Jürgen Müller

Analysis einer und mehrerer Veränderlicher

Skript zur Vorlesung Sommersemster 2020
Universität Trier
Fachbereich IV
Mathematik/Analysis

Inhaltsverzeichnis

1 Funktionenfolgen und Funktionenreihen

Wichtige Funktionen wie die Exponentialfunktion und die trigonometrischen Funktionen sind über gewisse Reihenwerte definiert. Ziel des ersten Abschnitts ist es, allgemeine Strukturaussagen über Funktionen zu machen, die sich als Grenzwerte sogenannter Funktionenfolgen oder -reihen ergeben.

Definition 1.1 Es seien $X \neq \varnothing$ eine Menge und (Y, d) ein metrischer Raum. Eine Folge $(f_n)_{n\in N}$ in $\mathrm{Abb}(X, Y)$ nennt man eine **Funktionenfolge**.[1] 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 1.2 Wir betrachten die Funktionen $f_n : \mathbb{R} \to \mathbb{R}$ mit

$$f_n(x) := x^n \qquad (x \in \mathbb{R},\, n \in \mathbb{N}) .$$

Dann gilt

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

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 1.3 1. Es seien $M \neq \varnothing$ eine Menge und

$$B(M) := B(M, \mathbb{C}) := \{f : M \to \mathbb{C} : f \text{ beschränkt}\}.$$

Dann ist $B(M)$ mit der Supremumsnorm

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

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

ein normierter Raum. Ist M ein kompakter metrischer Raum, so ist der Raum $C(M)$ der stetigen Funktionen von M nach $\mathbb{C}$ ein Teilraum von $B(M)$ (siehe Einführung in die Mathematik).
2. Eine Folge $(f_n)_{n\in N}$ in $\text{Abb}(X, \mathbb{C})$ heißt **gleichmäßig konvergent** auf der Menge $M \subset X$, falls eine Funktion $f : M \to \mathbb{C}$ existiert mit

$$\|f_n - f\|_\infty \to 0 \quad (n \to \infty).$$

Wir schreiben dann

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

oder auch

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

Gilt $f_n \to f$ gleichmäßig auf M, so folgt wegen $|f_n(x) - f(x)| \leq \|f_n - f\|_\infty$ auch $f_n \to f$ punktweise auf $x \in M$, mit anderen Worten: gleichmäßige Konvergenz impliziert punktweise Konvergenz.

Beispiel 1.4 Wir betrachten wieder f_n und f aus Beispiel 1.2. Ist $M = [-1/2, 1/2]$, so gilt

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

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

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

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

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

Satz 1.5 *Es seien (X, d) ein metrischer Raum und $a \in X$. Weiter sei $(f_n)_{n\in N}$ eine Folge in $\text{Abb}(X, \mathbb{C})$, die gleichmäßig auf einer Umgebung*[2] *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

$$\sup_{x\in U} |f(x) - f_n(x)| < \varepsilon/3\ .$$

Da f_n stetig an der Stelle a ist, existiert ein $\delta = \delta_\varepsilon > 0$ so, dass $U_\delta(a) \subset U$ und

$$|f_n(x) - f_n(a)| < \varepsilon/3 \qquad (x \in U_\delta(a))\ .$$

[2] $U \subset X$ heißt **Umgebung** von a, falls $U_\delta(a) \subset U$ für ein $\delta > 0$, wobei $U_\delta(a) := \{x : d(x, a) < \delta\}$.

Damit gilt für $x \in U_\delta(a)$

$$|f(x) - f(a)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(a) \ + |f_n(a) - f(a)| < \varepsilon \ .$$

□

Ist $M \neq \varnothing$ eine Menge, so ist durch $d_\infty(f,g) = d_{\infty,M}(f,g) := \|f-g\|_\infty$ für $f, g \in B(M)$ eine Metrik auf $B(M)$ gegeben (die von der Supremumsnorm induzierte Metrik). Es gilt damit

Satz 1.6 (Cauchy-Kriterium für gleichmäßige Konvergenz)
Es sei $M \neq \varnothing$ eine Menge. Dann ist jede Cauchy-Folge in $(B(M), d_\infty)$ konvergent.[3]

Beweis. Es sei $(f_n)_{n\in N}$ eine Cauchy-Folge in $B(M)$. Dann ist insbesondere für jedes feste $x \in M$ die Folge $(f_n(x))_{n\in N}$ eine Cauchyfolge in $\mathbb{C}$. Da $\mathbb{C}$ 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 die Abbildung $y \mapsto |f_n(x) - y|$ stetig (man beachte: $u \mapsto |u|$ ist stetig). Also gilt

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

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 beschränkt, 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 1.7 Sind $X \neq \varnothing$ eine Menge und $(f_n)_{n\geq m}$ eine Folge in $\mathrm{Abb}(X, \mathbb{C})$. Dann heißt die Folge $(s_n)_{n\geq m}$ der Partialsummen $s_n := \sum\limits_{\nu=m}^{n} f_\nu$ die mit (f_n) gebildete **Funktionenreihe**. Man schreibt wie bei Zahlenreihen $\sum\limits_{\nu=m}^{\infty} f_\nu$ statt $(s_n)_{n\geq m}$. Die Funktionenreihe $\sum\limits_{\nu=m}^{\infty} f_\nu$ 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 $\sum\limits_{\nu=m}^{\infty} f_\nu$ dann auch für die Grenzfunktion.

[3] Man kann die Aussage so interpretieren, dass der normierte Raum $(B(M), \|\cdot\|_\infty)$ ein Banachraum ist, d. h. $(B(M), d_\infty)$ ist ein vollständiger metrischer Raum.

Ein einfaches hinreichendes Kriterium für gleichmäßige Konvergenz ist[4]

Satz 1.8 (Weierstraß-Kriterium)
Es seien $M \neq \varnothing$ eine Menge und $(f_n)_{n\geq m}$ eine Folge in $B(M)$. 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 $|f_n(x)| \leq b_n$ für alle $x \in M$, $n \geq n_0$, so konvergiert $\sum_{\nu=m}^{\infty} f_\nu$ gleichmäßig auf M.

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

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

Damit folgt für $n > n' > R$

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

Mit dem Cauchy-Kriterium für gleichmäßige Konvergenz (Satz 1.6) ergibt sich die Behauptung. □

Definition 1.9 Es seien $a \in \mathbb{K}$ und $(c_n)_{n=0}^{\infty}$ eine Folge in $\mathbb{C}$. Dann heißt die Funktionenreihe $\sum_{\nu=0}^{\infty} f_\nu$ mit $f_n : \mathbb{K} \to \mathbb{C}$, definiert durch

$$f_n(x) := c_n(x-a)^n \quad (x \in \mathbb{K},\ n \in \mathbb{N}_0),$$

Potenzreihe mit **Entwicklungsmitte** a und **Koeffizientenfolge** (c_n). Weiter heißen

$$R := \sup\{|h| : \sum_{\nu=0}^{\infty} c_\nu h^\nu \text{ konvergent}\} \in [0,\infty]$$

der **Konvergenzradius** und $U_R(a)$ der **Konvergenzkreis** (im Falle $\mathbb{K} = \mathbb{R}$ meist das **Konvergenzintervall**) der Potenzreihe.

Beispiel 1.10 1. Die geometrische Reihe $\sum_{\nu=0}^{\infty} z^\nu$ ist eine Potenzreihe mit $a = 0$ und $c_n = 1$ für $n \in \mathbb{N}_0$. Hier ist $R = 1$ und

$$\sum_{\nu=0}^{\infty} z^\nu = \frac{1}{1-z} \qquad (|z| < 1).$$

[4] 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.

2. Die Exponentialreihe

$$\sum_{\nu=0}^{\infty} \frac{z^\nu}{\nu!} = \exp(z)$$

ist eine Potenzreihe mit $a = 0$ und $c_n = 1/n!$ für $n \in \mathbb{N}_0$. Da die Reihe für alle $z \in \mathbb{C}$ konvergiert, gilt $R = \infty$.

Bemerkung 1.11 Es sei $\sum\limits_{\nu=0}^{\infty} c_\nu (x-a)^\nu$ eine Potenzreihe.

1. Ist der Konvergenzradius $R > 0$, so ist die Potenzreihe für alle $r < R$ gleichmäßig konvergent auf $U_r(a)$ und damit punktweise konvergent auf $U_R(a) = \bigcup_{r<R} U_r(a)$.

> Denn: Es sei $h \in \mathbb{K}$ mit $|h| > r$ so, dass $\sum\limits_{\nu=0}^{\infty} c_\nu h^\nu$ konvergiert. Dann ist $(c_n h^n)$ eine Nullfolge, also existiert ein n_0 mit $|c_n h^n| \le 1$ für $n \ge n_0$. Damit ergibt sich für $x \in U_r(a)$
>
> $$|f_n(x)| = |c_n(x-a)^n| \le |x-a|^n/|h|^n \le (r/|h|)^n \quad (n \ge n_0).$$
>
> Wegen $q := r/|h| < 1$ ist $\sum\limits_{\nu=0}^{\infty} q^\nu < \infty$. Mit dem Weierstraß-Kriterium (Satz 1.8) folgt die Behauptung.

2. Definiert man $f(x) := \sum\limits_{\nu=0}^{\infty} c_\nu (x-a)^\nu$ für $x \in U_R(a)$, so ist die Funktion $f : U_R(a) \to \mathbb{C}$ stetig.

> Denn: Ist $y \in U_R(a)$, so existiert ein $r < R$ mit $y \in U := U_r(a)$. Da U offen ist, ist U eine Umgebung von y, und da Polynome stetig sind, ist f nach 1. und Satz 1.5 stetig an der Stelle y.

3. Aus dem Quotientenkriterium (siehe Einführung in die Mathematik) folgt leicht: Ist $c_n \neq 0$ für n genügend groß und gilt $|c_{n+1}/c_n| \to d$ für $n \to \infty$, so ist der Konvergenzradius $R = 1/d$ (mit $1/0 := \infty$ und $1/\infty := 0$).

Beispiel 1.12 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

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

Da für $\alpha > 1$

$$\sum_{\nu=1}^{\infty} 1/\nu^\alpha < \infty$$

gilt (Einführung in die Mathematik), ist nach dem Weierstraß-Kriterium die Funktionenreihe $\sum\limits_{\nu=1}^{\infty} f_\nu$ gleichmäßig konvergent auf M_α und damit punktweise auf $M := \bigcup_{\alpha>1} M_\alpha$. Die Funktion $\zeta : M \to \mathbb{C}$, definiert durch $\zeta(z) := \sum\limits_{\nu=1}^{\infty} 1/\nu^z$ für $z \in M$, heißt **(Riemannsche) Zetafunktion**. Da M_α eine Umgebung jedes Punktes $a \in M_\alpha$ ist, folgt aus Satz 1.5 die Stetigkeit von ζ auf M.

2 Differenzialrechnung

Wir untersuchen wieder Funktionen $f : X \to \mathbb{C}$, wobei $X \subset \mathbb{K}$, also Funktionen einer reellen oder komplexen Variablen. Um die feinere Struktur des Veränderungsverhaltens solcher Funktionen untersuchen zu können, brauchen wir einen über die Stetigkeit hinausgehenden Glattheitsbegriff. Grob gesagt wollen wir Funktionen definieren, die lokal sehr gut durch affin-lineare Abbildungen approximiert werden können.

Definition 2.1 Es seien $X \subset \mathbb{K}$ und $f : X \to \mathbb{C}$.

1. f heißt **differenzierbar an der Stelle** $a \in X$, falls $a \in X'$ gilt[5] und der Grenzwert
$$f'(a) := \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$
existiert. Man bezeichnet $f'(a)$ als **Ableitung** von f an der Stelle a.

2. f heißt **differenzierbar** (auf X), falls f in jedem Punkt $x \in X$ differenzierbar ist. In diesem Fall heißt die Funktion $f' : X \to \mathbb{C}$ **Ableitung** von f.[6]

Bemerkung 2.2 Es seien $X \subset \mathbb{K}$, $a \in X \cap X'$ und $f : X \to \mathbb{C}$. Definiert man die Funktion $\tau_a f : (X - a) \to \mathbb{C}$ durch
$$\tau_a f(h) := (\tau_a f)(h) := f(a + h) - f(a) \qquad (h \in X - a).$$
so ist $(\tau_a f)(0) = 0$ und f genau dann differenzierbar an a, wenn
$$\lim_{h \to 0} \frac{1}{h}\, \tau_a f(h) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$$
existiert, also $\tau_a f$ differenzierbar an 0 ist, und in diesem Fall ist
$$f'(a) = (\tau_a f)'(0).$$
Man kann sich also bei Bedarf bei der Untersuchung von Ableitungen stets auf den Fall $a = f(a) = 0$ zurückziehen.

Beispiel 2.3 1. Sind $b, c \in \mathbb{C}$ und $f(x) = cx + b$ für $x \in \mathbb{K}$, so ist $f(x + h) - f(x) = ch$, also f differenzierbar auf $\mathbb{K}$ mit $f'(x) = c$. Ist $f(x) := x^2$, so gilt für $x \in \mathbb{K}$
$$\frac{(x + h)^2 - x^2}{h} = 2x + h \to 2x \quad (h \to 0),$$
also $f'(x) = 2x$. Allgemeiner ergibt sich: Sind $k \in \mathbb{N}$ und $f(x) = x^k$ für $x \in \mathbb{K}$, so gilt (Einführung in die Mathematik)
$$\frac{(x + h)^k - x^k}{h} = \sum_{j=0}^{k-1} (x + h)^j x^{k-1-j} \to \sum_{j=0}^{k-1} x^{k-1} = kx^{k-1} \quad (h \to 0),$$

[5] X' ist die Menge der Häufungspunkte von X, also die $x \in \mathbb{K}$ so, dass $U_\delta(x) \setminus \{x\} \neq \varnothing$ für alle $\delta > 0$ ist.
[6] Weitere Schreibweisen sind etwa Df oder df oder auch (df/dx).

also $f'(x) = kx^{k-1}$.
2. Ist $f(x) := |x|$ für $x \in \mathbb{R}$, so ist f stetig auf $\mathbb{R}$, aber nicht differenzierbar an der Stelle $a = 0$, da $f(h)/h = 1$ für $h > 0$ und $f(h)/h = -h/h = -1$ für $h < 0$ gilt. Also hat $h \mapsto f(h)/h$ keinen (beidseitigen) Grenzwert an 0. Das Beispiel zeigt, dass Stetigkeit an einer Stelle im Allgemeinen *nicht* die Differenzierbarkeit an dieser Stelle impliziert.

