**Jürgen Müller**

# Analysis einer und mehrerer Veränderlicher

Skript zur Vorlesung Sommersemster 2021
Universität Trier
Fachbereich IV
Mathematik/Analysis

# Inhaltsverzeichnis

# 1 Metrische Räume und Funktionenfolgen

Wie wir bereits in der Einführung in die Mathematik gesehen haben, spielt in der Analysis das Konzept der Grenzwerte eine zentrale Rolle. Dabei ist es wesentlich, von Abständen zwischen zwei Elementen in einer Menge sprechen zu können. Wir betrachten nun ganz allgemein Mengen, in denen ein Abstand zwischen jeweils zwei Elementen definiert ist.

**Definition 1.1** Es sei $X \neq \varnothing$ eine Menge. Eine Abbildung $d : X \times X \to \mathbb{R}$ heißt **Metrik** (oder **Abstand**) auf $X$, falls folgende Bedingungen erfüllt sind:

(d1) (Definitheit) Für alle $x, y \in X$ ist $d(x,x) = 0$ und $d(x,y) > 0$, falls $x \neq y$.

(d2) (Symmetrie) Für alle $x, y \in X$ ist $d(x,y) = d(y,x)$.

(d3) (Dreiecksungleichung) Für alle $x, y, z \in X$ gilt $d(x,y) \leq d(x,z) + d(z,y)$.

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

**Bemerkung und Definition 1.2** 1. Ist $X \neq \varnothing$ eine beliebige Menge, so definiert

$$\delta(x,y) := \begin{cases} 0, & \text{falls} \quad x = y \\ 1, & \text{falls} \quad x \neq y \end{cases}$$

eine Metrik auf $X$, die sogenannte **diskrete Metrik**. Insbesondere kann also jede nichtleere Menge mit einer Metrik versehen werden.
2. Ist $(X, d)$ ein metrischer Raum und ist $M \subset X$ nichtleer, so ist durch $d_M := d|_{M \times M}$ eine Metrik auf $M$ gegeben. Man nennt $d_M$ die **Spurmetrik** von $d$ auf $M$.
3. Sind $(X_1, d_1), \ldots, (X_m, d_m)$ metrische Räume, so sind durch

$$d_{\max}(x,y) := \max\{d_j(x_j, y_j) : j = 1, \ldots, m\} \quad (x = (x_1, \ldots, x_m), y = (y_1, \ldots, y_m))$$

und

$$d_\Sigma(x,y) := \sum_{j=1}^{m} d_j(x_j, y_j) \qquad (x = (x_1, \ldots, x_m), y = (y_1, \ldots, y_m))$$

Metriken auf $X_1 \times \ldots \times X_m$ definiert mit ([Ü])

$$d_{\max} \leq d_\Sigma \leq m \cdot d_{\max} \,. \tag{1.1}$$

Wir betrachten eine Klasse von Räumen, die eine lineare Struktur und eine metrische Struktur besitzen.

**Definition 1.3** Es sei $V = (V, +, \cdot)$ ein Vektorraum über $\mathbb{K}$. Eine Abbildung $||\cdot|| : V \to \mathbb{R}$ heißt **Norm** (auf $V$), falls folgende Bedingungen erfüllt sind

(N1) (Positivität) Für alle $v \neq 0$ ist $||v|| > 0$.

(N2) (Homogenität) Für alle $v \in V$ und alle $\lambda \in \mathbb{K}$ ist $||\lambda v|| = |\lambda| \cdot ||v||$.

(N3) (Dreiecksungleichung) Für alle $u, v \in V$ gilt $||u + v|| \leq ||u|| + ||v||$.

Das Paar $(V, ||\cdot||)$ heißt dann ein **normierter Raum** (über $\mathbb{K}$).

**Bemerkung und Definition 1.4** Wir setzen

$$||x||_2 := \Big(\sum_{j=1}^m |x_j|^2\Big)^{1/2} \qquad (x = (x_1, \dots, x_m) \in \mathbb{K}^m).$$

Sind $a, b \in \mathbb{R}$, so ist $2ab \le a^2 + b^2$ wegen $(a-b)^2 \ge 0$. Damit ergibt sich für $x, y \in \mathbb{K}^m \setminus \{0\}$

$$\sum_{j=1}^m \frac{|x_j|}{||x||_2} \cdot \frac{|y_j|}{||y||_2} \le \frac{1}{2} \sum_{j=1}^m \Big(\frac{|x_j|^2}{||x||_2^2} + \frac{|y_j|^2}{||y||_2^2}\Big) = 1.$$

Also folgt (auch für $x = 0$ oder $y = 0$) die **Cauchy-Schwarzsche Ungleichung**

$$\sum_{j=1}^m |x_j y_j| \le ||x||_2 ||y||_2.$$

Hieraus ergibt sich wiederum

$$||x+y||_2^2 \le \sum_{j=1}^m |x_j| \cdot |x_j + y_j| + \sum_{j=1}^m |y_j| \cdot |x_j + y_j| \le (||x||_2 + ||y||_2) \cdot ||x+y||_2,$$

also auch $||x+y||_2 \le ||x||_2 + ||y||_2$. Damit ist (N3) für $||\cdot||_2$ erfüllt. Da auch (N1) und (N2) erfüllt sind, ist $||\cdot||_2$ eine Norm auf $\mathbb{K}^m$. Im Fall $\mathbb{K} = \mathbb{R}$ nennt man $||x||_2$ die **euklidsche Länge** von $x$. Wie man leicht sieht, sind durch

$$||x||_1 := \sum_{j=1}^m |x_j| \qquad (x = (x_1, \dots, x_m) \in \mathbb{K}^m)$$

und

$$||x||_\infty := \max\{|x_j| : j = 1, \dots, m\} \qquad (x = (x_1, \dots, x_m) \in \mathbb{K}^m)$$

weitere Normen auf $\mathbb{K}^m$ gegeben. Aus

$$|x_k|^2 \le \sum_{j=1}^m |x_j|^2 \le \sum_{j,\ell=1}^m |x_j x_\ell| = \Big(\sum_{j=1}^m |x_j|\Big)^2$$

für alle $x \in \mathbb{K}^m$ und $k \in \{1, \dots, m\}$ folgt

$$||x||_\infty \le ||x||_2 \le ||x||_1 \ .$$

Falls nicht anders angegeben, soll $\mathbb{K}^m$ stets mit der Norm $||\cdot||_2$ versehen sein. Wir schreiben auch kurz $|\cdot| := ||\cdot||_2$.

**Bemerkung und Definition 1.5** Ist $(V, \|\cdot\|)$ ein normierter Raum, so ist durch

$$d_{\|\cdot\|}(u,v) := \|u - v\| \qquad (u, v \in V)$$

eine Metrik auf $V$ gegeben, die von der Norm $\|\cdot\|$ **induzierte Metrik**. Insbesondere ist nach Bemerkung und Definition 1.4 durch

$$d(x,y) := |x-y| = ||x-y||_2 \qquad (x,y \in \mathbb{K}^m)$$

eine Metrik auf $\mathbb{K}^m$ definiert. Man spricht dann im Falle $m=1$ von der **Betragsmetrik** und im Falle $\mathbb{K}=\mathbb{R}$ auch von der **euklidschen Metrik**. Wenn wir im Weiteren von Teilmengen $X$ von $\mathbb{K}^m$ als metrischem Raum sprechen, soll stets deren Spurmetrik gemeint sein, falls nichts Anderes gesagt wird. Weiter gilt hier $d_{||\cdot||_\infty} = d_{\max}$ sowie $d_{||\cdot||_1} = d_\Sigma$ und, wieder nach Bemerkung und Definition 1.4,

$$d_{\max}(x,y) \le d(x,y) \le d_\Sigma(x,y) \qquad (x,y \in \mathbb{K}^d). \tag{1.2}$$

**Definition 1.6** Es sei $(X,d)$ ein metrischer Raum.
1. Für $a \in X$ und $0 \le \rho \le \infty$ setzen wir

$$U_\rho(a) := U_{\rho,d}(a) := U_{\rho,X}(a) := \{x \in X : d(x,a) < \rho\}.$$

Im Fall $\rho > 0$ heißt $U_\rho(a)$ die (offene) $\rho$-**Umgebung** von $a$. Weiter setzen wir

$$B_\rho(a) := B_{\rho,d}(a) := B_{\rho,X}(a) := \{x \in X : d(x,a) \le \rho\}.$$

2. Sind $a \in X$ und $M \subset X$, so heißt $a$ ein **innerer Punkt** von $M$ (in $(X,d)$), falls ein $\rho > 0$ existiert mit $U_\rho(a) \subset M$. In diesem Fall heißt zudem $M$ eine **Umgebung** von $a$ (in $(X,d)$). Weiter heißt $a$ ein **Häufungspunkt** von $M$, falls $M \cap (U_\rho(a) \setminus \{a\}) \neq \varnothing$ für alle $\rho > 0$ ist. Wir schreiben $M'$ für die Menge der Häufungspunkte von $M$ in $X$.

**Bemerkung und Definition 1.7** Es seien $(X,d)$ ein metrischer Raum und $(x_n)_{n\in N}$[1] eine Folge in $X$. Die Folge heißt $d$-**konvergent** oder kurz **konvergent** in $X$, falls die Metrik aus dem Kontext heraus klar ist, falls ein $c \in X$ so existiert, dass $(d(x_n,c))_{n\in N}$ abklingend (in $\mathbb{R}$) ist. Dann ist $c$ eindeutig bestimmt. Wir nennen $c$ den $d$-**Grenzwert** oder kurz **Grenzwert** und schreiben dann $x_n \to c$ für $n \to \infty$ (in $(X,d)$) oder auch

$$\lim x_n := \lim_{n\to\infty} x_n := d - \lim_{n\to\infty} x_n := c.$$

Im Fall $d = d_{||\cdot||}$ schreiben wir auch kurz $||\cdot||$ statt $d_{||\cdot||}$, also etwa $||\cdot||$-konvergent.

**Bemerkung 1.8** Es seien $X$ eine Menge und $\delta$ die diskrete Metrik auf $X$. Eine Folge $(x_n)$ in $X$ ist genau dann $\delta$-konvergent mit Grenzwert $c$, wenn $x_n = c$ für alle genügend großen $n$ gilt ([Ü]). Ist $X = \mathbb{R}$, so ist die Folge $(1/n)$ also *nicht* $\delta$-konvergent (obwohl $(1/n)$ natürlich $|\cdot|$-konvergent ist).

**Bemerkung und Definition 1.9** Es seien $(X,d)$ ein metrischer Raum und $(x_n)_{n\in N}$ eine Folge in $X$. Dann heißt $(x_n)_{n\in N}$ eine ($d$-)**Cauchyfolge**, falls zu jedem $\varepsilon > 0$ ein $R > 0$ existiert mit $d(x_n, x_{n'}) < \varepsilon$ für $n, n' > R$. Ganz allgemein gilt (mit entsprechendem Beweis wie im Fall $X \subset \mathbb{K}$; siehe Einführung in die Mathematik):

[1] Falls nicht anders gesagt, ist die Indexmenge $N$ bei Folgen stets eine nach oben unbeschränkte Menge in $\mathbb{N}_0$.

1. Jede konvergente Folge ist eine Cauchyfolge.
2. Jede Cauchyfolge, die eine konvergente Teilfolge besitzt, ist konvergent.

**Bemerkung 1.10** Ist $X = (0, \infty)$, so ist die Folge $(1/n)$ eine Cauchyfolge in $X$, aber nicht konvergent in $X$. Im Allgemeinen sind also in metrischen Räumen nicht alle Cauchyfolgen konvergent.

**Definition 1.11** Eine Metrik $d$ auf $X$ beziehungsweise der metrische Raum $(X, d)$ heißt **folgenvollständig** oder kurz **vollständig**, falls jede Cauchyfolge konvergiert. Ist $V$ ein normierter Raum über $\mathbb{K}$ mit der Eigenschaft, dass $d_{||\cdot||}$ vollständig ist, so nennt man $(V, ||\cdot||)$ einen **Banachraum** (über $\mathbb{K}$).

**Bemerkung 1.12** Ist $X \subset \mathbb{C}$ mit $X' \subset X$, so ist jede Cauchyfolge in $X$ nach dem Cauchykriterium für Folgen (siehe Einführung in die Mathematik) konvergent in $X$. Insbesondere ist $(\mathbb{K}, d_{|\cdot|})$ vollständig, also $(\mathbb{K}, |\cdot|)$ ein Banachraum.[2]

**Bemerkung 1.13** 1. Es seien $(X_1, d_1), \dots, (X_m, d_m)$ metrische Räume und es sei

$$(x_n)_{n \in N} = (x_{1,n}, \dots, x_{m,n})_{n \in N}$$

eine Folge in $X_1 \times \cdots \times X_m$. Aus den Abschätzungen (1.1) folgt unmittelbar, dass folgende Aussagen äquivalent sind:

a) $(x_n)_{n \in N}$ ist $d_{\max}$-konvergent.

b) $(x_n)_{n \in N}$ ist $d_\Sigma$-konvergent.

c) Jede **Komponentenfolge** $(x_{j,n})_{n \in N}$ ist $d_j$-konvergent.

Außerdem gilt im Falle der Konvergenz

$$\lim_{n \to \infty} x_n = \big( \lim_{n \to \infty} x_{1,n}, \dots, \lim_{n \to \infty} x_{m,n} \big).$$

Die Äquivalenz von a), b) und c) gilt auch, wenn man *konvergent* durch *Cauchyfolge* ersetzt. Ist speziell $X_1 \times \cdots \times X_m = \mathbb{K}^m$, so ergibt sich mit (1.2), dass a)-c) auch äquivalent dazu sind, dass die Folge $|\cdot|$-konvergent (beziehungsweise eine $|\cdot|$-Cauchyfolge) ist. Mit Bemerkung 1.12 ergibt sich die Vollständigkeit von $(\mathbb{K}^m, d_{|\cdot|})$. Also ist $(\mathbb{K}^m, |\cdot|)$ ein Banachraum.

Wie wir in der Einführung in die Mathematik gesehen haben, sind wichtige Funktionen wie die Exponentialfunktion und die trigonometrischen Funktionen über gewisse Reihenwerte definiert. Wir wollen nun allgemeine Strukturaussagen über Funktionen machen, die sich als Grenzwerte sogenannter Funktionenfolgen oder -reihen ergeben.

[2] Das Cauchykriterium ist im Wesentlichen eine Folgerung aus der Ordnungsvollständigkeit von $\mathbb{R}$. Man kann also sagen, dass die Ordnungsvollständigkeit die Folgenvollständigkeit von $\mathbb{R}$ impliziert.

**Definition 1.14** Es seien $X \neq \varnothing$ eine Menge, $(Y, d)$ ein metrischer Raum und $(f_n)_{n \in N}$ eine Folge in $\mathrm{Abb}(X, Y)$.[3] Die Funktionenfolge $(f_n)_{n \in N}$ heißt **punktweise konvergent** auf der Menge $M \subset X$, falls für alle $x \in M$ die Folge $(f_n(x))_{n \in N}$ in $Y$ konvergiert. Die Funktion $f : M \to Y$ mit

$$f(x) := \lim_{n \to \infty} f_n(x)$$

heißt **Grenzfunktion** der Folge $(f_n)_{n \in N}$ (auf $M$). Wir schreiben dann auch

$$f_n \to f \ (n \to \infty) \quad \text{punktweise auf } M.$$

**Beispiel 1.15** Wir betrachten die Funktionen $f_n : \mathbb{R} \to \mathbb{R}$ mit $f_n(x) := x^n$ für $x \in \mathbb{R}$ und $n \in \mathbb{N}$.

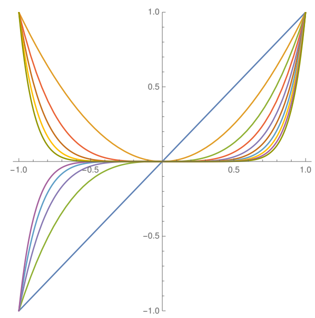


Abbildung 1: $x \mapsto x^n$ auf $[-1, 1]$ für $n = 1, \ldots, 10$.

Dann gilt

$$f_n(x) \to \begin{cases} 0 & , \quad \text{falls} \quad x \in (-1, 1) \\ 1 & , \quad \text{falls} \quad x = 1 \end{cases} ,$$

das heißt, $(f_n)$ konvergiert punktweise auf $(-1, 1]$ gegen der Grenzfunktion $f : (-1, 1] \to \mathbb{R}$ mit $f(x) = 0$ für $x \in (-1, 1)$ und $f(1) = 1$. Außerdem divergiert die Folge $(f_n(x))$ für alle anderen $x$. Das Beispiel zeigt insbesondere, dass die Grenzfunktion unstetig (an der Stelle $1$) ist, obwohl alle Folgenglieder $f_n$ stetige Funktionen auf $\mathbb{R}$ sind.

Wir führen nun einen strengeren Konvergenzbegriff für Funktionenfolgen ein, der den entscheidenden Vorteil hat, dass sich Stetigkeit auf die Grenzfunktion überträgt.

**Bemerkung und Definition 1.16** Es seien $X \neq \varnothing$ eine Menge und $(Y, d)$ ein metrischer Raum. Eine Folge $(f_n)_{n \in N}$ in $\mathrm{Abb}(X, Y)$ heißt **gleichmäßig konvergent** auf der Menge $M \subset X$, falls eine Funktion $f : M \to Y$ so existiert, dass die Folge $\left(\sup_{x \in M} d(f_n(x), f(x))\right)_{n \in N}$ in $\mathbb{R}$ abklingend ist. Wir schreiben dann

$$f_n \to f \ (n \to \infty) \quad \text{gleichmäßig auf } M$$

oder auch $f_n(x) \to f(x)$ für $n \to \infty$ gleichmäßig auf $M$. Aus $f_n \to f$ gleichmäßig auf $M$ folgt wegen $d(f_n(x), f(x)) \leq \sup_M d(f_n, f)$ für $x \in M$ auch $f_n \to f$ punktweise auf $M$, mit anderen Worten: gleichmäßige Konvergenz gegen $f$ impliziert punktweise Konvergenz gegen $f$.

[3] Man spricht dann von einer Funktionenfolge.

**Beispiel 1.17** Wir betrachten wieder $f_n$ und $f$ aus Beispiel 1.15. Ist $M = [-1/2, 1/2]$, so gilt $f|_M = 0$ und

$$\sup_{x\in M} |x^n - 0| = 1/2^n \to 0 \qquad (n \to \infty) \,,$$

also $f_n \to 0$ $(n \to \infty)$ gleichmäßig auf $[-1/2, 1/2]$. Für $M = [0, 1)$ ist ebenfalls $f|_M = 0$ und

$$\sup_{x\in[0,1)} |x^n - 0| = \sup_{x\in[0,1)} x^n = 1 \quad (n \in \mathbb{N}) \,.$$

Also ist $(f_n)$ nicht gleichmäßig konvergent auf $[0, 1)$.

**Bemerkung und Definition 1.18** Es seien $(X, d_X)$ und $(Y, d)$ metrische Räume und $f : X \to Y$. Ist $a \in X$, so heißt $f$ **stetig** an der Stelle $a \in X$, falls zu jedem $\varepsilon > 0$ ein $\delta = \delta_\varepsilon > 0$ existiert mit $d(f(x), f(a)) < \varepsilon$ für alle $x \in U_{\delta,X}(a)$. Weiterhin heißt $f$ stetig auf $M \subset X$, falls $f$ stetig in jedem Punkt $a \in M$ ist und kurz stetig, falls $f$ stetig auf $X$ ist. Aus der Definition ergibt sich leicht ([Ü]): Ist $(Z, d_Z)$ ein weiterer metrischer Raum und sind $f$ stetig an $a$ und $g : Y \to Z$ stetig an $f(a)$, so ist $g \circ f$ stetig an $a$.

Wir kommen nun zu dem bereits angedeuteten Ergebnis über die Vererbung der Stetigkeit auf die Grenzfunktion.

**Satz 1.19** *Es seien $(X, d_X)$, $(Y, d)$ metrische Räume und $a \in X$. Weiter sei $(f_n)_{n\in N}$ eine Folge in $\mathrm{Abb}(X, Y)$, die gleichmäßig auf einer Umgebung $U$ von $a$ gegen $f$ konvergiert. Sind die Funktionen $f_n$ stetig an $a$, so ist auch $f$ stetig an $a$.*

**Beweis.** Es sei $\varepsilon > 0$ gegeben. Aufgrund der gleichmäßigen Konvergenz von $(f_n)$ gegen $f$ auf $U$ existiert ein $n = n_\varepsilon \in N$ mit $d(f(x), f_n(x)) < \varepsilon/3$ *für alle* $x \in U$. Da $f_n$ stetig an $a$ ist, existiert ein $\delta = \delta_\varepsilon > 0$ mit $U_\delta(a) \subset U$ und $d(f_n(x), f_n(a)) < \varepsilon/3$ für $x \in U_\delta(a)$. Mit der Dreiecksungleichung gilt für $x \in U_\delta(a)$

$$\begin{aligned} d(f(x), f(a)) &\le d(f(x), f_n(x)) + d(f_n(x), f(a)) \\ &\le d(f(x), f_n(x)) + d(f_n(x), f_n(a)) + d(f_n(a), f(a)) < \varepsilon \,. \end{aligned}$$

□

**Bemerkung und Definition 1.20** Es seien $M \neq \emptyset$ eine Menge und $(E, |\cdot|)$ ein normierter Raum. Eine Funktion $f : M \to E$ heißt **beschränkt**, falls $\sup_{x\in M} |f(x)| < \infty$ gilt. Wir schreiben $B(M, E)$ für die Menge der beschränkten Funtionen $f : M \to E$.[4] Dann ist durch

$$\|f\|_\infty := \|f\|_{\infty,M} := \sup_{x\in M} |f(x)| \qquad (f \in B(M))$$

eine Norm auf $B(M, E)$ definiert ([Ü]), gennant **Supremumsnorm**. Sind $(f_n)$ eine Folge in $B(M, E)$ und $f \in B(M, E)$, so ist $(f_n)$ genau dann $\|\cdot\|_\infty$-konvergent gegen $f$, wenn $f_n \to f$ gleichmäßig auf $M$ gilt (mit $(Y, d) = (E, d_{|\cdot|})$).

[4] Vorerst interessiert uns eigentlich nur der Fall $E = \mathbb{C}$. Wir schreiben dann kurz $B(M)$ statt $B(M, E)$.

**Satz 1.21** *Es seien $M \neq \varnothing$ eine Menge und $(E, |\cdot|)$ ein Banachraum. Dann ist der normierte Raum $(B(M,E), \|\cdot\|_\infty)$ ebenfalls ein Banachraum.*

**Beweis.** Es sei $(f_n)_{n\in N}$ eine Cauchy-Folge in $B(M,E)$. Dann ist insbesondere für jedes feste $x \in M$ die Folge $(f_n(x))_{n\in N}$ eine Cauchyfolge in $E$. Da $(E, d_{|\cdot|})$ vollständig ist, ist $(f_n(x))_{n\in N}$ konvergent. Also konvergiert $(f_n)$ punktweise. Es sei $f : M \to \mathbb{C}$ die Grenzfunktion. Wir zeigen, dass $(f_n)$ gleichmäßig auf $M$ gegen $f$ konvergiert.
Dazu sei $\varepsilon > 0$ gegeben. Dann existiert ein $R = R_\varepsilon > 0$ mit $\|f_n - f_{n'}\|_\infty < \varepsilon$ für $n, n' > R$. Es sei $n > R$ fest. Ist $x \in M$, so ist nach der umgekehrten Dreiecksungleichung

$$\big||f_n(x) - y| - |f_n(x) - y'|\big| \leq |y - y'| \qquad (y, y' \in E).$$

Wegen $|f_m(x) - f(x)| \to 0$ für $m \to \infty$ gilt damit

$$|f_n(x) - f_m(x)| \to |f_n(x) - f(x)| \qquad (m \to \infty)\ .$$

Aus $|f_n(x) - f_m(x)| < \varepsilon$ für $m > R$ folgt $|f_n(x) - f(x)| \leq \varepsilon$, und da $x \in M$ beliebig war, ist damit auch $\|f - f_n\|_\infty \leq \varepsilon$.
Schießlich ist $f \in B(M,E)$, denn wählt man ein $n \in N$ mit $||f - f_n||_\infty < 1$, so gilt $|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq 1 + ||f_n||_\infty$ für alle $x \in X$. □

**Definition 1.22** E seien $X \neq \varnothing$ eine Menge, $(E; |\cdot|)$ ein normierter Raum und $(f_n)_{n\geq m}$ eine Folge in $\mathrm{Abb}(X,E)$. Dann heißt die Folge $(s_n)_{n\geq m}$ der Partialsummen $s_n := \sum_{\nu=m}^{n} f_\nu$ die mit $(f_n)$ gebildete **Funktionenreihe**. Man schreibt wie bei Zahlenreihen $\sum_{\nu=m}^{\infty} f_\nu$ statt $(s_n)_{n\geq m}$. Die Funktionenreihe $\sum_{\nu=m}^{\infty} f_\nu$ heißt **punktweise konvergent** bzw. **gleichmäßig konvergent** auf $M \subset X$, falls die Funktionenfolge $(s_n)$ auf $M$ punktweise bzw. gleichmäßig konvergiert. Man verwendet das Symbol $\sum_{\nu=m}^{\infty} f_\nu$ dann auch für die Grenzfunktion.

