

# Analysis II

Walter Bergweiler

Sommersemester 2012

Fassung vom 22. Juni 2012

Diese Vorlesung ist eine Fortsetzung der Vorlesung Analysis I aus dem Wintersemester 2011/2012. Die Nummerierung dieser Vorlesung wird hier fortgesetzt; Verweise wie “nach Satz  $x.y.z$ ” mit  $x \leq 4$  beziehen sich darauf.

## Inhaltsverzeichnis

<b>5</b>	<b>Integration</b>	<b>1</b>
5.1	Das Riemann-Integral . . . . .	1
5.2	Der Hauptsatz der Differential- und Integralrechnung . . . . .	7
5.3	Integrationstechniken . . . . .	10
5.4	Integralform des Taylorrestglieds . . . . .	14
5.5	Folgen und Reihen integrierbarer Funktionen . . . . .	16
5.6	Fourierreihen . . . . .	17
5.7	Uneigentliche Integrale . . . . .	22
<b>6</b>	<b>Topologie metrischer Räume</b>	<b>26</b>
6.1	Metrik, Norm und Skalarprodukt . . . . .	26
6.2	Konvergenz . . . . .	29
6.3	Offene und abgeschlossene Mengen . . . . .	31
6.4	Stetigkeit . . . . .	33
6.5	Kompaktheit . . . . .	37
6.6	Zusammenhang . . . . .	41
<b>7</b>	<b>Differentialrechnung in mehreren Veränderlichen</b>	<b>43</b>
7.1	Partielle Ableitungen . . . . .	43
7.2	Differenzierbarkeit . . . . .	48
7.3	Lokale Extrema . . . . .	52
7.4	Höhere Ableitungen und Taylorformel . . . . .	58
7.5	Die Mittelwertungleichung . . . . .	59
7.6	Der Satz über implizite Funktionen . . . . .	61
7.7	Extrema unter Nebenbedingungen . . . . .	67



# 5 Integration

## 5.1 Das Riemann-Integral

Der Flächeninhalt eines Rechtecks ist das Produkt der Seitenlängen. Die Idee beim Riemann-Integral ist die Approximation allgemeinerer Flächen durch Rechtecke und die Definition des Flächeninhaltes durch einen Grenzübergang.

**Definition 5.1.1** Seien  $a, b \in \mathbb{R}$  mit  $a < b$ . Weiter seien  $n \in \mathbb{N}$  und  $x_0, \dots, x_n$  mit  $a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b$ . Dann heißt  $Z := \{x_0, x_1, \dots, x_n\}$  eine *Zerlegung* von  $[a, b]$ . Für  $k \in \{1, \dots, n\}$  heißt  $I_k := [x_{k-1}, x_k]$  das  $k$ -te Teilintervall der Zerlegung und  $|I_k| := x_k - x_{k-1}$  die *Länge* von  $I_k$ . Weiter heißt  $|Z| := \max_k |I_k|$  die *Feinheit* der Zerlegung  $Z$ .

Sei  $\xi = (\xi_1, \xi_2, \dots, \xi_n) \in I_1 \times I_2 \times \dots \times I_n$ , also  $\xi_k \in I_k$  für alle  $k$ . Dann heißen die  $\xi_k$  (oder auch das  $n$ -tupel  $\xi$ ) *Stützstellen* zur Zerlegung  $Z$ .

Sei zusätzlich noch  $f : [a, b] \rightarrow \mathbb{C}$ . Dann heißt

$$S(Z, \xi) := S(f, Z, \xi) := \sum_{k=1}^n f(\xi_k) |I_k|$$

*Riemannsche Summe* von  $f$  bzgl.  $Z$  und  $\xi$ .

Eine Zerlegung eines Intervalls  $[a, b]$  ist nach Definition also nichts anderes als eine endliche Teilmenge des Intervalls, die die Endpunkte enthält. Schreiben wir im folgenden aber eine Zerlegung  $Z$  von  $[a, b]$  in der Form  $Z = \{x_0, x_1, \dots, x_n\}$ , so werden wir immer  $a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b$  annehmen. Betrachten wir auf einem Intervall  $[a, b]$  definierte Funktionen, werden wir immer  $a < b$  voraussetzen, wenn nichts anderes gesagt wird.

Für die Zerlegung  $Z = \{1, 2, 2,5, 3,2, 4\}$  des Intervalls  $[1, 4]$ , geeignete Stützstellen  $\xi$  und die durch  $x \mapsto 2 + \sin x$  gegebene Funktion  $f$  sind in Abbildung 20 die Rechtecke dargestellt, deren Flächensumme  $S(f, Z, \xi)$  ist.

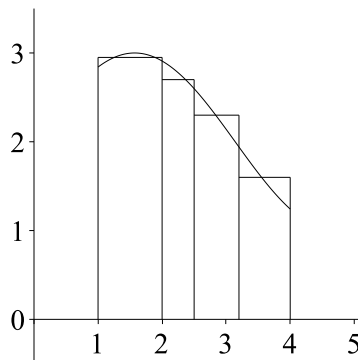


Abbildung 20: Eine Riemannsche Summe.

Die Idee ist nun, einen Grenzübergang  $|Z| \rightarrow 0$  durchzuführen.

**Definition 5.1.2** Eine Funktion  $f : [a, b] \rightarrow \mathbb{C}$  heißt *Riemann-integrierbar* (kurz: *integrierbar*), falls es ein  $S \in \mathbb{C}$  gibt, so dass für alle  $\varepsilon > 0$  ein  $\delta > 0$  mit folgender Eigenschaft existiert: Ist  $Z$  eine Zerlegung von  $[a, b]$  mit  $|Z| < \delta$  und sind  $\xi$  zugehörige Stützstellen, so gilt  $|S(f, Z, \xi) - S| < \varepsilon$ .

Existiert so eine Zahl  $S$ , so ist sie eindeutig und heißt *Riemann-Integral* (kurz: *Integral*) von  $f$  über  $[a, b]$  und wird mit  $\int_a^b f(x)dx$  oder  $\int_a^b f$  bezeichnet. Setzt man  $I := [a, b]$ , schreibt man auch  $\int_I f$ .

In Quantorenschreibweise lautet die Bedingung wie folgt:

$$\exists S \in \mathbb{C} \forall \varepsilon > 0 \exists \delta > 0 \forall \text{ Zerlegungen } Z \text{ von } [a, b] \forall \text{ Stützstellen } \xi \text{ zu } Z : \\ |Z| < \delta \Rightarrow |S(f, Z, \xi) - S| < \varepsilon.$$

Die Eindeutigkeit von  $S$  ist klar; man vgl. den Beweis, dass eine konvergente Zahlenfolge nur einen Grenzwert hat.

Man kann  $\int_a^b f$  nicht direkt als Grenzwert von  $S(Z, \xi)$  für  $|Z| \rightarrow 0$  definieren, da  $Z$  und  $\xi$  durch  $|Z|$  ja nicht eindeutig festgelegt sind. Dennoch übertragen sich viele Resultate über Grenzwerte ohne weiteres. Beispielsweise ist  $f$  genau dann integrierbar mit  $S = \int_a^b f$ , wenn für jede Folge  $(Z_n)$  von Zerlegungen mit  $|Z_n| \rightarrow 0$  und jede zugehörige Stützstellenfolge  $(\xi_n)$  gilt, dass  $S(Z_n, \xi_n) \rightarrow S$ .

Auch das Cauchy Kriterium gilt analog:

**Satz 5.1.1 (Cauchy Kriterium für Integrierbarkeit)** Sei  $f : [a, b] \rightarrow \mathbb{C}$  eine Funktion. Dann ist  $f$  genau dann integrierbar, wenn für alle  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass gilt: Sind  $Z_1$  und  $Z_2$  Zerlegungen von  $[a, b]$  mit  $|Z_1| < \delta$  und  $|Z_2| < \delta$  und sind  $\xi_1$  und  $\xi_2$  Stützstellen zu  $Z_1$  bzw.  $Z_2$ , so gilt  $|S(Z_1, \xi_1) - S(Z_2, \xi_2)| < \varepsilon$ .

In Quantorenschreibweise lautet die Bedingung wie folgt:

$$\forall \varepsilon > 0 \exists \delta > 0 \forall \text{ Zerlegungen } Z_1, Z_2 \text{ von } [a, b] \forall \text{ Stützstellen } \xi_1, \xi_2 \text{ zu } Z_1, Z_2 : \\ |Z_1| < \delta \text{ und } |Z_2| < \delta \Rightarrow |S(Z_1, \xi_1) - S(Z_2, \xi_2)| < \varepsilon.$$

Der Beweis ist analog zum Cauchy Kriterium für Funktionengrenzwerte (Satz 3.3.4).

**Satz 5.1.2** Jede stetige Funktion  $f : [a, b] \rightarrow \mathbb{C}$  ist integrierbar.

Zum Beweis benutzen wir folgenden

**Hilfssatz 5.1.1** Sei  $f : [a, b] \rightarrow \mathbb{C}$ . Seien  $\varepsilon, \delta > 0$  und es gelte  $|f(x) - f(y)| < \varepsilon$  für alle  $x, y \in [a, b]$  mit  $|x - y| < \delta$ . Seien weiter  $Z, Z'$  Zerlegungen mit Stützstellen  $\xi, \xi'$ . Ist dann  $Z \subset Z'$  und  $|Z| < \delta$ , so gilt  $|S(Z, \xi) - S(Z', \xi')| < \varepsilon(b - a)$ .

*Beweis.* Seien  $I_1, I_2, \dots, I_m$  die Teilintervalle von  $Z$  und  $I'_1, I'_2, \dots, I'_n$  die Teilintervalle von  $Z'$ . Dann gilt  $m \leq n$  und es gilt  $I_1 = I'_1 \cup I'_2 \cup \dots \cup I'_{\ell_1}$ ,  $I_2 = I'_{\ell_1+1} \cup \dots \cup I'_{\ell_2}$ ,  $\dots$ ,  $I_m = I'_{\ell_{m-1}+1} \cup \dots \cup I'_{\ell_m}$  mit  $0 < \ell_1 < \dots < \ell_m := n$ . Es folgt mit  $\ell_0 := 0$ , dass

$$|S(Z, \xi) - S(Z', \xi')| = \left| \sum_{k=1}^m f(\xi_k) |I_k| - \sum_{k=1}^n f(\xi'_k) |I'_k| \right|$$

$$\begin{aligned}
&= \left| \sum_{k=1}^m f(\xi_k) \sum_{j=\ell_{k-1}+1}^{\ell_k} |I'_j| - \sum_{k=1}^m \sum_{j=\ell_{k-1}+1}^{\ell_k} f(\xi'_j) |I'_j| \right| \\
&\leq \sum_{k=1}^m \sum_{j=\ell_{k-1}+1}^{\ell_k} |f(\xi_k) - f(\xi'_j)| |I'_j| \\
&< \varepsilon \sum_{k=1}^m \sum_{j=\ell_{k-1}+1}^{\ell_k} |I'_j| \\
&= \varepsilon(b-a).
\end{aligned}$$

*Beweis von Satz 5.1.2.* Sei  $\varepsilon > 0$ . Da  $f$  nach Satz 3.6.5 gleichmäßig stetig ist, existiert  $\delta > 0$  mit  $|f(x) - f(y)| < \varepsilon' := \frac{\varepsilon}{2(b-a)}$  für alle  $x, y \in [a, b]$  mit  $|x - y| < \delta$ .

Es seien nun  $Z_1, Z_2$  Zerlegungen von  $[a, b]$  mit Stützstellen  $\xi_1, \xi_2$ , so dass  $|Z_1| < \delta$  und  $|Z_2| < \delta$ . Für  $Z := Z_1 \cup Z_2$  und beliebige Stützstellen  $\xi$  zu  $Z$  folgt mit Hilfssatz 5.1.1 dann

$$|S(Z_1, \xi_1) - S(Z_2, \xi_2)| \leq |S(Z_1, \xi_1) - S(Z, \xi)| + |S(Z, \xi) - S(Z_2, \xi_2)| < 2\varepsilon'(b-a) = \varepsilon.$$

Die Behauptung folgt mit dem Cauchy Kriterium (Satz 5.1.1).  $\square$

*Beispiel.* Die Funktion  $f : [0, 1] \rightarrow [0, 1]$ ,  $x \mapsto x$ , ist stetig. Mit

$$Z_n = \left\{ 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1 \right\} \quad \text{und} \quad \xi_n = \left( \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1 \right)$$

gilt  $|Z_n| = 1/n \rightarrow 0$  und

$$S(Z_n, \xi_n) = \sum_{k=1}^n \frac{k}{n} \cdot \frac{1}{n} = \frac{1}{n^2} \sum_{k=1}^n k = \frac{1}{n^2} \frac{n(n+1)}{2} = \frac{1}{2} \left( 1 + \frac{1}{n} \right) \rightarrow \frac{1}{2},$$

also  $\int_0^1 x \, dx = \frac{1}{2}$ .

Die Umkehrung von Satz 5.1.2 gilt nicht, d.h., integrierbare Funktionen müssen nicht stetig sein. Zum Beispiel zeigt man leicht, dass beschränkte Funktionen mit endlich (oder sogar abzählbar unendlich) vielen Unstetigkeitsstellen integrierbar sind. Es gilt aber der folgende Satz.

**Satz 5.1.3** *Integrierbare Funktionen sind beschränkt.*

Wir skizzieren nur die *Beweisidee*: Ist  $f : [a, b] \rightarrow \mathbb{C}$  unbeschränkt, so existiert eine Folge  $(x_k)$  in  $[a, b]$  mit  $|f(x_k)| \rightarrow \infty$ . Sei nun  $Z$  eine Zerlegung mit Stützstellen  $\xi = (\xi_1, \dots, \xi_n)$ . Sei  $I_{j_k}$  das Teilintervall, welches  $x_k$  enthält. (Falls  $x_k \in Z$  ist, aber  $x_k$  kein Randpunkt von  $[a, b]$  ist, existieren zwei Teilintervalle mit dieser Eigenschaft. In diesem Falle sei  $I_{j_k}$  irgendeines dieser beiden.) Wir betrachten die Stützstellen  $\xi^{(k)}$ , die wir dadurch erhalten, dass wir  $\xi_{j_k}$  durch  $x_k$  ersetzen, also  $\xi^{(k)} = (\xi_1, \dots, \xi_{j_k-1}, x_k, \xi_{j_k+1}, \dots, \xi_n)$ . Dann folgt  $|S(Z, \xi^{(k)})| \rightarrow \infty$ .  $\square$

**Satz 5.1.4** *Seien  $f, g : [a, b] \rightarrow \mathbb{C}$  integrierbar und  $c \in \mathbb{C}$ . Dann sind auch  $f + g$ ,  $c \cdot f$  und  $\overline{f}$  integrierbar und es gilt*

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g, \quad \int_a^b c \cdot f = c \cdot \int_a^b f \quad \text{und} \quad \int_a^b \overline{f} = \overline{\int_a^b f}.$$

*Beweis.* Die erste Behauptung folgt aus

$$S(f + g, Z, \xi) = S(f, Z, \xi) + S(g, Z, \xi).$$

Die anderen erhält man analog.  $\square$

**Satz 5.1.5** *Eine Funktion  $f : [a, b] \rightarrow \mathbb{C}$  ist genau dann integrierbar, wenn  $\operatorname{Re}f$  und  $\operatorname{Im}f$  integrierbar sind. In diesem Fall gilt*

$$\int_a^b f = \int_a^b \operatorname{Re}f + i \int_a^b \operatorname{Im}f.$$

Der *Beweis* folgt unmittelbar aus Satz 5.1.4.

Man kann sich also auf die Integration reellwertiger Funktionen beschränken. Für beschränktes  $f : [a, b] \rightarrow \mathbb{R}$  und eine Zerlegung  $Z$  von  $I$  nennt man

$$\overline{S}(Z) := \overline{S}(f, Z) := \sup_{\xi} S(Z, \xi)$$

bzw.

$$\underline{S}(Z) := \underline{S}(f, Z) := \inf_{\xi} S(Z, \xi)$$

*Riemannsche Ober- bzw. Untersumme* von  $f$ . Dabei sind Supremum bzw. Infimum über alle Stützstellen  $\xi$  zur Zerlegung  $Z$  zu nehmen. Für  $Z \subset Z'$  gilt dann

$$\underline{S}(Z) \leq \underline{S}(Z') \leq \overline{S}(Z') \leq \overline{S}(Z).$$

Man nennt

$$\overline{\int_a^b} f := \inf_Z \overline{S}(Z)$$

bzw.

$$\underline{\int_a^b} f := \sup_Z \underline{S}(Z)$$

*oberes bzw. unteres Riemann-Integral* von  $f$ . Hier sind Supremum bzw. Infimum über alle Zerlegungen  $Z$  von  $I$  zu nehmen. Man beachte, dass oberes und unteres Riemann-Integral (für beschränktes  $f$ ) immer existieren.

**Satz 5.1.6 (Riemannsches Integrierbarkeitskriterium)** *Sei  $f : [a, b] \rightarrow \mathbb{R}$  beschränkt. Dann sind die folgenden Aussagen äquivalent:*

- (i)  $f$  ist integrierbar,
- (ii)  $\underline{\int_a^b} f = \overline{\int_a^b} f$ ,
- (iii) Für alle  $\varepsilon > 0$  existiert eine Zerlegung  $Z$  mit  $\overline{S}(Z) - \underline{S}(Z) < \varepsilon$ .

Im Falle der Integrierbarkeit gilt  $\int_a^b f = \underline{\int_a^b} f = \overline{\int_a^b} f$ .

Wir verzichten hier auf einen *Beweis*. (Zumindest teilweise wird dies aber in der Übung behandelt.)

**Satz 5.1.7** Seien  $f : [a, b] \rightarrow [c, d]$  integrierbar und  $\phi : [c, d] \rightarrow \mathbb{R}$  stetig. Dann ist  $\phi \circ f$  integrierbar.

*Beweis.* Sei  $\varepsilon > 0$ . Da  $\phi$  nach Satz 3.6.5 gleichmäßig stetig ist, existiert  $\delta > 0$  mit  $|\phi(x) - \phi(y)| < \varepsilon' := \frac{\varepsilon}{2(b-a)}$  für alle  $x, y \in [a, b]$  mit  $|x - y| < \delta$ . Wir setzen  $K := \max \phi(J) - \min \phi(J)$ . Man kann annehmen, dass  $\delta < \varepsilon/(2K + 1)$ .

Nach dem Riemannsches Integrierbarkeitskriterium existiert eine Zerlegung  $Z = \{x_0, x_1, \dots, x_n\}$  von  $[a, b]$  mit  $\overline{S}(f, Z) - \underline{S}(f, Z) < \delta^2$ . Mit  $I_k := [x_{k-1}, x_k]$  wie vorher definieren wir nun  $M_k := \sup f(I_k)$ ,  $m_k := \inf f(I_k)$ ,  $M_k^* := \sup \phi(f(I_k))$ ,  $m_k^* := \inf \phi(f(I_k))$ ,  $A = \{k : M_k - m_k < \delta\}$  und  $B = \{k : M_k - m_k \geq \delta\}$ . Für  $k \in A$  ist dann  $M_k^* - m_k^* < \varepsilon'$ . Außerdem gilt für alle  $k$ , insbesondere also für  $k \in B$ , auch  $M_k^* - m_k^* \leq K$ . Desweiteren ist

$$\sum_{k \in B} |I_k| \leq \frac{1}{\delta} \sum_{k \in B} (M_k - m_k) |I_k| \leq \frac{1}{\delta} \sum_{k=1}^n (M_k - m_k) |I_k| = \frac{1}{\delta} (\overline{S}(f, Z) - \underline{S}(f, Z)) < \delta.$$

Es folgt

$$\begin{aligned} \overline{S}(\phi \circ f, Z) - \underline{S}(\phi \circ f, Z) &= \sum_{k=1}^n (M_k^* - m_k^*) |I_k| \\ &\leq \sum_{k \in A} \varepsilon' |I_k| + \sum_{k \in B} K |I_k| \\ &< \varepsilon' |I| + K \delta \\ &< \varepsilon. \quad \square \end{aligned}$$

**Satz 5.1.8** Sind  $f, g : [a, b] \rightarrow \mathbb{C}$  integrierbar, so sind auch  $|f|$  und  $f \cdot g$  integrierbar.

*Beweis.* Für reellwertiges  $f$  folgt die Integrierbarkeit von  $|f|$  direkt aus Satz 5.1.7 (mit  $\phi(x) = |x|$ ). Mit  $\phi(x) = x^2$  folgt aus diesem Hilfssatz auch die Integrierbarkeit von  $f^2$ . Wegen  $fg = \frac{1}{4}((f+g)^2 - (f-g)^2)$  erhält man mit Satz 5.1.4 dann auch die Integrierbarkeit von  $f \cdot g$ , falls  $f$  und  $g$  beide reellwertig sind.

Für komplexwertiges  $f, g$  erhält man die Integrierbarkeit von  $|f|$  und  $f \cdot g$  jetzt mit Satz 5.1.5 (und wiederum Satz 5.1.7).  $\square$

Durch Abschätzung der Riemannsches Summen zeigt man leicht folgenden Satz.

**Satz 5.1.9** Ist  $f : [a, b] \rightarrow \mathbb{C}$  integrierbar, so gilt

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx \leq (b-a) \sup_{x \in [a, b]} |f(x)|.$$

Sind  $f, g : [a, b] \rightarrow \mathbb{R}$  integrierbar mit  $f(x) \leq g(x)$  für alle  $x \in [a, b]$ , so gilt

$$\int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

Insbesondere gilt für integrierbares  $f : [a, b] \rightarrow \mathbb{R}$

$$(b-a) \inf_{x \in [a, b]} f(x) \leq \int_a^b f(x) dx \leq (b-a) \sup_{x \in [a, b]} f(x).$$

Die linke bzw. rechte Seite in der letzten Ungleichung sind auch die Riemannsche Unter- bzw. Obersumme zu der Zerlegung von  $I$ , die nur aus den beiden Randpunkten besteht.

**Satz 5.1.10 (Mittelwertsatz der Integralrechnung)** Sei  $f : [a, b] \rightarrow \mathbb{R}$  integrierbar. Dann existiert  $\mu \in \mathbb{R}$  mit  $\inf f([a, b]) \leq \mu \leq \sup f([a, b])$  und  $\int_a^b f = \mu(b - a)$ . Ist  $f$  stetig, so existiert  $\xi \in (a, b)$  mit  $\mu = f(\xi)$ .

*Beweis.* Die erste Behauptung folgt aus obiger Folgerung, die zweite aus dem Zwischenwertsatz.  $\square$

Analog beweist man das folgende Ergebnis.

**Satz 5.1.11 (Erweiterter Mittelwertsatz der Integralrechnung)** Seien  $f, g : [a, b] \rightarrow \mathbb{R}$  integrierbar. Sei  $g(x) \geq 0$  für alle  $x \in [a, b]$ . Dann existiert  $\mu \in \mathbb{R}$  mit  $\inf f([a, b]) \leq \mu \leq \sup f([a, b])$  und  $\int_a^b fg = \mu \int_a^b g$ . Ist  $f$  stetig, so existiert  $\xi \in (a, b)$  mit  $\mu = f(\xi)$ .

**Satz 5.1.12** Jede monotone Funktion  $f : [a, b] \rightarrow \mathbb{R}$  ist integrierbar.

*Beweis.* Ohne Beschränkung der Allgemeinheit sei  $f$  monoton steigend und nicht konstant. Dann ist  $f(b) > f(a)$ . Sei  $\varepsilon > 0$  und sei  $Z$  eine Zerlegung von  $[a, b]$  mit  $|Z| < \delta := \varepsilon / (f(b) - f(a))$ . Dann gilt

$$\begin{aligned} \overline{S}(f, Z) - \underline{S}(f, Z) &= \sum_{k=1}^n (f(x_k) - f(x_{k-1})) |I_k| \\ &< \delta \sum_{k=1}^n (f(x_k) - f(x_{k-1})) \\ &= \delta (f(b) - f(a)) \\ &= \varepsilon. \quad \square \end{aligned}$$

Sei  $[a, b] \subset M \subset \mathbb{R}$  und  $f : M \rightarrow \mathbb{C}$  eine Funktion. Ist  $f|_{[a, b]}$  integrierbar, so sagt man, dass  $f$  über das Intervall  $[a, b]$  integrierbar ist.

**Satz 5.1.13** (i) Sei  $f : [a, b] \rightarrow \mathbb{C}$  integrierbar und  $a \leq c < d \leq b$ , also  $[c, d] \subset [a, b]$ . Dann ist  $f$  über  $[c, d]$  integrierbar.

(ii) Seien  $a, b, c \in \mathbb{R}$ ,  $a < b < c$  und sei  $f : [a, c] \rightarrow \mathbb{C}$ . Ist  $f$  integrierbar über  $[a, b]$  und  $[b, c]$ , so auch über  $[a, c]$  und es gilt

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

*Beweis.* Übung.

Wegen (ii) setzt man für  $a < b$  und integrierbares  $f : [a, b] \rightarrow \mathbb{C}$  auch  $\int_b^a f := -\int_a^b f$ . Außerdem setzt man  $\int_a^a f = 0$  für  $f : \{a\} \rightarrow \mathbb{C}$ . Mit diesen Setzungen gilt die Formel in (ii) für alle  $a, b, c \in \mathbb{R}$ , wenn die auftretenden Integrale definiert sind.



## 5.2 Der Hauptsatz der Differential- und Integralrechnung

**Definition 5.2.1** Sei  $I$  Intervall,  $F : I \rightarrow \mathbb{C}$  differenzierbar und  $f : I \rightarrow \mathbb{C}$ . Dann heißt  $F$  *Stammfunktion* von  $f$ , falls  $F' = f$ .

Seien  $F, G$  Stammfunktionen von  $f, g$  und sei  $c \in \mathbb{C}$ . Unmittelbar sieht man ein, dass dann  $F + G$  Stammfunktion von  $f + g$  und  $c \cdot F$  Stammfunktion von  $c \cdot f$  ist.

Mit  $F$  ist auch  $F + c$  Stammfunktion von  $f$ . Umgekehrt gilt nach Satz 4.3.1, dass wenn  $F_1$  und  $F_2$  Stammfunktionen von  $f$  sind, so existiert  $c \in \mathbb{C}$  mit  $F_1 = F_2 + c$ .

*Beispiel.* Sei  $I := (-1, 1)$ ,  $f : I \rightarrow \mathbb{R}$ ,  $f(x) = -\frac{1}{\sqrt{1-x^2}}$ . Eine Stammfunktion von  $f$  ist  $\arccos|_I$ , eine andere ist  $-\arcsin|_I$ . Man beachte, dass  $\arccos = \frac{\pi}{2} - \arcsin$ .

**Satz 5.2.1 (Hauptsatz der Differential- und Integralrechnung)** Seien  $a, b \in \mathbb{R}$ ,  $a < b$  und  $f : [a, b] \rightarrow \mathbb{C}$  integrierbar.

- (i) Besitzt  $f$  eine Stammfunktion  $F$ , so gilt  $\int_a^b f = F(b) - F(a)$ .
- (ii) Sei  $F : [a, b] \rightarrow \mathbb{C}$  definiert durch  $F(x) = \int_a^x f$ . Ist  $f$  stetig in  $y \in [a, b]$ , so ist  $F$  differenzierbar in  $y$  mit  $F'(y) = f(y)$ . Ist  $f$  stetig in  $[a, b]$ , so ist  $F$  Stammfunktion von  $f$ .

*Beweis.* Ohne Beschränkung der Allgemeinheit sei  $f : [a, b] \rightarrow \mathbb{R}$ .

zu (i): Sei  $Z = \{x_0, x_1, \dots, x_n\}$  Zerlegung von  $[a, b]$ . Nach dem Mittelwertsatz (der Differentialrechnung) existiert zu  $k \in \{1, \dots, n\}$  dann  $\xi_k \in (x_{k-1}, x_k)$  mit

$$F(x_k) - F(x_{k-1}) = F'(\xi_k)(x_k - x_{k-1}) = f(\xi_k)(x_k - x_{k-1}).$$

Mit  $\xi = (\xi_1, \dots, \xi_n)$  erhält man

$$\begin{aligned} F(b) - F(a) &= F(x_n) - F(x_0) \\ &= \sum_{k=1}^n F(x_k) - F(x_{k-1}) \\ &= \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}) \\ &= S(f, Z, \xi). \end{aligned}$$

Hieraus folgt die Behauptung.

zu (ii): Seien  $x, y \in [a, b]$ ,  $x \neq y$ . Dann gilt nach Satz 5.1.13

$$F(x) - F(y) = \int_a^x f - \int_a^y f = \int_x^y f$$

und damit nach Satz 5.1.9

$$\begin{aligned} \left| \frac{F(x) - F(y)}{x - y} - f(y) \right| &= \left| \frac{1}{x - y} \int_y^x f(t) dt - f(y) \frac{1}{x - y} \int_y^x 1 dt \right| \\ &= \frac{1}{|x - y|} \left| \int_y^x (f(t) - f(y)) dt \right| \end{aligned}$$

$$\leq \sup_{t \in [y, x]} |f(t) - f(y)|.$$

Ist  $f$  stetig in  $y$ , so strebt die rechte Seite für  $x \rightarrow y$  gegen 0 und es folgt

$$\lim_{x \rightarrow y} \frac{F(x) - F(y)}{x - y} = f(y). \quad \square$$

Die im Hauptsatz auftretende Differenz  $F(b) - F(a)$  schreiben wir im folgenden oft in der Form

$$F \Big|_a^b, \quad F(x) \Big|_a^b \quad \text{oder} \quad F(x) \Big|_{x=a}^{x=b}.$$

*Beispiel.* Es ist  $\arctan' x = \frac{1}{1+x^2}$  für  $x \in \mathbb{R}$  und damit

$$\int_0^1 \frac{dx}{1+x^2} = \arctan \Big|_0^1 = \arctan 1 - \arctan 0 = \frac{\pi}{4}.$$

Man beachte, dass aus der Integrierbarkeit nicht die Existenz einer Stammfunktion folgt, und aus der Existenz einer Stammfunktion auch nicht die Integrierbarkeit.

*Beispiele.* 1. Sei  $f : [-1, 1] \rightarrow \mathbb{R}$ ,

$$f(x) = \begin{cases} -1 & \text{falls } -1 \leq x < 0, \\ 1 & \text{falls } 0 \leq x \leq 1. \end{cases}$$

Dann ist  $f$  integrierbar und  $\int_{-1}^1 f = \int_{-1}^0 f + \int_0^1 f = -1 + 1 = 0$ . Wäre  $F$  Stammfunktion, so wäre nach dem Hauptsatz  $F(1) - F(-1) = \int_{-1}^1 f = 0$ , nach dem Satz von Rolle würde also  $\xi \in (-1, 1)$  mit  $f(\xi) = F'(\xi) = 0$  existieren, ein Widerspruch. Also hat  $f$  keine Stammfunktion.

2. Sei  $F : [-1, 1] \rightarrow \mathbb{R}$ ,

$$F(x) = \begin{cases} x^2 \sin \frac{1}{x^2} & \text{falls } x \neq 0, \\ 0 & \text{falls } x = 0 \end{cases}$$

Dann ist  $F$  differenzierbar (vgl. das Beispiel in §4.1) mit

$$F'(x) = \begin{cases} 2x \sin \frac{1}{x^2} - \frac{2}{x} \cos \frac{1}{x^2} & \text{falls } x \neq 0, \\ 0 & \text{falls } x = 0 \end{cases}$$

Setzt man  $f := F'$ , so ist also  $F$  Stammfunktion von  $f$ . Es ist aber  $f$  nicht integrierbar, da  $f\left(\frac{1}{\sqrt{2\pi k}}\right) = 2\sqrt{2\pi k} \rightarrow \infty$  für  $k \rightarrow \infty$ ,  $k \in \mathbb{N}$ , und damit  $f$  unbeschränkt ist.

Der Hauptsatz reduziert also die Berechnung von Integralen auf das Auffinden von Stammfunktionen. Statt “ $F$  ist Stammfunktion von  $f$ ” schreibt man auch

$$\int f = F \quad \text{oder} \quad \int f(x) dx = F(x).$$

Da für konstantes  $C$  mit  $F$  auch  $F + C$  Stammfunktion ist, schreibt man manchmal auch

$$\int f(x)dx = F(x) + C.$$

Man nennt Stammfunktionen auch *unbestimmte Integrale*.

Obige Schreibweisen sind mit Vorsicht zu benutzen! So gilt etwa

$$\int \frac{dx}{1+x^2} = \arctan x \quad \text{und} \quad \int \frac{dx}{1+x^2} = -\operatorname{arccot} x,$$

aber es gilt  $\arctan x \neq -\operatorname{arccot} x$ . (Es ist  $\arctan x + \operatorname{arccot} x = \frac{\pi}{2}$ .)

Falls nichts anderes angegeben ist, gelten die in den folgenden *Beispielen* angegebenen Stammfunktionen in jedem Intervall, in denen der Integrand definiert ist.

- $\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1}$  für  $\alpha \in \mathbb{C}$ ,  $\alpha \neq -1$ ,
- $\int \frac{1}{x} dx = \ln|x|$ ,
- $\int e^x dx = e^x$ ,
- $\int \cos x dx = \sin x$ ,
- $\int \sin x dx = -\cos x$ ,
- $\int \frac{dx}{1+x^2} = \arctan x$ ,
- $\int \frac{dx}{1-x^2} = \frac{1}{2} \ln \left| \frac{1+x}{1-x} \right|$ ,
- $\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x$ ,
- $\int \frac{dx}{\cos^2 x} = \tan x$ ,
- $\int \frac{dx}{\sin^2 x} = -\cot x$ ,
- $\int \frac{dx}{\sqrt{1+x^2}} = \ln \left( x + \sqrt{1+x^2} \right)$ ,
- $\int \frac{dx}{\sqrt{x^2-1}} = \ln \left( x + \sqrt{x^2-1} \right)$  für  $x \in (1, \infty)$ .