Von zentraler Bedeutung für die Analysis ist

Satz 2.4 *Die Funktion* exp *ist differenzierbar auf* $\mathbb{C}$ *mit*

$$\exp' = \exp\,.$$

Beweis. Nach Bemerkung 1.11 ist $g : \mathbb{C} \to \mathbb{C}$ mit

$$g(h) := \sum_{\mu=0}^{\infty} \frac{h^\mu}{(\mu+1)!}$$

stetig auf $\mathbb{C}$ mit $\varepsilon(0) = 1$. Damit gilt für[7] $h \in \mathbb{C}^*$

$$\frac{1}{h}(e^h - 1) = \sum_{\nu=1}^{\infty} \frac{h^{\nu-1}}{\nu!} = g(h) \to 1 \qquad (h \to 0).$$

Für beliebiges $a \in \mathbb{C}$ folgt damit $(e^{a+h} - e^a)/h = e^a \cdot (e^h - 1)/h \to e^a$ für $h \to 0$. $\square$

Satz 2.5 *Es seien* $X \subset \mathbb{K}$, $a \in X \cap X'$ *und* $f : X \to \mathbb{C}$. *Mit* $X_a := (X - a) \setminus \{0\}$ *gilt*

1. (**Zerlegungsformel**, **affin-lineare Approximation**)

   f ist differenzierbar an a genau dann, wenn ein $c = c_{f,a} \in \mathbb{C}$ und eine Funktion $\varepsilon = \varepsilon_{f,a} : X_a \to \mathbb{C}$ existieren mit $\varepsilon(h) \to 0$ für $h \to 0$ und

   $$f(a+h) = f(a) + c \cdot h + \varepsilon(h) \cdot h \qquad (h \in X_a)\ .$$

   In diesem Fall ist $f'(a) = c$.

2. *Ist f differenzierbar an a, so ist f auch stetig an a.*

Beweis. 1. $\Rightarrow$: Setzt man für $h \in X_a$

$$\varepsilon(h) := (\tau_a f)(h)/h - f'(a),$$

[7] Ist $(R, +, \cdot)$ ein Ring, so schreiben wir $R^* := R \setminus \{0\}$.

so gilt $\varepsilon(h) \to 0$ für $h \to 0$ aufgrund der Differenzierbarkeit von f an a und

$$f(a+h) = f(a) + f'(a) \cdot h + \varepsilon(h) \cdot h \qquad (h \in X_a),$$

also ist $c = f'(a)$ geeignet.
$\Leftarrow$: Es gilt $(\tau_a f)(h)/h = c + \varepsilon(h) \to c$ für $h \to 0$.
2. Die Aussage folgt aus der Zerlegungsformel für $h \to 0$. □

Die Zerlegungsformel zeigt, dass Differenzierbarkeit einer Funktion f an einer Stelle a bedeutet, dass f lokal an a mit einer gewissen Güte durch eine affin-lineare Funktion der Form

$$h \mapsto f(a) + c \cdot h$$

approximiert werden kann.

Satz 2.6 (Summenregel, Produktregel, Quotientenregel)
Es seien $X \subset \mathbb{K}$ und $f, g : X \to \mathbb{C}$ differenzierbar an der Stelle $a \in X$. Dann gilt

1. *$f + g$ ist differenzierbar an a mit*

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

2. *$f \cdot g$ ist differenzierbar an a mit*

$$(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a)\ .$$

3. *Ist g nullstellenfrei, so ist f/g differenzierbar an a mit*

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}\ .$$

Beweis. Wie man leicht nachrechnet, gilt für $h \in X - a$

$$\tau_a(f+g)(h) = \tau_a f(h) + \tau_a g(h)$$

und

$$\tau_a(f \cdot g)(h) = g(a+h) \cdot \tau_a f(h) + f(a) \cdot \tau_a g(h).$$

Nach Satz 2.5 ist g stetig an a. Damit ergeben sich die Summenregel und die Produktregel jeweils nach Division durch h und Grenzwertbildung für $h \to 0$. Ist zusätzlich g nullstellenfrei, so gilt auch

$$\tau_a(1/g)(h) = \frac{-\tau_a g(h)}{g(a+h)g(a)}\ .$$

Wieder nach Division durch h und Grenzwertbildung für $h \to 0$ ergibt sich die Quotientenregel für $f = 1$. Die Aussage für allgemeines f folgt mit der Produktregel. □

Beispiel 2.7 Für $f(x) = xe^x$ folgt aus der Produktregel $f'(x) = 1 \cdot e^x + x \cdot e^x = (x+1)e^x$.

Satz 2.8 (Kettenregel)
Es seien $X, Y \subset \mathbb{K}$, $f : X \to Y$ und $g : Y \to \mathbb{C}$. Ist f differenzierbar an $a \in X$ und ist g differenzierbar an $f(a)$, so ist $g \circ f$ differenzierbar an a mit

$$(g \circ f)'(a) = g'\big(f(a)\big) f'(a) \,.$$

Beweis. Wir betrachten zunächst den Spezialfall $a = 0$ und $f(0) = g(0) = 0$. Ist $\varepsilon := \varepsilon_{g,0}$ wie in der Zerlegungsformel und $\varepsilon(0) := 0$, so ist $\varepsilon \circ f$ abklingend an 0 und damit gilt

$$\begin{aligned} \frac{1}{h}(g \circ f)(h) &= \frac{1}{h}\big(g(f(h)) - g'(0)f(h)\big) + g'(0)\frac{f(h)}{h} \\ &= \varepsilon(f(h))\frac{f(h)}{h} + g'(0)\frac{f(h)}{h} \to g'(0)f'(0) \quad (h \to 0). \end{aligned}$$

Also ist $g \circ f$ differenzierbar an 0 mit $(g \circ f)'(0) = g'(0)f'(0)$.
Sind nun a sowie $f(a)$ und $g(f(a))$ beliebig, so gilt

$$\tau_a(g \circ f) = \tau_{f(a)} g \circ \tau_a f \,,$$

und damit ergibt sich die Behauptung durch Anwendung des Spezialfalls auf die rechte Seite. □

Beispiel 2.9 Für $p : \mathbb{K} \to \mathbb{C}$ mit

$$p(x) = (x^3 + 2x + 1)^5 \quad (x \in \mathbb{K})$$

gilt $p = g \circ f$ mit $f(x) = x^3 + 2x + 1$ und $g(y) = y^5$. Also ergibt sich aus der Kettenregel

$$p'(x) = g'(f(x))f'(x) = 5(x^3 + 2x + 1)^4(3x^2 + 2) \quad (x \in \mathbb{K}).$$

Satz 2.10 *Die Funktionen* sin *und* cos *sind differenzierbar auf* $\mathbb{C}$ *mit*

$$\sin' = \cos \quad \textit{und} \quad \cos' = -\sin \,.$$

Beweis. Für $z \in \mathbb{C}$ ist nach der Kettenregel

$$\sin'(z) = \frac{1}{2i}(ie^{iz} + ie^{-iz}) = \frac{1}{2}(e^{iz} + e^{-iz}) = \cos z \,.$$

Entsprechend ergibt sich $\cos' = -\sin$. □

Beispiel 2.11 1. Aus Satz 2.10 und der Quotientenregel folgt

$$\tan' = \frac{1}{\cos^2}\big(\cos^2 + \sin^2\big) = \frac{1}{\cos^2} = 1 + \tan^2$$

auf $\mathbb{C} \setminus \pi(\mathbb{Z} + 1/2)$ und entsprechend auf $\mathbb{C} \setminus \pi\mathbb{Z}$

$$\cot' = -\sin^{-2} = -1 - \cot^2.$$

2. Ist $a > 0$ fest und $f(z) := a^z = e^{z \ln a}$ für $z \in \mathbb{C}$, so folgt aus der Kettenregel

$$f'(z) = a^z \ln a \qquad (z \in \mathbb{C}).$$

Satz 2.12 (Umkehrregel)
Es seien $X \subset \mathbb{K}$ und $f : X \to Y \subset \mathbb{C}$ bijektiv. Ist f differenzierbar an der Stelle $a \in X$ mit $f'(a) \neq 0$ und ist die Umkehrfunktion f^{-1} stetig an $c := f(a)$, so ist f^{-1} differenzierbar an c mit

$$(f^{-1})'(c) = 1/f'(a) = 1/(f'(f^{-1}(c))\ .$$

Beweis. Ist (x_n) eine Folge in X mit $a \neq x_n \to a$ $(n \to \infty)$, so gilt aufgrund der Stetigkeit von f an a und der Injektivität auch

$$f(a) \neq f(x_n) \to c \qquad (n \to \infty)\ .$$

Also ist c ein Häufungspunkt von Y. Es sei zunächst $a = c = 0$. Dann ist f^{-1} stetig an 0. Also gilt $f^{-1}(u) \to 0$ $(u \to 0)$ und folglich wegen $f^{-1}(u) \neq 0$ für $u \neq 0$

$$\frac{f^{-1}(u)}{u} = \frac{f^{-1}(u)}{f(f^{-1}(u))} \to \frac{1}{f'(0)} \qquad (u \to 0)\ .$$

Der allgemeine Fall ergibt sich aus Bemerkung 2.2 und $\tau_c f^{-1} = (\tau_a f)^{-1}$. □

Bemerkung 2.13 Sind $X = [0, 2\pi)$ und $f : X \to \{z : |z| = 1\}$ mit $f(x) = e^{ix}$, so ist f bijektiv und differenzierbar mit $f'(x) = ie^{ix} \neq 0$ für alle x. Die Umkehrfunktion ist allerdings nicht stetig an der Stelle 1 ([Ü]) und damit auch nicht differenzierbar. Man sieht also, dass man im Allgemeinen auf die Stetigkeitsvoraussetzung an f^{-1} in Satz 2.12 nicht verzichten kann.
Ist X *kompakt* und f stetig, so ist die Umkehrfunktion f^{-1} ist „automatisch“ stetig (Einführung in die Mathematik).

Bemerkung 2.14 1. Es seien $X \subset \mathbb{K}$ und $f : X \to Y \subset \mathbb{C}$ bijektiv mit stetiger Umkehrfunktion. Ist $g : Y \to \mathbb{C}$ nullstellenfrei mit $f' = g \circ f$ auf X, so ist nach der Umkehrregel

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

2. Mit $g(t) := t$ für $t > 0$ und $f(s) := e^s$ für $s \in \mathbb{R}$ folgt aus 1.

$$\ln'(t) = 1/t \qquad (t > 0) \ .$$

Damit ergibt sich für festes $\alpha \in \mathbb{C}$ mit der Kettenregel auch die Differenzierbarkeit von $t \mapsto t^\alpha = e^{\alpha \ln t}$ auf $(0, \infty)$ mit Ableitung $t \mapsto \alpha t^{\alpha-1}$.

3. Es gilt

$$\arctan'(t) \quad = \quad \frac{1}{1+t^2}, \qquad \operatorname{arccot}'(t) = -\frac{1}{1+t^2} \qquad (t \in \mathbb{R}) \ .$$

Denn: Nach 1. und Beispiel 2.11, angewandt auf $g(t) := 1 + t^2$ für $t \in \mathbb{R}$ und $f(s) = \tan(s)$ für $s \in (-\pi/2, \pi/2)$, ist $\arctan' = (f^{-1})' = 1/g$. Entsprechendes gilt für arccot.

4. Es gilt ([Ü])

$$\arcsin'(t) = \frac{1}{\sqrt{1-t^2}} \ , \qquad \arccos'(t) = -\frac{1}{\sqrt{1-t^2}} \qquad (t \in (-1, 1)).$$

Wir beschäftigen uns nun damit, wie man die Differenzialrechnung nutzen kann, um insbesondere das lokale Verhalten differenzierbarer Funktionen genauer zu untersuchen. Dazu definieren wir zunächst, was wir unter Extremstellen verstehen.

Definition 2.15 Es seien (X, d) ein metrischer Raum und $f : X \to \mathbb{R}$.

1. Ein Punkt $a \in X$ heißt **Maximalstelle** (von f), falls eine Umgebung U von a existiert mit
$$f(x) \le f(a) \qquad (x \in U) \ ,$$
falls also $f|_U$ an der Stelle a maximal wird. In diesem Fall nennt man den Funktionswert $f(a)$ ein **lokales Maximum**. Gilt $<$ statt $\le$ für $x \ne a$, so spricht man von einem **strikten** lokalen Maximum und einer strikten Maximalstelle.

2. Ein Punkt $a \in X$ heißt **Minimalstelle** (von f), falls eine Umgebung U von a existiert mit
$$f(x) \ge f(a) \qquad (x \in U) \ ,$$
falls also $f|_U$ an der Stelle a minimal wird. Dann heißt $f(a)$ ein **lokales Minimum**. Gilt $>$ statt $\ge$ für $x \ne a$, so spricht man von einem **strikten** lokalen Minimum und einer strikten Minimalstelle.

Ist a eine Maximal- oder eine Minimalstelle von f, so nennt man a auch eine **Extremstelle** von f.