Ein einfaches hinreichendes Kriterium für gleichmäßige Konvergenz ist[5]

**Satz 1.23 (Weierstraß-Kriterium)**
*Es seien $M \neq \varnothing$ eine Menge, $(E, |\cdot|)$ ein Banachraum und $(f_n)_{n\geq m}$ eine Folge in $B(M,E)$. Existieren eine Folge $(b_n)_{n\geq m}$ in $[0,\infty)$ mit $\sum_{\nu=m}^{\infty} b_\nu < \infty$ und ein $n_0 \geq m$ mit $\|f_n\|_\infty \leq b_n$ für $n \geq n_0$, so konvergiert $\sum_{\nu=m}^{\infty} f_\nu$ gleichmäßig auf $M$.*

**Beweis.** Mit $f_n$ sind auch $s_n = \sum_{\nu=m}^{n} f_\nu \in B(M,E)$. Ist $\varepsilon > 0$ gegeben, so existiert nach dem Cauchy-Kriterium (siehe Einführung in die Mathematik) ein $R \geq n_0$ mit

$$\sum_{\nu=n'+1}^{n} b_\nu < \varepsilon \qquad (n > n' > R).$$

[5] Für Reihen $\sum_{\nu=m}^{\infty} b_\nu$ mit $b_\nu \geq 0$ schreiben wir oft kurz $\sum_{\nu=m}^{\infty} b_\nu < \infty$ im Falle der Konvergenz.

Damit ergibt sich für $n > n' > R$

$$||s_n - s_{n'}||_\infty = \Big\| \sum_{\nu=n'+1}^{n} f_\nu \Big\|_\infty \leq \sum_{\nu=n'+1}^{n} b_\nu < \varepsilon.$$

Also ist $(s_n)$ eine Cauchyfolge in $B(M,E)$ und folglich nach Satz 1.21 konvergent in $B(M,E)$. □

**Beispiel 1.24** Wir betrachten die Funktionen $f_n : \mathbb{C} \to \mathbb{C}$ mit $f_n(z) := 1/n^z = e^{-z \ln n}$. Wegen der Stetigkeit von $\exp$ ist jedes $f_n$ stetig. Sind $\alpha > 1$ und $M_\alpha := \{z \in \mathbb{C} : \operatorname{Re} z > \alpha\}$, so ist

$$|f_n(z)| = |e^{-z \ln n}| = e^{-\operatorname{Re}(z) \ln n} = 1/n^{\operatorname{Re}(z)} \leq 1/n^\alpha \quad (z \in M_\alpha,\, n \in \mathbb{N}).$$

Da $\sum\limits_{\nu=1}^{\infty} 1/\nu^\alpha < \infty$ für $\alpha > 1$ gilt ([Ü]), ist die Funktionenreihe $\sum\limits_{\nu=1}^{\infty} f_\nu$ nach dem Weierstraß-Kriterium gleichmäßig konvergent auf $M_\alpha$ und damit punktweise auf $M := \{z \in \mathbb{C} : \operatorname{Re} z > 1\}$. Die Funktion $\zeta : M \to \mathbb{C}$, definiert durch

$$\zeta(z) := \sum_{\nu=1}^{\infty} 1/\nu^z \qquad (z \in M),$$

heißt **(Riemannsche) Zetafunktion**. Da $M_\alpha$ eine Umgebung jedes Punktes $a \in M_\alpha$ ist, folgt aus Satz 1.19 die Stetigkeit von $\zeta$ auf $M$.

**Bemerkung 1.25** Es seien $\sum\limits_{\nu=0}^{\infty} c_\nu z^\nu$ eine Potenzreihe mit Konvergenzradius

$$R := 1/\inf\{q \geq 0 : \sqrt[n]{|c_n|} \leq q \text{ für } n \text{ genügend groß}\} > 0.$$

Aus den Überlegungen in der Einführung in die Mathematik[6] und dem Weierstraß-Kriterium ergibt sich, dass die Potenzreihe für alle $r < R$ gleichmäßig konvergent auf $U_r(0)$ ist.

Aus Bemerkung 1.25 Satz 1.19 folgt, dass dass Funktionen, die durch Potenzreihen darstellbar sind, stetig auf dem Konvergenzkreis sind. Wir zeigen nun viel schärfer:

**Satz 1.26** *Es sei* $\sum\limits_{\nu=0}^{\infty} c_\nu z^\nu$ *eine Potenzreihe mit Konvergenzradius* $R > 0$*. Ist* $f(z) := \sum\limits_{\nu=0}^{\infty} c_\nu z^\nu$ *für* $z \in U = U_R(0)$*, so ist* $f$ *beliebig oft differenzierbar auf* $U$ *und*

$$\frac{1}{k!} f^{(k)}(z) = \sum_{\nu=0}^{\infty} \binom{\nu+k}{k} c_{\nu+k}\, z^\nu \qquad (z \in U,\, k \in \mathbb{N}_0)\ .$$

*Insbesondere ist* $f^{(k)}(0)/k! = c_k$ *für* $k \in \mathbb{N}_0$*, also* $f(z) = \sum\limits_{\nu=0}^{\infty} \frac{f^{(\nu)}(0)}{\nu!} z^\nu$ *für* $z \in U$*.*

[6] siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, B/D 6.1

**Beweis.** 1. Wir zeigen: $f$ ist differenzierbar auf $U$ und

$$f'(x) = \sum_{\nu=1}^{\infty} \nu c_\nu z^{\nu-1} = \sum_{\nu=0}^{\infty} (\nu+1) c_{\nu+1} z^\nu \qquad (z \in U)\,.$$

Dazu sei $z \in U$. Dann ist $z \in U_r(0)$ für ein $r < R$. Wir wählen ein $\delta > 0$ mit $U_\delta(z) \subset U_r(0)$. Für $h \in U_\delta(0)$ gilt $|z+h| < r$ und[7]

$$(z+h)^n - z^n = h \sum_{k=0}^{n-1} (z+h)^k z^{n-k-1} =: h\phi_n(h) \quad (n \in \mathbb{N})$$

mit $|\phi_n(h)| \leq n r^{n-1}$. Nach Definition des Konvergenzradius existiert ein $s < 1$ mit $\sqrt[n]{|c_n|} r \leq s$ für $n$ genügend groß. Wegen $\sqrt[n]{n} \to 1$ für $n \to \infty$ ist die Reihe $\sum\limits_{\nu=1}^{\infty} \nu |c_\nu| r^{\nu-1}$ nach dem Wurzelkriterium[8] konvergent. Aus dem Weierstraß-Kriterium (Satz 1.23) folgt damit die gleichmäßige Konvergenz der Funktionenreihe $\sum\limits_{\nu=1}^{\infty} c_\nu \phi_\nu$ auf $U_\delta(0)$. Nach Satz 1.19 ist die Funktion $\phi := \sum\limits_{\nu=1}^{\infty} c_\nu \phi_\nu$ stetig an 0. Also gilt

$$\frac{1}{h}(f(z+h) - f(z)) = \sum_{\nu=1}^{\infty} c_\nu \frac{1}{h}\big((z+h)^\nu - z^\nu\big) = \phi(h) \to \phi(0) \quad (h \to 0).$$

Aus $\phi(0) = \sum\limits_{\nu=1}^{\infty} \phi_\nu(0) = \sum\limits_{\nu=1}^{\infty} \nu c_\nu z^{\nu-1}$ folgt die Behauptung.

2. Induktiv erhält man mit Beweisschritt 1, dass $f^{(k-1)}$ für alle $k \in \mathbb{N}$ differenzierbar auf $U$ ist mit

$$f^{(k)}(z) = \sum_{\nu=0}^{\infty} (\nu+1)(\nu+2)\dots(\nu+k) c_{\nu+k} z^\nu\,.$$

Also folgt die Behauptung nach Division durch $k!$. □

**Beispiel 1.27** Da $\exp'(0) = 1$ ist, gilt $(e^z - 1)/z \to 1$ für $z \to 0$. Unter Verwendung des Satzes 1.26 lässt sich viel mehr aussagen: Die Funktion $f : \mathbb{C} \to \mathbb{C}$ mit

$$f(z) = \sum_{\nu=0}^{\infty} \frac{z^\nu}{(\nu+1)!} = \begin{cases} z^{-1} \sum_{\mu=1}^{\infty} (\mu!)^{-1} z^\mu = (e^z - 1)/z\,, & z \neq 0 \\ 1\,, & z = 0 \end{cases}$$

ist beliebig oft differenzierbar auf $\mathbb{C}$ (also insbesondere an 0) und es gilt $f^{(k)}(0) = 1/(k+1)$ für alle $k \in \mathbb{N}_0$.

**Bemerkung und Definition 1.28** Es seien $X \subset \mathbb{K}$ und $f : X \to \mathbb{C}$. Eine Funktion $F : X \to \mathbb{C}$ heißt **Stammfunktion** zu $f$ (auf $X$), falls $F$ differenzierbar ist mit $F' = f$ auf $X$. Ist $X \subset \mathbb{R}$ ein Intervall[9] und sind $F$ und $G$ Stammfunktionen zu $f$ auf $X$, so ist $(F - G)' = 0$

[7] siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, Satz 2.17
[8] siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, Satz 5.17
[9] oder allgemeiner $X \subset \mathbb{C}$ sternförmig

auf $X$, und damit ist $F-G$ nach dem Schrankensatz (siehe Einführung in die Mathematik) konstant auf $X$, mit anderen Worten, $F$ und $G$ unterscheiden sich lediglich durch eine additive Konstante.

**Bemerkung 1.29** Ist $\sum_{\nu=0}^{\infty} c_\nu z^\nu$ eine Potenzreihe mit Konvergenzradius $R > 0$ und ist

$$f(z) := \sum_{\nu=0}^{\infty} c_\nu z^\nu \qquad (z \in U_R(0)),$$

so ist nach Satz 1.26 durch

$$F(z) := \sum_{\nu=0}^{\infty} \frac{c_\nu}{\nu+1} z^{\nu+1} = \sum_{\nu=1}^{\infty} \frac{c_{\nu-1}}{\nu} z^\nu \qquad (z \in U_R(0))$$

eine Stammfunktion zu $f$ auf $U_R(0)$ definiert.[10]

**Beispiel 1.30** Ist $f(z) := 1/(1-z)$ für $z \in \mathbb{C} \setminus \{1\}$, so ist nach Bemerkung 1.29 wegen $f(z) = \sum_{\nu=0}^{\infty} z^\nu$ für $z \in \mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ durch

$$F(z) = \sum_{\nu=1}^{\infty} \frac{z^\nu}{\nu} \qquad (z \in \mathbb{D})$$

eine Stammfunktion zu $f$ auf $\mathbb{D}$ gegeben. Auf $(-\infty, 1)$ definiert

$$G(x) = \ln\left(\frac{1}{1-x}\right) = -\ln(1-x) \qquad (x < 1)$$

ebenfalls eine Stammfunktion zu $f$. Nach Bemerkung 1.28 sind $F$ und $G$ auf $I = (-1,1)$ bis auf eine additive Konstante gleich. Da $F(0) = 0 = G(0)$ gilt, ist die additive Konstante $= 0$, und damit stimmen $F$ und $G$ auf $I$ überein. Also folgt

$$\ln\left(\frac{1}{1-x}\right) = \sum_{\nu=1}^{\infty} \frac{x^\nu}{\nu} \qquad (-1 < x < 1).$$

Man spricht daher auch von der **Logarithmusreihe**.

Am Rande ihres Konvergenzkreises können Funktionen, die durch Potenzreihen definiert sind, ein sehr kompliziertes Verhalten zeigen. Konvergiert die Potenzreihe an einem Randpunkt $\zeta$ des Konvergenzkreises, so existiert jedenfalls der sogenannte radiale Randwert der Grenzfunktion an der Stelle $\zeta$. Wir formulieren das entsprechende Ergebnis für Potenzreihen mit Konvergenzradius 1. Der allgemeine Fall kann leicht darauf zurückgeführt werden.

**Satz 1.31 (Abelscher Grenzwertsatz)**
*Es sei $\sum_{\nu=0}^{\infty} c_\nu z^\nu$ eine Potenzreihe mit Konvergenzradius 1 und $f(z) := \sum_{\nu=0}^{\infty} c_\nu z^\nu$ für $|z| < 1$. Ist $\zeta$ mit $|\zeta| = 1$ so, dass die Reihe $\sum_{\nu=0}^{\infty} c_\nu \zeta^\nu$ konvergiert, so gilt*

$$\lim_{r \to 1^-} f(r\zeta) = \sum_{\nu=0}^{\infty} c_\nu \zeta^\nu .$$

[10] Aus dem Wurzelkriterium folgt, dass der Konvergenzradius auch $R$ ist.

**Beweis.** Ohne Einschränkung können $\zeta = 1$ annehmen (ansonsten betrachte man $g(z) := f(\zeta z)$). Wir setzen $s_n := \sum_{\nu=0}^{n} c_\nu$ für $n \in \mathbb{N}_0$ und $s := \sum_{\nu=0}^{\infty} c_\nu = \lim s_n$. Da $(s_n)$ beschränkt ist, hat die Potenzreihe $\sum_{\nu=0}^{\infty} s_\nu z^\nu$ Konvergenzradius $\geq 1$. Also gilt mit $s_{-1} := 0$ für $|z| < 1$

$$(1-z)\sum_{\nu=0}^{\infty} s_\nu z^\nu = \sum_{\nu=0}^{\infty} s_\nu z^\nu - \sum_{\nu=0}^{\infty} s_\nu z^{\nu+1} = \sum_{\nu=0}^{\infty}(s_\nu - s_{\nu-1})z^\nu = \sum_{\nu=0}^{\infty} c_\nu z^\nu = f(z)\ .$$

Es sei $\varepsilon > 0$ gegeben. Dann existiert ein $n \in \mathbb{N}$ mit $|s_\nu - s| < \varepsilon$ für alle $\nu > n$. Für $0 < r < 1$ ergibt sich mit $1 = (1-r)\sum_{\nu=0}^{\infty} r^\nu$

$$\begin{aligned} |f(r) - s| &= |(1-r)\sum_{\nu=0}^{\infty}(s_\nu - s)r^\nu| \\ &\leq (1-r)\sum_{\nu=0}^{n}|s_\nu - s|r^\nu + \varepsilon(1-r)\sum_{\nu=n+1}^{\infty} r^\nu \leq (1-r)\sum_{\nu=0}^{n}|s_\nu - s| + \varepsilon. \end{aligned}$$

Aus $(1-r)\sum_{\nu=0}^{n}|s_\nu - s| \to 0$ für $r \to 1^-$ folgt die Existenz eines $\delta > 0$ mit $|f(r) - s| < 2\varepsilon$ für $1 - \delta < r < 1$. □

**Beispiel 1.32** Nach Beispiel 1.30 ist

$$\ln\Big(\frac{1}{1-x}\Big) = \sum_{\nu=1}^{\infty} \frac{x^\nu}{\nu} \qquad \big(x \in (-1,1)\big).$$

Da die alternierende harmonische Reihe $\sum_{\nu=1}^{\infty}(-1)^\nu/\nu$ nach dem Leibniz-Kriterium konvergiert, ergibt sich mit dem Abelschen Grenzwertsatz für $\zeta = -1$

$$\sum_{\nu=1}^{\infty} \frac{(-1)^\nu}{\nu} = \lim_{r\to 1^-} \ln(1/(1+r)) = \ln(1/2) = -\ln(2).$$

**Bemerkung und Definition 1.33** Es seien $X \subset \mathbb{K}$ und $a \in X$ ein innerer Punkt. Dann heißt $f : X \to \mathbb{C}$ **analytisch** an der Stelle $a$, falls eine Umgebung $U$ von $a$ und eine Folge $(c_k)$ in $\mathbb{C}$ so existieren, dass

$$f(x) = \sum_{\nu=0}^{\infty} c_\nu (x-a)^\nu \qquad (x \in U).$$

Ist $f$ analytisch an der Stelle $a$, so sieht man mit Satz 1.26, dass $c_k = f^{(k)}(a)/k!$ und damit $f(a+h) = \sum_{\nu=0}^{\infty} \frac{f^{(k)}(a)}{k!}h^\nu$ für $|h|$ genügend klein gilt. Die Koeffizienten $f^{(k)}(a)/k!$ nennt man die **Taylor-Koeffizienten** von $f$ und die Potenzreihe $\sum_{\nu=0}^{\infty} \frac{f^{(k)}(a)}{k!}h^\nu$ die **Taylor-Reihe** von $f$ bezüglich $a$. Weiter heißt $f$ analytisch auf $X$, falls $f$ analytisch an jedem Punkt $a \in X$ ist. In diesem Fall ist $f$ nach Satz 1.26 beliebig oft differenzierbar auf $X$.

**Beispiel 1.34** Sind $c \in \mathbb{C}$ und $f : \mathbb{C} \to \mathbb{C}$ mit $f(z) = e^{cz}$, so gilt für beliebiges $a \in \mathbb{C}$

$$f(z) = e^{ca}e^{c(z-a)} = \sum_{\nu=0}^{\infty} \frac{c^\nu e^{ca}}{\nu!}(z-a)^\nu \quad (z \in \mathbb{C}).$$

Also ist $f$ analytisch auf $\mathbb{C}$ mit $f^{(k)}(a) = c^k e^{ca}$ für alle $a$. Als Linearkombinationen analytischer Funktionen sind damit auch $\sin$, $\cos$, $\sinh$, $\cosh$ analytisch auf $\mathbb{C}$.

# 2 Integralrechnung

Die Integralrechnung entstand ursprünglich aus der Frage nach der Definition und der Berechnung von Flächeninhalten. Ähnlich wie bei der Differenzialrechnung werden wir Integrale über einen gewissen Grenzprozess einführen. Dazu betrachten wir zunächst besonders einfache Funktionen, für die wir die „orientierte Fläche unter den Graphen“ in sehr natürlicher Weise über die Flächen von Rechtecken definieren können.
Ist $I \subset \mathbb{R}$ ein Intervall, so schreiben wir $|I| := \operatorname{diam}(I)$[11] und nennen $|I|$ die **Länge** von $I$.

**Bemerkung und Definition 2.1** Es seien $a, b \in \mathbb{R}$ mit $a \leq b$.
1. Eine endliche Menge $E$ nichtleerer Intervalle $I \subset [a, b]$ heißt eine **Intervallzerlegung** oder kurz **Zerlegung** von $[a, b]$, falls die Intervalle paarweise disjunkt sind (also $I_1 \cap I_2 = \varnothing$ für $I_1, I_2 \in E$ mit $I_1 \neq I_2$) und $[a, b] = \bigcup_{I \in E} I$ gilt. Ist $E$ eine Zerlegung von $[a, b]$ und ist $J \subset [a, b]$ ein weiteres Intervall, so gilt

$$|J| = \sum_{I \in E} |I \cap J|. \tag{2.1}$$

2. Eine Funktion $\varphi \in B[a, b] := B([a, b], \mathbb{C})$ heißt **Treppenfunktion** (auf $[a, b]$), falls eine Zerlegung $E$ von $[a, b]$ so existiert, dass $\varphi$ konstant auf jedem Intervall $I \in E$ ist, d. h. falls für $I \in E$ Konstanten $c(I) = c_\varphi(I) \in \mathbb{C}$ existieren mit

$$\varphi = \sum_{I \in E} c(I) \cdot 1_I.$$

Eine Zerlegung, zu der entsprechende Konstanten $c(I)$ existieren, nennen wir **zulässig** für die Funktion $\varphi$. Wir schreiben $T[a, b]$ für die Menge der Treppenfunktionen auf $[a, b]$.

**Beispiel 2.2** 1. Wir betrachten $[a, b] = [0, 1]$ und die Funktion $\varphi = 1_{(1/2,1]}$. Dann ist $\varphi$ eine Treppenfunktion und etwa

$$E := \{[0, 1/2], (1/2, 1]\}$$

eine zulässige Zerlegung für $\varphi$, wobei hier $c([0, 1/2]) = 0$ und $c((1/2, 1]) = 1$ gilt. Eine weitere ist etwa

$$F = \{[0, 1/2], (1/2, 3/4], (3/4, 1]\},$$

wobei dann $c([0, 1/2]) = 0$ und $c((1/2, 3/4]) = c((3/4, 1]) = 1$ gilt. Übrigens ist auch $\{[0, 1/2], (1/2, 1), \{1\}\}$ zulässig, da einpunktige Intervalle nicht ausgeschlossen sind.
2. Für $[a, b] \subset \mathbb{R}$, $n \in \mathbb{N}$ und $j = 0, \ldots, n$ sei $t_{j,n} := a + j(b - a)/n$. Dann ist mit $I_{0,n} := \{a\}$ und $I_{j,n} := (t_{j-1,n}, t_{j,n}]$ für $j = 1, \ldots, n$ durch $E_n([a, b]) := \{I_{j,n} : j = 0, \ldots, n\}$ eine Zerlegung von $[a, b]$ gegeben. Hier gilt $|I_{j,n}| = (b - a)/n$ für $j = 1, \ldots, n$ [12] (und $|I_{0,n}| = 0$).

**Bemerkung und Definition 2.3** Es seien $E$ und $F$ Zerlegungen von $[a, b]$. Dann heißt

$$E \wedge F := \{I \cap J : I \in E,\, J \in F,\, I \cap J \neq \varnothing\},$$

[11] Sind $(X, d)$ ein metrischer Raum und $M \subset X$, so heißt $\operatorname{diam}(M) := \sup\{d(x, y) : x, y \in M\}$ (mit $\operatorname{diam} \varnothing := 0$) der Durchmesser von $M$.

[12] Man spricht daher auch von einer äquidistanten Zerlegung von $[a, b]$

die **gemeinsame Verfeinerung** von $E$ und $F$. Die gemeinsame Verfeinerung ist ebenfalls eine Zerlegung von $[a,b]$. Ist $\varphi : [a,b] \to \mathbb{C}$ eine Treppenfunktion und ist $E$ zulässig für $\varphi$, so ist auch $E \wedge F$ zulässig für $\varphi$. Ist auch $F$ zulässig für $\varphi$, so gilt $\varphi|_{I\cap J} = c(I) = c(J)$ für $I \in E$ und $J \in F$ mit $I \cap J \neq \varnothing$. Wegen (2.1) ergibt sich

$$\sum_{I\in E} c(I)|I| = \sum_{K\in E\wedge F} c(K)|K| = \sum_{J\in F} c(J)|J| \; .$$

Ist $E$ zulässig für $\varphi$, so heißt

$$\int_a^b \varphi := \int_a^b \varphi(t)\,dt := \sum_{I\in E} c(I)\cdot |I| \; ,$$

**Integral** von $\varphi$ (auf $[a,b]$). Wichtig ist dabei: Die Summe auf der rechten Seite ist unabhängig von der Wahl der zulässigen Zerlegung!

**Beispiel 2.4** In der Situation von Beispiel 2.2 gilt $\int_0^1 \varphi = \sum\limits_{I\in E} c(I)\cdot|I| = 1/2$.

**Definition 2.5** Es seien $M \neq \varnothing$ eine Menge und $L$ ein linearer Teilraum von $\mathrm{Abb}(M,\mathbb{K})$. Ist $\ell : L \to \mathbb{K}$ eine lineare Abbildung, so sagen wir $\ell$ sei **nichtnegativ**, falls $\ell(f) \geq 0$ für alle $f$ mit $f \geq 0$ gilt.[13] Aufgrund der Linearität ist in diesem Fall $\ell$ auch **monoton** in dem Sinne, dass $\ell(g) \leq \ell(f)$ für alle reellwertigen $f, g$ mit $g \leq f$ gilt.

**Satz 2.6** *Die Abbildung $\int_a^b : T[a,b] \to \mathbb{C}$ ist linear und monoton. Außerdem gilt für $\varphi \in T[a,b]$ und $\tau \in [a,b]$:*

1. *$|\varphi|$ ist eine Treppenfunktion und*

$$\Big|\int_a^b \varphi\Big| \leq \int_a^b |\varphi| \leq (b-a)\max_{[a,b]}|\varphi| \; .$$

2. *$\int_a^b \varphi = \int_a^\tau \varphi + \int_\tau^b \varphi$ .*

**Beweis.** Es seien $\varphi, \psi$ Treppenfunktionen und $\lambda \in \mathbb{C}$. Sind $E$ beziehungsweise $F$ zulässige Zerlegungen für $\varphi$ beziehungsweise $\psi$, so ist die gemeinsame Verfeinerung $E \wedge F$ sowohl für $\varphi$ als auch für $\psi$ zulässig. Sind $c_\varphi(K) \in \mathbb{C}$ beziehungsweise $c_\psi(K) \in \mathbb{C}$ für $K \in E \wedge F$ wie in Bemerkung und Definition 2.1, so ist $\lambda\varphi + \psi$ konstant $= \lambda c_\varphi(K) + c_\psi(K)$ auf $K$. Also ist $\lambda\varphi + \psi$ eine Treppenfunktion. Damit ist $T[a,b]$ ein Teilraum von $\mathrm{Abb}([a,b],\mathbb{C})$. Außerdem gilt

$$\begin{aligned}\int_a^b (\lambda\varphi + \psi) &= \sum_{K\in E\wedge F} (\lambda c_\varphi(K) + c_\psi(K))|K| \\ &= \lambda \sum_{K\in E\wedge F} c_\varphi(K)|K| + \sum_{K\in E\wedge F} c_\psi(K)|K| = \lambda \int_a^b \varphi + \int_a^b \psi \; .\end{aligned}$$

[13] Für $f, g : X \to \mathbb{R}$ schreiben wir kurz $g \leq f$, falls $g(x) \leq f(x)$ für alle $x \in X$ gilt.