Die in den letzten beiden Beispielen angegebenen Stammfunktionen sind Umkehrfunktionen der hyperbolischen Funktionen  $\sinh$  und  $\cosh$ , d. h., die durch

$$\operatorname{arsinh} : \mathbb{R} \rightarrow \mathbb{R}, \quad x \rightarrow \ln \left( x + \sqrt{1 + x^2} \right),$$

und

$$\operatorname{arcosh} : [1, \infty) \rightarrow [0, \infty), \quad x \rightarrow \ln \left( x + \sqrt{x^2 - 1} \right),$$

definierten Funktionen erfüllen  $\operatorname{arsinh}(\sinh x) = x$  für  $x \in \mathbb{R}$  bzw.  $\operatorname{arcosh}(\cosh x) = x$  für  $x \in [0, \infty)$ . Sie heißen *Area sinus (bzw. cosinus) hyperbolicus*; vgl. § 3.4. Im Intervall  $(-\infty, -1)$  gilt

$$\bullet \int \frac{dx}{\sqrt{x^2 - 1}} = -\operatorname{arcosh}(-x).$$

### 5.3 Integrationstechniken

Kennt man die Ableitungen zweier Funktionen  $f$  und  $g$ , so kann man die von  $f \cdot g$ ,  $f/g$ ,  $f \circ g$ , usw. daraus berechnen. Entsprechendes gilt *nicht* für die Berechnung von Stammfunktionen. Dennoch gibt es einige nützliche Regeln.

**Satz 5.3.1 (Partielle Integration)** *Seien  $f, g : [a, b] \rightarrow \mathbb{C}$  stetig differenzierbar. Dann gilt*

$$\int_a^b fg' = fg \Big|_a^b - \int_a^b f'g.$$

*Beweis.* Sei  $F : [a, b] \rightarrow \mathbb{C}$ ,  $F(x) = fg \Big|_a^x - \int_a^x f'g$ . Nach Produktregel und Hauptsatz, Teil (ii), folgt dann  $F' = f'g + fg' - f'g = fg'$ . Die Behauptung folgt jetzt mit dem Hauptsatz, Teil (i).  $\square$

Man schreibt die Regel der partiellen Integration auch als

$$\int f(x)g(x)dx = f(x)g(x) - \int f'(x)g(x)dx.$$

*Beispiele.* 1. Für  $a \in \mathbb{R}$  ist

$$\int_0^a xe^x dx = xe^x \Big|_0^a - \int_0^a e^x dx = ae^a - e^x \Big|_0^a = ae^a - e^a + 1.$$

Es gilt also

$$\int xe^x dx = xe^x - e^x = (x - 1)e^x.$$

(Die additive Konstante 1 ist hier weggelassen.)

2. Für  $x > 0$  ist

$$\int_1^x \ln t dt = \int_1^x \underbrace{1}_{g'} \cdot \underbrace{\ln t}_f dt = t \cdot \ln t \Big|_1^x - \int_1^x t \cdot \frac{1}{t} dt = x \ln x - x + 1.$$

Wiederum schreibt man auch  $\int \ln x = x \ln x - x = x(\ln x - 1)$ .

Neben der partiellen Integration ist die wichtigste Integrationsmethode die Substitution.

**Satz 5.3.2 (Substitutionsregel)** Sei  $g : [a, b] \rightarrow [c, d]$  stetig differenzierbar und  $f : [c, d] \rightarrow \mathbb{C}$  stetig. Dann gilt

$$\int_a^b f(g(t))g'(t)dt = \int_{g(a)}^{g(b)} f(x)dx.$$

*Beweis.* Nach dem Hauptsatz hat  $f$  eine Stammfunktion  $F$ . Nach Kettenregel ist  $(F \circ g)'(t) = F'(g(t))g'(t) = f(g(t))g'(t) = ((f \circ g) \cdot g')(t)$  und damit ist  $F \circ g$  Stammfunktion von  $(f \circ g) \cdot g'$ . Mit dem Hauptsatz folgt

$$\int_a^b (f \circ g) \cdot g' = F \circ g \Big|_a^b = F \Big|_{g(a)}^{g(b)} = \int_{g(a)}^{g(b)} f. \quad \square$$

Die Substitutionsregel kann man sich merken, indem man  $x = g(t)$  setzt. Dann ist  $\frac{dx}{dt} = g'(t)$ .

*Beispiele.* 1. Mit der Substitution  $x = t^2$  folgt

$$\int_0^2 \frac{t dt}{1+t^4} = \frac{1}{2} \int_0^4 \frac{dx}{1+x^2} = \frac{1}{2} \arctan 4.$$

2. Substituiert man zuerst  $x = \ln t$  und dann  $y = 1 + x$ , erhält man

$$\int_1^e \frac{\sqrt{1+\ln t}}{t} dt = \int_0^1 \sqrt{1+x} dx = \int_1^2 \sqrt{y} dy = \frac{2}{3} y^{3/2} \Big|_1^2 = \frac{2}{3} (2\sqrt{2} - 1).$$

3. Wir betrachten für  $x \in [-1, 1]$  das Integral

$$\int_0^x \sqrt{1-y^2} dy.$$

Mit  $y = \sin t$ ,  $t = \arcsin y$ , wobei  $y \in [-1, 1]$ ,  $t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ , ist dann  $\sqrt{1-y^2} = \sqrt{1-\sin^2 t} = \sqrt{\cos^2 t} = |\cos t| = \cos t$ . Die Substitutionsregel (mit  $g(t) = \sin t$ ,  $g'(t) = \cos t$ ) liefert also

$$\int_0^s \cos^2 t dt = \int_0^{\sin s} \sqrt{1-y^2} dy$$

für  $s \in [-1, 1]$ . Es folgt also

$$\begin{aligned} \int_0^x \sqrt{1-y^2} dy &= \int_0^{\arcsin x} \cos^2 t dt \\ &= \int_0^{\arcsin x} \frac{1}{2}(1 + \cos 2t) dt \\ &= \frac{1}{2} \left( t + \frac{\sin 2t}{2} \right) \Big|_0^{\arcsin x} \\ &= \frac{1}{2} (t + \sin t \cos t) \Big|_0^{\arcsin x} \\ &= \frac{1}{2} \left( \arcsin x + x\sqrt{1-x^2} \right). \end{aligned}$$

Wir beschreiben jetzt die *Integration rationaler Funktionen*, d. h., Funktionen  $f$  der Form  $f = \frac{P}{Q}$  mit teilerfremden Polynomen  $P, Q$ . Wir dürfen  $\text{grad}(P) < q := \text{grad}(Q)$  annehmen. (Andernfalls führen wir zunächst eine Polynomdivision durch.) Nach dem Fundamentalsatz der Algebra (Satz 1.8.2) existieren  $c \in \mathbb{C}$ ,  $k \in \mathbb{N}$ ,  $a_1, \dots, a_k \in \mathbb{C}$  und  $m_1, \dots, m_k \in \mathbb{N}$  mit

$$Q(z) = c \prod_{j=1}^k (z - a_j)^{m_j}$$

für  $z \in \mathbb{C}$ . Es gilt dann  $\sum_{j=1}^k m_j = q$ . Wir dürfen  $c = 1$  annehmen. Wir setzen nun die Koeffizienten von  $Q$  als reell voraus. Dann gilt  $\overline{Q(z)} = Q(\bar{z})$  und damit ist mit  $a$  auch  $\bar{a}$  Nullstelle von  $Q$ . Wir fassen in dem Produkt für  $Q$  nun konjugiert komplexe Nullstellen zusammen und erhalten wegen

$$(z - a)(z - \bar{a}) = z^2 - (a + \bar{a})z + a\bar{a} = z^2 - 2 \operatorname{Re} a \cdot z + |a|^2$$

für  $Q$  dann eine Darstellung

$$Q(x) = \prod_{j=1}^M (x - a_j)^{m_j} \prod_{j=1}^N (x^2 + b_j x + c_j)^{n_j}$$

mit  $a_j, b_j, c_j \in \mathbb{R}$ ,  $c_j > b_j^2/4$ ,  $m_j, n_j \in \mathbb{N}$  und  $\sum_{j=1}^M m_j + 2 \sum_{j=1}^N n_j = q$ .

Der *Satz von der Partialbruchzerlegung* besagt nun, dass  $A_{i,j} \in \mathbb{R}$ ,  $i = 1, \dots, M$ ,  $j = 1, \dots, m_i$ , und  $B_{i,j}, C_{i,j} \in \mathbb{R}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$ , existieren, so dass

$$\begin{aligned} \frac{P(x)}{Q(x)} = & \frac{A_{1,1}}{x - a_1} + \frac{A_{1,2}}{(x - a_1)^2} + \dots + \frac{A_{1,m_1}}{(x - a_1)^{m_1}} \\ & + \frac{A_{2,1}}{x - a_2} + \frac{A_{2,2}}{(x - a_2)^2} + \dots + \frac{A_{2,m_2}}{(x - a_2)^{m_2}} \\ & + \dots \\ & + \frac{A_{M,1}}{x - a_M} + \frac{A_{M,2}}{(x - a_M)^2} + \dots + \frac{A_{M,m_M}}{(x - a_M)^{m_M}} \\ & + \frac{B_{1,1}x + C_{1,1}}{x^2 + b_1x + c_1} + \frac{B_{1,2}x + C_{1,2}}{(x^2 + b_1x + c_1)^2} + \dots + \frac{B_{1,n_1}x + C_{1,n_1}}{(x^2 + b_1x + c_1)^{n_1}} \\ & + \dots \\ & + \frac{B_{N,1}x + C_{N,1}}{x^2 + b_Nx + c_N} + \frac{B_{N,2}x + C_{N,2}}{(x^2 + b_Nx + c_N)^2} + \dots + \frac{B_{N,n_N}x + C_{N,n_N}}{(x^2 + b_Nx + c_N)^{n_N}}. \end{aligned}$$

Der *Beweis* lässt sich durch vollständige Induktion über  $q$  führen; hier nicht. Die Koeffizienten  $A_{i,j}, B_{i,j}, C_{i,j}$  kann man durch Multiplikation mit  $Q(x)$  und Koeffizientenvergleich bestimmen.

Die in der Partialbruchzerlegung rechts stehenden Terme lassen sich nun integrieren. Es ist zunächst

$$\int \frac{A}{x - a} dx = A \ln |x - a|$$

und

$$\int \frac{A}{(x - a)^m} dx = -\frac{A}{(m - 1)(x - a)^{m-1}}$$

für  $m \geq 2$ . Wegen  $x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + c - \frac{b^2}{4}$  lassen sich die Terme mit  $(x^2 + bx + c)^n$  im Nenner durch eine Substitution auf die Form

$$\frac{\beta x + \gamma}{(x^2 + \alpha^2)^n}$$

bringen. Nun ist

$$\int \frac{\beta x}{x^2 + \alpha^2} dx = \frac{\beta}{2} \ln(x^2 + \alpha^2)$$

und

$$\int \frac{\beta x}{(x^2 + \alpha^2)^n} dx = -\frac{\beta}{2(n-1)} \frac{1}{(x^2 + \alpha^2)^{n-1}}$$

für  $n \geq 2$ . Schließlich ist noch

$$\int \frac{dx}{x^2 + \alpha^2} = \frac{1}{\alpha} \arctan \frac{x}{\alpha}$$

und für  $n \geq 2$  zeigt partielle Integration, dass

$$\begin{aligned} \int \frac{dx}{(x^2 + \alpha^2)^n} &= \frac{1}{\alpha^2} \int \frac{\alpha^2 + x^2 - x^2}{(x^2 + \alpha^2)^n} dx \\ &= \frac{1}{\alpha^2} \int \frac{dx}{(x^2 + \alpha^2)^{n-1}} - \frac{1}{\alpha^2} \int \underbrace{x}_f \cdot \underbrace{\frac{x}{(x^2 + \alpha^2)^n}}_{g'} dx \\ &= \frac{1}{\alpha^2} \int \frac{dx}{(x^2 + \alpha^2)^{n-1}} - \frac{1}{\alpha^2} \left[ x \left( -\frac{1}{2(n-1)(x^2 + \alpha^2)^{n-1}} \right) \right. \\ &\quad \left. + \int \frac{dx}{2(n-1)(x^2 + \alpha^2)^{n-1}} \right]. \end{aligned}$$

Damit können die Integrale

$$\int \frac{dx}{(x^2 + \alpha^2)^n}$$

rekursiv berechnet werden.

*Beispiel.* Man berechne

$$\int_1^2 \frac{x^4 + 1}{x^3 + x} dx.$$

Zunächst ist

$$f(x) := \frac{x^4 + 1}{x^3 + x} = x - \frac{x^2 - 1}{x^3 + x} = x - \frac{x^2 - 1}{x(x^2 + 1)}.$$

Es existieren nun  $A, B, C \in \mathbb{R}$  mit

$$\frac{x^2 - 1}{x(x^2 + 1)} = \frac{A}{x} + \frac{Bx + C}{x^2 + 1}.$$

Zur Bestimmung der Konstanten  $A, B, C$  kann man natürlich mit dem Nenner durchmultiplizieren und anschließend Koeffizientenvergleich machen, was auf ein lineares Gleichungssystem für  $A, B, C$  führt. Einfacher ist, obige Gleichung mit  $x$

zu multiplizieren und anschließend  $x = 0$  zu setzen. Dies führt sofort auf  $A = -1$  und damit auf

$$\frac{Bx + C}{x^2 + 1} = \frac{x^2 - 1}{x(x^2 + 1)} + \frac{1}{x} = \frac{x^2 - 1 + x^2 + 1}{x(x^2 + 1)} = \frac{2x}{x^2 + 1}.$$

Dies ergibt

$$\begin{aligned} \int_1^2 f(x) dx &= \int_1^2 \left( x + \frac{1}{x} - \frac{2x}{x^2 + 1} \right) dx \\ &= \left. \frac{x^2}{2} + \ln x - \ln(x^2 + 1) \right|_1^2 \\ &= 2 + \ln 2 - \ln 5 - \frac{1}{2} + \ln 2 \\ &= \frac{3}{2} + 2 \ln 2 - \ln 5 \quad \left( = \frac{3}{2} + \ln \frac{4}{5} = 1,276\dots \right) \end{aligned}$$

Viele Integrale lassen sich durch geeignete Substitutionen auf Integrale über rationale Funktionen zurückführen. Sei etwa  $R$  eine rationale Funktionen von zwei Veränderlichen, d. h.,  $R(x, y) = P(x, y)/Q(x, y)$ , wobei  $P, Q$  bei festem  $x$  Polynome in  $y$  und bei festem  $y$  Polynome in  $x$  sind. Dann kann

$$\int R(\sin x, \cos x) dx$$

durch die Substitution  $t = \tan \frac{x}{2}$ ,  $x = 2 \arctan t$ , in ein Integral über eine rationale Funktion überführt werden. Die Rechnungen sind allerdings aufwendig.

Auch Integrale der Form

$$\int R(x, \sqrt{ax^2 + bx + c}) dx$$

lassen sich durch geeignete Substitutionen in Integrale über rationale Funktionen umformen.

## 5.4 Integralform des Taylorrestglieds

Mit partieller Integration lässt sich auch eine Formel für das Restglied der Taylorentwicklung gewinnen. Für eine  $n$ -mal differenzierbare Funktion  $f : I \rightarrow \mathbb{R}$  und  $\xi, x \in I$  hatten wir (siehe §4.4)

$$T_n(x) := \sum_{k=0}^n \frac{f^{(k)}(\xi)}{k!} (x - \xi)^k$$

als Taylorpolynom (vom Grad  $n$ ) bezeichnet. Für das *Restglied*

$$R_n(x) := f(x) - T_n(x)$$

hatten wir für  $(n + 1)$ -mal differenzierbares  $f$  die Darstellung

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



mit  $\eta$  zwischen  $\xi$  und  $x$  bewiesen (Satz 4.4.3). Man nennt dies auch die *Lagrangesche Form des Restglieds*. Wir leiten noch eine *Integralform des Restglieds* her.

Wir bezeichnen für ein Intervall  $I$  und  $n \in \mathbb{N}$  die Menge der  $n$ -mal stetig differenzierbaren Funktionen von  $I$  nach  $\mathbb{C}$  mit  $C^n(I)$ . Mit  $C(I)$  bezeichnen wir die Menge der stetigen (englisch: continuous) Funktionen von  $I$  nach  $\mathbb{C}$ . Statt  $C^n([a, b])$  bzw.  $C([a, b])$  schreiben wir  $C^n[a, b]$  bzw.  $C[a, b]$ .

**Satz 5.4.1** Für  $f \in C^{n+1}(I)$  und  $\xi, x \in I$  gilt

$$R_n(x) = \frac{1}{n!} \int_{\xi}^x f^{(n+1)}(t)(x-t)^n dt.$$

*Beweis.* Es ist

$$\begin{aligned} f(x) &= f(\xi) + \int_{\xi}^x f'(t) dt \\ &= f(\xi) + f'(t)(t-x) \Big|_{t=\xi}^{t=x} - \int_{\xi}^x f''(t)(t-x) dt \\ &= f(\xi) + f'(\xi)(x-\xi) + \int_{\xi}^x f''(t)(x-t) dt \\ &= f(\xi) + f'(\xi)(x-\xi) + \frac{1}{2} f''(\xi)(x-\xi)^2 + \frac{1}{2} \int_{\xi}^x f'''(t)(x-t)^2 dt \\ &= \dots, \end{aligned}$$

und die Behauptung folgt durch vollständige Induktion.  $\square$

Ist  $f$  reellwertig, so erhält man mit dem erweiterten Mittelwertsatz der Integralrechnung (Satz 5.1.11) aus der Integralform die Lagrangesche Form des Restglieds. (In Satz 4.4.3 wurde allerdings nicht die Stetigkeit von  $f^{(n+1)}$  vorausgesetzt.)

In §4.4 hatten wir das Beispiel  $f(x) = \ln x$ ,  $\xi = 1$ , untersucht und gezeigt, dass  $\lim_{n \rightarrow \infty} R_n(x) = 0$  für  $\frac{1}{2} \leq x \leq 2$  gilt. Mit Satz 5.4.1 können wir zeigen, dass dies sogar für  $0 < x \leq 2$  gilt, denn für  $0 < x < 1$  ist

$$\begin{aligned} |R_n(x)| &= \left| \frac{1}{n!} \int_x^1 (x-t)^n f^{(n+1)}(t) dt \right| \\ &= \int_x^1 \frac{(t-x)^n}{t^{n+1}} dt \\ &= \int_x^1 \left(1 - \frac{x}{t}\right)^n \frac{dt}{t} \\ &\leq (1-x)^n \int_x^1 \frac{dt}{t} \\ &= (1-x)^n \ln \frac{1}{x}. \end{aligned}$$

Für  $f \in C^{n+1}(I)$  folgt aus Satz 5.4.1, und für reellwertiges  $f$  auch aus Satz 4.4.3, dass positive Konstanten  $K, \varepsilon \in \mathbb{R}$  mit

$$|R_n(x)| \leq K|x-\xi|^{n+1}$$

für  $|x-\xi| < \varepsilon$  existieren. Nun können  $T_n$  und  $R_n$  aber bereits für  $n$ -mal differenzierbares  $f$  definiert werden. Dann erhalten wir aber nur schlechtere Abschätzungen.

**Satz 5.4.2** Für  $f \in C^n(I)$  und  $\xi, x$  in  $I$  gilt

$$\lim_{x \rightarrow \xi} \frac{R_n(x)}{(x - \xi)^n} = 0.$$

*Beweis.* Es gilt

$$\begin{aligned} |R_n(x)| &= |f(x) - T_n(x)| \\ &= \left| f(x) - T_{n-1}(x) - \frac{1}{n!} f^{(n)}(\xi)(x - \xi)^n \right| \\ &= \left| \frac{1}{(n-1)!} \int_{\xi}^x f^{(n)}(t)(x-t)^{n-1} dt - \frac{1}{n!} f^{(n)}(\xi)(x - \xi)^n \right| \\ &= \left| \frac{1}{(n-1)!} \int_{\xi}^x (f^{(n)}(t) - f^{(n)}(\xi))(x-t)^{n-1} dt \right| \\ &\leq |x - \xi|^n \frac{1}{(n-1)!} \sup_t |f^{(n)}(t) - f^{(n)}(\xi)|, \end{aligned}$$

wobei das Supremum über alle  $t$  zwischen  $\xi$  und  $x$  zu nehmen ist. Die Behauptung folgt.  $\square$

Für reellwertiges  $f$  erhalten wir die Behauptung von Satz 5.4.2 auch direkt aus der Lagrangeschen Form des Restglieds:

$$R_n(x) = R_{n-1}(x) - \frac{1}{n!} f^{(n)}(\xi)(x - \xi)^n = \frac{1}{n!} (f^{(n)}(\eta) - f^{(n)}(\xi)) (x - \xi)^n.$$

Die obigen Restgliedabschätzungen lassen sich gut mit den sogenannten *Landau-Symbolen* beschreiben. Ist  $M \subset \mathbb{C}$  und  $\xi$  Häufungspunkt von  $M$ , wobei für  $M \subset \mathbb{R}$  auch  $\xi = \pm\infty$  zugelassen ist, und sind  $\alpha : M \rightarrow \mathbb{C}$  und  $\beta : M \rightarrow (0, \infty)$ , so schreibt man

$$\alpha(x) = o(\beta(x)) \text{ für } x \rightarrow \xi, \quad \text{falls } \lim_{x \rightarrow \xi} \frac{\alpha(x)}{\beta(x)} = 0,$$

und

$$\alpha(x) = O(\beta(x)) \text{ für } x \rightarrow \xi, \quad \text{falls } \limsup_{x \rightarrow \xi} \frac{|\alpha(x)|}{\beta(x)} < \infty.$$

Dabei lässt man den Zusatz “für  $x \rightarrow \xi$ ” oft weg, wenn dieser aus dem Zusammenhang hervorgeht. Offensichtlich gilt  $\alpha(x) = O(\beta(x))$  für  $x \rightarrow \xi$  genau dann, wenn positive Konstanten  $K, \varepsilon$  existieren, so dass  $|\alpha(x)|/\beta(x) \leq K$  für  $|x - \xi| < \varepsilon$ , bzw. falls  $x > 1/\varepsilon$  oder  $x < -1/\varepsilon$  falls  $\xi = \pm\infty$ .

Für  $\xi, x \in I$  und  $f : I \rightarrow \mathbb{C}$  (mit einem Intervall  $I$ ) erhalten wir damit:

$$\begin{aligned} f \in C^{n+1}(I) &\Rightarrow R_n(x) = O(|x - \xi|^{n+1}) \text{ für } x \rightarrow \xi, \\ f \in C^n(I) &\Rightarrow R_n(x) = o(|x - \xi|^n) \text{ für } x \rightarrow \xi. \end{aligned}$$

## 5.5 Folgen und Reihen integrierbarer Funktionen

In §3.7 bzw. §4.4 hatten wir untersucht, wann sich bei einer konvergenten Funktionenfolge aus der Stetigkeit bzw. Differenzierbarkeit der Folgenglieder die der Grenzfunktion folgern lässt. Wir betrachten hier das entsprechende Problem für Integrierbarkeit.

**Satz 5.5.1** Sei  $(f_n)$  eine Folge über  $[a, b]$  integrierbarer Funktionen, die gleichmäßig gegen eine Funktion  $f$  konvergiert. Dann ist  $f$  über  $[a, b]$  integrierbar und es gilt  $\lim_{n \rightarrow \infty} \int_a^b f_n = \int_a^b f$ .

*Beweis.* Ist bereits bekannt, dass  $f$  integrierbar ist, so folgt  $\lim_{n \rightarrow \infty} \int_a^b f_n = \int_a^b f$  sehr leicht. Denn ist  $\varepsilon > 0$ , so existiert  $N \in \mathbb{N}$  mit  $\sup_{x \in [a, b]} |f_n(x) - f(x)| < \varepsilon / (b - a)$  für alle  $n \geq N$ . Es folgt

$$\left| \int_a^b f_n - \int_a^b f \right| = \left| \int_a^b (f_n - f) \right| \leq (b - a) \sup_{x \in [a, b]} |f_n(x) - f(x)| < \varepsilon$$

und damit die Behauptung.

Zu zeigen ist also lediglich die Integrierbarkeit von  $f$ . Diese zeigen wir mit dem Cauchy Kriterium (Satz 5.1.1). Sei dazu  $\varepsilon > 0$ . Dann existiert  $N \in \mathbb{N}$  mit  $\sup_{x \in [a, b]} |f_n(x) - f(x)| < \frac{\varepsilon}{3(b-a)}$  für  $n \geq N$ . Für jede Zerlegung  $Z$  von  $[a, b]$  mit Stützstellen  $\xi$  folgt dann

$$|S(f_N, Z, \xi) - S(f, Z, \xi)| = |S(f_N - f, Z, \xi)| < \frac{\varepsilon}{3(b-a)}(b-a) = \frac{\varepsilon}{3}.$$

Da  $f_N$  integrierbar existiert  $\delta > 0$ , so dass für alle Zerlegungen  $Z_1, Z_2$  mit Stützstellen  $\xi_1, \xi_2$  gilt: Ist  $|Z_1| < \delta$  und  $|Z_2| < \delta$ , so ist

$$|S(f_N, Z_1, \xi_1) - S(f_N, Z_2, \xi_2)| < \frac{\varepsilon}{3}.$$

Für  $|Z_1| < \delta$  und  $|Z_2| < \delta$  folgt damit

$$\begin{aligned} & |S(f, Z_1, \xi_1) - S(f, Z_2, \xi_2)| \\ & \leq |S(f, Z_1, \xi_1) - S(f_N, Z_1, \xi_1)| + |S(f_N, Z_1, \xi_1) - S(f_N, Z_2, \xi_2)| \\ & \quad + |S(f_N, Z_2, \xi_2) - S(f, Z_2, \xi_2)| \\ & < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ & = \varepsilon. \quad \square \end{aligned}$$

Natürlich gilt auch wieder eine Analogon von Satz 5.5.1 für Reihen.

## 5.6 Fourierreihen

Sei  $\omega > 0$ . Eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{C}$  heißt  $\omega$ -periodisch, falls  $f(x + \omega) = f(x)$  für alle  $x \in \mathbb{R}$ . Beispiele  $2\pi$ -periodischer Funktionen sind der Cosinus und Sinus und allgemeiner für  $k \in \mathbb{Z}$  auch die durch

$$x \mapsto \cos kx, \quad x \mapsto \sin kx, \quad x \mapsto e_k(x) := e^{ikx} = \cos kx + i \sin kx$$

definierten Funktionen von  $\mathbb{R}$  nach  $\mathbb{C}$ . Wir untersuchen, wann eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{C}$  eine Darstellung der Form  $f = \sum_{k=-\infty}^{\infty} c_k e_k$ , also  $f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$  für alle  $x \in \mathbb{R}$ , besitzt.

Dabei heißt für  $u : \mathbb{Z} \rightarrow \mathbb{C}$  die Reihe  $\sum_{k=-\infty}^{\infty} u(k)$  konvergent, falls die Reihen  $\sum_{k=0}^{\infty} u(k)$  und  $\sum_{k=1}^{\infty} u(-k)$  beide konvergieren, und man setzt dann  $\sum_{k=-\infty}^{\infty} u(k) =$

$\sum_{k=0}^{\infty} u(k) + \sum_{k=1}^{\infty} u(-k)$ . Analog definiert man Begriffe wie absolute Konvergenz, gleichmäßige Konvergenz, usw. für Reihen von  $-\infty$  nach  $-\infty$ .

Man nennt für  $n \in \mathbb{N}_0$  und  $c_k \in \mathbb{C}$ ,  $-n \leq k \leq n$ , eine Funktion der Form

$$x \mapsto \sum_{k=-n}^n c_k e^{ikx}$$

ein *trigonometrisches Polynom*. Wegen  $e^{ikx} = \cos kx + i \sin kx$  gilt

$$\sum_{k=-n}^n c_k e^{ikx} = c_0 + \sum_{k=1}^n ((c_k + c_{-k}) \cos kx + i(c_k - c_{-k}) \sin kx),$$

mit  $a_k := c_k + c_{-k}$  und  $b_k := i(c_k - c_{-k})$  also

$$\sum_{k=-n}^n c_k e^{ikx} = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx).$$

Man erkennt leicht, dass  $a_k$  und  $b_k$  genau dann reell sind, wenn  $c_{-k} = \overline{c_k}$  gilt.

Konvergiert die Reihe  $\sum_{k=-\infty}^{\infty} c_k e_k$  gleichmäßig, ist durch  $x \mapsto \sum_{k=-\infty}^{\infty} c_k e^{ikx}$  nach Satz 3.7.2 (bzw. dem Analogon dieses Satzes für Reihen) eine stetige Funktion  $f : \mathbb{R} \rightarrow \mathbb{C}$  gegeben. Mit Satz 5.5.1 (bzw. dem Analogon dieses Satzes für Reihen) folgt für  $j \in \mathbb{Z}$ , dass

$$\begin{aligned} \int_{-\pi}^{\pi} f(x) e^{-ijx} dx &= \int_{-\pi}^{\pi} \left( \sum_{k=-\infty}^{\infty} c_k e^{ikx} \right) e^{-ijx} dx \\ &= \int_{-\pi}^{\pi} \left( \sum_{k=-\infty}^{\infty} c_k e^{i(k-j)x} \right) dx \\ &= \sum_{k=-\infty}^{\infty} c_k \int_{-\pi}^{\pi} e^{i(k-j)x} dx. \end{aligned}$$

Leicht rechnet man nach, dass für  $m \in \mathbb{Z}$

$$\int_{-\pi}^{\pi} e^{imx} dx = \begin{cases} 2\pi & \text{falls } m = 0, \\ 0 & \text{falls } m \neq 0, \end{cases}$$

gilt. Es folgt  $\int_{-\pi}^{\pi} f(x) e^{-ijx} dx = c_j 2\pi$ , also

$$c_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ijx} dx.$$

Dies motiviert folgende Definition.

**Definition 5.6.1** Sei  $f : \mathbb{R} \rightarrow \mathbb{C}$  integrierbar über  $[-\pi, \pi]$  und  $2\pi$ -periodisch. Für  $k \in \mathbb{Z}$  heißt dann

$$\widehat{f}(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx$$

der  $k$ -te *Fourierkoeffizient* von  $f$ . Für  $n \in \mathbb{N}$  heißt

$$F_n(x) := F_n(x, f) := \sum_{k=-n}^n \widehat{f}(k) e^{ikx}$$

das  $n$ -te *Fourierpolynom* von  $f$  und

$$F(x) := F(x, f) := \sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikx}$$

heißt *Fourierreihe* von  $f$ .

Die Überlegungen vor der Definition zeigen, dass wenn eine  $2\pi$ -periodische Funktion  $f : \mathbb{R} \rightarrow \mathbb{C}$  eine gleichmäßig konvergente Reihenentwicklung  $f = \sum_{k=-\infty}^{\infty} c_k e_k$  hat, dann gilt  $c_k = \widehat{f}(k)$ . Wenn die Fourierreihe  $F$  einer  $2\pi$ -periodischen Funktion  $f : \mathbb{R} \rightarrow \mathbb{C}$  gleichmäßig konvergiert, gilt also  $\widehat{f}(k) = \widehat{F}(k)$ . Außerdem ist  $F$  stetig. Ohne Beweis notieren wir folgenden Satz.

**Satz 5.6.1** *Seien  $f, g : \mathbb{R} \rightarrow \mathbb{C}$  stetige,  $2\pi$ -periodische Funktionen mit  $\widehat{f}(k) = \widehat{g}(k)$  für alle  $k \in \mathbb{Z}$ . Dann gilt  $f = g$ .*

Hieraus erhält man direkt das folgende Ergebnis.

**Satz 5.6.2** *Sei  $f : \mathbb{R} \rightarrow \mathbb{C}$  stetig und  $2\pi$ -periodisch. Die Fourierreihe von  $f$  konvergiere gleichmäßig. Dann gilt  $f(x) = F(x, f)$  für alle  $x \in \mathbb{R}$ .*

Als Anwendung erhalten wir das folgende Resultat.

**Satz 5.6.3** *Sei  $f : \mathbb{R} \rightarrow \mathbb{C}$  zweimal stetig differenzierbar und  $2\pi$ -periodisch. Dann konvergiert die Fourierreihe von  $f$  gleichmäßig und es gilt  $f(x) = F(x, f)$  für alle  $x \in \mathbb{R}$ .*

Zum Beweis benutzen wir folgenden Hilfssatz.

**Hilfssatz 5.6.1** *Sei  $f : \mathbb{R} \rightarrow \mathbb{C}$  stetig differenzierbar und  $2\pi$ -periodisch. Dann gilt  $\widehat{f}'(k) = ik\widehat{f}(k)$ .*

*Beweis.* Mit partieller Integration folgt

$$2\pi \widehat{f}'(k) = \int_{-\pi}^{\pi} f'(x) e^{-ikx} dx = f(x) e^{-ikx} \Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f(x) (-ik) e^{-ikx} dx = 2\pi ik \widehat{f}(k)$$

und daraus die Behauptung.  $\square$

*Beweis von Satz 5.6.3.* Mit  $M := \max_{x \in [-\pi, \pi]} |f''(x)|$  folgt nach Hilfssatz 5.6.1

$$\left| \widehat{f}(k) \right| = \frac{1}{k^2} \left| \widehat{f''}(k) \right| = \frac{1}{2\pi k^2} \left| \int_{-\pi}^{\pi} f''(x) e^{-ikx} dx \right| \leq \frac{M}{k^2}$$

für  $k \in \mathbb{Z} \setminus \{0\}$ . Hieraus folgt die gleichmäßige Konvergenz der Fourierreihe und damit die Behauptung wegen Satz 5.6.2.  $\square$

Tatsächlich konvergiert die Fourierreihe sogar für einmal stetig differenzierbare Funktionen. Punktweise Konvergenz hat man sogar unter wesentlich schwächeren Voraussetzungen.

