tau)\,, \quad \tau \in [\alpha,\beta]\,,$$

das heißt: Die Tangentialvektoren von g in τ und von f in $\varphi(\tau)$ sind parallel, die Tangenteneinheitsvektoren sind entweder gleich oder entgegengesetzt.

Satz 6.9 *Sei $f : [a,b] \to \mathbb{R}^n$ stetig differenzierbar, sei $\varphi : [\alpha,\beta] \to [a,b]$ eine C^1-Parametertransformation. Dann gilt*

$$L(f) = L(f \circ \varphi) \tag{6.18}$$

Beweis: Aus Satz 6.7 folgt mit der Kettenregel (komponentenweise angewendet) und der Substitutionsregel

$$\begin{aligned}
L(f \circ \varphi) &= \int_\alpha^\beta \|(f \circ \varphi)'(\tau)\|\, d\tau = \int_\alpha^\beta \|(f'(\varphi(\tau))\varphi'(\tau)\|\, d\tau \\
&= \int_\alpha^\beta \|(f'(\varphi(\tau))\|\, |\varphi'(\tau)|\, d\tau = \operatorname{sign}(\varphi') \int_\alpha^\beta \|(f'(\varphi(\tau))\|\, \varphi'(\tau)\, d\tau \\
&= \operatorname{sign}(\varphi') \int_{\varphi(\alpha)}^{\varphi(\beta)} \|f'(t)\|\, dt \\
&= L(f)\,.
\end{aligned} \tag{6.19}$$

□

Sei $f : [a,b] \to \mathbb{R}^n$ eine reguläre Kurve. Dann ist

$$l(t) = \int_a^t \|f'(s)\|\, ds$$

die Länge des Kurvenstücks vom Anfangspunkt $f(a)$ bis zum Punkt $f(t)$. Es ist $l \in C^1[a,b]$ und, da f regulär,

$$l'(t) = \|f'(t)\| > 0\,,$$

also ist

$$l^{-1} : [0, L(f)] \to [a,b]$$

eine C^1-Parametertransformation. Die Kurve

$$g = f \circ l^{-1}$$

entsteht aus f durch "Parametrisierung mit der Bogenlänge". Es gilt

$$g'(\tau) = f'(l^{-1}(\tau))(l^{-1})'(\tau) = \frac{f'(l^{-1}(\tau))}{l'(l^{-1}(\tau))} = \frac{f'(l^{-1}(\tau))}{\|f'(l^{-1}(\tau))\|}\,,$$

also ist $\|g'(\tau)\| = 1$ für alle τ und

$$\int_0^\tau \|g'(s)\|\, ds = \tau\,.$$

7 Partielle Ableitungen, Skalar- und Vektorfelder

Definition 7.1 (Skalarfeld)
Sei $\Omega \subset \mathbb{R}^n$. Ein $f : \Omega \to \mathbb{R}$ heißt skalares Feld, oder Skalarfeld, auf Ω. □

Skalarfelder kann man sich durch ihre Niveaumengen

$$N_c(f) = \{x : x \in \Omega,\ f(x) = c\}$$

veranschaulichen, besonders gut dann, wenn $\Omega \subset \mathbb{R}^2$.

Sei $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$. Wir betrachten f "längs einer Geraden", d.h. für einen gegebenen Punkt $x \in \Omega$ und eine gegebene "Richtung" $v \in \mathbb{R}^n$ betrachten wir

$$g(t) = f(x + tv)\,, \quad g : D_g \to \mathbb{R}\,, \quad D_g = \{t : t \in \mathbb{R},\ x + tv \in \Omega\}\,.$$

Falls $g'(0)$ existiert, gibt diese Zahl die "Ableitung von f in die Richtung v" an.

Definition 7.2 (Richtungsableitung, partielle Ableitung)
Sei $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$. Ist $x \in \Omega$ und $v \in \mathbb{R}^n$, so heißt

$$\partial_v f(x) = \lim_{\substack{t \to 0, t \neq 0 \\ x + tv \in \Omega}} \frac{f(x + tv) - f(x)}{t} \tag{7.1}$$

die Richtungsableitung von f an der Stelle x in Richtung v, falls der Grenzwert existiert. Ist $v = e_i$ der i-te Einheitsvektor, so sprechen wir von der i-ten partiellen Ableitung von f an der Stelle x, wir bezeichnen sie mit

$$\partial_i f(x)\,.$$

Andere gebräuchliche Schreibweisen für $\partial_i f(x)$ sind

$$\frac{\partial}{\partial x_i} f(x)\,, \quad D_i f(x)\,.$$

Die Funktion f heißt partiell differenzierbar in Ω, falls $\partial_i f(x)$ existiert für alle $x \in \Omega$ und alle i, $1 \leq i \leq n$. Eine in Ω partiell differenzierbare Funktion f heißt stetig differenzierbar in Ω, falls die Funktionen

$$\partial_i f : \Omega \to \mathbb{R}$$

stetig sind für alle i. □

Wir können die i-te partielle Ableitung auch definieren als

$$\partial_i f(x) = g_i'(x_i)\,, \quad g_i(\xi) := f(x_1, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_n)\,,$$

diese Definition ist äquivalent zur obigen.

Beispiel 7.3 (i) $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1 \sin x_3 + x_1^2 x_2^2 \exp(x_2 x_3)$. Es ist

$$\begin{aligned}
\partial_1 f(x_1, x_2, x_3) &= \sin x_3 + 2x_1 x_2^2 \exp(x_2 x_3) \\
\partial_2 f(x_1, x_2, x_3) &= 2x_1^2 x_2 \exp(x_2 x_3) + x_1^2 x_2^2 \exp(x_2 x_3) x_3 \\
\partial_3 f(x_1, x_2, x_3) &= x_1 \cos x_3 + x_1^2 x_2^2 \exp(x_2 x_3) x_2\,.
\end{aligned}$$

(ii) Wir betrachten

$$r : \mathbb{R}^n \to \mathbb{R}\,, \quad r(x) = \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}\,.$$

r ist stetig differenzierbar in $\mathbb{R}^n \setminus \{0\}$,

$$\partial_i r(x) = \frac{1}{2\sqrt{\sum_{j=1}^n x_j^2}} 2x_i = \frac{x_i}{r(x)}\,. \tag{7.2}$$

(iii) Sei $g : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ (stetig) differenzierbar, sei $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ definiert durch

$$f(x) = g(r(x))\,. \tag{7.3}$$

Dann ist auch f (stetig) differenzierbar, und

$$\partial_i f(x) = g'(r(x))\partial_i r(x) = g'(r(x))\frac{x_i}{r(x)}\,. \tag{7.4}$$

□

Aus der Existenz aller partiellen Ableitungen einer Funktion f folgt noch nicht die Stetigkeit von f! Beispiel dazu: $f : \mathbb{R}^n \to \mathbb{R}$, $n \geq 2$, definiert durch

$$f(0) = 0\,, \quad f(x) = \frac{x_1 \cdots x_n}{\|x\|_2^{2n}}\,,\; x \neq 0\,.$$

In $x = 0$ existieren alle partiellen Ableitungen von f und sind gleich Null, da

$$\frac{f(0 + te_i) - f(0)}{t} = \frac{0 - 0}{t} = 0\,.$$

f ist aber nicht stetig in 0: Betrachten wir

$$a_k = \left(\frac{1}{k}, \ldots, \frac{1}{k}\right)\,,$$

so ist

$$\|a_k\|_2 = \frac{\sqrt{n}}{k}\,, \quad f(a_k) = \left(\frac{1}{k}\right)^n \cdot \left(\frac{k}{\sqrt{n}}\right)^{2n} = \frac{k^n}{n^n}\,,$$

also

$$\lim_{k\to\infty} a_k = 0\,, \quad \lim_{k\to\infty} f(a_k) = \infty\,, \quad f(0) = 0\,.$$

(Die partiellen Ableitungen werten nur das Verhalten von f entlang der Koordinatenachsen aus.)

Definition 7.4 (Höhere partielle Ableitungen)
Sei $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$. f heißt $(k+1)$-mal partiell differenzierbar in Ω, falls f k-mal partiell differenzierbar in Ω ist und alle k-ten partiellen Ableitungen der Form

$$\partial_{i_k}\partial_{i_{k-1}} \ldots \partial_{i_1} f\,, \quad 1 \leq i_j \leq n\,, \quad 1 \leq j \leq k\,,$$

partiell differenzierbar sind. f heißt k-mal stetig differenzierbar, falls alle partielle Ableitungen der Ordnung $\leq k$ existieren und stetig sind. Wir definieren

$$\begin{aligned} C^0(\Omega) &= C(\Omega)\,, \\ C^k(\Omega) &= \{f \mid f : \Omega \to \mathbb{R} \text{ } k\text{-mal stetig differenzierbar}\}\,, \\ C^\infty(\Omega) &= \bigcap_{k\in\mathbb{N}} C^k(\Omega)\,. \end{aligned} \tag{7.5}$$

Funktionen in $C^\infty(\Omega)$ heißen unendlich oft differenzierbar. □

Satz 7.5 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^2(\Omega)$. Dann gilt*

$$\partial_j \partial_i f(a) = \partial_i \partial_j f(a)\,, \tag{7.6}$$

für alle $a \in \Omega$ und alle $1 \leq i, j \leq n$.

Beweis: Sei o.B.d.A. $n = 2$, $i = 1$, $j = 2$, $a = 0 \in \Omega$. Wir wählen $\delta > 0$ mit $B(0,\delta) \subset \Omega$ (Kugel bezüglich der Maximumnorm). Sei $(x_1, x_2) \in B(0,\delta)$. Wir wenden den Mittelwertsatz der Differentialrechnung an auf

$$g(t) = f(t, x_2) - f(t, 0)$$

und erhalten die Existenz eines $\xi \in (0, x_1)$ mit

$$g(x_1) = g(0) + x_1 g'(\xi)\,,$$

also

$$f(x_1, x_2) - f(x_1, 0) = f(0, x_2) - f(0,0) + x_1(\partial_1 f(\xi, x_2) - \partial_1 f(\xi, 0))\,. \tag{7.7}$$

Wir wenden nun den Mittelwertsatz an auf $h(t) = \partial_1 f(\xi, t)$ und erhalten die Existenz eines $\eta \in (0, x_2)$ mit

$$h(x_2) = h(0) + x_2 h'(\eta)\,,$$

also aus (7.7)

$$f(x_1, x_2) - f(x_1, 0) = f(0, x_2) - f(0,0) + x_1 x_2 \partial_2 \partial_1 f(\xi, \eta)\,. \tag{7.8}$$

Vertauschen der Rollen von x_1 und x_2 liefert $\tilde{\xi} \in (0, x_1)$, $\tilde{\eta} \in (0, x_2)$ mit

$$f(x_1, x_2) - f(0, x_2) = f(x_1, 0) - f(0,0) + x_1 x_2 \partial_1 \partial_2 f(\tilde{\xi}, \tilde{\eta})\,, \tag{7.9}$$

also

$$\partial_2 \partial_1 f(\xi, \eta) = \partial_1 \partial_2 f(\tilde{\xi}, \tilde{\eta})\,.$$

Grenzübergang $(x_1, x_2) \to (0,0)$ ergibt $(\xi, \eta) \to (0,0)$ und $(\tilde{\xi}, \tilde{\eta}) \to (0,0)$, also wegen der Stetigkeit der zweiten partiellen Ableitungen

$$\partial_2 \partial_1 f(0,0) = \partial_1 \partial_2 f(0,0)\,.$$

□

Folgerung 7.6 *Für $f \in C^n(\Omega)$ können wir die Reihenfolge von bis zu n partiellen Ableitungen beliebig vertauschen.* $\square$

Insbesondere können wir partielle Ableitungen sortieren, z.B.:

$$\partial_1\partial_4\partial_1\partial_2\partial_4\partial_1 f = \partial_1\partial_1\partial_1\partial_2\partial_4\partial_4 f = \partial_1^3\partial_2\partial_4^2 f \,.$$

Wir verwenden dabei die Schreibweise

$$\partial_i^k f = \underbrace{\partial_i\partial_i\ldots\partial_i}_{k \text{ mal}} f \,.$$

Definition 7.7 (Gradient)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}$ partiell differenzierbar. Dann heißt

$$\operatorname{grad} f(x) = \nabla f(x) = (\partial_1 f(x), \ldots, \partial_n f(x)) \tag{7.10}$$

der Gradient von f in x. $\square$

Für radialsymmetrische Funktionen $f(x) = g(\|x\|_2)$ haben wir in (7.4) berechnet

$$\partial_i f(x) = g'(\|x\|_2)\frac{x_i}{\|x\|_2}\,,$$

also

$$\operatorname{grad} f(x) = \nabla f(x) = g'(\|x\|_2)\frac{x}{\|x\|_2}\,. \tag{7.11}$$

Für $f, g : \Omega \to \mathbb{R}$ und $\lambda \in \mathbb{R}$ gelten die Rechenregeln

$$\operatorname{grad}(f+g) = \operatorname{grad} f + \operatorname{grad} g\,, \quad \operatorname{grad}(\lambda f) = \lambda \operatorname{grad} f\,, \tag{7.12}$$

$$\operatorname{grad}(f \cdot g) = g \cdot \operatorname{grad} f + f \cdot \operatorname{grad} g\,. \tag{7.13}$$

(Folgt aus der Definition und den Rechenregeln aus Analysis I.)

Definition 7.8 (Vektorfeld)
Sei $\Omega \subset \mathbb{R}^n$. Ein $F : \Omega \to \mathbb{R}^n$ heißt Vektorfeld auf Ω. F heißt (stetig) partiell differenzierbar, falls alle Komponentenfelder $F_i : \Omega \to \mathbb{R}$ es sind. $\square$

Definition 7.9 (Rotation)
Sei $\Omega \subset \mathbb{R}^3$ offen, $F : \Omega \to \mathbb{R}^3$ partiell differenzierbar. Dann heißt

$$\operatorname{rot} F : \Omega \to \mathbb{R}^3\,, \tag{7.14}$$

$$\operatorname{rot} F(x) = (\partial_2 F_3(x) - \partial_3 F_2(x), \partial_3 F_1(x) - \partial_1 F_3(x), \partial_1 F_2(x) - \partial_2 F_1(x)) \tag{7.15}$$

die Rotation von F. $\square$

Definition 7.10 (Gradientenfeld)
Ein Vektorfeld $F : \Omega \to \mathbb{R}^n$, $\Omega \subset \mathbb{R}^n$, heißt Gradientenfeld, falls es ein skalares Feld $f : \Omega \to \mathbb{R}$ gibt mit

$$F = \operatorname{grad} f\,. \tag{7.16}$$

f heißt ein Potential von F. $\square$

Definition 7.11 (Divergenz)
Sei $\Omega \subset \mathbb{R}^n$, $F : \Omega \to \mathbb{R}^n$ ein partiell differenzierbares Vektorfeld. Dann heißt

$$\operatorname{div} F : \Omega \to \mathbb{R}\,, \tag{7.17}$$

$$\operatorname{div} F(x) = \sum_{i=1}^{n} \partial_i F_i(x)\,, \tag{7.18}$$

die Divergenz von F. □

Beispiel 7.12

(i) Sei $\Omega \subset \mathbb{R}^n$ offen, seien $f : \Omega \to \mathbb{R}$, $F : \Omega \to \mathbb{R}^n$ partiell differenzierbar. Dann gilt

$$\begin{aligned} \operatorname{div}(fF) &= \sum_{i=1}^{n} \partial_i (fF_i) = \sum_{i=1}^{n} (\partial_i f \cdot F_i + f \cdot \partial_i F_i) \\ &= \langle \operatorname{grad} f, F\rangle + f \cdot \operatorname{div} F\,. \end{aligned} \tag{7.19}$$

(ii) Wir betrachten $G : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}^n$,

$$G(x) = \frac{x}{\|x\|_2}\,. \tag{7.20}$$

Wir setzen

$$f(x) = \frac{1}{\|x\|_2} = \left(\sum_{i=1}^{n} x_i^2\right)^{-\frac{1}{2}}\,, \quad F(x) = x\,,$$

Dann ist

$$\partial_i f(x) = -\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2\right)^{-\frac{3}{2}} \cdot 2x_i = -\frac{x_i}{\|x\|_2^3}\,,$$

$$\operatorname{div} F(x) = n\,,$$

also erhalten wir aus (7.19)

$$\begin{aligned} (\operatorname{div} G)(x) &= \langle (\operatorname{grad} f)(x), F(x)\rangle + f(x) \cdot (\operatorname{div} F)(x) \\ &= \left\langle -\frac{x}{\|x\|_2^3}, x \right\rangle + \frac{1}{\|x\|_2} n = -\frac{\|x\|_2^2}{\|x\|_2^3} + \frac{1}{\|x\|_2} n \\ &= \frac{n-1}{\|x\|_2}\,. \end{aligned}$$

□

Definition 7.13 (Laplace-Operator)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^2(\Omega)$. Wir definieren

$$\begin{aligned} \Delta f(x) &= (\operatorname{div}(\operatorname{grad} f))(x) = \sum_{i=1}^{n} \partial_i (\operatorname{grad} f)_i(x) \\ &= \sum_{i=1}^{n} \partial_i^2 f(x)\,. \end{aligned} \tag{7.21}$$

Der Operator $\Delta = \sum_{i=1}^{n} \partial_i^2$ heißt der Laplace-Operator. □

Die Gleichung

$$-\Delta f(x) = 0\,, \quad x \in \Omega \subset \mathbb{R}^n\,, \tag{7.22}$$

heißt Laplace-Gleichung oder Potentialgleichung. Für $\Omega \subset \mathbb{R}^n$, $I \subset \mathbb{R}$, $f : \Omega \times I \to \mathbb{R}$, heißt

$$\partial_t^2 f(x,t) - c^2 \Delta f(x,t) = 0\,, \tag{7.23}$$

die Wellengleichung (∂_t partielle Ableitung nach t, Δ bezogen auf die ersten n Komponenten x von (x,t)). Die Zahl $c > 0$ beschreibt die Ausbreitungsgeschwindigkeit der Welle. Die Gleichung

$$\partial_t f(x,t) - k\Delta f(x,t) = 0\,, \tag{7.24}$$

k Wärmeleitkoeffizient, heißt die Wärmeleitungsgleichung oder die Diffusionsgleichung.