Folglich ist $\int_a^b$ linear. Ist $\varphi \geq 0$, so ist $c(I) \geq 0$ für alle $I \in E$ und damit $\int_a^b \varphi \geq 0$. Also ist $\int_a^b$ nichtnegativ und damit auch monoton. Außerdem ist

$$|\varphi| = \sum_{I \in E} |c(I)| 1_I \leq \|\varphi\|_\infty 1_{[a,b]}$$

und insbesondere $|\varphi| \in T[a,b]$. Mit der Dreiecksungleichung ergibt sich die Abschätzung in 1. Die Aussage 2. folgt mit $\varphi = \varphi \cdot 1_{[a,\tau]} + \varphi \cdot 1_{(\tau,b]}$ und $\int_a^\tau \varphi = \int_a^b \varphi \cdot 1_{[a,\tau]}$ sowie $\int_\tau^b \varphi = \int_a^b \varphi \cdot 1_{(\tau,b]}$ aus der Linearität. □

Wir werden nun allgemeinere Funktionen betrachten, die sich in geeigneter Weise durch Treppenfunktionen annähern lassen. Für diese Funktionen können wir dann das Integral über die Integrale der entsprechenden Treppenfunktionen definieren.

**Bemerkung und Definition 2.7** Eine Funktion $f \in B[a,b]$ heißt **Regelfunktion** (auf dem Intervall $[a,b]$), falls eine Folge $(\varphi_n)$ von Treppenfunktionen existiert mit $||f - \varphi_n||_\infty \to 0$ für $n \to \infty$, also $\varphi_n \to f$ gleichmäßig auf $[a,b]$. Wir schreiben $R[a,b]$ für die Menge der Regelfunktionen.

Wir wollen zeigen, dass stetige Funktionen auf $[a,b]$ Regelfunktionen sind. Vorbereitend beweisen wir

**Satz 2.8** *Ist $I$ ein kompaktes Intervall [14] und ist $f \in C(I)$, so ist $f$ gleichmäßig stetig[15].*

**Beweis.** Angenommen, die Behauptung ist falsch. Dann gibt es ein $\varepsilon > 0$ so, dass zu jedem $n \in \mathbb{N}$ Punkte $t_n, s_n \in I$ existieren mit $|f(t_n) - f(s_n)| \geq \varepsilon$ und $|t_n - s_n| < 1/n$. Nach dem Satz von Bolzano-Weierstraß hat die Folge $(t_n)$ eine konvergente Teilfolge $(t_n)_{n \in I}$. Ist $c$ der Grenzwert, so ist $c \in I$. Wegen $|s_n - c| \leq |s_n - t_n| + |t_n - c| < 1/n + |t_n - c|$ konvergiert auch $(s_n)_{n \in I}$ gegen $c$. Da $f$ (folgen-)stetig ist, erhält man die widersprüchliche Aussage

$$\varepsilon \leq |f(t_n) - f(s_n)| \leq |f(t_n) - f(c)| + |f(c) - f(s_n)| \to 0 \quad (n \to \infty,\, n \in I).$$

□

Im Folgenden seien $I_{j,n}$ und $t_{j,n}$ wie in Beispiel 2.2.

**Satz 2.9** *Es seien $\tau_{j,n} \in [t_{j-1,n}, t_{j,n}]$ für $n \in \mathbb{N}$ und $j = 1, \ldots, n$. Ist $f \in C[a,b]$ und ist $\varphi_n \in T[a,b]$ definiert durch $\varphi_n(a) := f(a)$ sowie*

$$\varphi_n(t) := f(\tau_{j,n}) \qquad (t \in I_{j,n},\, j = 1, \ldots, n),$$

*so konvergiert die Folge $(\varphi_n)$ gleichmäßig gegen $f$.*

[14] also von der Form $[a,b]$ mit $a \leq b$

[15] Man nennt $f : I \to \mathbb{C}$ **gleichmäßig stetig**, falls zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $|f(t) - f(s)| < \varepsilon$ für $|t - s| < \delta$.

**Beweis.** Es sei $\varepsilon > 0$ gegeben. Nach Satz 2.8 existiert ein $\delta > 0$ mit $|f(t) - f(s)| < \varepsilon$ für $|t - s| < \delta$. Ist $n > (b-a)/\delta$, so gilt $|I_{j,n}| \leq (b-a)/n < \delta$. Ist nun $t \in [a, b]$, so existiert ein $j \in \{0, \ldots, n\}$ mit $t \in I_{j,n}$ und damit wegen $|t - \tau_{j,n}| < \delta$

$$|f(t) - \varphi_n(t)| = |f(t) - f(\tau_{j,n})| < \varepsilon\,,$$

also $||f - \varphi_n||_\infty \leq \varepsilon$. □

**Bemerkung 2.10** Nach Satz 2.9 sind insbesondere auf $[a, b]$ stetige Funktionen Regelfunktionen. Man kann zeigen ([Ü]), dass auch monotone Funktionen stets Regelfunktionen sind.

**Bemerkung und Definition 2.11** Es seien $f$ eine Regelfunktion und $(\varphi_n)$ eine Folge von Treppenfunktionen mit $\varphi_n \to f$ gleichmäßig auf $[a, b]$. Dann gilt:

1. Die Folge $(\int_a^b \varphi_n)_n$ konvergiert in $\mathbb{C}$, denn für $n, n' \in \mathbb{N}$ gilt nach Satz 2.6

   $$\Big|\int_a^b \varphi_n - \int_a^b \varphi_{n'}\Big| = \Big|\int_a^b (\varphi_n - \varphi_{n'})\Big| \leq ||\varphi_n - \varphi_{n'}||_\infty (b-a)\,,$$

   und da $(\varphi_n)$ eine Cauchyfolge in $B[a, b]$ ist, ist auch $(\int_a^b \varphi_n)$ eine Cauchyfolge in $\mathbb{C}$, also konvergent.

2. Ist $(\psi_n)$ eine weitere Folge von Treppenfunktionen mit $\psi_n \to f$ gleichmäßig auf $[a, b]$, so gilt

   $$\Big|\int_a^b \varphi_n - \int_a^b \psi_n\Big| \leq ||\varphi_n - \psi_n||_\infty (b-a) \leq \big(||\varphi_n - f||_\infty + ||f - \psi_n||_\infty\big)(b-a) \to 0\,,$$

   also $\lim\limits_{n\to\infty} \int_a^b \varphi_n = \lim\limits_{n\to\infty} \int_a^b \psi_n$.

Damit setzen wir

$$\int_a^b f := \int_a^b f(t)\,dt := \lim_{n\to\infty} \int_a^b \varphi_n$$

und nennen $\int_a^b f$ das **Regelintegral** oder auch kurz **Integral** von $f$ auf $[a, b]$. Nach 2. ist dabei der Wert unabhängig von der speziellen Wahl der Treppenfunktionenfolge! Zudem setzen wir noch[16]

$$\int_b^a f := -\int_a^b f.$$

**Beispiel 2.12** Wir betrachten $f(t) = t$ auf $[0, 1]$. Dann ist nach Satz 2.9 mit $\tau_{j,n} = j/n$ durch $\varphi_n(0) = 0$ und $\varphi_n(t) := j/n$ für $t \in I_{j,n}$ und $j = 1, \ldots, n$ eine Folge von Treppenfunktionen auf $[0, 1]$ gegeben, die gleichmäßig auf $[0, 1]$ gegen $f$ konvergiert. Hier gilt

$$\int_0^1 \varphi_n = \sum_{j=1}^n \frac{j}{n} \cdot |I_{j,n}| = \frac{1}{n^2} \sum_{j=1}^n j = \frac{n(n+1)}{2n^2} = \frac{1}{2} + \frac{1}{2n} \to \frac{1}{2} \quad (n \to \infty)\,,$$

also ist $\int_0^1 f = \int_0^1 t\,dt = 1/2$.

[16] Man beachte: nach Definition ist $\int_a^a f = 0$.

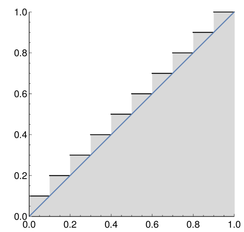


Abbildung 2: Treppenfunktion $\varphi_{10}$ und $\int_0^1 \varphi_{10}$ als Approximation von $\int_0^1 t\,dt$.

Wir stellen einige Rechenregeln für Regelintegrale zusammen, die sich aus der Approximation durch Treppenfunktionen ergeben.

**Satz 2.13** *Die Abbildung* $\int_a^b : R[a,b] \to \mathbb{C}$ *ist linear und monoton. Außerdem gilt für* $f \in R[a,b]$ *und* $\tau \in [a,b]$*:*

1. $|f|$ *ist eine Regelfunktion mit*
$$\Big|\int_a^b f\Big| \le \int_a^b |f| \le (b-a)\sup_{[a,b]} |f|.$$

2. $f|_{[a,\tau]} \in R[a,\tau]$ *und* $f|_{[\tau,b]} \in R[\tau,b]$ *mit* $\int_a^b f = \int_a^\tau f + \int_\tau^b f$.

*Außerdem gilt: Aus* $R[a,b] \ni f_n \to f$ *gleichmäßig auf* $[a,b]$ *für* $n \to \infty$ *folgt* $f \in R[a,b]$ *und* $\int_a^b f_n \to \int_a^b f$ *für* $n \to \infty$.

**Beweis.** Sind $f, g \in R[a,b]$ und $\lambda \in \mathbb{C}$, so existieren Folgen von Treppenfunktionen $(\varphi_n)$ und $(\psi_n)$ mit $\varphi_n \to f$ und $\psi_n \to g$ gleichmäßig auf $[a,b]$ und damit

$$||\lambda f + g - (\lambda\varphi_n + \psi_n)||_\infty \le |\lambda| \cdot ||f - \varphi_n||_\infty + ||g - \psi_n||_\infty \to 0$$

für $n \to \infty$. Also ist $\lambda f + g \in R[a,b]$ und mit Satz 2.6 gilt

$$\int_a^b \lambda f + g = \lim_{n\to\infty} \int_a^b (\lambda\varphi_n + \psi_n) = \lambda \int_a^b f + \int_a^b g\,.$$

Ist $f \ge 0$, so betrachten wir $\varphi_n^+ := \varphi_n + ||f - \varphi_n||_\infty$. Dann gilt $\varphi_n^+ \in T[a,b]$ mit $\varphi_n^+ \ge f \ge 0$ sowie $||f - \varphi_n^+||_\infty \to 0$. Also folgt mit Satz 2.6

$$\int_a^b f = \lim_{n\to\infty} \int_a^b \varphi_n^+ \ge 0\,.$$

Damit ist $\int_a^b : R[a,b] \to \mathbb{C}$ linear und nichtnegativ, also auch monoton.
Aus $\big|\,|f| - |\varphi_n|\,\big| \le |f - \varphi_n|$ folgt, dass auch $|f|$ eine Regelfunktion ist und dass $|\varphi_n| \to |f|$ gleichmäßig auf $[a,b]$ gilt. Mit Satz 2.6.1 ergibt sich

$$\Big|\int_a^b f\Big| = \lim_{n\to\infty} \Big|\int_a^b \varphi_n\Big| \le \lim_{n\to\infty} \int_a^b |\varphi_n| = \int_a^b |f|\,.$$

Wegen der Monotonie von $\int_a^b$ folgt aus $|f| \le \|f\|_\infty$ auch $\int_a^b |f| \le \|f\|_\infty(b-a)$. Also gilt 1. Die Aussage 2. ergibt sich in ähnlicher Weise aus Satz 2.6.2, angewandt auf $\varphi_n$.
Schließlich sieht man damit auch: Gilt $\|f_n \to f\|_\infty \to 0$ für $n \to \infty$ und sind $\varphi_n$ Treppenfunktionen mit $\|f_n - \varphi_n\|_\infty < 1/n$, so gilt $\|f - \varphi_n\|_\infty \to 0$ für $n \to \infty$, also $f \in R[a,b]$, und

$$\Big| \int_a^b f - \int_a^b f_n \Big| \le \|f - f_n\|_\infty (b-a) \to 0 \quad (n \to \infty).$$

□

Wir kommen zu zentralen Sätzen der eindimensionalen Analysis, die die Beziehung zwischen der Differenzial- und der Integralrechnung herstellen.

**Bemerkung und Definition 2.14** Wir setzen für *allgemeine* Intervalle $I$

$$R(I) := \{f : I \to \mathbb{C} : f|_{[a,b]} \in R[a,b] \text{ für alle } [a,b] \subset I\}.$$

Sind $f \in R(I)$ und $u, v, w \in I$, so gilt mit Satz 2.13.2

$$\int_u^w f = \int_u^v f + \int_v^w f,$$

zunächst im Fall $u \le v \le w$, aber wegen der Setzung $\int_b^a = -\int_a^b$ auch allgemein. Ist $u \in I$ fest, so nennen wir die Funktion $Vf = V_u f : I \to \mathbb{C}$, definiert durch

$$(Vf)(x) := \int_u^x f \qquad (x \in I),$$

die **Integralfunktion** von $f$ (bezüglich $u$). Ist $w \in I$, so unterscheiden sich die Funktionen $V_u f$ und $V_w f$ lediglich durch eine additive Konstante (genauer ist $V_u f = V_w f + \int_u^w f$).

**Beispiel 2.15** Sind $I = \mathbb{R}$ und $f = 1_{[0,\infty)}$, so gilt

$$(V_0 f)(x) = \begin{cases} \int_0^x 1\,dt = x, & \text{falls } x \ge 0 \\ \int_0^x 0\,dt = 0, & \text{falls } x < 0 \end{cases},$$

also $V_0 f = \mathrm{id} \cdot 1_{[0,\infty)}$. Man sieht: Anders als $f$ ist $V_0 f$ stetig.

**Satz 2.16 (Hauptsatz über Integralfunktionen)** [17]
*Es seien $I$ ein Intervall, $f \in R(I)$ und $u \in I$. Dann ist die Integralfunktion $Vf = V_u f$ stetig auf $I$.*[18] *Außerdem gilt: Ist $f$ stetig an der Stelle $x \in I$, so ist $Vf$ differenzierbar an $x$ mit*

$$(Vf)'(x) = f(x)\,.$$

[17] wird auch als Hauptsatz der Differenzial- und Integralrechnung, Teil 1, bezeichnet
[18] Die lineare Abbildung $V : R(I) \to C(I)$ nennt man Volterra-Operator auf $R(I)$; daher das $V$.

**Beweis.** Es sei $x \in I$ beliebig. Dann existiert ein $\delta > 0$ so, dass $J := I \cap [x-\delta, x+\delta]$ ein kompaktes Intervall ist. Als Regelfunktion ist damit $f$ beschränkt auf $J$. Für $h \in J - x$ gilt

$$|(Vf)(x+h) - (Vf)(x)| = \Big| \int_u^{x+h} f - \int_u^x f \Big| = \Big| \int_x^{x+h} f \Big| \le \sup_J |f| \cdot |h| \to 0 \quad (h \to 0).$$

Also ist $Vf$ stetig an $x$. Wegen $\int_x^{x+h} f(x)\,dt = f(x)h$ ist zudem für $h \neq 0$ [19]

$$\Big| \frac{(Vf)(x+h) - (Vf)(x)}{h} - f(x) \Big| = \frac{1}{|h|} \Big| \int_x^{x+h} (f(t) - f(x))\,dt \Big| \le \sup_{[x,x+h]} |f - f(x)|.$$

Ist $f$ stetig an $x$, so ist die rechte Seite abklingend für $h \to 0$, und damit auch die linke. □

**Bemerkung und Definition 2.17** Nach Satz 2.16 sind die Integralfunktionen $V_u f$ im Falle einer *stetigen* Funktion $f$ auch Stammfunktionen zu $f$ auf $I$. Insbesondere existieren also in diesem Fall Stammfunktionen. Für (unstetige) Regelfunktionen sind Integralfunktionen im Allgemeinen *keine* Stammfunktionen, wie etwa Beispiel 2.15 zeigt.

Der folgende Satz beinhaltet *das* zentrale Ergebnis zur Berechnung von Integralen.

**Satz 2.18 (Hauptsatz der Differenzial- und Integralrechnung)**
*Es seien $I$ ein Intervall und $f : I \to \mathbb{C}$ stetig. Ist $F$ eine Stammfunktion zu $f$ auf $I$, so ist*

$$\int_u^v f = F(v) - F(u) =: F(t)\Big|_u^v =: F\Big|_u^v$$

*für alle $u, v \in I$.*

**Beweis.** Da $f$ stetig ist, ist nach Bemerkung 2.17 auch $V_u f$ eine Stammfunktion zu $f$ auf $I$. Nach Bemerkung 1.28 ist die Differenz $F - V_u f$ konstant auf $I$. Damit ergibt sich

$$\int_u^v f = (V_u f)(v) = (V_u f)(v) - (V_u f)(u) = F(v) - F(u)$$

für alle $u, v \in I$. □

**Beispiel 2.19** 1. Es sei $f(t) = 1/t$ für $t > 0$. Dann ist $t \mapsto \ln(t)$ eine Stammfunktion zu $f$ auf $(0, \infty)$. Nach dem Hauptsatz der Differenzial- und Integralrechnung gilt für $0 < u, v < \infty$

$$\int_u^v f = \int_u^v \frac{1}{t}\,dt = \ln t\Big|_u^v = \ln(v) - \ln(u)\ .$$

2. Es seien $\alpha \in \mathbb{C} \setminus \{-1\}$ und $f(t) = t^\alpha$ für $t > 0$. Dann ist $t \mapsto \frac{1}{\alpha+1} t^{\alpha+1}$ eine Stammfunktion zu $f$ auf $(0, \infty)$ und folglich ist für $0 < u, v < \infty$

$$\int_u^v t^\alpha\,dt = \frac{1}{\alpha+1} t^{\alpha+1}\Big|_u^v = \frac{1}{\alpha+1}\left(v^{\alpha+1} - u^{\alpha+1}\right)\ .$$

Im Fall $\mathrm{Re}(\alpha) \ge 0$ gilt dies auch für $u = 0$.

[19] Wir schreiben $[u,v] := \{u + t(v-u) : t \in [0,1]\}$ bzw. $(u,v) := \{u + t(v-u) : t \in (0,1)\}$ für die Stecke zwischen $u$ und $v$ (mit bzw. ohne Anfangs- und Endpunkt), ganz allgemein für $u, v$ in einem Vektorraum.

**Bemerkung 2.20** (**partielle Integration**) Sind $f, g : I \to \mathbb{C}$, ist $F$ eine Stammfunktion zu $f$ und ist $g$ differenzierbar auf $I$, so folgt aus der Produktregel, dass $Fg$ eine Stammfunktion zu $fg + Fg'$ auf $I$ ist. Sind $f, g'$ stetig, so ergibt sich mit dem Hauptsatz der Differenzial- und Integralrechnung für $u, v \in I$
$$\int_u^v fg = Fg\big|_u^v - \int_u^v Fg' \,.$$
Man kann also die Berechnung des Integrals $\int_u^v fg$ auf die von $\int_u^v Fg'$ zurückführen.

**Beispiel 2.21** Für $\alpha \neq -1$ und $u, v > 0$ gilt mit partieller Integration, angewandt auf $f(t) = t^\alpha$ und $F(t) = t^{\alpha+1}/(\alpha+1)$
$$\int_u^v t^\alpha \ln t \, dt = \frac{t^{\alpha+1}}{\alpha+1} \ln t\big|_u^v - \frac{1}{\alpha+1}\int_u^v t^\alpha \, dt = \frac{t^{\alpha+1}}{\alpha+1}\big(\ln t - \frac{1}{\alpha+1}\big)\big|_u^v \,.$$
Insbesondere ist
$$\int_u^v \ln t \, dt = \int_u^v 1 \cdot \ln t \, dt = t(\ln(t) - 1)\big|_u^v.$$

**Bemerkung 2.22 (Substitution)** Es seien $I$ ein Intervall und $\gamma : I \to \mathbb{C}$ differenzierbar. Ist $f : \gamma(I) \to \mathbb{C}$ und ist $F$ eine Stammfunktion zu $f$ auf $\gamma(I)$, so ist $F \circ \gamma$ nach der Kettenregel eine Stammfunktion zu $(f \circ \gamma)\gamma\,'$ auf $I$. Sind $f$ und $\gamma'$ stetig, so gilt nach dem Hauptsatz der Differenzial- und Integralrechnung für $u, v \in I$
$$\int_u^v f(\gamma(t))\gamma\,'(t)\, dt = \int_u^v (f \circ \gamma)\gamma\,' = F\big|_{\gamma(u)}^{\gamma(v)}. \tag{2.2}$$
Ist $\gamma$ reellwertig, so ist $\gamma(I)$ nach dem Zwischenwertsatz ein Intervall und wieder mit dem Hauptsatz der Differenzial- und Integralrechnung auch
$$\int_{\gamma(u)}^{\gamma(v)} f = \int_u^v (f \circ \gamma)\gamma\,'.$$

**Beispiel 2.23** Mit $\gamma(t) := 1 + t^2$ auf $\mathbb{R}$ und $f(s) := 1/\sqrt{s}$ auf $(0, \infty)$ sowie $F(s) = 2\sqrt{s}$ gilt nach der Substitutionsregel für $u, v \in \mathbb{R}$
$$\int_u^v \frac{t}{\sqrt{1+t^2}}\, dt = \frac{1}{2}\int_u^v \frac{\gamma'(t)}{\sqrt{\gamma(t)}}\, dt = \sqrt{s}\big|_{1+u^2}^{1+v^2} \,.$$
Sind $a, h \in \mathbb{K}$ und ist $f$ stetig differenzierbar auf $[a, a+h]$, so ist mit $\gamma(t) := a + th$ für $t \in [0,1]$ wegen $\gamma(0) = a$ und $\gamma(1) = a + h$ nach der Substitutionsregel
$$f(a+h) = f(a) + \int_0^1 (f' \circ \gamma)\gamma' = f(a) + h\int_0^1 f'(a+th)\, dt.$$
Ist $X \subset \mathbb{K}$ und ist $f : X \to \mathbb{C}$ analytisch an $a \in X$, so gilt nach Satz 1.26 für $|h|$ genügend klein
$$f(a+h) = \sum_{\nu=0}^\infty \frac{f^{(\nu)}(a)}{\nu!} h^\nu.$$
Der folgende Satz kann als eine Art Brücke zwischen den beiden obigen Aussagen angesehen werden.

**Satz 2.24** (**Taylor**)
*Es seien $a, h \in \mathbb{K}$ und $n \in \mathbb{N}_0$. Ist $f : [a, a+h] \to \mathbb{C}$ eine $(n+1)$-mal stetig differenzierbare Funktion, so gilt*

$$f(a+h) = \sum_{\nu=0}^{n} \frac{f^{(\nu)}(a)}{\nu!} h^\nu + \frac{h^{n+1}}{n!} \int_0^1 (1-t)^n f^{(n+1)}(a+th)\, dt.$$

**Beweis.** Wir beweisen die Behauptung per Induktion nach $n$. Für $n = 0$ ist die Aussage Inhalt der Vorbermerkung. Gilt die Behauptung für $n-1$ und ist $f$ $(n+1)$-mal stetig differenzierbar, so folgt mit partieller Integration

$$\begin{aligned} f(a+h) \quad - \quad \sum_{\nu=0}^{n-1} \frac{f^{(\nu)}(a)}{\nu!} h^\nu &= \frac{h^n}{(n-1)!} \int_0^1 (1-t)^{n-1} f^{(n)}(a+th)\, dt \\ &= \frac{h^n}{(n-1)!} \Big( - \frac{(1-t)^n}{n} f^{(n)}(a+th)\Big|_0^1 + h \int_0^1 \frac{(1-t)^n}{n} f^{(n+1)}(a+th)\, dt \Big) \\ &= \frac{h^n}{n!} f^{(n)}(a) + \frac{h^{n+1}}{n!} \int_0^1 (1-t)^n f^{(n+1)}(a+th)\, dt\ . \end{aligned}$$

□

**Bemerkung und Definition 2.25** In der Situation von Satz 2.24 wird $f^{(k)}(a)/k!$ für $k = 0, \ldots, n+1$ der $k$-te **Taylor-Koeffizient** von $f$ bezüglich $a$ genannt. Das Polynom $T_n(f, a)$, definiert durch

$$T_n(f,a)(h) := \sum_{\nu=0}^{n} \frac{f^{(\nu)}(a)}{\nu!} h^\nu \qquad (h \in \mathbb{C}),$$

heißt das $n$-te **Taylor-Polynom** von $f$ bezüglich $a$ und

$$R_n(f,a)(h) := f(a+h) - T_n(f,a)(h) \qquad (h \in X - a)$$

das $n$-te **Restglied**. Aus Bemerkung und Definition 1.33 ergibt sich: Ist $f : X \to \mathbb{C}$ beliebig oft differenzierbar und $a$ ein innerer Punkt von $X$, so ist $f$ genau dann analytisch an $a$, wenn die Folge $(R_n(f,a))$ auf einer Umgebung von $a$ punktweise gegen $0$ konvergiert.