Bevor wir Beispiele betrachten, notieren wir, dass mit

$$a_k := a_k(f) := \widehat{f}(k) + \widehat{f}(-k) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx$$

und

$$b_k := b_k(f) := i \left( \widehat{f}(k) - \widehat{f}(-k) \right) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx$$

das  $n$ -te Fourierpolynom die Form

$$F_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx)$$

hat.

*Beispiel 1.* Sei  $f$  die  $2\pi$ -periodische Funktion, die für  $x \in [-\pi, \pi]$  durch  $f(x) = x^2$  gegeben ist. Für  $k \neq 0$  erhält man mit partieller Integration

$$\begin{aligned} 2\pi \widehat{f}(k) &= \int_{-\pi}^{\pi} x^2 e^{-ikx} \, dx \\ &= x^2 \frac{e^{-ikx}}{-ik} \Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} 2x \frac{e^{-ikx}}{-ik} \, dx \\ &= \frac{2}{ik} \int_{-\pi}^{\pi} x e^{-ikx} \, dx \\ &= \frac{2}{ik} \left( x \frac{e^{-ikx}}{-ik} \Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} \frac{e^{-ikx}}{-ik} \, dx \right) \\ &= \frac{2}{ik} 2\pi \frac{(-1)^k}{-ik} \\ &= 2\pi \frac{2(-1)^k}{k^2}, \end{aligned}$$

also  $\widehat{f}(k) = 2(-1)^k/k^2$  für  $k \neq 0$  und damit

$$a_k(f) = \widehat{f}(k) + \widehat{f}(-k) = 4 \frac{(-1)^k}{k^2}$$

für  $k \in \mathbb{N}$ . Außerdem gilt

$$a_0(f) = 2\widehat{f}(0) = \frac{1}{\pi} \int_{-\pi}^{\pi} x^2 \, dx = \frac{2}{3}\pi^2.$$

Da  $\sum_{k=1}^{\infty} 1/k^2$  konvergiert, folgt mit Satz 5.6.3, dass  $f(x) = F(x, f)$ , also

$$x^2 = \frac{1}{3}\pi^2 + 4 \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} \cos kx \quad \text{für } -\pi \leq x \leq \pi.$$

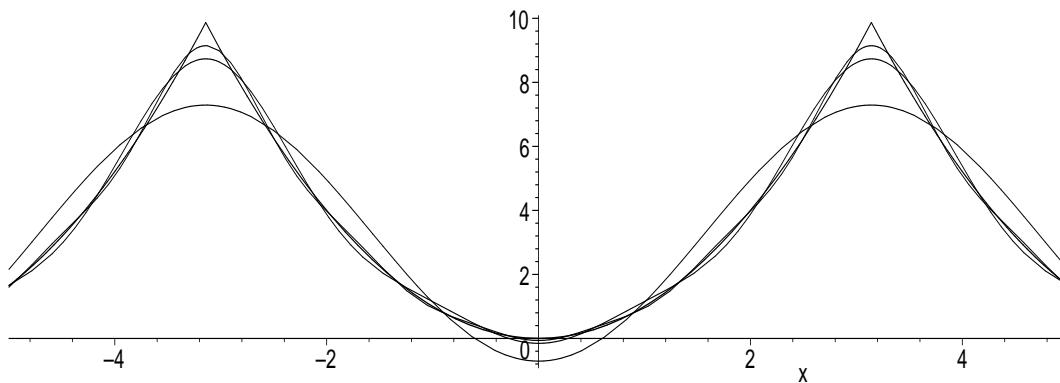


Abbildung 21: Fourierpolynome vom Grad 1, 3 und 5 für  $f(x) = x^2$

Einige Fourierpolynome dieser Funktion sind in Abbildung 21 dargestellt.

Für  $x = 0$  erhalten wir

$$0 = \frac{1}{3}\pi^2 + 4 \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}$$

und damit

$$\frac{1}{12}\pi^2 = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \frac{1}{25} - \dots$$

Für  $x = \pi$  erhalten wir

$$\pi^2 = \frac{1}{3}\pi^2 + 4 \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} (-1)^k$$

und damit

$$\frac{1}{6}\pi^2 = \sum_{k=1}^{\infty} \frac{1}{k^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \dots$$

*Beispiel 2.* Sei  $f$  die  $2\pi$ -periodische Funktion, die für  $x \in [-\pi, \pi]$  durch

$$f(x) = \begin{cases} 1 & \text{falls } 0 < x < \pi, \\ -1 & \text{falls } -\pi < x < 0, \\ 0 & \text{falls } x = 0 \text{ oder } x = \pm\pi, \end{cases}$$

gegeben ist. Eine einfache Rechnung zeigt, dass  $a_k = 0$  für alle  $k$  und

$$b_k(f) = \begin{cases} 0 & \text{falls } k \text{ gerade,} \\ \frac{4}{\pi k} & \text{falls } k \text{ ungerade.} \end{cases}$$

Man kann zeigen, dass wieder  $F(x, f) = f(x)$  für alle  $x \in \mathbb{R}$  gilt. Einige Fourierpolynome dieser Funktion sind in Abbildung 22 dargestellt.

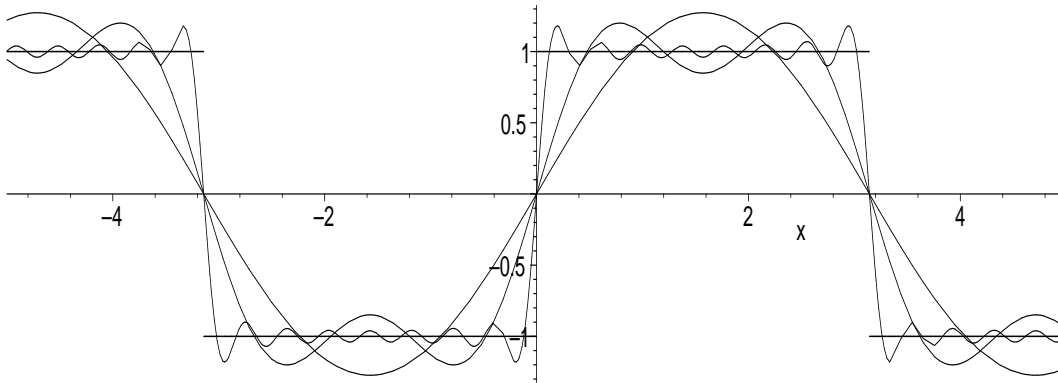


Abbildung 22: Fourierpolynome vom Grad 1, 3 und 15.

## 5.7 Uneigentliche Integrale

Bei der Definition von  $\int_I f$  haben wir vorausgesetzt, dass  $I$  kompakt ist, und aus der Existenz des Integrals konnte die Beschränktheit von  $f$  gefolgert werden (Satz 5.1.3). Wir werden jetzt  $\int_I f$  in einigen Fällen definieren, in denen  $I$  nicht kompakt oder  $f$  nicht beschränkt ist.

**Definition 5.7.1** Sei  $-\infty < a < b \leq \infty$  und  $f : [a, b) \rightarrow \mathbb{C}$ . Für alle  $c \in [a, b)$  sei  $f$  integrierbar über  $[a, c]$ . Existiert dann  $\lim_{c \rightarrow b} \int_a^c f$ , so heißt  $f$  *uneigentlich integrierbar* über  $[a, b)$ . Der Grenzwert heißt *uneigentliches Integral* von  $f$  über  $[a, b)$  und wird mit  $\int_a^b f$  bezeichnet.

Man sagt auch, dass das uneigentliche Integral *konvergiert* (bzw. *divergiert*). Analog definiert man für  $-\infty \leq a < b < \infty$  und  $f : (a, b] \rightarrow \mathbb{C}$  das uneigentliche Integral  $\int_a^b f$ . Im Falle  $f : (a, b) \rightarrow \mathbb{C}$  sagt man, dass das uneigentliche Integral  $\int_a^b f$  existiert, wenn für  $c \in (a, b)$  die beiden uneigentlichen Integrale  $\int_a^c f$  und  $\int_c^b f$  existieren. Man setzt dann  $\int_a^b f = \int_a^c f + \int_c^b f$ . (Dies hängt nicht von  $c$  ab.)

*Beispiele.* 1. Für  $\alpha \in \mathbb{R}$  und  $r > 0$  gilt

$$\int_1^r \frac{dx}{x^\alpha} = \begin{cases} \frac{1}{1-\alpha} (r^{1-\alpha} - 1) & \text{falls } \alpha \neq 1, \\ \ln r & \text{falls } \alpha = 1. \end{cases}$$

Es folgt, dass  $\int_1^\infty \frac{dx}{x^\alpha}$  für  $\alpha > 1$  konvergiert, mit  $\int_1^\infty \frac{dx}{x^\alpha} = \frac{1}{\alpha-1}$ , während  $\int_1^\infty \frac{dx}{x^\alpha}$  für  $\alpha \leq 1$  divergiert.

2. Wir können in obigem Beispiel auch den Grenzwert für  $r \rightarrow 0$  betrachten und erhalten, dass  $\int_0^1 \frac{dx}{x^\alpha}$  für  $\alpha < 1$  konvergiert und den Wert  $\frac{1}{1-\alpha}$  hat, während das Integral für  $\alpha \geq 1$  divergiert.

3. Es ist

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dx}{1+x^2} &= \lim_{a \rightarrow -\infty} \int_a^0 \frac{dx}{1+x^2} + \lim_{b \rightarrow \infty} \int_0^b \frac{dx}{1+x^2} \\ &= \lim_{a \rightarrow -\infty} (-\arctan a) + \lim_{b \rightarrow \infty} \arctan b \end{aligned}$$



$$\begin{aligned}
&= \frac{\pi}{2} + \frac{\pi}{2} \\
&= \pi.
\end{aligned}$$

#### 4. Das uneigentliche Integral

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2} dx$$

existiert nicht, denn

$$\lim_{r \rightarrow \infty} \int_0^r \frac{x}{1+x^2} dx = \lim_{r \rightarrow \infty} \frac{1}{2} \ln(1+r^2) = \infty,$$

womit

$$\int_0^{\infty} \frac{x}{1+x^2} dx$$

nicht existiert. (Auch

$$\int_{-\infty}^0 \frac{x}{1+x^2} dx$$

existiert nicht. Man beachte aber, dass

$$\lim_{r \rightarrow \infty} \int_{-r}^r \frac{x}{1+x^2} dx = 0$$

gilt.)

**Satz 5.7.1 (Cauchykriterium für uneigentliche Integrale)** Sei  $-\infty < a < b < \infty$  und  $f : [a, b) \rightarrow \mathbb{C}$ . Für alle  $c \in [a, b)$  sei  $f$  integrierbar über  $[a, c]$ . Dann konvergiert  $\int_a^b f$  genau dann, wenn für alle  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass für alle  $x, y$  aus dem Intervall  $(b - \delta, b)$  die Abschätzung  $|\int_x^y f| < \varepsilon$  gilt.

In Quantorenschreibweise lautet die Bedingung aus dem Cauchykriterium wie folgt:

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in (b - \delta, b) : \left| \int_x^y f \right| < \varepsilon.$$

*Beweis.* Man wende das Cauchykriterium für Funktionengrenzwerte (Satz 3.3.4) auf  $F(x) := \int_a^x f$  und beachte, dass  $F(y) - F(x) = \int_x^y f$ .  $\square$

Ist  $b = \infty$  in Satz 5.7.1, so lautet das Cauchykriterium:  $\int_a^\infty f$  konvergiert genau dann, wenn

$$\forall \varepsilon > 0 \exists r > a \forall x, y \in (r, \infty) : \left| \int_x^y f \right| < \varepsilon.$$

**Satz 5.7.2 (Vergleichskriterium für uneigentliche Integrale)** Seien  $-\infty < a < b \leq \infty$ ,  $f : [a, b) \rightarrow \mathbb{C}$  und  $g : [a, b) \rightarrow \mathbb{R}$ . Für alle  $c \in [a, b)$  seien  $f$  und  $g$  integrierbar über  $[a, c]$  und es gelte  $|f(x)| \leq g(x)$  für alle  $x \in \mathbb{R}$ . Konvergiert dann  $\int_a^b g$ , so konvergieren auch  $\int_a^b |f|$  und  $\int_a^b f$ .

*Beweis.* Wegen

$$\left| \int_x^y f \right| \leq \int_x^y |f| \leq \int_x^y g$$

für alle  $x, y \in (a, b)$  folgt die Behauptung direkt aus dem Cauchy Kriterium.  $\square$

Durch Kontraposition kann man mit dem Vergleichskriterium natürlich auch Divergenz nachweisen.

Die Sätze 5.7.1 und 5.7.2 und die Bemerkungen dazu gelten analog für Funktionen  $f : (a, b] \rightarrow \mathbb{C}$ ,  $-\infty \leq a < b < \infty$ .

*Beispiel.* Wegen  $|\sin t| \leq 1$  für  $t \in \mathbb{R}$  konvergiert

$$\int_1^\infty \frac{\sin t}{t^\alpha} dt$$

nach Vergleichskriterium und Beispiel 1 nach Definition 5.7.1 für  $\alpha > 1$ .

Tatsächlich konvergiert das Integral sogar für  $\alpha > 0$ , denn für  $y > x > (2/\varepsilon)^{1/\alpha} > 1$  folgt mit partieller Integration

$$\begin{aligned} \left| \int_x^y \frac{\sin t}{t^\alpha} dt \right| &= \left| -\frac{\cos t}{t^\alpha} \Big|_x^y - \alpha \int_x^y \frac{\cos t}{t^{\alpha+1}} dt \right| \\ &\leq \frac{|\cos y|}{y^\alpha} + \frac{|\cos x|}{x^\alpha} + \alpha \int_x^y \frac{|\cos t|}{t^{\alpha+1}} dt \\ &\leq \frac{1}{y^\alpha} + \frac{1}{x^\alpha} + \alpha \int_x^y \frac{dt}{t^{\alpha+1}} \\ &= \frac{2}{x^\alpha} \\ &< \varepsilon, \end{aligned}$$

und die Konvergenz folgt mit dem Cauchy Kriterium. Man kann hieraus folgern, dass

$$\int_0^\infty \sin t^2 dt$$

konvergiert, denn für  $0 < x < y < \infty$  folgt mit der Substitutionsregel

$$\int_x^y \sin t^2 dt = \frac{1}{2} \int_{\sqrt{x}}^{\sqrt{y}} \frac{\sin s}{\sqrt{s}} ds.$$

**Satz 5.7.3 (Integralkriterium für Reihen)** Sei  $f : [1, \infty) \rightarrow (0, \infty)$  monoton fallend. Dann konvergiert  $\int_1^\infty f(x) dx$  genau dann, wenn  $\sum_{k=1}^\infty f(k)$  konvergiert.

*Beweis.* Nach Satz 5.1.12 ist  $f$  über kompakte Teilintervalle von  $[1, \infty)$  integrierbar. Für  $k \in \mathbb{N}$  gilt

$$f(k+1) \leq \int_k^{k+1} f(x) dx \leq f(k)$$

und damit folgt

$$\sum_{k=m+1}^n f(k) \leq \int_m^n f(x) dx \leq \sum_{k=m}^{n-1} f(k)$$

für  $m, n \in \mathbb{N}$ ,  $n > m$ . Die Behauptung folgt aus dem Cauchy Kriterium.  $\square$

*Beispiel.* 1. Für  $\alpha > 1$  konvergiert

$$\sum_{k=2}^{\infty} \frac{1}{k(\ln k)^{\alpha}},$$

da

$$\int_2^r \frac{dx}{x(\ln x)^{\alpha}} = -\frac{1}{(\alpha+1)(\ln x)^{\alpha+1}} \Big|_2^r.$$

2. Die Reihe

$$\sum_{k=2}^{\infty} \frac{1}{k \ln k}$$

divergiert, da

$$\int_2^r \frac{dx}{x \ln x} = \ln \ln x \Big|_2^r.$$

## 6 Topologie metrischer Räume

### 6.1 Metrik, Norm und Skalarprodukt

Bisher haben wir Funktionen  $f : M \rightarrow N$  mit  $M, N \subset \mathbb{R}, \mathbb{C}$  betrachtet. Jetzt werden wir allgemeinere Mengen  $M, N$  zulassen, etwa  $\mathbb{R}^n$  oder  $\mathbb{C}^n$  mit  $n \in \mathbb{N}$ , oder auch Mengen von Funktionen. Wichtige Beispiele sind hier  $C^n(I)$  und  $C(I)$  mit einem Intervall  $I$ . Die genannten Mengen bilden Vektorräume. (Der Begriff des Vektorraums wird als aus der Linearen Algebra bekannt vorausgesetzt.) Wir benötigen aber noch weitere "Strukturen" auf diesen Mengen.

**Definition 6.1.1** Sei  $M$  nichtleere Menge. Eine Funktion  $d : M \times M \rightarrow [0, \infty)$  heißt *Metrik*, wenn für alle  $x, y, z \in M$  die folgenden drei Bedingungen erfüllt sind:

- (i)  $d(x, y) = 0 \Leftrightarrow x = y$ ,
- (ii)  $d(x, y) = d(y, x)$  (*Symmetrie*)
- (iii)  $d(x, z) \leq d(x, y) + d(y, z)$  (*Dreiecksungleichung*)

Das Paar  $(M, d)$  heißt dann *metrischer Raum*.

*Beispiele.* 1.  $M = \mathbb{R}$  oder  $M = \mathbb{C}$ ,  $d(x, y) = |x - y|$ . Wir nennen  $d$  die *euklidische Metrik* und  $d(x, y)$  den *euklidischen Abstand* von  $x$  und  $y$ .

2.  $M \neq \emptyset$  beliebig,

$$d(x, y) = \begin{cases} 1 & \text{falls } x \neq y, \\ 0 & \text{falls } x = y. \end{cases}$$

Dies nennt man die *diskrete Metrik* auf  $M$ .

3.  $M = \mathbb{C}$ ,

$$d(x, y) = \begin{cases} |x - y| & \text{falls } \lambda, \mu \in \mathbb{R} \text{ mit } \lambda x + \mu y = 0 \text{ existieren,} \\ |x| + |y| & \text{sonst.} \end{cases}$$

Liegen die Punkte  $x, y$  auf einer Geraden durch 0, so ist  $d(x, y)$  also der euklidische Abstand. Andernfalls ist  $d(x, y)$  die Summe der Abstände zu 0. (Aus naheliegenden Gründen nennt man das auch "französische Eisenbahnmetrik".)

Weitere wichtige Beispiele metrischer Räume erhält man durch folgende Definition. Hier und im folgenden sei  $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ .

**Definition 6.1.2** Sei  $V$  ein Vektorraum über  $\mathbb{K}$ . Eine Funktion  $\|\cdot\| : V \rightarrow [0, \infty)$  (also  $x \mapsto \|x\|$ ) heißt *Norm*, wenn für alle  $x, y \in V$  und  $\lambda \in \mathbb{K}$  die folgenden drei Bedingungen erfüllt sind:

- (i)  $\|x\| = 0 \Leftrightarrow x = 0$ ,
- (ii)  $\|\lambda \cdot x\| = |\lambda| \cdot \|x\|$ ,
- (iii)  $\|x + y\| \leq \|x\| + \|y\|$  (*Dreiecksungleichung*)

Das Paar  $(V, \|\cdot\|)$  heißt dann *normierter Raum*.

**Satz 6.1.1** Sei  $(V, \|\cdot\|)$  normierter Raum. Dann ist die durch  $d(x, y) := \|x - y\|$  definierte Funktion  $d : V \times V \rightarrow [0, \infty)$  eine Metrik auf  $V$ , d. h.,  $(V, d)$  ist metrischer Raum.

*Beweis.* Seien  $x, y, z \in V$ . Dann gilt

$$d(x, y) = 0 \Leftrightarrow \|x - y\| = 0 \Leftrightarrow x - y = 0 \Leftrightarrow x = y,$$

$$d(x, y) = \|x - y\| = \|(-1)(y - x)\| = |(-1)| \cdot \|(y - x)\| = \|(y - x)\| = d(y, x)$$

und

$$d(x, z) = \|x - z\| = \|x - y + y - z\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z). \quad \square$$

*Beispiele.* 1.  $V = \mathbb{K}$ ,  $\|x\| = |x|$  für  $x \in V$ .

2.  $V = \mathbb{K}^p$  mit  $p \in \mathbb{N}$ ,  $\|x\|_\infty := \max_j |x_j|$  für

$$x = (x_1, x_2, \dots, x_p)^T := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$$

Wir werden Vektoren im Vektorraum  $\mathbb{K}^p$  als Spaltenvektoren schreiben. Hier steht “ $T$ ” für “transponiert”. Es ist  $(V, \|\cdot\|_\infty)$  normierter Raum.

3.  $V = \mathbb{K}^p$  mit  $p \in \mathbb{N}$ ,  $\|x\|_1 := \sum_{j=1}^p |x_j|$  für  $x = (x_1, \dots, x_p)^T \in V$ . Dann ist  $(V, \|\cdot\|_1)$  normierter Raum.

4.  $V = C[a, b]$ ,  $\|f\|_\infty := \max_{x \in [a, b]} |f(x)|$  für  $f \in V$ . Dann ist  $(V, \|\cdot\|_\infty)$  normierter Raum. Die zugehörige Metrik  $d$  ist durch  $d(f, g) := \max_{x \in [a, b]} |f(x) - g(x)|$  gegeben.

5.  $V = C[a, b]$ ,  $\|f\|_1 := \int_a^b |f|$  für  $f \in V$ . Dann ist  $(V, \|\cdot\|_1)$  normierter Raum. Die zugehörige Metrik  $d_1$  ist durch  $d_1(f, g) := \int_a^b |f - g|$  gegeben.

Die einzige Bedingung aus Definition 6.1.2, die vielleicht nicht offensichtlich ist, ist “ $\Leftarrow$ ” in (i). Wir zeigen dies durch Kontraposition. Sei dazu  $f \in C[a, b]$ ,  $f \neq 0$ . Dann existiert  $\xi \in [a, b]$  mit  $f(\xi) \neq 0$ . Da  $f$  stetig, existiert  $\delta > 0$ , so dass  $|f(x) - f(\xi)| < \varepsilon := \frac{1}{2}|f(\xi)|$  für  $|x - \xi| < \delta$ ,  $x \in [a, b]$ . Mit  $I := [a, b] \cap [\xi - \delta, \xi + \delta]$  folgt  $|f(x)| \geq |f(\xi)| - |f(x) - f(\xi)| > |f(\xi)| - \varepsilon = \varepsilon$  für  $x \in I$  und damit  $\|f\|_1 = \int_a^b |f| \geq \int_I |f| \geq \varepsilon |I| > 0$ .

6. Sei  $V$  der Vektorraum der Riemann-integrierbaren Funktionen von  $[a, b]$  nach  $\mathbb{C}$  und sei  $\|f\|_1 := \int_a^b |f|$  für  $f \in V$ . Dann ist  $(V, \|\cdot\|_1)$  kein normierter Raum, denn Bedingung (i) aus Definition 6.1.2 gilt nicht, wie das Beispiel

$$f(x) = \begin{cases} 1 & \text{falls } x = a, \\ 0 & \text{falls } a < x \leq b, \end{cases}$$

zeigt.

**Definition 6.1.3** Sei  $V$  ein Vektorraum über  $\mathbb{K}$ . Eine Funktion  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$  (also  $(x, y) \mapsto \langle x, y \rangle$ ) heißt *Skalarprodukt* (oder *inneres Produkt*), wenn für alle  $x, y, z \in V$  und  $\lambda \in \mathbb{K}$  die folgenden fünf Bedingungen erfüllt sind:

- (i)  $\langle x, y \rangle = \overline{\langle y, x \rangle}$ ,
- (ii)  $\langle x, x \rangle \geq 0$ ,
- (iii)  $\langle x, x \rangle = 0 \Leftrightarrow x = 0$ ,
- (iv)  $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ ,
- (v)  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ .

Dabei steht der Querstrich in (i) für komplexe Konjugation. (Im Falle  $\mathbb{K} = \mathbb{R}$  kann also darauf verzichtet werden.) Das Paar  $(V, \langle \cdot, \cdot \rangle)$  heißt *Prähilbertraum*.

Statt  $\langle x, y \rangle$  sind auch die Schreibweisen  $x \cdot y$  und  $(x, y)$  üblich.

**Satz 6.1.2** Sei  $(V, \langle \cdot, \cdot \rangle)$  Prähilbertraum. Dann ist durch  $\|x\| := \sqrt{\langle x, x \rangle}$  eine Norm auf  $V$  definiert, d. h.,  $(V, \|\cdot\|)$  ist normierter Raum.

Wir stellen den Beweis von Satz 6.1.2 noch etwas zurück und betrachten zunächst zwei Beispiele.

*Beispiel 1.* Sei  $V = \mathbb{K}^p$  mit  $p \in \mathbb{N}$ . Man rechnet leicht nach, dass mit  $\langle x, y \rangle := \sum_{j=1}^p x_j \overline{y_j}$  für  $x = (x_1, \dots, x_n)^T$ ,  $y = (y_1, \dots, y_n)^T \in V$  ein Skalarprodukt gegeben ist. (Dieses Skalarprodukt ist für  $\mathbb{K} = \mathbb{R}$  das aus der Schule bekannte Skalarprodukt.) Die zugehörige Norm heißt *euklidische Norm* und wird mit  $\|\cdot\|_2$  bezeichnet, also  $\|x\|_2 = \sqrt{\sum_{j=1}^p x_j \overline{x_j}} = \sqrt{\sum_{j=1}^p |x_j|^2}$ . Die zugehörige Metrik heißt *euklidische Metrik*.

*Beispiel 2.* Sei  $V = C[a, b]$  und  $\langle f, g \rangle := \int_a^b f \overline{g}$ . Die Skalarprodukteigenschaften sind einfach nachzurechnen; für (iii) vgl. Beispiel 5 nach Satz 6.1.1.

Wir ziehen einige einfache Folgerungen aus den Eigenschaften (i)-(v) von Definition 6.1.3. Dabei seien wieder  $x, y, z \in V$  und  $\lambda \in \mathbb{K}$ :

- (a)  $\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle$ , denn  $\langle x, \lambda y \rangle = \overline{\langle \lambda y, x \rangle} = \overline{\lambda \langle y, x \rangle} = \overline{\lambda} \overline{\langle y, x \rangle} = \overline{\lambda} \langle x, y \rangle$ ,
- (b)  $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ ,
- (c)  $\langle 0, x \rangle = \langle x, 0 \rangle = 0$ , denn  $\langle 0, x \rangle = \langle 0 - 0, x \rangle = \langle 0, x \rangle - \langle 0, x \rangle = 0$ .

Drei weitere wichtige Eigenschaften fassen wir in folgendem Satz zusammen.

**Satz 6.1.3** Sei  $(V, \langle \cdot, \cdot \rangle)$  Prähilbertraum,  $x, y \in V$ . Dann gilt (mit  $\|\cdot\|$  wie in Satz 6.1.2)

- (i)  $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2 \operatorname{Re} \langle x, y \rangle$  (*Satz des Pythagoras*).
- (ii)  $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$  (*Parallelogrammidentität*).
- (iii)  $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$  (*Cauchy-Schwarz-Ungleichung*).

*Beweis.* Die Beweise von (i) und (ii) erfolgen durch einfaches Ausrechnen des Ausdrucks  $\|x + y\|^2 = \langle x + y, x + y \rangle$  mit Hilfe von Bedingung (iv) und (v) aus Definition 6.1.3 sowie obiger Rechenregeln (b) und (c). Sie werden hier weggelassen.

Zum Beweis von (iii) notieren wir, dass für  $\lambda \in \mathbb{K}$

$$0 \leq \langle x - \lambda y, x - \lambda y \rangle = \langle x, x \rangle - \lambda \langle y, x \rangle - \bar{\lambda} \langle x, y \rangle + \lambda \bar{\lambda} \langle y, y \rangle$$

gilt. Mit  $\lambda = \langle x, y \rangle / \langle y, y \rangle = \langle x, y \rangle / \|y\|^2$  folgt

$$0 \leq \langle x, x \rangle - \frac{\langle x, y \rangle \cdot \langle y, x \rangle}{\langle y, y \rangle} = \|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^2}$$

und hieraus die Behauptung.  $\square$

*Beweis von Satz 6.1.2.* Seien  $x, y \in V$  und  $\lambda \in \mathbb{K}$ . Dann gilt

$$\|x\| = 0 \Leftrightarrow \langle x, x \rangle = 0 \Leftrightarrow x = 0$$

und

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x \rangle} = \sqrt{\lambda \bar{\lambda} \langle x, x \rangle} = \sqrt{|\lambda|^2 \langle x, x \rangle} = |\lambda| \cdot \|x\|.$$

Mit der Cauchy-Schwarz-Ungleichung und dem Satz des Pythagoras gilt

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2 \operatorname{Re} \langle x, y \rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2,$$

also die Dreiecksungleichung.  $\square$

Eine Analyse des Beweises zeigt, dass für  $x, y \neq 0$  in der Cauchy-Schwarz-Ungleichung genau dann Gleichheit gilt, wenn  $\lambda \in \mathbb{K}$  mit  $x = \lambda y$  existiert. (Es gilt dann  $\lambda = \langle x, y \rangle / \langle y, y \rangle$ .) Gleichheit in der Dreiecksungleichung gilt genau dann, wenn  $\lambda > 0$  mit  $x = \lambda y$  existiert.

## 6.2 Konvergenz

**Definition 6.2.1** Sei  $(M, d)$  metrischer Raum und  $(a_n)$  Folge in  $M$ . Dann heißt  $(a_n)$  *konvergent*, falls  $a \in M$  mit folgender Eigenschaft existiert:

Zu jedem  $\varepsilon > 0$  existiert ein  $N \in \mathbb{N}$ , so dass  $d(a_n, a) < \varepsilon$  für alle  $n \in \mathbb{N}$  mit  $n \geq N$ .

Dieses  $a$  heißt dann *Grenzwert* der Folge  $(a_n)$  und wir schreiben  $a = \lim_{n \rightarrow \infty} a_n$  oder  $a_n \rightarrow a$  (für  $n \rightarrow \infty$ ).

Weiter heißt  $b \in M$  *Häufungswert* von  $(a_n)$ , falls für jedes  $\varepsilon > 0$  unendlich viele  $n \in \mathbb{N}$  mit  $d(a_n, b) < \varepsilon$  existieren.

Diese Definition ist völlig analog zu den Definitionen 2.1.2 und 2.3.1 in Analysis I. Viele der dort gemachten Aussagen über Grenzwerte und Häufungswerte übertragen sich leicht. So gilt wieder die Aussage von Satz 2.1.1: *Eine konvergente Folge hat genau einen Grenzwert.* Es gilt auch wieder die Aussage von Satz 2.3.2:

- (i)  $a$  ist Häufungswert von  $(a_n) \Leftrightarrow$  Es existiert eine Teilfolge von  $(a_n)$ , die gegen  $a$  konvergiert.
- (ii)  $a$  ist Grenzwert von  $(a_n) \Leftrightarrow$  Alle Teilfolgen von  $(a_n)$  konvergieren gegen  $a$ .

*Beispiel.* Sei  $V := C[a, b]$  und  $d$  die von der Norm  $\|\cdot\|_\infty$  erzeugte Metrik, also  $d(f, g) = \|f - g\|_\infty = \max_{x \in [a, b]} |f(x) - g(x)|$  für  $f, g \in V$ . Sei  $(f_n)$  Folge in  $V$  und  $f \in V$ . Dann gilt

$$\begin{aligned} & f_n \rightarrow f \text{ (bzgl. der von } \|\cdot\|_\infty \text{ erzeugten Metrik in } V) \\ \Leftrightarrow & \forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n \geq N : \|f_n - f\|_\infty < \varepsilon \\ \Leftrightarrow & \forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n \geq N \forall x \in [a, b] : |f_n(x) - f(x)| < \varepsilon \\ \Leftrightarrow & f_n \rightarrow f \text{ gleichmäßig} \end{aligned}$$

Die folgende Definition ist analog zu Definition 2.3.3.

**Definition 6.2.2** Sei  $(M, d)$  metrischer Raum und  $(x_n)$  Folge in  $M$ . Dann heißt  $(x_n)$  *Cauchyfolge*, wenn zu jedem  $\varepsilon > 0$  ein  $N \in \mathbb{R}$  existiert, so dass für alle  $m, n \in \mathbb{N}$  mit  $m \geq N$  und  $n \geq N$  gilt, dass  $d(x_m, x_n) < \varepsilon$  ist.

Vom Cauchy Kriterium für Folgenkonvergenz (Satz 2.3.5) gilt aber nur “eine Hälfte”.