Wir wenden den Laplace-Operator auf eine radialsymmetrische Funktion an. Sei $f(x) = g(\|x\|_2)$. Dann gilt für $x \neq 0$

$$\begin{aligned}
\Delta f(x) &= (\operatorname{div}(\operatorname{grad} f))(x) = \operatorname{div}\left(\frac{x}{\|x\|_2} g'(\|x\|_2)\right) \\
&= \left\langle \operatorname{grad}(g'(\|x\|_2)), \frac{x}{\|x\|_2} \right\rangle + g'(\|x\|_2)\operatorname{div}\left(\frac{x}{\|x\|_2}\right) \\
&= \left\langle g''(\|x\|_2)\frac{x}{\|x\|_2}, \frac{x}{\|x\|_2} \right\rangle + g'(\|x\|_2)\frac{n-1}{\|x\|_2} \\
&= g''(\|x\|_2) + \frac{n-1}{\|x\|_2} g'(\|x\|_2)\,.
\end{aligned}$$

Im Spezialfall $n = 2$, $g(r) = \ln r$, gilt

$$g''(r) = -\frac{1}{r^2} = -\frac{1}{r} g'(r)\,,$$

also für $n = 2$

$$\Delta \ln(\|x\|_2) = 0\,, \quad x \neq 0\,.$$

Im Fall $n \geq 3$, $g(r) = r^{2-n}$, gilt

$$g'(r) = (2-n)r^{1-n}\,, \quad g''(r) = (1-n)(2-n)r^{-n} = \frac{1-n}{r} g'(r)\,,$$

also wieder

$$\Delta(\|x\|_2^{2-n}) = 0\,, \quad x \neq 0\,.$$

8 Differenzierbarkeit im Mehrdimensionalen

Ist $f : \mathbb{R}^n \to \mathbb{R}^m$ eine Funktion mit Komponentenfunktionen $f_i : \mathbb{R}^n \to \mathbb{R}$, so können wir die partiellen Ableitungen $\partial_j f_i : \mathbb{R}^n \to \mathbb{R}$ bilden und damit rechnen. Was uns aber noch fehlt, ist ein allgemeiner Begriff der Differenzierbarkeit von f. Die Formel

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

und deren Interpretation (Tangentensteigung = Grenzwert der Sekantensteigungen) sind hierfür nicht geeignet. Was sich aber verallgemeinern läßt, ist die Vorstellung, f "lokal durch eine lineare Abbildung zu approximieren", also die Darstellung

$$f(x+h) = f(x) + f'(x)h + r(h)\,, \quad \lim_{\substack{h\to 0\\ h\neq 0}} \frac{r(h)}{h} = 0\,.$$

Definition 8.1 (Differenzierbarkeit)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}^m$, $x \in \Omega$. f heißt differenzierbar in x, falls eine lineare Abbildung $T : \mathbb{R}^n \to \mathbb{R}^m$ existiert mit

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{\|f(x+h)-f(x)-T(h)\|}{\|h\|} = 0\,. \tag{8.1}$$

Statt "differenzierbar" sagt man auch "Fréchet-differenzierbar" oder "total differenzierbar". Die Abbildung T heißt Ableitung (oder Fréchet-Ableitung oder totale Ableitung) von f in x. Wir bezeichnen die Ableitung von f in x mit $Df(x)$. $\square$

In Satz 8.3 werden wir mitbeweisen, daß es höchstens eine solche Abbildung T gibt (dadurch wird die Bezeichnung $Df(x)$ gerechtfertigt).

Wegen der Äquivalenz aller Normen im $\mathbb{R}^n$ ist es gleichgültig, welche Norm man in Definition 8.1 zugrundelegt.

Verallgemeinert man Definition 8.1 auf Abbildungen $f : X \to Y$ zwischen beliebigen normierten Räumen X, Y, so verlangt man zusätzlich, daß T stetig ist (im Endlichdimensionalen ist jede lineare Abbildung stetig).

Beispiel 8.2

(i) Sei $f : \mathbb{R}^n \to \mathbb{R}^m$ linear, $x \in \mathbb{R}^n$. Dann ist

$$f(x+h) - f(x) - f(h) = f(x) + f(h) - f(x) - f(h) = 0\,,$$

also ist f differenzierbar in x, und $Df(x) = f$ (Gleichheit im Sinn von Abbildungen, wie f ist auch $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$, und $(Df(x))(h) = f(h)$ für alle $h \in \mathbb{R}^n$). Die Ableitung einer linearen Abbildung hängt also nicht von x ab,

$$Df : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$$

ist eine konstante Abbildung (hier bezeichnet $L(\mathbb{R}^n, \mathbb{R}^m)$ die Menge aller linearen Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$).

(ii) Sei $f : \mathbb{R}^n \to \mathbb{R}^m$ definiert durch

$$f(x) = Ax + b\,, \quad A \in \mathbb{R}^{(m,n)}\,,\, b \in \mathbb{R}^m\,,$$

das heißt

$$\begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Es ist

$$f(x+h) - f(x) - Ah = A(x+h) + b - Ax - b - Ah = 0\,,$$

also ist f differenzierbar in x, und $Df(x)(h) = Ah$.

(iii) Sei $A \in \mathbb{R}^{(n,n)}$ symmetrisch und $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch

$$f(x) = \frac{1}{2}\langle x, Ax\rangle = \frac{1}{2}x^T Ax = \frac{1}{2}\sum_{i,j=1}^{n} x_i a_{ij} x_j\,.$$

Für festes $x \in \mathbb{R}^n$ gilt

$$f(x+h) - f(x) = \frac{1}{2}\langle x+h, A(x+h)\rangle - \frac{1}{2}\langle x, Ax\rangle \tag{8.2}$$

$$= \frac{1}{2}\langle h, Ax\rangle + \frac{1}{2}\langle x, Ah\rangle + \frac{1}{2}\langle h, Ah\rangle \tag{8.3}$$

$$= \langle Ax, h\rangle + \frac{1}{2}\langle h, Ah\rangle\,. \tag{8.4}$$

Mit $T(h) = \langle Ax, h\rangle$ gilt

$$\|f(x+h) - f(x) - T(h)\|_2 = \left\|\frac{1}{2}\langle h, Ah\rangle\right\|_2 \leq \frac{1}{2}\|h\|_2\|Ah\|_2$$

$$\leq \frac{1}{2}\|h\|_2 C\|h\|_2\,, \tag{8.5}$$

wobei C eine von A, aber nicht von h abhängende Konstante ist (A ist stetig !), also

$$0 \leq \frac{\|f(x+h) - f(x) - T(h)\|_2}{\|h\|_2} \leq \frac{1}{2}C\|h\|_2 \to 0$$

für $h \to 0$, also ist f differenzierbar in x und

$$(Df(x))(h) = T(h) = \langle Ax, h\rangle\,.$$

□

Satz 8.3 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \mathbb{R}^n \to \mathbb{R}^m$, f in $x \in \Omega$ differenzierbar. Dann ist f stetig in x und partiell differenzierbar in x, und*

$$(Df(x))(h) = (J_f(x))h\,, \quad J_f(x) = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix}. \tag{8.6}$$

Die Matrix $J_f(x) \in \mathbb{R}^{(m,n)}$ heißt Jacobi-Matrix, oder auch Funktionalmatrix, von f in x.

Beweis: Sei $T : \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung, welche (8.1) erfüllt.
f ist stetig in x: Mit $r(h) = f(x+h) - f(x) - T(h)$ gilt

$$\|f(x+h) - f(x)\| = \|r(h) + T(h)\| \le \|r(h)\| + \|T(h)\| \le \|r(h)\| + C\|h\|\,,$$

wobei C eine von h unabhängige Konstante ist, also

$$\lim_{h \to 0} \|f(x+h) - f(x)\| = 0\,.$$

f ist partiell differenzierbar in x: Sei $A \in \mathbb{R}^{(m,n)}$ die Matrix, welche die lineare Abbildung T bezüglich der kanonischen Basis darstellt, also

$$T(h) = Ah\,, \quad h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

Wir definieren $r : \mathbb{R}^n \to \mathbb{R}^m$ durch

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

also

$$r_i(h) = f_i(x+h) - f_i(x) - \sum_{k=1}^n a_{ik}h_k\,, \quad 1 \le i \le m\,.$$

Für $h = te_j$, $t \in \mathbb{R}$, $t \neq 0$, e_j der j-te Einheitsvektor, gilt

$$r_i(te_j) = f_i(x+te_j) - f_i(x) - ta_{ij}\,,$$

also

$$\frac{f_i(x+te_j) - f_i(x)}{t} = a_{ij} + \frac{r_i(te_j)}{t}\,,$$

und

$$\left|\frac{r_i(te_j)}{t}\right| = \frac{|r_i(te_j)|}{\|te_j\|_2} \le \frac{\|r(te_j)\|_2}{\|te_j\|_2} \to 0$$

für $t \to 0$, da $\|r(h)\|_2/\|h\|_2 \to 0$ für $h \to 0$. Also

$$\partial_j f_i(x) = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f_i(x+te_j) - f_i(x)}{t} = a_{ij}\,.$$

□

Der soeben gegebene Beweis zeigt außerdem, daß die lineare Abbildung T in Definition 8.1 eindeutig bestimmt ist, also ist die Definition von $Df(x)$ sinnvoll.

Definition 8.4 (Stetige Differenzierbarkeit)
Sei $\Omega \subset \mathbb{R}^n$ offen. Eine Abbildung $f : \Omega \to \mathbb{R}^m$ heißt stetig differenzierbar in Ω, falls alle partiellen Ableitungen $\partial_j f_i : \Omega \to \mathbb{R}$ existieren und in Ω stetig sind.

Satz 8.5 *Sei $\Omega \subset \mathbb{R}^n$ offen, sei $f : \Omega \to \mathbb{R}^m$ stetig differenzierbar in Ω. Dann ist f differenzierbar in allen Punkten x von Ω.*

Beweis: Wir setzen

$$r(h) = f(x+h) - f(x) - J_f(x)h\,,$$

also

$$r_i(h) = f_i(x+h) - f_i(x) - \sum_{j=1}^{n} \partial_j f_i(x) h_j\,, \quad 1 \le i \le m\,.$$

Es genügt zu zeigen, daß für alle i gilt

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{|r_i(h)|}{\|h\|_\infty} = 0\,. \tag{8.7}$$

Sei $x \in \Omega$. Wähle zunächst $\delta > 0$ so, dass $x + h \in \Omega$ gilt für alle h mit $\|h\|_\infty \le \delta$ (Ω ist offen). Sei nun $h \in \mathbb{R}^n$ mit $\|h\|_\infty \le \delta$. Wir definieren

$$z^j = x + \sum_{k=1}^{j} h_k e_k\,, \quad z^0 = x\,.$$

Wir wenden den Mittelwertsatz im $\mathbb{R}$ an auf

$$g_{ij}(t) = f_i(z^{j-1} + te_j)\,.$$

Es gibt also τ_{ij} zwischen 0 und h_j mit

$$g_{ij}(h_j) = g_{ij}(0) + g_{ij}'(\tau_{ij})h_j\,,$$

also

$$f_i(z^j) = f_i(z^{j-1}) + \partial_j f_i(\eta_{ij})h_j\,, \quad \eta_{ij} = z^{j-1} + \tau_{ij}e_j\,.$$

Weiter folgt

$$f_i(x+h) - f_i(x) = \sum_{j=1}^{n} (f_i(z^j) - f_i(z^{j-1})) = \sum_{j=1}^{n} \partial_j f_i(\eta_{ij})h_j\,,$$

also

$$r_i(h) = \sum_{j=1}^{n} \partial_j f_i(\eta_{ij})h_j - \sum_{j=1}^{n} \partial_j f_i(x)h_j\,,$$

also

$$|r_i(h)| \le \|h\|_\infty \sum_{j=1}^{n} |\partial_j f_i(\eta_{ij}) - \partial_j f_i(x)|\,.$$

Es ist $\|\eta_{ij} - x\|_\infty \le \|h\|_\infty$, also folgt aus der Stetigkeit der partiellen Ableitungen

$$\lim_{h\to 0} \sum_{j=1}^{n} |\partial_j f_i(\eta_{ij}) - \partial_j f_i(x)| = 0\,.$$

und damit die Behauptung (8.7). □

In den beiden vorangehenden Sätzen haben wir die Implikationen bewiesen:

$$\begin{array}{lcl}
f \text{ stetig differenzierbar in } \Omega & \Rightarrow & f \text{ differenzierbar in } \Omega \\
f \text{ differenzierbar in } \Omega & \Rightarrow & f \text{ partiell differenzierbar in } \Omega \\
f \text{ differenzierbar in } \Omega & \Rightarrow & f \text{ stetig in } \Omega
\end{array}$$

Alle anderen (außer die sich durch Transitivität ergebenden) möglichen Implikationen zwischen diesen 4 Begriffen gelten nicht !

Beispiel 8.6 (Polarkoordinaten in der Ebene)
Die Abbildung

$$f : \mathbb{R}^2 \to \mathbb{R}^2 \,, \quad f(r, \varphi) = (r \cos \varphi, r \sin \varphi) \tag{8.8}$$

repräsentiert die Transformation von Polarkoordinaten auf kartesische Koordinaten. Es gilt

$$J_f(r, \varphi) = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix} . \tag{8.9}$$

Alle $\partial_j f_i$ existieren und sind stetig, also ist f differenzierbar.

Satz 8.7 (Kettenregel im Mehrdimensionalen)
Seien $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ offen, seien $f : U \to \mathbb{R}^m$, $g : V \to \mathbb{R}^l$ mit $f(U) \subset V$. Ist f differenzierbar in $x \in U$ und g differenzierbar in $f(x)$, so ist auch $g \circ f$ differenzierbar in x, und es gilt

$$D(g \circ f)(x) = Dg(f(x)) \circ Df(x) \,, \quad J_{g \circ f}(x) = J_g(f(x)) \cdot J_f(x) \,. \tag{8.10}$$

Beweis: Wir betrachten die Restglieder

$$r_f(h) = f(x+h) - f(x) - Df(x)(h) \,, \quad r_g(\eta) = g(f(x) + \eta) - g(f(x)) - Dg(f(x))(\eta) \,.$$

Es gilt

$$\lim_{\substack{h \to 0 \\ h \neq 0}} \frac{\|r_f(h)\|}{\|h\|} = 0 \,, \quad \lim_{\substack{\eta \to 0 \\ \eta \neq 0}} \frac{\|r_g(\eta)\|}{\|\eta\|} = 0 \,.$$

Wir definieren nun

$$\eta(h) = f(x+h) - f(x) = Df(x)(h) + r_f(h) \,.$$

Dann gilt

$$\begin{aligned}
(g \circ f)(x+h) - (g \circ f)(x) &= g(f(x) + \eta(h)) - g(f(x)) \\
&= Dg(f(x))(Df(x))(h) + r_f(h)) + r_g(\eta(h)) \,,
\end{aligned}$$

also

$$(g \circ f)(x+h) - (g \circ f)(x) - (Dg(f(x)) \circ Df(x))(h) = r_g(\eta(h)) + Dg(f(x))(r_f(h)) \,. \tag{8.11}$$

Wir wollen zeigen, daß die rechte Seite von (8.11) schneller gegen 0 geht als $\|h\|$. Zunächst gilt (C_1 von h unabhängig)

$$\frac{\|Dg(f(x))(r_f(h))\|}{\|h\|} \leq \frac{C_1 \|r_f(h)\|}{\|h\|} \to 0$$

für $h \to 0$. Falls $\eta(h) = 0$, so ist auch $r_g(\eta(h)) = 0$. Falls $\eta(h) \neq 0$, so gilt

$$\frac{\|r_g(\eta(h))\|}{\|h\|} = \frac{\|r_g(\eta(h))\|}{\|\eta(h)\|} \cdot \frac{\|\eta(h)\|}{\|h\|}\,. \tag{8.12}$$

Der zweite Bruch auf der rechten Seite ist wegen

$$\|\eta(h)\| \leq \|Df(x)(h)\| + \|r_f(h)\| \leq C_2\|h\| + \|r_f(h)\|$$

beschränkt, der erste Bruch auf der rechten Seite von (8.12) konvergiert gegen 0 für $h \to 0$, da $\eta(h) \to 0$ für $h \to 0$. Insgesamt folgt also aus (8.11)

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{\|(g\circ f)(x+h) - (g\circ f)(x) - (Dg(f(x))\circ Df(x))(h)\|}{\|h\|} = 0\,, \tag{8.13}$$

und damit die erste Gleichung in (8.10). Die zweite ergibt sich, da beim Übergang von linearen Abbildungen zu Matrizen die Komposition in die Matrixmultiplikation übergeht. □

Schreibt man die Gleichung

$$J_{g\circ f}(x) = J_g(f(x)) \cdot J_f(x)$$

elementweise auf, so ergibt sich

$$\partial_j (g\circ f)_i(x) = \sum_{k=1}^{m} (\partial_k g_i)(f(x)) \cdot \partial_j f_k(x)\,. \tag{8.14}$$

Im Spezialfall $n = l = 1$, also $f : \mathbb{R} \to \mathbb{R}^m$ Kurve, $g : \mathbb{R}^m \to \mathbb{R}$ skalares Feld, ist $g \circ f : \mathbb{R} \to \mathbb{R}$ und

$$(g\circ f)'(t) = \sum_{k=1}^{m} (\partial_k g)(f(t)) f_k'(t) = \langle (\operatorname{grad} g)(f(t)), f'(t)\rangle\ . \tag{8.15}$$

Verläuft die Spur der Kurve f ganz in einer Niveaufläche $N_c(g) = \{x : g(x) = c\}$ von g, dann ist

$$g(f(t)) = c$$

für alle t, also folgt aus (8.15)

$$0 = (g\circ f)'(t) = \langle (\operatorname{grad} g)(f(t)), f'(t)\rangle\ ,$$

das heißt, der Gradient von g in $f(t)$ steht senkrecht auf den Tangentialvektor $f'(t)$ von f in t. Da dies für jede Kurve gilt, deren Spur ganz in $N_c(g)$ verläuft, steht also der Gradient senkrecht auf der von allen möglichen Tangentialvektoren gebildeten tangentialen Hyperebene der Niveaufläche.