**Bemerkung 2.26** Sind $g, w : [0,1] \to \mathbb{R}$ stetig und $w \geq 0$, so gilt nach Satz 2.13

$$\min_{[0,1]} g \cdot \int_0^1 w \leq \int_0^1 gw \leq \max_{[0,1]} g \cdot \int_0^1 w$$

Wegen $\int_0^1 (1-t)^n\, dt = 1/(n+1)$ ergibt sich unter den Voraussetzungen des Taylorsatzes mit $g(t) = |f^{(n+1)}(a+th)|$ und $w(t) = (1-t)^n$

$$|R_n(f,a)(h)| \leq \frac{|h|^{n+1}}{(n+1)!} \max_{[a,a+h]} |f^{(n+1)}|.$$

Ist $f$ reellwertig und sind $a, h \in \mathbb{R}$, so kann man genauer zeigen ([Ü]), dass ein $\xi \in [a, a+h]$ (abhängig von $f, a, h, n$) existiert mit

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

Diese Darstellung des Restgliedes nennt man **Lagrange-Form**. Für $n = 0$ ergibt sich die Aussage des Mittelwertsatzes.

**Beispiel 2.27** Für $f = \cos$ gilt $T_3(f, 0)(h) = T_2(f, 0)(h) = 1 - h^2/2$. Wegen $\cos^{(4)} = \cos$ folgt $|\cos(h) - 1 + h^2/2| = |R_3(f, 0)(h)| \leq |h|^4/4!$ für $h \in \mathbb{R}$ mit Bemerkung 2.26.

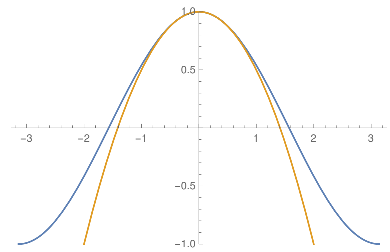


Abbildung 3: cos und $T_3(\cos, 0)$ auf $(-\pi, \pi)$.

Wir haben bisher nur Integrale auf kompakten Intervallen definiert. Wir wollen jetzt beliebige Intervalle betrachten – eine nicht zu unterschätzende Erweiterung.

**Bemerkung und Definition 2.28** Es seien $I$ ein Intervall und $a := \inf I$, $b := \sup I$. Eine Funktion $f \in R(I)$ heißt **integrierbar** auf $I$, falls $(V_u f)(a^+)$ und $(V_u f)(b^-)$ für ein $u \in I$ existieren.[20] In diesem Fall existieren die beiden Grenzwerte für jedes $u \in I$, und die Differenz $(V_u f)(b^-) - (V_u f)(a^+)$ ist nach Bemerkung 2.14 unabhängig von $u$. Die Zahl

$$\int_{a^+}^{b^-} f := \int_{a^+}^{b^-} f(t)\, dt := (V_u f)(b^-) - (V_u f)(a^+)$$

heißt dann **uneigentliches Integral** von $f$ auf $I$. Aus Satz 2.13 folgt leicht, dass durch $f \mapsto \int_{a^+}^{b^-} f$ eine lineare und monotone Abbildung auf dem Raum der auf $I$ integrierbaren Funktionen definiert ist.

**Bemerkung 2.29** 1. Ist $b \in I$, so gilt $(V_u f)(b) = (V_u f)(b^-)$ nach dem Hauptsatz über Integralfunktionen. Entsprechend ist $(V_u f)(a) = (V_u f)(a^+)$ im Falle $a \in I$. Für $I = [a, b]$ ist also

$$\int_{a^+}^{b^-} f = (V_u f)(b^-) - (V_u f)(a^+) = \int_a^b f,$$

das heißt, „eigentliches" und uneigentliches Integral stimmen überein. Daher spricht man auch wieder kurz vom **Integral** von $f$. Ist $a \in I$ oder $a = +\infty$, so schreibt man zudem meist $\int_a^{b^+}$ und entsprechend $\int_{a^+}^b$, falls $b \in I$ oder $b = +\infty$ ist.

2. (**Erweiterter Hauptsatz der Differenzial- und Integralrechnung**) Es sei $f : I \to \mathbb{C}$ stetig. Ist $F$ eine beliebige Stammfunktion zu $f$ auf $I$, so ist $F - V_u f$ konstant auf $I$. Also ist $f$ genau dann integrierbar, wenn $F(b^-)$ und $F(a^+)$ existieren. Außerdem gilt dann

$$\int_{a^+}^{b^-} f = F(b^-) - F(a^+) =: F|_a^b.$$

[20] Dabei ist $g(-\infty^+) := \lim\limits_{t \to -\infty} g(t)$ bzw. $g(+\infty^-) := \lim\limits_{t \to +\infty} g(t)$, im Falle der Existenz.

**Beispiel 2.30** 1. Für $\alpha \in \mathbb{R}$ sei $f_\alpha(t) := t^{-\alpha}$ auf $I = (0, \infty)$. Dann ist durch

$$F_\alpha(t) := \begin{cases} \frac{t^{1-\alpha}}{1-\alpha}, & \alpha \neq 1 \\ \ln(t), & \alpha = 1 \end{cases}$$

eine Stammfunktion $F_\alpha$ zu $f_\alpha$ auf $I$ definiert. Außerdem erhalten wir für $t \to \infty$

$$F_\alpha(t) \to \begin{cases} 0, & \alpha > 1 \\ \infty, & \alpha \leq 1 \end{cases}$$

und für $t \to 0^+$

$$F_\alpha(t) \to \begin{cases} 0, & \alpha < 1 \\ -\infty, & \alpha \geq 1 \end{cases}.$$

Also ist $f_\alpha$ nach Bemerkung 2.29.2 genau dann integrierbar auf $[1, \infty)$, wenn $\alpha > 1$ ist, und in diesem Falle ist

$$\int_1^\infty t^{-\alpha}\, dt = F_\alpha(t)\big|_1^\infty = 0 - \frac{1}{1-\alpha} = \frac{1}{\alpha - 1}\ .$$

Entsprechend ist $f_\alpha$ genau dann integrierbar auf $(0, 1]$, wenn $\alpha < 1$ ist, und dann gilt

$$\int_{0^+}^1 t^{-\alpha}\, dt = F_\alpha(t)\big|_0^1 = \frac{1}{1-\alpha}\ .$$

Hieraus folgt auch, dass $f_\alpha$ für kein $\alpha \in \mathbb{R}$ auf $(0, \infty)$ integrierbar ist.
2. Wir betrachten $f(t) = 1/\sqrt{1-t^2}$ auf $I = (-1, 1)$. Dann gilt $\arcsin' = f$ auf $(-1, 1)$. Da $\arcsin$ stetig auf $[-1, 1]$ ist, ist $f$ nach Bemerkung 2.29.2 integrierbar und es gilt

$$\int_{-1^+}^{1^-} \frac{dt}{\sqrt{1-t^2}} = \arcsin(t)\big|_{-1}^1 = \pi\ .$$

**Satz 2.31** (**Uneigentliche partielle Integration und Substitution**)
*Es seien $I \subset \mathbb{R}$ ein Intervall und $a := \inf I$, $b := \sup I$.*

1. *Sind $f \in C(I)$, $g \in C^1(I)$, ist $F$ eine Stammfunktion zu $f$ auf $I$ und existieren $(Fg)(b^-)$ sowie $(Fg)(a^+)$, so gilt: $fg$ ist genau dann integrierbar auf $I$, wenn $Fg'$ integrierbar auf $I$ ist, und in diesem Fall ist*

$$\int_{a^+}^{b^-} fg = Fg\big|_a^b - \int_{a^+}^{b^-} Fg'.$$

2. *Ist $\gamma : I \to \mathbb{R}$ stetig differenzierbar und monoton, und ist $f \in C(\gamma(I))$ integrierbar auf $\gamma(I)$, so ist $(f \circ \gamma)\gamma'$ integrierbar auf $I$ und mit $c := \inf \gamma(I)$, $d := \sup \gamma(I)$ gilt*

$$\int_{a^+}^{b^-} (f \circ \gamma)\gamma' = \begin{cases} \displaystyle\int_{c^+}^{d^-} f, & \textit{falls } \gamma \textit{ wachsend ist,} \\ \displaystyle -\int_{c^+}^{d^-} f, & \textit{falls } \gamma \textit{ fallend ist.} \end{cases}$$

**Beweis.** 1. $Fg$ ist eine Stammfunktion zur (stetigen) Funktion $fg + Fg'$. Nach Bemerkung 2.29.2 ist $fg + Fg'$ integrierbar mit $\int_{a^+}^{b^-}(fg + Fg') = Fg\big|_a^b$. Ist nun etwa $Fg'$ integrierbar, so ist auch $fg = (fg + Fg') - Fg'$ integrierbar und es gilt

$$\int_{a^+}^{b^-} fg = Fg\big|_a^b - \int_{a^+}^{b^-} Fg'.$$

Entsprechendes gilt, falls $fg$ integrierbar ist.
2. Es sei $u \in I$ fixiert. Dann gilt für $t \in I$ nach der Substitutionsregel

$$(V_u(f \circ \gamma)\gamma')(t) = \int_u^t (f \circ \gamma)\gamma' = \int_{\gamma(u)}^{\gamma(t)} f = V_{\gamma(u)}f(\gamma(t)).$$

Für $t \to b^-$ konvergiert rechte Seite gegen $V_{\gamma(u)}f(d^-)$ falls $\gamma$ wachsend ist, und gegen $V_{\gamma(u)}f(c^+)$ falls $\gamma$ fallend ist. Entsprechend konvergiert die rechte Seite für $t \to a^+$ gegen $V_{\gamma(u)}f(c^+)$ falls $\gamma$ wachsend ist, und gegen $V_{\gamma(u)}f(d^-)$ falls $\gamma$ fallend ist. Damit ist $(f \circ \gamma)\gamma'$ integrierbar auf $I$ und zudem gilt die behauptete Gleichheit. □

**Beispiel 2.32** (Fläche der Einheitskreisscheibe) Wir betrachten die auf $[-1, 1]$ stetigen Funktionen $f(t) := 1$ und $g(t) := \sqrt{1 - t^2}$. Dann existiert das (eigentliche) Integral $\int_{-1}^1 fg$. Mit $F(t) := t$ auf $[-1, 1]$ und $g'(t) = -t/\sqrt{1 - t^2}$ für $t \in (-1, 1)$ ergibt sich nach Satz 2.31 und $t^2 = 1 - (1 - t^2)$

$$\int_{-1}^1 \sqrt{1 - t^2}\,dt = t\sqrt{1 - t^2}\big|_{-1}^1 + \int_{-1^+}^{1^-} \frac{t^2}{\sqrt{1 - t^2}}\,dt = \int_{-1^+}^{1^-} \frac{dt}{\sqrt{1 - t^2}} - \int_{-1}^1 \sqrt{1 - t^2}\,dt$$

und damit

$$2\int_{-1}^1 \sqrt{1 - t^2}\,dt = \int_{-1^+}^{1^-} \frac{dt}{\sqrt{1 - t^2}} = \pi.$$

**Satz 2.33** (**Majorantenkriterium für Integrale**)
*Es sei $I$ ein Intervall und $a := \inf I$, $b := \sup I$. Ist $f \in R(I)$ und ist $g$ integrierbar auf $I$ mit $|f| \le g$, so ist auch $f$ integrierbar auf $I$ mit*

$$\Big|\int_{a^+}^{b^-} f\Big| \le \int_{a^+}^{b^-} g\,.$$

**Beweis.** Es seien $F := V_u f$ und $G := V_u g$ für ein $u \in I$. Sind $s, t \in I$ mit $s < t$, so gilt

$$|F(t) - F(s)| = \Big|\int_s^t f\Big| \le \int_s^t |f| \le \int_s^t g = G(t) - G(s) \tag{2.3}$$

und damit $|F(t) - F(s)| \le |G(t) - G(s)|$ für beliebige $s, t \in I$. Ist nun $(t_n)$ eine Folge in $I$ mit $b > t_n \to b$, so gilt $|F(t_n) - F(t_{n'})| \le |G(t_n) - G(t_{n'})|$ für $n, n' \in \mathbb{N}$. Da $(G(t_n))$ eine Cauchyfolge ist, ist auch $(F(t_n))$ eine Cauchyfolge und damit konvergent. Hieraus ergibt sich die Existenz von $F(b^-)$.[21] Genauso sieht man, dass $F(a^+)$ existiert. Aus (2.3) folgt dann auch $|F(b^-) - F(a^+)| \le G(b^-) - G(a^+)$. □

[21] siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, Bem. 5.25

**Bemerkung und Definition 2.34** Insbesondere ergibt sich aus Satz 2.33 mit $g := |f|$: Ist $f \in R(I)$ (und damit auch $|f|$) und ist $|f|$ integrierbar, so ist auch $f$ integrierbar. Ist $|f|$ integrierbar, so sagen wir $f$ sei **absolut integrierbar**. Wie bei Reihen gilt also: Ist $f$ absolut integrierbar, so ist $f$ integrierbar. Außerdem gilt dann nach Satz 2.33

$$\Big| \int_{a^+}^{b^-} f \Big| \leq \int_{a^+}^{b^-} |f| \, .$$

**Beispiel 2.35** Für $\alpha > 1$ betrachten wir die Funktion $f : [1, \infty) \to \mathbb{R}$ mit

$$f(t) := \cos(t)/t^\alpha \qquad (t \geq 1).$$

Es gilt $|\cos t| t^{-\alpha} \leq t^{-\alpha}$ für $t \geq 1$. Da $t \mapsto t^{-\alpha}$ nach Beispiel 2.30.1 integrierbar ist, folgt die absolute Integrierbarkeit von $f$ aus dem Majorantenkriterium (Satz 2.33). Entsprechendes gilt für die Funktion $t \mapsto \sin(t)/t^\alpha$ auf $[1, \infty)$.

**Bemerkung 2.36** Es sei $f : [1, \infty) \to \mathbb{C}$ stetig. Hat $f$ eine beschränkte Stammfunktion $F$, so ist $t \mapsto f(t)/t$ integrierbar auf $[1, \infty)$.

> Denn: Mit $g(t) = 1/t$ gilt $(Fg)(t) = F(t)/t \to 0$ für $t \to \infty$. Da $F$ beschränkt ist, ist $t \mapsto F(t)t^{-2}$ nach dem Majorantenkriterium integrierbar. Wegen $g'(t) = -1/t^2$ folgt die Behauptung mit Satz 2.31.

Wir betrachten $f = \sin$. Hier ist $F = -\cos$ eine beschränkte Stammfunktion. Also ist $t \mapsto \sin(t)/t$ integrierbar auf $[1, \infty)$. Man kann zeigen, dass die Funktion $t \mapsto |\sin t|/t$ nicht integrierbar auf $[1, \infty)$ ist ([Ü]). Also: Absolute Integrierbarkeit ist eine echt stärkere Eigenschaft als Integrierbarkeit.

Im folgenden Satz wird ein Zusammenhang zwischen der Konvergenz von Reihen und der Existenz uneigentlicher Integrale hergestellt.[22]

**Satz 2.37 (Integralkriterium)**
*Es sei $f : [1, \infty) \to \mathbb{R}$ fallend und $f \geq 0$. Dann ist $0 \leq d_n := \sum\limits_{\nu=1}^{n-1} f(\nu) - \int_1^n f \leq f(1)$ für $n \in \mathbb{N}$ und $(d_n)$ wachsend, also auch konvergent nach dem Hauptsatz über monotone Folgen.*

**Beweis.** Wir setzen $a_k := f(k)$ für $k \in \mathbb{N}$. Aus $a_k \geq f(t) \geq a_{k+1}$ für $t \in [k, k+1]$ folgt $a_k \geq \int_k^{k+1} f \geq a_{k+1}$ und damit

$$0 \leq a_k - \int_k^{k+1} f \leq a_k - a_{k+1}.$$

Also ist

$$d_n = \sum_{\nu=1}^{n-1} a_\nu - \int_1^n f = \sum_{\nu=1}^{n-1} \Big( a_\nu - \int_\nu^{\nu+1} f \Big)$$

wachsend mit $0 \leq d_n \leq \sum\limits_{\nu=1}^{n-1} (a_\nu - a_{\nu+1}) = a_1 - a_n \leq a_1$. □

[22] Man beachte: Monotone Funktionen sind stets in $R(I)$.

**Beispiel 2.38** Es seien $\alpha > 0$ und $f(t) := t^{-\alpha}$ für $t \geq 1$. Dann ist $f$ fallend auf $[1, \infty)$ und $f \geq 0$. Nach Satz 2.37 ist $(d_n)$ mit $d_n := \sum_{\nu=1}^{n-1} \nu^{-\alpha} - \int_1^n f$ konvergent. Im Fall $\alpha > 1$ gilt

$$\lim_{n\to\infty} \int_1^n f = \int_1^\infty t^{-\alpha}\, dt = \frac{1}{\alpha - 1}$$

und damit ergibt sich (wieder) die Konvergenz der Reihe $\sum_{\nu=1}^{\infty} \nu^{-\alpha} = \zeta(\alpha)$. Dabei ist $0 \leq \zeta(\alpha) - 1/(\alpha-1) \leq 1$. Ist $\alpha = 1$, so gilt $\int_1^n t^{-1}\, dt = \ln n$ und damit konvergiert

$$d_n = \sum_{\nu=1}^{n-1} \frac{1}{\nu} - \ln n\, .$$

Der Grenzwert heißt **Euler-Mascheroni Konstante**.[23] Ist $0 < \alpha < 1$, so ergibt sich entsprechend die Konvergenz von

$$d_n = \sum_{\nu=1}^{n-1} \nu^{-\alpha} - \int_1^n t^{-\alpha}\, dt = \sum_{\nu=1}^{n-1} \nu^{-\alpha} - \frac{n^{1-\alpha} - 1}{1 - \alpha}\, .$$

**Satz 2.39** *Die Funktion $t \mapsto e^{-t} t^{z-1}$ ist für alle $z \in \mathbb{C}$ absolut integrierbar auf $[1, \infty)$ und für $\mathrm{Re}(z) > 0$ absolut integrierbar auf $(0, \infty)$.*

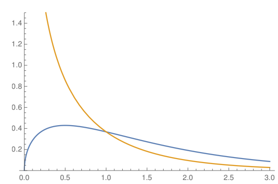


Abbildung 4: $t \mapsto e^{-t} t^{1/2}$ und $t \mapsto e^{-t} t^{-1/2}$.

**Beweis.** Wir setzen $f(t) := e^{-t} t^{z-1}$ für $t > 0$. Aus $t^{\mathrm{Re}(z)+1} e^{-t} \to 0$ für $t \to \infty$ folgt wegen $|t^w| = t^{\mathrm{Re}\, w}$ für $w \in \mathbb{C}$, dass $t \mapsto |e^{-t} t^{z+1}|$ maximal auf $[1, \infty)$ wird.[24] Also existiert eine Konstante $c > 0$ mit $|f(t)| \leq c/t^2$ für alle $t \in [1, \infty)$. Aus dem Majorantenkriterium, angewandt mit $g(t) = c/t^2$, folgt die absolute Integrierbarkeit von $f$ auf $[1, \infty)$.
Für $t \in (0, 1]$ gilt zudem $|f(t)| \leq |t^{z-1}| = t^{\mathrm{Re}(z)-1}$. Ist $\mathrm{Re}(z) > 0$, so ist $t \mapsto t^{\mathrm{Re}(z)-1}$ nach Beispiel 2.30.1 integrierbar auf $(0, 1]$. Wieder mit dem Majorantenkriterium ergibt sich die absolute Integrierbarkeit von $f$ auf $(0, 1]$ und damit auch auf $(0, \infty)$. □

**Bemerkung und Definition 2.40** Es sei $\mathbb{C}_+ := \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\}$ die rechte Halbebene. Die Funktion $\Gamma : \mathbb{C}_+ \to \mathbb{C}$ mit

$$\Gamma(z) := \int_{0+}^{\infty} e^{-t} t^{z-1} dt \qquad (\mathrm{Re}(z) > 0)$$

[23] Zur Bedeutung siehe etwa `https://de.wikipedia.org/wiki/Euler-Mascheroni-Konstante`.
[24] siehe `https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf`, Bem. 6.31

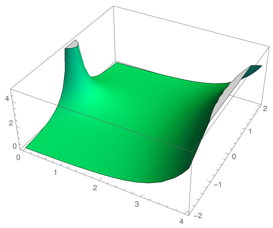

Abbildung 5: $z \mapsto |\Gamma(z)|$.

heißt **(Eulersche) Gammafunktion**. Durch uneigentliche partielle Integration erhält man wegen $t^z \to 0$ für $t \to 0$ und $e^{-t}t^z \to 0$ für $t \to \infty$

$$\int_{0+}^{\infty} e^{-t} t^z \, dt = -e^{-t}t^z \Big|_0^{\infty} - \int_{0+}^{\infty} -e^{-t} z t^{z-1} \, dt = z \int_{0+}^{\infty} e^{-t} t^{z-1} \, dt,$$

also

$$\Gamma(z+1) = z \cdot \Gamma(z) \qquad (\mathrm{Re}(z) > 0) \,. \tag{2.4}$$

Speziell gilt $\Gamma(1) = \int_0^\infty e^{-t} dt = -e^{-t}\big|_0^\infty = 1$, woraus sich wiederum mit (2.4) induktiv

$$\Gamma(n) = (n-1)!$$

für alle $n \in \mathbb{N}$ ergibt. Die Gammafunktion „interpoliert“ also die Fakultäten; man kann die Werte $\Gamma(z)$ als verallgemeinerte Fakultäten auffassen.

# 3 Differenzierbarkeit in normierten Räumen

Wir wollen jetzt Ableitungen für Funktionen zwischen normierten Räumen definieren und untersuchen.

**Bemerkung und Definition 3.1** Es seien $(X, d)$ ein metrischer Raum und $M \subset X$. Man nennt $M$ **offen** (in $X$), falls jeder Punkt $x \in M$ ein innerer Punkt von $M$ ist, und **abgeschlossen** (in $X$), falls $X \setminus M$ offen ist. Man sieht leicht ([Ü]): $M$ ist abgeschlossen genau dann, wenn jeder Häufungspunkt von $M$ in $M$ liegt, d. h. $M' \subset M$ gilt. Für $a \in X$ und $\rho > 0$ schreiben wir

$$\dot{U}_\rho(a) := \dot{U}_{\rho,d}(a) := \{x \in X : 0 < d(x, a) < \rho\}.$$

Mit der Dreiecksungleichung sieht man: Sowohl $U_\rho(a)$ als auch $\dot{U}_\rho(a)$ sind offen in $X$.
Sind $(E, |\cdot|)$ ein normierter Raum, $a \in X'$ und $f : X \to E$, so sagen wir $f$ sei **abklingend** an $a$, falls zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $|f(x)| < \varepsilon$ für alle $x \in \dot{U}_\delta(a)$. Ist $c \in E$, so schreiben wir wie im Fall $E = \mathbb{C}$ wieder $f(x) \to c$ für $x \to a$, falls $f - c$ abklingend an $a$ ist, und nennen (das dadurch eindeutig bestimmte) $c$ den **Grenzwert** von $f$ an $a$. Weiter schreiben wir

$$B_E := \{y \in E : |y| \leq 1\}.$$

für die **Einheitskugel** in $E$. Eine Menge $Y \subset E$ heißt **beschränkt**, falls ein $\rho > 0$ existiert mit $Y \subset \rho B_E$, also $|y| \leq \rho$ für alle $y \in Y$. Ist $X$ eine Menge, so heißt $f : X \to E$ **beschränkt auf** $M \subset X$, falls die Bildmenge $f(M)$ beschränkt ist (vgl. Bemerkung und Definition 1.20).

**Bemerkung und Definition 3.2** Es seien $(V, |\cdot| = |\cdot|_V)$ und $(E, |\cdot| = |\cdot|_E)$ normierte Räume über $\mathbb{K}$. Eine lineare Abbildung $T : V \to E$ heißt **(lokal) beschränkt**, falls $T$ beschränkt auf $B_V$ ist.[25] Wir schreiben $L(V, E)$ für den Raum der lokal beschränkten linearen Abbildungen von $V$ nach $E$. Indem man $T$ mit $T|_{B_V}$ identifiziert, kann man $L(V, E)$ als Unterraum von $B(B_V, E)$ auffassen und dann ist durch[26]

$$\|T\| := \sup_{x \in B_V} |Tx| = \|T|_{B_V}\|_\infty$$

eine Norm auf $L(V, E)$ gegeben. Man nennt $\|\cdot\|$ die **Operatornorm** auf $L(V, E)$.

**Satz 3.3** *Es seien $V, W, E$ normierte Raüme und $T \in L(V, W)$, $S \in L(W, E)$.*

1. *Für $x \in V$ ist $|Tx| \leq \|T\| \cdot |x|$.*
2. *(**Submultiplikativität**) Mit $ST := S \circ T$ gilt $\|ST\| \leq \|S\| \cdot \|T\|$.*

**Beweis.** 1. $T0 = 0$ und für $x \neq 0$ gilt $|Tx| = \left|T\Big(|x| \dfrac{x}{|x|}\Big)\right| = |x| \left|T\Big(\dfrac{x}{|x|}\Big)\right| \leq \|T\| \cdot |x|$.
2. Nach 1. gilt $|STx| \leq \|S\| \cdot |Tx| \leq \|S\| \cdot \|T\|$ für $|x| \leq 1$. □

Im Weiteren sei $V$ stets ein normierter Raum und $E$ ein Banachraum über (dem gleichen) $\mathbb{K}$, wenn nichts anderes gesagt ist.