**Satz 6.2.1** *Konvergente Folgen sind Cauchyfolgen.*

*Beweis* (vgl. Beweis zu Satz 2.3.5). Sei  $(M, d)$  metrischer Raum und sei  $(x_n)$  konvergente Folge in  $M$ , etwa  $x_n \rightarrow \xi$ . Sei  $\varepsilon > 0$ . Dann existiert  $N \in \mathbb{N}$  mit  $d(x_n, \xi) < \frac{\varepsilon}{2}$  für  $n \geq N$ . Für  $m, n \in \mathbb{N}$  mit  $m, n \geq N$  folgt dann  $d(x_m, x_n) \leq d(x_m, \xi) + d(\xi, x_n) < \varepsilon$ .  $\square$

Die Umkehrung gilt nicht, wie etwa das Beispiel  $M = \mathbb{Q}$  (mit der euklidischen Metrik) zeigt, und wie wir später an weiteren Beispielen sehen werden.

**Definition 6.2.3** Ein metrischer Raum heißt *vollständig*, wenn jede Cauchyfolge in ihm konvergiert. Ein vollständiger normierter Raum heißt *Banachraum* und ein vollständiger Prähilbertraum heißt *Hilbertraum*.

*Beispiele.* 1.  $(C[a, b], \|\cdot\|_\infty)$  ist vollständig. Denn sei  $(f_n)$  Cauchyfolge in  $C[a, b]$ , d.h., (vgl. das Beispiel nach Definition 6.2.1)

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall x \in [a, b] \forall m, n \in \mathbb{N} : m \geq N \wedge n \geq N \Rightarrow |f_m(x) - f_n(x)| < \varepsilon.$$

Nach dem Cauchy Kriterium für Funktionenfolgen (Satz 3.7.1) folgt, dass  $(f_n)$  gleichmäßig gegen eine Funktion  $f : [a, b] \rightarrow \mathbb{C}$  konvergiert. Nach Satz 3.7.2 ist  $f$  stetig, also  $f \in C[a, b]$ . Es folgt, dass  $(f_n)$  in  $(C[a, b], \|\cdot\|_\infty)$  gegen  $f$  konvergiert.

2.  $(C[a, b], \|\cdot\|_1)$  ist nicht vollständig. Ohne Beschränkung der Allgemeinheit sei  $[a, b] = [-1, 1]$ . Wir betrachten die durch

$$f(x) = \begin{cases} 0 & \text{falls } -1 \leq x \leq 0, \\ nx & \text{falls } 0 < x \leq \frac{1}{n}, \\ 1 & \text{falls } \frac{1}{n} \leq x \leq 1 \end{cases}$$

definierte Folge  $(f_n)$ . Für  $m > n$  gilt

$$\|f_m - f_n\|_1 = \int_{-1}^1 |f_m - f_n| \leq \int_0^{\frac{1}{n}} |f_m - f_n| \leq \frac{1}{n}$$



und damit ist  $(f_n)$  Cauchyfolge. Die Folge  $(f_n)$  ist aber nicht konvergent. Dies ist intuitiv klar, da der der einzige "Kandidat"  $f$  für eine Grenzfunktion  $f(x) = 0$  für  $x < 0$  und  $f(x) = 1$  für  $x > 0$  erfüllen sollte, und damit in 0 nicht stetig sein kann.

Für einen formalen Beweis nehmen wir an, dass  $(f_n)$  konvergent ist, also  $f_n \rightarrow f$  für ein  $f \in C[-1, 1]$ . Es existiert  $\delta \in (0, 1]$  mit  $|f(x) - f(0)| \leq \frac{1}{4}$  für  $|x| < \delta$ . Falls  $|f(0)| \leq \frac{1}{2}$  folgt

$$\|f - f_n\|_1 \geq \int_{\frac{1}{n}}^{\delta} |f(x) - 1| dx \geq \int_{\frac{1}{n}}^{\delta} 1 - |f(0)| - |f(x) - f(0)| dx \geq \int_{\frac{1}{n}}^{\delta} \frac{1}{4} dx = \frac{\delta - \frac{1}{n}}{4}$$

für  $n \geq \frac{1}{\delta}$  und im Falle  $|f(0)| > \frac{1}{2}$  gilt

$$\|f - f_n\|_1 \geq \int_{-\delta}^0 |f(x)| dx \geq \int_{-\delta}^0 \frac{1}{4} dx = \frac{\delta}{4}.$$

In beiden Fällen erhält man einen Widerspruch.

3.  $(\mathbb{K}^p, \|\cdot\|_{\infty})$  ist vollständig (für  $p \in \mathbb{N}$ ). Denn sei  $(x^k)$  Cauchyfolge in  $\mathbb{K}^p$ , mit  $x^k = (x_1^k, \dots, x_p^k)^T$ . Wegen  $|x_j^k - x_j^{\ell}| \leq \|x^k - x^{\ell}\|_{\infty}$  für  $k, \ell \in \mathbb{N}$  und  $j \in \{1, \dots, p\}$  folgt, dass  $(x_j^k)_{k \in \mathbb{N}}$  für jedes  $j \in \{1, \dots, p\}$  eine Cauchyfolge ist, und damit also konvergent, etwa  $x_j^k \rightarrow \xi_j$ . Mit  $\xi = (\xi_1, \dots, \xi_p)$  folgt dann

$$\|x^k - \xi\|_{\infty} = \max_j |x_j^k - \xi_j| \rightarrow 0,$$

also  $x^k \rightarrow \xi$  für  $k \rightarrow \infty$ .

Genauso kann man zeigen, dass  $(\mathbb{K}^p, \|\cdot\|_1)$  und  $(\mathbb{K}^p, \|\cdot\|_2)$  vollständig sind. Dies kann man sich aber auch an Hand des folgenden Begriffes überlegen.

**Definition 6.2.4** Sei  $V$  Vektorraum über  $\mathbb{K}$  und seien  $\|\cdot\|'$  und  $\|\cdot\|''$  Normen auf  $V$ . Dann heißen  $\|\cdot\|'$  und  $\|\cdot\|''$  *äquivalent*, falls  $\alpha, \beta > 0$  existieren, so dass  $\alpha\|x\|' \leq \|x\|'' \leq \beta\|x\|'$  für alle  $x \in V$  gilt.

Sind die Normen  $\|\cdot\|'$  und  $\|\cdot\|''$  auf dem Vektorraum  $V$  äquivalent, so ist eine Folge  $(x_n)$  in  $V$  offensichtlich genau dann konvergent (bzw. Cauchyfolge) in  $(V, \|\cdot\|')$ , wenn sie konvergent (bzw. Cauchyfolge) in  $(V, \|\cdot\|'')$  ist. Leicht rechnet man nun nach, dass in  $\mathbb{K}^p$  die Normen  $\|\cdot\|_1$  und  $\|\cdot\|_2$  äquivalent zur Norm  $\|\cdot\|_{\infty}$  sind. Hieraus folgt, dass mit  $(\mathbb{K}^p, \|\cdot\|_{\infty})$  auch  $(\mathbb{K}^p, \|\cdot\|_1)$  und  $(\mathbb{K}^p, \|\cdot\|_2)$  vollständig sind.

Schließlich notieren wir noch das folgende Analogon zu Definition 3.3.1.

**Definition 6.2.5** Sei  $(M, d)$  metrischer Raum,  $A \subset M$  und  $\xi \in M$ . Dann heißt  $\xi$  *Häufungspunkt* von  $A$ , falls eine Folge  $(a_n)$  in  $A \setminus \{\xi\}$  mit  $a_n \rightarrow \xi$  existiert.

## 6.3 Offene und abgeschlossene Mengen

**Definition 6.3.1** Sei  $(M, d)$  metrischer Raum und  $x \in M$ . Für  $\varepsilon > 0$  heißt dann  $U(x, \varepsilon) := \{y \in M : d(x, y) < \varepsilon\}$  die  $\varepsilon$ -Umgebung von  $x$ . Sei weiter  $A \subset M$ . Existiert  $\varepsilon > 0$  mit  $U(x, \varepsilon) \subset A$ , so heißt  $A$  *Umgebung* von  $x$ . Weiter heißt  $A$  *offen*, falls  $A$  Umgebung jedes Punktes von  $A$  ist, und  $A$  heißt *abgeschlossen*, falls  $M \setminus A$  offen ist.

Die Bezeichnungen für die  $\varepsilon$ -Umgebung sind in der Literatur sehr uneinheitlich, statt  $U(a, \varepsilon)$  findet man auch  $B(a, \varepsilon)$ ,  $K(a, \varepsilon)$ ,  $D(a, \varepsilon)$ ,  $U_\varepsilon(a)$ ,  $B_\varepsilon(a)$ ,  $K_\varepsilon(a)$  und anderes.

Sei  $(M, d)$  metrischer Raum und  $N \subset M$ . Dann ist auch  $(N, d|_{N \times N})$  metrischer Raum. Für  $A \subset N$  können wir Offenheit und Abgeschlossenheit außer in  $(M, d)$  jetzt auch in  $(N, d|_{N \times N})$  betrachten. Wir sagen dann, dass  $A$  *relativ offen* (bzw. *abgeschlossen*) in  $N$  ist. Den Zusatz "relativ" lassen wir dabei in der Regel weg.

Aus der Offenheit von  $A$  in  $M$  folgt die Offenheit von  $A$  in  $N$ . Allgemeiner gilt:

**Satz 6.3.1** *Sei  $(M, d)$  metrischer Raum und  $A \subset N \subset M$ . Dann ist  $A$  offen in  $N$  genau dann, wenn ein (in  $M$ ) offenes  $B \subset M$  mit  $A = B \cap N$  existiert.*

*Beweis.* Wir bezeichnen die  $\varepsilon$ -Umgebungen in  $M$  bzw.  $N$  mit  $U_M(\cdot, \varepsilon)$  bzw.  $U_N(\cdot, \varepsilon)$ . Dann gilt  $U_N(x, \varepsilon) = U_M(x, \varepsilon) \cap N$  für alle  $x \in N$ .

Sei nun  $B \subset M$  offen in  $M$  und  $A := B \cap N$ . Ist  $x \in A$ , so ist auch  $x \in B$  und damit existiert  $\varepsilon > 0$  mit  $U_M(x, \varepsilon) \subset B$ . Es folgt

$$U_N(x, \varepsilon) = U_M(x, \varepsilon) \cap N \subset B \cap N = A.$$

Also ist  $A$  offen in  $N$ .

Ist umgekehrt  $A \subset N$  offen in  $N$ , so existiert für jedes  $x \in A$  ein  $\varepsilon_x > 0$  mit  $U_N(\varepsilon_x, x) \subset A$ . Hieraus erhält man  $A = \bigcup_{x \in A} U_N(\varepsilon_x, x)$  und damit leistet  $B := \bigcup_{x \in A} U_M(\varepsilon_x, x)$  das Verlangte.

**Satz 6.3.2** (i) *Vereinigungen offener Mengen sind offen und Durchschnitte von endlich vielen offenen Mengen sind offen.*

(ii) *Durchschnitte abgeschlossener Mengen sind abgeschlossen und Vereinigungen von endlich vielen abgeschlossenen Mengen sind abgeschlossen.*

*Beweis.* (i). Sei  $M$  Menge offener Mengen und  $V = \bigcup_{U \in M} U$ . Zu zeigen ist, dass  $V$  offen ist. Sei dazu  $\xi \in V$ . Dann existiert  $U \in M$  mit  $\xi \in U$ . Damit ist  $U$  Umgebung von  $\xi$ , und wegen  $V \supset U$  also auch  $V$  Umgebung von  $\xi$ .

Seien jetzt  $U_1, \dots, U_n$  offen und  $V = \bigcap_{j=1}^n U_j$ . Sei  $\xi \in V$ . Dann ist  $\xi \in U_j$  für alle  $j \in \{1, \dots, n\}$  und damit existiert  $\varepsilon_j > 0$  mit  $U(\xi, \varepsilon_j) \subset U_j$ . Wir definieren  $\varepsilon := \min\{\varepsilon_1, \dots, \varepsilon_n\}$ . Es folgt dann  $U(\xi, \varepsilon) \subset U_j$  für alle  $j \in \{1, \dots, n\}$ , also  $U(\xi, \varepsilon) \subset \bigcap_{j=1}^n U_j = V$ . Damit ist  $V$  offen.

(ii). Dies folgt aus (i) und den Regeln von de Morgan.  $\square$

*Beispiele.* 1. Für  $a, b \in \mathbb{R}$  mit  $a < b$  ist das offene Intervall  $(a, b)$  offen in  $\mathbb{R}$ , aber nicht offen in  $\mathbb{C}$ . Denn bezeichnen wir mit  $U(x, \varepsilon)$  die  $\varepsilon$ -Umgebung in  $\mathbb{C}$ , so gilt folgt mit  $\varepsilon := \min\{x - a, b - x\}$ , dass  $U(x, \varepsilon) \cap \mathbb{R} = (x - \varepsilon, x + \varepsilon) \subset (a, b)$ . Andererseits ist  $(a, b)$  nicht offen in  $\mathbb{C}$ , da für  $x \in (a, b)$  und  $\varepsilon > 0$  ja  $x + \frac{1}{2}\varepsilon i \in U(x, \varepsilon) \setminus (a, b)$  gilt. Entsprechendes gilt für Intervalle  $(-\infty, b)$  und  $(a, \infty)$ .

2. Abgeschlossene Intervalle sind abgeschlossen in  $\mathbb{R}$  (und damit auch in  $\mathbb{C}$ ).

**Satz 6.3.3** *Sei  $(M, d)$  metrischer Raum und  $A \subset M$ . Dann sind äquivalent:*

(i)  *$A$  ist abgeschlossen.*

(ii) *Für jede konvergente Folge in  $A$  ist auch ihr Grenzwert in  $A$ .*

(iii)  $A$  enthält alle seine Häufungspunkte.

*Beweis.* (i)  $\Rightarrow$  (ii): Sei  $A$  abgeschlossen und  $(x_n)$  Folge in  $A$  mit  $x_n \rightarrow \xi$ . Zu zeigen ist, dass  $\xi \in A$ . Dazu nehmen wir an, dass  $\xi \in M \setminus A$ . Da  $M \setminus A$  offen ist, existiert  $\varepsilon > 0$  mit  $U(\xi, \varepsilon) \subset M \setminus A$ , also  $d(x, \xi) \geq \varepsilon$  für alle  $x \in A$ , insbesondere damit  $d(x_n, \xi) \geq \varepsilon$  für alle  $n \in \mathbb{N}$ . Es folgt  $x_n \not\rightarrow \xi$ , ein Widerspruch.

(ii)  $\Rightarrow$  (iii): Sei  $\xi$  Häufungspunkt von  $A$ . Dann existiert eine Folge  $(x_n)$  in  $A \setminus \{\xi\}$  mit  $x_n \rightarrow \xi$ . Es folgt  $\xi \in A$ .

(iii)  $\Rightarrow$  (i): Wir nehmen an, dass  $A$  nicht abgeschlossen ist, also  $M \setminus A$  nicht offen ist. Dann existiert  $\xi \in M \setminus A$ , so dass für jedes  $\varepsilon > 0$  gilt, dass  $U(\xi, \varepsilon) \not\subset M \setminus A$ , also  $U(\xi, \varepsilon) \cap A \neq \emptyset$ . Insbesondere existiert dann  $x_n \in U(\xi, \frac{1}{n}) \cap A$ . Es folgt  $x_n \rightarrow \xi$ . Da  $x_n \in A \setminus \{\xi\}$  für alle  $n \in \mathbb{N}$ , ist  $\xi$  Häufungspunkt von  $A$ , also  $\xi \in A$ , ein Widerspruch.  $\square$

**Definition 6.3.2** Sei  $(M, d)$  metrischer Raum und  $A \subset M$ . Sei  $X$  die Menge aller abgeschlossenen Teilmengen von  $M$ , die  $A$  enthalten. Dann heißt

$$\bar{A} := \bigcap_{T \in X} T$$

der *Abschluss* von  $A$ .

Nach Satz 6.3.2 ist  $\bar{A}$  abgeschlossen. Damit ist  $\bar{A}$  also die kleinste abgeschlossene Menge, die  $A$  enthält. Es gilt

$$\bar{A} = A \cup \{x \in M : x \text{ ist Häufungspunkt von } A\}.$$

**Satz 6.3.4** Sei  $(M, d)$  vollständiger metrischer Raum und  $A \subset M$  abgeschlossen. Dann ist  $A$  vollständig (d. h.,  $(A, d|_{A \times A})$  ist vollständiger metrischer Raum).

*Beweis.* Ist  $(x_n)$  Cauchfolge in  $A$ , so ist  $(x_n)$  Cauchfolge in  $M$ , also gilt  $x_n \rightarrow \xi$  für ein  $\xi \in M$ . Mit Satz 6.3.3 folgt  $\xi \in A$ .  $\square$

## 6.4 Stetigkeit

**Definition 6.4.1** Seien  $(M, d_M), (N, d_N)$  metrische Räume und sei  $f : M \rightarrow N$  eine Funktion. Sei  $\xi \in M$ . Dann heißt  $f$  *stetig in  $\xi$* , falls für jede Folge  $(x_n)$  in  $M$  mit  $x_n \rightarrow \xi$  gilt, dass  $f(x_n) \rightarrow f(\xi)$ . Für  $A \subset M$  heißt  $f$  *stetig in  $A$* , falls  $f$  stetig in jedem Punkt von  $A$  ist. Schließlich heißt  $f$  *stetig*, wenn  $f$  stetig in  $M$  ist.

Für  $M, N \subset \mathbb{C}$ , mit der euklidischen Metrik versehen, ist dies genau Definition 3.1.1. Die Sätze 3.1.1 und 3.1.2 übertragen sich unmittelbar:

**Satz 6.4.1** Seien  $(M, d_M), (N, d_N), (P, d_P)$  metrische Räume,  $g : M \rightarrow N$  stetig (in  $\xi \in M$ ) und  $f : N \rightarrow P$  stetig (in  $g(\xi) \in g(M)$ ). Dann ist  $f \circ g$  stetig (in  $\xi$ ).

**Satz 6.4.2** Seien  $(M, d_M), (N, d_N)$  metrische Räume,  $\xi \in M$  und  $f : M \rightarrow N$ . Dann ist  $f$  stetig in  $\xi$  genau dann, wenn zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass für jedes  $x \in M$  mit  $d_M(x, \xi) < \delta$  gilt, dass  $d_N(f(x), f(\xi)) < \varepsilon$  ist.

*Beispiel.* Sei  $(M, d)$  metrischer Raum,  $A \subset M$  und  $f : A \rightarrow \mathbb{R}$  definiert durch  $f(x) := \inf_{a \in A} d(a, x)$ . Es ist also  $f(x)$  der Abstand von  $x$  zur Menge  $A$ .

*Behauptung.*  $f$  ist stetig.

*Beweis.* Sei  $x, y \in M$  und  $a \in A$ . Dann gilt  $d(x, a) \leq d(x, y) + d(y, a)$  und damit  $f(x) \leq d(x, y) + f(y)$ . Es folgt  $f(x) - f(y) \leq d(x, y)$  und aus Symmetriegründen auch  $f(y) - f(x) \leq d(x, y)$ , also  $|f(y) - f(x)| \leq d(x, y)$ . Die Stetigkeit von  $d$  folgt jetzt aus Satz 6.4.2. (Man kann dort  $\delta := \varepsilon$  wählen.)  $\square$

**Satz 6.4.3** Seien  $(M, d_M), (N, d_N)$  metrische Räume und  $f : M \rightarrow N$ . Dann gilt:

- (i) Sei  $\xi \in M$ . Dann ist  $f$  stetig in  $\xi$  genau dann, wenn für jede Umgebung  $U$  von  $f(\xi)$  das Urbild  $f^{-1}(U)$  eine Umgebung von  $\xi$  ist.
- (ii) Die Funktion  $f$  ist genau dann stetig, wenn für jedes offene  $U \subset N$  das Urbild  $f^{-1}(U)$  offen ist.

*Beweis.* Behauptung (ii) folgt leicht aus (i). Es reicht also, (i) zu beweisen. Dazu benutzen wir die Charakterisierung der Stetigkeit durch die  $\varepsilon$ - $\delta$ -Bedingung (Satz 6.4.2 bzw. Satz 3.1.2).

Sei zunächst  $f$  stetig in  $\xi$  und sei  $U$  Umgebung von  $f(\xi)$ . Dann existiert  $\varepsilon > 0$  mit  $U(f(\xi), \varepsilon) \subset U$ . Da  $f$  stetig in  $\xi$  ist, existiert  $\delta > 0$  mit  $d_N(f(x), f(\xi)) < \varepsilon$  für alle  $x \in M$  mit  $d_M(x, \xi) < \delta$ . Es folgt, dass  $f(x) \in U(f(\xi), \varepsilon) \subset U$  für alle  $x \in M$  mit  $d_M(x, \xi) < \delta$ , also  $U(\xi, \delta) \subset f^{-1}(U)$ . Damit ist  $f^{-1}(U)$  Umgebung von  $\xi$ .

Sei nun umgekehrt für jede Umgebung  $U$  von  $f(\xi)$  das Urbild  $f^{-1}(U)$  eine Umgebung von  $\xi$ . Sei  $\varepsilon > 0$ . Da  $U(f(\xi), \varepsilon)$  Umgebung von  $f(\xi)$  ist, ist  $f^{-1}(U(f(\xi), \varepsilon))$  Umgebung von  $\xi$ . Damit existiert  $\delta > 0$  mit  $U(\xi, \delta) \subset f^{-1}(U(f(\xi), \varepsilon))$ . Für  $x \in M$  mit  $d_M(x, \xi) < \delta$  gilt nun  $x \in U(\xi, \delta) \subset f^{-1}(U(f(\xi), \varepsilon))$ , also  $f(x) \in U(f(\xi), \varepsilon)$  und damit  $d_N(f(x), f(\xi)) < \varepsilon$ .  $\square$

In Satz 2.2.1 wurde für Zahlenfolgen  $(a_n)$  und  $(b_n)$  mit  $a_n \rightarrow a$  und  $b_n \rightarrow b$  gezeigt, dass  $a_n + b_n \rightarrow a + b$  und  $a_n \cdot b_n \rightarrow a \cdot b$  gilt. Hieraus wurde in §3.1 die Stetigkeit von Summe und Produkt stetiger Funktionen gefolgert. Entsprechende Aussagen gelten auch hier wieder (wobei die Beweise analog sind).

**Satz 6.4.4** Sei  $(M, d)$  metrischer Raum,  $(V, \|\cdot\|)$  normierter Raum (über  $\mathbb{K}$ ),  $\xi \in M$ ,  $\lambda \in \mathbb{K}$  und  $f, g : M \rightarrow V$  stetig in  $\xi$ . Dann sind  $f + g$  und  $\lambda \cdot f$  stetig in  $\xi$ . Ist speziell  $V = \mathbb{K}$  und  $\|\cdot\| = |\cdot|$ , so ist auch  $f \cdot g$  stetig in  $\xi$ . Im Falle  $g(\xi) \neq 0$  ist auch die (in  $M \setminus g^{-1}(0)$  definierte) Funktion  $f/g$  stetig in  $\xi$ .

Wir untersuchen Abbildungen  $f : \mathbb{K}^p \rightarrow \mathbb{K}$  auf Stetigkeit bzgl. der Norm  $\|\cdot\|_\infty$ . Für  $j \in \{1, \dots, p\}$  heißt die durch  $\pi_j(x_1, \dots, x_p) = x_j$  gegebene Abbildung  $\pi_j : \mathbb{K}^p \rightarrow \mathbb{K}$  die *Projektion* auf die  $j$ -te Koordinate. Offensichtlich ist  $\pi_j$  stetig. (Man wähle  $\delta := \varepsilon$  in Satz 6.4.2.) Nach Satz 6.4.4 sind dann auch Funktionen  $f$  der Form

$$f(x_1, \dots, x_p) = \sum_{(k_1, \dots, k_p) \in I} a_{k_1, \dots, k_p} x_1^{k_1} \cdot \dots \cdot x_p^{k_p}$$

stetig, wobei  $I \subset \mathbb{N}_0^p$  endlich und  $a_{k_1, \dots, k_p} \in \mathbb{K}$  für  $(k_1, \dots, k_p) \in I$ . Solche Funktionen heißen *Polynome* (in  $p$  Veränderlichen). Quotienten von Polynomen heißen *rationale Funktionen*. Sie sind dort definiert, wo der Nenner nicht verschwindet.

Wir untersuchen jetzt Abbildungen  $f : \mathbb{K}^p \rightarrow \mathbb{K}^q$  auf Stetigkeit, wieder bzgl. der Norm  $\|\cdot\|_\infty$  (sowohl im Definitions- wie im Zielbereich). Für  $j \in \{1, \dots, q\}$  heißt  $f_j := \pi_j \circ f$  die  $j$ -te *Koordinatenfunktion*. Es ist also  $f(x) = (f_1(x), \dots, f_q(x))^T$  für  $x \in \mathbb{K}^p$ .

**Satz 6.4.5** *Eine Funktion von  $\mathbb{K}^p$  nach  $\mathbb{K}^q$  ist genau dann stetig (in  $\xi \in \mathbb{K}^p$ ), wenn alle ihre Koordinatenfunktionen stetig (in  $\xi$ ) sind.*

*Beweis.* Sei  $f = (f_1, \dots, f_q)^T : \mathbb{K}^p \rightarrow \mathbb{K}^q$ . Aus der Stetigkeit von  $f$  folgt die der  $f_j$  mit Satz 6.4.1, da  $f_j = \pi_j \circ f$  und  $\pi_j$  stetig.

Seien nun alle  $f_j$  stetig in  $\xi$ . Sei  $\varepsilon > 0$ . Dann existiert zu  $j \in \{1, \dots, q\}$  ein  $\delta_j > 0$  mit  $|f_j(x) - f_j(\xi)| < \varepsilon$  für  $\|x - \xi\|_\infty < \delta_j$ . Mit  $\delta := \min_j \delta_j$  folgt hieraus, dass  $\|f(x) - f(\xi)\|_\infty = \max_j |f_j(x) - f_j(\xi)| < \varepsilon$  für  $\|x - \xi\|_\infty < \delta$ .  $\square$

Man könnte nun vermuten, dass eine Funktion  $f = (f_1, \dots, f_q)^T : \mathbb{K}^p \rightarrow \mathbb{K}^q$  stetig ist, wenn sie für jedes  $j$  bei festgehaltenem  $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p$  stetig in  $x_j$  ist. Dies ist aber nicht der Fall.

*Beispiel.* Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,

$$f(x, y) = \begin{cases} \frac{2xy}{x^2 + y^2} & \text{falls } (x, y) \neq (0, 0), \\ 0 & \text{falls } (x, y) = (0, 0). \end{cases}$$

Dann ist  $f$  unstetig in  $(0, 0)$ , da  $f(x, x) = 1$  für alle  $x \in \mathbb{R} \setminus \{0\}$ , aber  $f(0, 0) = 0$ . Wird eine Variable festgehalten, ist  $f$  aber stetig in der anderen.

Es folgt aus Satz 6.4.5, dass lineare Abbildungen von  $\mathbb{K}^p$  nach  $\mathbb{K}^q$  stetig sind, da jede Koordinatenfunktion ein Polynom ist. Denn ist  $f = (f_1, \dots, f_q)^T : \mathbb{K}^p \rightarrow \mathbb{K}^q$  linear, so existieren  $a_{ij} \in \mathbb{K}$ ,  $1 \leq i \leq q$ ,  $1 \leq j \leq p$ , mit  $f_i(x_1, \dots, x_p) = \sum_{j=1}^p a_{ij} x_j$ . Entscheidend ist dabei, dass  $\mathbb{K}^p$  und  $\mathbb{K}^q$  endlichdimensional sind. Lineare Abbildungen zwischen unendlichdimensionalen normierten Räumen brauchen nicht stetig zu sein.

*Beispiel.* Es ist  $(C^1[-1, 1], \|\cdot\|_\infty)$  normierter Raum und die durch  $f \mapsto f'(0)$  gegebene Abbildung  $L : C^1[-1, 1] \rightarrow \mathbb{C}$  ist linear. Mit  $f_n(x) = \frac{1}{n} \sin nx$  gilt  $\|f_n\|_\infty \leq \frac{1}{n}$ , also  $f_n \rightarrow 0$ , aber  $L(f_n) = f_n'(0) = \cos n0 = 1 \neq 0 = L(0)$ . Damit ist  $L$  nicht stetig.

**Definition 6.4.2** (vgl. Definition 3.6.2) Seien  $(M, d_M), (N, d_N)$  metrische Räume und  $f : M \rightarrow N$ . Dann heißt  $f$  *gleichmäßig stetig*, falls zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass  $d_N(f(x), f(y)) < \varepsilon$  für alle  $x, y \in M$  mit  $d_M(x, y) < \delta$ .

**Satz 6.4.6** *Seien  $(V, \|\cdot\|_V), (W, \|\cdot\|_W)$  normierte Räume und sei  $f : V \rightarrow W$  linear. Dann sind äquivalent:*

- (i)  $f$  ist gleichmäßig stetig,
- (ii)  $f$  ist stetig,
- (iii)  $f$  ist stetig in 0,
- (iv)  $\exists C > 0 \forall x \in V : \|f(x)\|_W \leq C\|x\|_V$ .

*Beweis.* Wir zeigen, dass (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii)  $\Rightarrow$  (iv)  $\Rightarrow$  (i). Dabei ist (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii) trivial.

Zu (iii)  $\Rightarrow$  (iv): Sei  $f$  stetig in 0. Zu  $\varepsilon := 1$  existiert dann  $\delta > 0$  mit  $\|f(x)\|_W \leq 1$  für  $\|x\|_V \leq \delta$ . Für  $x \in V \setminus \{0\}$  folgt dann

$$\|f(x)\|_W = \frac{\|x\|_V}{\delta} \left\| f\left(\frac{\delta}{\|x\|_V}x\right) \right\|_W \leq \frac{\|x\|_V}{\delta}.$$

Mit  $C := \frac{1}{\delta}$  folgt also  $\|f(x)\|_W \leq C\|x\|_V$ , und das gilt auch für  $x = 0$ .

Zu (iv)  $\Rightarrow$  (i): Sei  $\varepsilon > 0$  und  $\delta := \frac{\varepsilon}{C}$ . Für  $x, y \in V$  mit  $\|x - y\|_V < \delta$  gilt dann  $\|f(x) - f(y)\|_W = \|f(x - y)\|_W \leq C\|x - y\|_V < \varepsilon$ .  $\square$

Seien  $(V, \|\cdot\|_V)$ ,  $(W, \|\cdot\|_W)$  normierte Räume. Den Vektorraum der stetigen linearen Abbildungen von  $V$  nach  $W$  bezeichnen wir mit  $L(V, W)$ . Nach Satz 6.4.6, (iv), ist für  $f \in L(V, W)$  dann

$$\|f\|_{L(V, W)} := \sup_{x \in V \setminus \{0\}} \frac{\|f(x)\|_W}{\|x\|_V} = \sup_{\substack{x \in V \\ \|x\|_V = 1}} \|f(x)\|_W < \infty.$$

Man nennt  $\|f\|_{L(V, W)}$  *Operatornorm* von  $f$ . Es ist nicht schwer zu zeigen, dass  $(L(V, W), \|\cdot\|_{L(V, W)})$  normierter Raum ist. Darüberhinaus ist  $(L(V, W), \|\cdot\|_{L(V, W)})$  Banachraum, falls  $(W, \|\cdot\|_W)$  Banachraum ist.

*Beispiel.* Seien  $(V, \|\cdot\|_V) = (\mathbb{R}^p, \|\cdot\|_\infty)$ ,  $(W, \|\cdot\|_W) = (\mathbb{R}^q, \|\cdot\|_\infty)$  und  $f = (f_1, \dots, f_q)^T : \mathbb{R}^p \rightarrow \mathbb{R}^q$ . Dann existieren  $a_{ij} \in \mathbb{K}$ ,  $1 \leq i \leq q$ ,  $1 \leq j \leq p$ , mit  $f_i(x_1, \dots, x_p) = \sum_{j=1}^p a_{ij}x_j$  für alle  $i$ . In Matrixschreibweise haben wir also

$$\begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_q(x) \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ a_{21} & \dots & a_{2p} \\ \vdots & & \vdots \\ a_{q1} & \dots & a_{qp} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$

Nun ist für  $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$

$$|f_i(x_1, \dots, x_p)| \leq \sum_{j=1}^p |a_{ij}| \cdot |x_j| \leq \|x\|_\infty \sum_{j=1}^p |a_{ij}|$$

für alle  $i$  und damit

$$\|f(x)\|_\infty \leq \|x\|_\infty \max_i \sum_{j=1}^p |a_{ij}|.$$

Wählt man, bei festem  $i$ , nun  $x_j = \pm 1$ , so dass  $x_j$  und  $a_{ij}$  das gleiche Vorzeichen haben, so hat man in obiger Abschätzung Gleichheit. Es folgt

$$\|f\|_{L(V, W)} = \max_i \sum_{j=1}^p |a_{ij}|.$$

Man nennt dies auch die *Zeilensummennorm* der Matrix  $(a_{ij})_{\substack{1 \leq i \leq q \\ 1 \leq j \leq p}}$ .

**Definition 6.4.3** (vgl. Definition 3.3.2) Seien  $(M, d_M), (N, d_N)$  metrische Räume,  $\xi$  Häufungspunkt von  $M$  und  $f : M \rightarrow N$ . Wir sagen, dass  $f(x)$  für  $x \rightarrow \xi$  konvergiert, falls  $\eta \in N$  existiert, so dass für jede Folge  $(x_n)$  in  $M \setminus \{\xi\}$  mit  $x_n \rightarrow \xi$  auch  $f(x_n) \rightarrow \eta$  gilt. Dieses  $\eta$  heißt dann *Grenzwert* von  $f$  für  $x \rightarrow \xi$  und wird mit  $\lim_{x \rightarrow \xi} f(x)$  bezeichnet. Wir schreiben auch  $f(x) \rightarrow \eta$  für  $x \rightarrow \xi$ .