Definition 8.8 (Tangentialvektor an eine Niveaufläche)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}$ differenzierbar, $x \in \Omega$ mit $f(x) = c$ und $\operatorname{grad} f(x) \neq 0$. *Ein Vektor $v \in \mathbb{R}^n$ heißt Tangentialvektor an $N_c(f)$ in x, falls*

$$\langle (\operatorname{grad} f)(x), v\rangle = 0\,. \tag{8.16}$$

□

Satz 8.9 *In der Situation von Definition 8.8 gilt:*

$$T(x) = \{v : v \text{ ist Tangentialvektor an } N_c(f) \text{ in } x\} \tag{8.17}$$

ist ein Unterraum des $\mathbb{R}^n$ *der Dimension* $n-1$*.* $T(x)$ *heißt der Tangentialraum von* $N_c(f)$ *in* x*. Der affine Unterraum* $x + T(x)$ *heißt Tangentialebene an* $N_c(f)$ *in* x*.*

Beweis: Siehe Lineare Algebra. □

Als Beispiel für Satz 8.9 betrachten wir $f : \mathbb{R}^3 \to \mathbb{R}$,

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 .$$

$N_1(f)$ ist der Rand der Einheitskugel im $\mathbb{R}^3$. Für $x \in N_1(f)$ gilt

$$(\operatorname{grad} f)(x) = 2x , \quad T(x) = \{v : v \in \mathbb{R}^3, \langle v, x\rangle = 0\} ,$$

das heißt, $T(x)$ ist der auf x senkrecht stehende zweidimensionale Unterraum.

Satz 8.10 *Sei* $\Omega \subset \mathbb{R}^n$ *offen,* $f : \Omega \to \mathbb{R}$ *differenzierbar,* $x \in \Omega$*. Dann existiert für jedes* $v \in \mathbb{R}^n$ *die Richtungsableitung* $\partial_v f(x)$*, und es gilt*

$$\partial_v f(x) = \langle (\operatorname{grad} f)(x), v\rangle . \tag{8.18}$$

Beweis: Wir betrachten $h : (-\delta, \delta) \to \mathbb{R}^n$, $h(t) = x + tv$. Für hinreichend kleines δ ist $h((-\delta, \delta)) \subset \Omega$, also existiert nach der Kettenregel $(f \circ h)'(0)$,

$$(f \circ h)'(0) = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(x + tv) - f(x)}{t} = \partial_v f(x) ,$$

und aus Formel (8.15) folgt

$$\partial_v f(x) = \langle \operatorname{grad} f(x), h'(0)\rangle = \langle \operatorname{grad} f(x), v\rangle .$$

□

Wir können die Formel (8.18) interpretieren: Ist $v \in \mathbb{R}^n$ mit $\|v\|_2 = 1$, so ist

$$\partial_v f(x) = \langle (\operatorname{grad} f)(x), v\rangle = \|(\operatorname{grad} f)(x)\|_2 \cdot \cos\varphi ,$$

wobei φ der Winkel zwischen $\operatorname{grad} f(x)$ und v ist. Die Richtungsableitung $\partial_v f(x)$ wird also maximal, wenn $\cos\varphi = 1$ ist, also wenn v in dieselbe Richtung wie $\operatorname{grad} f(x)$ zeigt. Der Gradient von f zeigt also in die Richtung des steilsten Anstiegs von f.

Definition 8.11 (Verbindungsstrecke)
Ist X *Vektorraum, so bezeichnen wir die Verbindungsstrecke zweier Punkte* x *und* y *in* X *mit*

$$[x, y] = \{tx + (1 - t)y : t \in [0, 1]\} . \tag{8.19}$$

□

Satz 8.12 (Mittelwertsatz für skalare Felder)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}$ differenzierbar, seien $x, y \in \Omega$ mit $[x, y] \subset \Omega$. Dann gibt es ein $\xi \in [x, y]$ mit

$$f(y) - f(x) = \langle \operatorname{grad} f(\xi), y - x \rangle \ . \tag{8.20}$$

Beweis: Wir definieren $g : [0, 1] \to \mathbb{R}$ durch $g(t) = f(tx + (1 - t)y)$, dann ist g stetig in $[0, 1]$ und differenzierbar in $(0, 1)$. Aus dem Mittelwertsatz im $\mathbb{R}$ folgt, daß ein $t \in (0, 1)$ existiert mit

$$g(1) - g(0) = g'(t) \, ,$$

also

$$f(y) - f(x) = \langle \operatorname{grad} f(tx + (1 - t)y), y - x \rangle \ .$$

$\square$

Satz 8.13 (Mittelwertsatz im Mehrdimensionalen)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}^m$ stetig differenzierbar, seien $x, y \in \Omega$ mit $[x, y] \subset \Omega$. Dann gilt

$$\|f(y) - f(x)\|_\infty \leq L\|y - x\|_1 \leq nL\|y - x\|_\infty \, , \tag{8.21}$$

wobei

$$L = \max\{|\partial_j f_i(\xi)| : 1 \leq i \leq m, \, 1 \leq j \leq n, \, \xi \in [x, y]\} \, . \tag{8.22}$$

Beweis: Das Maximum in (8.22) existiert, da alle partiellen Ableitungen stetig sind und $[x, y]$ kompakt ist. Für alle i gibt es nach Satz 8.12 ein $\xi_i \in [x, y]$ mit

$$f_i(y) - f_i(x) = \langle \operatorname{grad} f_i(\xi_i), y - x \rangle = \sum_{j=1}^{n} \partial_j f_i(\xi_i)(y_j - x_j) \, ,$$

also gilt für alle i

$$|f_i(y) - f_i(x)| \leq L \sum_{j=1}^{n} |y_j - x_j| \leq nL \max_{1 \leq j \leq n} |y_j - x_j| \, .$$

$\square$

9 Taylorentwicklung im Mehrdimensionalen

Zur Erinnerung: Für $f : \mathbb{R} \to \mathbb{R}$, $f \in C^{k+1}(\mathbb{R})$ kennen wir die Taylorentwicklung

$$f(x+h) = \sum_{\alpha=0}^{k} \frac{f^{(\alpha)}(x)}{\alpha!} h^\alpha + \frac{f^{(k+1)}(\xi)}{(k+1)!} h^{k+1} , \tag{9.1}$$

wobei ξ zwischen x und $x+h$ liegt. Sei nun $f : \mathbb{R}^n \to \mathbb{R}$. Um $f(x+h)$ durch Ableitungen von f in x zu approximieren, betrachten wir

$$g : \mathbb{R} \to \mathbb{R} , \quad g(t) = f(x+th) .$$

Ist g hinreichend oft differenzierbar, so gilt gemäß (9.1), angewendet auf g,

$$f(x+h) = g(1) = \sum_{\alpha=0}^{k} \frac{g^{(\alpha)}(0)}{\alpha!} + \frac{g^{(k+1)}(\tau)}{(k+1)!} .$$

Die Ableitungen von g lassen sich durch Ableitungen von f ausdrücken,

$$g'(t) = \langle \operatorname{grad} f(x+th), h \rangle = \sum_{i=1}^{n} \partial_i f(x+th) h_i ,$$

$$g''(t) = \sum_{i=1}^{n} \langle (\operatorname{grad} (\partial_i f))(x+th), h \rangle \, h_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \partial_j \partial_i f(x+th) h_j \right) h_i ,$$

entsprechend die höheren Ableitungen.

Notation 9.1 (Multiindex) *Ein $\alpha = (\alpha_1, \dots, \alpha_n)$ mit $\alpha_i \in \mathbb{N} \cup \{0\}$ heißt Multiindex. Für einen Multiindex $\alpha = (\alpha_1, \dots, \alpha_n)$ definieren wir*

$$|\alpha| = \sum_{i=1}^{n} \alpha_i , \quad \alpha! = \prod_{i=1}^{n} \alpha_i! . \tag{9.2}$$

Für $f \in C^{|\alpha|}(\Omega)$, $\Omega \subset \mathbb{R}^n$ offen, setzen wir

$$\partial^\alpha f = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_n^{\alpha_n} f .$$

Für $x \in \mathbb{R}^n$ definieren wir

$$x^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i} .$$

Satz 9.2 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^k(\Omega)$, $x \in \Omega$, $h \in \mathbb{R}^n$ mit $[x, x+h] \subset \Omega$. Dann wird durch*

$$g(t) = f(x+th)$$

eine Funktion $g \in C^k([0,1])$ definiert, und es gilt

$$g^{(k)}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \partial^\alpha f(x+th) h^\alpha . \tag{9.3}$$

Beweis: Mit vollständiger Induktion folgt aus der Kettenregel die Existenz von $g^{(k)}$ und die Formel

$$g^{(k)}(t) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} \partial_{i_k} \cdots \partial_{i_2} \partial_{i_1} f(x+th) h_{i_k} \cdots h_{i_2} h_{i_1} . \tag{9.4}$$

Jede der auftretenden partiellen Ableitungen $\partial_{i_k} \cdots \partial_{i_2} \partial_{i_1} f$ entspricht nach Umsortieren und Zusammenfassen einer Ableitung der Form $\partial^\alpha f$ mit einem Multiindex α mit $|\alpha| = k$. Jeder solche Multiindex kommt in der k-fachen Summe (9.4) mit der Anzahl

$$\binom{k}{\alpha_1} \binom{k-\alpha_1}{\alpha_2} \cdots \binom{k-\alpha_1-\ldots-\alpha_{n-1}}{\alpha_n} = \frac{k!}{\alpha_1! \alpha_2! \ldots \alpha_n!} = \frac{k!}{\alpha!}$$

vor, woraus die Behauptung folgt. □

Satz 9.3 (Taylorformel für Skalarfelder)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^{k+1}(\Omega)$, $x \in \Omega$, $h \in \mathbb{R}^n$ mit $[x, x+h] \subset \Omega$. Dann gibt es ein $\xi \in [x, x+h]$ mit

$$f(x+h) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(\xi)}{\alpha!} h^\alpha . \tag{9.5}$$

(Summiert wird über die Multiindizes.)

Beweis: Für $g(t) = f(x+th)$ gilt mit geeignetem $\tau \in (0,1)$

$$\begin{aligned} f(x+h) = g(1) &= \sum_{j=0}^{k} \frac{g^{(j)}(0)}{j!} + \frac{g^{(k+1)}(\tau)}{(k+1)!} \\ &= \sum_{j=0}^{k} \frac{1}{j!} \left(\sum_{|\alpha|=j} \frac{j!}{\alpha!} \partial^\alpha f(x) h^\alpha \right) + \frac{1}{(k+1)!} \sum_{|\alpha|=k+1} \frac{(k+1)!}{\alpha!} \partial^\alpha f(x+\tau h) h^\alpha . \end{aligned} \tag{9.6}$$

□

Folgerung 9.4 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^k(\Omega)$, $x \in \Omega$ mit $K_\delta(x) \subset \Omega$. Dann gilt für alle $h \in \mathbb{R}^n$ mit $\|h\| \le \delta$*

$$f(x+h) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + R_{k+1}(h) , \quad \lim_{\substack{h \to 0 \\ h \ne 0}} \frac{R_{k+1}(h)}{\|h\|_\infty^k} = 0 . \tag{9.7}$$

Beweis: Es gilt nach Satz 9.3 für geeignetes $\xi \in [x, x+h]$

$$\begin{aligned} f(x+h) &= \sum_{|\alpha| \le k-1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=k} \frac{\partial^\alpha f(\xi)}{\alpha!} h^\alpha \\ &= \sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \underbrace{\sum_{|\alpha|=k} \frac{\partial^\alpha f(\xi) - \partial^\alpha f(x)}{\alpha!} h^\alpha}_{=:R_{k+1}(h)} , \end{aligned} \tag{9.8}$$

und

$$|R_{k+1}(h)| \le \sum_{|\alpha|=k} \frac{|\partial^\alpha f(\xi) - \partial^\alpha f(x)|}{\alpha!} |h^\alpha| \le \|h\|_\infty^k \underbrace{\sum_{|\alpha|=k} \frac{|\partial^\alpha f(\xi) - \partial^\alpha f(x)|}{\alpha!}}_{\to 0 \text{ für } h \to 0}, \tag{9.9}$$

da $\|\xi - x\|_\infty \le \|h\|_\infty$. □

Für $k = 2$ erhalten wir

$$f(x+h) = f(x) + \sum_{j=1}^n \partial_j f(x) h_j + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(x) h_i h_j + R_3(h) . \tag{9.10}$$

Der quadratische Term läßt sich schreiben als

$$\sum_{i,j=1}^n \partial_i \partial_j f(x) h_i h_j = \langle h, H_f(x) h \rangle = h^T H_f(x) h , \tag{9.11}$$

wobei die Matrix $H_f(x) \in \mathbb{R}^{(n,n)}$ gegeben ist durch

$$H_f(x) = \begin{pmatrix} \partial_1^2 f(x) & \partial_2 \partial_1 f(x) & \cdots & \partial_n \partial_1 f(x) \\ \partial_1 \partial_2 f(x) & & & \vdots \\ \vdots & & & \vdots \\ \partial_1 \partial_n f(x) & \cdots & \cdots & \partial_n^2 f(x) \end{pmatrix} . \tag{9.12}$$

Definition 9.5 (Hesse-Matrix)
Die Matrix $H_f(x)$ in (9.12) heißt Hesse-Matrix von f in x. □

Die Hesse-Matrix ist symmetrisch nach Satz 7.5, wenn $f \in C^2(\Omega)$.

Definition 9.6 (Lokales Maximum, Minimum, Extremum)
Sei $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$. Ein $x \in \Omega$ heißt lokales Minimum (Maximum) von f, falls es eine Umgebung U von x gibt mit

$$f(x) \le f(y) \qquad (\text{bzw. } f(x) \ge f(y)) \qquad \text{für alle } y \in U \cap \Omega.$$

Ein $x \in \Omega$ heißt lokales Extremum, falls x lokales Minimum oder lokales Maximum ist. Ein lokales Extremum heißt isoliert, falls außerdem

$$f(y) \ne f(x)$$

gilt für alle $y \in U \cap \Omega$ mit $y \ne x$. □

Satz 9.7 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}$ partiell differenzierbar. Ist $x \in \Omega$ lokales Extremum von f, so gilt* $\operatorname{grad} f(x) = 0$.

Beweis: Sei $h \in \mathbb{R}^n$ beliebig. Wähle $\delta > 0$ mit $[x - \delta h, x + \delta h] \subset \Omega$. Dann hat $g : (-\delta, \delta) \to \mathbb{R}$, $g(t) = f(x + th)$, ein lokales Extremum in 0. Setzen wir $h = e_j$, so folgt

$$0 = g'(0) = \partial_j f(x) .$$

□

Definition 9.8 *Sei $A \in \mathbb{R}^{(n,n)}$. A heißt*

positiv definit, falls $h^T A h > 0$ gilt für alle $h \in \mathbb{R}^n$, $h \neq 0$,
positiv semidefinit, falls $h^T A h \geq 0$ gilt für alle $h \in \mathbb{R}^n$,
negativ (semi)definit, falls $-A$ positiv (semi)definit ist,
indefinit, falls A weder positiv semidefinit noch negativ semidefinit ist.

□

Satz 9.9 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^2(\Omega)$, $x \in \Omega$. Dann gilt*

x lokales Minimum $\Rightarrow$ $H_f(x)$ positiv semidefinit,
x lokales Maximum $\Rightarrow$ $H_f(x)$ negativ semidefinit.

Ist außerdem $\operatorname{grad} f(x) = 0$*, so gilt*

$H_f(x)$ positiv definit $\Rightarrow$ x isoliertes lokales Minimum,
$H_f(x)$ negativ definit $\Rightarrow$ x isoliertes lokales Maximum.

Beweis: Sei x lokales Minimum. Sei $h \in \mathbb{R}^n$ beliebig, sei $g(t) = f(x + th)$. Dann ist $g \in C^2((-\delta, \delta))$ für hinreichend kleines $\delta > 0$, und es gilt

$$g'(t) = \langle \operatorname{grad} f(x + th), h \rangle \,,$$

$$g''(t) = \sum_{i,j=1}^{n} \partial_i \partial_j f(x + th) h_i h_j = h^T H_f(x + th) h \,.$$

Da 0 ein lokales Minimum ist für g, gilt

$$0 \leq g''(0) = h^T H_f(x) h \,.$$

Da h beliebig war, folgt die Behauptung. Ist x lokales Maximum von f, so wenden wir das eben Bewiesene auf $-f$ an.
Sei nun $\operatorname{grad} f(x) = 0$, $H_f(x)$ positiv definit. Aus (9.10) folgt, falls $\|h\|$ hinreichend klein ist,

$$f(x + h) = f(x) + \frac{1}{2} h^T H_f(x) h + R_3(h) \,, \quad \lim_{\substack{h \to 0 \\ h \neq 0}} \frac{R_3(h)}{\|h\|^2} = 0 \,.$$

Sei

$$Q(h) = \frac{1}{2} h^T H_f(x) h \,, \quad \alpha = \min_{\|h\|=1} Q(h)$$

Das Minimum existiert, da Q stetig ist und die Einheitssphäre kompakt ist. Es ist $\alpha > 0$, da $H_f(x)$ positiv definit ist. Daher gilt für alle $h \neq 0$ mit $x + h \in \Omega$

$$f(x + h) - f(x) = \|h\|^2 \left(\frac{1}{2} Q(\frac{h}{\|h\|}) + \frac{R_3(h)}{\|h\|^2} \right) \geq \|h\|^2 \left(\frac{1}{2}\alpha + \frac{R_3(h)}{\|h\|^2} \right) \,.$$

Wir wählen $\delta > 0$ so, dass

$$\frac{1}{2}\alpha > \frac{|R_3(h)|}{\|h\|^2}$$

für alle $\|h\| < \delta$, dann gilt $f(x + h) > f(x)$ für alle $\|h\| < \delta$. □

In der Linearen Algebra wird bewiesen: Ist $A \in \mathbb{R}^{(n,n)}$ symmetrisch, so gilt

A positiv definit $\Leftrightarrow$ alle Eigenwerte von A sind > 0,
A positiv semidefinit $\Leftrightarrow$ alle Eigenwerte von A sind ≥ 0,
A negativ (semi)definit $\Leftrightarrow$ alle Eigenwerte von A sind $< 0 (\leq 0)$.