[25] Man beachte, dass dies keinesfalls bedeutet, dass $T$ beschränkt auf $V$ ist!
[26] Meist schreibt man bei linearen Abbildungen kurz $Tx := T(x)$.

**Bemerkung und Definition 3.4** Es seien $X \subset V$ und $f : X \to E$.
1. Ist $a \in X$ innerer Punkt von $X$, so heißt $f$ **(Fréchet-)differenzierbar** an der Stelle $a$, falls ein $T = T_{f,a} \in L(V, E)$ so existiert, dass

$$\frac{1}{|h|}\big(f(a+h) - f(a) - Th\big) \to 0 \qquad (h \to 0)$$

gilt[27], also mit

$$\tau_a f(h) := f(a+h) - f(a)$$

für $h \in X - a$ die Funktion $|\cdot|^{-1}(\tau_a f - T) : (X - a) \setminus \{0\} \to E$ abklingend an $0$ ist. Man sieht leicht, dass $T$ im Falle der Differenzierbarkeit von $f$ an $a$ eindeutig bestimmt ist, und nennt $f'(a) := T$ die **Fréchetableitung** oder kurz **Ableitung** von $f$ an $a$. Dabei ist zu beachten, dass man im schon betrachteten skalaren Fall $V = \mathbb{K}$ und $E = \mathbb{C}$ (als Vektorraum über $\mathbb{K}$) die Zahl $c \in \mathbb{C}$ mit der linearen Abbildung $T : \mathbb{K} \to \mathbb{C}$, definiert durch $Th := h \cdot c$, identifiziert (also $c = T(1)$).
2. Ist $f$ differenzierbar an $a$, so ist $f$ ist stetig an $a$.

> Denn: Nach Voraussetzung existiert ein $\delta > 0$ mit $|\tau_a f(h) - Th|/|h| \le 1$ für $h \in \dot{U}_\delta(0)$. Wegen $|h|^{-1}|Th| \le \|T\|$ ist nach der Dreiecksungeichung $|h|^{-1}|\tau_a f(h)| \le 1 + ||T||$ und damit $|\cdot|^{-1}\tau_a f$ beschränkt auf $\dot{U}_\delta(0)$. Also ist $\tau_a f$ abklingend an $0$ und damit $f$ stetig an $a$.

Sind $g : X \to E$ differenzierbar an $a$ und $\lambda \in \mathbb{K}$, so ist auch $\lambda f + g$ differenzierbar an $a$ mit $(\lambda f + g)'(a) = \lambda f'(a) + g'(a)$.

> Denn: Die Funktion
>
> $$|\cdot|^{-1}(\tau_a(\lambda f + g) - \lambda f'(a) - g'(a)) = \lambda|\cdot|^{-1}(\tau_a f - f'(a)) + |\cdot|^{-1}(\tau_a g - g'(a))$$
>
> ist abklingend an $0$.

3. Ist $X$ offen und ist $f$ differenzierbar an allen Stellen $a \in X$, so heißt $f$ kurz **differenzierbar** (auf $X$). Die dann definierte Abbildung $f' : X \to L(V, E)$ mit heißt **Ableitung** von $f$ (auf $X$).[28] Ist $f' : X \to L(V, E)$ (wobei $L(V, E)$ mit der Operatornorm versehen ist) stetig, so sagen wir, $f$ sei **stetig differenzierbar**. Im Falle $\mathbb{K} = \mathbb{R}$ schreiben wir $C^1(X, E)$ für die Menge aller stetig differenzierbaren $f : X \to E$ und $C^1(X) := C^1(X, \mathbb{C}) = C^1(X, \mathbb{R}^2)$.

**Bemerkung und Definition 3.5** 1. Ist die Norm $|\cdot|$ auf dem Banachraum $E$ von einem Skalarprodukt $\langle\cdot, \cdot\cdot\rangle$ auf $E$ induziert (d. h. $|\cdot|^2 = \langle\cdot, \cdot\rangle$), so nennt man $E = (E, \langle\cdot, \cdot\cdot\rangle)$ einen **Hilbertraum**. Von zentraler Bedeutung ist die **Cauchy-Schwarzsche Ungleichung** (siehe Lineare Algebra), also

$$|\langle u, v\rangle| \le |u| \cdot |v| \qquad (u, v \in V).$$

Ist $V = \mathbb{R}^d$ und $\langle x, y\rangle := x^\top y$ das kanonische Skalarprdunkt auf $\mathbb{R}^d$, so ist die induzierte Norm die euklidsche.

[27] Man beachte: Hier ist $1/|h| \in \mathbb{R}$ ein Skalar, also hat man es mit einer Skalarmultiplikation zu tun. Eine Division durch $h \in V$ und damit eine Definition über Differenzenquotienten ist hier nicht mehr möglich.

[28] Weitere Schreibweisen sind wieder $Df$ oder $df$ oder auch $df/dx$.

2. Ist $A \in \mathbb{R}^{m\times d}$, so ist durch $Tx := Ax$ eine lineare Abbildung $T : \mathbb{R}^d \to \mathbb{R}^m$ definiert. Ist umgekehrt $T : \mathbb{R}^d \to \mathbb{R}^m$ linear, so gilt $Tx = Ax$ für die Matrix $A = (T\mathbf{e}_1, \ldots, T\mathbf{e}_k) \in \mathbb{R}^{m\times d}$, wobei $\mathbf{e}_k = (\delta_{j,k})_{j=1,\ldots,d}$ den $k$-ten Einheitsvektor in $\mathbb{R}^d$ bezeichnet. In dieser Weise entspricht $\mathbb{R}^{m\times d}$ dem Raum der linearen Abbildungen von $\mathbb{R}^d$ nach $\mathbb{R}^m$. Mit $(a_{1,k}, \ldots, a_{m,k})^\top := T\mathbf{e}_k$ gilt

$$\max_{j,k} |a_{j,k}| \leq \sup_{|x|\leq 1} |Tx| \leq d\sqrt{m} \max_{j,k} |a_{j,k}|$$

([Ü]). Insbesondere ist jede lineare Abbildung lokal beschränkt. Indem man die Matrix $A = (a_{j,k}) \in \mathbb{R}^{m\times d}$ mit $T$ identifiziert, kann man den Raum $\mathbb{R}^{m\times d}$ mit der Operatornorm versehen.

**Beispiel 3.6** 1. (affin-lineare Funktionen) Es seien $T \in L(V, E)$ und $c \in E$. Ist $f : V \to E$ definiert durch $f(x) := Tx+c$, so ist $\tau_x f(h)-Th = 0$ für alle $x, h \in V$ und damit $f'(x)h = Th$ für alle $x, h \in V$, also kurz $f'(x) = T$ für alle $x \in V$.
2. (quadratische Funktionen) Es seien $V$ ein reeller Hilbertraum und $T \in L(V) := L(V, V)$. Ist $f : V \to \mathbb{R}$ definiert durch $f(x) := \langle x, Tx\rangle$, so gilt für $x, h \in V$

$$(\tau_x f)(h) = \langle x + h, T(x + h)\rangle - \langle x, Tx\rangle = \langle x, Th\rangle + \langle h, Tx\rangle + \langle h, Th\rangle.$$

Mit $\varepsilon(h) := \langle h, Th\rangle/|h|$ für $h \neq 0$ gilt $|\varepsilon(h)| \leq |Th| \leq \|T\| \cdot |h|$ nach der Cauchy-Schwarzschen Ungleichung. Also ist $\varepsilon$ abklingend an $0$. Da

$$h \mapsto \langle x, Th\rangle + \langle h, Tx\rangle \in L(V, \mathbb{R})$$

ist, folgt

$$f'(x)h = \langle x, Th\rangle + \langle h, Tx\rangle \qquad (x, h \in V),$$

also kurz $f'(x) = \langle x, T\rangle + \langle \cdot, Tx\rangle$ für $x \in V$. Ist speziell $E = \mathbb{R}^d$ mit dem kanonischen Skalarprodukt und ist $A \in \mathbb{R}^{d\times d}$, so gilt für $f : \mathbb{R}^d \to \mathbb{R}$ mit $f(x) := x^\top Ax$ nach obiger Überlegung

$$f'(x)h = x^\top Ah + h^\top Ax = x^\top (A + A^\top)h \qquad (x, h \in \mathbb{R}^d).$$

Ist $d = 1$ und $A = (a)$, so ergibt sich $f(x) = ax^2$ und $f'(x)h = 2ax \cdot h$.

**Bemerkung und Definition 3.7** Es sei nun $\mathbb{K} = \mathbb{R}$. Einen Vektor $\mathbf{v} \in V \setminus \{0\}$ nennen wir eine **Richtung** in $V$. Für $a \in V$ ist dann $a + \mathbb{R}\mathbf{v} = \{a + t\mathbf{v} : t \in \mathbb{R}\}$ die Gerade durch den Punkt $a$ in Richtung des Vektors $\mathbf{v}$. Ist $X \subset V$ und $f : X \to E$ differenzierbar an $a$, so gilt für $t \in \mathbb{R}^*$ mit $a + t\mathbf{v} \in X$

$$\Big|\frac{\tau_a f(t\mathbf{v})}{t} - f'(a)(\mathbf{v})\Big| = |\mathbf{v}|\frac{1}{|t\mathbf{v}|}|(\tau_a f - f'(a))(t\mathbf{v})| \to 0 \quad (t \to 0)$$

und damit

$$\frac{f(a + t\mathbf{v}) - f(a)}{t} = \frac{\tau_a f(t\mathbf{v})}{t} \to f'(a)(\mathbf{v}) \quad (t \to 0).$$

Allgemein heißt $f$ **richtungsdifferenzierbar** an der Stelle $a \in X$ in Richtung $\mathbf{v}$, falls

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

existiert. In diesem Fall heißt $\partial_{\mathbf{v}} f(a)$ die **Richtungsableitung** oder auch **Gâteaux-Ableitung** von $f$ an der Stelle $a$ in Richtung $\mathbf{v}$. Die obige Überlegung zeigt, dass aus der Differenzierbarkeit von $f$ an $a$ insbesondere die Differenzierbarkeit in alle Richtungen $\mathbf{v}$ folgt, und dass dann

$$\partial_{\mathbf{v}} f(a) = f'(a)\mathbf{v} \tag{3.1}$$

gilt. Ist speziell $V = \mathbb{R}^d$ und $\mathbf{v} = \mathbf{e}_k$ der $k$-te Einheitsvektor, so sagt man auch, $f$ sei **partiell differenzierbar** an $a$ nach der $k$-ten Variablen. Dann schreiben wir auch $\partial_k f(a)$ statt $\partial_{\mathbf{e}_k} f(a)$ und sprechen von der **partiellen Ableitung** von $f$ an $a$ nach der $k$-ten Variablen. Man sieht also: $\partial_k f$ ist die Ableitung der Funktion $x_k \mapsto f(x_1, \dots, x_k, \dots, x_d)$ bei festgehaltenen Variablen $x_1, \dots, x_{k-1}$ und $x_{k+1}, \dots, x_d$.[29] Zudem nennt man

$$\nabla f(a) := \mathrm{grad} f(a) := \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_d f(a) \end{pmatrix} \in E^d$$

den **Gradient** von $f$ an $a$. Die Definition zeigt, dass im Falle und $E = \mathbb{C}$ (als $\mathbb{R}$-Vektorraum) Richtungs- und partielle Ableitungen nichts anderes als Ableitungen von Funktionen einer reellen Variable sind. Folglich stehen damit die Rechenregeln und Ergebnisse der eindimensionalen Differenzialrechnung zur Verfügung.

**Beispiel 3.8** 1. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x,y) = x^2 + y$ für $(x,y) \in \mathbb{R}^2$. Dann gilt für $\mathbf{v} = (u,w)$ und $a = (0,0)$

$$f\big(t(u,w)\big) = t^2u^2 + tw \qquad (t \in \mathbb{R})\ .$$

und damit

$$\partial_{\mathbf{v}} f(0,0) = \left(\frac{\partial f}{\partial \mathbf{v}}(0,0) = D_{\mathbf{v}} f(0,0)\right) = \lim_{t\to 0} \frac{t^2u^2 + tw}{t} = w\ .$$

Für $(x,y) \in \mathbb{R}^2$ sind

$$\partial_1 f(x,y) \left(= \frac{\partial f}{\partial x}(x,y) = f_x(x,y)\right) = 2x$$

und

$$\partial_2 f(x,y) \left(= \frac{\partial f}{\partial y}(x,y) = f_y(x,y)\right) = 1\ .$$

2. Es sei $f : \mathbb{R}^3 \to \mathbb{R}$ definiert durch $f(x,y,z) = xy^2z^3$ für $(x,y,z) \in \mathbb{R}^3$. Dann gilt für $(x,y,z) \in \mathbb{R}^3$

$$\begin{aligned}
\partial_1 f(x,y,z) \quad &\Big(= \tfrac{\partial f}{\partial x}(x,y,z) \quad = \quad f_x(x,y,z)\Big) \quad = \quad y^2z^3 \\
\partial_2 f(x,y,z) \quad &\Big(= \tfrac{\partial f}{\partial y}(x,y,z) \quad = \quad f_y(x,y,z)\Big) \quad = \quad 2xyz^3 \\
\partial_3 f(x,y,z) \quad &\Big(= \tfrac{\partial f}{\partial z}(x,y,z) \quad = \quad f_z(x,y,z)\Big) \quad = \quad 3xy^2z^2\ .
\end{aligned}$$

[29] Im Falle $d = 2$ schreibt man für die Variablen traditionell oft $(x,y)$ statt $(x_1,x_2)$. In diesem Falle spricht man von den partiellen Ableitungen nach $x$ bzw. $y$ und schreibt auch $\partial f/\partial x$ oder $f_x$ sowie $\partial f/\partial y$ oder $f_y$. Entsprechend schreibt man im Falle $d = 3$ oft $(x,y,z)$ statt $(x_1,x_2,x_3)$ und $\partial f/\partial x$, $\partial f/\partial y$ sowie $\partial f/\partial z$ beziehungsweise $f_x, f_y, f_z$.

**Bemerkung und Definition 3.9** Es seien $X \subset \mathbb{R}^d$, $a \in X$ und $f = (f_1, \dots, f_m) : X \to \mathbb{R}^m$. Ist $f$ an $a$ partiell differenzierbar nach allen Variablen, so heißt die Matrix

$$Jf(a) := (\nabla f)^\top(a) = (\partial_1 f(a), \dots, \partial_d f(a)) = \begin{pmatrix} (\nabla f_1)^\top(a) \\ \vdots \\ (\nabla f_m)^\top(a) \end{pmatrix} = (\partial_k f_j(a))_{j,k} \in \mathbb{R}^{m \times d}$$

die **Jacobi-Matrix** von $f$ an $a$. Ist $f$ sogar differenzierbar an $a$, so gilt nach (3.1)

$$\partial_k f(a) = f'(a)\mathbf{e}_k \qquad (k = 1, \dots, d)$$

und damit ist $Jf(a)$ die Matrix, die die lineare Abbildung $f'(a)$ in der Standardbasis darstellt, also $f'(a) = (h \mapsto Jf(a)h)$.[30] Außerdem gilt dann, wieder nach (3.1),

$$\partial_{\mathbf{v}} f(a) = Jf(a)\mathbf{v}$$

für alle Richtungen $\mathbf{v}$.

**Satz 3.10** *Es seien $X \subset \mathbb{R}^d$, $a \in X$ ein innerer Punkt und $f : X \to \mathbb{R}^m$ so, dass $Jf$ existiert. Sind alle $\partial_k f_j : X \to \mathbb{R}$ stetig an $a$, so ist $f$ differenzierbar an $a$ mit $f'(a) = (h \mapsto Jf(a)h)$.*

**Beweis.** Ohne Einschränkung können wir $a = 0$ und $f(0) = 0$ annehmen (der allgemeine Fall ergibt sich mit $\tau_a f$ statt $f$). Weiterhin setzen wir $m = 1$ voraus, also $f$ reellwertig. Den Fall $\mathbb{R}^m$ kann man durch Betrachtung der Komponentenfunktionen $f_1, \dots, f_m$ von $f$ darauf zurückführen.
Ist $\varepsilon > 0$, so existiert ein $\delta > 0$ mit $\left|\partial_k f(x) - \partial_k f(0)\right| < \varepsilon/d$ für alle $x \in U_\delta(0)$ und $k = 1, \dots, d$. Für $0 < |h| < \delta$ setzen wir $h^{(0)} := 0$ und

$$h^{(k)} := h^{(k-1)} + h_k\mathbf{e}_k = (h_1, \dots, h_k, 0, \dots, 0) \qquad (k = 1, \dots, d).$$

Dann ist $h^{(k)} = h$ und wegen $f(0) = 0$ (Teleskopsumme)

$$f(h) = \sum_{k=1}^{d} \big(f(h^{(k)}) - f(h^{(k-1)})\big).$$

Wegen $|h^{(k)}| \le |h| < \delta$ ist $g_k : [0,1] \to \mathbb{R}$ mit $g_k(t) := f(h^{(k-1)} + th_k\mathbf{e}_k)$ für $t \in [0,1]$ nach der Definition partieller Ableitungen differenzierbar mit $g_k'(t) = \partial_k f(h^{(k-1)} + th_k\mathbf{e}_k)h_k$. Nach dem Mittelwertsatz existiert zu jedem $k$ ein $\tau_k \in (0,1)$ mit

$$f(h^{(k)}) - f(h^{(k-1)}) = g_k(1) - g_k(0) = g_k'(\tau_k) = \partial_k f(\xi^{(k)})h_k,$$

wobei $\xi^{(k)} := h^{(k-1)} + \tau_k h_k\mathbf{e}_k$. Also folgt wegen $\xi^{(k)} \in U_\delta(0)$

$$\begin{aligned} |f(h) - Jf(0)h| &\le \sum_{k=1}^{d} \left|f(h^{(k)}) - f(h^{(k-1)}) - \partial_k f(0) \cdot h_k\right| \\ &= \sum_{k=1}^{d} \left|\big(\partial_k f(\xi^{(k)}) - \partial_k f(0)\big)h_k\right| \le \frac{\varepsilon}{d} \sum_{k=1}^{d} |h_k| \le \varepsilon|h| \end{aligned}$$

[30] Im Sinne von Bemerkung 3.5.2 identifiziert man $f'(a)$ auch mit $Jf(a)$.

und damit ist $f$ differenzierbar an $a$ mit Ableitung $h \mapsto \nabla Jf(0)h$. $\square$

**Bemerkung 3.11** Ist $X \subset \mathbb{R}^d$ und ist $g = (g_{j,k}) : X \to \mathbb{R}^{m \times d}$ eine matrixwertige Funktion, so folgt aus Bemerkung 3.5, dass $g$ genau dann stetig an $a$ ist, wenn alle $g_{j,k} : X \to \mathbb{R}$ stetig sind. Mit Satz 3.10 sieht man: Ist $X$ offen, so ist $f : X \to \mathbb{R}^m$ stetig differenzierbar (also $f \in C^1(X, \mathbb{R}^m)$) genau dann wenn alle partiellen Ableitungen $\partial_k f_j$ stetig auf $X$ sind.

**Beispiel 3.12** 1. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x,y) = e^{xy^2}$ für $x, y \in \mathbb{R}$. Dann sind die partiellen Ableitungen

$$\partial_1 f(x,y) = y^2 e^{xy^2} \quad \text{und} \quad \partial_2 f(x,y) = 2xye^{xy^2}$$

stetig auf $\mathbb{R}^2$. Also ist $f$ nach Bemerkung 3.11 (stetig) differenzierbar auf $\mathbb{R}^2$.
2. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

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

Dann gilt für $(x,y) \neq (0,0)$

$$\partial_1 f(x,y) = \frac{y}{x^2+y^2} - \frac{2x^2 y}{(x^2+y^2)^2} = \frac{y}{(x^2+y^2)^2}(y^2 - x^2)$$

und

$$\partial_2 f(x,y) = \frac{x}{x^2+y^2} - \frac{2xy^2}{(x^2+y^2)^2} = \frac{x}{(x^2+y^2)^2}(x^2 - y^2) .$$

Aus $f(t,0) = f(0,t) = 0$ für alle $t \in \mathbb{R}$ folgt $\partial_1 f(0,0) = \partial_2 f(0,0) = 0$. Also existieren die partiellen Ableitungen in allen Punkten $(x,y)$. Die Funktion $f$ ist allerdings nicht stetig an der Stelle $(0,0)$, denn für $t \in \mathbb{R}^*$ gilt

$$f(t,t) = 1/2 \to 1/2 \neq 0 = f(0,0) \qquad (n \to \infty) .$$

Man beachte, dass die partiellen Ableitungen an $a = (0,0)$ nicht stetig sind. Das Beispiel zeigt, dass die Existenz der partiellen Ableitungen auf einer Umgebung von $a$ im Allgemeinen noch nicht die Stetigkeit von $f$ an $a$ impliziert (und damit auch nicht die Differenzierbarkeit).

**Satz 3.13 (Kettenregel)**
*Es seien $X \subset V$ und $f : X \to Y \subset E$ differenzierbar an $a$. Sind $W$ ein Banachraum und $g : Y \to W$ differenzierbar an $f(a)$, so ist $g \circ f$ differenzierbar an $a$ mit $(g \circ f)'(a) = g'(f(a))f'(a)$.*

**Beweis.** Es sei zuächst speziell $a = 0$, $f(0) = 0$ und $= g(0) = 0$. Dann gilt

$$\varepsilon(u) := \frac{1}{|u|}(g(u) - g'(0)u) \to 0 =: \varepsilon(0) \quad (u \to 0)$$

und $|h|^{-1}(f - f'(0))(h) \to 0$ für $h \to 0$. Nach Bemerkung 3.4.2 ist $|\cdot|^{-1} f$ beschränkt auf $\dot{U}_\delta(0)$ für ein $\delta > 0$ und $\varepsilon \circ f$ abklingend an 0. Mit $|g'(0)u| \leq \|g'(0)\| \cdot |u|$ ergibt sich

$$\frac{1}{|h|}((g \circ f)(h) - g'(0)f'(0)h) = \frac{|f(h)|}{|h|}(\varepsilon \circ f)(h) + g'(0)\frac{1}{|h|}(f(h) - f'(0)h) \to 0.$$

für $h \to 0$. Also ist $(g \circ f)'(0) = g'(0)f'(0)$. Der allgemeine Fall ergibt sich wegen $\tau_a(g \circ f) = \tau_{f(a)} g \circ \tau_a f$ aus dem Spezialfall, angewandt auf $\tau_{f(a)} g$ statt $g$ und $\tau_a f$ statt $f$. □

**Bemerkung 3.14** Im Spezialfall $g = d + S$ mit $S \in L(E, W)$ und $d \in W$ ist

$$(g \circ f)'(a) = Sf'(a)$$

nach der Kettenregel und Beispiel 3.6.1. Ist speziell $f = b + T|_X$ mit $T \in L(V, E)$ und $b \in E$, so ist entsprechend

$$(g \circ f)'(a) = g'(f(a))T.$$

Ist dabei $V = \mathbb{K}$ und $T(x) = xc$, so gilt

$$(g \circ f)'(a)h = hg'(f(a))c \qquad (h \in \mathbb{K}).$$

**Satz 3.15** *Es seien $V$ Banachraum, $X \subset V$ offen und $a, b \in V$ mit $(a, b) \subset X$. Weiter sei $f : X \cup [a, b] \to E$ differenzierbar auf $X$ und stetig auf $[a, b]$.*

1. ***(Mittelwertsatz)*** *Ist $E = \mathbb{K}$ und ist $f$ reellwertig, so existiert ein $\xi \in (a, b)$ mit*

$$f(b) - f(a) = f'(\xi)(b - a).$$

2. ***(Schrankensatz)*** *Ist $E$ ein Hilbertraum, so existiert ein $\xi \in (a, b)$ mit*

$$|f(b) - f(a)| \leq |f'(\xi)(b - a)| \, .$$

**Beweis.** Es sei $s : \mathbb{K} \to V$ definiert durch $s(u) := a + u(b - a)$ für $u \in \mathbb{K}$. Nach Bemerkung 3.14 gilt

$$(f \circ s)'(t)h = hf'(s(t))(b - a) \qquad (t \in (0, 1),\ h \in \mathbb{K}).$$

1. Ist $f$ reellwertig, so existiert nach dem skalaren Mittelwertsatz ein $\tau \in (0, 1)$ mit

$$f(b) - f(a) = (f \circ s)(1) - (f \circ s)(0) = (f \circ s)'(\tau)(1 - 0) = f'(s(\tau))(b - a).$$

Also folgt 1. mit $\xi := s(\tau)$.