Wie in Satz 3.3.1 ist  $f$  genau dann stetig in  $\xi$  ist, wenn  $\lim_{x \rightarrow \xi} f(x) = f(\xi)$  gilt. Auch Satz 3.3.3 gilt entsprechend: *Es gilt  $\lim_{x \rightarrow \xi} f(x) = \eta$  genau dann, wenn*

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in M \setminus \{\xi\} : d_M(x, \xi) < \delta \Rightarrow d_N(f(x), \eta) < \varepsilon.$$

## 6.5 Kompaktheit

**Definition 6.5.1** Sei  $(M, d)$  metrischer Raum.

- (i)  $(M, d)$  heißt *kompakt*, falls jede Folge in  $M$  eine konvergente Teilfolge besitzt.
- (ii)  $(M, d)$  heißt *totalbeschränkt* (oder *präkompakt*), falls für alle  $\varepsilon > 0$  eine endliche Teilmenge  $E$  von  $M$  existiert, so dass  $M \subset \bigcup_{x \in E} U(x, \varepsilon)$ .
- (iii)  $(M, d)$  heißt *beschränkt*, falls  $\sup_{x, y \in M} d(x, y) < \infty$ .

Kurz sagen wir auch, dass  $M$  (statt  $(M, d)$ ) kompakt bzw. (total)beschränkt ist. Für  $K \subset M$  sagen wir, dass  $K$  kompakt bzw. (total)beschränkt ist, wenn der metrische Raum  $(K, d|_{K \times K})$  diese Eigenschaft hat.

Teil (i) obiger Definition entspricht damit Definition 3.6.1. Statt die Existenz einer konvergenten Teilfolge zu fordern, kann man natürlich auch die Existenz eines Häufungswertes fordern; vgl. Satz 2.3.2, (i).

Aus der Totalbeschränktheit folgt die Beschränktheit, denn aus

$$M \subset \bigcup_{x \in E} U(x, \varepsilon)$$

folgt  $d(a, b) \leq \max_{x, y \in E} d(x, y) + 2\varepsilon$  für  $a, b \in M$ .

Sind  $I, M$  nichtleere Mengen, so nennt man eine durch  $i \mapsto x_i$  gegebene Funktion von  $I$  nach  $M$  auch *Familie* und bezeichnet sie mit  $(x_i)_{i \in I}$ . (Für  $I = \mathbb{N}$  sind dies Folgen.)

**Definition 6.5.2** Sei  $(M, d)$  metrischer Raum,  $K \subset M$  und  $(U_i)_{i \in I}$  eine Familie offener Teilmengen von  $M$  mit  $K \subset \bigcup_{i \in I} U_i$ . Dann heißt  $(U_i)_{i \in I}$  eine *offene Überdeckung* von  $K$ . Existiert eine endliche Teilmenge  $E \subset I$  mit  $K \subset \bigcup_{i \in E} U_i$ , so sagt man, dass  $(U_i)_{i \in I}$  eine *endliche Teilüberdeckung* besitzt.

Die Totalbeschränktheit besagt also, dass für jedes  $\varepsilon > 0$  die offene Überdeckung  $(U(x, \varepsilon))_{x \in M}$  eine endliche Teilüberdeckung besitzt.

**Satz 6.5.1** *Sei  $(M, d)$  metrischer Raum und  $K \subset M$ . Dann sind äquivalent:*

- (i) *Jede offene Überdeckung von  $K$  besitzt eine endliche Teilüberdeckung;*
- (ii)  *$K$  ist kompakt;*

(iii)  $K$  ist vollständig und totalbeschränkt.

Die Äquivalenz von (i) und (ii) heißt auch *Satz von Heine-Borel*.

In sogenannten *topologischen Räumen* – die eine Verallgemeinerung der metrischen Räume sind – gilt Satz 6.5.1 nicht. Dort wird die in (i) formulierte Eigenschaft als Kompaktheit bezeichnet und die in Definition 6.5.1, (i), angegebene als Folgenkompaktheit.

*Beweis von Satz 6.5.1.* Zu (i)  $\Rightarrow$  (ii): Sei  $(x_n)$  Folge in  $K$ , die keine konvergente Teilfolge (also auch keinen Häufungswert) besitzt. Dann ist  $A := \{x_n : n \in \mathbb{N}\}$  unendlich.

Sei nun  $y \in K$ . Da  $y$  kein Häufungswert von  $(x_n)$  ist, existiert  $\varepsilon_y > 0$  und  $N \in \mathbb{N}$ , so dass  $d(x_n, y) \geq \varepsilon_y$  für alle  $n \geq N$  gilt. Folglich ist  $A \cap U(y, \varepsilon_y)$  endlich.

Nun ist  $(U(y, \varepsilon_y))_{y \in K}$  offene Überdeckung von  $K$ . Nach (i) existiert nun eine endliche Teilüberdeckung, d.h., es existiert eine endliche Teilmenge  $E$  von  $K$  mit  $K \subset \bigcup_{y \in E} U(y, \varepsilon_y)$ . Es folgt  $A = A \cap K \subset \bigcup_{y \in E} A \cap U(y, \varepsilon_y)$ . Damit ist  $A$  endlich, ein Widerspruch.

Zu (ii)  $\Rightarrow$  (iii): Sei  $K$  kompakt. Wir zeigen zuerst, dass  $K$  vollständig ist. Sei dazu  $(x_n)$  Cauchyfolge in  $K$ . Dann hat  $(x_n)$  eine konvergente Teilfolge  $(x_{n_k})$ , etwa  $x_{n_k} \rightarrow \xi \in K$ . Wir zeigen, dass  $x_n \rightarrow \xi$ . Sei dazu  $\varepsilon > 0$ . Da  $(x_n)$  Cauchyfolge ist, existiert  $N \in \mathbb{N}$  mit  $d(x_n, x_m) < \frac{\varepsilon}{2}$  für  $m, n \geq N$ . Da  $x_{n_k} \rightarrow \xi$  existiert  $K \in \mathbb{N}$  mit  $d(x_{n_k}, \xi) < \frac{\varepsilon}{2}$  für  $k \geq K$ . Sei nun  $k \geq K$  mit  $n_k \geq N$ . Für  $n \geq N$  folgt dann  $d(x_n, \xi) \leq d(x_n, x_{n_k}) + d(x_{n_k}, \xi) < \varepsilon$ . Es gilt also  $x_n \rightarrow \xi$ .

Wir zeigen jetzt, dass  $K$  totalbeschränkt ist. Dazu nehmen wir an, dass dies nicht der Fall ist. Dann existiert  $\varepsilon > 0$  mit  $K \not\subset \bigcup_{x \in E} U(x, \varepsilon)$  für jede endliche Teilmenge  $E$  von  $K$ . Es sei nun  $x_1 \in K$  beliebig. Dann gilt  $K \not\subset U(x_1, \varepsilon)$  und damit existiert  $x_2 \in K \setminus U(x_1, \varepsilon)$ . Wegen  $K \not\subset U(x_1, \varepsilon) \cup U(x_2, \varepsilon)$  existiert jetzt  $x_3 \in K \setminus (U(x_1, \varepsilon) \cup U(x_2, \varepsilon))$ . Induktiv findet man so eine Folge  $(x_n)$  in  $K$  mit  $x_{n+1} \in K \setminus \bigcup_{j=1}^n U(x_j, \varepsilon)$ . Es folgt  $d(x_m, x_n) \geq \varepsilon$  für  $m \neq n$ . Dies impliziert, dass keine Teilfolge von  $(x_n)$  eine Cauchyfolge ist. Damit konvergiert aber auch keine Teilfolge von  $(x_n)$ , ein Widerspruch.

Zu (iii)  $\Rightarrow$  (i): Sei  $(U_i)_{i \in I}$  offene Überdeckung von  $K$ . Wir nehmen an, dass keine endliche Teilüberdeckung existiert.

Wegen der Totalbeschränktheit existiert eine endliche Teilmenge  $E_1$  von  $K$  mit  $K \subset \bigcup_{x \in E_1} U(x, \frac{1}{2})$ . Nun ist  $(U_i)_{i \in I}$  auch offene Überdeckung von  $K \cap U(x, \frac{1}{2})$  für alle  $x \in K$ . Es folgt, dass  $x_1 \in E_1$  existiert, so dass keine endliche Teilüberdeckung von  $K \cap U(x_1, \frac{1}{2})$  existiert.

Wegen der Totalbeschränktheit existiert weiter eine endliche Teilmenge  $E_2$  von  $K$  mit  $K \cap U(x_1, \frac{1}{2}) \subset \bigcup_{x \in E_2} U(x, \frac{1}{4})$ . Dabei kann man  $U(x_1, \frac{1}{2}) \cap U(x, \frac{1}{4}) \neq \emptyset$ , also  $d(x_1, x) < \frac{1}{2} + \frac{1}{4}$ , für alle  $x \in E_2$  annehmen. Es existiert nun  $x_2 \in E_2$ , so dass keine endliche Teilüberdeckung von  $K \cap U(x_2, \frac{1}{4})$  existiert.

Induktiv findet man so eine Folge  $(x_n)$ , so dass keine endliche Teilüberdeckung von  $K \cap U(x_n, \frac{1}{2^n})$  existiert und außerdem

$$d(x_n, x_{n+1}) < \frac{1}{2^n} + \frac{1}{2^{n+1}} = \frac{3}{2^{n+1}}$$



gilt. Für  $m > n$  folgt

$$d(x_n, x_m) < \sum_{k=n+1}^m \frac{3}{2^k} = \frac{3}{2^{n+1}} \sum_{j=0}^{\infty} \frac{1}{2^j} = \frac{3}{2^n}.$$

Damit ist  $(x_n)$  Cauchyfolge und wegen der Vollständigkeit von  $K$  also konvergent, etwa  $x_n \rightarrow \xi \in K$ . Da  $(U_i)_{i \in I}$  offene Überdeckung von  $K$  ist, existiert  $i_0 \in I$  mit  $\xi \in U_{i_0}$ .

Nun existiert  $\varepsilon > 0$  mit  $U(\xi, \varepsilon) \subset U_{i_0}$ . Für genügend großes  $n \in \mathbb{N}$  gilt nun  $d(x_n, \xi) < \frac{\varepsilon}{2}$  und  $\frac{1}{2^n} < \frac{\varepsilon}{2}$ . Es folgt

$$K \cap U\left(x_n, \frac{1}{2^n}\right) \subset U\left(x_n, \frac{1}{2^n}\right) \subset U\left(\xi, d(x_n, \xi) + \frac{1}{2^n}\right) \subset U(\xi, \varepsilon) \subset U_{i_0}.$$

Damit existiert eine endliche Teilüberdeckung von  $K \cap U\left(x_n, \frac{1}{2^n}\right)$ , ein Widerspruch.  $\square$

**Satz 6.5.2** (i) *Kompakte Mengen sind beschränkt und abgeschlossen.*

(ii) *Abgeschlossene Teilmengen kompakter Mengen sind kompakt.*

*Beweis.* Sei  $K$  kompakt (im metrischen Raum  $(M, d)$ ).

Zu (i): Zunächst ist  $K$  wegen Satz 6.5.1 totalbeschränkt und damit auch beschränkt.

Um die Abgeschlossenheit zu zeigen, benutzen wir Satz 6.3.3. Sei dazu  $(x_n)$  Folge in  $K$  mit  $x_n \rightarrow \xi \in M$ . Zu zeigen ist, dass  $\xi \in K$ . Dies folgt aber, da  $(x_n)$  Cauchyfolge und  $K$  nach Satz 6.5.1 vollständig ist.

Zu (ii): Sei  $A \subset K$  abgeschlossen. Wir geben zwei Beweise, dass  $A$  kompakt ist.

*Beweisvariante 1.* Sei  $(x_n)$  Folge in  $A$ . Dann besitzt  $(x_n)$  eine Teilfolge, die in  $K$  konvergiert, etwa  $x_{n_k} \rightarrow \xi \in K$ . Da  $A$  abgeschlossen, folgt mit Satz 6.3.3, dass  $\xi \in A$ . Also ist  $A$  kompakt.

*Beweisvariante 2.* Sei  $(U_i)_{i \in I}$  offene Überdeckung von  $A$ . Da  $M \setminus A$  offen ist, ist dann durch  $K \subset \bigcup_{i \in I} U_i \cup M \setminus A$  eine offene Überdeckung von  $K$  gegeben. Damit existiert eine endliche Teilmenge  $E \subset I$  mit  $K \subset \bigcup_{i \in E} U_i \cup M \setminus A$ , also  $A \subset \bigcup_{i \in E} U_i$ .  $\square$

Wir werden gleich an einem Beispiel sehen, dass die Umkehrung von Satz 6.5.2, (i), nicht gilt. Es gilt aber der folgende Satz.

**Satz 6.5.3** *Sei  $p \in \mathbb{N}$  und  $K \subset \mathbb{R}^p$ . Bezüglich der durch die Norm  $\|\cdot\|_\infty$  erzeugten Metrik ist  $K$  genau dann kompakt, wenn  $K$  beschränkt und abgeschlossen ist.*

Auch dieser Satz wird nach Heine und Borel benannt. Er gilt auch für andere Normen in  $\mathbb{R}^p$ ; siehe Satz 6.5.6 später.

*Beweis von Satz 6.5.3.* “ $\Rightarrow$ ”. Satz 6.5.2, (i).

“ $\Leftarrow$ ”. Nach Satz 6.5.1 reicht es zu zeigen, dass  $K$  vollständig und totalbeschränkt ist. Die Vollständigkeit folgt dabei wegen Satz 6.3.4 (und der Vollständigkeit von  $\mathbb{R}^p$ ) unmittelbar aus der Abgeschlossenheit.

Wegen der Beschränktheit von  $K$  existiert  $r > 0$  mit  $\|x\|_\infty \leq r$  für alle  $x \in K$ . Ist dann  $\varepsilon > 0$  und  $N \in \mathbb{N}$  mit  $N\varepsilon \geq r$ , so folgt

$$K \subset \bigcup_{j_1, \dots, j_n = -N}^N U((j_1\varepsilon, \dots, j_n\varepsilon), \varepsilon)$$

und damit die Totalbeschränktheit von  $K$ .  $\square$

*Beispiel.* Wir betrachten den Prähilbertraum  $(C[0, 2\pi], \langle \cdot, \cdot \rangle)$  und bezeichnen die zugehörige Norm mit  $\|\cdot\|_2$ . Wie in §5.6 sei  $e_n \in C[0, 2\pi]$  definiert durch  $e_n(x) = e^{inx}$ , für  $n \in \mathbb{Z}$ . Dann gilt, vgl. §5.6,

$$\langle e_n, e_m \rangle = \int_0^{2\pi} e_n \overline{e_m} = \int_0^{2\pi} e^{i(n-m)x} dx = 0$$

für  $n \neq m$  und  $\|e_n\|_2 = \sqrt{\langle e_n, e_n \rangle} = \sqrt{2\pi}$ . Nach dem Satz von Pythagoras ist damit  $\|e_n - e_m\|_2^2 = \|e_n\|_2^2 + \|e_m\|_2^2 = 4\pi$ , also  $\|e_n - e_m\|_2 = \sqrt{4\pi}$ .

Wir betrachten jetzt  $K := \{e_n : n \in \mathbb{N}\}$ . Dann ist  $K$  offensichtlich beschränkt. Außerdem ist  $K$  abgeschlossen, denn ist  $(x_k)$  eine Folge in  $K$ , die in  $(C[0, 2\pi], \langle \cdot, \cdot \rangle)$  konvergiert, so existieren  $M, N \in \mathbb{N}$  mit  $x_k = e_N$  für  $k \geq M$ .

Andererseits ist  $K$  nicht kompakt, da  $(e_n)_{n \in \mathbb{N}}$  keine konvergente Teilfolge besitzt. Es folgt, dass die Umkehrung von Satz 6.5.2, (i), im allgemeinen gilt.

**Satz 6.5.4** *Seien  $(M, d_M)$  und  $(N, d_N)$  metrische Räume,  $K \subset M$  kompakt und  $f : K \rightarrow N$  stetig. Dann ist  $f(K)$  kompakt.*

Für  $M = N = \mathbb{C}$  (mit euklidischer Metrik) ist dies Satz 3.6.3. Der dortige Beweis kann wörtlich übernommen werden.

Ebenso überträgt sich Satz 3.6.4:

**Satz 6.5.5** *Sei  $(M, d)$  metrischer Raum,  $K \subset M$  kompakt und  $f : K \rightarrow \mathbb{R}$  stetig. Dann nimmt  $f$  auf  $K$  sein Minimum und Maximum an.*

**Satz 6.5.6** *Alle Normen auf dem  $\mathbb{R}$ -Vektorraum  $\mathbb{R}^p$  (mit  $p \in \mathbb{N}$ ) sind äquivalent.*

*Beweis.* Es reicht zu zeigen, dass jede Norm äquivalent zu  $\|\cdot\|_\infty$  ist. Sei also  $\|\cdot\|$  Norm auf  $\mathbb{R}^p$ .

Sei  $\{e_1, \dots, e_p\}$  die Standardbasis des  $\mathbb{R}^p$ . Mit  $\beta := \sum_{j=1}^p \|e_j\|$  ist dann

$$\|x\| = \left\| \sum_{j=1}^p x_j e_j \right\| \leq \sum_{j=1}^p |x_j| \|e_j\| \leq \beta \|x\|_\infty$$

für  $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ . Insbesondere folgt, dass die durch  $x \mapsto \|x\|$  gegebene Abbildung von  $(\mathbb{R}^p, \|\cdot\|_\infty)$  nach  $(\mathbb{R}, |\cdot|)$  stetig ist; vgl. auch Satz 6.4.6. Da  $K := \{x \in \mathbb{R}^p : \|x\|_\infty = 1\}$  beschränkt und abgeschlossen und damit kompakt in  $(\mathbb{R}^p, \|\cdot\|_\infty)$  ist, existiert deshalb nach Satz 6.4.5 ein  $\xi \in K$  mit  $\alpha := \|\xi\| = \min_{x \in K} \|x\|$ . Für  $x \in \mathbb{R}^p \setminus \{0\}$  folgt

$$\|x\| = \left\| \frac{x}{\|x\|_\infty} \right\| \cdot \|x\|_\infty \geq \alpha \|x\|_\infty.$$

Damit erhalten wir  $\alpha \|x\|_\infty \leq \|x\| \leq \beta \|x\|_\infty$  für alle  $x \in \mathbb{R}^p$ .  $\square$

Analog zu Satz 3.6.5 ist das folgende Resultat (und sein Beweis).

**Satz 6.5.7** *Seien  $(M, d_M)$  und  $(N, d_N)$  metrische Räume,  $M$  kompakt und  $f : M \rightarrow N$  stetig. Dann ist  $f$  gleichmäßig stetig.*

## 6.6 Zusammenhang

**Definition 6.6.1** Ein metrischer Raum  $(M, d)$  heißt *unzusammenhängend*, falls nichtleere, offene Teilmengen  $U, V$  von  $M$  existieren, so dass  $M = U \cup V$  und  $U \cap V = \emptyset$  gilt. Andernfalls heißt  $(M, d)$  *zusammenhängend*.

*Bemerkungen.* 1. Aus  $U, V \neq \emptyset$  und  $U \cap V = \emptyset$  folgt  $U, V \neq M$ .

2. Statt zu fordern, dass  $U, V$  beide offen sind, kann man auch fordern, dass  $U, V$  beide abgeschlossen sind. Denn sind  $U, V$  wie in Definition 6.6.1, so sind  $A := M \setminus U$  und  $B := M \setminus V$  abgeschlossen, nichtleer und es gilt  $A \cup B = M$  und  $A \cap B = \emptyset$ .

3. Eine Teilmenge  $N$  von  $M$  heißt (un)zusammenhängend, falls dies für den metrischen Raum  $(N, d|_{N \times N})$  gilt. Es gilt dann, dass  $N$  genau dann unzusammenhängend ist, wenn (in  $M$ ) offene Mengen  $U', V' \subset M$  existieren, so dass  $U' \cap N \neq \emptyset$ ,  $V' \cap N \neq \emptyset$ ,  $N \subset U' \cup V'$  und  $U' \cap V' \cap N = \emptyset$ ; vgl. Satz 6.3.1.

**Satz 6.6.1** Eine nichtleere Teilmenge von  $\mathbb{R}$  ist genau dann zusammenhängend, wenn sie ein Intervall ist.

*Beweis.* Einpunktige Mengen sind offensichtlich zusammenhängend. Sei  $I \subset \mathbb{R}$  und  $I$  enthalte mindestens zwei Punkte.

Sei zunächst  $I$  zusammenhängend. Seien  $\alpha, \beta \in I$  und  $\gamma \in (\alpha, \beta)$ . Dann folgt  $\gamma \in I$ , da sonst mit  $U' := (-\infty, \gamma)$  und  $V' = (\gamma, \infty)$  gilt, dass  $I \subset U' \cup V'$ ,  $U' \cap I \neq \emptyset$ ,  $V' \cap I \neq \emptyset$  und  $U' \cap V' = \emptyset$ . Mit  $a := \inf I$  und  $b := \sup I$  folgt hieraus, dass  $I$  eines der Intervalle  $[a, b]$ ,  $[a, b)$ ,  $(a, b]$  oder  $(a, b)$  ist.

Sei nun  $I$  Intervall. Wir nehmen an, dass  $I$  nicht zusammenhängend ist, etwa  $I = A \cup B$  mit nichtleeren, disjunkten, in  $I$  abgeschlossenen Mengen  $A, B \subset I$ . Sei  $a \in A$  und  $b \in B$ . Ohne Beschränkung der Allgemeinheit sei  $a < b$ . Wegen  $a, b \in I$  folgt dann  $[a, b] \subset I$ . Sei  $c := \sup(A \cap [a, b])$ . Dann gilt  $c \in A$  (und damit insbesondere  $c < b$ ), da  $A$  abgeschlossen. Andererseits ist  $(c, b] \subset B$  und wegen der Abgeschlossenheit von  $B$  also auch  $c \in B$ . Insgesamt folgt  $c \in A \cap B$ , ein Widerspruch.  $\square$

**Satz 6.6.2** Seien  $(M, d_M)$  und  $(N, d_N)$  metrische Räume,  $M$  zusammenhängend und  $f : M \rightarrow N$  stetig. Dann ist  $f(M)$  zusammenhängend.

*Beweis.* Ohne Beschränkung der Allgemeinheit sei  $f(M) = N$ . Sei  $N$  unzusammenhängend, etwa  $N = U \cup V$  mit nichtleeren, disjunkten, offenen Teilmengen  $U, V$  von  $N$ . Mit  $\tilde{U} := f^{-1}(U)$  und  $\tilde{V} := f^{-1}(V)$  gilt dann  $M \subset \tilde{U} \cup \tilde{V}$ , und es sind  $\tilde{U}$  und  $\tilde{V}$  offen (vgl. Satz 6.4.3, (ii)), nichtleer und disjunkt. Damit ist  $M$  unzusammenhängend.  $\square$

**Definition 6.6.2** Sei  $(M, d)$  metrischer Raum.

- (i) Seien  $a, b \in \mathbb{R}$ ,  $a < b$  und  $\gamma : [a, b] \rightarrow M$  stetig. Dann heißt  $\gamma$  *Weg* von  $\gamma(a)$  nach  $\gamma(b)$ . Die Punkte  $\gamma(a)$  bzw.  $\gamma(b)$  heißen *Anfangs-* bzw. *Endpunkt* des Weges  $\gamma$ .
- (ii)  $(M, d)$  heißt *wegzusammenhängend*, wenn für alle  $x, y \in M$  ein Weg von  $x$  nach  $y$  existiert.

**Satz 6.6.3** *Ein wegzusammenhängender metrischer Raum ist zusammenhängend.*

*Beweis.* Sei  $(M, d)$  wegzusammenhängender metrischer Raum. Wir nehmen an, dass  $M$  unzusammenhängend ist, etwa  $M = U \cup V$  mit nichtleeren, disjunkten, offenen Teilmengen  $U, V$  von  $M$ . Sei  $u \in U, v \in V$ . Dann existiert ein Weg  $\gamma : [a, b] \rightarrow M$  von  $u$  nach  $v$ . Es ist dann  $\gamma([a, b])$  unzusammenhängend, nach Satz 6.6.2 also  $[a, b]$  unzusammenhängend, im Widerspruch zu Satz 6.6.1.  $\square$

Die Umkehrung von Satz 6.6.3 gilt nicht, wie das folgende *Beispiel* zeigt. Sei  $M := \{(x, \sin \frac{1}{x}) : x \in (0, 1]\} \cup \{(0, y) : y \in [-1, 1]\}$  und sei  $d$  die euklidische Metrik, eingeschränkt auf  $M$ . Dann ist  $(M, d)$  zusammenhängend, aber nicht wegzusammenhängend.

Die Menge  $M$  ist in Abbildung 23 dargestellt.

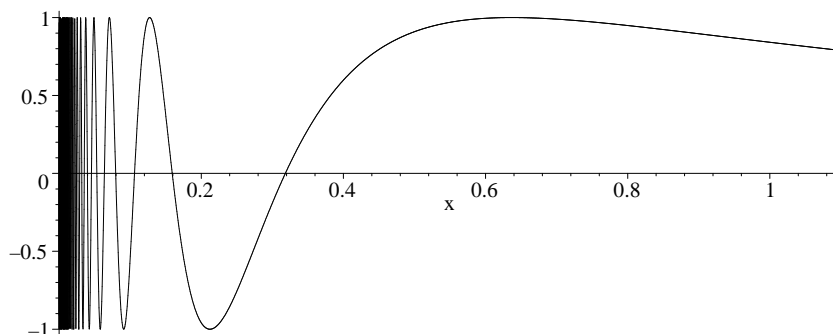


Abbildung 23: Eine zusammenhängende Menge.

**Definition 6.6.3** Sei  $V$  Vektorraum (über  $\mathbb{K}$ ).

- (i) Für  $x, y \in V$  heißt  $[x, y] := \{(1 - t)x + ty : t \in [0, 1]\}$  *Strecke* (von  $x$  nach  $y$ ).
- (ii) Für  $x_1, x_2, \dots, x_n \in V$  heißt  $[x_1, x_2, \dots, x_n] := [x_1, x_2] \cup [x_2, x_3] \cup \dots \cup [x_{n-1}, x_n]$  *Streckenzug* (von  $x_1$  nach  $x_n$ ).

In einem normierten Raum  $(V, \|\cdot\|)$  ist für  $x, y \in V$  die durch  $t \mapsto (1 - t)x + ty$  gegebene Funktion  $\gamma : [0, 1] \rightarrow V$  stetig, also ein Weg von  $x$  nach  $y$ . Damit ist die Strecke  $[x, y] = \gamma([0, 1])$  das Bild eines Weges. Auch Streckenzüge sind Bilder von Wegen.

**Satz 6.6.4** *Sei  $(V, \|\cdot\|)$  normierter Raum und  $M \subset V$  offen und zusammenhängend. Dann existiert für alle  $x, y \in M$  ein in  $M$  verlaufender Streckenzug von  $x$  nach  $y$ , d. h., es existieren  $x_1, x_2, \dots, x_n \in M$  mit  $x_1 = x, x_n = y$  und  $[x_1, x_2, \dots, x_n] \subset M$ . Insbesondere ist  $M$  wegzusammenhängend.*

*Beweis.* Sei  $x \in M$ . Wir betrachten die Menge  $U$  aller  $y \in M$ , für die ein in  $M$  verlaufender Streckenzug von  $x$  nach  $y$  existiert. Zu zeigen ist, dass  $U = M$  gilt.

Wir zeigen zunächst, dass  $U$  offen ist. Sei dazu  $y \in U$  und  $[x_1, x_2, \dots, x_n]$  ein in  $M$  verlaufender Streckenzug von  $x$  nach  $y$ . Da  $M$  offen und  $y \in U \subset M$ , existiert  $\varepsilon > 0$  mit  $U(y, \varepsilon) \subset M$ . Für  $z \in U(y, \varepsilon)$  ist dann  $[x_1, x_2, \dots, x_n, z]$  ein in  $M$  verlaufender Streckenzug von  $x$  nach  $z$ . Es folgt  $U(y, \varepsilon) \subset U$  und damit ist  $U$  offen.

Ein analoges Argument zeigt, dass auch  $V := M \setminus U$  offen ist. Wegen  $U \neq \emptyset, U \cup V = M$  und  $U \cap V = \emptyset$  folgt nun aus dem Zusammenhang von  $M$ , dass  $V = \emptyset$ , also  $U = M$ .  $\square$

# 7 Differentialrechnung in mehreren Veränderlichen

## 7.1 Partielle Ableitungen

**Definition 7.1.1** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi = (\xi_1, \dots, \xi_p)^T \in U$  und  $f : U \rightarrow \mathbb{R}$ . Für  $j \in \{1, \dots, p\}$  heißt  $f$  *partiell differenzierbar* in  $\xi$  nach der  $j$ -ten Variablen (oder nach  $x_j$ ), falls

$$\begin{aligned} & \partial_j f(\xi) \\ := & \lim_{x_j \rightarrow \xi_j} \frac{f(\xi_1, \dots, \xi_{j-1}, x_j, \xi_{j+1}, \dots, \xi_p) - f(\xi_1, \dots, \xi_{j-1}, \xi_j, \xi_{j+1}, \dots, \xi_p)}{x_j - \xi_j} \\ = & \lim_{h \rightarrow 0} \frac{f(\xi_1, \dots, \xi_{j-1}, \xi_j + h, \xi_{j+1}, \dots, \xi_p) - f(\xi_1, \dots, \xi_{j-1}, \xi_j, \xi_{j+1}, \dots, \xi_p)}{h} \end{aligned}$$

existiert. Der Grenzwert heißt dann *partielle Ableitung* (von  $f$  nach der  $j$ -ten Variablen an der Stelle  $\xi$ ) und wird auch mit  $\frac{\partial f}{\partial x_j}(\xi)$  oder  $f_{x_j}(\xi)$  bezeichnet.

Definiert man für ein geeignetes Intervall  $I$ , welches  $\xi_j$  als inneren Punkt enthält,  $g : I \rightarrow \mathbb{R}$  durch  $g(x) = f(\xi_1, \dots, \xi_{j-1}, x, \xi_{j+1}, \dots, \xi_p)$ , so existiert also  $\partial_j f(\xi)$  genau dann, wenn  $g'(\xi_j)$  existiert, und es gilt dann  $\partial_j f(\xi) = g'(\xi_j)$ . Man berechnet also  $\partial_j f$  durch Differenzieren nach  $x_j$  bei festgehaltenem  $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p$ .

Ist  $p = 2$  bzw.  $p = 3$ , so bezeichnet man die Variablen statt mit  $x_1, x_2$  (und  $x_3$ ) in der Regel mit  $x, y$  bzw.  $x, y, z$ . Dementsprechend schreibt man auch zum Beispiel  $\frac{\partial f}{\partial y}$  und  $f_y$  statt  $\frac{\partial f}{\partial x_2}$  und  $f_{x_2}$ . Analog verfährt man bei anderen Bezeichnungen für die Variablen.

Statt  $\partial_j f(\xi)$  findet man in der Literatur auch die Bezeichnung  $D_j f(\xi)$ . Üblich (aber gelegentlich problematisch) sind auch Schreibweisen wie  $\frac{\partial f(x,y)}{\partial x}$  oder  $\frac{\partial f}{\partial x}(x, y)$  für  $\partial_1 f(x, y)$ .

*Beispiele.* 1. Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $f(x, y) = xe^y + \sin(xy)$ . Dann ist

$$\partial_1 f(x, y) = e^y + y \cos(xy) \quad \text{und} \quad \partial_2 f(x, y) = xe^y + x \cos(xy).$$

2. Sei  $f : \mathbb{R}^p \setminus \{0\} \rightarrow \mathbb{R}$ ,  $f(x) = 1/\|x\|_2^2$ , also

$$f(x_1, \dots, x_p) = \frac{1}{x_1^2 + \dots + x_p^2}.$$

Dann gilt

$$\partial_j f(x_1, \dots, x_p) = -\frac{2x_j}{(x_1^2 + \dots + x_p^2)^2}.$$

Ist  $U \subset \mathbb{R}^p$  offen,  $k \in \{1, \dots, p\}$ ,  $f : U \rightarrow \mathbb{R}$  und existiert  $\partial_k f(x)$  für alle  $x \in U$ , so ist durch  $x \mapsto \partial_k f(x)$  eine Funktion  $\partial_k f : U \rightarrow \mathbb{R}$  gegeben. Man kann auch diese wieder auf partielle Differenzierbarkeit untersuchen. Für  $\partial_j \partial_k f$  schreiben wir dabei kurz  $\partial_{jk}$  (oder  $\partial_{j,k}$ ), bzw.  $\frac{\partial^2 f}{\partial x_j \partial x_k}$  oder  $f_{x_k x_j}$ .

Entsprechend definiert man höhere partielle Ableitungen, etwa

$$\partial_1 \partial_3^2 \partial_2 f = \frac{\partial^4 f}{\partial x_1 \partial x_3^2 \partial x_2} = f_{x_2 x_3 x_3 x_1}.$$

Die Anzahl der auftretenden Ableitungen wird also *Ordnung* der Ableitung bezeichnet, im angegebenen Beispiel handelt es sich also um eine partielle Ableitung 4. Ordnung.