Wir betrachten als Beispiel

$$f : \mathbb{R}^2 \to \mathbb{R}\,, \quad f(x_1, x_2) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2}\,, \quad a, b > 0\,.$$

Die Niveaumengen $N_c(f)$ sind die Ellipsen

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = c\,, \quad c > 0\,.$$

Der Graph von f stellt ein elliptisches Paraboloid im $\mathbb{R}^3$ dar. Es ist

$$\operatorname{grad} f(x_1, x_2) = \left(\frac{2x_1}{a^2}\,, \frac{2x_2}{b^2} \right)\,, \quad \operatorname{grad} f(0) = 0\,,$$

$$H_f(x_1, x_2) = \begin{pmatrix} \frac{2}{a^2} & 0 \\ 0 & \frac{2}{b^2} \end{pmatrix}\,.$$

Die Eigenwerte von $H_f(x)$ sind $2/a^2$ und $2/b^2$, also ist $H_f(0)$ positiv definit und damit 0 ein isoliertes lokales Minimum von f (in diesem Fall sogar ein globales Minimum, d.h. ein Minimum bezüglich des gesamten Definitionsbereichs von f). Als weiteres Beispiel betrachten wir

$$f : \mathbb{R}^2 \to \mathbb{R}\,, \quad f(x_1, x_2) = \frac{x_1^2}{a^2} - \frac{x_2^2}{b^2}\,, \quad a, b > 0\,.$$

Die Niveaumengen $N_c(f)$ sind die Hyperbeln

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = c\,, \quad c \in \mathbb{R}\,.$$

Der Graph von f stellt ein hyperbolisches Paraboloid im $\mathbb{R}^3$ dar. Es ist

$$\operatorname{grad} f(x_1, x_2) = \left(\frac{2x_1}{a^2}\,, -\frac{2x_2}{b^2} \right)\,, \quad \operatorname{grad} f(0) = 0\,,$$

$$H_f(x_1, x_2) = \begin{pmatrix} \frac{2}{a^2} & 0 \\ 0 & -\frac{2}{b^2} \end{pmatrix}\,.$$

Die Eigenwerte von $H_f(x)$ sind $2/a^2$ und $-2/b^2$, also ist $H_f(0)$ indefinit und daher 0 kein lokales Extremum.

Im semidefiniten Fall kann man an der Hessematrix nicht erkennen, ob ein Extremum vorliegt oder nicht: Für die Funktionen

$$f(x_1, x_2) = x_1^2 + x_2^4\,, \quad g(x_1, x_2) = x_1^2\,, \quad h(x_1, x_2) = x_1^2 + x_2^3$$

gilt $\operatorname{grad} f(0) = \operatorname{grad} g(0) = \operatorname{grad} h(0) = 0$,

$$H_f(0) = H_g(0) = H_h(0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\,,$$

aber f hat in 0 ein isoliertes, lokales Minimum; g hat in 0 ein nicht isoliertes, lokales Minimum; h hat kein lokales Extremum in 0.

Definition 9.10 (Sattelpunkt)
Sei $\Omega \subset \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}$ partiell differenzierbar, $x \in \Omega$. Ist $\operatorname{grad} f(x) = 0$, aber x kein lokales Extremum von f, so heißt x Sattelpunkt von f. □

Wir können Taylorentwicklungen höherer Ordnung für $f : \Omega \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ offen, komponentenweise erhalten. Nach Satz 9.3 gilt für alle i, falls $f_i \in C^{k+1}(\Omega)$,

$$f_i(x+h) = \sum_{|\alpha|\le k} \frac{\partial^\alpha f_i(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=k+1} \frac{\partial^\alpha f_i(\xi_i)}{\alpha!} h^\alpha \,. \tag{9.13}$$

Führen wir die Notation

$$\partial^\alpha f(x) = (\partial^\alpha f_1(x), \dots, \partial^\alpha f_m(x))$$

ein, so erhalten wir wie in Folgerung 9.4, falls f k-mal stetig differenzierbar ist,

$$f(x+h) = \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + R_{k+1}(h) \,, \tag{9.14}$$

wobei $R_{k+1}(h) \in \mathbb{R}^m$ und

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{\|R_{k+1}(h)\|}{\|h\|^k} = 0 \,.$$

Ohne hier in nähere Einzelheiten einzusteigen, bemerken wir, dass wir auch "totale" Ableitungen höherer Ordnung definieren können. Ist $f : \mathbb{R}^n \to \mathbb{R}^m$ in $\mathbb{R}^n$ differenzierbar, so ist $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung für jedes $x \in \mathbb{R}^n$, also

$$Df : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m) = \{\varphi | \, \varphi : \mathbb{R}^n \to \mathbb{R}^m \text{ linear}\} \,.$$

Wir können die zweite Ableitung $D^2 f$ definieren als Ableitung von Df, das heißt für jedes $x \in \mathbb{R}^n$ ist

$$D^2 f(x) : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$$

eine lineare Abbildung, welche Df im Sinne der Definition der totalen Ableitung lokal approximiert. Insgesamt ist

$$D^2 f : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \,.$$

In der Linearen Algebra ergibt es sich, daß der Raum $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ kanonisch isomorph ist zum Raum $\operatorname{Bil}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$ der bilinearen Abbildungen von $\mathbb{R}^n \times \mathbb{R}^n$ nach $\mathbb{R}^m$. Im Falle $m = 1$ handelt es sich gerade um die Bilinearformen auf $\mathbb{R}^n$, darstellbar durch quadratische Matrizen, und es ergibt sich

$$(D^2 f(x))(u, v) = u^T H_f(x) v$$

für alle $u, v \in \mathbb{R}^n$, wenn man die Matrixdarstellung in der kanonischen Basis wählt. Diese Überlegungen kann man induktiv fortsetzen und für beliebiges $k \in \mathbb{N}$ eine k-te Ableitung $D^k f$ als multilineare Abbildung vom Grad k definieren.

10 Der Fixpunktsatz von Banach

Definition 10.1 (Fixpunkt)
Sei X Menge, $f : X \to X$. Ein $x \in X$ mit

$$x = f(x) \tag{10.1}$$

heißt Fixpunkt von f. Die Iteration

$$x_{k+1} = f(x_k)\,, \quad x_0 \in X \text{ gegeben,} \tag{10.2}$$

heißt Fixpunktiteration. □

"Fixpunktsätze" machen Aussagen über Lösungen der Fixpunktgleichung $x = f(x)$, etwa über Existenz und Eindeutigkeit von Fixpunkten, unter gewissen Voraussetzungen an X und f. Fixpunktsätze stellen ein zentrales Werkzeug der Analysis dar. Als Beispiel betrachten wir ein Anfangswertproblem für eine gewöhnliche Differentialgleichung,

$$x'(t) = \sin(tx(t))\,, \quad x(0) = 1\,. \tag{10.3}$$

Wir können (10.3) durch Integration umformen zu

$$x(t) = 1 + \int_0^t \sin(\tau x(\tau))\, d\tau\,. \tag{10.4}$$

Wir können (10.4) als Fixpunktgleichung der Form (10.1) interpretieren, wenn wir $X = C[0,1]$ setzen und $f : C[0,1] \to C[0,1]$ definieren durch

$$(f(x))(t) = 1 + \int_0^t \sin(\tau x(\tau))\, d\tau\,.$$

(Veranschaulichung der Fixpunktiteration für $X = \mathbb{R}$, siehe Bilder in der Vorlesung.)

Definition 10.2 (Lipschitz-Stetigkeit)
Seien (X, d_1), (Y, d_2) metrische Räume. Ein $f : X \to Y$ heißt Lipschitz-stetig mit der Lipschitz-Konstante L, falls gilt

$$d_2(f(x_1), f(x_2)) \le L d_1(x_1, x_2)\,, \quad \text{für alle } x_1, x_2 \in X. \tag{10.5}$$

Falls $Y = X$, $d_1 = d_2$ und $L < 1$, so heißt f Kontraktion. □

Jede Lipschitz-stetige Abbildung ist gleichmäßig stetig (das folgt unmittelbar aus den Definitionen).

Satz 10.3 (Banachscher Fixpunktsatz, Kontraktionssatz)
Sei (X, d) vollständiger metrischer Raum, $f : X \to X$ Kontraktion. Dann gilt:

(i) f hat genau einen Fixpunkt x.

(ii) Für jedes $x_0 \in X$ konvergiert die durch die Fixpunktiteration $x_{k+1} = f(x_k)$ definierte Folge $(x_k)_{k\in\mathbb{N}}$ gegen x.

(iii) Ist $L < 1$ eine Lipschitz-Konstante für f, so gilt die Fehlerabschätzung

$$d(x_k, x) \le \frac{L}{1-L} d(x_{k-1}, x_k)\,. \tag{10.6}$$

Beweis: Sei $x_0 \in X$ beliebig, sei $L < 1$ Lipschitz-Konstante für f. Dann gilt für die durch die Fixpunktiteration definierte Folge $(x_k)_{k\in\mathbb{N}}$

$$d(x_k, x_{k+1}) = d(f(x_{k-1}), f(x_k)) \le L d(x_{k-1}, x_k)\,, \qquad \text{für alle } k \in \mathbb{N},$$

also (mit Induktion)

$$d(x_k, x_{k+1}) \le L^k d(x_0, x_1)\,, \qquad \text{für alle } k \in \mathbb{N},$$

also für alle $k, m \in \mathbb{N}$

$$\begin{aligned} d(x_k, x_{k+m}) &\le d(x_k, x_{k+1}) + \ldots + d(x_{k+m-1}, x_{k+m}) \\ &\le L^k d(x_0, x_1) + \ldots + L^{k+m-1} d(x_0, x_1) = L^k \left(\sum_{j=0}^{m-1} L^j \right) d(x_0, x_1) \\ &\le \frac{L^k}{1-L} d(x_0, x_1) \qquad \text{da } L < 1. \end{aligned}$$

Die Folge $(x_k)_{k\in\mathbb{N}}$ ist also eine Cauchyfolge, welche wegen der Vollständigkeit von X gegen ein $x \in X$ konvergiert. Aus der Stetigkeit von f folgt

$$x = \lim_{k\to\infty} x_k = \lim_{k\to\infty} f(x_{k-1}) = f(x)\,.$$

Ist y ebenfalls Fixpunkt von x, so gilt

$$d(x, y) = d(f(x), f(y)) \le L d(x, y)\,,$$

also $(1 - L)d(x, y) \le 0$, also $d(x, y) = 0$ und damit $x = y$. Zum Beweis von (10.6) zeigen wir wie oben

$$d(x_k, x_{k+m}) \le L \left(\sum_{j=0}^{m-1} L^j \right) d(x_{k-1}, x_k) \le \frac{L}{1-L} d(x_{k-1}, x_k)\,,$$

also

$$d(x_k, x) = \lim_{m\to\infty} d(x_k, x_{k+m}) \le \frac{L}{1-L} d(x_{k-1}, x_k)\,.$$

$\square$

11 Inverse Funktionen im Mehrdimensionalen

Zur Motivation eine Wiederholung aus dem Eindimensionalen: Sei $f : (a,b) \to \mathbb{R}$ stetig differenzierbar. Ist $x \in (a,b)$ mit $f'(x) \neq 0$, so ist die Einschränkung $f : I_\delta \to \mathbb{R}$, $I_\delta = (x-\delta, x+\delta)$, für hinreichend kleines $\delta > 0$ streng monoton und damit, aufgefasst als Abbildung

$$f : I_\delta \to f(I_\delta)$$

bijektiv. Die Bildmenge $f(I_\delta)$ ist ebenfalls ein offenes Intervall im $\mathbb{R}$, und die Umkehrung $f^{-1} : f(I_\delta) \to I_\delta$ ist stetig differenzierbar. Aus der Voraussetzung $f'(x) \neq 0$ folgt also, dass f "lokal invertierbar" und die Inverse ebenfalls stetig differenzierbar ist.

Wir können auch einen globalen Satz formulieren: Ist $f'(x) \neq 0$ für alle $x \in (a,b)$, so ist f invertierbar, und $f^{-1} : f((a,b)) \to (a,b)$ ist ebenfalls stetig differenzierbar.

Wir betrachten nun Funktionen $f : \mathbb{R}^n \to \mathbb{R}^n$. In diesem Abschnitt beweisen wir einen Satz über die lokale Invertierbarkeit von f. Aussagen über globale Invertierbarkeit sind i.a. schwieriger und nicht Gegenstand dieser Vorlesung.

Falls die Inverse f^{-1} auf irgendeinem Teilgebiet existiert und differenzierbar ist, können wir ihre Ableitung sofort aus der Kettenregel berechnen. Aus $f^{-1} \circ f = id$ folgt

$$(D(f^{-1} \circ f))(x) = (D(id))(x) = id\,,$$

und weiter aus der Kettenregel

$$id = (D(f^{-1} \circ f))(x) = (Df^{-1})(f(x)) \circ (Df)(x)$$

beziehungsweise

$$I = J_{f^{-1}}(f(x)) \cdot J_f(x)\,.$$

Eine differenzierbare Inverse von f kann also nur dann existieren, wenn die Ableitung bzw. die Funktionalmatrix von f invertierbar ist.

Ist f eine lineare Abbildung, also $f(x) = Ax$ mit einer Matrix $A \in \mathbb{R}^{(n,n)}$, so ist $J_f(x) = A$ in allen Punkten x, und $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ existiert genau dann, wenn A invertierbar ist; in diesem Fall ist f^{-1} auch differenzierbar. Ist A nicht invertierbar, so ist die Einschränkung von f auf eine offene Kugel $B_\delta(x)$ nicht invertierbar, egal wie klein δ ist und wo x liegt. Damit ist für lineare Abbildungen im Endlichdimensionalen die Frage der Invertierbarkeit geklärt. In der Linearen Algebra wird weiter bewiesen, daß gilt

$$A \text{ invertierbar} \quad \Leftrightarrow \quad \operatorname{rang}(A) = n \quad \Leftrightarrow \quad \det(A) \neq 0$$

Satz 11.1 *Die Determinante, aufgefasst als Funktion*

$$\det : \mathbb{R}^{(n,n)} \to \mathbb{R} \tag{11.1}$$

ist beliebig oft stetig differenzierbar. Die Menge

$$GL(n,\mathbb{R}) = \{A : A \in \mathbb{R}^{(n,n)},\ A \text{ invertierbar}\} \tag{11.2}$$

ist eine offene Teilmenge des $\mathbb{R}^{(n,n)}$.

Beweis: In der Linearen Algebra wird die folgende Formel bewiesen:

$$\det(A) = \sum_{\pi} \operatorname{sign}(\pi) \prod_{i=1}^{n} a_{\pi(i),i} \,,$$

wobei die Summe über alle Permutationen π der Indexmenge $\{1, \ldots, n\}$ gebildet wird. Die Determinante, aufgefasst als Funktion der Matrixelemente, ist also ein Polynom und damit beliebig oft stetig differenzierbar. Die zweite Behauptung folgt aus der Darstellung

$$GL(n, \mathbb{R}) = \{A : A \in \mathbb{R}^{(n,n)}, \det(A) \neq 0\} \,,$$

d.h. $GL(n, \mathbb{R})$ ist das Urbild der offenen Menge $\mathbb{R} \setminus \{0\}$ unter der stetigen Abbildung "det". □

Folgerung 11.2 *Die Abbildung*

$$T : GL(n, \mathbb{R}) \to \mathbb{R}^{(n,n)} \,, \quad T(A) = A^{-1} \,, \tag{11.3}$$

ist stetig differenzierbar.

Beweis: Folgt aus Satz 11.1, da sich die Inverse mit einer Formel darstellen lässt, in der nur Determinanten vorkommen, nämlich (siehe Lineare Algebra)

$$A^{-1} = \frac{1}{\det A} (\operatorname{adj} A)^T \,,$$

wobei adj A diejenige Matrix im $\mathbb{R}^{(n,n)}$ ist, deren (i, j)-tes Element gegeben ist durch

$$(-1)^{i+j} \det(\tilde{A}_{ij}) \,,$$

und $\tilde{A}_{ij}$ diejenige Matrix im $\mathbb{R}^{(n-1),(n-1)}$ ist, die aus A durch Streichen der i-ten Zeile und j-ten Spalte entsteht. □

Satz 11.3 (Satz über inverse Funktionen, Spezialfall)
Sei $\Omega \subset \mathbb{R}^n$ offen mit $0 \in \Omega$, sei $f : \Omega \to \mathbb{R}^n$ stetig differenzierbar, es gelte $f(0) = 0$ und $Df(0) = id$, d.h. $J_f(0) = I$. Dann gibt es offene Mengen $U, V \subset \mathbb{R}^n$ mit $0 \in U \subset \Omega$, so dass gilt:

(i) $f|U : U \to V$ *ist bijektiv.*

(ii) $(f|U)^{-1} : V \to U$ *ist differenzierbar in* 0, $Df^{-1}(0) = id$.