Wir werden sehen, dass die Differenzialrechnung ein effizientes Instrumentarium zur Bestimmung von Extremstellen zur Verfügung stellt. Ist $f : X \to \mathbb{C}$ differenzierbar, so nennt man eine Nullstelle von f' auch **kritische Stelle** von f.

Satz 2.16 *Es seien $X \subset \mathbb{R}$ 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 eine kritische Stelle, also $f'(a) = 0$.*

Beweis. Ohne Einschränkung sei a eine Minimalstelle. Dann existiert ein $\delta > 0$ mit $\tau_a f(h) \geq 0$ für $|h| < \delta$. Da f differenzierbar an a ist, gilt

$$0 \leq \lim_{h \to 0^+} (\tau_a f)(h)/h = f'(a) = \lim_{h \to 0^-} (\tau_a f)(h)/h \leq 0.$$

□

Bemerkung 2.17 1. Das Verschwinden von f' an einer Stelle a ist lediglich eine *notwendige* Bedingung dafür, dass a eine Maximal- oder Minimalstelle ist. So ist etwa für $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = x^3$ der Nullpunkt eine kritische Stelle, aber keine Extremstelle. Daher benutzt man Satz 2.16 typischerweise dafür, die Extremalität von f an den inneren Punkten a, die nicht kritisch sind, auszuschließen. Kritische Stellen sind die einzigen „Kandidatinnen" für Extremstellen im Inneren von X.
2. Ist a eine Extremstelle, aber kein innerer Punkt von X, so ist a nicht notwendig eine kritische Stelle. So sind ± 1 für $f : [-1, 1] \to \mathbb{R}$ mit $f(x) = |x|$ Maximalstellen, aber keine kritischen Stellen. Außerdem ist 0 eine Minimalstelle, aber auch keine kritische Stelle (da f an 0 gar nicht differenzierbar ist).

Satz 2.18 (Rolle)
Es seien $I = (\alpha, \beta)$ ein offenes Intervall und $f : I \to \mathbb{R}$ differenzierbar. Existieren $f(\alpha^+)$ und $f(\beta^-)$ und ist $f(\alpha^+) = f(\beta^-)$,[8] so hat f eine kritische Stelle.

Beweis. Man kann zeigen ([Ü]), dass f auf (α, β) maximal oder minimal wird.[9] Also hat f insbesondere eine Extremstelle und damit nach Satz 2.16 eine kritische Stelle. □

Als Folgerung erhalten wir

Satz 2.19 *Es seien $I = (\alpha, \beta)$ ein offenes Intervall und $f, g : I \to \mathbb{R}$ differenzierbar. Existieren $f(\alpha^+)$ und $f(\beta^-)$ sowie $g(\alpha^+)$ und $g(\beta^-)$, so existiert ein $\tau \in (\alpha, \beta)$ mit*

$$(f(\beta^-) - f(\alpha^+))g'(\tau) = f'(\tau)(g(\beta^-) - g(\alpha^+))$$

(**erweiterter Mittelwertsatz**). *Im Fall $g(t) = t$ für $t \in (\alpha, \beta)$ ist*

$$f(\beta^-) - f(\alpha^+) = f'(\tau)(\beta - \alpha)$$

(**Mittelwertsatz**).

[8] Hier bezeichnen $f(s^+) := \lim_{t \to s^+} f(t)$ den rechtsseitigen und $f(s^-) := \lim_{t \to s^-} f(t)$ den linksseitigen Grenzwert an s.

[9] Wesentlich ist dabei, dass stetige reellwertige Funktionen auf kompakten Mengen maximal und minimal werden, also Maximum und Minimun haben.; eine der zentralen Aussagen aus der Einführung in die Mathematik!

Beweis. Wir betrachten die Funktion $\varphi : I \to \mathbb{R}$ mit

$$\varphi(t) := \big(f(\beta^-) - f(\alpha^+)\big)g(t) - f(t)\big(g(\beta^-) - g(\alpha^+)\big) \qquad (t \in I).$$

Da $\varphi(\beta^-) = \varphi(\alpha^+) = f(\alpha^+)g(\beta^-) - f(\beta^-)g(\alpha^+)$ gilt, folgt die erste Aussage durch Anwendung des Satzes von Rolle mit φ statt f. □

Bemerkung 2.20 Ist $f : (\alpha, \beta) \to \mathbb{C}$ differenzierbar, so ist auch $\mathrm{Re}f$ differenzierbar mit $(\mathrm{Re}f)' = \mathrm{Re}(f')$. Anwendung des Mittelwertsatzes auf $\mathrm{Re}f$ ergibt: Existieren $\mathrm{Re}f(\alpha^+)$ und $\mathrm{Re}f(\beta^-)$, so gilt es ein $\tau \in (\alpha, \beta)$ mit $\mathrm{Re}\big(f(\beta^-) - f(\alpha^+)\big) = (\beta - \alpha) \cdot \mathrm{Re}(f')(\tau)$.

Es seien V ein Vektorraum und $u, v \in V$. Wir definieren $s_u^v : [0,1] \to V$ durch $s_u^v(t) := u + t(v - u)$ für $t \in [0,1]$ und nennen s_u^v **orientierte Strecke** von u nach v. Außerdem setzen wir

$$[u, v] := s_u^v([0,1]) = \{u + t(v - u) : t \in [0,1]\}$$

und

$$(u, v) := s_u^v((0,1)) = \{u + t(v - u) : t \in (0,1)\}.$$

Satz 2.21 (Schrankensatz)
Es seien $a, b \in \mathbb{K}$, $a \neq b$. Ist $f : [a, b] \to \mathbb{C}$ stetig auf $[a, b]$ und differenzierbar auf (a, b), so existiert ein $\xi \in (a, b)$ mit

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

Beweis. Ohne Einschränkung können wir $c := f(b) - f(a) > 0$ annehmen (ist $c = 0$, so ist die Behauptung kar; ist $c \neq 0$, so kann man statt f die Funktion $(\bar{c}/|c|)f$ betrachten).
Ist $s(t) := a + t(b - a) = s_a^b(t)$ für $t \in [0,1]$, so gilt $(f \circ s)' = (f' \circ s) \cdot (b - a)$ auf $(0,1)$ nach der Kettenregel. Mit Bemerkung 2.20, angewandt auf $f \circ s$ mit $I = (0,1)$, folgt die Existenz eines $\tau \in (0,1)$ so, dass

$$c = \mathrm{Re}\big((f \circ s)(1) - (f \circ s)(0)\big) = \mathrm{Re}(f'(s(\tau)) \cdot (b - a)) \leq |f'(s(\tau))| \cdot |b - a|.$$

Mit $\xi := s(\tau) = a + \tau(b - a) \in (a, b)$ ergibt sich die Behauptung. □

Bemerkung 2.22 Ist V ein Vektorraum, so heißt $X \subset V$ **sternförmig bezüglich** $a \in X$, falls

$$X = \bigcup_{x \in X} [a, x]$$

gilt, und kurz **sternförmig**, falls sie sternförmig bezüglich eines Punktes a ist. Die Menge X heißt **konvex**, falls sie sternförmig bezüglich aller Punkte $a \in X$ ist, also falls $[a, x] \subset X$ für alle $a, x \in X$ gilt. Eine nichtleere Menge in $\mathbb{R}$ ist genau dann sternförmig, wenn sie konvex ist, und dies gilt genau dann, wenn sie ein Intervall ist.
Aus dem Schrankensatz folgt leicht: Ist $X \subset \mathbb{K}$ sternförmig und ist $f : X \to \mathbb{C}$ differenzierbar mit $f' = 0$, so ist f konstant.

Satz 2.23 *Es sei I ein Intervall und $f : I \to \mathbb{R}$ sei stetig auf I und differenzierbar auf I°. Dann gilt*

1. *Ist $f' \geq 0$ auf I°, so ist f wachsend auf I.*
2. *Ist $f' \leq 0$ auf I°, so ist f fallend auf I.*

Ist dabei $f' > 0$ beziehungsweise $f' < 0$, so ist die Monotonie streng.

Beweis. 1. Sind $s, t \in I$ mit $s < t$, so ergibt sich aus dem Mittelwertsatz (Satz 2.19) mit einem $\tau \in (s, t)$
$$f(t) - f(s) = f'(\tau)(t - s) \geq 0 ,$$
also $f(s) \leq f(t)$. Ist dabei $f'(\tau) > 0$, so ist $f(s) < f(t)$.
2. Ergibt sich durch Anwendung von 1. auf $-f$. $\square$

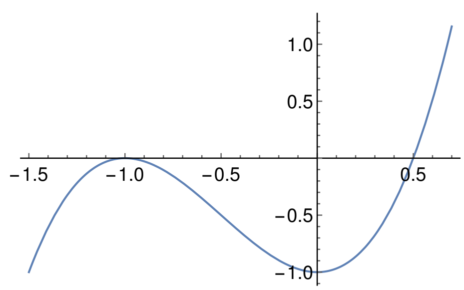

Abbildung 1: Polynom $x \mapsto 2x^3 + 3x^2 - 1$

Beispiel 2.24 1. Es sei $f : \mathbb{R} \to \mathbb{R}$ mit
$$f(x) = 2x^3 + 3x^2 - 1 \qquad (x \in \mathbb{R}) .$$
Dann gilt
$$f'(x) = 6x(x+1) \begin{cases} > 0, & \text{für } x \in (-\infty, -1) \\ < 0, & \text{für } x \in (-1, 0) \\ > 0, & \text{für } x \in (0, \infty) \end{cases} .$$

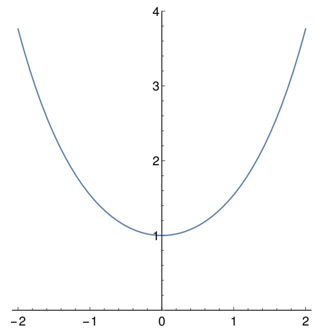

Abbildung 2: cosh auf dem Intervall $[-2, 2]$

Nach Satz 2.23 ist f streng wachsend auf $(-\infty, -1]$, streng fallend auf $[-1, 0]$ und streng wachsend auf $[0, \infty)$.
2. Wegen $\sinh > 0$ auf $(0, \infty)$ und $\sinh < 0$ auf $(-\infty, 0)$ sowie $\cosh' = \sinh$ ist die Funktion $\cosh|_{\mathbb{R}}$ nach Satz 2.23 streng wachsend auf $[0, \infty)$ und streng fallend auf $(-\infty, 0]$.[10]

Der folgende Satz gibt ein *hinreichendes* Kriterium für Extremstellen.

Satz 2.25 (Vorzeichenwechsel-Kriterium)
Es seien I ein Intervall, $a \in I$ und $f : I \to \mathbb{R}$ stetig auf I und differenzierbar auf I°. Dann gilt

1. *Existiert ein $\delta > 0$ mit $f' \leq 0$ auf $I \cap (a - \delta, a)$ und $f' \geq 0$ auf $I \cap (a, a + \delta)$ so ist a eine Minimalstelle von f.*
2. *Existiert ein $\delta > 0$ mit $f' \geq 0$ auf $I \cap (a - \delta, a)$ und $f' \leq 0$ auf $I \cap (a, a + \delta)$, so ist a eine Maximalstelle von f.*

Gilt hierbei $f' < 0$ beziehungsweise $f' > 0$, so ist das Extremum strikt.

Beweis. 1. Nach Satz 2.23 ist f wachsend auf $I \cap [a, a+\delta)$ und fallend auf $I \cap (a-\delta, a]$. Damit ist a eine Minimalstelle. Gilt dabei jeweils die strikte Ungleichung, so ist das Minimum strikt.
2. Wieder durch Anwendung von 1. auf $-f$. □

Beispiel 2.26 (vgl. Beispiel 2.24) Nach Beispiel 2.24 und Satz 2.25 sind für $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = 2x^3 + 3x^2 - 1$ der Nullpunkt eine strikte Minimalstelle und -1 eine strikte Maximalstelle. Für $\cosh|_{\mathbb{R}}$ ist 0 eine strikte Minimalstelle. Hier wird $\cosh|_{\mathbb{R}}$ sogar minimal an 0, d. h. die Funktion nimmt an 0 das Minimum an.

[10] Sind $a < 0 < b$, so ist der Graph von $\cosh|_{[a,b]}$ eine sogenannte Kettenlinie. Sie beschreibt den Durchhang einer an ihren Enden $(a, \cosh(a))$ und $(b, \cosh(b))$ aufgehängten Kette unter Einfluss der Schwerkraft.

Eine elegante Methode für die mögliche Berechnung gewisser Grenzwerte ergibt sich aus

Satz 2.27 (Regeln von de l'Hospital)
Es seien $I = (\alpha, \beta) \subset \mathbb{R}$ *ein offenes Intervall und* $f, g : I \to \mathbb{R}$ *differenzierbar mit*

$$f(\alpha^+) = g(\alpha^+) = 0 \quad \textit{oder} \quad g(x) \to \pm\infty \quad (x \to \alpha).$$

Ist g' *nullstellenfrei und gilt*

$$f'(t)/g'(t) \to c \in \mathbb{R} \cup \{\pm\infty\} \qquad (t \to \alpha),$$

so folgt

$$f(x)/g(x) \to c \qquad (x \to \alpha)\ .$$

Eine entsprechende Aussage gilt für Grenzwerte $x \to \beta$.