*Beispiele.* 1. Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $f(x, y) = xe^y + \sin(xy)$  wie oben, mit

$$\partial_1 f(x, y) = e^y + y \cos(xy) \quad \text{und} \quad \partial_2 f(x, y) = xe^y + x \cos(xy).$$

Es ist

$$\begin{aligned} \partial_{11} f(x, y) &= -y^2 \sin(xy), \\ \partial_{21} f(x, y) &= e^y + \cos(xy) - xy \sin(xy), \\ \partial_{12} f(x, y) &= e^y + \cos(xy) - xy \sin(xy) = \partial_{21} f(x, y), \\ \partial_{22} f(x, y) &= xe^y - x^2 \sin(xy). \end{aligned}$$

2. Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & \text{falls } (x, y) \neq (0, 0), \\ 0 & \text{falls } (x, y) = (0, 0), \end{cases}$$

Dann gilt

$$\partial_1 f(0, y) = \lim_{x \rightarrow 0} \frac{f(x, y)}{x} = \begin{cases} 0 & \text{falls } y = 0, \\ \lim_{x \rightarrow 0} y \frac{x^2 - y^2}{x^2 + y^2} & \text{falls } y \neq 0 \end{cases} = -y$$

und analog  $\partial_2 f(x, 0) = x$ . Es folgt  $\partial_{12} f(0, 0) = 1 \neq -1 = \partial_{21} f(0, 0)$ .

**Satz 7.1.1 (Schwarz)** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$ ,  $f : U \rightarrow \mathbb{R}$  und  $j, k \in \{1, \dots, p\}$ . Es sei weiter vorausgesetzt, dass die zweiten partiellen Ableitungen  $\partial_{jk}$  und  $\partial_{kj}$  in  $U$  existieren und in  $\xi$  stetig sind. Dann gilt  $\partial_{jk} f(\xi) = \partial_{kj} f(\xi)$ .

*Beweis.* Ohne Beschränkung der Allgemeinheit sei  $p = 2$ ,  $j = 1$ ,  $k = 2$ . Zu zeigen ist also  $\partial_{12} f(\xi) = \partial_{21} f(\xi)$ .

Sei  $\varepsilon > 0$  mit  $U(\xi, \varepsilon) \subset U$  (wobei  $U(\xi, \varepsilon) \subset U$  mit der Norm  $\|\cdot\|_\infty$  gebildet sei). Es sei  $\xi = (a, b)$  und

$$F : U(\xi, \varepsilon) \rightarrow \mathbb{R}, \quad F(x, y) = f(x, y) - f(x, b) - f(a, y) + f(a, b).$$

Für festes  $y \in (b - \varepsilon, b + \varepsilon)$ ,  $y \neq b$ , betrachten wir

$$\phi : (a - \varepsilon, a + \varepsilon) \rightarrow \mathbb{R}, \quad \phi(x) = f(x, y) - f(x, b).$$

Es ist dann  $F(x, y) = \phi(x) - \phi(a)$ . Dann ist  $\phi$  differenzierbar und nach Mittelwertsatz existiert für  $x \neq a$  dann  $\alpha_1$  zwischen  $a$  und  $x$  mit

$$F(x, y) = \phi(x) - \phi(a) = (x - a)\phi'(\alpha_1) = (x - a)(\partial_1 f(\alpha_1, y) - \partial_1 f(\alpha_1, b)).$$

Erneute Anwendung des Mittelwertsatzes liefert  $\beta_1$  zwischen  $b$  und  $y$  mit

$$F(x, y) = (x - a)(y - b)\partial_{21} f(\alpha_1, \beta_1).$$

Jetzt betrachten wir für festes  $x \in (a - \varepsilon, a + \varepsilon)$ ,  $x \neq a$ , die Funktion

$$\psi : (b - \varepsilon, b + \varepsilon) \rightarrow \mathbb{R}, \quad \psi(x) = f(x, y) - f(a, y).$$

Man findet für  $y \neq b$  dann  $\beta_2$  zwischen  $b$  und  $y$  mit

$$F(x, y) = \psi(y) - \psi(b) = (y - b)\psi'(\beta_2) = (y - b)(\partial_2 f(x, \beta_2) - \partial_2 f(a, \beta_2))$$

und anschließend  $\alpha_2$  zwischen  $a$  und  $x$  mit

$$F(x, y) = (x - a)(y - b)\partial_{12}f(\alpha_2, \beta_2).$$

Es folgt

$$\partial_{21}f(\alpha_1, \beta_1) = \partial_{12}f(\alpha_2, \beta_2).$$

Da für  $(x, y) \rightarrow (a, b)$  auch  $(\alpha_1, \beta_1) \rightarrow (a, b)$  und  $(\alpha_2, \beta_2) \rightarrow (a, b)$  gilt, folgt die Behauptung jetzt aus der Stetigkeit von  $\partial_{12}f$  und  $\partial_{21}f$  in  $\xi = (a, b)$ .  $\square$

**Satz 7.1.2** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$  und  $f : U \rightarrow \mathbb{R}$ . Die partiellen Ableitungen von  $f$  mögen existieren und sie seien in  $\xi = (\xi_1, \dots, \xi_p)^T$  stetig. Dann gilt

$$\frac{f(x) - f(\xi) - \sum_{j=1}^p \partial_j f(\xi)(x_j - \xi_j)}{\|x - \xi\|} \rightarrow 0$$

für  $x = (x_1, \dots, x_p)^T \rightarrow \xi$ . Desweiteren ist  $f$  stetig in  $\xi$ .

*Bemerkungen.* 1. Wegen der Äquivalenz der Normen in  $\mathbb{R}^p$  (Satz 6.5.6) ist es irrelevant, welche Norm in Satz 7.1.2 betrachtet wird.

2. Aus der Existenz der partiellen Ableitungen allein folgt noch nicht die Stetigkeit, man vgl. etwa das Beispiel nach Satz 6.4.5 oder betrachte eine beliebige Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , die  $f(0, t) = f(t, 0) = 0$  für  $t \in \mathbb{R}$  erfüllt. Dann gilt  $\partial_1 f(0, 0) = \partial_2 f(0, 0) = 0$ , aber  $f$  braucht nicht stetig sein.

3. Mit den Landau-Symbolen lässt sich die Aussage von Satz 7.1.2 auch in der Form

$$f(x) - f(\xi) = \sum_{j=1}^p \partial_j f(\xi)(x_j - \xi_j) + o(\|x - \xi\|)$$

für  $x \rightarrow \xi$  schreiben. Sie lässt sich wie folgt interpretieren: Die durch

$$h = (h_1, \dots, h_n)^T \mapsto \sum_{j=1}^p \partial_j f(\xi)h_j$$

definierte lineare Funktion  $T : \mathbb{R}^p \rightarrow \mathbb{R}$  ist (in der Nähe von 0) eine gute Approximation der durch  $h \mapsto f(\xi + h) - f(\xi)$  definierten Funktion, denn

$$f(\xi + h) - f(\xi) = T(h) + o(\|h\|)$$

für  $h \rightarrow 0$ . Wir können dies auch als

$$f(x) - f(\xi) = T(x - \xi) + o(\|x - \xi\|)$$

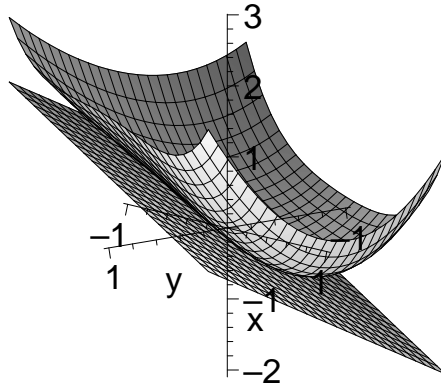


Abbildung 24: Approximation des Graphen durch eine Ebene.

für  $x \rightarrow \xi$  schreiben. Leicht sieht man, dass  $T$  die einzige lineare Funktion mit dieser Eigenschaft ist. Wir werden diese lineare Abbildung  $T$  später Ableitung von  $f$  nennen.

Der Graph der linearen Funktion  $T : \mathbb{R}^p \rightarrow \mathbb{R}$  ist ein  $p$ -dimensionaler Teilraum des  $\mathbb{R}^{p+1}$ , eine sogenannte Hyperebene (durch 0), im Falle  $p = 2$  also eine Ebene im  $\mathbb{R}^3$  (durch 0). Für  $f(x, y) = \sin(x - y) + x^2 + y^4$  und  $\xi = (0, 0)$  ist der Graph von  $f$  und  $T$  in Abbildung 24 dargestellt.

*Beweis von Satz 7.1.2.* Sei  $\varepsilon > 0$  mit  $U(\xi, \varepsilon) \subset U$  (wobei  $U(\xi, \varepsilon) \subset U$  wieder mit der Norm  $\|\cdot\|_\infty$  gebildet sei). Für  $0 < \|x - \xi\|_\infty < \varepsilon$  gilt dann

$$\begin{aligned}
 f(x) - f(\xi) &= f(x_1, \dots, x_p) - f(\xi_1, \dots, \xi_p) \\
 &= f(x_1, \dots, x_p) - f(\xi_1, \dots, \xi_{p-1}, x_p) \\
 &\quad + f(\xi_1, \dots, \xi_{p-1}, x_p) - f(\xi_1, \dots, \xi_{p-2}, x_{p-1}, x_p) \\
 &\quad + \dots \\
 &\quad + f(\xi_1, x_2, \dots, x_p) - f(x_1, \dots, x_p) \\
 &= \sum_{j=1}^p f(\xi_1, \dots, \xi_j, x_{j+1}, \dots, x_p) - f(\xi_1, \dots, \xi_{j-1}, x_j, \dots, x_p) \\
 &= \sum_{j=1}^p (x_j - \xi_j) \partial_j f(\xi_1, \dots, \xi_{j-1}, y_j, x_{j+1}, \dots, x_p) \\
 &= \sum_{j=1}^p (x_j - \xi_j) \partial_j f(\eta^j)
 \end{aligned}$$

wobei  $y_j$  zwischen  $\xi_j$  und  $x_j$  liegt und  $\eta^j := (\xi_1, \dots, \xi_{j-1}, y_j, x_{j+1}, \dots, x_p)^T$ , also  $\eta^j \in U(\xi, \|x - \xi\|_\infty)$  gilt. Mit  $T$  wie in obiger Bemerkung folgt also

$$f(x) - f(\xi) - T(x - \xi) = \sum_{j=1}^p (x_j - \xi_j) (\partial_j f(\eta^j) - \partial_j f(\xi))$$

und damit

$$\frac{|f(x) - f(\xi) - T(x - \xi)|}{\|x - \xi\|_\infty} \leq \sum_{j=1}^p |\partial_j f(\eta^j) - \partial_j f(\xi)|.$$



Die erste Behauptung folgt jetzt aus der Stetigkeit der  $\partial_j f$ . Die Stetigkeit von  $f$  folgt da  $T(x - \xi) \rightarrow 0$  für  $x \rightarrow \xi$ .  $\square$

Die partiellen Ableitungen beschreiben das Änderungsverhalten einer Funktion in Richtung der Koordinatenachsen. Analoges lässt sich auch für andere Richtungen machen.

**Definition 7.1.2** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$ ,  $f : U \rightarrow \mathbb{R}$  und  $v \in \mathbb{R}^p$  mit  $\|v\|_2 = 1$ . Dann heißt

$$\partial_v f(\xi) := \frac{\partial f}{\partial v}(\xi) := \lim_{t \rightarrow 0} \frac{f(\xi + tv) - f(\xi)}{t}$$

die *Richtungsableitung* von  $f$  an der Stelle  $\xi$  in Richtung  $v$ , falls der Grenzwert existiert.

Mit der Standardbasis  $\{e_1, \dots, e_p\}$  des  $\mathbb{R}^p$  gilt also  $\partial_{e_j} = \partial_j$  für alle  $j$ .

**Satz 7.1.3** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$ ,  $f : U \rightarrow \mathbb{R}$  und  $v \in \mathbb{R}^p$  mit  $\|v\|_2 = 1$ . Die partiellen Ableitungen von  $f$  mögen existieren und sie seien in  $\xi$  stetig. Dann existiert  $\partial_v f(\xi)$  und es gilt

$$\partial_v f(\xi) = \sum_{j=1}^p \partial_j f(\xi) v_j.$$

*Beweis.* Nach Satz 7.1.2 gilt

$$\lim_{t \rightarrow 0} \frac{f(\xi + tv) - f(\xi) - \sum_{j=1}^p \partial_j f(\xi) v_j t}{t} = 0,$$

und daraus folgt die Behauptung.  $\square$

**Definition 7.1.3** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$  und  $f : U \rightarrow \mathbb{R}$ . Dann heißt  $\text{grad} f(\xi) := (\partial_1 f(\xi), \dots, \partial_p f(\xi))$  der *Gradient* von  $f$  an der Stelle  $\xi$ , falls die partiellen Ableitungen existieren.

Unter den Voraussetzungen von Satz 7.1.3 gilt dann  $\partial_v f(\xi) = \text{grad} f(\xi) \cdot v$ . (Hierbei steht “ $\cdot$ ” für das Matrixprodukt der  $(1 \times p)$ -Matrix  $\text{grad} f(\xi)$  mit der  $(p \times 1)$ -Matrix  $(v_1, \dots, v_p)^T$ . Wir können dies auch als Skalarprodukt auffassen:  $\partial_v f(\xi) = \langle \text{grad} f(\xi), v \rangle$ .)

In Satz 7.1.2 erhalten wir

$$\frac{f(x) - f(\xi) - \text{grad} f(\xi)(x - \xi)}{\|x - \xi\|} \rightarrow 0$$

für  $x \rightarrow \xi$ .

**Definition 7.1.4** Seien  $p, q \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$  und  $f = (f_1, \dots, f_q)^T : U \rightarrow \mathbb{R}^q$ . Dann heißt die  $(q \times p)$ -Matrix

$$J_f(\xi) := \begin{pmatrix} \partial_1 f_1(\xi) & \dots & \partial_p f_1(\xi) \\ \partial_1 f_2(\xi) & \dots & \partial_p f_2(\xi) \\ \vdots & & \vdots \\ \partial_1 f_q(\xi) & \dots & \partial_p f_q(\xi) \end{pmatrix}$$

die *Funktionalmatrix* (oder *Jacobimatrix*) von  $f$  an der Stelle  $\xi$ , falls die partiellen Ableitungen existieren.

Die  $k$ -te Zeile von  $J_f(\xi)$  ist also durch  $\text{grad}f_k(\xi)$  gegeben.

**Satz 7.1.4** Seien  $p, q \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $\xi \in U$  und  $f = (f_1, \dots, f_q)^T : U \rightarrow \mathbb{R}^q$ . Die partiellen Ableitungen der  $f_k$  mögen existieren und sie seien in  $\xi$  stetig. Dann gilt

$$\frac{f(x) - f(\xi) - J_f(\xi)(x - \xi)}{\|x - \xi\|} \rightarrow 0$$

für  $x \rightarrow \xi$ .

*Beweis.* Die Behauptung folgt unmittelbar aus Satz 7.1.2 (bzw. der auf Definition 7.1.3 folgenden Bemerkung). Denn die  $k$ -te Koordinate des Zählers ist durch  $f_k(x) - f_k(\xi) - \text{grad}f_k(\xi)(x - \xi)$  gegeben.  $\square$

Für  $U \subset \mathbb{R}^p$  offen und  $n \in \mathbb{N}$  bezeichnet man die Menge aller Funktionen von  $U$  nach  $\mathbb{R}^q$ , für die alle partiellen Ableitungen bis zur Ordnung  $n$  existieren und stetig sind, mit  $C^n(U, \mathbb{R}^q)$  oder, falls der Zielraum aus dem Zusammenhang klar ist, auch mit  $C^n(U)$ . Für solche Funktionen sind nach dem Satz von Schwarz die partiellen Ableitungen bis zur Ordnung  $n$  unabhängig von der Reihenfolge der Differentiation. Weiter setzt man  $C^\infty(U, \mathbb{R}^q) := \bigcap_{n=0}^\infty C^n(U, \mathbb{R}^q)$ .

## 7.2 Differenzierbarkeit

Reelle  $(q \times p)$ -Matrizen korrespondieren zu linearen Abbildungen von  $\mathbb{R}^p$  nach  $\mathbb{R}^q$ . Insbesondere entspricht die Funktionalmatrix einer linearen Abbildung. Beachtet man noch, dass lineare Abbildungen von  $\mathbb{R}^p$  nach  $\mathbb{R}^q$  stetig sind (siehe §6.4), so führt dies auf folgende Definition.

**Definition 7.2.1** Seien  $(V, \|\cdot\|_V)$ ,  $(W, \|\cdot\|_W)$  Banachräume,  $U \subset V$  offen,  $\xi \in U$  und  $f : U \rightarrow W$ . Dann heißt  $f$  (total) differenzierbar in  $\xi$ , falls  $T \in L(V, W)$  existiert, so dass

$$\lim_{x \rightarrow \xi} \frac{f(x) - f(\xi) - T(x - \xi)}{\|x - \xi\|_V} = \lim_{h \rightarrow 0} \frac{f(\xi + h) - f(\xi) - T(h)}{\|h\|_V} = 0.$$

Die Abbildung  $T$  ist dann eindeutig bestimmt (siehe die 2. Bemerkung unten) und heißt (totale) Ableitung (oder Fréchet-Ableitung) von  $f$  an der Stelle  $\xi$ . Sie wird mit  $Df(\xi)$  bezeichnet.

Die Funktion  $f$  heißt (total) differenzierbar, falls  $f$  in jedem Punkt von  $U$  differenzierbar ist. Falls die durch  $x \mapsto Df(x)$  gegebene Funktion  $Df : U \rightarrow L(V, W)$  dann stetig ist, heißt  $f$  stetig differenzierbar.

*Bemerkungen.* 1. Mit  $L(V, W)$  hatten wir den Vektorraum der stetigen linearen Abbildungen von  $V$  nach  $W$  bezeichnet; vgl. §6.4.

2. Wir zeigen, dass  $T$  in obiger Definition eindeutig bestimmt ist. Dazu nehmen wir an, dass  $T_1, T_2 \in L(V, W)$  mit  $T_1 \neq T_2$  existieren, die die in der Definition angegebene Eigenschaft haben. Dann gilt

$$\lim_{h \rightarrow 0} \frac{T_1(h) - T_2(h)}{\|h\|_V} = \lim_{h \rightarrow 0} \frac{(T_1 - T_2)(h)}{\|h\|_V} = 0.$$

Andererseits existiert  $y \in V$  mit  $T_1(y) \neq T_2(y)$ . Es folgt

$$\frac{(T_1 - T_2)(ty)}{\|ty\|_V} = \frac{t(T_1 - T_2)(y)}{t\|y\|_V} = \frac{(T_1 - T_2)(y)}{\|y\|_V}$$

für  $t \in \mathbb{R} \setminus \{0\}$  und damit

$$\lim_{t \rightarrow 0} \frac{(T_1 - T_2)(ty)}{\|ty\|_V} = \frac{(T_1 - T_2)(y)}{\|y\|_V} \neq 0,$$

ein Widerspruch.

3. Die Interpretation der in obiger Definition gegebenen Eigenschaft ist analog zu der von Satz 7.1.2; vgl. Bemerkung 3 dort. Die lineare Abbildung  $Df(\xi)$  ist (in der Nähe von 0) eine gute Approximation der durch  $h \mapsto f(\xi + h) - f(\xi)$  gegebenen Funktion. Denn ist  $R(h) := f(\xi + h) - f(\xi) - Df(\xi)(h)$ , so gilt  $R(h) = o(\|h\|_V)$  für  $h \rightarrow 0$ . Es ist  $Df(\xi)$  die einzige lineare Abbildung mit dieser Eigenschaft; vgl. die vorige Bemerkung.

4. Im allgemeinen hängt  $Df$  von den Normen  $\|\cdot\|_V, \|\cdot\|_W$  ab. Dies gilt aber nicht, falls  $V = \mathbb{R}^p, W = \mathbb{R}^q$ , da dort alle Normen äquivalent sind (Satz 6.5.6).

*Beispiel.* Sei  $f \in L(V, W)$ . Dann ist  $f$  differenzierbar und es gilt  $Df(\xi) = f$  für alle  $\xi \in V$ .

**Satz 7.2.1** *Differenzierbare Funktionen sind stetig.*

*Beweis.* Seien  $U, V, W, f$  wie in Definition 7.2.1 und sei  $f$  differenzierbar in  $\xi \in U$ . Für  $x \in U$  sei  $R(x) := f(x) - f(\xi) - Df(\xi)(x - \xi)$ . Nach Definition der Differenzierbarkeit gilt dann  $R(x)/\|x - \xi\|_V \rightarrow 0$  für  $x \rightarrow \xi$ . Es folgt  $R(x) \rightarrow 0$  für  $x \rightarrow \xi$ . Da  $Df(\xi)$  stetig ist, folgt auch  $Df(\xi)(x - \xi) \rightarrow 0$  für  $x \rightarrow \xi$ . Insgesamt folgt  $f(x) \rightarrow f(\xi)$  für  $x \rightarrow \xi$ . Damit ist  $f$  stetig in  $\xi$ .  $\square$

Man beachte, dass im Beweis die Stetigkeit von  $Df(\xi)$  benutzt wird. Dies ist ein Grund dafür, in Definition 7.2.1  $T$  als linear *und stetig* zu fordern.

**Satz 7.2.2** *Seien  $p, q \in \mathbb{N}, U \subset \mathbb{R}^p$  offen,  $\xi \in U$  und  $f = (f_1, \dots, f_q)^T : U \rightarrow \mathbb{R}^q$ . Dann sind äquivalent:*

- (i)  $f$  ist differenzierbar in  $\xi$ .
- (ii) Alle partiellen Ableitungen  $\partial_k f_j(\xi)$  existieren und es gilt

$$\frac{f(x) - f(\xi) - J_f(\xi)(x - \xi)}{\|x - \xi\|} \rightarrow 0$$

für  $x \rightarrow \xi$ .

Es gilt dann  $Df(\xi)(h) = J_f(\xi)(h)$  für alle  $h \in \mathbb{R}^p$ , d. h., die Matrix  $J_f(\xi)$  entspricht der linearen Abbildung  $Df(\xi)$  bzgl. der Standardbasis.

*Beweis.* (ii)  $\Rightarrow$  (i). Diese Richtung ist trivial, denn offensichtlich hat die durch  $T(h) = J_f(\xi)h$  definierte Funktion  $T \in L(V, W)$  die in Definition 7.2.1 verlangte Eigenschaft.

(i)  $\Rightarrow$  (ii). Sei  $A = (a_{ij})_{\substack{1 \leq i \leq q \\ 1 \leq j \leq p}}$  die zu  $Df(\xi)$  gehörige Matrix (bzgl. der Standardbasen  $\{e_1, e_2, \dots\}$ ). Dann folgt

$$\frac{f(\xi + h) - f(\xi) - Ah}{\|h\|} \rightarrow 0$$

für  $h \rightarrow 0$  und damit

$$\frac{f_j(\xi + te_k) - f_j(\xi) - a_{jk}t}{t} \rightarrow 0$$

für  $t \rightarrow 0$ . Hieraus folgt, dass die partiellen Ableitungen  $\partial_k f_j(\xi)$  existieren und dass  $a_{jk} = \partial_k f_j(\xi)$  und  $A = J_f(\xi)$  gilt.  $\square$

Aus der totalen Differenzierbarkeit in  $\xi$  folgt also mit Satz 7.2.2 die partielle Differenzierbarkeit in  $\xi$ , aber nicht umgekehrt (vgl. die Beispiele nach Satz 7.1.2, wo die partiellen Ableitungen existieren, aber die Funktion sogar unstetig ist). Existieren aber die partiellen Ableitungen in einer Umgebung von  $\xi$  und sind sie in  $\xi$  stetig, so ist  $f$  in  $\xi$  total differenzierbar; vgl. Satz 7.1.4 und 7.2.2.

Für stetige Differenzierbarkeit sind die Verhältnisse einfacher.

**Satz 7.2.3** *Seien  $p, q \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen und  $f = (f_1, \dots, f_q)^T : U \rightarrow \mathbb{R}^q$ . Dann ist  $f$  genau dann stetig differenzierbar, wenn  $f \in C^1(U, \mathbb{R}^q)$ , d. h., wenn alle partiellen Ableitungen der Koordinatenfunktionen von  $f$  existieren und stetig sind.*

*Beweis.* Da aus der Stetigkeit der partiellen Ableitungen die Differenzierbarkeit folgt, können wir  $f$  als differenzierbar annehmen. Für  $h \in \mathbb{R}^p$  und  $x, \xi \in U$  ist dann  $(Df(x) - Df(\xi))(h) = (J_f(x) - J_f(\xi))(h)$ . Legen wir in  $\mathbb{R}^p, \mathbb{R}^q$  die Norm  $\|\cdot\|_\infty$  zugrunde, so erhalten wir (vgl. das Beispiel nach Satz 6.4.6)

$$\|Df(x) - Df(\xi)\|_{L(\mathbb{R}^p, \mathbb{R}^q)} = \max_j \sum_{k=1}^p |\partial_k f_j(x) - \partial_k f_j(\xi)|.$$

Hieraus folgt leicht die Behauptung.  $\square$

Die üblichen Regeln für die Differenzierbarkeit von  $f + g$ , und  $\lambda \cdot f$ , wobei  $f, g : U \rightarrow W$  und  $\lambda \in \mathbb{K}$ , übertragen sich unmittelbar. Entsprechendes gilt auch für  $f \cdot g$  und  $f/g$ , falls  $W = \mathbb{R}$  ist; vgl. auch Satz 6.4.4. Wir formulieren die Kettenregel:

**Satz 7.2.4** *Seien  $(V, \|\cdot\|_V)$ ,  $(W, \|\cdot\|_W)$ ,  $(X, \|\cdot\|_X)$  Banachräume,  $U \subset V$  offen,  $U' \subset W$  offen,  $\xi \in U$ ,  $g : U \rightarrow W$  differenzierbar in  $\xi$ ,  $g(U) \subset U'$  und  $f : U' \rightarrow X$  differenzierbar in  $g(\xi)$ . Dann ist  $f \circ g$  differenzierbar in  $\xi$  und*

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

*Beweis.* Es ist

$$R_g(h) := g(\xi + h) - g(\xi) - Dg(\xi)(h) = o(\|h\|_V)$$

für  $h \rightarrow 0$  und

$$R_f(k) := f(g(\xi) + k) - f(g(\xi)) - Df(g(\xi))(k) = o(\|k\|_W)$$

für  $k \rightarrow 0$ . Mit  $\Delta(h) := g(\xi + h) - g(\xi) = Dg(\xi)(h) + R_g(h)$  folgt

$$\begin{aligned} & f(g(\xi + h)) - f(g(\xi)) \\ &= f(g(\xi) + \Delta(h)) - f(g(\xi)) \\ &= Df(g(\xi))(\Delta(h)) + R_f(\Delta(h)) \\ &= (Df(g(\xi)) \circ Dg(\xi))(h) + Df(g(\xi))(R_g(h)) + R_f(\Delta(h)). \end{aligned}$$

Zu zeigen ist, dass  $R(h) := Df(g(\xi))(R_g(h)) + R_f(\Delta(h)) = o(\|h\|_V)$  für  $h \rightarrow 0$ . Wegen

$$\|Df(g(\xi))(R_g(h))\|_X \leq \|Df(g(\xi))\|_{L(W,X)} \|R_g(h)\|_W = o(\|h\|_V)$$

für  $h \rightarrow 0$  müssen wir nur noch  $\|R_f(\Delta(h))\|_X = o(\|h\|_V)$  für  $h \rightarrow 0$  zeigen. Dies folgt aber, da  $\Delta(h) = O(\|h\|_V)$  und damit  $\Delta(h) \rightarrow 0$  für  $h \rightarrow 0$ , und da

$$\frac{\|R_f(\Delta(h))\|_X}{\|h\|_V} = \frac{\|R_f(\Delta(h))\|_X}{\|\Delta(h)\|_W} \cdot \frac{\|\Delta(h)\|_W}{\|h\|_V}$$

falls  $\Delta(h) \neq 0$ . Für  $h \rightarrow 0$  strebt der erste Faktor rechts gegen 0, während der zweite beschränkt bleibt.  $\square$

Im Falle  $V = \mathbb{R}^p$ ,  $W = \mathbb{R}^q$ ,  $X = \mathbb{R}^r$  folgt aus Satz 7.2.4, dass

$$J_{f \circ g}(\xi) = J_f(g(\xi)) \cdot J_g(\xi).$$

*Beispiel.* Sei  $U \subset \mathbb{R}^p$  offen und  $I$  Intervall. Es seien  $f : U \rightarrow \mathbb{R}$  und  $\gamma : I \rightarrow U$  differenzierbar. Dann ist  $f \circ \gamma$  differenzierbar und es gilt

$$(f \circ \gamma)'(t) = (\partial_1 f(\gamma(t)), \partial_2 f(\gamma(t)), \dots, \partial_p f(\gamma(t))) \cdot \begin{pmatrix} \gamma'_1(t) \\ \gamma'_2(t) \\ \vdots \\ \gamma'_p(t) \end{pmatrix} = \sum_{j=1}^p (\partial_j f \circ \gamma)(t) \gamma'_j(t).$$

Existieren auch die zweiten Ableitungen, so folgt

$$\begin{aligned} (f \circ \gamma)''(t) &= \sum_{j=1}^p (\partial_j f \circ \gamma)'(t) \gamma'_j(t) + (\partial_j f \circ \gamma)(t) \gamma''_j(t) \\ &= \sum_{j=1}^p \left( \sum_{k=1}^p \partial_k (\partial_j f \circ \gamma)(t) \gamma'_k(t) \right) \gamma'_j(t) + \sum_{j=1}^p (\partial_j f \circ \gamma)(t) \gamma''_j(t) \\ &= \sum_{j=1}^p \sum_{k=1}^p \partial_{k,j} (f \circ \gamma)(t) \gamma'_k(t) \gamma'_j(t) + \sum_{j=1}^p (\partial_j f \circ \gamma)(t) \gamma''_j(t) \end{aligned}$$

**Definition 7.2.2** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $f : U \rightarrow \mathbb{R}$  und  $\xi \in U$ . Existieren die zweiten partiellen Ableitungen von  $f$  in  $\xi$ , so heißt

$$H_f(\xi) := (\partial_{ij}f(\xi))_{\substack{1 \leq i \leq p \\ 1 \leq j \leq p}}$$

Hesse-Matrix von  $f$  an der Stelle  $\xi$ .

Nach dem Satz von Schwarz ist die Hesse-Matrix für  $f \in C^2(U, \mathbb{R})$  symmetrisch.

Mit dieser Definition erhält man im obigen Beispiel

$$(f \circ \gamma)'(t) = \text{grad}f(\gamma(t)) \cdot \gamma'(t)$$

und

$$(f \circ \gamma)''(t) = \gamma'(t)^T \cdot H_f(\gamma(t)) \cdot \gamma'(t) + \text{grad}f(\gamma(t)) \cdot \gamma''(t).$$

Sei nun  $\xi \in U$  und  $v \in \mathbb{R}^p$  mit  $[\xi, \xi + v] \subset U$ . Mit  $\gamma : [0, 1] \rightarrow U$ ,  $\gamma(t) = \xi + tv$  folgt  $\gamma'(t) = v$ ,  $\gamma''(t) = 0$ ,

$$(f \circ \gamma)'(0) = \text{grad}f(\xi) \cdot v$$

und

$$(f \circ \gamma)''(0) = v^T \cdot H_f(\xi) \cdot v.$$

Der Taylorsche Satz (Satz 5.4.2) liefert nun

$$\begin{aligned} f(\xi + tv) &= (f \circ \gamma)(t) \\ &= (f \circ \gamma)(0) + (f \circ \gamma)'(0)t + \frac{1}{2}(f \circ \gamma)''(0)t^2 + o(t^2) \\ &= f(\xi) + (\text{grad}f(\xi) \cdot v)t + \frac{1}{2}(v^T \cdot H_f(\xi) \cdot v)t^2 + o(t^2) \end{aligned}$$

für  $t \rightarrow 0$ . Mit  $h = tv$  erhält man folgenden Satz.

**Satz 7.2.5** Sei  $p \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $f \in C^2(U, \mathbb{R})$  und  $\xi \in U$ . Dann gilt

$$f(\xi + h) = f(\xi) + \text{grad}f(\xi) \cdot h + \frac{1}{2}h^T \cdot H_f(\xi) \cdot h + o(\|h\|^2)$$

für  $h \rightarrow 0$ ,  $h \in \mathbb{R}^p$ .

### 7.3 Lokale Extrema

In Definition 4.2.1 haben wir für Funktionen  $f : M \rightarrow \mathbb{R}$  mit  $M \subset \mathbb{C}$  definiert, wann  $f$  ein lokales bzw. globales Extremum hat. Diese Begriffe können wörtlich auf den Fall, dass  $M$  metrischer Raum ist, übertragen werden.

Für  $f \in C^2(I, \mathbb{R})$  mit einem Intervall  $I$  und  $\xi \in \text{int}(I)$  hatten wir in Satz 4.3.3 und Satz 4.2.1 gezeigt:

$$\begin{aligned} & f'(\xi) = 0 \text{ und } f''(\xi) > 0 \text{ (bzw. } < 0) \\ \Rightarrow & f \text{ hat lokales Minimum (bzw. Maximum) in } \xi \\ \Rightarrow & f'(\xi) = 0 \end{aligned}$$

Wir wollen diese Aussagen jetzt auf Funktionen  $f \in C^2(U, \mathbb{R})$  mit  $U \subset \mathbb{R}^p$  übertragen. Dabei ersetzen wir  $f'$  durch den Gradienten und  $f''$  durch die Hesse-Matrix.

**Satz 7.3.1** Sei  $U \subset \mathbb{R}^p$  offen und  $f \in C^1(U, \mathbb{R})$  habe in  $\xi \in U$  ein lokales Extremum. Dann gilt  $\text{grad}f(\xi) = 0$ .