Beweis: In diesem Beweis steht $\|\cdot\|$ für $\|\cdot\|_\infty$. Zur Konstruktion der lokalen Inverse wird der Banachsche Fixpunktsatz herangezogen. Zu diesem Zweck definieren wir für jedes $y \in \mathbb{R}^n$ eine Abbildung $T_y : \Omega \to \mathbb{R}^n$ durch

$$T_y(x) = x + y - f(x) \,. \tag{11.4}$$

Es gilt dann

$$y = f(x) \qquad \Leftrightarrow \qquad x \text{ ist Fixpunkt von } T_y$$

und weiter ist $T_y \in C^1(\Omega)$,

$$J_{T_y}(x) = J_{T_0}(x) = I - J_f(x)\,, \quad J_{T_y}(0) = 0\,.$$

Wir wählen ein $\delta > 0$, so dass gilt

$$\|T_y(x) - T_y(\xi)\| \le \frac{1}{2}\|x - \xi\|\,, \qquad \text{für alle } x, \xi \in K_\delta(0),\ y \in \mathbb{R}^n. \tag{11.5}$$

(Dass das möglich ist, folgt aus dem Mittelwertsatz 8.13, der Stetigkeit der Abbildung $x \mapsto J_{T_0}(x)$ und der Kompaktheit der abgeschlossenen Kugeln $K_\delta = K_\delta(0)$.) Für $x \in K_\delta$ gilt also

$$\|T_y(x)\| = \|x + y - f(x)\| \le \|T_0(x)\| + \|y\| \le \frac{1}{2}\|x\| + \|y\|\,. \tag{11.6}$$

Aus (11.5) und (11.6) folgt:

$$T_y : K_\delta \to K_\delta \text{ ist Kontraktion, falls } \|y\| \le \frac{\delta}{2}\,. \tag{11.7}$$

Aus dem Banachschen Fixpunktsatz folgt nun, dass alle solchen T_y einen eindeutigen Fixpunkt $x \in K_\delta$ haben, also gilt

$$\text{zu jedem } y \in K_{\frac{\delta}{2}} \text{ gibt es genau ein } x \in K_\delta \text{ mit } f(x) = y\,, \tag{11.8}$$

und für dieses x gilt wegen (11.6) und $x = T_y(x)$, dass

$$\|x\| \le 2\|y\|\,. \tag{11.9}$$

Wir definieren ($B_\delta = B_\delta(0)$ offene Kugel)

$$V = B_{\frac{\delta}{2}}\,, \quad U = f^{-1}(V) \cap B_\delta\,.$$

Es ist $0 \in U$, $U \subset \mathbb{R}^n$ offen, $f(U) \subset V$, und wegen (11.8) ist $f|U$ injektiv. Es ist auch $V \subset f(U)$ wegen (11.9), also ist (i) bewiesen. Zum Beweis von (ii) genügt es zu zeigen: Für jede Folge $(y_k)_{k\in\mathbb{N}}$ in $B_{\frac{\delta}{2}}$ mit $y_k \to 0$, $y_k \neq 0$ für alle k, gilt

$$\lim_{k\to\infty} \frac{\|f^{-1}(y_k) - f^{-1}(0) - I(y_k - 0)\|}{\|y_k - 0\|} = 0\,. \tag{11.10}$$

Sei (y_k) eine solche Folge, sei $x_k = f^{-1}(y_k)$, dann ist $x_k \neq 0$ für alle k, und es gilt

$$\begin{aligned}
\frac{\|f^{-1}(y_k) - f^{-1}(0) - I(y_k - 0)\|}{\|y_k - 0\|} &= \frac{\|f^{-1}(y_k) - y_k\|}{\|y_k\|} \\
&= \frac{\|x_k - f(x_k)\|}{\|x_k\|} \cdot \frac{\|x_k\|}{\|y_k\|} \\
&\le \frac{\|f(x_k) - f(0) - I(x_k - 0)\|}{\|x_k\|} \cdot 2\,,
\end{aligned}$$

und die rechte Seite dieser Ungleichung konvergiert gegen 0 für $k \to \infty$, da $J_f(0) = I$. Damit ist (11.10) bewiesen. □

Satz 11.4 (Satz über inverse Funktionen)
Sei $\Omega \subset \mathbb{R}^n$ offen, sei $f : \Omega \to \mathbb{R}^n$ stetig differenzierbar, sei $x_* \in \Omega$, sei $Df(x_*)$ invertierbar. Dann gibt es offene Mengen $U, V \subset \mathbb{R}^n$ mit $x_* \in U \subset \Omega$, so dass gilt:

(i) $f|U : U \to V$ ist bijektiv.

(ii) $(f|U)^{-1} : V \to U$ ist stetig differenzierbar in V, und für alle $x \in U$ gilt

$$Df^{-1}(f(x)) = (Df(x))^{-1}\,, \quad J_{f^{-1}}(f(x)) = J_f(x)^{-1}\,. \tag{11.11}$$

Beweis: Die Menge

$$\begin{aligned}\tilde{\Omega} &= \{x : x \in \Omega,\ Df(x) \text{ ist invertierbar}\} \\ &= \Omega \cap \{x : J_f(x) \in GL(n,\mathbb{R})\} = \Omega \cap (J_f)^{-1}(GL(n,\mathbb{R}))\end{aligned}$$

ist offen nach Satz 11.1, da J_f stetig von x abhängt. Da es genügt, den Beweis für $\tilde{\Omega}$ statt Ω zu führen, können wir o.B.d.A. annehmen, dass $Df(x)$ invertierbar ist für alle $x \in \Omega$. Wir setzen $\Omega_* = \Omega - x_*$ und definieren $f_* : \Omega_* \to \mathbb{R}^n$ durch

$$f_*(h) = (Df(x_*))^{-1}(f(x_* + h) - f(x_*))\,. \tag{11.12}$$

Für f_* sind alle Voraussetzungen von Satz 11.3 erfüllt, da $f_*(0) = 0$ und $Df_*(0) = Df(x_*)^{-1} \circ Df(x_*) = id$, also gibt es $U_*, V_* \subset \mathbb{R}^n$ offen mit $0 \in U_* \subset \Omega_*$, $f_* : U_* \to V_*$ bijektiv, f_*^{-1} differenzierbar in 0 mit $Df_*^{-1}(0) = id$. Wir setzen

$$U = U_* + x_*\,.$$

Aus (11.12) folgt, dass für alle $x \in U$ gilt

$$f(x) = f(x_*) + Df(x_*)(f_*(x - x_*))\,. \tag{11.13}$$

Da f_* auf U_* bijektiv und $Df(x_*)$ invertierbar, ist f auf U injektiv. Wir setzen

$$V = f(U) = f(x_*) + Df(x_*)(f_*(U_*)) = f(x_*) + Df(x_*)(V_*)\,.$$

Mit V_* ist auch V offen, und $f|U : U \to V$ bijektiv. Nach (11.12) gilt für alle $x \in U$

$$f_*(x - x_*) = Df(x_*)^{-1}(f(x) - f(x_*))\,,$$

also

$$x = x_* + f_*^{-1}[Df(x_*)^{-1}(f(x) - f(x_*)]\,,$$

also für alle $y \in V$

$$f^{-1}(y) = x_* + f_*^{-1}[Df(x_*)^{-1}(y - f(x_*)]\,.$$

Aus der Kettenregel folgt

$$(Df^{-1})(f(x_*)) = (Df_*^{-1})(0) \circ (Df(x_*))^{-1} = (Df(x_*))^{-1}\,, \quad J_{f^{-1}}(f(x_*)) = J_f(x_*)^{-1}\,.$$

Da für jedes $x \in U$ die Voraussetzungen des Satzes (mit x statt x_*) erfüllt sind, ist f^{-1} auf ganz V differenzierbar, und (11.11) gilt. Es ist außerdem f^{-1} stetig auf V nach Satz 8.3, und wegen Folgerung 11.2 auch die Abbildung

$$y \mapsto (J_f(f^{-1}(y)))^{-1}\,.$$

$\square$

Wir kommen auf die Polarkoordinaten

$$f : \mathbb{R}^2 \to \mathbb{R}^2 \,, \quad f(r, \varphi) = (r \cos \varphi, r \sin \varphi)$$

zurück. Es ist

$$\det(J_f(r, \varphi)) = r \,,$$

also ist f lokal invertierbar in allen Punkten (r, φ) mit $r \neq 0$. Es gilt darüber hinaus, dass $f : U \to V$ bijektiv ist für

$$U = \{(r, \varphi) : r > 0,\, 0 < \varphi < 2\pi\} \,, \quad V = \{(x, y) : y \neq 0 \text{ oder } x < 0\} \,.$$

12 Implizite Funktionen

Im einfachsten Fall wollen wir eine Gleichung der Form

$$f(x,y) = 0\,, \quad f : \mathbb{R}^2 \to \mathbb{R}\,,$$

nach y auflösen. Beispiel:

$$xy - 1 = 0\,,$$

Lösung:

$$y = \frac{1}{x}\,.$$

Wir haben also eine Funktion $g : \mathbb{R} \to \mathbb{R}$ gefunden, nämlich $g(x) = 1/x$, so dass gilt

$$f(x, g(x)) = 0\,. \tag{12.1}$$

Im allgemeinen ist das nicht in einer so explizit angebbaren Form möglich, oder vielleicht auch nicht sinnvoll. Es stellt sich dann u.a. die Frage, unter welchen Voraussetzungen an f ein solches g existiert. Weiteres Beispiel: Der Einheitskreis

$$0 = f(x,y) = x^2 + y^2 - 1\,. \tag{12.2}$$

Hier liefern

$$g(x) = +\sqrt{1-x^2}\,, \quad g(x) = -\sqrt{1-x^2}\,,$$

Lösungen von (12.1). Hier ist also die Lösung "global" mehrdeutig, aber "lokal" (d.h. in einer hinreichend kleinen Umgebung eines Punktes (x,y) mit $f(x,y) = 0$) eindeutig für $|x| < 1$; lokal mehrdeutig für $|x| = 1$, dort ist g nicht differenzierbar; nicht definiert für $|x| > 1$. Existiert eine differenzierbare Funktion g, so dass (12.1) gilt, so erhält man aus der Kettenregel

$$0 = \frac{d}{dx} f(x, g(x)) = \partial_x f(x, g(x)) + \partial_y f(x, g(x)) g'(x)\,,$$

also

$$g'(x) = -\frac{\partial_x f(x, g(x))}{\partial_y f(x, g(x))}\,,$$

jedenfalls dann, wenn $\partial_y f(x, g(x)) \neq 0$. Andernfalls kann es Probleme geben. Für (12.2) haben wir

$$\partial_x f(x,y) = 2x\,, \quad \partial_y f(x,y) = 2y\,.$$

Die allgemeine Situation im Endlichdimensionalen sieht so aus: Wir haben m Gleichungen mit $n+m$ Unbekannten, von denen wir mit Hilfe der Gleichungen m Unbekannte eliminieren wollen, indem wir sie als Funktionen der anderen n Unbekannten ausdrücken.

$$\begin{aligned} f_1(x_1, \ldots, x_n, y_1, \ldots, y_m) &= 0\,, \\ f_2(x_1, \ldots, x_n, y_1, \ldots, y_m) &= 0\,, \\ &\vdots \\ f_m(x_1, \ldots, x_n, y_1, \ldots, y_m) &= 0\,. \end{aligned}$$

Gesucht sind Funktionen g_i, "$y_i = g_i(x_1, \ldots, x_n)$", für $1 \le i \le m$, mit

$$f_j(x_1, \ldots, x_n, g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n)) = 0\,, \quad 1 \le j \le m\,.$$

Erheblich übersichtlicher ist die vektorielle Schreibweise: Gegeben ist

$$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m\,,$$

gesucht ist

$$g : \mathbb{R}^n \to \mathbb{R}^m \qquad \text{mit} \qquad f(x, g(x)) = 0$$

für alle (oder doch möglichst viele) $x \in \mathbb{R}^n$. In einem solchen Fall empfiehlt es sich, auch für die Jacobi-Matrix eine Block-Notation einzuführen. Ist $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ differenzierbar, und schreiben wir $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_m)$ für die Elemente in $\mathbb{R}^n \times \mathbb{R}^m$, so definieren wir

$$\partial_x f(x,y) = \begin{pmatrix} \partial_{x_1} f_1(x,y) & \cdots & \partial_{x_n} f_1(x,y) \\ \vdots & & \vdots \\ \partial_{x_1} f_k(x,y) & \cdots & \partial_{x_n} f_k(x,y) \end{pmatrix} \in \mathbb{R}^{(k,n)}\,, \tag{12.3}$$

$$\partial_y f(x,y) = \begin{pmatrix} \partial_{y_1} f_1(x,y) & \cdots & \partial_{y_m} f_1(x,y) \\ \vdots & & \vdots \\ \partial_{y_1} f_k(x,y) & \cdots & \partial_{y_m} f_k(x,y) \end{pmatrix} \in \mathbb{R}^{(k,m)}\,. \tag{12.4}$$

Die gesamte Jacobi-Matrix setzt sich aus zwei nebeneinanderstehenden Blöcken zusammen,

$$J_f(x,y) = \begin{pmatrix} \partial_x f(x,y) & \partial_y f(x,y) \end{pmatrix} \in \mathbb{R}^{(k,n+m)}\,.$$

Satz 12.1 (Implizite Funktionen)
Sei $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$ offen, sei $f : \Omega \to \mathbb{R}^m$ stetig differenzierbar, sei $(x_*, y_*) \in \Omega$ mit $f(x_*, y_*) = 0$. Ist $\partial_y f(x_*, y_*)$ invertierbar, so gibt es eine Umgebung W von x_* und eine stetig differenzierbare Funktion $g : W \to \mathbb{R}^m$ mit $y_* = g(x_*)$ und

$$f(x, g(x)) = 0\,, \quad (x, g(x)) \in \Omega\,, \quad \textit{für alle } x \in W, \tag{12.5}$$

und es gilt

$$J_g(x) = -\partial_y f(x, g(x))^{-1} \partial_x f(x, g(x)) \tag{12.6}$$

für alle $x \in W$.

Beweis: Wir definieren

$$F : \Omega \to \mathbb{R}^n \times \mathbb{R}^m\,, \quad F(x,y) = (x, f(x,y))\,.$$

Wir zeigen, dass F in (x_*, y_*) die Voraussetzungen des Satzes 11.4 über inverse Funktionen erfüllt. Dazu genügt es zu zeigen, dass die Funktionalmatrix

$$J_F(x,y) = \begin{pmatrix} I_n & 0 \\ \partial_x f(x,y) & \partial_y f(x,y) \end{pmatrix}$$

für $(x, y) = (x_*, y_*)$ invertierbar ist. In der Tat, die Matrix

$$J_F(x, y) = \begin{pmatrix} I_n & 0 \\ B(x, y) & (\partial_y f(x, y))^{-1} \end{pmatrix}, \quad B(x, y) = -(\partial_y f(x, y))^{-1} \partial_x f(x, y),$$

ist für $(x, y) = (x_*, y_*)$ die Inverse von $J_F(x, y)$. Aus Satz 11.4 folgt nun: Es gibt offene Mengen $U, V \subset \mathbb{R}^n \times \mathbb{R}^m$ mit $(x_*, y_*) \in U \subset \Omega$, so dass $F : U \to V$ bijektiv und $F^{-1} : V \to U$ stetig differenzierbar ist. Wir zerlegen F^{-1} in zwei Blöcke,

$$F^{-1}(x, y) = (\varphi_x(x, y), \varphi_y(x, y)),$$

mit $\varphi_x : V \to \mathbb{R}^n$, $\varphi_y : V \to \mathbb{R}^m$. Für $(x, y) \in V$ gilt dann

$$\begin{aligned}(x, y) &= F(F^{-1}(x, y)) = F(\varphi_x(x, y), \varphi_y(x, y)) \\ &= (\varphi_x(x, y), f(\varphi_x(x, y), \varphi_y(x, y))),\end{aligned}$$

also

$$x = \varphi_x(x, y), \quad y = f(x, \varphi_y(x, y)). \tag{12.7}$$

Wir wählen nun offene Mengen $W \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, mit

$$F(x_*, y_*) = (x_*, 0) \in W \times Y \subset V,$$

und definieren $g : W \to U$ durch

$$g(x) = \varphi_y(x, 0). \tag{12.8}$$

Ausserdem verlangen wir, daß $\partial_y f(x, g(x))$ invertierbar ist für alle $x \in W$ (das können wir immer durch Verkleinern von W erreichen). Es gilt dann wegen (12.7)

$$0 = f(x, \varphi_y(x, 0)) = f(x, g(x)), \quad \text{für alle } x \in W,$$

und

$$(x_*, y_*) = F^{-1}(x_*, 0) = (x_*, \varphi_y(x_*, 0)) = (x_*, g(x_*)),$$

also

$$y_* = g(x_*).$$

Da F^{-1} stetig differenzierbar ist, ist auch g stetig differenzierbar. Die Formel (12.6) für J_g folgt aus der Kettenregel, angewendet auf

$$x \mapsto \begin{pmatrix} x \\ g(x) \end{pmatrix} \mapsto f(x, g(x)) = 0,$$

da wir durch Differenzieren erhalten

$$\begin{aligned}0 &= \begin{pmatrix} \partial_x f(x, g(x)) & \partial_y f(x, g(x)) \end{pmatrix} \begin{pmatrix} I_n \\ J_g(x) \end{pmatrix} \\ &= \partial_x f(x, g(x)) + \partial_y f(x, g(x)) J_g(x).\end{aligned}$$

$\square$

Als Beispiel betrachten wir die Niveaulinien einer stetig differenzierbaren Funktion $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^2$. Sei $(x_*, y_*) \in \Omega$, $f(x_*, y_*) = c$, also $(x_*, y_*) \in N_c(f)$. Die Frage ist: Wie sieht $N_c(f)$ in der Nähe von (x_*, y_*) aus? Wir betrachten zunächst den regulären Fall

$$\operatorname{grad} f(x_*, y_*) \neq 0 \,. \tag{12.9}$$

In diesem Fall ist mindestens eine der beiden partiellen Ableitungen nicht Null. Ist $\partial_y f(x_*, y_*) \neq 0$, so gibt es nach Satz 12.1 ein offenes Intervall $I = (x_* - \delta, x_* + \delta)$ und ein $g \in C^1(I)$ mit $g(x_*) = y_*$ und $f(x, g(x)) = 0$, d.h. y lässt sich in der Nähe von (x_*, y_*) nach x auflösen. Ist $\partial_x f(x_*, y_*) \neq 0$, so gibt es nach Satz 12.1 ein offenes Intervall $J = (y_* - \delta, y_* + \delta)$ und ein $h \in C^1(J)$ mit $h(y_*) = x_*$ und $f(h(y), y) = 0$, d.h. x lässt sich in der Nähe von (x_*, y_*) nach y auflösen. Das ist der Fall in Beispiel (12.2), in dem

$$\begin{aligned} \partial_x f(x,y) &\neq 0 \,, \quad \text{ausser in } (0,1),\ (0,-1), \\ \partial_y f(x,y) &\neq 0 \,, \quad \text{ausser in } (1,0),\ (-1,0). \end{aligned}$$

Man kann ausserdem zeigen, dass es im regulären Fall (12.9) in einer hinreichend kleinen Umgebung von (x_*, y_*) keine anderen Punkte von $N_c(f)$ geben kann.