Beweis. 1. Es gelte $f(\alpha^+) = g(\alpha^+) = 0$. Ist $x \in I$, so existiert nach dem erweiterten Mittelwertsatz (Satz 2.19) ein $\tau(x) \in (\alpha, x)$ mit

$$\frac{f'(\tau(x))}{g'(\tau(x))} = \frac{f(x) - f(\alpha^+)}{g(x) - g(\alpha^+)} = \frac{f(x)}{g(x)}$$

(man beachte: nach dem Mittelwertsatz ist $g(x) - g(\alpha^+) \neq 0$). Dabei gilt $\alpha < \tau(x) \to \alpha$ $(x \to \alpha)$. Also folgt

$$\frac{f'(\tau(x))}{g'(\tau(x))} \to c \quad (x \to \alpha).$$

2. Es gelte $g(x) \to \infty$. Ist $\varepsilon > 0$ gegeben, so existiert ein $\delta > 0$ mit $g(t) > 0$ und $f'(t)/g'(t) \in U_\varepsilon(c)$ für $t \in U_\delta(\alpha)$. Wir wählen ein $s \in U_\delta(\alpha)$. Ist $\alpha < x < s$, so existiert, wieder nach dem erweiterten Mittelwerstz (2.19) ein $\tau(x) \in (x, s)$ mit

$$f'(\tau(x))(g(s) - g(x)) = (f(s) - f(x))g'(\tau(x))\ ,$$

also

$$f(x) = f(s) + \frac{f'(\tau(x))}{g'(\tau(x))}(g(x) - g(s))$$

und folglich

$$\frac{f(x)}{g(x)} = \frac{f(s)}{g(x)} + \frac{f'(\tau(x))}{g'(\tau(x))}\left(1 - \frac{g(s)}{g(x)}\right)\ .$$

Nach Voraussetzung gilt $f(s)/g(x) \to 0$ und $g(s)/g(x) \to 0$ $(x \to \alpha^+)$. Da

$$f'(\tau(x))/g'(\tau(x)) \in U_\varepsilon(c)$$

gilt, existiert ein $\eta > 0$ mit $f(x)/g(x) \in U_{2\varepsilon}(c)$ für $x \in U_\eta(\alpha)$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung.
3. Gilt $g(x) \to -\infty$, so folgt $-g(x) \to \infty$ und damit die Behauptung aus 2. $\square$

Beispiel 2.28 Für alle $\alpha > 0$ gilt

$$\lim_{x\to\infty} \frac{\ln x}{x^\alpha} = 0 \quad \text{und} \quad \lim_{x\to 0^+} x^\alpha \ln x = 0,$$

das heißt, die Logarithmusfunktion wächst langsamer als jede Potenzfunktion $x \mapsto x^\alpha$ mit positivem α. Dies ist eine Variante der Aussage, dass exponentielles Wachstum stärker ist als polynomielles.

Denn: Es gilt $\lim_{x\to\infty} x^\alpha = \infty$ und

$$\lim_{x\to\infty} \frac{x^{-1}}{\alpha x^{\alpha-1}} = \lim_{x\to\infty} \frac{1}{\alpha x^\alpha} = 0 \; ,$$

also folgt die erste Behauptung mit Satz 2.27. Die zweite ergibt sich aus der ersten durch Ersetzen von x durch $1/x$.

Bemerkung und Definition 2.29 Es seien $X \subset \mathbb{K}$ mit $X \subset X'$ und $f : X \to \mathbb{C}$.
1. Mit $f^{(0)} := f$ definiert man höhere Ableitungen rekursiv: Ist $n \in \mathbb{N}$ und $f^{(n-1)}$ differenzierbar, so heißt f **n-mal differenzierbar** und die Funktion

$$f^{(n)} := (f^{(n-1)})' : X \to \mathbb{C}$$

die **n-te Ableitung** von f. Dabei schreibt man meist $f'' := f^{(2)}$ und $f''' := f^{(3)}$. Ist $f^{(n)}$ stetig, so sagt man, f sei n-mal **stetig differenzierbar** und existiert $f^{(n)}$ für alle $n \in \mathbb{N}_0$, so sagt man, f sei beliebig oft differenzierbar.
2. Im Fall $X \subset \mathbb{R}$ schreiben wir $C^n(X)$ für die Menge aller n-mal stetig diffenzierbaren Funktion $f : X \to \mathbb{C}$, wobei $n \in \mathbb{N}_0$, und $C^\infty(X) := \bigcap_{n\in\mathbb{N}_0} C^n(X)$. Damit haben wir die Inklusionskette

$$C^\infty(X) \subset C^{n+1}(X) \subset C^n(X) \subset C^0(X) = C(X) \quad (n \in \mathbb{N}),$$

wobei sämtliche Inklusionen strikt sind ([Ü]).

Ein weiteres *hinreichendes* Kriterium für Extremstellen ist

Satz 2.30 *Es seien $I \subset \mathbb{R}$ ein Intervall, $f : I \to \mathbb{R}$ zweimal stetig differenzierbar und $a \in I$ eine kritische Stelle.*

1. *Gilt $f''(a) > 0$, so ist a eine strikte Minimalstelle.*
2. *Gilt $f''(a) < 0$, so ist a eine strikte Maximalstelle.*

Beweis. 1. Da f'' stetig an a ist, existiert ein $\delta > 0$ mit $f''(x) > 0$ für alle $x \in U_\delta(a) \cap I$. Damit ist f' nach Satz 2.23 streng wachsend auf $U_\delta(a) \cap I$. Aus $f'(a) = 0$ folgt $f' > 0$ auf $I \cap (a, a+\delta)$ und $f' < 0$ auf $I \cap (a-\delta, a)$. Nach dem Vorzeichenwechsel-Kriterium (Satz 2.25) hat f an a ein striktes lokales Minimum.
2. Ergibt sich aus 1. durch Anwendung auf $-f$. $\square$

Beispiel 2.31 Wir betrachten noch einmal das Polynom f aus Beispiel 2.24 und Beispiel 2.26, also $f(x) = 2x^3 + 3x^2 - 1$. Hier ist

$$f''(x) = 12x + 6$$

und damit gilt $f''(0) = 6 > 0$ an der kritischen Stelle 0 und $f''(-1) = -6 < 0$ an der kritischen Stelle -1. Also hat f an 0 ein striktes lokales Minimum und an -1 ein striktes lokales Maximum.

Wir haben bereits gesehen, dass Funktionen, die durch Potenzreihen darstellbar sind, stetig auf dem Konvergenzkreis sind. Wir zeigen nun viel schärfer, dass solche Funktionen tatsächlich beliebig oft differenzierbar sind.

Definition 2.32 Es seien $X \subset \mathbb{K}$ offen und $f : X \to \mathbb{C}$. Dann heißt f **analytisch an der Stelle** $a \in X$, falls eine Umgebung U von a und eine Folge (c_k) in $\mathbb{C}$ so existieren, dass

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

Weiter heißt f **analytisch** (auf X), falls f analytisch an jedem Punkt $x \in X$ ist.

Der folgende Satz zeigt insbesondere, dass analytische Funktionen stets beliebig oft differenzierbar sind.

Satz 2.33 *Es seien $X \subset \mathbb{K}$ offen und $f : X \to \mathbb{C}$ analytisch an $a \in X$. Ist U eine offene Umgebung von a mit $f(x) = \sum_{\nu=0}^{\infty} c_\nu (x-a)^\nu$ für $x \in U$, so ist f beliebig oft differenzierbar auf U und*

$$f^{(k)}(x) = \sum_{\nu=0}^{\infty} (\nu+1)(\nu+2)\cdots(\nu+k)c_{\nu+k}\,(x-a)^\nu \qquad (x \in U,\, k \in \mathbb{N}_0)\,.$$

Insbesondere ist $f^{(k)}(a) = k!c_k$ für $k \in \mathbb{N}_0$, also $f(x) = \sum_{\nu=0}^{\infty} \frac{f^{(\nu)}(a)}{\nu!}(x-a)^\nu$ für $x \in U$.

Beweis. Ist $R > 0$ der Konvergenzradius der Potenzreihe, so gilt $U \subset U_R(a)$ nach Definition des Konvergenzkreises. Ohne Einschränkung können wir $a = 0$ annehmen.
1. Wir zeigen: f ist differenzierbar auf U und

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

Dazu sei $x \in U$. Dann ist $x \in U_r(0)$ für ein $r < R$. Wir wählen ein $\delta > 0$ mit $U_\delta(x) \subset U \cap U_r(0)$. Für $h \in U_\delta(0)$ gilt mit $\phi_n(h) := \sum_{k=0}^{n-1} (x+h)^k x^{n-1-k}$ wie in Beispiel 2.3.1

$$\frac{1}{h}(f(x+h) - f(x)) = \sum_{\nu=1}^{\infty} c_\nu \frac{1}{h}((x+h)^\nu - x^\nu) = \sum_{\nu=1}^{\infty} c_\nu \phi_\nu(h).$$

Ist $\phi := \sum_{\nu=1}^{\infty} c_\nu \phi_\nu$ stetig an 0, so ergibt sich aus

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

die Behauptung. Nach Satz 1.5 reicht es dazu zu zeigen, dass die Funktionenreihe $\sum_{\nu=1}^{\infty} c_\nu \phi_\nu$ gleichmäßig auf $U_\delta(0)$ konvergiert: Aus $|x| < r$ und $|x+h| < r$ folgt $|\phi_n(h)| \leq nr^{n-1}$. In Bemerkung 1.11 haben wir gesehen, dass ein $q < 1$ existiert mit $|c_n|r^n \leq q^n$ für n genügend groß. Wegen $n^{1/n} \to 1$ für $n \to \infty$ konvergiert nach dem Wurzelkriterium (siehe Einführung in die Mathematik) damit $\sum_{\nu=1}^{\infty} \nu|c_\nu|r^{\nu-1}$. Aus dem Weierstraß-Kriterium (Satz 1.8) folgt damit die gleichmäßige Konvergenz von $\sum_{\nu=1}^{\infty} c_\nu \phi_\nu$ auf $U_\delta(0)$.

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

$$f^{(k)}(x) = \sum_{\nu=0}^{\infty} (\nu+1)(\nu+2)\ldots(\nu+k)c_{\nu+k}(x-a)^\nu \,.$$

Die Zusatzbehauptung $k!c_k = f^{(k)}(a)$ ergibt sich für $x = a$. □

Beispiel 2.34 1. 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 nach Satz 2.33. Damit sind auch sin, cos, sinh, cosh (als Linearkombinationen analytischer Funktionen) analytisch auf $\mathbb{C}$.

2. Mit Satz 2.27 (de l'Hospital) sieht man, dass $\sin(x)/x \to 1$ für $x \to 0$ in $\mathbb{R}$ gilt. Unter Verwendung des Satzes 2.33 lässt sich viel mehr aussagen: Die Funktion $f : \mathbb{C} \to \mathbb{C}$ mit

$$f(z) = \sum_{\mu=0}^{\infty} \frac{(-1)^\mu z^{2\mu}}{(2\mu+1)!} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} \ldots = \begin{cases} \sin(z)/z\,, & z \neq 0 \\ 1\,, & z = 0 \end{cases}$$

ist analytisch an 0 und es gilt für $\mu \in \mathbb{N}_0$

$$f^{(2\mu)}(0) = \frac{(-1)^\mu}{2\mu+1} \qquad \text{und} \qquad f^{(2\mu+1)}(0) = 0.$$

Bemerkung und Definition 2.35 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 $X \subset \mathbb{K}$ sternförmig und sind F und G Stammfunktionen zu f auf X, so ist $(F-G)' = 0$ auf X, und damit ist $F - G$ nach Bemerkung 2.22 konstant auf X, mit anderen Worten, F und G unterscheiden sich lediglich durch eine additive Konstante.

Bemerkung 2.36 Ist $\sum\limits_{\nu=0}^{\infty} c_\nu (x-a)^\nu$ eine Potenzreihe mit Konvergenzradius $R > 0$ und ist

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

so ist nach Satz 2.33 durch

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

eine Stammfunktion zu f auf $U_R(a)$ definiert.[11]

Beispiel 2.37 1. Ist $f(z) := 1/(1-z)$ für $z \in \mathbb{C} \setminus \{1\}$, so ist nach Bemerkung 2.36 wegen $f(z) = \sum\limits_{\nu=0}^{\infty} z^\nu$ 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 $(-\infty, 1)$ ist durch

$$G(x) - \ln(1-x) \qquad (x < 1)$$

ebenfalls eine Stammfunktion zu f gegeben ist. Nach Bemerkung 2.35 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

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

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

$$f(z) := e^{-z^2} = \sum_{\mu=0}^{\infty} \frac{(-1)^\mu}{\mu!} z^{2\mu} \qquad (z \in \mathbb{C}),$$

so ist durch

$$F(z) = \sum_{\mu=0}^{\infty} \frac{(-1)^\mu}{\mu!(2\mu+1)} z^{2\mu+1} \qquad (z \in \mathbb{C})$$

eine Stammfunktion zu f auf $\mathbb{C}$ gegeben. Die Funktion $\operatorname{erf} := (2/\sqrt{\pi})F$ heißt **Fehlerfunktion** (error function).[12]

Am Rande ihres Konvergenzkreises können Funktionen, die durch Potenzreihen definiert sind, ein sehr kompliziertes Verhalten zeigen. Konvergiert die Potenzreihe an einem Randpunkt ζ des Konvergenzkreises, so existiert jedenfalls der sogenannte radiale Randwert der

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

[12] Die Fehlerfunktion ist von zentraler Bedeutung in der Statistik aufgrund ihrer direkten Beziehung zu den Verteilungsfunktionen der Normalverteilungen.

Grenzfunktion an der Stelle ζ. Wir formulieren das entsprechende Ergebnis für Potenzreihen mit Entwicklungsmitte 0 und Konvergenzradius 1. Der allgemeine Fall kann leicht darauf zurückgeführt werden.