*Beweis.* Für  $j \in \{1, \dots, p\}$  hat die durch  $t \mapsto f(\xi + te_j)$  definierte Funktion  $g_j$  ein lokales Extremum in 0. Nach Satz 4.3.3 folgt  $\partial_j f(\xi) = g_j'(0) = 0$ .  $\square$

*Beispiel.* Sei

$$K := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, y \geq 0\}$$

und

$$f : K \rightarrow \mathbb{R}, \quad f(x, y) = x^2 + y^2 - \sqrt{3}x - y.$$

Man bestimme  $\max_{(x,y) \in K} f(x, y)$  und  $\min_{(x,y) \in K} f(x, y)$ . (Die Existenz von Maximum und Minimum folgt, da  $K$  kompakt und  $f$  stetig.)

Für ein im Inneren von  $K$  gelegenes Extremum  $(x, y)$  folgt nach Satz 7.3.1, dass  $(0, 0) = \text{grad}f(x, y) = (2x - \sqrt{3}, 2y - 1)$ , also  $(x, y) = (\frac{1}{2}\sqrt{3}, \frac{1}{2})$ .

Wir untersuchen jetzt den Rand von  $K$  auf Extrema. Mit

$$\Gamma_1 := \{(x, 0) : -2 \leq x \leq 2\} \quad \text{und} \quad \Gamma_2 := \{(2 \cos t, 2 \sin t) : 0 \leq t \leq \pi\}$$

gilt  $\partial K = \Gamma_1 \cup \Gamma_2$ . Desweiteren ist

$$\max_{(x,y) \in \Gamma_1} f(x, y) = \max_{-2 \leq x \leq 2} f(x, 0) = \max_{-2 \leq x \leq 2} g(x)$$

mit  $g(x) := x^2 - \sqrt{3}x$  und

$$\max_{(x,y) \in \Gamma_2} f(x, y) = \max_{0 \leq t \leq \pi} f(2 \cos t, 2 \sin t) = \max_{0 \leq t \leq \pi} h(t)$$

mit  $h(t) := 4 - 2\sqrt{3} \cos t - 2 \sin t$ . Analoges gilt für die Minima.

Wegen  $g'(x) = 2x - \sqrt{3}$  hat  $g'$  nur die Nullstelle  $\frac{1}{2}\sqrt{3}$  und damit sind Extremwerte von  $g$  nur in  $-2, \frac{1}{2}\sqrt{3}$  und  $2$  möglich. Wegen  $h'(t) = 2\sqrt{3} \sin t - 2 \cos t$  hat  $h'$  in  $(0, \pi)$  nur die Nullstelle  $\frac{\pi}{6}$  und damit sind Extremwerte von  $h$  nur in  $0, \frac{\pi}{6}$  und  $\pi$  möglich.

Insgesamt erhält man 5 Kandidaten für mögliche Extremstellen:  $\xi_1 = (\frac{1}{2}\sqrt{3}, \frac{1}{2})$ ,  $\xi_2 = (\frac{1}{2}\sqrt{3}, 0)$ ,  $\xi_3 = (2 \cos \frac{\pi}{6}, 2 \sin \frac{\pi}{6}) = (\sqrt{3}, 1)$ ,  $\xi_4 = (-2, 0)$  und  $\xi_5 = (2, 0)$ .

Wegen  $f(\xi_1) = -1$ ,  $f(\xi_2) = g(\frac{1}{2}\sqrt{3}) = -\frac{3}{4}$ ,  $f(\xi_3) = h(\frac{\pi}{6}) = 0$ ,  $f(\xi_4) = g(-2) = h(\pi) = 4 + 2\sqrt{3}$  und  $f(\xi_5) = g(2) = h(0) = 4 - 2\sqrt{3}$  folgt, dass  $f$  sein globales Maximum  $4 + 2\sqrt{3}$  in  $\xi_4 = (-2, 0)$  und sein globales Minimum  $-1$  in  $\xi_1 = (\frac{1}{2}\sqrt{3}, \frac{1}{2})$  annimmt.

Wir haben hier die Extrema zur Übung mit Hilfe der Differentialrechnung bestimmt. Im konkreten Beispiel hätte man das leichter tun können, in dem man  $f(x, y)$  in der Form  $f(x, y) = (x - \frac{1}{2}\sqrt{3})^2 + (y - \frac{1}{2})^2 - 1$  schreibt.

**Definition 7.3.1** Sei  $A$  reelle, symmetrische  $(p \times p)$ -Matrix. Dann heißt  $A$

- (i) *positiv definit*, falls  $x^T A x > 0$  für alle  $x \in \mathbb{R}^p \setminus \{0\}$ ;
- (ii) *negativ definit*, falls  $x^T A x < 0$  für alle  $x \in \mathbb{R}^p \setminus \{0\}$ ;
- (iii) *positiv semidefinit*, falls  $x^T A x \geq 0$  für alle  $x \in \mathbb{R}^p$ ;

- (iv) *negativ semidefinit*, falls  $x^T Ax \leq 0$  für alle  $x \in \mathbb{R}^p$ ;
- (v) *indefinit*, falls  $A$  weder positiv noch negativ semidefinit ist, das heißt, falls  $x, y \in \mathbb{R}^p$  mit  $x^T Ax > 0$  und  $y^T Ay < 0$  existieren.

Ist  $A$  positiv definit, so ist wegen Satz 6.5.5

$$c := \min_{\substack{x \in \mathbb{R}^p \\ \|x\|=1}} x^T Ax > 0$$

und es gilt

$$x^T Ax = \left( \frac{x}{\|x\|} A \frac{x}{\|x\|} \right) \|x\|^2 \geq c \|x\|^2$$

für alle  $x \in \mathbb{R}^p$ . Entsprechendes gilt im negativ definiten Fall.

**Satz 7.3.2** *Sei  $A$  reelle, symmetrische  $(p \times p)$ -Matrix. Dann sind alle Eigenwerte von  $A$  reell und es gilt:*

- (i)  $A$  ist positiv definit  $\Leftrightarrow$  Alle Eigenwerte von  $A$  sind positiv.
- (ii)  $A$  ist negativ definit  $\Leftrightarrow$  Alle Eigenwerte von  $A$  sind negativ.
- (iii)  $A$  ist positiv semidefinit  $\Leftrightarrow$   $A$  hat keine negativen Eigenwerte.
- (iv)  $A$  ist negativ semidefinit  $\Leftrightarrow$   $A$  hat keine positiven Eigenwerte.
- (v)  $A$  ist indefinit  $\Leftrightarrow$   $A$  hat sowohl positive wie negative Eigenwerte.

**Satz 7.3.3** *Sei*

$$A = (a_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq p}}$$

*reelle, symmetrische  $(p \times p)$ -Matrix. Für  $1 \leq k \leq p$  sei  $A_k$  die  $(k \times k)$ -Matrix, die man aus  $A$  durch Weglassen der letzten  $p - k$  Zeilen und Spalten erhält, also*

$$A_k = (a_{ij})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq k}}.$$

*Dann gilt:*

- (i)  $A$  ist positiv definit  $\Leftrightarrow$  Für alle  $k \in \{1, \dots, p\}$  gilt  $\det(A_k) > 0$ .
- (ii)  $A$  ist negativ definit  $\Leftrightarrow$  Für alle  $k \in \{1, \dots, p\}$  gilt  $(-1)^k \det(A_k) > 0$ .

Für den Beweis der Sätze 7.3.2 und 7.3.3 sei auf die Lineare Algebra verwiesen.

**Satz 7.3.4** *Sei  $U \subset \mathbb{R}^p$  offen,  $f \in C^2(U, \mathbb{R})$ ,  $\xi \in U$  und  $\text{grad}f(\xi) = 0$ . Dann gilt:*

- (i) Ist  $H_f(\xi)$  positiv definit, so hat  $f$  ein lokales Minimum in  $\xi$ .
- (ii) Ist  $H_f(\xi)$  negativ definit, so hat  $f$  ein lokales Maximum in  $\xi$ .
- (iii) Ist  $H_f(\xi)$  indefinit, so hat  $f$  kein lokales Extremum in  $\xi$ .



Im semidefiniten Fall ist keine allgemeine Aussage möglich. (Dies entspricht dem Fall  $f''(\xi) = 0$  im Falle einer Veränderlichen.)

*Beweis von Satz 7.3.4.* Sei  $h \in \mathbb{R}^p$  mit  $\xi + h \in U$ . Aufgrund der Taylorformel (Satz 7.2.5), und da  $\text{grad}f(\xi) = 0$  nach Voraussetzung, gilt

$$f(\xi + h) - f(\xi) = \frac{1}{2}h^T H_f(\xi)h + R(h)$$

mit  $R(h) = o(\|h\|^2)$  für  $h \rightarrow 0$ .

Zu (i). Ist  $H_f(\xi)$  positiv definiert, so existiert  $c > 0$  mit  $h^T H_f(\xi)h \geq c\|h\|^2$  für alle  $h \in \mathbb{R}^p$ . Weiter existiert  $\delta > 0$  mit  $|R(h)| < \frac{c}{2}\|h\|^2$  für  $0 < \|h\| \leq \delta$ . Es folgt

$$f(\xi + h) - f(\xi) \geq \frac{1}{2}h^T H_f(\xi)h - |R(h)| > 0$$

für  $0 < \|h\| \leq \delta$ . Also ist  $\xi$  lokales Minimum.

Zu (ii). Dies folgt analog zu (i).

Zu (iii). Ist  $H_f(\xi)$  indefinit, so existieren  $h_1, h_2 \in \mathbb{R}^p$  mit  $c_1 := h_1^T H_f(\xi)h_1 < 0$  und  $c_2 := h_2^T H_f(\xi)h_2 > 0$ . Für  $t \in \mathbb{R}$  und  $j \in \{1, 2\}$  folgt  $(th_j)^T H_f(\xi)(th_j) = c_j t^2$  und  $R(th_j) = o(\|th_j\|^2) = o(t^2)$  für  $t \rightarrow 0$ . Für genügend kleines  $t$  ist dann  $|R(th_j)| < \frac{1}{2}|c_j|t^2 = \frac{1}{2}|(th_j)^T H_f(\xi)(th_j)|$  und damit

$$f(\xi + th_1) - f(\xi) = \frac{1}{2}(th_1)^T H_f(\xi)(th_1) + R(th_1) \leq \frac{1}{2}c_1 t^2 + |R(th_1)| < 0$$

und

$$f(\xi + th_2) - f(\xi) = \frac{1}{2}(th_2)^T H_f(\xi)(th_2) + R(th_2) \geq \frac{1}{2}c_2 t^2 - |R(th_2)| > 0.$$

Es folgt, dass  $\xi$  kein lokales Extremum ist.  $\square$

Für  $p = 2$  sieht im Falle (iii) der Graph von  $f$ , also die Menge

$$\{(x, y, f(x, y)) : (x, y) \in U\} \subset \mathbb{R}^3,$$

lokal wie ein ‘‘Sattel’’ aus, im Beispiel  $f(x, y) = x^2 - y^2$ ,  $\xi = (0, 0)$  ist dies in Abbildung 25 dargestellt. Man spricht daher auch von einem *Sattelpunkt*.

Wir betrachten den Spezialfall  $p = 2$  genauer. Für eine symmetrische  $(2 \times 2)$ -Matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

ist die Determinante  $\det(A)$  durch  $\det(A) = ac - b^2$  gegeben und es gilt:

(i)  $A$  ist positiv definit  $\Leftrightarrow a > 0$  und  $\det(A) > 0$ .

(ii)  $A$  ist negativ definit  $\Leftrightarrow a < 0$  und  $\det(A) > 0$ .

(iii)  $A$  ist indefinit  $\Leftrightarrow \det(A) < 0$ .

Damit erhalten wir das folgende Resultat direkt aus Satz 7.3.4.

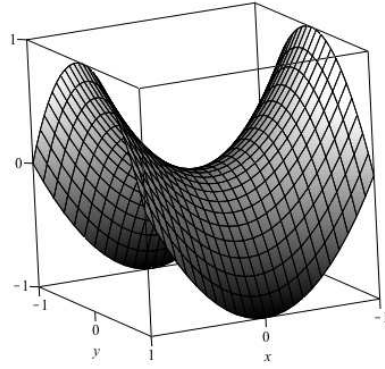


Abbildung 25: Ein Sattelpunkt.

**Satz 7.3.5** Sei  $U \subset \mathbb{R}^2$  offen,  $f \in C^2(U, \mathbb{R})$ ,  $\xi \in U$  und  $\text{grad}f(\xi) = 0$ . Sei

$$\Delta(\xi) := \partial_{11}f(\xi)\partial_{22}f(\xi) - (\partial_{12}f(\xi))^2.$$

Dann gilt:

- (i) Ist  $\Delta(\xi) > 0$  und  $\partial_{11}f(\xi) > 0$ , so hat  $f$  ein lokales Minimum in  $\xi$ .
- (ii) Ist  $\Delta(\xi) > 0$  und  $\partial_{11}f(\xi) < 0$ , so hat  $f$  ein lokales Maximum in  $\xi$ .
- (iii) Ist  $\Delta(\xi) < 0$ , so hat  $f$  kein lokales Extremum (sondern einen Sattelpunkt) in  $\xi$ .

*Beispiel 1.* Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $f(x, y) = xy - x^3 - y^2$ . Dann ist  $f \in C^2(\mathbb{R}^2, \mathbb{R})$  und  $\partial_1 f(x, y) = y - 3x^2$  und  $\partial_2 f(x, y) = x - 2y$ . Aus  $\text{grad}f(x, y) = 0$  folgt  $y = 3x^2$  und  $x = 2y$ , also  $y = 3(2y)^2 = 12y^2$  und damit  $y = 0$  oder  $y = \frac{1}{12}$ . Zusammen mit  $x = 2y$  folgt  $(x, y) = \xi_1 := (0, 0)$  oder  $(x, y) = \xi_2 := (\frac{1}{6}, \frac{1}{12})$ . Tatsächlich gilt auch  $\text{grad}f(\xi_j) = 0$  für  $j \in \{1, 2\}$ .

Wegen  $\partial_{11}f(x, y) = -6x$ ,  $\partial_{12}f(x, y) = \partial_{21}f(x, y) = 1$  und  $\partial_{22}f(x, y) = -2$  folgt

$$H_f(x, y) = \begin{pmatrix} -6x & 1 \\ 1 & -2 \end{pmatrix}$$

und damit  $\Delta(x, y) = \det H_f(x, y) = 12x - 1$ .

Somit ist  $\Delta(\xi_1) = \Delta(0, 0) = -1 < 0$  und folglich ist  $\xi_1$  Sattelpunkt. Weiter ist  $\Delta(\xi_2) = \Delta(\frac{1}{6}, \frac{1}{12}) = 1 > 0$  und wegen  $\partial_{11}f(\xi_2) = -1 < 0$  hat  $f$  in  $\xi_2$  ein lokales Maximum.

Der Graph von  $f$  – mit den  $\xi_1, \xi_2$  entsprechenden Punkten markiert – ist in Abbildung 26 dargestellt.

*Beispiel 2.* Sei  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ ,  $f(x, y, z) = 2x^4 + y^4 + z^4 - 4xyz - 2x^2$ . Dann ist  $f \in C^2(\mathbb{R}^3, \mathbb{R})$  und

$$\text{grad}f(x, y, z) = (8x^3 - 4yz - 4x, 4y^3 - 4xz, 4z^3 - 4xy).$$

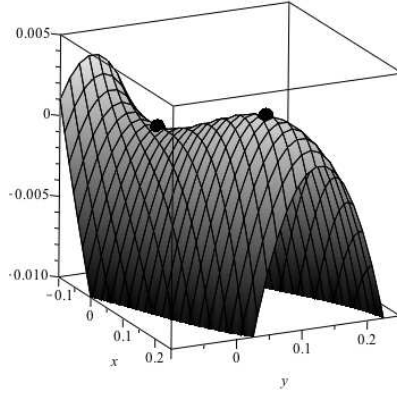


Abbildung 26: Der Graph der Funktion aus Beispiel 1.

Eine längere Rechnung zeigt, dass  $\text{grad}f(x, y, z) = 0$  für genau 7 Punkte  $(x, y, z)$  gilt, nämlich für  $(x, y, z) = \xi_j$  wobei  $j \in \{1, \dots, 7\}$ , mit  $\xi_1 := (0, 0, 0)$ ,  $\xi_2 := (\frac{1}{\sqrt{2}}, 0, 0)$ ,  $\xi_3 := (-\frac{1}{\sqrt{2}}, 0, 0)$ ,  $\xi_4 := (1, 1, 1)$ ,  $\xi_5 := (1, -1, -1)$ ,  $\xi_6 := (-1, 1, -1)$  und  $\xi_7 := (-1, -1, 1)$ . Es ist

$$H_f(x, y, z) = \begin{pmatrix} 24x^2 - 4 & -4z & -4y \\ -4z & 12y^2 & -4x \\ -4y & -4x & 12z^2 \end{pmatrix}$$

und damit

$$H_f(\xi_1) = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Diese Matrix ist negativ semidefinit, Satz 7.3.4 also nicht anwendbar. Es ist aber  $f(0, 0, z) > 0$  für alle  $z \in \mathbb{R}$  und  $f(x, 0, 0) < 0$  für  $0 < x < 1$ , und daher hat  $f$  in  $\xi_1$  kein lokales Extremum. Weiter gilt

$$H_f(\xi_{2,3}) = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 0 & \pm 2\sqrt{2} \\ 0 & \pm 2\sqrt{2} & 0 \end{pmatrix}.$$

Das charakteristische Polynom dieser Matrix ist durch  $(8 - \lambda)(\lambda^2 - 8)$  gegeben. Damit hat  $H_f(\xi_{2,3})$  sowohl positive wie negative Eigenwerte, ist also indefinit, und folglich ist in  $\xi_{2,3}$  kein lokales Extremum. Weiter ist

$$H_f(\xi_4) = \begin{pmatrix} 20 & -4 & -4 \\ -4 & 12 & -4 \\ -4 & -4 & 12 \end{pmatrix}.$$

Diese Matrix ist positiv definit, wie man etwa durch Betrachtung der Abschnittsdeterminanten (Satz 7.3.3) feststellt. Damit liegt in  $\xi_4$  ein lokales Minimum vor.

Analog – oder durch Symmetrieüberlegungen – zeigt man, dass  $f$  auch in  $\xi_5, \xi_6, \xi_7$  lokale Minima hat.

## 7.4 Höhere Ableitungen und Taylorformel

Satz 7.2.5 liefert eine Taylorentwicklung vom Grad 2. Für Taylorentwicklungen höheren Grades muss man die entsprechenden höheren Ableitungen bilden.

Sei  $f \in C^n(U, \mathbb{R})$  und  $\xi \in U$ . Wie in §7.2 sei  $v \in \mathbb{R}^p$  mit  $[\xi, \xi + v] \subset U$  und  $\gamma : [0, 1] \rightarrow U$ ,  $\gamma(t) = \xi + tv$ . Mit  $\gamma'(t) = v$  und  $\gamma^{(k)}(t) = 0$  für  $k \geq 2$  erhalten wir genau wie in §7.2 dann

$$\begin{aligned} (f \circ \gamma)^{(n)}(t) &= \sum_{i_1=1}^p \cdots \sum_{i_n=1}^p (\partial_{i_1, \dots, i_n} f \circ \gamma)(t) v_{i_1} \cdots v_{i_p} \\ &= \sum_{i_1=1}^p \cdots \sum_{i_n=1}^p \partial_{i_1, \dots, i_n} f(\xi + tv) v_{i_1} \cdots v_{i_p} \end{aligned}$$

Nach dem Satz von Schwarz gilt zum Beispiel  $\partial_{112}f = \partial_{121}f = \partial_{211}f$ . Um gleiche Ableitungen wie diese zusammenzufassen, sind folgende Notationen hilfreich.

Für  $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathbb{N}_0^p$  und  $x = (x_1, \dots, x_p) \in \mathbb{R}^p$  setzt man

$$|\alpha| = \sum_{j=1}^p \alpha_j, \quad \alpha! = \prod_{j=1}^p \alpha_j! \quad \text{und} \quad x^\alpha = \prod_{j=1}^p x_j^{\alpha_j}.$$

Man nennt  $\alpha$  hier auch *Multiindex*. Desweiteren sei für  $f \in C^{|\alpha|}(U, \mathbb{R})$

$$\partial^\alpha f = \partial_1^{\alpha_1} \cdots \partial_p^{\alpha_p} f.$$

Sei nun  $n \in \mathbb{N}$  und  $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathbb{N}_0^p$  mit  $|\alpha| = n$ . Wir betrachten die Menge  $I_\alpha$  der  $n$ -Tupel  $(i_1, \dots, i_n)$ , bei denen für alle  $k \in \{1, \dots, p\}$  die Menge  $\{m : i_m = k\}$  genau  $\alpha_k$  Elemente hat. Nach dem Satz von Schwarz gilt für  $(i_1, \dots, i_n) \in I_\alpha$  dann  $\partial_{i_1, \dots, i_n} f = \partial^\alpha f$ . Eine kombinatorische Überlegung zeigt, dass  $I_\alpha$  aus  $n!/|\alpha|!$  Elementen besteht. Damit erhält obige Darstellung von  $(f \circ \gamma)^{(n)}(t)$  folgende Form.

**Satz 7.4.1** Seien  $p, n \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $f \in C^n(U, \mathbb{R})$  und  $\xi \in U$ . Sei  $v \in \mathbb{R}^p$  mit  $[\xi, \xi + v] \subset U$  und  $\gamma : [0, 1] \rightarrow U$ ,  $\gamma(t) = \xi + tv$ . Dann gilt

$$(f \circ \gamma)^{(n)}(t) = \sum_{|\alpha|=n} \frac{n!}{\alpha!} \partial^\alpha f(\xi + tv) v^\alpha$$

Völlig analog zu Satz 7.2.5 erhält man nun folgendes Resultat.

**Satz 7.4.2** Seien  $p, n \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $f \in C^n(U, \mathbb{R})$  und  $\xi \in U$ . Dann gilt

$$f(\xi + h) = \sum_{k=0}^n \sum_{|\alpha|=k} \frac{1}{\alpha!} \partial^\alpha f(\xi) h^\alpha + o(\|h\|^n)$$

für  $h \rightarrow 0$ ,  $h \in \mathbb{R}^p$ .

Die auf der rechten Seite stehende Summe bezeichnet man wieder als *Taylorpolynom* vom Grad  $n$ .

Wir haben hier Satz 5.4.2 benutzt. Stattdessen kann man natürlich auch die Lagrange- oder Integralform des Restglieds benutzen. So erhält man zum Beispiel folgendes Resultat.

**Satz 7.4.3** Seien  $p, n \in \mathbb{N}$ ,  $U \subset \mathbb{R}^p$  offen,  $f \in C^{n+1}(U, \mathbb{R})$  und  $\xi \in U$ . Dann existiert  $\theta \in [0, 1]$  mit

$$f(\xi + h) = \sum_{k=0}^n \sum_{|\alpha|=k} \frac{1}{\alpha!} \partial^\alpha f(\xi) h^\alpha + \sum_{|\alpha|=n+1} \frac{1}{\alpha!} \partial^\alpha f(\xi + \theta h) h^\alpha$$

Insbesondere gilt für  $f \in C^{n+1}(U, \mathbb{R})$ , schärfer als in Satz 7.4.2, dass

$$f(\xi + h) = \sum_{k=0}^n \sum_{|\alpha|=k} \frac{1}{\alpha!} \partial^\alpha f(\xi) h^\alpha + O(\|h\|^{n+1})$$

für  $h \rightarrow 0$ .

## 7.5 Die Mittelwertungleichung

Wir betrachten Abbildungen  $f : I \rightarrow V$  wobei  $I$  Intervall und  $V$  Banachraum. Hier können wir die Ableitung wieder – wie in Analysis I – über den Differenzenquotienten bilden.

**Definition 7.5.1** Sei  $I$  Intervall,  $(V, \|\cdot\|)$  Banachraum,  $f : I \rightarrow V$  und  $\xi \in I$ . Dann heißt

$$f'(\xi) := \lim_{x \rightarrow \xi} \frac{f(x) - f(\xi)}{x - \xi} = \lim_{h \rightarrow 0} \frac{f(\xi + h) - f(\xi)}{h}$$

*Ableitung* von  $f$  an der Stelle  $\xi$ , falls der Grenzwert existiert.

Andererseits können wir ein offenes Intervall (oder das Innere eines beliebigen Intervalls) als offene Teilmenge des Banachraums  $\mathbb{R}$  auffassen und die Fréchet-Ableitung betrachten.

**Satz 7.5.1** Seien  $I, V, f$  wie in Definition 7.5.1 und  $\xi \in \text{int}(I)$ . Dann existiert  $f'(\xi)$  genau dann, wenn  $Df(\xi)$  existiert, und es gilt dann  $f'(\xi) = Df(\xi)(1)$ .

*Beweis.* Existiert  $f'(\xi)$ , so gilt

$$\lim_{h \rightarrow 0} \frac{f(\xi + h) - f(\xi) - hf'(\xi)}{h} = 0,$$

und damit folgt, dass  $Df(\xi)$  existiert und durch  $h \mapsto hf'(\xi)$  gegeben ist. Es folgt also  $f'(\xi) = Df(\xi)(1)$ . Analog folgt aus der Existenz von  $Df(\xi)$  die von  $f'(\xi)$ .  $\square$

Mit den Bezeichnungen von Satz 7.5.1 gilt  $\|f'(\xi)\| = \|Df(\xi)\|_{L(\mathbb{R}, V)}$ , falls  $f'(\xi)$  (und damit  $Df(\xi)$ ) existiert.

Der Mittelwertsatz der Differentialrechnung (Satz 4.2.2) gilt nur für reellwertige Funktionen. Für komplexwertige Funktionen hatten wir nur eine Ungleichung erhalten (Satz 4.2.5). Diese können wir auf die Banachraumsituation übertragen.

**Satz 7.5.2** Sei  $(V, \|\cdot\|)$  Banachraum und  $f : [a, b] \rightarrow V$  differenzierbar. Es sei  $M := \sup_{t \in [a, b]} \|f'(t)\| < \infty$ . Dann gilt  $\|f(b) - f(a)\| \leq M(b - a)$ .

*Beweis.* Ohne Einschränkung der Allgemeinheit sei  $[a, b] = [0, 1]$  und  $f(0) = 0$ . Für  $\varepsilon > 0$  betrachten wir

$$A := \{t \in [0, 1] : \|f(t)\| \leq (M + \varepsilon)t\}.$$

Offensichtlich gilt  $0 \in A$ . Wir werden zeigen, dass  $A = [0, 1]$  gilt. Insbesondere gilt dann  $1 \in A$ , und mit  $\varepsilon \rightarrow 0$  folgt dann die Behauptung.

Um zu zeigen, dass  $A = [0, 1]$  gilt, setzen wir  $c := \sup\{t \in [0, 1] : [0, t] \subset A\}$ . Aus der Stetigkeit von  $f$  folgt, dass  $c \in A$ . Wir nehmen an, dass  $c < 1$  gilt. Nach Definition der Differenzierbarkeit existiert  $\delta > 0$  mit

$$\|f(c+h) - f(c) - f'(c)h\| < \varepsilon h$$

falls  $c+h \in [0, 1]$  und  $|h| < \delta$ . Für  $0 < h < \min\{\delta, 1-c\}$  gilt damit

$$\begin{aligned} \|f(c+h)\| &\leq \|f(c) + f'(c)h\| + \varepsilon h \\ &\leq \|f(c)\| + \|f'(c)\|h + \varepsilon h \\ &\leq (M + \varepsilon)c + Mh + \varepsilon h \\ &= (M + \varepsilon)(c+h) \end{aligned}$$

und damit  $c+h \in A$ , im Widerspruch zur Wahl von  $c$ . Es folgt  $c = 1$ , also  $1 \in A$ .  $\square$

**Satz 7.5.3** *Seien  $(V, \|\cdot\|_V)$ ,  $(W, \|\cdot\|_W)$  Banachräume,  $U \subset V$  offen,  $a, b \in U$  mit  $[a, b] \in U$  und  $f : U \rightarrow W$  differenzierbar. Ist  $M := \sup_{x \in [a, b]} \|Df(x)\|_{L(V, W)} < \infty$ , so gilt  $\|f(b) - f(a)\|_W \leq M\|b - a\|_V$ .*

*Beweis.* Sei  $\gamma : [0, 1] \rightarrow [a, b]$ ,  $t \mapsto a + t(b - a)$ , und  $g := f \circ \gamma$ . Dann ist  $g$  differenzierbar in  $(0, 1)$  mit  $Dg(t) = Df(\gamma(t)) \circ D\gamma(t)$  und damit

$$g'(t) = Dg(t)(1) = Df(\gamma(t))(D\gamma(t)(1)) = Df(\gamma(t))(\gamma'(t)) = Df(\gamma(t))(b - a).$$

Es folgt

$$\|g'(t)\|_W \leq \|Df(\gamma(t))\|_{L(V, W)} \|b - a\|_V \leq M \|b - a\|_V.$$

Mit Satz 7.5.2 folgt

$$\|f(b) - f(a)\|_W = \|g(1) - g(0)\|_W \leq M \|b - a\|_V. \quad \square$$

Auch die Ungleichung in Satz 7.5.3 nennt man Mittelwertungleichung. Eine typische Anwendung ist der folgende Satz (vgl. Satz 4.3.1).

**Satz 7.5.4** *Seien  $(V, \|\cdot\|_V)$ ,  $(W, \|\cdot\|_W)$  Banachräume,  $U \subset V$  offen und zusammenhängend,  $f : U \rightarrow W$  differenzierbar und  $Df(x) = 0$  für alle  $x \in U$ . Dann ist  $f$  konstant, d. h., es existiert  $c \in W$  mit  $f(x) = c$  für alle  $x \in U$ .*

*Beweis.* Sei  $x \in U$ . Nach Satz 7.5.3 gilt dann  $f(y) = f(x)$  für alle  $y \in U$  mit  $[x, y] \subset U$ . Daraus folgt, dass  $f(z) = f(x)$  für alle  $z \in U$  gilt, die mit  $x$  durch einen in  $U$  verlaufenden Streckenzug verbunden werden können. Dies sind nach Satz 6.6.4 aber alle  $z \in U$ .  $\square$

## 7.6 Der Satz über implizite Funktionen

Wir betrachten die Frage, wann sich eine Gleichung der Form  $F(x, y) = 0$  (mit einer differenzierbaren, auf einer Teilmenge von  $\mathbb{R}^2$  definierten Funktion  $F$ ) nach  $y$  auflösen lässt, d.h., wann eine Funktion  $f$  existiert, so dass die Gleichung  $F(x, y) = 0$  äquivalent zur Gleichung  $y = f(x)$  ist. Es gilt dann  $F(x, f(x)) = 0$ . Wir nennen dann  $F(x, y) = 0$  eine *implizite* Form der *explizit* gegebenen Gleichung  $y = f(x)$ .

Etwas genauer werden wir folgende Fragestellung untersuchen: Gegeben sei  $(\xi, \eta) \in \mathbb{R}^2$  mit  $F(\xi, \eta) = 0$ . Existieren dann Umgebungen  $U$  und  $V$  von  $\xi$  und  $\eta$  und eine (differenzierbare) Funktion  $f : U \rightarrow V$  mit  $f(\xi) = \eta$  und  $F(x, f(x)) = 0$  für alle  $x \in U$ ? Inwieweit ist  $f$  eindeutig bestimmt?

Für eine konkrete Funktion  $F$  könnte  $\{(x, y) \in \mathbb{R}^2 : F(x, y) = 0\}$  etwa wie in Abbildung 27 aussehen. Hier scheint es so zu sein, dass für  $(\xi, \eta) = P$  Umgebungen  $U, V$  und eine Funktion  $f$  mit den verlangten Eigenschaften existieren, während das für  $(\xi, \eta) = Q$  nicht der Fall ist.

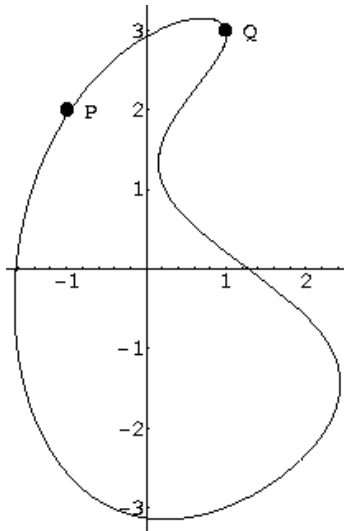


Abbildung 27: Bild von  $\{(x, y) \in \mathbb{R}^2 : F(x, y) = 0\}$  für eine Funktion  $F$ .

Man erkennt nun, dass für  $(\xi, \eta) = Q$  die durch  $y \mapsto F(\xi, y)$  definierte Funktion in einer Umgebung von  $\eta$  immer  $\leq 0$  oder immer  $\geq 0$  ist, also in  $\eta$  ein lokales Extremum hat. Es folgt, dass  $\partial_2 F(Q) = 0$  gilt.

Wir werden zeigen, dass (für stetig differenzierbares  $F$ ) Umgebungen  $U, V$  und eine Funktion  $f$  mit den verlangten Eigenschaften existieren, wenn  $\partial_2 F(\xi, \eta) \neq 0$  gilt.

Zur Umsetzung dieser Beweisidee benötigen wir den folgenden Satz.

**Satz 7.6.1 (Banachscher Fixpunktsatz)** Sei  $(M, d)$  vollständiger metrischer Raum und  $T : M \rightarrow M$ . Es existiere  $\alpha \in [0, 1)$  mit  $d(T(x), T(y)) \leq \alpha d(x, y)$  für alle  $x, y \in M$ . Dann hat  $T$  genau einen Fixpunkt, d. h., es existiert genau ein  $\xi \in M$  mit  $T(\xi) = \xi$ .