Der singuläre Fall ist charakterisiert durch

$$\operatorname{grad} f(x_*, y_*) = 0 \,. \tag{12.10}$$

Hier kann alles mögliche passieren. Sei etwa

$$f(x,y) = x^3 - y^2 = 0 \,, \tag{12.11}$$

dann ist

$$\operatorname{grad} f(x,y) = (3x^2, -2y) \,, \quad \operatorname{grad} f(0) = 0 \,.$$

Wir können x nach y auflösen,

$$x = h(y) = \sqrt[3]{y^2} \,,$$

aber h ist nicht differenzierbar in 0, und $N_0(f)$ hat eine Spitze in 0. Für

$$f(x,y) = x^2 - y^2 = 0 \,, \tag{12.12}$$

gilt

$$\operatorname{grad} f(x,y) = (2x, -2y) \,, \quad \operatorname{grad} f(0) = 0 \,,$$

und $N_0(f)$ besteht aus den beiden Winkelhalbierenden (die sich im Nullpunkt schneiden). Eine kompliziertere Situation liegt vor bei

$$f(x,y) = r^6 \sin \frac{1}{r^2} \,, \quad r = \sqrt{x^2 + y^2} \,, \quad f(0) = 0 \,. \tag{12.13}$$

Hier ist ebenfalls $\operatorname{grad} f(0) = 0$, und $N_0(f)$ besteht aus unendlich vielen Kreisen um 0 mit den Radien r, wobei

$$\frac{1}{r^2} = \pi n \,, \quad n \in \mathbb{N} \,,$$

sowie den Nullpunkt selbst. Es gilt allgemein (ohne Beweis):

Zu jeder abgeschlossenen Menge $A \subset \mathbb{R}^n$ gibt es eine stetig differenzierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ mit $A = N_0(f)$.

Wir wenden den Satz über implizite Funktionen an auf das Problem, das Minimum einer Funktion "unter einer Nebenbedingung" zu bestimmen.

Problem 12.2
Sei $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$, $J : \Omega \to \mathbb{R}$. Wir betrachten die Optimierungsaufgabe

$$\min J(x)\,, \quad \text{wobei} \quad x \in \Omega,\ f(x) = 0\,. \tag{12.14}$$

Die Bedingung "$f(x) = 0$" heißt Nebenbedingung (genauer: Gleichungsnebenbedingung). Ein $x_* \in \Omega$ heißt lokale Lösung von (12.14), falls es eine Umgebung U von x_* gibt, so dass

$$J(x_*) \le J(x)\,, \quad \text{für alle } x \in U \cap \Omega \text{ mit } f(x) = 0.$$

Satz 12.3 *Sei $\Omega \subset \mathbb{R}^n$ offen, seien $f, J \in C^1(\Omega)$, sei x_* lokale Lösung von (12.14), es gelte* $\operatorname{grad} f(x_*) \neq 0$. *Dann gibt es ein $\lambda \in \mathbb{R}$ mit*

$$\operatorname{grad} J(x_*) = \lambda \operatorname{grad} f(x_*)\,. \tag{12.15}$$

Die Zahl λ heißt Lagrange-Multiplikator.

Beweis: Sei o.B.d.A $\partial_n f(x_*) \neq 0$ (lässt sich durch Umnumerieren der Unbekannten immer erreichen). Wir schreiben

$$x_* = (x'_*, x_{*n})\,, \quad x'_* = (x_{*1}, \ldots, x_{*n-1})\,.$$

Nach Satz 12.1 gibt es eine stetig differenzierbare Funktion $g : W \to \mathbb{R}$, $W \subset \mathbb{R}^{n-1}$ offen mit $x'_* \in W$, so dass

$$f(x', g(x')) = 0\,, \quad (x', g(x')) \in \Omega\,, \quad \text{für alle } x' \in W.$$

Wenn wir die durch

$$h(x_i) = f(x', g(x'))\,, \quad x' = (x'_1, \ldots, x'_{n-1})\,,$$

definierte Abbildung h_i in $x' = x'_*$ differenzieren, erhalten wir

$$\partial_i f(x_*) + \partial_n f(x_*) \partial_i g(x'_*) = 0\,, \quad 1 \le i \le n-1\,. \tag{12.16}$$

Wir betrachten

$$\tilde{J} : W \to \mathbb{R}\,, \quad \tilde{J}(x') = J(x', g(x'))\,.$$

x'_* ist lokales Minimum von $\tilde{J}$, also gilt

$$0 = \partial_i J(x_*) + \partial_n J(x_*) \partial_i g(x'_*)\,, \quad 1 \le i \le n-1\,. \tag{12.17}$$

Wir multiplizieren (12.17) mit $\partial_n f(x_*)$, (12.16) mit $\partial_n J(x_*)$ und subtrahieren die resultierenden Gleichungen, dann

$$\partial_i J(x_*) \partial_n f(x_*) - \partial_i f(x_*) \partial_n J(x_*) = 0\,, \quad 1 \le i \le n\,. \tag{12.18}$$

Wir setzen

$$\lambda = \frac{\partial_n J(x_*)}{\partial_n f(x_*)}$$

und erhalten

$$\partial_i J(x_*) = \lambda \partial_i f(x_*)\,, \quad 1 \le i \le n\,.$$

$\square$

Als erstes Beispiel betrachten wir

$$\min J(x,y) = \frac{1}{2}(x^2+y^2)\,, \quad \text{wobei} \quad e^{xy} = x+y\,. \tag{12.19}$$

Hier ist

$$f(x,y) = e^{xy} - x - y\,,$$
$$\operatorname{grad} J(x,y) = (x,y)\,, \quad \operatorname{grad} f(x,y) = (ye^{xy}-1, xe^{xy}-1)\,.$$

Gemäß Satz 12.3 suchen wir Kandidaten $(x,y,\lambda) \in \mathbb{R}^3$, welche die Gleichungen

$$x = \lambda(ye^{xy}-1)\,, \tag{12.20}$$
$$y = \lambda(xe^{xy}-1)\,. \tag{12.21}$$

sowie

$$e^{xy} = x+y\,. \tag{12.22}$$

erfüllen. Wir haben also das Optimierungsproblem darauf zurückgeführt, das System (12.20) - (12.22) von 3 nichtlinearen Gleichungen mit den 3 Unbekannten (x,y,λ) zu lösen – was im allgemeinen nicht durch explizite Angabe einer Lösung möglich ist.

Als weiteres Beispiel betrachten wir das Problem, die Extrema einer quadratischen Funktion auf dem Rand der Einheitskugel zu bestimmen. Sei $J : \mathbb{R}^n \to \mathbb{R}$ definiert durch

$$J(x) = x^T A x\,, \quad A \in \mathbb{R}^{(n,n)} \text{ symmetrisch,} \tag{12.23}$$

die Nebenbedingung sei

$$0 = f(x) = \sum_{i=1}^n x_i^2 - 1\,.$$

Es ist

$$\operatorname{grad} J(x) = 2Ax\,, \quad \operatorname{grad} f(x) = 2x\,.$$

Die Funktion J nimmt auf dem Rand der Einheitskugel ihr Maximum und ihr Minimum an, in beiden Fällen folgt aus Satz 12.3, im Falle des Maximums angewendet auf $-f$, die Existenz eines $\lambda \in \mathbb{R}$ mit $\operatorname{grad} J(x) = \lambda \operatorname{grad} f(x)$, also

$$Ax = \lambda x\,, \tag{12.24}$$

das heißt, λ ist Eigenwert von A zum Eigenvektor x. Aus (12.24) folgt

$$J(x) = x^T A x = x^T \lambda x = \lambda \|x\|_2^2 = \lambda\,. \tag{12.25}$$

Da diese Formel für jeden Eigenwert gilt, wenn wir einen zugehörigen Eigenvektor auf 1 normieren, haben wir das folgende Ergebnis erhalten:

$$\max_{\|x\|_2=1} J(x) = \lambda_{max}\,, \quad \min_{\|x\|_2=1} J(x) = \lambda_{min}\,, \tag{12.26}$$

wobei λ_{max} und λ_{min} der größte bzw. kleinste Eigenwert von A sind.

13 Parameterabhängige Integrale

Satz 13.1 *Seien (X, d_1), (Y, d_2) metrische Räume, sei $f : X \times Y \to \mathbb{R}$ stetig bezüglich der Produktmetrik d. Dann wird durch*

$$(F(y))(x) = f(x, y) \tag{13.1}$$

für jedes $y \in Y$ eine stetige Funktion $F(y) : X \to \mathbb{R}$ definiert. Ist X kompakt, so ist

$$F : Y \to (C(X), \|\cdot\|_\infty) \tag{13.2}$$

stetig.

Beweis: Sei $y \in Y$ beliebig, sei $(x_k)_{k\in\mathbb{N}}$ Folge in X mit $x_k \to x$. Dann gilt

$$d((x_k, y), (x, y)) = d_1(x_k, x) \to 0\,,$$

also $(x_k, y) \to (x, y)$ in $X \times Y$ und damit

$$(F(y))(x_k) = f(x_k, y) \to f(x, y) = (F(y))(x)\,,$$

also ist $F(y)$ stetig. Sei nun X kompakt, sei $(y_k)_{k\in\mathbb{N}}$ Folge in Y mit $y_k \to y$. Zu zeigen ist

$$\|F(y_k) - F(y)\|_\infty = \sup_{x\in X} |f(x, y_k) - f(x, y)| \to 0 \tag{13.3}$$

für $k \to \infty$. Wir definieren $K \subset Y$ durch

$$K = \{y_k : k \in \mathbb{N}\} \cup \{y\}\,.$$

Dann ist $(K, d_2|K)$ kompakt (Übung). Nach Satz 5.2 ist $X \times K$ kompakt, nach Satz 5.12 ist $f|(X \times K)$ gleichmäßig stetig. Sei nun $\varepsilon > 0$ beliebig, sei $\delta > 0$ gemäß der Definition der gleichmäßigen Stetigkeit gewählt. Sei N gewählt mit $d_2(y_k, y) < \delta$ für alle $k \geq N$. Dann gilt

$$d((x, y_k), (x, y)) = d_2(y_k, y) < \delta\,, \quad \text{für alle } k \geq N \text{ und alle } x \in X,$$

also

$$|(f(x, y_k) - f(x, y)| < \varepsilon\,, \quad \text{für alle } k \geq N \text{ und alle } x \in X,$$

also

$$\sup_{x\in X} |f(x, y_k) - f(x, y)| \leq \varepsilon\,.$$

Damit ist (13.3) bewiesen. □

Satz 13.2 *Sei (Y, d_2) metrischer Raum, sei $[a, b] \subset \mathbb{R}$ kompaktes Intervall, sei $f : [a, b] \times Y \to \mathbb{R}$ stetig. Dann wird durch*

$$\varphi(y) = \int_a^b f(x, y)\, dx \tag{13.4}$$

eine stetige Funktion $\varphi : Y \to \mathbb{R}$ definiert.

Beweis: Es ist $\varphi = I \circ F$, F wie in Satz 13.1,

$$I : (C[a,b], \|\cdot\|_\infty) \to \mathbb{R}\,, \quad I(g) = \int_a^b g(x)\,dx\,.$$

F ist stetig nach Satz 13.1, I ist stetig nach Beispiel 4.24. □

Satz 13.3 (Differenzieren unter dem Integral)
Sei $f : [a,b] \times [c,d] \to \mathbb{R}$ stetig, $\partial_2 f : [a,b] \times [c,d] \to \mathbb{R}$ existiere und sei stetig. Dann ist $\varphi : [c,d] \to \mathbb{R}$,

$$\varphi(y) = \int_a^b f(x,y)\,dx\,, \tag{13.5}$$

stetig differenzierbar, und

$$\varphi'(y) = \int_a^b \partial_2 f(x,y)\,dx\,. \tag{13.6}$$

Beweis: Sei $y \in [c,d]$, (y_k) Folge in $[c,d]$ mit $y_k \to y$, $y_k \neq y$. Wir setzen

$$g_k(x) = \frac{f(x,y_k) - f(x,y)}{y_k - y} - \partial_2 f(x,y)\,.$$

Aus dem Mittelwertsatz folgt die Existenz von $\eta_k(x) \in [y, y_k]$ mit

$$g_k(x) = \partial_2 f(x, \eta_k(x)) - \partial_2 f(x,y)\,.$$

Sei $\varepsilon > 0$. Wähle $\delta > 0$ so, dass gilt

$$|y - z| < \delta \qquad \Rightarrow \qquad |\partial_2 f(x,y) - \partial_2 f(x,z)| < \varepsilon \quad \text{für alle } x \in [a,b].$$

(Das ist möglich, da $\partial_2 f$ gleichmäßig stetig ist auf $[a,b] \times [c,d]$.) Wähle N so, dass

$$|y_k - y| < \delta\,, \quad \text{für alle } k \geq N.$$

Dann gilt $|\eta_k(x) - y| < \delta$ für alle $x \in [a,b]$ und

$$\|g_k\|_\infty = \sup_{x \in [a,b]} |g_k(x)| \leq \varepsilon\,, \quad \text{für alle } k \geq N,$$

also gilt $g_k \to 0$ gleichmäßig, und es folgt

$$\left| \frac{\varphi(y_k) - \varphi(y)}{y_k - y} - \int_a^b \partial_2 f(x,y)\,dx \right| \leq \int_a^b |g_k(x)|\,dx \to 0$$

für $k \to \infty$. □

14 Kurvenintegrale und Potentiale

Zur Motivation: Ein konstantes Kraftfeld, repräsentiert durch einen Vektor $F \in \mathbb{R}^3$, leistet entlang der Strecke, welche einen Punkt $x_0 \in \mathbb{R}^3$ mit einem Punkt $x_1 \in \mathbb{R}^3$ verbindet, die Arbeit

$$W = \langle F, x_1 - x_0 \rangle = \|F\|_2 \|x_1 - x_0\|_2 \cos\varphi$$

wobei φ der von den Vektoren F und $x_1 - x_0$ eingeschlossene Winkel ist. Betrachten wir den Streckenzug, welcher nacheinander die Punkte $x_0, x_1, \ldots x_N$ verbindet, und nehmen wir an, dass längs der Verbindung von x_{i-1} nach x_i die Kraft F_i wirkt, so ergibt sich die Gesamtarbeit zu

$$W = \sum_{i=1}^{N} \langle F_i, x_i - x_{i-1} \rangle \ . \tag{14.1}$$

Wenn wir diesen Streckenzug als Kurve $x : [a, b] \to \mathbb{R}^3$ mit $x(t_i) = x_i$ zu einer geeigneten Zerlegung auffassen und $F(x_i) = F_i$ setzen, so wird (14.1) zu

$$W = \sum_{i=1}^{N} \left\langle F(x(t_i)), \frac{x(t_i) - x(t_{i-1})}{t_i - t_{i-1}} \right\rangle (t_i - t_{i-1}) \,, \tag{14.2}$$

Ist $F : \mathbb{R}^3 \to \mathbb{R}^3$ stetig und $x : [a, b] \to \mathbb{R}^3$ stetig differenzierbar, so ist die längs x verrichtete Arbeit gleich

$$\int_a^b \langle F(x(t)), x'(t) \rangle \, dt \,.$$

Definition 14.1 (Kurvenintegral)
Sei $\Omega \subset \mathbb{R}^n$, seien $F : \Omega \to \mathbb{R}^n$ stetig und $x : [a, b] \to \mathbb{R}^n$ stetig differenzierbar, sei $C = x([a, b])$. Dann heißt

$$\int_C F \cdot dx := \int_a^b \langle F(x(t)), x'(t) \rangle \, dt \tag{14.3}$$

das Kurvenintegral von F entlang C von $x(a)$ nach $x(b)$. □

Diese Definition ist sinnvoll, da das Kurvenintegral sich nicht ändert, wenn wir eine andere Parametrisierung von C wählen, die die Orientierung erhält. Ist etwa $\varphi : [\alpha, \beta] \to [a, b]$ eine C^1-Parametertransformation, so gilt

$$\begin{aligned}\int_\alpha^\beta \langle F((x \circ \varphi)(\tau)), (x \circ \varphi)'(\tau) \rangle \, \varphi'(\tau) \, d\tau &= \int_\alpha^\beta \langle F(x(\varphi(\tau))), x'(\varphi(\tau)) \rangle \, d\tau \\ &= \pm \int_{\varphi(\alpha)}^{\varphi(\beta)} \langle F(x(t)), x'(t) \rangle \, d\tau = \pm \int_a^b \langle F(x(t)), x'(t) \rangle \, d\tau \,,\end{aligned}$$

je nachdem, ob φ orientierungserhaltend oder orientierungsumkehrend ist.

Für stückweise stetig differenzierbare Kurven $x : [a, b] \to \mathbb{R}^n$ setzen wir

$$\int_C F \cdot dx = \sum_{i=1}^{k} \int_{C_i} F \cdot dx \,, \tag{14.4}$$

falls $x|[t_{i-1}, t_i]$ stetig differenzierbar ist und $C_i = x([t_{i-1}, t_i])$. (Man zeigt, dass diese Definition unabhängig ist von der Wahl der Zerlegung.) Aus der Linearität des Integrals folgt unmittelbar

$$\int_C (F+G) \cdot dx = \int_C F \cdot dx + \int_C G \cdot dx \,, \quad \int_C \lambda F \cdot dx = \lambda \int_C F \cdot dx \,, \tag{14.5}$$

für $\lambda \in \mathbb{R}$.