2. Ohne Einschränkung sei $c := f(b) - f(a) \neq 0$. Wir betrachten $\varphi : E \to \mathbb{K}$ mit $\varphi y := \langle y, c\rangle$. Dann ist $\varphi \in L(E, \mathbb{K})$ mit $\|\varphi\| = |c|$ ([Ü]). Wieder mit Bemerkung 3.14 ist

$$(\varphi \circ f \circ s)'(t)h = h\varphi f'(s(t))(b - a) \quad (t \in (0, 1),\ h \in \mathbb{K}).$$

Nach dem skalaren Schrankensatz existiert ein $\tau \in (0, 1)$ mit

$$|c|^2 = \varphi c = (\varphi \circ f \circ s)(1) - (\varphi \circ f \circ s)(0) \leq |\varphi f'(s(\tau))(b - a)| \leq \|\varphi\| \cdot |f'(s(\tau))(b - a)|.$$

Wegen $||\varphi|| = |c|$ ergibt sich die Behauptung nach Division durch $|c|$ mit $\xi := s(\tau)$. □

**Bemerkung 3.16** Wir wollen kurz auf die geometrische Bedeutung der Ableitung eingehen.[31] Sind $X \subset \mathbb{R}^d$ und $f : X \to \mathbb{R}$ differenzierbar an $a \in X$, so gilt $\partial_{\mathbf{v}} f(a) = (\nabla f)^\top(a)\mathbf{v}$ für alle Richtungen $\mathbf{v}$ nach Bemerkung und Definition 3.9. Ist $|\mathbf{v}| = 1$, so ist nach der Cauchy-Schwarzschen Ungleichung $|(\nabla f)^\top(a)\mathbf{v}| \leq |\nabla f(a)|$, also

$$-|\nabla f(a)| \leq \partial_{\mathbf{v}} f(a) \leq |\nabla f(a)|.$$

Ist dabei $\nabla f(a) \neq 0$, so nennt man $\mathbf{v}^* := \nabla f(a)/|\nabla f(a)|$ die **Gradientenrichtung** von $f$ an $a$. Hier gilt $\partial_{\pm\mathbf{v}^*} f(a) = \pm|\nabla f(a)|$, also

$$\partial_{\mathbf{v}^*} f(a) = \max_{|\mathbf{v}|=1} \partial_{\mathbf{v}} f(a) \quad \text{und} \quad \partial_{-\mathbf{v}^*} f(a) = \min_{|\mathbf{v}|=1} \partial_{\mathbf{v}} f(a).$$

Außerdem gilt für $\mathbf{v} \perp \mathbf{v}^*$

$$\partial_{\mathbf{v}} f(a) = 0\,,$$

d. h. die Richtungsableitungen der zur Gradientenrichtung senkrechten Richtungen verschwinden. Ist $f'$ stetig an $a$, so existiert nach dem Mittelwertsatz und (3.1) für jede Richtung $\mathbf{v}$ mit $|\mathbf{v}| = 1$ und $t$ genügend klein ein $\xi \in (a, a + t\mathbf{v})$ so, dass

$$f(a + t\mathbf{v}) - f(a) = tf'(\xi)\mathbf{v} = t\partial_{\mathbf{v}} f(\xi) \approx t\partial_{\mathbf{v}} f(a)$$

Daher kann man die Gradientenrichtung als *Richtung des steilsten Anstiegs* von $f$ an $a$ und die negative Gradientenrichtung als *Richtung des steilsten Abstiegs* von $f$ an $a$ ansehen.

**Beispiel 3.17** Es sei $f : \mathbb{R}^d \to \mathbb{R}$ mit $f(x) = x^\top x = \sum_{k=1}^{n} x_k^2 = |x|^2$ für $x \in \mathbb{R}^d$. Dann gilt

$$\nabla f(x) = 2x \qquad (x \in \mathbb{R}^d).$$

Für $x \neq 0$ ist $\mathbf{v}^* = x/|x|$ Richtung des steilsten Anstiegs von $f$ an $x$.

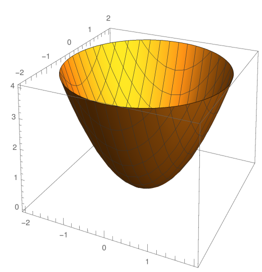

Abbildung 6: $f(x) = f(x_1, x_2) = x_1^2 + x_2^2$ für $|x| \leq 1$ (Rotationsparaboloid)

In der Einführung in die Mathematik hatten wir schon kurz kompakte Mengen in $\mathbb{K}$ betrachtet. Wir definieren nun wesentlich allgemeiner kompakte metrische Räume.

[31] Diese Überlegungen haben insbesondere für die Optimierung große Bedeutung.

**Bemerkung und Definition 3.18** Es sei $(X,d)$ ein metrischer Raum.
1. $(X,d)$ heißt **(folgen-)kompakt**, falls jede Folge in $X$ eine konvergente Teilfolge besitzt. Insbesondere sind kompakte metrische Räume vollständig.
2. Eine Teilmenge $M$ von $X$ heißt **kompakt**, falls $M$ mit der Spurmetrik kompakt ist. Man überlegt sich leicht, dass dies genau dann der Fall ist, wenn jede Folge in $M$ eine konvergente Teilfolge *mit Grenzwert in* $M$ besitzt. Insbesondere sind kompakte Teilmengen von $X$ abgeschlossen in $X$. Ist $(X,d)$ kompakt, so ist umgekehrt auch jede abgeschlossene Teilmenge kompakt ([Ü]).
3. Aus der Definition ergibt sich leicht:[32] Sind $(Y,d_Y)$ ein weiterer metrischer Raum und $f: X \to Y$ stetig, so ist für jede kompakte Menge $M \subset X$ auch die Bildmenge $f(M) \subset Y$ kompakt; kurz: Bilder kompakter Mengen unter stetigen Funktionen sind kompakt. Dies hat die überaus wichtige Konsequenz, dass stetige reellwertige Funktionen auf kompakten Mengen stets maximal und minimal werden, d. h. für kompakte $M$ existieren $\max_{x\in M} f(x) = \max f(M)$ und $\max_{x\in M} f(x) = \max f(M)$. Außerdem ist $f$ geichmäßig stetig auf jeder kompakten Teilmenge $M$, d. h. für jedes $\varepsilon > 0$ existiert ein $\delta > 0$ so, dass $d_Y(f(x), f(x')) < \varepsilon$ für alle $x, x' \in M$ mit $d(x,x') < \delta$. Der Beweis ist modulo Bezeichnungen der gleiche wie zu Satz 2.8.

**Satz 3.19** *Es seien $(X_1,d_1),\dots,(X_m,d_m)$ kompakte metrische Räume. Dann ist auch der metrische Raum $(X_1\times\cdots\times X_m, d_\infty)$ kompakt.*

**Beweis.** Wir beweisen die Behauptung per Induktion nach $m$.
Für $m = 1$ ist nichts zu zeigen.
$m \to m+1$: Wir setzen $X := X_1 \times \cdots \times X_m$ und definieren $p : X \times X_{m+1} \to X$, $q : X \times X_{m+1} \to X_{m+1}$ durch

$$p(x) := (x_1,\dots,x_m), \quad q(x) := x_{m+1} \quad (x = (x_1,\dots,x_m,x_{m+1}) \in X \times X_{m+1}).$$

Ist $(x_n)_{n\in N}$ eine Folge in $X_1\times\cdots\times X_{m+1}$, so ist $(p(x_n))_{n\in N}$ eine Folge in $X$. Also existiert nach Induktionsvoraussetzung eine $d_\infty$-konvergente Teilfolge $(p(x_n))_{n\in I}$. Weiter ist $(q(x_n))_{n\in I}$ eine Folge in $X_{m+1}$. Also existiert eine $d_{m+1}$-konvergente Teilfolge $(q(x_n))_{n\in J}$. Aus

$$d_\infty(x_n,c) = \max\{d_\infty(p(x_n),p(c)), d_{m+1}(q(x_n),q(c))\}$$

für $c \in X \times X_{m+1}$ ergibt sich mit Bemerkung 1.13 die $d_\infty$-Konvergenz von $(x_n)_{n\in J} = (p(x_n),q(x_n))_{n\in J}$. □

**Bemerkung 3.20** Es sei $(V, |\cdot|_V)$ ein normierter Raum. Ist $M \subset V$ kompakt (also kompakt im metrischen Raum $(V, d_{|\cdot|})$), so ist $M$ beschränkt und abgeschlossen ([Ü]).

Für Teilmengen von $\mathbb{R}^d$ ergibt sich aus dem Satz con Bolzano-Weierstraß

**Satz 3.21 (Heine-Borel)**
*$M \subset \mathbb{R}^d$ ist kompakt genau dann, wenn $M$ beschränkt und abgeschlossen ist.*

[32]siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, Satz 6.30

**Beweis.** Ist $M$ kompakt, so ist $M$ nach Bemerkung 3.20 beschränkt und abgeschlossen. Ist umgekehrt $M$ beschränkt und abgeschlossen in $\mathbb{R}^d$, so ist $M \subset [-R, R]^d$ für ein $R > 0$ und $M$ auch abgeschlossen in $[-R, R]^d$. Aus dem Satz von Bolzano-Weierstaß folgt, dass $[-R, R]$ kompakt ist. Nach Satz 3.19 und Bemerkung 1.13 ist $[-R, R]^d$ kompakt und mit Bemerkung 3.18 dann auch $M$. □

**Satz 3.22** (**Parameterintegrale**)
*Es seien $X \subset \mathbb{K}^d$ offen, $[a, b] \subset \mathbb{R}$ und $\varphi : X \times [a, b] \to \mathbb{C}$ stetig. Ist $\Phi : X \to \mathbb{C}$ definiert durch*

$$\Phi(x) := \int_a^b \varphi(x, t)\, dt \qquad (x \in X),$$

*so ist $\Phi$ stetig. Ist zudem $\varphi(\cdot, t)$ differenzierbar für alle $t \in [a, b]$ und $\partial_1\varphi : X \times [a, b] \to L(\mathbb{K}^d, \mathbb{C})$, definiert durch $\partial_1\varphi(x, t) := \varphi(\cdot, t)'(x)$, stetig, so ist $\Phi$ differenzierbar mit*

$$\Phi'(x) = \Big(h \mapsto \int_a^b \partial_1\varphi(x, t)h\, dt\Big) \qquad (x \in X).$$

**Beweis.** Es sei $x \in X$. Wir wählen ein $r > 0$ mit $K := B_r(x) = x + B_r(0) \subset X$. Dann ist $K \times [a, b]$ nach dem Satz von Heine-Borel kompakt (auch in Fall $\mathbb{K} = \mathbb{C}$), also $\varphi$ gleichmäßig stetig auf $K \times [a, b]$. Ist $\varepsilon > 0$ gegeben, so existiert ein $0 < \delta(\leq r)$ so, dass

$$|\varphi(x + h, t) - \varphi(x, t)| < \varepsilon \qquad (|h| < \delta,\ t \in [a, b]).$$

Dann ist für $|h| < \delta$

$$|\Phi(x + h) - \Phi(x)| = \Big|\int_a^b \varphi(x + h, t) - \varphi(x, t)\, dt\Big| \leq \varepsilon(b - a)\,.$$

Also ist $\Phi$ stetig an der Stelle $x$.
Nun sei $\varphi(\cdot, t)$ differenzierbar für alle $t \in [a, b]$ und $\partial_1\varphi$ stetig. Wir setzen für $t \in [a, b]$

$$\psi_t(h) := \varphi(x + h, t) - \partial_1\varphi(x, t)h \qquad (h \in X - x).$$

Nach Beispiel 3.6.1 und der Kettenregel ist $\psi_t$ stetig differenzierbar mit

$$\psi_t'(h) = \partial_1\varphi(x + h, t) - \partial_1\varphi(x, t).$$

Wieder sei $\varepsilon > 0$ gegeben. Da $(u, t) \mapsto \psi_t'(u)$ stetig auf $B_r(0) \times [a, b]$ ist und $\psi_t'(0) = 0$ gilt, existiert wie oben ein $\delta > 0$ so, dass $||\psi_t'(u)|| < \varepsilon$ für $|u| < \delta$ und $t \in [a, b]$. Sind $|h| < \delta$ und $t \in [a, b]$, so existiert nach dem Schrankensatz ein $\xi \in [0, h]$ mit

$$|\varphi(x + h, t) - \varphi(x, t) - \partial_1\varphi(x, t)h| = |\psi_t(h) - \psi_t(0)| \leq \|\psi_t'(\xi)\| \cdot |h| < \varepsilon|h|\,.$$

Da $\|\partial_1\varphi\|$ als stetige Funktion beschränkt auf $K \times [a, b]$ ist, definiert $Th := \int_a^b \partial_1\varphi(x, t)h\, dt$ für $h \in \mathbb{K}^d$ ein $T \in L(\mathbb{K}^d, \mathbb{C})$. Außerdem gilt

$$|\Phi(x + h) - \Phi(x) - Th| = \Big|\int_a^b \big(\varphi(x + h, t) - \varphi(x, t)\big)\, dt - Th\Big| \leq \varepsilon|h|(b - a)$$

für $|h| < \delta$. Da $\varepsilon > 0$ beliebig war, ist $|\cdot|^{-1}(\tau_x\Phi - T)$ abklingend an 0 und damit $\Phi$ differenzierbar an $x$ mit $\Phi'(x) = T$. $\square$

Als Anwendung des Satzes 3.22 beweisen wir

**Satz 3.23** *Es gilt* $\int\limits_0^\infty e^{-t^2}\,dt = \dfrac{\sqrt{\pi}}{2}$.

**Beweis.** Wir betrachten die Funktion $\Phi : \mathbb{R} \to [0, \infty)$, definiert durch

$$\Phi(x) := \int_0^1 e^{-x^2(1+t^2)} \frac{dt}{1+t^2} \qquad (x \in \mathbb{R}).$$

Mit $f(t) := e^{-t^2}$ für $t \in \mathbb{R}$ und $Vf = V_0 f =$ gilt nach Satz 3.22, der Substitutionsregel, der Kettenregel und dem Hauptsatz über Integralfunktionen

$$\Phi'(x)1 = -2e^{-x^2} \int_0^1 xe^{-(xt)^2}\,dt = -2e^{-x^2} \int_0^x e^{-s^2}\,ds = -((Vf)^2)'(x) \quad (x \in \mathbb{R})\,.$$

Also ist $(Vf)^2 + \Phi$ konstant auf $\mathbb{R}$. Wegen $(Vf)^2(0) = 0$ und

$$\Phi(0) = \int_0^1 \frac{dt}{1+t^2} = \arctan(1) = \pi/4$$

ist $(Vf)^2 + \Phi = \pi/4$. Aus $e^{-x^2t^2} \leq 1$ für $t \in [0, 1]$ ergibt sich $0 \leq \Phi(x) \leq e^{-x^2} \to 0$ für $x \to \infty$, also $(Vf)^2(x) \to \pi/4$ für $x \to \infty$ und damit $\int_0^\infty e^{-t^2}\,dt = Vf\big|_0^\infty = \sqrt{\pi}/2$. $\square$

# 4 Ableitungen höherer Ordnung und Taylorformeln

Wir beschäftigen uns nun mit Ableitungen höherer Ordnung. Wieder seien $V$ ein normierter Raum und $E$ ein Banachraum, jetzt jeweils über $\mathbb{R}$.

**Bemerkung und Definition 4.1** Es seien $X \subset V$ offen und $f : X \to E$. Sind $\mathbf{v}$ und $\mathbf{v}_1, \mathbf{v}_2, \ldots$ Richtungen in $V$, so definiert man $\partial^0_{\mathbf{v}} f := f$ und damit induktiv (soweit existent!)

$$\partial_{\mathbf{v}_n} \ldots \partial_{\mathbf{v}_1} f := \partial_{\mathbf{v}_n} \big( \partial_{\mathbf{v}_{n-1}} \ldots \partial_{\mathbf{v}_1} f \big).$$

Für $\mathbf{v}_1 = \ldots = \mathbf{v}_n =: \mathbf{v}$ schreibt man kurz $\partial^n_{\mathbf{v}} f$ und spricht dann von der **Richtungsableitung der Ordnung** $n$ von $f$ in Richtung $\mathbf{v}$. Sind speziell $V = \mathbb{R}^d$ und $\mathbf{v}_1 = \mathbf{e}_{k_1}, \ldots, \mathbf{v}_n = \mathbf{e}_{k_n}$, so schreibt man

$$\partial_{k_n} \ldots \partial_{k_1} f := \partial_{\mathbf{v_n}} \ldots \partial_{\mathbf{v}_1} f$$

und spricht von den **partiellen Ableitungen der Ordnung** $n$. Für $k_1 = \ldots = k_n =: k$ schreibt man kurz $\partial^n_k f$ (und $\partial^0_k f := f$).[33]

**Beispiel 4.2** Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x,y) = x^2 y$ für $x, y \in \mathbb{R}$. Dann gilt

$$\begin{aligned} \partial_1 f(x,y)(= f_x(x,y)) &= 2xy \\ \partial_2 f(x,y)(= f_y(x,y)) &= x^2 \end{aligned}$$

und

$$\begin{aligned} \partial_1^2 f(x,y)(= f_{xx}(x,y)) &= 2y \\ \partial_2\partial_1 f(x,y)(= f_{xy}(x,y)) &= 2x = \partial_1\partial_2 f(x,y) \\ \partial_2^2 f(x,y)(= f_{yy}(x,y)) &= 0\,. \end{aligned}$$

Weiter erhalten wir etwa

$$\partial_1\partial_2\partial_1 f(x,y) = 2 = \partial_2\partial_1^2 f(x,y).$$

In Beispiel 4.2 ist es so, dass die Reihenfolge, in der die partiellen Ableitungen gebildet werden, vertauscht werden kann. Allgemein gilt

**Satz 4.3 (Schwarz)**
*Es seien $X \subset \mathbb{R}^d$, $a \in X$ ein innerer Punkt und $j, k \in \{1, \ldots, d\}$. Weiter sei $f : X \to \mathbb{R}^m$ so, dass $\partial_k f$, $\partial_j f$ und $\partial_k\partial_j f$ existieren. Ist $\partial_k\partial_j f$ stetig an der Stelle $a \in X$, so existiert auch $\partial_j\partial_k f(a)$ und es gilt*

$$\partial_k\partial_j f(a) = \partial_j\partial_k f(a)\,.$$

**Beweis.** Es reicht, die Behauptung für $m = 1$ zu beweisen (der allgemeine Fall ergibt sich durch Anwendung auf die Komponentenfunktionen). Ohne Einschränkung können wir zudem $d = 2$, $j = 1$, $k = 2$ sowie $a = (0,0)$ annehmen.

[33] Andere Schreibweisen sind $\frac{\partial^n f}{\partial x_{k_n} \ldots \partial x_{k_1}}$ oder auch $f_{x_{k_1} \ldots x_{k_n}}$, sowie $\frac{\partial^n f}{\partial x_k^n}$ im Fall $k_1 = \ldots, k_n =: k$.

Zunächst existiert ein $r > 0$ mit $(-r,r)^2 \subset U$. Für $(x,y) \in (-r,r)^2$ sei

$$\Delta(x,y) := f(x,y) - f(x,0) - f(0,y) + f(0,0) = g(x,y) - g(0,y)$$

mit $g(x,y) := f(x,y) - f(x,0)$. Zwei Anwendungen des Mittelwertsatzes zeigen, dass ein $\xi = \xi(x,y) \in (0,x)$ und ein $\eta = \eta(x,y) \in (0,y)$ existieren mit

$$\Delta(x,y) = \partial_1 g(\xi,y) \cdot x = [\partial_1 f(\xi,y) - \partial_1 f(\xi,0)] \cdot x = \partial_2\partial_1 f(\xi,\eta) \cdot x \cdot y\ .$$

Nun sei $\varepsilon > 0$ gegeben. Da $\partial_2\partial_1 f$ stetig an $(0,0)$ ist, existiert ein $0 < \delta \le r$ so, dass für alle $(u,v) \in (-\delta,\delta)^2$

$$|\partial_2\partial_1 f(u,v) - \partial_2\partial_1 f(0,0)| < \varepsilon$$

gilt. Also erhalten wir für $0 < |x|, |y| < \delta$

$$\left|\frac{\Delta(x,y)}{x \cdot y} - \partial_2\partial_1 f(0,0)\right| < \varepsilon\ .$$

Beachtet man, dass bei festem $x$

$$\lim_{y\to 0} \frac{\Delta(x,y)}{xy} = \frac{\partial_2 f(x,0) - \partial_2 f(0,0)}{x}$$

gilt, so erhält man auch

$$\left|\frac{\partial_2 f(x,0) - \partial_2 f(0,0)}{x} - \partial_2\partial_1 f(0,0)\right| \le \varepsilon$$

für alle $x$ mit $0 < |x| < \delta$. Da $\varepsilon > 0$ beliebig war, existiert $\partial_1\partial_2 f(0,0)$ und es gilt $\partial_1\partial_2 f(0,0) = \partial_2\partial_1 f(0,0)$. □

**Bemerkung 4.4** Die Aussage des Satzes 4.3 wird im Allgemeinen falsch, wenn man auf die Voraussetzung der Stetigkeit der partiellen Ableitung $\partial_k\partial_j f$ an $a$ verzichtet. Ist etwa $f : \mathbb{R}^2 \to \mathbb{R}$

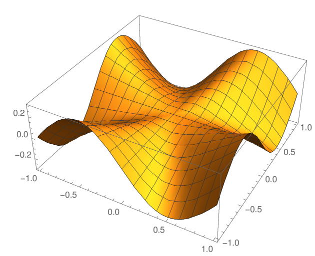


Abbildung 7: $f(x,y)$ für $|x|, |y| \le 1$

definiert durch

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

(vgl. Beispiel 3.12.2), so kann man zeigen ([Ü]): Alle partiellen Ableitungen zweiter Ordnung existieren auf $\mathbb{R}^2$, aber es gilt $\partial_2\partial_1 f(0,0) = 1$ und $\partial_1\partial_2 f(0,0) = -1$.

**Bemerkung und Definition 4.5** Sind $X \subset \mathbb{R}^d$ offen und $n \in \mathbb{N}_0$, so bezeichnen wir mit $C^n(X,E)$ die Menge aller Funktionen $f : X \to E$ mit der Eigenschaft, dass alle partiellen Ableitungen der Ordnung $\leq n$ auf $X$ existieren und dort stetig sind.[34] Außerdem setzen wir $C^\infty(X,E) := \bigcap_{n\in\mathbb{N}} C^n(X,E)$ und $C^n(X) := C^n(X,\mathbb{C}) = C^1(X,\mathbb{R}^2)$. Durch mehrfache Anwendung des Satzes von Schwarz (Satz 4.3) sieht man: Ist $f \in C^n(X,\mathbb{R}^m)$ und sind $k_1,\dots,k_n \in \{1,\dots,d\}$, so gilt für jede Permutation $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$:

$$\partial_{k_n} \dots \partial_{k_1} f = \partial_{k_{\sigma(n)}} \dots \partial_{k_{\sigma(1)}} f \,,$$

d. h. die Reihenfolge, in der die partiellen Ableitungen $\partial_{k_1},\dots,\partial_{k_n}$ gebildet werden, spielt keine Rolle.

Nach (3.1) gilt im Falle $X \subset \mathbb{R}^d$ und $f \in C^1(X,E)$ für alle Richtungen $\mathbf{v} = (v_1,\dots,v_d)$

$$\partial^1_{\mathbf{v}} f(x) = \partial_{\mathbf{v}} f(x) = f'(x)\mathbf{v} = \sum_{k=1}^d v_k f'(x)\mathbf{e}_k = \sum_{k=1}^d v_k \partial_k f(x) \qquad (x \in X)\,.$$

Wir zeigen nun, dass man für beliebiges $n$ die Richtungsableitungen $\partial^n_{\mathbf{v}} f$ als Linearkombination der partiellen Ableitungen der Ordnung $n$ von $f$ ausdrücken kann. Dazu setzen wir für $\alpha = (\alpha_1,\dots,\alpha_d) \in \mathbb{N}_0^d$ und $h = (h_1,\dots,h_d) \in \mathbb{R}^d$

$$\alpha! = \prod_{k=1}^d \alpha_k!\,, \qquad |\alpha|_1 := \sum_{k=1}^d \alpha_k\,, \qquad \partial^\alpha := \partial_1^{\alpha_1}\cdots\partial_d^{\alpha_d}\,, \qquad h^\alpha := \prod_{k=1}^d h_k^{\alpha_k}.$$

**Satz 4.6** *Es seien $X \subset \mathbb{R}^d$ offen und $f \in C^n(X,E)$. Dann ist $\partial^n_{\mathbf{v}} f$ stetig für alle Richtungen $\mathbf{v} = (v_1,\dots,v_d)$ und es gilt*

$$\partial^n_{\mathbf{v}} f = \sum_{(k_1,\dots,k_n)\in\{1,\dots,d\}^n} v_{k_1}\cdots v_{k_n} (\partial_{k_n}\dots\partial_{k_1} f).$$

*Ist $E = \mathbb{R}^d$, so gilt zudem*

$$\frac{1}{n!}\partial^n_{\mathbf{v}} f = \sum_{|\alpha|_1 = n} \frac{1}{\alpha!}\mathbf{v}^\alpha \partial^\alpha f\,.$$

**Beweis.** Wir beweisen die erste Aussage per Induktion nach $n$.
Für $n = 1$ entspricht die Behauptung der Gleichung (3.1) (siehe Vorbemerkung).
$n \to n+1$: Es sei $f \in C^{n+1}(X,\mathbb{R}^m)$ gegeben. Nach Induktionsvoraussetzung ist $\partial^n_{\mathbf{v}} f$ Linearkombination der partiellen Ableitungen der Ordnung $n$ von $f$, also $\partial^n_{\mathbf{v}} f \in C^1(X,\mathbb{R}^m)$. Wieder mit der Induktionsvoraussetzung und mit (3.1) ergibt sich

$$\begin{aligned}
\partial^{n+1}_{\mathbf{v}} f(x) &= \partial_{\mathbf{v}}(\partial^n_{\mathbf{v}} f)(x) = (\partial^n_{\mathbf{v}} f)'(x)\cdot\mathbf{v} = \sum_{k=1}^d v_k \partial_k(\partial^n_{\mathbf{v}} f)(x) \\
&= \sum_{k=1}^d \sum_{k_1,\dots,k_n=1}^d v_{k_1}\cdots v_{k_n}\cdot v_k \partial_k(\partial_{k_n}\dots\partial_{k_1} f)(x) \quad (x \in X).
\end{aligned}$$

[34] Nach Bemerkung 3.11 stimmt dies für $n = 1$ mit der alten Definition überein.