Satz 2.38 (Abelscher Grenzwertsatz)
Es sei $\sum\limits_{\nu=0}^{\infty} c_\nu z^\nu$ *eine Potenzreihe mit Konvergenzradius* 1 *und* $f(x) := \sum\limits_{\nu=0}^{\infty} c_\nu x^\nu$ *für* $|x| < 1$. *Ist* ζ *mit* $|\zeta| = 1$ *so, dass die Reihe* $\sum\limits_{\nu=0}^{\infty} c_\nu \zeta^\nu$ *konvergiert, so gilt*

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

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

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

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 $1 = (1-r)\sum\limits_{\nu=0}^{\infty} r^\nu$

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

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

Beispiel 2.39 Nach Beispiel 2.37 ist

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

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

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

3 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 $|I| := \operatorname{diam}(I)$[13] und nennen $|I|$ die **Länge** von I.

Bemerkung und Definition 3.1 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

$$|J| = \sum_{I \in E} |I \cap J|. \tag{3.1}$$

2. Eine Funktion $\varphi \in B[a, \beta] := B([a, b])$ heißt **Treppenfunktion** (auf $[a, b]$), falls eine Zerlegung E von $[a, b]$ und für $I \in E$ Konstanten $c(I) = c_\varphi(I) \in \mathbb{C}$ existieren mit

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

also so, dass φ konstant mit Wert $c(I)$ auf I ist. Eine Zerlegung, zu der entsprechende Konstanten $c(I)$ existieren, nennen wir **zulässig** für die Funktion φ. Wir schreiben $T[a, b]$ für die Menge der Treppenfunktionen auf $[a, b]$.

Beispiel 3.2 Wir betrachten $[a, b] = [0, 1]$ und die Funktion $\varphi = 1_{(1/2,1]}$. Dann ist φ eine Treppenfunktion und etwa

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

eine zulässige Zerlegung für φ, 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.

Bemerkung und Definition 3.3 Es seien E und F Zerlegungen von $[a, b]$. Dann heißt

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

[13] 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.

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 φ, so ist auch $E \wedge F$ zulässig für φ, und ist auch F zulässig für φ, 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 (3.1) 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 φ, so heißt

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

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

Beispiel 3.4 In der Situation von Beispiel 3.2 gilt

$$\int_0^1 \varphi = \sum_{I\in E} c(I)\cdot|I| = \frac{1}{2}$$

Definition 3.5 Es seien $X \subset \mathbb{K}$ und $L \subset \mathrm{Abb}(X,\mathbb{C})$ ein Unterraum. Ist $\ell : L \to \mathbb{C}$ eine lineare Abbildung, so sagen wir, ℓ sei **nichtnegativ**, falls $\ell(f) \geq 0$ für alle f mit $f \geq 0$ gilt. Aufgrund der Linearität ist in diesem Fall ℓ auch **monoton** in dem Sinne, dass $\ell(f) \leq \ell(g)$ für alle reellwertigen f, g mit $f \leq g$ gilt.

Satz 3.6 *Die Abbildung $\int_a^b : T[a,b] \to \mathbb{C}$ ist linear und nichtnegativ, also auch 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 φ, ψ Treppenfunktionen und $\lambda \in \mathbb{C}$. Sind E beziehungsweise F zulässige Zerlegungen für φ beziehungsweise ψ, so ist die gemeinsame Verfeinerung $E \wedge F$ sowohl für φ als auch für ψ zulässig. Sind $c_\varphi(K) \in \mathbb{C}$ beziehungsweise $c_\psi(K) \in \mathbb{C}$ für $K \in E \wedge F$ wie in Bemerkung 3.1, 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 Unterraum von $B[a,b]$. Außerdem gilt

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

Die Nichtnegativität und 1. ergeben sich unmittelbar aus der Definition und entsprechenden Eigenschaften von Summen. 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 3.7 Eine Funktion $f \in B[a,b]$ heißt **Regelfunktion** (auf $[a,b]$), falls eine Folge (φ_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.[14]

Für $[a,b] \subset \mathbb{R}$, $n \in \mathbb{N}$ und $j = 0,\dots,n$ 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 $j = 1,\dots,n$ durch $E_n := E_n[a,b] := \{I_{j,n} : j = 0,\dots,n\}$ eine Zerlegung von $[a,b]$ gegeben. Hier gilt $|I_{j,n}| = (b-a)/n$ für $j = 1,\dots,n$ (und $|I_{0,n}| = 0$).

Satz 3.8 *Ist $f \in C[a,b]$ so konvergiert die Folge (φ_n) in $T[a,b]$, definiert durch*

$$\varphi_n(t) := f(t_{j,n}) \qquad (t \in I_{j,n},\ j = 0,\dots,n),$$

gleichmäßig gegen f. Insbesondere ist $C[a,b] \subset R[a,b]$.

Beweis. Es sei $\varepsilon > 0$ gegeben. Da $[a,b]$ kompakt ist, ist f gleichmäßig stetig (Einführung in die Mathematik). Daher existiert ein $\delta > 0$ mit $|f(t) - f(s)| < \varepsilon$ für $|t-s| < \delta$.
Es sei $n > (b-a)/\delta$. Dann gilt $|I_{j,n}| \le (b-a)/n < \delta$. Ist $t \in [a,b]$, so existiert ein $j \in \{0,\dots,n\}$ mit $t \in I_{j,n}$ und damit wegen $|t - t_{j,n}| < \delta$

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

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

Beispiel 3.9 Wir betrachten $f(t) = t$ auf $[0,1]$. Dann ist nach Satz 3.8 wegen $t_{j,n} = j/n$ durch $\varphi_n(0) = 0$ und $\varphi_n(t) := j/n$ für $t \in I_{j,n}$ und $j = 1,\dots,n$ eine Folge von Treppenfunktionen auf $[0,1]$ gegeben mit $\varphi_n \to f$ gleichmäßig auf $[0,1]$.[15]

[14] Mit anderen Worten: $R[a,b]$ ist der Abschluss von $T[a,b]$ im Banachraum $(B[a,b], ||\cdot||_\infty)$ und damit ebenfalls ein Banachraum.

[15] Der Beweis zeigt, dass die Aussage auch gilt, wenn man bei der Definition der φ_n statt der rechten Randpunkte $t_{j,n}$ beliebige Punkte $\tau_{j,n} \in I_{j,n}$ wählt.

Bemerkung und Definition 3.10 Es seien f eine Regelfunktion und (φ_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 3.6
$$\Big|\int_a^b \varphi_n - \int_a^b \varphi_{n'}\Big| = \Big|\int_a^b (\varphi_n - \varphi_{n'})\Big| \le ||\varphi_n - \varphi_{n'}||_\infty (b-a)\,,$$
und da (φ_n) eine Cauchyfolge in $B[a,b]$ ist, ist auch $(\int_a^b \varphi_n)$ eine Cauchyfolge in $\mathbb{C}$, also konvergent.

2. Ist (ψ_n) eine weitere Folge von Treppenfunktionen mit $\psi_n \to f$ gleichmäßig auf $[a,b]$, so gilt
$$\Big|\int_a^b \varphi_n - \int_a^b \psi_n\Big| \le ||\varphi_n - \psi_n||_\infty (b-a) \le \big(||\varphi_n - f||_\infty + ||f - \psi_n||_\infty\big)(b-a) \to 0\,,$$
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
$$\int_b^a f := -\int_a^b f.$$

Beispiel 3.11 Es seien f und φ_n wie in Beispiel 3.9. Dann 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
$$\int_0^1 f = \int_0^1 t\,dt = \frac{1}{2}.$$

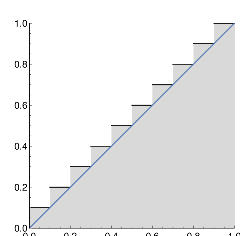

Abbildung 3: Treppenfunktion φ_{10} und $\int_0^1 \varphi_{10}$ als Approximation von $\int_0^1 t\,dt$.

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

Satz 3.12 *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*

$$\Big|\int_a^b f\Big| \leq \int_a^b |f| \leq (b-a)\sup_{[a,b]}|f|.$$

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

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

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

für $n \to \infty$. Also ist $\lambda f + g \in R[a,b]$ und mit Satz 3.6 gilt

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

Es sei nun $f \geq 0$. Wir setzen $\varphi_n^+ := \varphi_n + ||f - \varphi_n||_\infty$. Dann sind $\varphi_n^+ \in T[a,b]$ mit $\varphi_n^+ \geq f \geq 0$ sowie $||f - \varphi_n^+||_\infty \to 0$. Also folgt mit Satz 3.6

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

1. Aus $\big|\,|f| - |\varphi_n|\,\big| \leq |f - \varphi_n|$ folgt, dass auch $|f|$ eine Regelfunktion ist und dass $|\varphi_n| \to |f|$ gleichmäßig auf $[a,b]$ gilt. Mit Satz 3.6.1 ergibt sich

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

Wegen der Monotonie von $\int_a^b$ folgt aus $|f| \leq \|f\|_\infty$ auch $\int_a^b |f| \leq \|f\|_\infty(b-a)$.
2. Die Aussage ergibt sich in ähnlicher Weise aus Satz 3.6.2, angewandt auf φ_n. □

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

Bemerkung und Definition 3.13 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 3.12.2

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

zunächst im Fall $u \leq v \leq w$, 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 3.14 Sind $I = \mathbb{R}$ und $f = 1_{[0,\infty)}$, so gilt

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

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

Satz 3.15 (Hauptsatz über Integralfunktionen) [16]
Es seien I ein Intervall, $f \in R(I)$ und $u \in I$. Dann ist die Integralfunktion $Vf = V_u f$ stetig auf I.[17] *Außerdem gilt: Ist f stetig an der Stelle $a \in I$, so ist Vf differenzierbar an a mit*

$$(Vf)'(a) = f(a)\ .$$

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. Also ist f beschränkt auf J. Für $h \in J - x$ gilt dann

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

Es sei nun f stetig an der Stelle a. Dann folgt aus $\int_a^{a+h} f(a)\,dt = f(a)h$

$$\Big| \frac{(Vf)(a+h) - (Vf)(a)}{h} - f(a) \Big| = \Big| \frac{1}{h} \int_a^{a+h} (f - f(a)) \Big| \leq \sup_{[a,a+h]} |f - f(a)| \to 0$$

für $h \to 0$. □

Bemerkung und Definition 3.16 Nach Satz 3.15 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 3.14 zeigt.

[16] wird auch als Hauptsatz der Differenzial- und Integralrechnung, Teil 1, bezeichnet

[17] Die lineare Abbildung $V : R(I) \to C(I)$ nennt man Volterra-Operator auf $R(I)$; daher das V.

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

Satz 3.17 (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 3.16 auch $V_u f$ eine Stammfunktion zu f auf I. Nach Bemerkung 2.35 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 3.18 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 $a \in \mathbb{C} \setminus \{-1\}$ und $f(t) = t^a$ für $t > 0$. Dann ist $t \mapsto \frac{1}{a+1} t^{a+1}$ eine Stammfunktion zu f auf $(0, \infty)$ und folglich ist für $0 < u, v < \infty$

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

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

Satz 3.19 (Substitutionsregel) *Es sei I ein Intervall. Sind $\gamma : I \to \mathbb{R}$ stetig differenzierbar und $f : \gamma(I) \to \mathbb{C}$ stetig, so gilt für $u, v \in I$*

$$\int_u^v f(\gamma(t))\gamma'(t)\, dt = \int_u^v (f \circ \gamma)\gamma' = \int_{\gamma(u)}^{\gamma(v)} f \,.$$

Beweis. Nach dem Zwischenwertsatz ist $\gamma(I)$ ein Intervall und damit existiert nach Bemerkung 3.16 eine Stammfunktion F zu f auf $\gamma(I)$. Außerdem ist $F \circ \gamma$ nach der Kettenregel eine Stammfunktion zur auf I stetigen Funktion $(f \circ \gamma)\gamma'$. Nach dem Hauptsatz der Differenzial- und Integralrechnung haben beide Integrale den Wert $F(\gamma(v)) - F(\gamma(u))$. □

Beispiel 3.20 Mit $\gamma(t) := 1 + t^2$ auf $\mathbb{R}$ und $f(s) := 1/\sqrt{s}$ auf $(0, \infty)$ gilt nach der Substitutionsregel für $u, v \in \mathbb{R}$

$$\int_u^v \frac{t}{\sqrt{1+t^2}}\,dt = \frac{1}{2}\int_u^v \frac{\gamma'(t)}{\sqrt{\gamma(t)}}\,dt = \frac{1}{2}\int_{1+u^2}^{1+v^2} \frac{1}{\sqrt{s}}\,ds = \sqrt{s}\Big|_{1+u^2}^{1+v^2}\,.$$

Bemerkung 3.21 (**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 3.22 Für $a \neq -1$ und $u, v > 0$ gilt mit partieller Integration, angewandt auf $f(t) = t^a$ und $F(t) = t^{a+1}/(a+1)$

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

Insbesondere ist

$$\int_u^v \ln t\,dt = \int_u^v 1\cdot\ln t\,dt = t(\ln(t) - 1)\Big|_u^v.$$

Bemerkung 3.23 Sind X sternförmig bzgl. a und f stetig differenzierbar auf X, so ist für $h \in X - a$ die Funktion $f \circ s_a^{a+h}$ eine Stammfunktion zu $h(f' \circ s_a^{a+h})$ auf X und damit nach dem Hauptsatz der Differenzial- und Integralrechnung (Satz 3.17)

$$f(a+h) = f(a) + h\int_0^1 f'(a+th)\,dt.$$

Ist a ein innerer Punkt von X und ist f analytisch an a, so gilt nach Satz 2.33 für $|h|$ genügend klein

$$f(a+h) = \sum_{\nu=0}^{\infty} \frac{f^{(\nu)}(a)}{\nu!}h^\nu.$$

Der folgende Satz, der insbesondere für die numerische Mathematik von großer Bedeutung ist, kann als eine Art Brücke zwischen den beiden obigen Aussagen angesehen werden.