Satz 14.2 *Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^1(\Omega)$, $x : [a,b] \to \mathbb{R}^n$ stückweise stetig differenzierbar. Dann gilt für $C = x([a,b])$*

$$\int_C \operatorname{grad} f \cdot dx = f(x(b)) - f(x(a)) \,. \tag{14.6}$$

Beweis: Sei zunächst $x \in C^1([a,b])$. Dann gilt

$$\begin{aligned} \int_C \operatorname{grad} f \cdot dx &= \int_a^b \langle \operatorname{grad} f(x(t)), x'(t) \rangle \; dt = \int_a^b (f \circ x)'(t) \, dt \\ &= f(x(b)) - f(x(a)) \end{aligned}$$

nach Kettenregel und Hauptsatz. Ist x stückweise stetig differenzierbar und (t_i) eine entsprechende Zerlegung von $[a,b]$, so folgt

$$\int_C \operatorname{grad} f \cdot dx = \sum_{i=1}^k \int_{C_i} \operatorname{grad} f \cdot dx = \sum_{i=1}^k (f(x(t_i)) - f(x(t_{i-1}))) = f(x(b)) - f(x(a)) \,.$$

□

Aus Satz 14.2 folgt: Ist $F : \Omega \to \mathbb{R}^n$ ein Gradientenfeld, so ist das Kurvenintegral zwischen zwei Punkten $P, Q \in \mathbb{R}^n$ unabhängig davon, wie die Kurve von P nach Q verläuft. Insbesondere ist dann

$$\int_C F \cdot dx = 0$$

für jede geschlossene Kurve C.

Lemma 14.3 *Sei $\Omega \subset \mathbb{R}^n$ offen, $F : \Omega \to \mathbb{R}^n$ stetig differenzierbar. Ist F ein Gradientenfeld, so gilt*

$$\partial_i F_j(x) = \partial_j F_i(x) \tag{14.7}$$

für alle $x \in \Omega$ und alle i, j mit $1 \le i, j \le n$.

Beweis: Ist $F = \operatorname{grad} f$, so gilt

$$\partial_i F_j(x) = \partial_i(\partial_j f)(x) = \partial_j(\partial_i f)(x) = \partial_j F_i(x) \,.$$

□

Definition 14.4 *Sei X Vektorraum, sei $Y \subset X$. Y heißt sternförmig, falls es ein $y \in Y$ gibt mit $[x, y] \subset Y$ für alle $x \in Y$.* □

Satz 14.5 *Sei $\Omega \subset \mathbb{R}^n$ offen und sternförmig, sei $F : \Omega \to \mathbb{R}^n$ stetig differenzierbar. Gilt*

$$\partial_i F_j(x) = \partial_j F_i(x) \tag{14.8}$$

für alle $x \in \Omega$ und alle i, j, $1 \leq i, j \leq n$, so ist F ein Gradientenfeld, also gibt es $f \in C^2(\Omega)$ mit

$$F = \operatorname{grad} f \,. \tag{14.9}$$

Beweis: Sei $y \in \Omega$ mit $[y, x] \subset \Omega$ für alle $x \in \Omega$. Wir definieren f als das Kurvenintegral von F entlang der Strecke von y nach x, wir setzen also

$$f(x) = \int_0^1 \langle F(y + t(x - y)), x - y \rangle \, dt \,.$$

Nach Satz 13.3 existieren alle partiellen Ableitungen $\partial_i f$ und sind stetig. Für

$$\tilde{f}(x) = \langle F(y + t(x - y)), x - y \rangle$$

gilt

$$\partial_i \tilde{f}(x) = \langle t \partial_i F(y + t(x - y)), x - y \rangle + \langle F(y + t(x - y)), e_i \rangle \,,$$

also

$$\partial_i f(x) = \int_0^1 \langle t \partial_i F(y + t(x - y)), x - y \rangle + F_i(y + t(x - y)) \, dt \,.$$

Für

$$g_i(t) = t F_i(y + t(x - y))$$

gilt

$$g_i'(t) = \langle t (\operatorname{grad} F_i)(y + t(x - y)), x - y \rangle + F_i(y + t(x - y)) \,,$$

und mit $z = y + t(x - y)$

$$\begin{aligned} \langle (\operatorname{grad} F_i)(z), x - y \rangle &= \sum_{j=1}^n \partial_j F_i(z)(x_j - y_j) = \sum_{j=1}^n \partial_i F_j(z)(x_j - y_j) \\ &= \langle \partial_i F(z), x - y \rangle \,. \end{aligned}$$

Es folgt

$$g_i'(t) = \langle t(\partial_i F)(y + t(x - y)), x - y \rangle + F_i(y + t(x - y)) \,,$$

und damit

$$\partial_i f(x) = \int_0^1 g_i'(t) \, dt = g_i(1) - g_i(0) = F_i(x) \,.$$

□

Folgerung 14.6 *Sei $\Omega \subset \mathbb{R}^3$ offen und sternförmig, sei $F : \Omega \to \mathbb{R}^3$ stetig differenzierbar. Dann ist F ein Gradientenfeld auf Ω genau dann, wenn* $\operatorname{rot} F = 0$.

Definition 14.7 (Wegzusammenhang)
Ein metrischer Raum (X, d) heißt wegzusammenhängend, wenn es für alle $x_a, x_b \in X$ eine Kurve $r : [0, 1] \to X$ gibt mit $r(0) = x_a$ und $r(1) = x_b$. □

Definition 14.8 (Konservatives Vektorfeld)
Sei $\Omega \subset \mathbb{R}^n$ offen. Ein Vektorfeld $F : \Omega \to \mathbb{R}^n$ heißt konservativ auf Ω, falls "das Kurvenintegral wegunabhängig ist", d.h. falls für alle Punkte $x_a, x_b \in \Omega$ und alle stückweise C^1-Kurven C_1, C_2 von x_a nach x_b, welche in Ω verlaufen, gilt

$$\int_{C_1} F \cdot dx = \int_{C_2} F \cdot dx \,. \tag{14.10}$$

□

Satz 14.9 *Sei $\Omega \subset \mathbb{R}^n$ offen und wegzusammenhängend, sei $F : \Omega \to \mathbb{R}^n$ stetig. Dann gilt*

$$F \text{ ist Gradientenfeld auf } \Omega \qquad \Leftrightarrow \qquad F \text{ ist konservativ auf } \Omega \;.$$

Beweis: "$\Rightarrow$": Folgt aus Satz 14.2.
"$\Leftarrow$": Seien $x, y \in \Omega$ beliebig. Wir beweisen zunächst

$$\text{Es gibt eine stückweise } C^1\text{-Kurve von } y \text{ nach } x. \tag{14.11}$$

Sei $r : [0,1] \to \Omega$ stetig mit $r(0) = y$, $r(1) = x$. Sei

$$U_\varepsilon = \{z : z \in \mathbb{R}^n, \, \mathrm{dist}\,(z, r([0,1])) < \varepsilon\} \,.$$

Wir wählen $\varepsilon > 0$ so, dass $U_\varepsilon \subset \Omega$ gilt (das ist möglich, da $\mathrm{dist}\,(\partial\Omega, r([0,1]) > 0$ wegen Satz 5.15, falls $\partial\Omega \neq \emptyset$, was gleichbedeutend ist mit $\Omega \neq \mathbb{R}^n$). Für eine hinreichend feine Unterteilung (t_i) von $[0,1]$ gilt, daß der Polygonzug, welcher $r(0), r(t_1), \ldots, r(1)$ verbindet, ganz in U_ε und damit auch in Ω liegt. Damit ist (14.11) bewiesen. Wir wählen nun $x_* \in \Omega$ und definieren $f : \Omega \to \mathbb{R}$ durch

$$f(x) = \int_C F \cdot \, dx \,, \tag{14.12}$$

wobei C eine stückweise C^1-Kurve von x_* nach x ist (eine solche gibt es nach (14.11), und nach Voraussetzung hängt das Kurvenintegral nicht von der Wahl von C ab). Zu zeigen ist noch $F = \mathrm{grad}\, f$. Sei $x \in \Omega$ beliebig, sei $h > 0$ so gewählt, dass $[x, x + he_i] \subset \Omega$ gilt. Sei C_* eine stückweise C^1-Kurve von x_* nach x, sei $C(h)$ definiert durch

$$r_h : [0, h] \to \Omega \,, \quad r_h(t) = x + te_i \,.$$

Dann gilt

$$f(x + he_i) = \int_{C_*} F \cdot \, dx + \int_{C(h)} F \cdot \, dx = f(x) + \int_{C(h)} F \cdot \, dx \,.$$

Für

$$g(h) = f(x + he_i) - f(x)$$

gilt also

$$g(h) = \int_{C(h)} F \cdot \, dx = \int_0^h \langle F(r_h(t)), r_h'(t) \rangle \; dt = \int_0^h F_i(x + te_i) \, dt \,,$$

also ist g rechtsseitig differenzierbar in 0 und $g'_+(0) = F_i(x)$. Dasselbe Argument mit $-e_i$ statt e_i liefert die Existenz von $\partial_i f(x)$ und die Formel $\partial_i f(x) = F_i(x)$. □

Als Beispiel betrachten wir

$$\Omega = \mathbb{R}^2 \setminus \{0\}\,, \quad F : \Omega \to \mathbb{R}\,, \quad F(x,y) = \left(-\frac{y}{x^2+y^2}\,, \frac{x}{x^2+y^2} \right). \tag{14.13}$$

Es gilt für alle $(x,y) \neq (0,0)$

$$\partial_1 F_2(x,y) = \partial_2 F_1(x,y)\,.$$

Für jede sternförmige Teilmenge $\tilde{\Omega}$ von Ω gibt es also nach Satz 14.5 ein $f \in C^2(\tilde{\Omega})$ mit

$$F(x) = \operatorname{grad} f(x)\,, \quad \text{für alle } x \in \tilde{\Omega}\,.$$

Aber andererseits gilt, wenn wir den Einheitskreis als eine geschlossene Kurve C auffassen, mit $r : [0, 2\pi] \to \mathbb{R}^2$, $r(t) = (\cos t, \sin t)$

$$\int_C F \cdot\, dx = \int_0^{2\pi} \langle F(r(t)), r'(t) \rangle \; dt = \int_0^{2\pi} \sin^2 t + \cos^2 t \, dt = 2\pi\,,$$

also ist F nicht konservativ auf Ω. Aus Satz 14.9 (bzw. schon Satz 14.2) folgt, dass es kein $f \in C^2(\Omega)$ geben kann mit

$$F(x) = \operatorname{grad} f(x)\,, \quad \text{für alle } x \in \Omega\,.$$

15 Kompaktheit und $C(X)$

Wir charakterisieren kompakte Mengen in metrischen Räumen.

Satz 15.1 *Sei (X,d) metrischer Raum. Dann sind äquivalent:*

(i) X ist kompakt, d.h. jede Folge in X hat eine konvergente Teilfolge.

(ii) X ist vollständig, und zu jedem $\varepsilon > 0$ gibt es endlich viele ε-Kugeln, welche X überdecken, d.h.

$$\forall \varepsilon > 0 \, \exists n \in \mathbb{N} \, \exists x_1, \ldots, x_n \in X \quad \textit{mit} \quad X = \bigcup_{i=1}^{n} B(x_i, \varepsilon)\,. \tag{15.1}$$

(iii) ("endliche Überdeckungseigenschaft"): Ist $(U_i)_{i \in I}$ eine Familie von offenen Mengen in X mit

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

so gibt es ein $n \in \mathbb{N}$ und Indizes $i_1, \ldots, i_n \in I$ mit

$$X = \bigcup_{k=1}^{n} U_{i_k}\,.$$

("Jede Überdeckung hat eine endliche Teilüberdeckung.")

(iv) ("endliche Durchschnittseigenschaft"): Ist $(A_i)_{i \in I}$ eine Familie von abgeschlossenen Mengen in X mit der Eigenschaft, dass

$$\bigcap_{k=1}^{n} A_{i_k} \neq \emptyset$$

gilt für alle $n \in \mathbb{N}$ und alle $i_1, \ldots, i_n \in I$, so gilt

$$\bigcap_{i \in I} A_i \neq \emptyset\,.$$

Beweis: "(i)⇒(ii)": Sei $(x_k)_{k \in \mathbb{N}}$ Cauchyfolge in X. Wir wählen eine konvergente Teilfolge mit $x_{k_m} \to x \in X$. Sei $\varepsilon > 0$. Wir wählen $N \in \mathbb{N}$ mit

$$d(x_l, x_n) < \frac{\varepsilon}{2}\,, \quad \text{für alle } l, n \geq N,$$

und $m \in \mathbb{N}$ mit $k_m \geq N$ und

$$d(x_{k_m}, x) < \frac{\varepsilon}{2}\,,$$

dann gilt für alle $n \geq N$

$$d(x_n, x) \leq d(x_n, x_{k_m}) + d(x_{k_m}, x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Also ist X vollständig. Die Existenz endlich vieler überdeckender Kugeln ist bereits in einer Übungsaufgabe bewiesen worden.
"(ii)$\Rightarrow$(iii)": Wir zeigen, dass aus (ii) und der Negation von (iii) ein Widerspruch folgt. Sei $(U_i)_{i\in I}$ Familie offener Mengen mit

$$\bigcup_{i\in I} U_i = X\,, \quad \text{aber} \quad \bigcup_{k=1}^{n} U_{i_k} \neq X \tag{15.2}$$

für alle n und alle $i_1, \ldots, i_n \in I$. Wir konstruieren zunächst mit vollständiger Induktion Kugeln $B_n = B(x_n, 2^{-n})$ mit $B_n \cap B_{n-1} \neq \emptyset$, welche nicht von endlich vielen U_i überdeckt werden: Für $n = 1$ wähle endlich viele Kugeln mit Radius $1/2$, welche X überdecken. Sei $B_1 = B(x_1, 1/2)$ eine davon, welche nicht von endlich vielen U_i überdeckt werden kann (eine solche muss es geben, sonst gilt (15.2) nicht). Für den Induktionsschritt $n-1 \to n$ gehen wir von B_{n-1} aus und wählen endlich viele Kugeln mit Radius 2^{-n}, welche B_{n-1} überdecken und mit B_{n-1} einen nichtleeren Durchschnitt haben. Sei $B_n = B(x_n, 2^{-n})$ eine davon, welche nicht von endlich vielen U_i überdeckt werden kann (eine solche muss es geben nach Konstruktion von B_{n-1}). Die Folge $(x_n)_{n\in\mathbb{N}}$ ist Cauchyfolge, da

$$d(x_n, x_{n+1}) < 2^{-n} + 2^{-(n+1)} < 2^{-(n-1)}$$

für alle n, also

$$d(x_n, x_{n+p}) < 2^{-(n-2)}$$

für alle n, p mit $p \geq n$. Sei $x = \lim x_n$. Dann gibt es ein $i \in I$ mit $x \in U_i$. Da U_i offen ist, gibt es ein $\varepsilon > 0$ mit $x \in B(x, \varepsilon) \subset U_i$. Für hinreichend großes n gilt dann

$$B_n = B(x_n, 2^{-n}) \subset B(x, \varepsilon) \subset U_i$$

im Widerspruch zur Konstruktion von B_n.
"(iii)$\Rightarrow$(iv)": Sei $(A_i)_{i\in I}$ Familie abgeschlossener Mengen mit

$$\bigcap_{k=1}^{n} A_{i_k} \neq \emptyset$$

für je endlich viele Indizes $i_1, \ldots, i_n$. Sei $U_i = X \setminus A_i$. Dann ist

$$\bigcup_{k=1}^{n} U_{i_k} \neq X$$

für je endlich viele Indizes $i_1, \ldots, i_n$. Aus (iii) folgt $\cup_{i\in I} U_i \neq X$, also $\cap_{i\in I} A_i \neq \emptyset$.
"(iv)$\Rightarrow$(i)": Sei $(x_k)_{k\in\mathbb{N}}$ Folge in X. Wir setzen

$$A_i = \overline{\{x_i, x_{i+1}, \ldots\}} = \overline{\{x_k : k \geq i\}}\,, \quad i \in \mathbb{N}\,.$$

Dann gilt für alle endlichen Indexkombinationen

$$\bigcap_{m=1}^{n} A_{i_m} = A_j \neq \emptyset\,, \quad j = \max_{1\leq m\leq n} i_m\,,$$

also gibt es

$$x \in \bigcap_{i=1}^{\infty} A_i \,.$$

Wir konstruieren eine Teilfolge (x_{k_m}) mit

$$d(x_{k_m}, x) < \frac{1}{m} \,.$$

Für $m = 1$ wählen wir

$$x_{k_1} \in B(x, 1) \cap \{x_i : i \in \mathbb{N}\}$$

(der Schnitt ist nichtleer, da $x \in A_1$). Ist $x_{k_{m-1}}$ konstruiert, so wählen wir

$$x_{k_m} \in B(x, \frac{1}{m}) \cap \{x_i : i > k_{m-1}\}$$

(der Schnitt ist nichtleer, da $x \in A_{k_{m-1}+1}$). □

In allgemeinen topologischen Räumen wird der Begriff "kompakt" durch die endliche Überdeckungseigenschaft (iii) in Satz 15.1 definiert. Die endliche Überdeckungseigenschaft ist dort i.a. nicht mehr äquivalent damit, dass jede Folge eine konvergente Teilfolge hat.

Wir wollen kompakte Mengen im Raum der stetigen Funktionen charakterisieren.

Definition 15.2 (Gleichgradige Stetigkeit)
Sei (X, d) metrischer Raum. Eine Teilmenge F von $C(X)$ heißt gleichgradig stetig, falls für alle $x \in X$ und alle $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass für alle $y \in X$ gilt

$$d(x, y) < \delta \qquad \Rightarrow \qquad |f(x) - f(y)| < \varepsilon \quad \textit{für alle } f \in F.$$

("Das δ kann für alle $f \in F$ gemeinsam gewählt werden".)

Lemma 15.3 *Sei $\Omega \subset \mathbb{R}^n$, $F \subset C(\Omega)$, es gebe ein L mit*

$$|f(x) - f(y)| \le L\|x - y\| \,, \quad \textit{für alle } x, y \in \Omega, \ f \in F. \tag{15.3}$$

Dann ist F gleichgradig stetig.

Beweis: Mit $\delta = \varepsilon/L$ hat F die in Definition 15.2 verlangte Eigenschaft. □

Ist $F \subset C^1(\Omega)$, so folgt aus dem Mittelwertsatz, dass (15.3) gilt, falls es ein $C > 0$ gibt mit

$$|\partial_i f(x)| \le C \,, \quad \text{für alle } x \in \Omega, \ f \in F, \ i = 1, \ldots, n. \tag{15.4}$$

Satz 15.4 (Arzela-Ascoli)
Sei (X, d) kompakter metrischer Raum, sei $F \subset (C(X), \|\cdot\|_\infty)$. Dann sind äquivalent:

(i) F ist relativ kompakt (d.h. $\overline{F}$ ist kompakt).