Ist $E = \mathbb{R}^m$, so kann man nach Bemerkung und Definition 4.5 die Reihenfolgen der partiellen Ableitungen beliebig permutieren. Da $\frac{n!}{\alpha_1!\ldots\alpha_d!}$ Tupel $(k_1, \ldots, k_n) \in \{1, \ldots, d\}^n$ existieren, bei denen die Zahl $j \in \{1, \ldots, d\}$ genau $\alpha_j$-mal vorkommt,[35] ergibt sich

$$\partial_{\mathbf{v}}^n f(x) = \sum_{|\alpha|_1=n} \frac{n!}{\alpha_1! \cdots \alpha_d!} v_1^{\alpha_1} \cdots v_d^{\alpha_d} (\partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d} f)(x) = n! \sum_{|\alpha|_1=n} \frac{1}{\alpha!} \mathbf{v}^\alpha \partial^\alpha f(x) \quad (x \in X)\ .$$

□

**Bemerkung und Definition 4.7** Für $f \in C^2(X, E)$ und $x \in X$ ist

$$\partial_{\mathbf{v}}^2 f(x) = \sum_{j,k=1}^d v_j v_k (\partial_k \partial_j f)(x)$$

und im Fall $E = \mathbb{R}$ mit $(Hf)(x) := J\nabla f(x) = \big(\partial_k \partial_j f(x)\big)_{j,k=1,\ldots,d} \in \mathbb{R}^{d\times d}$ auch kurz

$$\partial_{\mathbf{v}}^2 f(x) = \mathbf{v}^\top Hf(x)\mathbf{v}.$$

Die **Hesse-Matrix** $Hf(x)$ von $f$ an der Stelle $x$ ist nach dem Satz von Schwarz (Satz 4.3) symmetrisch[36].

**Bemerkung und Definition 4.8 (Taylorformeln)** Sind $X \subset V$ offen, $f \in C^{n+1}(X)$, $a \in V$ und $\mathbf{v}$ eine Richtung mit $[a, a+\mathbf{v}] \subset X$, so betrachten wir die Funktion $g : [0,1] \to \mathbb{C}$, definiert durch

$$g(t) := f(a + t\mathbf{v}) \qquad (t \in [0,1]).$$

Aus der Definition der Richtungsableitungen ergibt sich induktiv $g^{(k)}(t) = \partial_{\mathbf{v}}^k f(a + t\mathbf{v})$ für $t \in [0,1]$ und $k \le n+1$ ([Ü]) und durch Anwendung des Satzes von Taylor (Satz 2.24) auf die Funktion $g$ damit

$$f(a + t\mathbf{v}) = \sum_{k=0}^n t^k \frac{\partial_{\mathbf{v}}^k f(a)}{k!} + \frac{t^{n+1}}{n!} \int_0^1 (1-s)^n \partial_{\mathbf{v}}^{n+1} f(a + st\mathbf{v})\, ds. \tag{4.1}$$

Im Fall $V = \mathbb{R}^d$ gilt nach Satz 4.6 mit $t\mathbf{v} = h$

$$\sum_{k=0}^n t^k \frac{\partial_{\mathbf{v}}^k f(a)}{k!} = \sum_{k=0}^n \sum_{|\alpha|_1=k} \frac{1}{\alpha!} t^k \mathbf{v}^\alpha \partial^\alpha f(a) = \sum_{0\le|\alpha|_1\le n} h^\alpha \frac{\partial^\alpha f(a)}{\alpha!}.$$

Man nennt $T_n(f,a)$, definiert durch

$$T_n(f,a)(h) := \sum_{0\le|\alpha|_1\le n} h^\alpha \frac{\partial^\alpha f(a)}{\alpha!} \quad (h \in \mathbb{R}^d).$$

[35] Multinomialkoeffizient; siehe etwa https://de.wikipedia.org/wiki/Multinomialkoeffizient.
[36] $A \in \mathbb{R}^{d\times d}$ heißt symmetrisch, falls $A = A^\top$ gilt.

$n$-tes **Taylor-Polynom** von $f$ bezüglich $a$.[37] Insbesondere ist $T_1(f,a)(h) = f(a) + (\nabla f)^\top(a)h$. Ist $f$ reellwertig, so existiert nach Bemerkung 2.26 ein $\xi \in [a, a+h]$ mit

$$\frac{t^{n+1}}{n!}\int_0^1 (1-s)^n \partial_{\mathbf{v}}^{n+1} f(a+st\mathbf{v})\,ds = t^{n+1}\frac{\partial_{\mathbf{v}}^{n+1} f(\xi)}{(n+1)!}$$

(**Lagrangeform des Restgliedes**). Wegen $\partial_{\mathbf{v}}^1 f = (\nabla f)^\top \mathbf{v}$ ergibt sich mit $h = t\mathbf{v}$ im Fall $n = 0$ die Existenz eines $\xi \in [a.a+h]$ mit

$$f(a+h) = f(a) + (\nabla f)^\top(\xi)h,$$

also wieder die Aussage des Mittelwertsatzes, für $n = 1$ die Existenz eines $\xi \in [a, a+h]$ mit

$$f(a+h) = f(a) + (\nabla f)^\top(a)h + \frac{1}{2}h^\top Hf(\xi)h\,. \tag{4.2}$$

und für allgemeines $n$ die eines $\xi \in [a, a+h]$ mit

$$f(a+h) = T_n(f,a)(h) + t^{n+1}\frac{\partial_{\mathbf{v}}^{n+1} f(\xi)}{(n+1)!} = T_n(f,a)(h) + \sum_{|\alpha|_1 = n+1} h^\alpha \frac{\partial^\alpha f(\xi)}{\alpha!}\,. \tag{4.3}$$

Als wesentliche Anwendung der Taylorformel (4.2) werden wir ein hinreichendes Kriterium für lokale Extrema herleiten. Zunächst formulieren wir ein einfaches notwendiges Kriterium.

**Definition 4.9** Sind $X \subset V$ und $f : X \to E$ differenzierbar an $a \in X$, so heißt $a$ **reguläre Stelle**, falls $f'(a)$ surjektiv ist. Ist $f'(a)$ nicht surjekiv, so heißt $a$ **kritisch** (oder **singulär**). Im Falle $E = \mathbb{R}$ ist $a$ genau dann kritisch, wenn $f'(a) = 0$ gilt.

**Satz 4.10** *Es seien $X \subset V$ und $a$ ein innerer Punkt von $X$. Ist $f : X \to \mathbb{R}$ differenzierbar an $a$ und ist $a$ eine Extremstelle von $f$, so ist $a$ ein kritische Stelle.*

**Beweis.** Es seien $\mathbf{v}$ eine Richtung und $I$ ein offenes Intervall mit $0 \in I$ sowie $a + I\mathbf{v} \subset X$. Ist $g : I \to \mathbb{R}$ definiert durch $g(t) := f(a+t\mathbf{v})$ für $t \in I$, so ist $0$ Extremstelle von $g$. Also gilt (siehe Einführung in die Mathematik) $0 = g'(0) = \partial_{\mathbf{v}} f(a) = f'(a)\mathbf{v}$. Da $\mathbf{v}$ beliebig war, ist $f'(a) = 0$. □

**Beispiel 4.11** 1. (vgl. Beispiel 3.17) Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x,y) = x^2 + y^2$$

für $x, y \in \mathbb{R}$. Dann wird $f$ an $(0,0)$ strikt minimal. Es gilt $\nabla f(x,y) = (2x, 2y)$, also tatsächlich $\nabla f(0,0) = 0$.
2. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x,y) = x^2 - y^2$$

für $x, y \in \mathbb{R}$. Dann gilt $\nabla f(x,y) = (2x, -2y)$, also $\nabla f(0,0) = (0,0)$. Allerdings ist $(0,0)$ keine Extremstelle von $f$.

[37] $T_n$ ist ein Polynom in den $d$ Variablen $h_1, \ldots, h_d$ vom Grad $\leq n$.

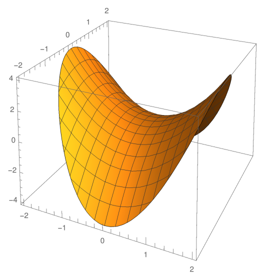


Abbildung 8: $f(x, y) = x^2 - y^2$ für $|x|, |y| \leq 1$

Das zweite Beispiel zeigt, dass auch im Höherdimensionalen an kritischen Stellen im Allgemeinen keine Extremstellen vorliegen. Um auf Extremstellen schließen zu können, bedarf es sogenannter Kriterien zweiter Ordnung, also Kriterien, die die Hesse-Matrix einbeziehen.

**Bemerkung und Definition 4.12** Ist $A \in \mathbb{R}^{d\times d}$ symmetrisch, so heißt $A$

1. **positiv (semi-)definit**, falls $\mathbf{v}^\top A\mathbf{v} > 0$ $(\geq 0)$ für alle Richtungen $\mathbf{v}$ gilt,
2. **negativ (semi-)definit**, falls $-A$ positiv (semi-)definit ist,
3. **indefinit**, falls $A$ weder positiv noch negativ semidefinit ist.

Für symmetrische $A = (a_{j,k})_{j,k=1,2} \in \mathbb{R}^{2\times 2}$ gilt ([Ü]): $A$ ist genau dann positiv (bzw. negativ) definit, wenn $\det(A) > 0$ und $a_{1,1} > 0$ (bzw. $a_{1,1} < 0$) gilt. Entsprechend ist $A$ genau dann positiv (bzw. negativ) semidefinit, wenn $\det(A) \geq 0$ und $a_{1,1}, a_{2,2} \geq 0$ (bzw. $\leq 0$) gilt.

**Bemerkung 4.13** Wir schreiben $\mathbb{S}^{d-1} := \{\mathbf{v} \in \mathbb{R}^d : |\mathbf{v}| = 1\}$ für die $(d-1)$-dimensionale **Einheitssphäre**. Dann ist $\mathbb{S}^{d-1}$ beschränkt und abgeschlossen, also kompakt nach dem Satz von Heine-Borel. Ist $A \in \mathbb{R}^{d\times d}$, so die Funktion $h \mapsto h^\top Ah$ stetig auf $\mathbb{R}^d$ (da sogar differenzierbar nach Beispiel 3.6.2). Damit existieren $\min\limits_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v}$ und $\max\limits_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v}$.
Ist $A$ symmetrisch, und ist $\sigma(A)$ die Menge der Eigenwerte von $A$ (das **Spektrum** von $A$), so ist

$$\min_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v} = \min \sigma(A) =: \lambda_{\min} \qquad \text{und} \qquad \max_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v} = \max \sigma(A) =: \lambda_{\max}$$

(siehe Lineare Algebra). Insbesondere folgt daraus: $A$ ist genau dann positiv (semi-)definit, wenn $\lambda_{\min} > 0$ $(\geq 0)$ gilt und genau dann negativ (semi-)definit wenn $\lambda_{\max} < 0$ $(\leq 0)$ gilt.

Wir kommen damit zu einer weitreichenden Verallgemeinerung des bekannten hinreichenden Kriteriums für Extremstellen über die zweiten Ableitungen aus der eindimensionalen Differenzialrechnung. [38]

**Satz 4.14** *Es seien $X \subset \mathbb{R}^d$ offen, $f \in C^2(X, \mathbb{R})$ und $a \in X$ eine kritische Stelle von $f$. Dann gilt*

[38] siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, Satz 5.17

1. *Ist $Hf(a)$ positiv definit, so ist $a$ eine strikte Minimalstelle.*
2. *Ist $a$ eine Minimalstelle, so ist $Hf(a)$ positiv semidefinit.*

*Entsprechende Aussagen gelten mit negativ definit statt positiv definit und Maximalstelle statt Minimalstelle.*

**Beweis.** 1. Es sei $A := Hf(a)$ positiv definit. Dann ist $c := \min_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v} > 0$ und

$$h^\top Ah = |h|^2(h/|h|)^\top A(h/|h|) \geq c|h|^2 \qquad (h \neq 0).$$

Nach Bemerkung 3.11 ist $Hf : X \to (\mathbb{R}^{d\times d}, \|\cdot\|)$ wegen $f \in C^2(X,\mathbb{R})$ stetig. Daher existiert ein $\delta > 0$ mit $||Hf(x) - A|| < c$ für $x \in U_\delta(a)$, also mit der Cauchy-Schwarzschen Ungleichung

$$|h^\top(Hf(x) - A)h| \leq |h|^2||Hf(x) - A|| < c|h|^2 \qquad (x \in U_\delta(a),\, h \neq 0).$$

Ist $h \in \dot{U}_\delta(a)$, so existiert nach (4.2) wegen $\nabla f(a) = 0$ ein $\xi \in [a, a+h] \subset U_\delta(a)$ mit

$$2(f(a+h) - f(a)) = h^\top Hf(\xi)h = h^\top Ah + h^\top(Hf(\xi) - A)h > c|h^2| - c|h|^2 = 0,$$

also $f(a+h) > f(a)$.
2. Es sei $\mathbf{v}$ eine Richtung. Ist $I$ ein offenes Intervall mit $0 \in I$ und $a + I\mathbf{v} \subset X$, und ist $g : I \to \mathbb{R}$ definiert durch $g(t) := f(a + t\mathbf{v})$ $(t \in I)$, so ist $g \in C^2(I,\mathbb{R})$ nach Bemerkung und Definition 4.8 mit $g''(0) = \partial^2_{\mathbf{v}} f(a) = \mathbf{v}^\top A\mathbf{v}$. Außerdem ist 0 Minimalstelle von $g$ und damit keine strikte Maximalstelle, also $\mathbf{v}^\top A\mathbf{v} = g''(0) \geq 0$[39].
3. Die Aussagen für Maximalstellen ergeben sich durch Betrachtung von $-f$. □

**Beispiel 4.15** 1. Ist $f(x,y) = x^2 + y^2$ für $x, y \in \mathbb{R}$ (vgl. Beispiel 4.11.1), so gilt

$$Hf(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

also ist $Hf$ stets positiv definit. Insbesondere ist $(0,0)$ eine strikte Minimalstelle.
2. Ist $f(x,y) = x^2 - y^2$ für $x, y \in \mathbb{R}$ (vgl. Beispiel 4.11.2) so gilt

$$Hf(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Hier ist $Hf$ stets indefinit. Also hat $f$ nach Satz 4.14 keine lokalen Extrema. Die kritische Stelle $(0,0)$ ist ein sogenannter Sattelpunkt (siehe Abbildung 8)
3. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x,y) = 2x^3 - 3x^2 + 2y^3 + 3y^2$ $(x, y \in \mathbb{R})$. Dann gilt

$$\nabla f(x,y) = (6(x^2 - x)\ ,\ 6(y^2 + y)) = (0,0)$$

genau dann, wenn $x \in \{0,1\}$ und $y \in \{0,-1\}$. Also haben wir die kritischen Stellen

$$(0,0)\ ,\ (0,-1)\ ,\ (1,0)\ ,\ (1,-1)\ .$$

Weiter gilt $Hf(x,y) = \begin{pmatrix} 12x - 6 & 0 \\ 0 & 12y + 6 \end{pmatrix}$. Also ist

[39] siehe https://www.math.uni-trier.de/~mueller/EinfMathe/Einf_Mathe_WS2020-21.pdf, Satz 7.32

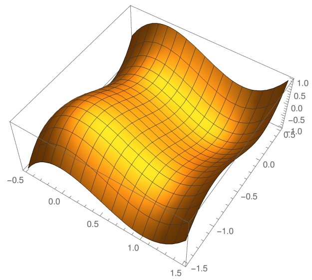


Abbildung 9: $f(x, y)$ für $|x - 1/2|, |y + 1/2| \leq 1$

$Hf(0,0) = \begin{pmatrix} -6 & 0 \\ 0 & 6 \end{pmatrix}$ indefinit,

$Hf(0,-1) = \begin{pmatrix} -6 & 0 \\ 0 & -6 \end{pmatrix}$ negativ definit,

$Hf(1,0) = \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}$ positiv definit und

$Hf(1,-1) = \begin{pmatrix} 6 & 0 \\ 0 & -6 \end{pmatrix}$ indefinit.

Damit sind $(0, -1)$ eine strikte Maximalstelle sowie $(1, 0)$ eine strikte Minimalstelle und ansonsten gibt es keine Extremstellen. In Abbildung 9 kann man die vier kritischen Stellen, von denen zwei Sattelpunkte sind, erkennen.

# 5 Hauptsätze der mehrdimensionalen Analysis

Wir starten wir mit einem allgemeinen Ergebnis über die Existenz von Fixpunkten.

**Bemerkung und Definition 5.1** Es seien $(Y, d)$ ein metrischer Raum, $X \subset Y$ und $\varphi : X \to Y$. Dann heißt $\varphi$ **kontraktiv** falls ein $\lambda < 1$ existiert mit

$$d(\varphi(x), \varphi(x')) \leq \lambda d(x, x') \qquad (x, x' \in X)\ .$$

In diesem Fall nennt man $\lambda$ eine **Kontraktionskonstante** von $\varphi$ und sagt auch $\varphi$ sei $\lambda$-kontraktiv. Insbesondere ist $\varphi$ dann stetig mit höchstens einem Fixpunkt $a$ (ist $a'$ ein weiterer Fixpunkt von $\varphi$ in $M$, so folgt $d(a, a') = d(\varphi(a), \varphi(a')) \leq \lambda d(a, a')$, also $d(a, a') = 0$). Aus dem Schrankensatz (Satz 3.15) ergibt sich folgende wichtige hinreichende Bedingung für Kontraktivität: Sind $V$ ein Hilbertraum, $X \subset V$ strikt konvex[40] sowie $\varphi : X \to V$ stetig auf $X$ und differenzierbar in allen inneren Punkten mit

$$\lambda := \sup\{\|\varphi'(x)\| : x \text{ innerer Punkt von } X\} < 1,$$

so ist $f$ kontraktiv mit Kontraktionskonstante $\lambda$.

Sind $X$ eine nichtleere Menge und $f : X \to X$, so schreiben wir $f^n := f^{\circ n}$ für das $n$-fache Produkt von $f$ bzgl. der Komposition $\circ$, also $f^0 := \mathrm{id_X}$ und $f^n := f \circ f^{n-1}$ für $n \in \mathbb{N}$.

**Satz 5.2 (Banachscher Fixpunktsatz)**
*Es seien $(Y, d)$ ein vollständiger metrischer Raum, $M \subset Y$ abgeschlossen und $\varphi : M \to Y$. Ist $\varphi$ kontraktiv mit $\varphi(M) \subset M$, so existiert ein Fixpunkt $a$ von $\varphi$ und die Folge $(\varphi^n)$ konvergiert punktweise gegen $a$ mit*

$$d(a, \varphi^n(x)) \leq \lambda^n d(a, x) \leq \frac{\lambda^n}{1-\lambda} d(\varphi(x), x) \quad (n \in \mathbb{N},\, x \in M)$$

*für jede Kontraktionskonstante $\lambda$.*

**Beweis.** Es seien $x \in M$ und $x_n := \varphi^n(x)$ für $n \in \mathbb{N}_0$. Dann gilt

$$d(x_{k+1}, x_k) = d(\varphi(x_k), \varphi(x_{k-1})) \leq \lambda \cdot d(x_k, x_{k-1}) \leq \ldots \leq \lambda^k \cdot d(x_1, x_0) \quad (k \in \mathbb{N}_0).$$

Also ist für $n, n' \in \mathbb{N}_0$ mit $n > n'$

$$d(x_n, x_{n'}) \leq \sum_{k=n'}^{n-1} d(x_{k+1}, x_k) \leq d(x_1, x_0) \sum_{k=n'}^{\infty} \lambda^k = \frac{\lambda^{n'}}{1-\lambda} d(x_1, x_0)\ . \tag{5.1}$$

Wegen $\lambda^{n'} \to 0$ für $n' \to \infty$ ist die rechte Seite abklingend und damit $(x_n)$ eine Cauchy-Folge in $M \subset Y$. Aufgrund der Vollständigkeit von $Y$ existiert ein $a \in M \cup M' \subset M$ mit $x_n \to a$ für $n \to \infty$. Mit (5.1) für $n' = 0$ ist $d(x_n, x_0) \leq d(x_1, x_0)/(1-\lambda)$ und durch

[40] $X$ heißt konvex, falls $[a, b] \subset X$ für alle $a, b \in X$ und strikt konvex, falls zusätzlich jedes $x \in (a, b)$ innerer Punkt von $X$ ist.

Grenzübergang $n \to \infty$ damit auch $d(a, x_0) \le d(x_1, x_0)/(1-\lambda)$. Aus der Stetigkeit von $\varphi$ folgt

$$a \leftarrow x_{n+1} = \varphi(x_n) \to \varphi(a) \qquad (n \to \infty),$$

also $a = \varphi(a)$, und schließlich

$$d(a, x_n) = d(\varphi(a), \varphi(x_{n-1})) \le \lambda d(a, x_{n-1}) \le \ldots \le \lambda^n d(a, x_0).$$

□

**Bemerkung und Definition 5.3** Es sei $V$ ein Banachraum und $L(V) := L(V, V)$. Dann ist $L(V)$ mit der Komposition $\circ$ eine Algebra. Aufgrund der Submultiplikativität der Operatornorm sind für $R, S \in L(V)$ die Abbildungen

$$L(V) \ni T \mapsto \begin{cases} RT + S \\ TR + S \end{cases} \tag{5.2}$$

stetig. Durch

$$\mathrm{Aut}(V) := \{T \in L(V) : T \text{ bijektiv und } T^{-1} \in L(V)\}$$

ist eine Gruppe definiert, die **Automorphismengruppe** von $V$. Ist $V$ *endlich-dimensional*, so ist jede lineare Abbildung beschränkt.[41] Also ist in diesem Fall $\mathrm{Aut}(V)$ die Menge der bijektiven linearen Abbildungen von $V$ nach $V$. Außerdem sind dann Surjektivität, Injektivität und Bijektivität äquivalent (Dimensionsformel; siehe Lineare Algebra).

**Satz 5.4** *Es sei $V \neq \{0\}$. Ist $T \in \mathrm{Aut}(V)$ und ist $S \in L(V)$ mit $||S - T|| < 1/||T^{-1}||$, so ist auch $S \in \mathrm{Aut}(V)$. Außerdem ist die Abbildung $\mathrm{Aut}(V) \ni T \mapsto T^{-1} \in \mathrm{Aut}(V)$ stetig.*

**Beweis.** Es seien $I := \mathrm{id}_V$ und $0 < r < 1$. Ist $R \in L(V)$ mit $||R|| \le r/||T^{-1}||$, so gilt $||RT^{-1}|| \le ||R|| \cdot ||T^{-1}|| \le r$. Hieraus ergibt sich $I - RT^{-1} \in \mathrm{Aut}(V)$ und[42]

$$\varphi(R) := \sum_{\nu=0}^{\infty} (RT^{-1})^{\nu} = (I - RT^{-1})^{-1}$$

mit gleichmäßiger Konvergenz auf $\{R : ||R|| \le r/||T^{-1}||\}$ (Neumannsche Reihe; [Ü]). Nach Satz 1.19 ist $\varphi$ stetig an 0 mit $\varphi(0) = I$. Also gilt $T^{-1}\varphi(R) \to T^{-1}I = T^{-1}$ für $||R|| \to 0$. Ist nun $S \in \mathrm{Aut}(V)$ mit $||S - T|| \le 1/||T^{-1}||$ und $R := T - S$, so folgt

$$S = T - R = (I - RT^{-1})T \in \mathrm{Aut}(V)$$

und $S^{-1} = T^{-1}\varphi(R) \to T^{-1}$ für $S \to T$. □

Wir beschäftigen uns nun mit der lokalen Umkehrbarkeit stetig differenzierbarer Funktionen. Im skalaren Fall $X \subset \mathbb{R}$ offen und $f \in C^1(X, \mathbb{R})$ gilt: Ist $a \in X$ mit $f'(a) \neq 0$, so

[41] Dies folgt aus der Äquivalenz von Normen auf endlich-dimensionalen Räumen; siehe Lineare Algebra

[42] Für $n \in \mathbb{N}$ und $T \in L(V)$ ist $T^n$ rekursiv definiert durch $T^n := T^{n-1}T$ mit $T^0 := I$.

existiert ein $\delta > 0$ so, dass $f'$ entweder $> 0$ auf $U := (a-\delta, a+\delta)$ ist oder $< 0$. Also ist $f$ streng monoton auf $U$ und $f_U : U \to f(U)$ mit $f_U(x) := f(x)$ bijektiv. Zudem ist $f(U)$ ein offenes Intervall und nach der Umkehrregel $g := f_U^{-1}$ stetig differenzierbar mit $g'(y) = 1/f'(g(y))$ für $y \in f(U)$, also kurz $g' = 1/(f' \circ g)$.