Satz 3.24 (**Taylor**)
Es seien $X \subset \mathbb{K}$ sternförmig bezüglich a und $n \in \mathbb{N}_0$. Dann gilt für alle $(n+1)$-mal stetig differenzierbaren Funktionen f auf X

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

Beweis. Wir beweisen die Behauptung per Induktion nach n. Für $n = 0$ ist die Aussage Inhalt der Bermerkung 3.23. 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 3.25 In der Situation von Satz 3.24 wird $f^{(k)}(a)/k!$ für $k = 0, \ldots, n$ 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**. Ist f sogar beliebig oft differenzierbar, so heißt die Funktionenreihe $\sum\limits_{\nu=0}^{\infty} \frac{f^{(\nu)}(a)}{\nu!} h^\nu$ die **Taylor-Reihe** von f bezüglich a. Aus Satz 2.33 ergibt sich: f ist analytisch an a genau dann, wenn $f(a+h)$ für $|h|$ genügend klein mit der Taylor-Reihe von f bezüglich a übereinstimmt.

Bemerkung 3.26 Sind $g, w : [0,1] \to \mathbb{R}$ stetig und $w \geq 0$, so gilt nach Satz 3.12

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

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 $w(t) = (1-t)^n$

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

für $h \in X - a$. Ist dabei f reellwertig, 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**.

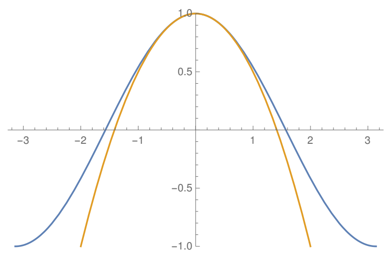

Abbildung 4: cos und $T_3(\cos, 0)$ auf $(-\pi, \pi)$.

Beispiel 3.27 Für $f = \cos$ gilt $T_3(f,0)(h) = T_2(f,0)(h) = 1 - h^2/2$. Wegen $\cos^{(4)} = \cos$ folgt aus Bemerkung 3.26

$$\left|\cos(h) - 1 + \frac{h^2}{2}\right| = \left|R_3(f,0)(h)\right| \leq \frac{|h|^4}{24} \qquad (h \in \mathbb{R}).$$

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

Bemerkung und Definition 3.28 Es seien I ein Intervall und $\alpha := \inf I$, $\beta := \sup I$.[18] Eine Funktion $f \in R(I)$ heißt **integrierbar** auf I, falls $(V_u f)(\alpha^+)$ und $(V_u f)(\beta^-)$ für ein $u \in I$ existieren. In diesem Fall existieren die beiden Grenzwerte für jedes $u \in I$, und die Differenz $(V_u f)(\beta^-) - (V_u f)(\alpha^+)$ ist nach Bemerkung 3.13 unabhängig von u. Die Zahl

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

heißt dann **uneigentliches Integral** von f auf I.[19] Aus Grenzwertsätzen und Satz 3.12 folgt leicht, dass durch $f \mapsto \int_{\alpha^+}^{\beta^-} f$ eine lineare und monotone Abbildung auf dem Raum der auf I integrierbaren Funktionen definiert ist.

Bemerkung 3.29 1. Ist $\beta \in I$, so gilt $(V_u f)(\beta) = (V_u f)(\beta^-)$ nach dem Hauptsatz über Integralfunktionen. Entsprechend ist $(V_u f)(\alpha) = (V_u f)(\alpha^+)$ im Falle $\alpha \in I$. Also ist im Falle $I = [\alpha, \beta]$

$$\int_{\alpha^+}^{\beta^-} f = (V_u f)(\beta^-) - (V_u f)(\alpha^+) = \int_{\alpha}^{\beta} f,$$

das heißt, „eigentliches“ und uneigentliches Integral stimmen überein. Daher spricht man auch wieder kurz vom **Integral** von f und schreibt auch β statt β^- (entsprechend für α).

[18] Dabei sind $\alpha = -\infty$ und $\beta = \infty$ sind zugelassen.

[19] Ähnlich wie bei Reihen spricht man im Falle der Existenz auch davon, dass das Integral $\int_{\alpha^+}^{\beta^-} f$ konvergiert.

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 $F(\beta^-)$ und $F(\alpha^+)$ existieren. Außerdem gilt dann

$$\int_{\alpha^+}^{\beta^-} f = F(\beta^-) - F(\alpha^+) =: F(t)|_\alpha^\beta.$$

Beispiel 3.30 1. Für $a \in \mathbb{R}$ sei $f_a(t) := t^{-a}$ auf $I = (0, \infty)$. Dann ist durch

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

eine Stammfunktion F_a zu f_a auf I definiert. Außerdem erhalten wir für $t \to \infty$

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

und für $t \to 0^+$

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

Also ist f_a nach Bemerkung 3.29.2 genau dann integrierbar auf $[1, \infty)$, wenn $a > 1$ ist, und in diesem Falle ist

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

Entsprechend ist f_a genau dann integrierbar auf $(0, 1]$, wenn $a < 1$ ist, und dann gilt

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

Hieraus folgt auch, dass f_a für kein $a \in \mathbb{R}$ 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 3.29.2 integrierbar und es gilt

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

Satz 3.31 (**Uneigentliche partielle Integration**)
Es seien $I \subset \mathbb{R}$ ein Intervall, $\alpha := \inf I$, $\beta := \sup I$ und $f \in C(I)$ sowie $g \in C^1(I)$. Ist F eine Stammfunktion zu f auf I und existieren $(Fg)(\beta^-)$ sowie $(Fg)(\alpha^+)$, so gilt: fg ist genau dann integrierbar auf I, wenn Fg' integrierbar auf I ist, und in diesem Fall ist

$$\int_{\alpha^+}^{\beta^-} fg = Fg\big|_\alpha^\beta - \int_{\alpha^+}^{\beta^-} Fg'.$$

Beweis. Fg ist eine Stammfunktion zur (stetigen) Funktion $fg + Fg'$. Nach Bemerkung 3.29.2 ist $fg + Fg'$ integrierbar mit $\int_{\alpha^+}^{\beta^-} (fg + Fg') = Fg\big|_\alpha^\beta$. Ist nun etwa Fg' integrierbar, so ist auch $fg = (fg + Fg') - Fg'$ integrierbar und es gilt

$$\int_{\alpha^+}^{\beta^-} fg = Fg\big|_\alpha^\beta - \int_{\alpha^+}^{\beta^-} Fg'.$$

Entsprechendes gilt, falls fg integrierbar ist. □

Beispiel 3.32 (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 3.31 (da $t^2 = 1 - (1 - t^2)$)

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

und damit

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

Satz 3.33 (**Majorantenkriterium für Integrale**)
Es sei I ein Intervall und $\alpha := \inf I$, $\beta := \sup I$. Ist $f \in R(I)$ und ist g integrierbar auf I mit $|f| \le g$, so ist auch f integrierbar auf I mit

$$\Big| \int_{\alpha^+}^{\beta^-} f \Big| \le \int_{\alpha^+}^{\beta^-} g\,.$$

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

$$|F(t) - F(s)| = \Big| \int_s^t f \Big| \le \int_s^t |f| \le \int_s^t g = G(t) - G(s) \tag{3.2}$$

und damit $|F(t) - F(s)| \le |G(t) - G(s)|$ für beliebige $s, t \in I$.
Es sei nun (t_n) eine Folge in I mit $\beta > t_n \to \beta$. Dann ist $|F(t_n) - F(t_{n'})| \le |G(t_n) - G(t_{n'})|$ 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 $F(\beta^-)$. Genauso sieht man, dass $F(\alpha^+)$ existiert. Aus (3.2) folgt dann auch $|F(\beta^-) - F(\alpha^+)| \le G(\beta^-) - G(\alpha^+)$. □

Bemerkung und Definition 3.34 Insbesondere ergibt sich aus Satz 3.33 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 3.33

$$\Big|\int_{\alpha^+}^{\beta^-} f\Big| \leq \int_{\alpha^+}^{\beta^-} |f|\,.$$

Beispiel 3.35 Für $a > 1$ betrachten wir die Funktion $f : [1, \infty) \to \mathbb{R}$ mit

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

Es gilt $|\cos t|t^{-a} \leq t^{-a}$ für $t \geq 1$. Da $t \mapsto t^{-a}$ nach Beispiel 3.30.1 integrierbar ist, folgt die absolute Integrierbarkeit von f aus dem Majorantenkriterium (Satz 3.33). Entsprechendes gilt für die Funktion $t \mapsto \sin(t)/t^a$ auf $[1, \infty)$.

Bemerkung 3.36 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 3.31.

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.[20]

Satz 3.37 (**Integralkriterium**)
Es sei $f : [1, \infty) \to \mathbb{R}$ fallend und $f \geq 0$. Dann ist die Folge (s_n) mit $s_n := \sum_{\nu=1}^{n} f(\nu) - \int_1^{n+1} f$ wachsend und beschänkt mit $0 \leq \lim s_n \leq f(1)$.

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

$$0 \leq a_n - \int_n^{n+1} f \leq a_n - a_{n+1}.$$

Also ist

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

wachsend mit $0 \leq s_n \leq a_1 - a_{n+1} \leq a_1$. Nach dem Hauptsatz über monotone Folgen ist (s_n) konvergent mit $0 \leq \lim s_n \leq a_1$. □

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

Beispiel 3.38 Es seien $a > 0$ und $f(t) := t^{-a}$ für $t \geq 1$. Dann ist f fallend auf $[1, \infty)$ und $f \geq 0$. Nach Satz 3.37 ist (s_n) mit $s_n := \sum\limits_{\nu=1}^{n} \nu^{-a} - \int_1^{n+1} t^{-a}$ konvergent und $\lim s_n \in [0,1]$ wegen $f(1) = 1$. Im Fall $a > 1$ konvergieren neben (s_n) auch die beiden Summandenfolgen mit Grenzwerten $\int_1^\infty t^{-a}\,dt = 1/(a-1)$ und $\zeta(a) = \sum\limits_{\nu=1}^{\infty} \nu^{-a}$. Also ist $0 \leq \zeta(a) - 1/(a-1) \leq 1$. Ist $a = 1$, so gilt $\int_1^{n+1} t^{-1}\,dt = \ln(n+1)$ und damit konvergiert

$$s_n = \sum_{\nu=1}^{n} \frac{1}{\nu} - \ln(n+1) \; .$$

Der Grenzwert $\lim s_n \in [0,1]$ heißt **Euler-Mascheroni Konstante**.[21] Ist $0 < a < 1$, so ergibt sich entsprechend die Konvergenz von

$$s_n = \sum_{\nu=1}^{n} \nu^{-a} - \int_1^{n+1} t^{-a}\,dt = \sum_{\nu=1}^{n} \nu^{-a} - \frac{(n+1)^{1-a} - 1}{1-a} \; .$$

Satz 3.39 *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* $(0, \infty)$.

Beweis. Wir setzen $f(t) := e^{-t} t^{z-1}$ für $t > 0$. Aus $t^{z+1} e^{-t} \to 0$ für $t \to \infty$ folgt, dass $t \mapsto |e^{-t} t^{z+1}|$ maximal auf $[1, \infty)$ wird. Also existiert eine Konstante $M > 0$ mit $|f(t)| \leq M/t^2$ für alle $t \in [1, \infty)$. Aus dem Majorantenkriterium, wieder angewandt mit $g(t) = 1/t^2$, folgt die absolute Integrierbarkeit von f auf $[1, \infty)$. Weiter gilt $|f(t)| \leq t^{\mathrm{Re}(z)-1}$ für $t \in (0,1]$. Ist $\mathrm{Re}(z) > 0$, so ist $t \mapsto t^{\mathrm{Re}(z)-1}$ nach Beispiel 3.30.1 integrierbar auf $(0,1]$. Wieder mit dem Majorantenkriterium ergibt sich die absolute Integrierbarkeit von f auf $(0,1]$ und damit auch auf $(0, \infty)$. $\square$

Bemerkung und Definition 3.40 Es sei $\mathbb{C}_+ := \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\}$ die offene rechte Halbebene. Die Funktion $\Gamma : \mathbb{C}_+ \to \mathbb{C}$ mit

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

heißt **(Eulersche) Gammafunktion**. Durch uneigentliche partielle Integration erhält man

$$\Gamma(z+1) = z \cdot \Gamma(z) \qquad (\mathrm{Re}(z) > 0) \; . \tag{3.3}$$

Speziell gilt $\Gamma(1) = \int_0^\infty e^{-t} dt = -e^{-t}\big|_0^\infty = 1$, woraus sich wiederum mit (3.3) 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.

[21] Zur Bedeutung siehe etwa `https://de.wikipedia.org/wiki/Euler-Mascheroni-Konstante`.