(ii) F ist beschränkt in $(C(X), \|\cdot\|_\infty)$ und gleichgradig stetig.

Beweis: "(i)⇒(ii)": Sei $\overline{F}$ kompakt. Nach Satz 15.1(ii) gibt es endlich viele $f_1, \ldots, f_n \in C(X)$ mit

$$\overline{F} \subset \bigcup_{i=1}^{n} B(f_i, 1)\,.$$

Da eine endliche Vereinigung von Kugeln beschränkt ist, ist auch F beschränkt. Wir zeigen nun, dass F gleichgradig stetig ist. Seien $x \in X$, $\varepsilon > 0$ beliebig. Wir wählen endlich viele $f_1, \ldots, f_n \in C(X)$ mit

$$\overline{F} \subset \bigcup_{i=1}^{n} B(f_i, \varepsilon)\,.$$

Wir wählen $\delta_i > 0$, so dass für alle $y \in X$ gilt

$$d(y, x) < \delta_i \qquad \Rightarrow \qquad |f_i(y) - f_i(x)| < \varepsilon\,,$$

und setzen $\delta = \min_i \delta_i$. Sei nun $f \in F$. Wir wählen ein k mit $\|f - f_k\|_\infty < \varepsilon$, dann gilt für alle $y \in X$ mit $d(y, x) < \delta$

$$|f(y) - f(x)| \le |f(y) - f_k(y)| + |f_k(y) - f_k(x)| + |f_k(x) - f(x)| < \varepsilon + \varepsilon + \varepsilon = 3\varepsilon\,.$$

Damit ist gezeigt, dass F gleichgradig stetig ist.
"(ii)⇒(i)" Zunächst ist $\overline{F}$ vollständig, da $C(X)$ vollständig und $\overline{F}$ abgeschlossen in $C(X)$ ist. Nach Satz 15.1 genügt es zu zeigen:

> Für alle $\varepsilon > 0$ lässt sich $\overline{F}$ durch endlich viele Teilmengen von $C(X)$ mit Durchmesser $\le \varepsilon$ überdecken.

Wir beweisen zuerst, dass auch $\overline{F}$ gleichgradig stetig ist. Sei $x \in X$, $\varepsilon > 0$. Wir wählen ein $\delta > 0$ mit

$$d(y, x) < \delta \qquad \Rightarrow \qquad |f(y) - f(x)| < \varepsilon \quad \text{für alle } f \in F.$$

Sei $g \in F$. Wir wählen $f \in F$ mit $\|g - f\|_\infty < \varepsilon$, dann gilt für alle $y \in X$ mit $d(y, x) < \delta$

$$|g(y) - g(x)| \le |g(y) - f(y)| + |f(y) - f(x)| + |f(x) - g(x)| < \varepsilon + \varepsilon + \varepsilon = 3\varepsilon\,,$$

also ist $\overline{F}$ gleichgradig stetig. Wir konstruieren nun die endliche Überdeckung. Sei wiederum $\varepsilon > 0$ gegeben. Zu jedem $x \in X$ wählen wir ein $\delta(x) > 0$ mit

$$d(y, x) < \delta(x) \qquad \Rightarrow \qquad |f(y) - f(x)| < \frac{\varepsilon}{4} \quad \text{für alle } f \in \overline{F}.$$

Wir wählen (X ist kompakt) endlich viele $x_i \in X$, $1 \le i \le k$, mit

$$X = \bigcup_{i=1}^{k} B(x_i, \delta(x_i))\,.$$

Sei $C > 0$ mit $\|f\|_\infty \le C$ für alle $f \in \overline{F}$ (F ist beschränkt). Sei

$$-C = c_1 < c_2 < \ldots < c_m = C$$

eine Zerlegung des möglichen Wertebereichs von Funktionen in $\overline{F}$ mit $c_{j+1} - c_j < \frac{\varepsilon}{2}$ für alle j, $1 < j \leq m$. Sei

$$\Phi = \{\varphi | \, \varphi : \{1, \ldots, k\} \to \{1, \ldots, m\}\} .$$

Für $\varphi \in \Phi$ definieren wir

$$L_\varphi = \{f : f \in \overline{F}, \, |f(x_i) - c_{\varphi(i)}| < \frac{\varepsilon}{4} \text{ für alle } i\} .$$

Dann ist

$$\overline{F} = \bigcup_{\varphi \in \Phi} L_\varphi .$$

Ist nämlich $f \in \overline{F}$, so ist $f \in L_\varphi$, wenn wir φ so wählen, dass

$$|f(x_i) - c_{\varphi(i)}| < \frac{\varepsilon}{4} .$$

Da Φ eine endliche Menge ist, genügt es also zu zeigen, dass $\operatorname{diam}(L_\varphi) \leq \varepsilon$ für alle φ. Seien $f, g \in L_\varphi$, sei $x \in X$. Wir wählen ein i, $1 \leq i \leq k$, mit $x \in B(x_i, \delta(x_i))$, dann gilt

$$\begin{aligned} |f(x) - g(x)| &\leq |f(x) - f(x_i)| + |f(x_i) - c_{\varphi(i)}| + |c_{\varphi(i)} - g(x_i)| + |g(x_i) - g(x)| \\ &< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon , \end{aligned}$$

also ist $\|f - g\|_\infty \leq \varepsilon$. □

Satz 15.5 (Satz von Dini)
Sei (X, d) kompakter metrischer Raum, sei $(f_n)_{n\in\mathbb{N}}$ Folge in $C(X)$, sei $f \in C(X)$. Die Folge (f_n) sei wachsend, d.h.

$$f_n(x) \leq f_{n+1}(x) , \quad \text{für alle } n \in \mathbb{N}, \, x \in X, \tag{15.5}$$

und f_n konvergiere punktweise gegen f,

$$\lim_{n\to\infty} f_n(x) = f(x) , \quad \text{für alle } x \in X. \tag{15.6}$$

Dann konvergiert f_n gleichmäßig gegen f.

Beweis: Sei $\varepsilon > 0$. Sei $x \in X$. Wir wählen $N(x) \in \mathbb{N}$ so, dass

$$0 \leq f(x) - f_n(x) < \frac{\varepsilon}{3} , \quad \text{für alle } n \geq N(x).$$

Wir wählen $\delta(x) > 0$ so, dass für alle $y \in B(x, \delta(x))$ gilt

$$|f(y) - f(x)| < \frac{\varepsilon}{3} , \quad |f_{N(x)}(y) - f_{N(x)}(x)| < \frac{\varepsilon}{3} .$$

Mit der Dreiecksungleichung folgt

$$0 \leq f(y) - f_{N(x)}(y) < \varepsilon , \quad \text{für alle } y \in B(x, \delta(x)).$$

Es ist

$$X = \bigcup_{x \in X} B(x, \delta(x))\,.$$

Da X kompakt ist, gibt es $k \in \mathbb{N}$, $x_1, \ldots, x_k \in X$ mit

$$X = \bigcup_{i=1}^{k} B(x_i, \delta(x_i))\,.$$

Für alle $n \geq \max_{1 \leq i \leq k} N(x_i)$ folgt

$$0 \leq f(y) - f_n(y) < \varepsilon\,, \quad \text{für alle } y \in X,$$

also $f_n \to f$ gleichmäßig. □

Definition 15.6 *Sei (X, d) metrischer Raum, sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Eine Teilmenge A von $C(X; \mathbb{K})$ heißt Unteralgebra von $C(X; \mathbb{K})$, falls gilt*

$$f, g \in A \qquad \Rightarrow f + g \in A\,, \quad f \cdot g \in A\,, \tag{15.7}$$

$$f \in A\,, \quad \lambda \in \mathbb{K} \qquad \Rightarrow \qquad \lambda f \in \mathbb{K}\,. \tag{15.8}$$

Wir sagen, dass A die Punkte von X trennt, falls es für alle $x, y \in X$ mit $x \neq y$ ein $f \in A$ gibt mit $f(x) \neq f(y)$.

Für $X \subset \mathbb{K}^n$ bildet die Menge A aller Polynome, eingeschränkt auf X, eine Unteralgebra von $C(X; \mathbb{K})$, welche die Punkte von X trennt. Dasselbe gilt für $X = [0, 2\pi]$, wenn A die Menge aller trigonometrischen Polynome ist, d.h. die Menge aller Funktionen f der Form

$$f(x) = \sum_{k=-n}^{n} a_k e^{ikx}\,,$$

wobei $n \in \mathbb{N}$ und $a_k \in \mathbb{C}$ beliebig sind (hier ist $\mathbb{K} = \mathbb{C}$).

Satz 15.7 (Stone-Weierstraß)
Sei (X, d) kompakter metrischer Raum. Ist A eine Unteralgebra von $C(X; \mathbb{R})$, welche die konstanten Funktionen enthält und die Punkte von X trennt, so gilt

$$\overline{A} = C(X; \mathbb{R})\,. \tag{15.9}$$

(D.h.: Für alle $f \in C(X; \mathbb{R})$ gibt es eine Folge (f_n) in A mit $f_n \to f$ gleichmäßig.)

Beweis: Der erste Teil des Beweises besteht darin, eine monoton wachsende Folge p_n von reellen Polynomen zu konstruieren, welche auf $[0, 1]$ gleichmäßig gegen $s(x) = \sqrt{x}$ konvergiert. Wir setzen $p_0 = 0$ und definieren

$$p_{n+1}(x) = p_n(x) + \frac{1}{2}(x - p_n(x)^2)\,. \tag{15.10}$$

Wir zeigen mit Induktion, dass

$$p_n(x) \leq \sqrt{x}\,, \quad \text{für alle } x \in [0, 1]. \tag{15.11}$$

Offensichtlich gilt (15.11) für $n = 0$. Induktionsschluss $n \to n+1$: Es ist

$$\begin{aligned}\sqrt{x} - p_{n+1}(x) &= \sqrt{x} - p_n(x) - \frac{1}{2}(x - p_n(x)^2) \\ &= (\sqrt{x} - p_n(x))\left(1 - \frac{1}{2}(\sqrt{x} + p_n(x))\right),\end{aligned}$$

und aus der Induktionsannahme folgt

$$\frac{1}{2}(\sqrt{x} + p_n(x)) \le \sqrt{x} \le 1\,, \quad x \in [0,1]\,,$$

und damit (15.11) für $n+1$. Aus (15.10) und (15.11) folgt nun

$$p_{n+1}(x) \ge p_n(x)$$

für alle $x \in [0,1]$ und alle $n \in \mathbb{N}$. Für alle x ist $(p_n(x))$ eine monoton wachsende beschränkte (durch $\sqrt{x}$) Folge, also existiert

$$p(x) = \lim_{n\to\infty} p_n(x)\,.$$

Grenzübergang $n \to \infty$ in (15.10) liefert nun $p(x) = \sqrt{x}$. Aus dem Satz von Dini folgt, dass $p_n \to p$ gleichmäßig in $[0,1]$. Im zweiten Teil des Beweises zeigen wir, dass A "hinreichend viele" Funktionen f enthält, so dass (15.9) folgt. Zunächst behaupten wir, dass gilt

$$f \in A \qquad \Rightarrow \qquad |f| \in \overline{A}\,. \tag{15.12}$$

Für $f \in A$, $f \ne 0$, setzen wir mit p_n aus (15.10)

$$f_n(x) = p_n\left(\frac{f(x)^2}{\|f\|_\infty^2}\right),$$

dann ist $f_n \in A$ (da A Unteralgebra), und aus dem ersten Teil des Beweises folgt

$$f_n(x) \to \sqrt{\frac{f(x)^2}{\|f\|_\infty^2}} = \frac{|f(x)|}{\|f\|_\infty}$$

gleichmäßig in $[0,1]$, also $\|f\|_\infty \cdot f_n \to |f|$ gleichmäßig und damit $|f| \in \overline{A}$. Da aus $f_n \to f$ gleichmäßig auch $|f_n| \to |f|$ gleichmäßig folgt, gilt weiter

$$f \in \overline{A} \qquad \Rightarrow \qquad |f| \in \overline{A}\,. \tag{15.13}$$

$\overline{A}$ ist ebenfalls eine Unteralgebra: Sind etwa $f, g \in \overline{A}$, so gibt es Folgen (f_n), (g_n) in A mit $f_n \to f$ und $g_n \to g$ gleichmäßig; wegen $f_n + g_n \in A$ und $f_n + g_n \to f + g$ gleichmäßig folgt $f + g \in \overline{A}$. Analog für Multiplikation und Skalarmultiplikation. Aus dem bereits Bewiesenen folgt weiter

$$f, g \in \overline{A} \qquad \Rightarrow \qquad \max\{f,g\} \in \overline{A}\,, \quad \min\{f,g\} \in \overline{A}\,, \tag{15.14}$$

da $\max\{f,g\} = \frac{1}{2}(f+g+|f-g|)$, analog für min. Zu beliebig vorgegebenen $x, y \in X$ und $\alpha, \beta \in \mathbb{R}$ gibt es ein $f \in A$ mit $f(x) = \alpha$ und $f(y) = \beta$. Ein solches f erhalten wir etwa, indem wir setzen

$$f(\xi) = \alpha + (\beta - \alpha)\frac{g(\xi) - g(x)}{g(y) - g(x)}\,, \quad \xi \in X\,,$$

wobei $g \in A$ so gewählt ist, dass $g(x) \neq g(y)$. Weiterhin können wir zu jedem $f \in C(X)$, $x \in X$ und $\varepsilon > 0$ ein $g \in \overline{A}$ finden mit

$$g(x) = f(x)\,, \quad g(y) \leq f(y) + \varepsilon \text{ für alle } y \in X. \tag{15.15}$$

Ein solches g finden wir, indem wir zu beliebigem $z \in X$ zunächst ein $h_z \in A$ wählen mit

$$h_z(x) = f(x)\,, \quad h_z(z) < f(z) + \frac{\varepsilon}{2}\,.$$

Sei nun $\delta(z) > 0$ so, dass

$$h_z(y) < f(y) + \varepsilon\,, \quad \text{für alle } y \in B(z, \delta(z)),$$

dann gibt es, da X kompakt ist, endlich viele $z_1, \ldots, z_k$ mit

$$X = \bigcup_{i=1}^{k} B(z_i, \delta(z_i))\,.$$

Wir setzen

$$g = \min\{h_{z_1}, \ldots, h_{z_k}\}\,,$$

dann ist $g \in \overline{A}$, und nach Konstruktion folgt (15.15). Wir zeigen nun die Behauptung $\overline{A} = C(X;\mathbb{R})$. Sei $f \in C(X;\mathbb{R})$ und $\varepsilon > 0$. Gemäß (15.15) wählen wir zu jedem $x \in X$ ein $g_x \in \overline{A}$ mit

$$g_x(x) = f(x)\,, \quad g_x(y) \leq f(y) + \varepsilon \text{ für alle } y \in X.$$

Sei $\tilde{\delta}(x) > 0$ so, dass

$$g_x(y) \geq f(y) - \varepsilon\,, \quad \text{für alle } y \in B(x, \tilde{\delta}(x)\,.$$

Wir wählen endlich viele $x_1, \ldots, x_m$ mit

$$X = \bigcup_{i=1}^{m} B(x_i, \tilde{\delta}(x_i))\,,$$

und setzen

$$g = \min\{g_{x_1}, \ldots, g_{x_m}\}\,.$$

Dann ist $g \in \overline{A}$, und

$$f(y) - \varepsilon \leq g(y) \leq f(y) + \varepsilon\,, \quad \text{für alle } y \in X.$$

Also ist $f \in \overline{A}$. $\square$

Folgerung 15.8 *Sei $K \subset \mathbb{R}^n$ kompakt, sei $f : K \to \mathbb{R}$ stetig. Dann gibt es zu jedem $\varepsilon > 0$ ein Polynom p mit*

$$|f(x) - p(x)| \le \varepsilon\,, \quad \textit{für alle } x \in K. \tag{15.16}$$

□

Es gilt auch eine komplexe Version des Satzes von Stone-Weierstraß.

Satz 15.9 *Sei (X,d) kompakter metrischer Raum, sei A eine Unteralgebra von $C(X;\mathbb{C})$, welche die Punkte von X trennt und alle konstanten Funktionen enthält, und für welche außerdem gilt*

$$f \in A \qquad \Rightarrow \qquad \overline{f} \in A\,.$$

($\overline{f}$ bezeichnet die zu f komplex konjugierte Funktion.) Dann gilt

$$\overline{A} = C(X;\mathbb{C})\,. \tag{15.17}$$

Beweis: Sei

$$A_0 = \{f : f \in A\,, f(x) \in \mathbb{R} \text{ für alle } x \in X\}\,.$$

Dann ist A_0 eine Unteralgebra von $C(X;\mathbb{R})$, welche alle (reellen) Konstanten enthält. Außerdem gilt für jedes $f \in A$, dass

$$\operatorname{Re} f = \frac{1}{2}(f + \overline{f}) \in A_0\,, \quad \operatorname{Im} f = \frac{1}{2i}(f - \overline{f}) \in A_0\,.$$

Trennt $f \in A$ die Punkte $x, y \in X$, also $f(x) \ne f(y)$, so ist entweder $(\operatorname{Re} f)(x) \ne (\operatorname{Re} f)(y)$ oder $(\operatorname{Im} f)(x) \ne (\operatorname{Im} f)(y)$, also trennt A_0 ebenfalls die Punkte von X. Aus Satz 15.7 folgt nun, dass

$$\overline{A_0} = C(X;\mathbb{R})\,.$$

Sei nun $f \in C(X;\mathbb{C})$. Wir wählen Folgen (g_n), (h_n) in A_0 mit $g_n \to \operatorname{Re} f$ und $h_n \to \operatorname{Im} f$ gleichmäßig, dann gilt

$$A \ni f_n = g_n + ih_n \to f$$

gleichmäßig, also $f \in \overline{A}$. □