**Satz 5.5 (Hauptsatz über lokale Umkehrbarkeit)**
*Es seien $V \neq \{0\}$ ein Hilbertraum*[43]*, $X \subset V$ offen, $f : X \to V$ stetig differenzierbar und $a \in X$ mit $f'(a) \in \mathrm{Aut}(V)$. Dann existiert eine offene Umgebung $U$ von $a$ mit folgenden Eigenschaften: $f'(x) \in \mathrm{Aut}(V)$ für $x \in U$, das Bild $f(U)$ ist offen, $f_U : U \to f(U)$ mit $f_U(x) := f(x)$ für $x \in U$ ist bijektiv, und $g := f_U^{-1}$ ist stetig differenzierbar mit*

$$g' = (f' \circ g)^{-1}.$$

**Beweis.** Wir setzen $T := f'(a)$. Für $z \in V$ sei $\varphi = \varphi_z : X \to V$ definiert durch

$$\varphi(x) := x + T^{-1}(z - f(x)) \qquad (x \in X).$$

Dann ist $\varphi_z(x) = x$ genau dann, wenn $f(x) = z$ gilt. Weiter ergibt sich mit Beispiel 3.6 und der Kettenregel

$$\varphi'(x) = \mathrm{id}_V - T^{-1}f'(x) = T^{-1}(T - f'(x)) \qquad (x \in X)$$

Da $f'$ stetig ist, existiert ein $\delta > 0$ so, dass mit $U := U_\delta(a)$

$$||f'(x) - T|| < c := \frac{1}{2||T^{-1}||} \qquad (x \in U)$$

gilt. Nach Satz 5.4 ist dann $f'(x) \in \mathrm{Aut}(V)$ für $x \in U$. Wegen

$$||\varphi'(x)|| \leq ||T^{-1}|| \cdot ||T - f'(x)|| < 1/2$$

für $x \in U$ ist $\varphi|_U$ 1/2-kontraktiv nach Bemerkung und Definition 5.1 mit höchstens einem Fixpunkt in $U$. Folglich ist $f|_U$ injektiv, d. h. die Umkehrfunktion $g : f(U) \to U$ von $f_U$ existiert. Wir zeigen:

1. $f(U)$ ist offen.
2. $g \in C^1(f(U), V)$ mit $g' = (f' \circ g)^{-1}$.

Dazu seien $x \in U$ und $y := f(x)$.
1. Da $U$ offen ist, existiert ein $\rho > 0$ mit $M := B_\rho(x) \subset U$. Ist $z \in U_{c\rho}(y)$ und $\varphi = \varphi_z$, so gilt für $u \in M$

$$\begin{aligned}|\varphi(u) - x| &\leq |\varphi(u) - \varphi(x)| + |\varphi(x) - x|\\ &\leq |u - x|/2 + |T^{-1}(z - f(x))|\\ &\leq |u - x|/2 + ||T^{-1}|| \cdot |z - y| \leq \rho/2 + ||T||^{-1}c\rho = \rho\ ,\end{aligned}$$

[43] Wir setzen hier und im folgenden Hauptsatz über implizite Funktionen *Hilbert* statt *Banach* nur deshalb voraus, um den Schrankensatzes zur Verfügung haben. Tatsächlich ist die Aussage des Schrankensatzes, wie man mit etwas mehr Aufwand oder mithilfe des Satzes von Hahn-Banach zeigen kann, für Banachräume erfüllt, und damit gilt dies auch für die beiden Hauptsätze.

also $\varphi(u) \in M$ und damit $\varphi(M) \subset M$. Aus dem Banachschen Fixpunktsatz folgt die Existenz eines $u \in M$ mit $\varphi(u) = u$, also $f(u) = z$. Damit ist $U_{c\rho}(y) \subset f(U)$.
2. Wir zeigen: $g'(y) = (f'(x))^{-1} = (f'(g(y))^{-1}$. Dabei können wir ohne Einschränkung $x = y = 0 (= f(0) = g(0))$ annehmen. Der allgemeine Fall ergibt sich dann wieder durch Anwendung des Spezialfalles auf $\tau_x f$.
Ist $\varphi = \varphi_0$, also $\varphi(u) = u - T^{-1}f(u)$, so ist $\varphi(0) = 0$ und $\varphi$ eine 1/2-Kontraktion. Nach der umgekehrten Dreiecksungleichung ist

$$|u|/2 \geq |\varphi(u)| \geq |u| - \|T^{-1}\| \cdot |f(u)|,$$

für $u \in U$ und folglich $c|u| \leq |f(u)|$. Also gilt $|g(h)| \leq |h|/c$ für $h \in f(U)$ und insbesondere $g(h) \to 0 = g(0)$ für $h \to 0$. Aus der Differenzierbarkeit von $f$ an 0 folgt mit $S := f'(0)$

$$\frac{c}{|h|}|g(h) - S^{-1}(h)| \leq \left\|S^{-1}\right\| \cdot \frac{|Sg(h) - h|}{|g(h)|} = \|S^{-1}\| \cdot \frac{\left|f(g(h)) - Sg(h)\right|}{|g(h)|} \to 0$$

für $h \to 0$. Also ist auch $g$ differenzierbar an 0 mit $g'(0) = S^{-1} = f'(g(0))^{-1}$. Da $g$ stetig an 0 ist und $f'$ nach Voraussetzung sowie $T \mapsto T^{-1}$ nach Satz 5.4 stetig sind, ist $g'$ stetig an 0. $\square$

**Bemerkung 5.6** Nach Bemerkung 5.3 ist für endlich-dimensionale $V$ schon dann $f'(a) \in \mathrm{Aut}(V)$, wenn $f'(a)$ surjektiv (also $a$ regulär) ist, und für $V = \mathbb{R}^d$ ist $f'(a)$ genau dann surjektiv, wenn $\det Jf(a) \neq 0$ gilt.

**Bemerkung und Definition 5.7** Es seien $X \subset V$ offen, $f : X \to V$ und $a \in X$. Existiert eine offene Umgebung $U$ von $a$ so, dass $f_U$ bijektiv ist, so sagt man $f$ sei **lokal umkehrbar** and $a$. Ist $f$ stetig differenzierbar und $f'(a) \in \mathrm{Aut}(V)$, so ist $f$ nach dem Hauptsatz über Umkehrfunktionen lokal umkehrbar an $a$ und zudem $f(U)$ offen mit stetig differenzierbarer lokaler Umkehrfunktion $g = (f_U)^{-1}$. Gilt dies, so sagen wir im Fall $\mathbb{K} = \mathbb{R}$ auch kurz, $f$ sei **lokal $C^1$-umkehrbar**, und im Fall $V = \mathbb{C}$ als Vektorraum über $\mathbb{C}$ kurz $f$ sei lokal holomorph umkehrbar.[44]

**Beispiel 5.8** (Polarkoordinaten) Wir betrachten $f : (0, \infty) \times \mathbb{R} \to \mathbb{R}^2$ mit

$$f(r, \theta) := \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} \qquad (r > 0,\, \theta \in \mathbb{R})\,.$$

Dann ist $f$ stetig differenzierbar mit $\det Jf(r, \theta) = r > 0$. Also ist $f$ nach Satz 5.5 lokal $C^1$-umkehrbar an allen Stellen $(r, \theta)$. Da etwa $f(r, 2k\pi) = f(r, 0)$ für alle $k \in \mathbb{Z}$ gilt, ist $f$ jedoch nicht injektiv.

In enger Beziehung zum Hauptsatz über Umkehrfunktionen steht ein weiterer Hauptsatz: der über implizite Funktionen. Worum geht es dabei? Sind $V, E$ Hilberträume über dem gleichen $\mathbb{K}$, so ist durch $\langle (x, y), (u, w) \rangle := \langle x, u \rangle + \langle u, w \rangle$ ein Skalarprodukt auf $V \times E$

[44] Sind $X \subset \mathbb{C}$ offen und $f : X \to \mathbb{C}$ stetig komplex differenzierbar, so nennt man $f$ holomorph.

gegeben, mit dem auch $V \times E$ zu einem Hibertraum wird.[45] Sind $\Omega \subset V \times E$ und $F : \Omega \to E$, so betrachten wir die Gleichung

$$F(x,y) = 0.$$

Ist $(a,b) \in \Omega$ eine Lösung, so stellt sich die Frage ist, ob die Gleichung lokal, also in einer Umgebung $W$ von $(a,b)$, nach $y$ auflösbar ist, also die Frage, ob die Lösungsmenge lokal der Graph einer Funktion $f$ der Variablen $x$ ist. Man sieht schnell, dass die Frage eng mit der Regularuität der Ableitung von $F$ nach $y$ verknüpft ist. Für $\Omega$ offen und stetig differenzierbares $F$ schreiben wir kurz

$$D_1F(x,y) := (F(\cdot,y))'(x) \quad \text{und} \quad D_2F(x,y) := (F(x,\cdot))'(y).$$

Aus der Definition der Differenzierbarkeit folgt mit $(h,k) = (h,0) + (0,k)$

$$F'(x,y)(h,k) = D_1F(x,y)h + D_2F(x,y)k \qquad ((h,k) \in V \times E).$$

**Beispiel 5.9** 1. (Lineare Gleichungen) Es seien $A \in \mathbb{R}^{m\times d}$ , $B \in \mathbb{R}^{m\times m}$. Wir betrachten das unterbestimmte lineare Gleichungssystem

$$F(x,y) := (A,B)\begin{pmatrix} x \\ y \end{pmatrix} = Ax + By = 0.$$

Ist $\det B \neq 0$, also $D_2F(x,y) = (k \mapsto Bk) \in \mathrm{Aut}(\mathbb{R}^m)$, so gilt $F(x,y) = 0$ genau dann, wenn $y = -B^{-1}Ax$, d. h.

$$\{(x,y) : F(x,y) = 0\} = \{(x, -B^{-1}Ax) : x \in \mathbb{R}^d\}.$$

2. (Einheitskreis) Wir betrachten die Kreisgleichung

$$F(x,y) := x^2 + y^2 - 1 = 0 \qquad (x,y \in \mathbb{R}) .$$

Für $(a,b)$ mit $a^2 + b^2 = 1$ sowie $b > 0$ und die Umgebung $W := (-1,1) \times (0,\infty)$ von $(a,b)$ gilt

$$\{(x,y) \in W : F(x,y) = 0\} = \{(x, \sqrt{1-x^2}) : x \in (-1,1)\}$$

Entsprechend ist $\{(x,y) \in (-1,1)\times(-\infty,0) : F(x,y) = 0\} = \{(x, -\sqrt{1-x^2}) : x \in (-1,1)\}$. Für die Lösungspunkte $(\pm 1, 0)$ existiert keine Umgebung $W$ auf der die Lösungsmenge der Graph einer Funktion von $x$ ist.

**Satz 5.10 (Hauptsatz über implizite Funktionen)**
*Es seien $\Omega \subset V \times E$ offen und $F : \Omega \to E$ stetig differenzierbar. Weiter sei $(a,b) \in \Omega$ mit $F(a,b) = 0$ und $D_2F(a,b) \in \mathrm{Aut}(E)$. Dann existieren offene Umgebungen $U$ von $a$ und $W$ von $(a,b)$ sowie eine stetig differenzierbare Funktion $f : U \to E$ so, dass*

$$\{(x,y) \in W : F(x,y) = 0\} = \{(x, f(x)) : x \in U\}.$$

*Dabei gilt*

$$f'(x) = -(D_2F(x,f(x))^{-1}D_1F(x,f(x)) \qquad (x \in U) .$$

[45] Addition und Skalarmultiplikation sind natürlich komponentenweise definiert. Man schreibt für den Produktraum dann auch $V \oplus E$.

**Beweis.** Es seien $p : V \times E \to V$ definiert durch $p(x, y) := x$ und $G := \binom{p}{F} : \Omega \to V \times E$, also

$$G(x, y) = \begin{pmatrix} x \\ F(x, y) \end{pmatrix} \qquad ((x, y) \in \Omega) .$$

Dann gilt $G(a, b) = (a, 0)$ und $G \in C^1(\Omega, V \times E)$. Weiterhin ist wegen $p'(x, y) = p$

$$G'(x, y)(h, k) = \begin{pmatrix} p'(x, y) \\ F'(x, y) \end{pmatrix}(h, k) = \begin{pmatrix} h \\ D_1F(x, y)h + D_2F(x, y)k \end{pmatrix}$$

Aus $D_2F(a, b) \in \mathrm{Aut}(E)$ ergibt sich damit leicht, dass $G'(a, b)$ bijektiv und $(G')^{-1}(a, b)$ lokal beschränkt ist, also $G'(a, b) \in \mathrm{Aut}(V \times E)$ gilt. Nach Satz 5.5 existieren offene Umgebungen $R$ von $(a, 0)$ und $W$ von $(a, b)$ so, dass $G_W : W \to R$ bijektiv ist mit $(G_W)'(x, y) \in \mathrm{Aut}(V \times E)$ für $(x, y) \in W$ und stetig differenzierbarer Umkehrfunktion. Ohne Einschränkung können wir (nach Verkleinerung) $R = U_\delta(a) \times U_\delta(0)$ annehmen und schreiben dann wieder kurz $G$ statt $G_W$. Aus der Definition von $G$ folgt, dass $G^{-1}$ von der Form

$$G^{-1}(x, z) = \begin{pmatrix} x \\ H(x, z) \end{pmatrix} \qquad ((x, z) \in R)$$

mit einer geeigneten stetig differenzierbaren Funktion $H : R \to E$ ist. Definiert man $q : V \times E \to E$ durch $q(u, z) := z$, so gilt

$$F \circ G^{-1} = q \circ G \circ G^{-1} = q|_R.$$

Ist $(x, y) \in W$, so ist $x \in U := U_\delta(a)$ und $y = H(x, z)$ für ein $z \in U_\delta(0)$ mit

$$F(x, y) = F(x, H(x, z)) = F(x, q(G^{-1}(x, z))) = (F \circ G^{-1})(x, z) = z.$$

Also ist $F(x, y) = 0$ genau dann wenn $y = H(x, 0) =: f(x)$. Mit $H$ ist auch $f$ stetig differenzierbar und wegen $\varphi(x) := F(x, f(x)) = 0$ für $x \in U$ folgt aus der Kettenregel

$$0 = \varphi'(x) = F'(x, f(x))\begin{pmatrix} \mathrm{id}_V \\ f'(x) \end{pmatrix} = D_1F(x, f(x)) + D_2F(x, f(x))f'(x).$$

Die Bijektivität von $G'(x, y)$ für alle $(x, y) \in W$ impliziert auch die von $D_2F(x, y)$. Also ergibt sich die Zusatzbehauptung durch Auflösen der letzten Gleichung nach $f'(x)$. □

**Beispiel 5.11** (Lemniskate) Es sei $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ definiert durch

$$F(x, y) = (x^2 + y^2)^2 - 2(x^2 - y^2) \qquad (x, y \in \mathbb{R}) .$$

Dann gilt $0 = \partial_2 F(x, y)$ und $0 = F(x, y)$ genau dann, wenn

$$(x, y) = (0, 0) \quad \text{oder} \quad (x, y) = (\pm\sqrt{2}, 0)$$

([Ü]). Nach Satz 5.10 ist für alle $(a, b)$ mit $F(a, b) = 0$ und $(a, b) \notin \{(0, 0), (\pm\sqrt{2}, 0)\}$ die Gleichung $F(x, y) = 0$ auf einer Umgebung $W$ von $(a, b)$ auflösbar nach $y$. Außerdem ergibt sich für die Funktion $f = f_{(a,b)}$

$$f'(x) = -\frac{\partial_1 F(x, f(x))}{\partial_2 F(x, f(x))} = -\frac{x(x^2 + f^2(x) - 1)}{f(x)(x^2 + f^2(x) + 1)}$$

auf einer Umgebung $U$ von $a$. Damit hat $f$ Extremstellen höchstens in Punkten $x$ mit

$$x^2 + f^2(x) = 1 ,$$

also in den Schnittpunkten mit dem Einheitskreis. Um eine Vorstellung von der Lösungsmenge zu bekommen, betrachten wir Polarkoordinaten: Für $(x, y) \neq (0, 0)$ und $(x, y) = (r \cos\theta, r \sin\theta)$ gilt

$$0 = F(x, y) = F(r\cos\theta, r\sin\theta) = r^4 - 2r^2(\cos^2\theta - \sin^2\theta) = r^2(r^2 - 2\cos 2\theta)$$

genau dann, wenn $r = r(\theta) = \sqrt{2\cos 2\theta}$ mit $\theta$ so, dass $\cos(2\theta) > 0$.

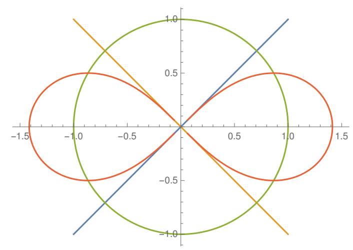


Abbildung 10: Lemniskate mit Einheitskreis und Winkelhalbierenden

Eine wichtige Anwendung des Hauptsatzes über implizite Funktionen ergibt sich im Bereich der Optimierung unter Nebenbedingungen.

**Definition 5.12** Es seien $(X, d)$ ein metrischer Raum, $E$ Vektorraum und $f : X \to \mathbb{R}$ sowie $g : X \to E$. Ist

$$L := \{g = 0\} := \{x \in X : g(x) = 0\}$$

und ist $a \in L$, so heißt $a$ eine **Extremstelle von $f$ unter der Nebenbedingung** $g = 0$, falls $a$ Extremstelle von $f|_L$ ist. Man spricht dann natürlich auch wieder von Maximal- bzw. Minimalstelle, je nach dem ob $a$ maximal oder minimal für $f|_L$ ist.

**Satz 5.13 (Lagrange)**
*Es seien $X \subset \mathbb{R}^d$ offen und $f \in C^1(X, \mathbb{R})$. Weiter seien $g \in C^1(X, \mathbb{R}^m)$ mit $m < d$ und $a \in X$ eine reguläre Stelle von $g$, also $g'(a)$ surjektiv. Ist $a$ eine Extremstelle von $f$ unter der Nebenbedingung $g(x) = 0$, so existiert eine lineare Abbildung $\lambda : \mathbb{R}^m \to \mathbb{R}$*

$$f'(a) = \lambda g'(a) .$$

**Beweis.** Da $g'(a)$ surjektiv ist, gilt $\dim \ker g'(a) = d - m$ (Dimensionsformel; Lineare Algebra). Also existieren $k_1, \ldots, k_m$ so, dass

$$\mathbb{R}^d = \ker g'(a) \times \operatorname{span}\{\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_m}\}$$

(Basisergänzungssatz; Lineare Algebra). Ohne Einschränkung kann man nach Permutation der Variablen $k_j = d - m + j$ für $j = 1, \ldots, m$ annehmen. Wir fassen $\mathbb{R}^d$ als $\mathbb{R}^{d-m} \times \mathbb{R}^m$ auf und schreiben $x = (u, w)$ mit $u \in \mathbb{R}^{d-m}$, $w \in \mathbb{R}^m$ sowie $a = (b, c)$ mit $b \in \mathbb{R}^{d-m}$, $c \in \mathbb{R}^m$. Dann ist $\ker g'(a) = \mathbb{R}^{d-m} \times \{0\}$, also

$$g'(a)(h,k) = g'(a)(h,0) + g'(a)(0,k) = D_2 g(a)k \quad ((h,k) \in \mathbb{R}^{d-m} \times \mathbb{R}^m).$$

Mit $g'(a)$ ist damit auch $D_2 g(a)$ surjektiv und nach der Dimensionsformel in $\mathrm{Aut}(\mathbb{R}^m)$. Definiert man $\lambda := D_2 f(a)(D_2 g(a))^{-1}$, so ist $\lambda : \mathbb{R}^m \to \mathbb{R}$ linear und

$$D_2 f(a) = \lambda D_2 g(a).$$

Nach dem Hauptsatz über implizite Funktionen ist die Gleichung $g(x) = g(u, w) = 0$ an $a = (b, c)$ lokal nach $w$ auflösbar. Insbesondere existieren eine offene Umgebung $U$ von $b$ sowie eine Funktion $\varphi \in C^1(U, \mathbb{R}^m)$ so, dass $\varphi(b) = c$ und $g(u, \varphi(u)) = 0$ für $u \in U$ gilt. Da die Ableitung von $u \mapsto g(u, \varphi(u))$ verschwindet, gilt nach der Kettenregel damit insbesondere

$$0 = g'(a)\begin{pmatrix} \mathrm{id}_{\mathbb{R}^{d-m}} \\ \varphi'(b) \end{pmatrix} = D_1 g(a) + D_2 g(a)\varphi'(b).$$

Ist $h \in C^1(U, \mathbb{R})$ definiert durch $h(u) := f(u, \varphi(u))$ für $u \in U$, so ist $b$ eine Extremstelle von $h$ (ohne Nebenbedingung). Nach Satz 4.10 gilt $h'(b) = 0$ und mit der Kettenregel wie oben

$$0 = h'(b) = D_1 f(a) + D_2 f(a)\varphi'(b)\ .$$

Hieraus ergibt sich wiederum

$$D_1 f(a) = -D_2 f(a)\varphi'(b) = -\lambda D_2 g(a)\varphi'(b) = \lambda D_1 g(a)$$

und damit insgesamt $f'(a) = \lambda g'(a)$. □

**Bemerkung 5.14** Da jede lineare Abbildung $\lambda : \mathbb{R}^m \to \mathbb{R}$ von der Form $h \mapsto (\lambda_1, \ldots, \lambda_m)^\top h$ ist, besagt der Satz von Lagrange, dass an (bezüglich $g$) regulären Extremstellen $a$ von $f$ unter der Nebenbedingung $g = 0$ *notwendigerweise*

$$\nabla f(a) = \sum_{j=1}^{m} \lambda_j \nabla g_j(a)$$

gilt, also der Gradient von $f$ eine Linearkombination der Gradienten von $g_1, \ldots, g_m$ ist. Ist $m = 1$, so ist die Regularitätbedingung schon dann erfüllt, wenn $\nabla g(a) \neq 0$ ist.
Um die entsprechenden kritischen Punkte $a$ zu bestimmen, hat man die Gleichungen

$$\partial_k f(x_1, \ldots, x_d) = \sum_{j=1}^{m} \lambda_j \partial_k g_j(x_1, \ldots, x_d) \qquad (k = 1, \ldots, d)$$

und

$$g_j(x_1, \ldots, x_d) = 0 \qquad (j = 1, \ldots, m)$$

zu lösen (also $(d+m)$ Gleichungen für die $(d+m)$ Unbekannten $x_1, \ldots, x_d$ und $\lambda_1, \ldots, \lambda_m$). Die Zahlen $\lambda_1, \ldots, \lambda_m$ nennt man **Lagrange-Multiplikatoren**.

**Beispiel 5.15** Die Produktion eines Unternehmens sei in Abhängigkeit der Produktionsfaktoren $x, y$ beschrieben durch die (Cobb-Douglas-)Funktion

$$P(x,y) = \sqrt{xy} \qquad (x,y > 0).$$

Die Produktionskosten seien gegeben durch eine lineare Kostenfunktion $K$ der Form

$$K(x,y) = px + qy \qquad (x,y > 0)$$

mit Konstanten $p, q > 0$. Gesucht ist eine kostenminimale Faktorkombination $(x, y)$ zu einem vorgegebenen Produktionsniveau $c > 0$, d. h. wir wollen $K(x, y)$ unter der Nebenbedingung $P(x, y) - c = 0$ minimieren. Nach Bemerkung 5.14 ist wegen $\nabla P(x, y) \neq 0$ für alle $(x, y)$ eine notwendige Bedingung gegeben durch

$$\nabla K(x,y) = \lambda\, \nabla P(x,y)$$

für ein $\lambda \in \mathbb{R}$, d. h. wir haben die drei Gleichungen

$$\begin{aligned} 2p &= \lambda\sqrt{y/x} \\ 2q &= \lambda\sqrt{x/y} \\ xy &= c^2 \end{aligned}$$

für die Unbekannten $x, y, \lambda$. Division der zweiten durch die erste Gleichung ergibt $x = qy/p$ und mit der dritten Gleichung erhalten wir $y = c\sqrt{p/q}$ und damit $x = c\sqrt{q/p}$. Also kann nur an $c(\sqrt{q/p}, \sqrt{p/q})$ ein Minimum unter der Nebenbedingung vorliegen. Man kann sich überlegen, dass dies tatsächlich der Fall ist. Für $c = \sqrt{2}$, $p = 1$ und $q = 2$ ergibt sich $(x, y) = (2, 1)$ (vgl. Abbildung 11). Man sieht, dass an der Minimalstelle die Höhenlinien von $P$ und $K$ tangential

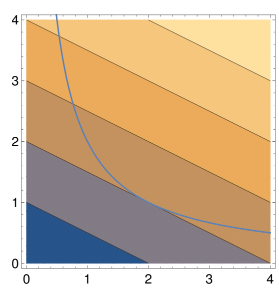


Abbildung 11: Niveaulinie $P(x,y) = \sqrt{2}$ und Niveaulinien zu $K(x,y) = x + 2y$

liegen. Dies liegt daran, dass die Gradienten linear abhängig sind und beide jeweils senkrecht auf den entsprechenden Höhenlinien stehen ([Ü]).

# Index