4 Mehrdimensionale Differenzialrechnung

Wir wollen jetzt Ableitungen für Funktionen mehrerer, meist reeller Variablen definieren und untersuchen. Dazu betrachten wir zunächst allgemeiner Banachräume V, E über $\mathbb{K}$ und Abbildungen $f : X \to E$, wobei $X \subset V$. Ist V ein normierter Raum über $\mathbb{K}$ mit Norm $|\cdot| = |\cdot|_V$ so schreiben wir

$$B := B_V := \{x \in V : |x|_V \leq 1\}.$$

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

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

Bemerkung und Definition 4.1 Esa seien V, E normierte Räume. Eine lineare Abbildung $A : V \to E$ heißt **beschränkt**, falls $A|_{B_V}$ beschränkt ist. Wir schreiben $L(V, E)$ für den Raum der beschränkten linearen Abbildungen von V nach E. Außerdem schreiben wir bei linearen Abbildungen oft kurz $Ax := A(x)$. Indem man A mit $A|_{B_V}$ identifiziert, kann man $L(V, E)$ als Unterraum von $B(B_V, E)$ auffassen[22] und dann ist durch

$$\|A\| := \sup_{x \in B_V} |Ax| = \|A|_{B_V}\|_\infty$$

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

Satz 4.2 *Es seien V, W, E normierte Raüme und $A \in L(V, E)$, $B \in L(E, W)$.*

1. *Für $x \in V$ ist $|Ax| \leq \|A\| \cdot |x|$.*[23]
2. *Mit $BA := B \circ A$ gilt $\|BA\| \leq \|B\| \cdot \|A\|$.*

Beweis. 1. Ohne Einschränkung sei $x \neq 0$. Dann gilt

$$|Ax| = \left|A\Big(|x| \frac{x}{|x|}\Big)\right| = |x| \left|A\Big(\frac{x}{|x|}\Big)\right| \leq \|A\| \cdot |x|.$$

2. Nach 1. gilt $|BAx| \leq \|B\| \cdot |Ax| \leq \|B\| \cdot \|A\|$ für $|x| \leq 1$. □

Im Weiteren seinen V, W, E stets Banachräume über (dem gleichen) $\mathbb{K}$, wenn nicht anderes gesagt ist.

Sind $X \subset V$, $a \in X'$ und $f : X \to E$, so sagen wir f sei **lokal beschränkt** an a, falls $M, \delta > 0$ existieren mit $|f(x)| \leq M$ für alle $x \in X$ mit $0 < |x - a| < \delta$. Weiter sagen wir f sei **abklingend** an a, falls zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $|f(x)| < \varepsilon$ für alle $x \in X$ mit $0 < |x - a| < \delta$.[24]

[22] Dabei ist $B(M, E)$ der Raum der beschränkten Funktionen von M nach E.

[23] Dies zeigt, dass beschränkte lineare Abbildungen stetig sind. Man sieht leicht, dass tatsächlich Stetigkeit äquivalent zu Beschränktheit ist.

[24] Ist $c \in E$, so schreiben wir kurz $f(x) \to c$ für $x \to a$, falls $f - c$ abklingend an a ist.

Bemerkung und Definition 4.3 Es seien $X \subset V$ und $f : X \to E$.
1. Ist $a \in X^\circ$, also a innerer Punkt von X, so heißt f **(Fréchet-)differenzierbar** an der Stelle a, falls ein $A = A_{f,a} \in L(V, E)$ und eine an 0 abklingende Funktion $\varepsilon = \varepsilon_{f,a} : X_a := (X - a) \setminus \{0\} \to E$ existieren mit

$$(\tau_a f)(h) := f(a + h) - f(a) = Ah + |h| \cdot \varepsilon(h) \quad (h \in X_a),$$

mit anderen Worten, falls $|\cdot|^{-1}(\tau_a f - A)$ abklingend an 0 ist (vgl. Satz 2.5, affin-lineare Approximation). Man sieht leicht, dass A im Falle der Differenzierbarkeit von f an a eindeutig bestimmt ist. Man nennt A die **(Fréchet-)Ableitung** von f an a und schreibt $f'(a) := A$. Dabei ist zu beachten, dass man im schon in Satz 2.5 betrachteten skalaren Fall $V = \mathbb{K}$ und $E = \mathbb{C}$ (als Vektorraum über $\mathbb{K}$) die Konstante $c \in \mathbb{C}$ mit der linearen Abbildung $A : \mathbb{K} \to \mathbb{C}$, definiert durch $Ah := h \cdot c$, identifiziert (also $c = A1$).
2. 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).[25] 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 $C^1(X) := C^1(X, \mathbb{C}) = C^1(X, \mathbb{R}^2)$.

Beispiel 4.4 1. (affin-lineare Abbildungen) Es seien $A \in L(V, E)$ und $c \in E$. Ist $f : V \to E$ definiert durch $f(x) := Ax + c$, so ist $f'(x)h = Ah$ für alle $x, h \in V$, also kurz $f'(x) = A$ für alle $x \in V$.
2. (quadratische Formen) Es seien V ein reeller Hilbertraum und $A \in L(V) := L(V, V)$. Ist $f : V \to \mathbb{R}$ definiert durch $f(x) := \langle x, Ax\rangle$, so gilt für $x, h \in V$

$$(\tau_x f)(h) = \langle x + h, A(x + h)\rangle - \langle x, Ax\rangle = \langle x, Ah\rangle + \langle h, Ax\rangle + \langle h, Ah\rangle.$$

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

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

ist, folgt $f'(x)h = \langle x, Ah\rangle + \langle h, Ax\rangle$ für $x, h \in V$, also kurz

$$f'(x) = \langle x, A\rangle + \langle \cdot, Ax\rangle \qquad (x \in V).$$

Satz 4.5 *Es seien $X \subset V$, $a \in X^\circ$ und $f : X \to E$ differenzierbar an a. Dann gilt*

1. $|\cdot|^{-1}\tau_a f$ *ist lokal beschränkt an* 0 *und* f *ist stetig an* a.
2. ***(Linearität der Ableitung)****: Sind $g : X \to E$ differenzierbar an a und $\lambda \in \mathbb{K}$, so ist auch $\lambda f + g$ differenzierbar an a mit $(\lambda f + g)'(a) = \lambda f'(a) + g'(a)$.*
3. ***(Kettenregel)*** *Ist $g : Y \to W$ mit $f(X) \subset Y \subset E$ differenzierbar an $f(a)$, so ist $g \circ f$ differenzierbar an a mit $(g \circ f)'(a) = (g'(f(a))f'(a)$.*

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

Beweis. 1. Nach Voraussetzung ist $\varepsilon = \varepsilon_{f,a}$ abklingend an a. Wegen

$$|h|^{-1}|\tau_a f(h)| \leq |\varepsilon(h)| + |h|^{-1}|Ah| \leq |\varepsilon(h)| + ||A|| \quad (h \in X_a)$$

ist $|\cdot|^{-1}\tau_a f$ lokal beschränkt an 0, also $\tau_a f$ abklingend an 0 und damit f stetig an a.
2. 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. Es sei zuächst speziell $a = 0$, $f(0) = 0$ und $= g(0) = 0$. Dann ist mit $\varepsilon := \varepsilon_{g,0}$ und $\varepsilon(0) := 0$

$$|f|(\varepsilon \circ f) + g'(0)(f - f'(0)) = g \circ f - g'(0)f'(0).$$

Da $\varepsilon \circ f$ sowie $|\cdot|^{-1}(f - f'(0))$ abklingend an 0 sind und $|\cdot|^{-1}f$ nach 1. lokal beschränkt an 0 ist, ist $|\cdot|^{-1}(g \circ f - g'(0)f'(0))$ abklingend an 0. Also ist $(g \circ f)'(0) = g'(0)f'(0)$. Der allgemeine Fall ergibt sich daraus mit $\tau_{f(a)} g \circ \tau_a f = \tau_a(g \circ f)$. □

Bemerkung und Definition 4.6 Es sei nun $\mathbb{K} = \mathbb{R}$. Ein Vektor $\mathbf{v} \in V^*$ nennen wir eine **Richtung** in V. Für $a \in V$ ist dann $a + \mathbb{R}\mathbf{v}$ 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 $t \in \mathbb{R}^*$ genügend klein

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

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

$$\partial_{\mathbf{v}} f(a) = f'(a)\mathbf{v} \tag{4.1}$$

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. Hier ist $\partial_k f$ ist die Ableitung der Funktion $x_k \mapsto f(x_1, \ldots, x_k, \ldots, x_d)$ bei festgehaltenen Variablen $x_1, \ldots, x_{k-1}$ und $x_{k+1}, \ldots, x_d$.[26] Die Definition zeigt, dass im Falle

[26] 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.

und $E = \mathbb{K}$ (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. Ist $E = \mathbb{K}^m$ und ist $f = (f_1, \ldots, f_m)^\top$ (also f_j die j-te Komponentenfunktion von f), so ist $\partial_{\mathbf{v}} f = (\partial_{\mathbf{v}} f_1, \ldots, \partial_{\mathbf{v}} f_m)^\top$.

Beispiel 4.7 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 $\mathbf{v} = (v_1, v_2)$ und $a = (0, 0)$

$$f\big(t(v_1, v_2)\big) = t^2 v_1^2 + t v_2 \qquad (t \in \mathbb{R}) \, .$$

und damit

$$\partial_{\mathbf{v}} f(0,0) = \left(\frac{\partial f}{\partial \mathbf{v}}(0,0) = D_{\mathbf{v}} f(0,0) \right) = \lim_{t \to 0} \frac{t^2 v_1^2 + t v_2}{t} = v_2 \, .$$

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 $(x, y, z) \in \mathbb{R}^3$

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

Bemerkung und Definition 4.8 Es seien $X \subset \mathbb{R}^d$, $a \in X$ und $f : X \to \mathbb{R}^m$. Ist f an a partiell differenzierbar nach allen Variablen, so heißt die Matrix $Jf(a) \in \mathbb{R}^{m \times d}$, deren k-te Spalte aus dem Vektor $\partial_k f(a)$ besteht, die **Jacobi-Matrix** von f an a. Ist speziell $m = 1$, so ist die Jacobi-Matrix der (Zeilen-)vektor $(\partial_1 f(a), \ldots, \partial_d f(a))$. Dann nennt man den Spaltenvektor

$$\nabla f(a) := \mathrm{grad} f(a) := (\partial_1 f(a), \ldots, \partial_d f(a))^\top = (Jf)^\top(a)$$

Gradient von f an a.[27] Es gilt also: Ist $f = (f_1, \ldots, f_d)$ mit den Komponentenfunktionen $f_j : X \to \mathbb{R}$, so ist die j-te Zeile von $Jf(a)$ der (transponierte) Gradient von f_j und

$$Jf(a) = (\partial_k f_j(a))_{j,k}.$$

[27] Wir werden im Weiteren Spaltenvektoren meistens wieder als Zeilenvektoren schreiben, was immer dann egal ist, wenn wir keine Matrizenarithmetik betreiben.

Ist f sogar differenzierbar an a, so gilt nach (4.1)

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

und damit ist $Jf(a)$ die Matrix, die die lineare Abbildung $f'(a)$ in der Standardbasis darstellt, also

$$f'(a) = (h \mapsto Jf(a)h).$$

Außerdem gilt dann, wieder nach (4.1),

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

für alle Richtungen $\mathbf{v}$.

Satz 4.9 *Es seien $X \subset \mathbb{R}^d$, $a \in X^\circ$ und $f : X \to \mathbb{R}^m$ so, dass Jf auf einer Umgebung U von a existiert. Sind alle $\partial_k f_j : U \to \mathbb{R}$ 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 wieder mit $\tau_a f$ statt f). Weiterhin setzen wir $m = 1$ voraus, also f reellwertig. Den Fall $\mathbb{R}^d$ kann man durch Betrachtung der Komponentenfunktionen $f_1, \ldots, f_d$ von f darauf zurückführen.
Ist $\varepsilon > 0$, so existiert ein $\delta > 0$ mit $\big|\partial_k f(x) - \partial_k f(0)\big| < \varepsilon/d$ für alle $x \in U_\delta(0)$ und $k = 1, \ldots, d$. Für $0 < |h| < \delta$ setzen wir $h^{(0)} := 0$ und

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

Dann ist(Teleskopsumme)

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

Wegen $|h^{(k)}| \leq |h| < \delta$ 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 $g'(t) = \partial_k f(h^{(k-1)} + th_k\mathbf{e}^{(k)})h_k$. Nach dem Mittelwertsatz existiert zu jedem k ein $\tau_k \in (0,1)$ mit

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

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

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

und damit ist f differenzierbar an a mit Ableitung $h \mapsto \nabla f(0)^\top h$. □

Bemerkung 4.10 Indem man die Matrix $A \in \mathbb{R}^{m \times d}$ mit der linearen Abbildung $h \mapsto Ah$ identifiziert, kann man den Raum $\mathbb{R}^{m \times d}$ mit der Operatornorm versehen. Ist $X \subset \mathbb{R}^d$ und ist $g = (g_{j,k}) : X \to \mathbb{R}^{m \times d}$ eine matrixwertige Funktion, so ist g genau dann stetig an a, wenn alle $g_{j,k} : X \to \mathbb{R}$ stetig sind ([Ü]). Mit Satz 4.9 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 4.11 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 4.10 (stetig) differenzierbar auf $\mathbb{R}^2$.
2. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

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

Dann gilt für $(x,y) \neq (0,0)$

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

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 4.12 *Es seien V ein Banachraum über $\mathbb{R}$, $X \subset V$ offen und $a, b \in V$ mit $[a,b] \subset X$.*

1. ***(Mittelwertsatz)** Ist $f : X \to \mathbb{R}$ differenzierbar, so existiert ein $\xi \in (a,b)$ mit*

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

2. ***(Schrankensatz)** Ist E ein Hilbertraum und ist $f : X \to E$ differenzierbar, so existiert ein $\xi \in (a,b)$ mit*

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

Beweis. Es sei $s := s_a^b$, also $s(t) := a + t(b-a)$ für $t \in [0,1]$. Nach Beispiel 4.4.1 gilt $s'(t)h = h(b-a)$ für $t \in [0,1]$, $h \in \mathbb{R}$ und damit nach der Kettenregel