Darüberhinaus gilt  $\lim_{n \rightarrow \infty} T^n(x) = \xi$  für alle  $x \in M$ . (Dabei ist  $T^n$  die  $n$ -te Iterierte von  $T$ . Diese sind rekursiv durch  $T^1 := T$  und  $T^{n+1} := T \circ T^n$  definiert.)

*Beweis.* Übung.

Der Operator  $T$ , auf den wir den Banachschen Fixpunktsatz anwenden, ist im Wesentlichen der des Newtonverfahrens. Dieses Verfahren zur Bestimmung der Nullstellen einer differenzierbaren Funktion  $\alpha$  besteht ja aus der Iteration der Funktion

$$y \mapsto y_1 := y - \frac{\alpha(y)}{\alpha'(y)}.$$

Die Newtonsche Iterationsfunktion  $y \rightarrow y_1$  erhält man durch Auflösen der Gleichung  $\alpha(y) + \alpha'(y)(y_1 - y) = 0$ . Wir werden die oben skizzierten Aussagen über implizite Funktionen in allgemeinerer Form beweisen. Dazu skizzieren wir kurz das Newtonverfahren in höheren Dimensionen. Für eine differenzierbare Funktion  $\alpha : H \rightarrow \mathbb{R}^q$ , wobei  $H \subset \mathbb{R}^q$  offen ist, löst man analog  $\alpha(y) + D\alpha(y)(y_1 - y) = 0$  und erhält

$$y \mapsto y_1 := y - (D\alpha(y))^{-1}(\alpha(y)).$$

Dabei wird natürlich vorausgesetzt, dass die lineare Abbildung  $D\alpha(y) : \mathbb{R}^q \rightarrow \mathbb{R}^q$  invertierbar ist. (Dies entspricht der Annahme  $\alpha'(y) \neq 0$  im eindimensionalen Fall.)

Seien nun  $G \subset \mathbb{R}^p$  offen,  $H \subset \mathbb{R}^q$  offen,  $\xi \in G$ ,  $\eta \in H$  und  $F \in C^1(G \times H, \mathbb{R}^q)$  mit  $F(\xi, \eta) = 0$ . Wir fragen nach der Existenz von Umgebungen  $U$  und  $V$  von  $\xi$  und  $\eta$  und einer differenzierbaren Funktion  $f : U \rightarrow V$  mit  $f(\xi) = \eta$  und  $F(x, f(x)) = 0$  für alle  $x \in U$ .

Wir suchen also – für festes  $x$  – eine Nullstelle der durch  $y \mapsto \alpha(y) := F(x, y)$  gegebenen Funktion. Für unsere Zwecke ist es etwas einfacher, statt des Newtonverfahrens die Funktion

$$y \mapsto y - (D\alpha(\eta))^{-1}(\alpha(y))$$

zu betrachten. Wir schreiben  $F_2^x$  statt  $\alpha$ , also  $F_2^x : H \rightarrow \mathbb{R}^q$ ,  $y \mapsto F(x, y)$  und setzen

$$D_2F(x, y) := DF_2^x(y)$$

und

$$\partial_2^*F(x, y) := J_{F_2^x}(y) = \begin{pmatrix} \partial_{p+1}F_1(x, y) & \dots & \partial_{p+q}F_1(x, y) \\ \partial_{p+1}F_2(x, y) & \dots & \partial_{p+q}F_2(x, y) \\ \vdots & & \vdots \\ \partial_{p+1}F_q(x, y) & \dots & \partial_{p+q}F_q(x, y) \end{pmatrix}.$$

Man schreibt auch  $\frac{\partial F}{\partial y}$  statt  $\partial_2^*F$ . Analog definiert man  $D_1F$  und  $\partial_1^*F$ , indem man  $F_1^y : G \rightarrow \mathbb{R}^p$ ,  $x \mapsto F(x, y)$ , betrachtet.

Wir benötigen jetzt, dass  $D\alpha(\eta) = D_2F(\xi, \eta)$  invertierbar ist. Dies ist nach Linearer Algebra äquivalent zu  $\det(\partial_2^*F(\xi, \eta)) \neq 0$ . Im eingangs diskutierten Fall  $p = q = 1$  erhält man die dort genannte Bedingung  $\partial_2F(\xi, \eta) \neq 0$

**Satz 7.6.2 (Satz über implizite Funktionen)** Sei  $G \subset \mathbb{R}^p$  offen,  $H \subset \mathbb{R}^q$  offen,  $\xi \in G$ ,  $\eta \in H$  und  $F \in C^1(G \times H, \mathbb{R}^q)$  mit  $F(\xi, \eta) = 0$ . Es sei  $D_2F(\xi, \eta)$  invertierbar. Dann existieren offene Umgebungen  $U \subset G$  und  $V \subset H$  von  $\xi$  und  $\eta$  und eine Funktion  $f \in C^1(U, V)$  mit  $f(\xi) = \eta$  und  $F(x, f(x)) = 0$  für  $x \in U$ , und umgekehrt folgt für  $(x, y) \in U \times V$  mit  $F(x, y) = 0$ , dass  $y = f(x)$  gilt.

Desweiteren gilt  $Df(x) = -D_2F(x, f(x))^{-1} \circ D_1F(x, f(x))$  für  $x \in U$ , insbesondere also  $Df(\xi) = -D_2F(\xi, \eta)^{-1} \circ D_1F(\xi, \eta)$ .



*Beweis.* Wir werden nur die Existenz von  $f \in C(U, V)$  mit den verlangten Eigenschaften beweisen, und den Beweis der stetigen Differenzierbarkeit von  $f$  nur skizzieren. Ist aber die Differenzierbarkeit von  $f$  bekannt, so folgt die Formel für  $Df$  aus der Kettenregel.

Zur Abkürzung sei  $A := D_2F(\xi, \eta)$ . Es ist  $A \in L(\mathbb{R}^q, \mathbb{R}^q)$  und nach Voraussetzung existiert  $A^{-1}$ . Seien  $\varepsilon, \delta > 0$ ,  $U := U(\xi, \delta)$  und  $V := U(\eta, \varepsilon)$ . Dabei können wir  $\varepsilon, \delta$  so klein wählen, dass die folgenden drei Bedingungen gelten:

(i)  $\bar{U} \subset G$  und  $\bar{V} \subset H$ .

(ii) Für  $x \in U$  und  $y \in V$  gilt

$$\|id_{\mathbb{R}^q} - A^{-1}(D_2F(x, y))\|_{L(\mathbb{R}^q, \mathbb{R}^q)} \leq \frac{1}{2}.$$

(iii) Für  $x \in U$  gilt

$$\|A^{-1}(F(x, \eta))\| \leq \frac{\varepsilon}{4}.$$

Dabei können (ii) und (iii) erreicht werden, weil die linken Seiten stetig in  $G \times H$  bzw.  $G$  sind und für  $(x, y) = (\xi, \eta)$  bzw.  $x = \xi$  verschwinden. Wir setzen nun  $X := C(\bar{U}, \mathbb{R}^q) = \{g : \bar{U} \rightarrow \mathbb{R}^q : g \text{ stetig}\}$  und  $\|g\|_X := \max_{x \in \bar{U}} \|g(x)\|$  für  $g \in X$ . Dann ist  $(X, \|\cdot\|_X)$  Banachraum; vgl. Beispiel 1 zu Satz 6.3.4. Weiter setzen wir

$$M := \left\{ g \in X : g(\xi) = \eta \text{ und } \|g(x) - \eta\|_X \leq \frac{\varepsilon}{2} \right\} \text{ für alle } x \in \bar{U} :$$

Dann ist  $M \neq \emptyset$ , da die konstante Funktion  $g = \eta$  in  $M$  ist. Außerdem ist  $M$  abgeschlossen. Nach Satz 6.3.4 ist  $M$  also vollständig.

Wir betrachten jetzt die Funktion  $T : M \rightarrow X$ ,

$$T(g)(x) := g(x) - A^{-1}(F(x, g(x))).$$

Wir werden zeigen:

(iv) Für  $g, h \in M$  gilt  $\|T(g) - T(h)\|_X \leq \frac{1}{2}\|g - h\|_X$ .

(v)  $T(M) \subset M$ .

Der Banachsche Fixpunktsatz liefert nun  $f \in M \subset C\left(\bar{U}, \overline{U(\eta, \frac{\varepsilon}{2})}\right)$  und damit  $f \in C(U, V)$  mit  $T(f) = f$ . Daraus folgt  $A^{-1}(F(x, f(x))) = 0$ , also  $F(x, f(x)) = 0$  für  $x \in U$ .

Zu (iv): Für  $x \in \bar{U}$  betrachten wir  $\phi^x : V \rightarrow \mathbb{R}^q$ ,  $y \mapsto y - A^{-1}(F(x, y))$ , also  $\phi^x = id_{\mathbb{R}^q}|_V - A^{-1} \circ F_2^x$ . Es ist dann  $\phi^x(g(x)) = T(g)(x)$  für  $g \in M$ .

Nach Kettenregel ist  $\phi^x$  differenzierbar mit

$$D\phi^x(y) = id_{\mathbb{R}^q} - A^{-1}(D_2F(x, y)).$$

Nach (ii) folgt  $\|D\phi^x(y)\|_{L(\mathbb{R}^q, \mathbb{R}^q)} \leq \frac{1}{2}$  für  $y \in V$ . Nach Mittelwertungleichung (Satz 7.5.3) folgt

$$\|\phi^x(y_1) - \phi^x(y_2)\| \leq \sup_{y \in [y_1, y_2]} \|D\phi^x(y)\|_{L(\mathbb{R}^q, \mathbb{R}^q)} \|y_1 - y_2\|$$

für  $y_1, y_2 \in V$ , und damit

$$\|\phi^x(y_1) - \phi^x(y_2)\| \leq \frac{1}{2}\|y_1 - y_2\|,$$

und das gilt aus Stetigkeitsgründen sogar für  $y_1, y_2 \in \bar{V}$ . Es folgt

$$\|T(g)(x) - T(h)(x)\| = \|\phi^x(g(x)) - \phi^x(h(x))\| \leq \frac{1}{2}\|g(x) - h(x)\|$$

für  $g, h \in M$  und  $x \in \bar{U}$  und damit (iv).

Zu (v): Für  $g \in M$  und  $x \in \bar{U}$  gilt mit (iv) und (iii)

$$\begin{aligned} \|T(g)(x) - \eta\| &\leq \|T(g)(x) - T(\eta)(x)\| + \|T(\eta)(x) - \eta\| \\ &\leq \frac{1}{2}\|g(x) - \eta\| + \|A^{-1}(F(x, \eta))\| \\ &\leq \frac{1}{2} \cdot \frac{\varepsilon}{2} + \frac{\varepsilon}{4}, \end{aligned}$$

und daraus folgt (v).

Wir haben also die Existenz einer Funktion  $f \in C(U, V)$  mit  $F(x, f(x)) = 0$  für  $x \in U$  nachgewiesen. Zum Nachweis der Eindeutigkeit seien nun  $x \in U$  und  $y \in V$  mit  $F(x, y) = 0$ . Dann ist  $\phi^x(y) = y$  und damit

$$\|f(x) - y\| = \|\phi^x(f(x)) - \phi^x(y)\| \leq \frac{1}{2}\|f(x) - y\|.$$

Hieraus folgt  $\|f(x) - y\| = 0$ , also  $f(x) = y$ .

Um die Differenzierbarkeit von  $f$  in  $\xi$  zu zeigen, betrachtet man die Entwicklung

$$\begin{aligned} 0 &= 0 - 0 \\ &= F(x, f(x)) - F(\xi, \eta) \\ &= D_1F(\xi, \eta)(x - \xi) + D_2F(\xi, \eta)(f(x) - \eta) + \dots, \end{aligned}$$

die wegen  $f(\xi) = \eta$

$$f(x) - f(\xi) = -(D_2F(\xi, \eta))^{-1} \circ D_1F(\xi, \eta)(x - \xi) + \dots$$

liefert. Die Differenzierbarkeit von  $f$  in  $\xi$  erhält man nun durch eine Analyse der hier nur durch  $\dots$  angedeuteten Restglieder. Die Differenzierbarkeit von  $f$  in einer Umgebung von  $\xi$  erhält man analog. (Dabei muss das im ersten Teil des Beweises gewählte  $\delta$  eventuell noch verkleinert werden.)  $\square$

*Beispiel 1.* Sei  $f : (0, \infty)^2 \rightarrow \mathbb{R}$ ,  $F(x, y) = x^y - y^x$ . Es gilt  $F(2, 4) = 0$  und  $\partial_2F(x, y) = x^y \ln x - xy^{x-1}$ , also  $\partial_2F(2, 4) = 2^4 \ln 2 - 2 \cdot 4 = 8(2 \ln 2 - 1) > 0$ . Nach dem Satz über implizite Funktionen existieren  $\varepsilon, \delta > 0$  und stetig differenzierbares  $f : (2 - \delta, 2 + \delta) \rightarrow (4 - \varepsilon, 4 + \varepsilon)$  mit  $F(x, f(x)) = 0$ , d. h.,  $x^{f(x)} = f(x)^x$ .

Es ist  $\partial_1F(x, y) = yx^{y-1} - y^x \ln y$ , also  $\partial_1F(2, 4) = 4 \cdot 2^3 - 4^2 \ln 4 = 32(1 - \ln 2)$ . Es folgt

$$f'(2) = -\frac{\partial_1F(2, 4)}{\partial_2F(2, 4)} = -4 \frac{1 - \ln 2}{2 \ln 2 - 1} = -3,177\dots$$

Eine Lösung der Gleichung  $F(x, y) = 0$  ist natürlich auch durch  $y = x$  gegeben. Wegen  $\partial_2 F(x, x) = x^x(\ln x - 1) \neq 0$  für  $x \neq e$  folgt aus dem Satz über implizite Funktionen, dass jeder Punkt  $x \neq e$  eine Umgebung  $U$  besitzt, so dass  $F(x, y) \neq 0$  für  $x, y \in U$  mit  $x \neq y$ . Dies ist für  $x = e$  nicht der Fall.

Die durch  $F(x, y) = 0$  gegebene Teilmenge des  $\mathbb{R}^2$  ist in Abbildung 28 dargestellt.

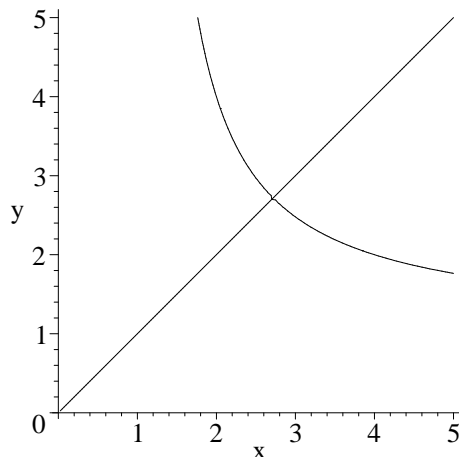


Abbildung 28: Die Menge der  $(x, y) \in \mathbb{R}^2$  mit  $x^y = y^x$ .

*Beispiel 2.* Sei  $G = \mathbb{R}$ ,  $H = (0, \infty)^2 \subset \mathbb{R}^2$  und  $F = (F_1, F_2)^T : G \times H \rightarrow \mathbb{R}^2$ ,

$$F(x, y_1, y_2) = \begin{pmatrix} F_1(x, y_1, y_2) \\ F_2(x, y_1, y_2) \end{pmatrix} = \begin{pmatrix} x \ln y_1 + y_1 e^{xy_2} - 1 \\ x e^{y_1} + \sqrt{y_1 y_2} - y_2 e^x \end{pmatrix}.$$

Es ist dann  $F \in C^1(G \times H, \mathbb{R}^2)$  und  $F(0, 1, 1) = 0$ . Es gilt

$$\begin{aligned} \partial_2^* F(x, y_1, y_2) &= \begin{pmatrix} \partial_2 F_1(x, y_1, y_2) & \partial_3 F_1(x, y_1, y_2) \\ \partial_2 F_2(x, y_1, y_2) & \partial_3 F_2(x, y_1, y_2) \end{pmatrix} \\ &= \begin{pmatrix} \frac{x}{y_1} + e^{xy_2} & y_1 x e^{xy_2} \\ x e^{y_1} + \frac{1}{2} \sqrt{\frac{y_2}{y_1}} & \frac{1}{2} \sqrt{\frac{y_1}{y_2}} - e^x \end{pmatrix}, \end{aligned}$$

also

$$A := \partial_2^* F(0, 1, 1) = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Wegen  $\det A = -\frac{1}{2} \neq 0$  ist  $A$  invertierbar. Es gilt

$$A^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix}.$$

Nach dem Satz über implizite Funktionen existiert also eine Umgebung  $U \subset G$  von 0 und eine Umgebung  $V \subset H$  von  $(1, 1)$  sowie eine Funktion  $f = (f_1, f_2)^T : U \rightarrow V$  mit  $F(x, f_1(x), f_2(x)) = 0$  für alle  $x \in U$ . Wegen

$$\partial_1^* F(x, y_1, y_2) = \begin{pmatrix} \partial_1 F_1(x, y_1, y_2) \\ \partial_1 F_2(x, y_1, y_2) \end{pmatrix} = \begin{pmatrix} \ln y_1 + y_1 y_2 e^{xy_2} \\ e^{y_1} - y_2 e^x \end{pmatrix}$$

ist

$$\partial_1^* F(0, 1, 1) = \begin{pmatrix} 1 \\ e - 1 \end{pmatrix}$$

und damit

$$f'(0) = -\partial_2^* F(0, 1, 1)^{-1} \partial_1^* F(0, 1, 1) = - \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ e - 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2e - 3 \end{pmatrix}.$$

**Satz 7.6.3 (Umkehrsatz)** Sei  $G \subset \mathbb{R}^p$  offen,  $f \in C^1(G, \mathbb{R}^p)$ ,  $\alpha \in G$  und  $Df(\alpha)$  invertierbar. Dann existieren offene Umgebungen  $U \subset G$  von  $\alpha$  und  $V \subset \mathbb{R}^p$  von  $\beta := f(\alpha)$ , so dass  $f : U \rightarrow V$  bijektiv ist (d. h.,  $f|_U$  ist injektiv und  $V := f(U)$  ist eine offene Umgebung von  $\beta$ ).

Die Umkehrfunktion  $g : V \rightarrow U$  von  $f : U \rightarrow V$  ist stetig differenzierbar und es gilt  $Dg(x) = Df(g(x))^{-1}$  für  $x \in V$ .

*Beweis.* Wir wenden den Satz über implizite Funktionen auf die Funktion

$$F : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^p, \quad F(x, y) = f(y) - x,$$

an. Es ist dann  $F(\beta, \alpha) = 0$  und  $D_2 F(x, y) = Df(y)$ , also  $D_2 F(\beta, \alpha) = Df(\alpha)$ .

Nach dem Satz über implizite Funktionen existieren offene Umgebungen  $W$  von  $\alpha$  und  $V$  von  $\beta$  sowie eine Funktion  $g \in C^1(V, W)$  mit  $0 = F(x, g(x)) = f(g(x)) - x$  für  $x \in V$ , und  $F(x, y) \neq 0$  falls  $x \in V$ ,  $y \in W$ ,  $y \neq g(x)$ .

Wegen  $f(g(x)) - x = 0$  für  $x \in V$  folgt, dass  $g$  injektiv ist. Mit  $U := g(V)$  folgt also, dass  $f : U \rightarrow V$  bijektiv mit Umkehrfunktion  $g : V \rightarrow U$  ist.

Noch zu zeigen ist, dass  $U$  offen ist. Dazu zeigen wir zunächst, dass  $U = f^{-1}(V) \cap W$  gilt. Sei dazu  $y \in U$ . Dann existiert  $x \in V$  mit  $y = g(x)$ . Es folgt  $f(y) = f(g(x)) = x \in V$ , also  $y \in f^{-1}(V)$ , und wegen  $U \subset W$  also  $U \subset f^{-1}(V) \cap W$ . Ist umgekehrt  $y \in f^{-1}(V) \cap W$ , so folgt  $x := f(y) \in V$ . Wegen  $F(x, y) = f(y) - x = 0$  folgt  $y = g(x) \in U$ . Insgesamt erhalten wir  $U = f^{-1}(V) \cap W$ . Da  $f$  stetig und  $V$  offen ist, ist  $f^{-1}(V)$  offen und damit auch  $f^{-1}(V) \cap W$  offen.

Die Formel  $Dg(x) = Df(g(x))^{-1}$  folgt aus dem Satz über implizite Funktionen – oder direkt aus der Kettenregel.  $\square$

Der Satz ist nur lokal, d. h., auch wenn  $Df(x)$  für alle  $x \in G$  invertierbar ist, muss  $f : G \rightarrow f(G)$  nicht bijektiv sein. Man betrachte etwa die Funktion  $f : \mathbb{C} \rightarrow \mathbb{C} \setminus \{0\}$ ,  $f(z) = e^z$ . In reeller Schreibweise haben wir

$$f(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}$$

und damit

$$\det J_f(x, y) = \det \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} = e^{2x} \neq 0,$$

womit  $Df(x)$  für alle  $x \in \mathbb{R}^2$  invertierbar ist. Andererseits gilt aber  $f(0, 2\pi) = f(0, 0)$ , womit  $f$  nicht injektiv ist.

**Definition 7.6.1** Seien  $(M, d_M)$ ,  $(N, d_N)$  metrische Räume. Eine Abbildung  $f : M \rightarrow N$  heißt *offen*, wenn  $f(U)$  für jedes offene  $U \subset M$  offen ist.

Man vergleiche diese Definition mit der durch Satz 6.4.3 gegebenen Charakterisierung der Stetigkeit. (“Urbilder offener Mengen sind offen.”)

Eine Folgerung aus dem Umkehrsatz ist das folgende Resultat.

**Satz 7.6.4** Sei  $G \subset \mathbb{R}^p$  offen,  $f \in C^1(G, \mathbb{R}^p)$  und  $Df(x)$  invertierbar für alle  $x \in G$ . Dann ist  $f$  offen.

## 7.7 Extrema unter Nebenbedingungen

In §4.3.3 haben wir für  $U \subset \mathbb{R}^p$  Funktionen  $f : U \rightarrow \mathbb{R}$  auf lokale Extrema untersucht. Jetzt betrachten wir den Fall, dass zusätzlich noch eine Funktion  $g : U \rightarrow \mathbb{R}^q$  gegeben ist, und wollen lokale Extrema von  $f$  unter der Nebenbedingung  $g = 0$  bestimmen, d. h., wir untersuchen  $f|_{\{x \in U : g(x) = 0\}} = f|_{g^{-1}(0)}$  auf lokale Extrema.

*Beispiel.* Man maximiere (oder minimiere)  $f(x, y) = x + y$  unter der Nebenbedingung  $g(x, y) = x^2 + y^2 - 1 = 0$ .

Eine Methode besteht darin, die Gleichung  $g(x, y) = 0$  nach  $y$  aufzulösen und das Resultat  $y = h(x)$  in  $f$  einzusetzen, also das Extremum von  $\tilde{f}(x) = f(x, h(x))$  zu bestimmen.

Im obigen *Beispiel* ist  $h(x) = \pm\sqrt{1-x^2}$  und  $\tilde{f}(x) = x \pm \sqrt{1-x^2}$ ,  $x \in [-1, 1]$ .

Der Nachteil dieser Methode ist, dass eine Auflösung der Gleichung  $g(x, y)$  nach  $y$  (oder  $x$ ) nicht immer möglich ist, oder die Auflösung oft kompliziert ist. Stattdessen werden wir jetzt den Satz über implizite Funktionen benutzen.

**Satz 7.7.1 (Lagrangesche Multiplikatorenregel)** Sei  $q < p$ ,  $G \subset \mathbb{R}^p$  offen,  $f \in C^1(G, \mathbb{R})$ ,  $g \in C^1(G, \mathbb{R}^q)$ ,  $\xi \in G$  und  $g(\xi) = 0$ . Die Funktion  $f$  habe in  $\xi$  ein lokales Extremum unter der Nebenbedingung  $g = 0$ . Hat dann  $Dg(\xi)$  Rang  $q$ , so existiert  $\Lambda \in L(\mathbb{R}^q, \mathbb{R})$  mit  $\Lambda \circ Dg(\xi) = Df(\xi)$ .

Ist  $\lambda = (\lambda_1, \dots, \lambda_q)$  die  $\Lambda$  darstellende  $(1 \times q)$ -Matrix bezüglich der Standardbasis, so erhält die Gleichung  $\Lambda \circ Dg(\xi) = Df(\xi)$  die Form

$$\text{grad} f(\xi) = \lambda \cdot J_g(\xi),$$

mit  $g = (g_1, \dots, g_q)^T$  also

$$\sum_{j=1}^q \lambda_j \partial_k g_j(\xi) = \partial_k f(\xi)$$

für  $1 \leq k \leq p$ . Die  $\lambda_j$  nennt man *Lagrangesche Multiplikatoren*.

*Beweis von Satz 7.7.1.* Nach der Voraussetzung über den Rang von  $Dg(\xi)$  existieren  $q$  linear unabhängige Spalten in  $J_g(\xi)$ . Ohne Beschränkung der Allgemeinheit seien die letzten  $q$  Spalten linear unabhängig. Wir setzen nun  $r := p - q$ . Dann ist  $\mathbb{R}^p = \mathbb{R}^r \times \mathbb{R}^q$  und  $f$  und  $g$  können als Abbildungen von  $\mathbb{R}^r \times \mathbb{R}^q$  nach  $\mathbb{R}$  aufgefasst werden. Dementsprechend schreiben wir  $x \in \mathbb{R}^p$  in der Form  $x = (z, y)$  mit  $z \in \mathbb{R}^r$  und  $y \in \mathbb{R}^q$  und damit  $f(z, y)$  bzw.  $g(z, y)$  statt  $f(x)$  bzw.  $g(x)$ . Wir definieren  $D_1 f, D_2 f, D_1 g$  und  $D_2 g$  wie zuvor.

Nach obiger Annahme über die Spalten von  $J_g(\xi)$  ist dann  $D_2g(\xi)$  invertierbar. Setzen wir noch  $\xi = (\zeta, \eta)$ , so folgt mit dem Satz über implizite Funktionen, dass Umgebungen  $U \subset \mathbb{R}^r$  von  $\zeta$  und  $V \subset \mathbb{R}^q$  von  $\eta$  und eine Funktion  $h \in C^1(U, V)$  existieren, so dass  $h(\zeta) = \eta$  und  $g(z, h(z)) = 0$  für  $z \in U$ . Damit ist in  $\zeta$  ein lokales Extremum der durch  $z \mapsto f(z, h(z))$  definierten Funktion  $\varphi : U \rightarrow \mathbb{R}$ . Es folgt  $D\varphi(\zeta) = 0$ .

Nun ist (nach Kettenregel)

$$D\varphi(\zeta) = D_1f(\zeta, \eta) + D_2f(\zeta, \eta) \circ Dh(\zeta)$$

und (nach Satz über implizite Funktionen)

$$Dh(\zeta) = -D_2g(\zeta, \eta)^{-1} \circ D_1g(\zeta, \eta).$$

Es folgt

$$0 = D\varphi(\zeta) = D_1f(\xi) - D_2f(\xi) \circ D_2g(\xi)^{-1} \circ D_1g(\xi).$$

Mit  $\Lambda := D_2f(\xi) \circ D_2g(\xi)^{-1}$  folgt also  $D_1f(\xi) = \Lambda \circ D_1g(\xi)$ . Trivialerweise ist auch  $D_2f(\xi) = \Lambda \circ D_2g(\xi)$ . Es folgt  $Df(\xi) = \Lambda \circ Dg(\xi)$ .  $\square$

*Beispiel 1.* Es seien  $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $f(x, y) = x + y$ ,  $g(x, y) = x^2 + y^2 - 1 = 0$ . Man bestimme das Maximum und Minimum von  $f$  unter der Nebenbedingung  $g = 0$ . (Die Existenz von Maximum und Minimum folgt, da  $f$  stetig und  $g^{-1}(0) = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$  kompakt.

Es sind  $f, g \in C^1(\mathbb{R}^2, \mathbb{R})$  mit  $\text{grad}f(x, y) = (1, 1)$  und  $\text{grad}g(x, y) = (2x, 2y)$ . Aus  $\text{grad}g(x, y) = (0, 0)$  folgt  $(x, y) = (0, 0)$ . Da aber  $g(0, 0) = -1 \neq 0$ , ist die Voraussetzung über den Rang von  $Dg$  in allen Punkte von  $g^{-1}(0)$  erfüllt.

Wir betrachten die Gleichung

$$\text{grad}f(x, y) = \lambda \text{grad}g(x, y),$$

also  $(1, 1) = \lambda(2x, 2y)$ . Es folgt  $\lambda \neq 0$  und damit  $x = y$ . Zusammen mit  $g(x, y) = x^2 + y^2 - 1 = 0$  erhalten wir  $(x, y) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) =: \xi_1$  oder  $(x, y) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) =: \xi_2$ . Es folgt, dass  $f$  sein Maximum  $\sqrt{2}$  in  $\xi_1$  und sein Minimum  $-\sqrt{2}$  in  $\xi_2$  annimmt.

*Beispiel 2.* Man maximiere das Volumen einer Konservendose bei konstanter Oberfläche, d.h., man maximiere die durch  $V(r, h) = \pi r^2 h$  gegebene Funktion  $V : (0, \infty)^2 \rightarrow \mathbb{R}$  unter der Nebenbedingung  $g(r, h) := 2\pi r^2 + 2\pi r h - F = 0$ , wobei  $F$  eine gegebene positive Konstante ist, und  $g : (0, \infty)^2 \rightarrow \mathbb{R}$ .

Zunächst ist wegen  $2\pi r^2 < F$  für  $g(r, h) = 0$  die Menge aller  $r > 0$ , für die ein  $h > 0$  mit  $g(r, h) = 0$  existiert, beschränkt. Damit gilt  $V(r, h) \rightarrow 0$  für  $h \rightarrow 0$  (und  $g(r, h) = 0$ ). Außerdem gilt für  $g(r, h) = 0$  auch  $2\pi r h < F$  und damit  $r < F/2\pi h$ , also  $V(r, h) < \pi h F^2 / (4\pi^2 h^2) = F^2 / 4\pi h$ . Es folgt  $V(r, h) \rightarrow 0$  für  $h \rightarrow \infty$  (und  $g(r, h) = 0$ ). Zusammen mit der Stetigkeit von  $V$  folgt aus diesen Überlegungen, dass das gesuchte Maximum existiert.

Es gilt nun  $V, g \in C^1(\mathbb{R}^2, \mathbb{R})$  mit  $\text{grad}V(r, h) = (2\pi r h, \pi r^2)$  und  $\text{grad}g(r, h) = (4\pi r + 2\pi h, 2\pi r)$ . Aus  $\text{grad}g(r, h) = (0, 0)$  folgt  $(r, h) = (0, 0)$ . Da  $(0, 0)$  nicht im Definitionsbereich von  $g$  liegt, ist die Voraussetzung über den Rang von  $Dg$  wieder in allen Punkte von  $g^{-1}(0)$  erfüllt. (Auch wenn man  $g$  und  $F$  im Definitionsbereich

$\mathbb{R} \times \mathbb{R}$  betrachten würde, wäre die Voraussetzung über den Rang von  $Dg$  wegen  $g(0,0) = -F \neq 0$  in allen Punkten von  $g^{-1}(0)$  erfüllt.) Die Gleichung

$$\text{grad } V(r, h) = \lambda \text{grad } g(r, h)$$

hat jetzt die Form

$$\begin{aligned} 2\pi r h &= \lambda(4\pi r + 2\pi h), \\ \pi r^2 &= \lambda 2\pi r. \end{aligned}$$

Aus der zweiten Gleichung folgt  $\lambda = \frac{r}{2}$  und Einsetzen in die erste liefert  $2\pi r h = \frac{r}{2}(4\pi r + 2\pi h) = 2\pi r^2 + \pi r h$  und damit  $r h = 2r^2$ , also  $h = 2r$ .

*Bemerkung.* In Beispiel 2 kann man natürlich auch die Nebenbedingung  $g(r, h) = 0$  nach  $h$  auflösen und erhält

$$h = \alpha(r) := \frac{F - 2\pi r^2}{2\pi r}.$$

Man hat dann

$$\beta(r) := V(r, \alpha(r)) = \pi r^2 \frac{F - 2\pi r^2}{2\pi r} = \frac{F}{2} r - \pi r^3$$

zu maximieren. Aus  $\beta'(r) = \frac{F}{2} - 3\pi r^2$  folgt, dass  $\beta$  sein Maximum bei  $r = r_0 := \sqrt{\frac{F}{6\pi}}$  annimmt. Der zugehörige Wert von  $h$  ist durch

$$h_0 := \alpha(r_0) = \frac{F - 2\pi r_0^2}{2\pi r_0} = \dots = 2\sqrt{\frac{F}{6\pi}}$$

gegeben.

Die interessante Aussage in diesem Beispiel ist, dass das Maximum für  $h = 2r$  angenommen wird. Diese erhält man mit Hilfe der Lagrangeschen Multiplikatoren direkt. Durch das Auflösen der Nebenbedingung nach einer Variablen erhält man zunächst die eher uninteressante Aussage, wie sich an der Stelle, wo das Maximum angenommen wird,  $r$  und  $h$  aus  $F$  berechnen lassen, und man kann die Bedingung  $r = 2h$  dann erst durch Elimination von  $F$  daraus gewinnen.