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

1. Ist $f : X \to \mathbb{R}$, so existiert nach dem skalaren Mittelwertsatz (Satz 2.19) ein $\tau \in (0,1)$ mit

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

Also folgt 1. mit $\xi := s(\tau)$.
2. Ohne Einschränkung sei $c := f(b) - f(a) \neq 0$. Wir betrachten $\varphi : E \to \mathbb{R}$ mit $\varphi y := \langle y, c\rangle$. Dann ist φ linear und nach der Cauchy-Schwarzschen Ungleichung gilt $||\varphi|| \leq |c|$, also $\varphi \in L(E, \mathbb{R})$. Wegen $\varphi c = |c|^2$ ist genauer $||\varphi|| = |c|$. Wieder mit Beispiel 4.4.1. ist zudem $\varphi'(y) = \varphi$ für alle $y \in E$, also $\varphi \circ f : X \to \mathbb{R}$ nach der Kettenregel differenzierbar mit

$$(\varphi \circ f)'(x) = \varphi f'(x) \quad (x \in X).$$

Nach 1. existiert ein $\xi \in (a,b)$ mit

$$|c|^2 = \varphi c = (\varphi \circ f)(b) - (\varphi \circ f)(a) = \varphi f'(\xi)(b-a) \leq \|\varphi\| \cdot \|f'(\xi)\| \cdot |b-a|.$$

Wegen $||\varphi|| = |c|$ ergibt sich die Behauptung nach Division durch $|c|$. □

Bemerkung 4.13 Ist $X \subset V$ offen und ist $f : X \to \mathbb{R}$ stetig differenzierbar, so existiert nach dem Mittelwertsatz und (4.1) für $a \in X$, jede Richtung $\mathbf{v}$ mit $|\mathbf{v}| = 1$ und t genügend klein ein $\xi \in (x, x + t\mathbf{v})$ so, dass

$$f(a + t\mathbf{v}) - f(a) = tf'(\xi)\mathbf{v} \approx tf'(a)\mathbf{v} = t\partial_{\mathbf{v}} f(a) \tag{4.2}$$

Ist nun $V = \mathbb{R}^d$, so gilt $\partial_{\mathbf{v}} f(a) = Jf(a)\mathbf{v} = (\nabla f)^\top(a)\mathbf{v}$ (siehe Bemerkung 4.8) und nach der Cauchy-Schwarzschen Ungleichung $|(\nabla f)^\top(a)\mathbf{v}| \leq |\nabla f(a)|$ für $|\mathbf{v}| = 1$, also

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

Ist $\nabla f(a) \neq 0$, 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

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

Angesichts von (4.2) kann man daher 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. Außerdem gilt für $\mathbf{v} \perp \mathbf{v}^*$

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

d. h. die Richtungsableitungen der zur Gradientenrichtung senkrechten Richtungen verschwinden.

Beispiel 4.14 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.

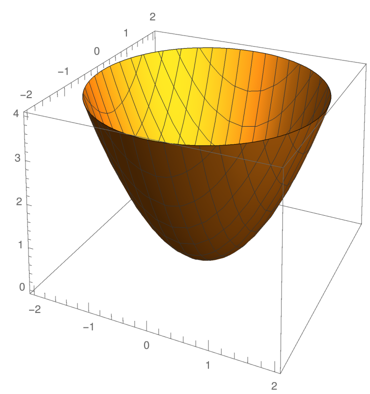

Abbildung 5: $f(x) = f(x_1, x_2) = x_1^2 + x_2^2$ für $|x| \leq 1$

Wir beschäftigen uns nun mit Ableitungen höherer Ordnung.

Bemerkung und Definition 4.15 Es seien V, E reelle Banachräume, $X \subset V$ offen und $f : X \to E$. Sind $\mathbf{v}$ und $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots$ Richtungen in V, so definiert man $\partial_{\mathbf{v}}^0 f := f$ 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 $\partial_{\mathbf{v}}^n f$ 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

$$\partial_{k_n} \ldots \partial_{k_1} f \quad \text{oder} \quad \frac{\partial^n f}{\partial x_{k_n} \ldots \partial x_{k_1}} \quad \text{oder} \quad f_{x_{k_1} \ldots x_{k_n}}$$

an Stelle von $\partial_{\mathbf{v}^{(n)}} \ldots \partial_{\mathbf{v}^{(1)}} f$ und spricht von den **partiellen Ableitungen der Ordnung** n. Für $k_1 = \ldots = k_n =: k$ schreibt man kurz $\partial_k^n f$ bzw. $\partial^n f / \partial x_k^n$ (und $\partial_k^0 f := f$).

Beispiel 4.16 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 4.16 ist es so, dass die Reihenfolge, in der die partiellen Ableitungen gebildet werden, vertauscht werden kann. Allgemein gilt

Satz 4.17 (Schwarz)
Es seien $X \subset \mathbb{R}^d$, $a \in X^\circ$ und $j,k \in \{1,\dots,d\}$. Weiter sei $f : X \to \mathbb{R}^m$ so, dass $\partial_k f$, $\partial_j f$ und $\partial_k\partial_j f$ auf einer Umgebung U von a 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.
Zunächst existiert ein $R > 0$ so, dass $(x,y) \in U$ für $x,y \in \mathbb{R}$ mit $0 < |x|,|y| < R$. Für solche (x,y) 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 $0 < \delta(\leq R)$ so, dass für alle $u,v \in \mathbb{R}$ mit $|u| < \delta$ und $|v| < \delta$

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

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 4.18 Die Aussage des Satzes 4.17 wird im Allgemeinen falsch, wenn man auf die Voraussetzung der Stetigkeit der partiellen Ableitung $\partial_k\partial_j f$ an a verzichtet. Ist etwa $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

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

(vgl. Beispiel 4.11.2), 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 $\partial_1\partial_2 f(0,0) = -1$.

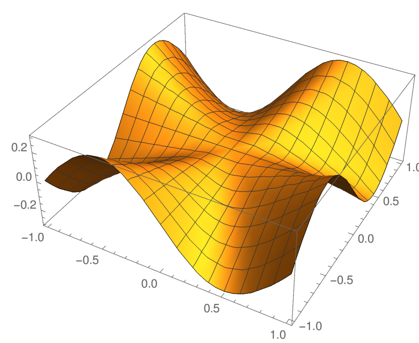

Abbildung 6: $f(x,y)$ für $|x|, |y| \le 1$

Bemerkung und Definition 4.19 1. Sind $X \subset \mathbb{R}^d$ offen und $n \in \mathbb{N}_0$, so bezeichnen wir mit $C^n(X, \mathbb{R}^m)$ die Menge aller Funktionen $f : X \to \mathbb{R}^m$ mit der Eigenschaft, dass alle partiellen Ableitungen der Ordnung $\le n$ auf X existieren und dort stetig sind.[28] Außerdem setzen wir $C^\infty(X, \mathbb{R}^m) := \bigcap_{n\in\mathbb{N}} C^n(X, \mathbb{R}^m)$ und $C^n(X) := C^n(X, \mathbb{C}) = C^1(X, \mathbb{R}^2)$. Durch mehrfache Anwendung des Satzes von Schwarz (Satz 4.17) 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.

[28] Nach Bemerkung 4.10 stimmt dies für $n = 1$ mit der alten Definition überein.

2. Nach (4.1) gilt im Falle $f \in C^1(X, \mathbb{R}^m)$

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

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

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

Satz 4.20 *Es seien $X \subset \mathbb{R}^d$ offen und $f \in C^n(X, \mathbb{R}^m)$. Dann ist $\partial^n_{\mathbf{v}} f$ stetig für alle Richtungen $\mathbf{v} = (v_1, \ldots, v_d)$ mit*

$$\partial^n_{\mathbf{v}} f = \sum_{(k_1,\ldots,k_n)\in\{1,\ldots,d\}^n} v_{k_1} \cdots v_{k_n} (\partial_{k_n} \ldots \partial_{k_1} f) = n! \sum_{|\alpha|_1 = n} \frac{1}{\alpha!} \mathbf{v}^\alpha \partial^\alpha f.$$

Beweis. 1. Wir beweisen den ersten Teil per Induktion nach n.
Für $n = 1$ entspricht die Behauptung (4.1) (siehe Bemerkung und Definition 4.19).
$n \to n+1$: Es sei $f \in C^{n+1}(X, \mathbb{R}^m)$ gegeben. Nach Induktionsvoraussetzung ist $\partial^n_{\mathbf{v}}$ 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 (4.1) ergibt sich

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

Damit ist $\partial^{n+1}_{\mathbf{v}} f$ stetig und die erste Gleichung gilt für $n+1$.
2. Nach Bemerkung und Definition 4.19 kann man die Reihenfolgen der partiellen Ableitungen in der ersten Summe 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 α_j-mal vorkommt,[29] ergibt sich auch die zweite Gleichung, also (ausgeschrieben)

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

□

Bemerkung und Definition 4.21 Für $f \in C^2(X, \mathbb{R})$ heißt die Matrix

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

[29] Multinomialkoeffizient; siehe etwa https://de.wikipedia.org/wiki/Multinomialkoeffizient.

die **Hesse-Matrix** von f an der Stelle x. Nach dem Satz von Schwarz (Satz 4.17) ist die Hesse-Matrix $Hf(x)$ für alle $x \in X$ symmetrisch.[30] und nach Satz 4.20 gilt

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

Bemerkung und Definition 4.22 Sind $X \subset \mathbb{R}^d$ offen, $a \in X$ und $f \in C^n(X)$, so nennt man

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

n-tes **Taylor-Polynom** von f bezüglich a.[31] Für $\mathbf{v} \neq 0$ ist nach Satz 4.20

$$T_n(f,a)(\mathbf{v}) = \sum_{\mu=0}^{n} \frac{1}{\mu!} \partial_{\mathbf{v}}^{\mu} f(a).$$

Sind $f \in C^{n+1}(X)$ und $I \supset [0,1]$ ein offenes Intervall mit $a + I\mathbf{v} \subset X$, so betrachten wir die Funktion $g : I \to \mathbb{C}$, definiert durch

$$g(t) := f(a + t\mathbf{v}) \qquad (t \in I).$$

Aus der Definition der Richtungsableitungen ergibt sich $g^{(k)}(t) = \partial_{\mathbf{v}}^k f(a+t\mathbf{v})$ für $t \in I$ und $k \le n+1$. Durch Anwendung des Satzes von Taylor (Satz 3.24) auf die Funktion g (mit $a = 0$ und $h = 1$) folgt

$$f(a+\mathbf{v}) = T_n(f,a)(\mathbf{v}) + \frac{1}{n!} \int_0^1 (1-t)^n \partial_{\mathbf{v}}^{n+1} f(a+t\mathbf{v})\, dt \tag{4.3}$$

(**Taylorformel für Richtungen**). Ist f reellwertig, so existiert zudem nach Bemerkung 3.26 ein $\xi \in [a, a+\mathbf{v}]$ so, dass

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

(**Lagrangeform des Restgliedes**). Im Fall $n = 0$ ergibt sich die Existenz eines $\xi \in [a, a+\mathbf{v}]$ so, dass

$$f(a+\mathbf{v}) = f(a) + \partial_{\mathbf{v}} f(\xi) = f(a) + (\nabla f)^\top(\xi)\mathbf{v},$$

also wieder die Aussage des Mittelwertsatzes, und für $n = 1$ die Existenz eines $\xi \in [a, a+\mathbf{v}]$ mit

$$f(a+\mathbf{v}) = f(a) + \partial_{\mathbf{v}} f(a) + \frac{1}{2}\partial_{\mathbf{v}}^2 f(\xi) = f(a) + (\nabla f)^\top(a)\mathbf{v} + \frac{1}{2}\mathbf{v}^\top Hf(\xi)\mathbf{v}\,. \tag{4.4}$$

[30] $A \in \mathbb{R}^{d \times d}$ heißt symmetrisch, falls $A = A^\top$ gilt.

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

Satz 4.23 (Taylor)
Es seien $X \subset \mathbb{R}^d$ sternförmig, $n \in \mathbb{N}_0$ und $f \in C^{n+1}(X, \mathbb{R})$. Dann existiert zu jedem $a \in X$ und jedem $h \in X - a$ ein $\xi \in [a, a+h]$ so, dass

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

Beweis. Ohne Einschränkung sei $h \neq 0$. Dann ergibt sich die Behauptung aus der Taylorformel für Richtungen mit Lagrange Restglied und Satz 4.20. □

Als wesentliche Anwendung der Taylorformel (4.4) werden wir ein hinreichendes Kriterium für lokale Extrema herleiten. Zunächst formulieren wir ein einfaches notwendiges Kriterium.

Definition 4.24 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 $V = \mathbb{R}^d$ und $E = \mathbb{R}$ ist a genau dann kritisch, wenn $\nabla f(a) = 0$ gilt.

Satz 4.25 *Es seien $X \subset \mathbb{R}^d$ 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 sei $k \in \{1, \ldots, d\}$. Ist I ein offenes Intervall mit $0 \in I$ und $a + I\mathbf{e}^{(k)} \subset X$, und ist $g_k : I \to \mathbb{R}$ definiert durch $g_k(t) := f(a + t\mathbf{e}^{(k)})$ für $t \in I$, so ist 0 Extremstelle von g_k. Nach Satz 2.16 gilt $0 = g_k'(0) = \partial_k f(a)$. □

Beispiel 4.26 1. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ mit

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

für $x, y \in \mathbb{R}$. Dann hat f an $(0,0)$ ein (offenbar sogar globales) Minimum. 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.

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.

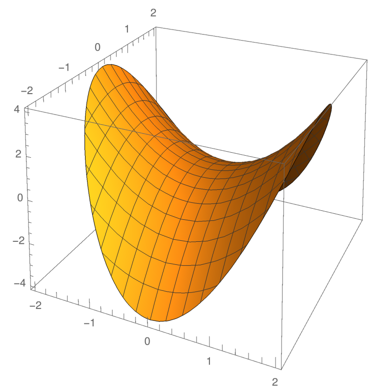

Abbildung 7: $f(x,y) = x^2 - y^2$ für $|x|, |y| \leq 1$

Bemerkung und Definition 4.27 Ist $A \in \mathbb{R}^{d\times d}$ symmetrisch, so heißt A

1. **positiv (semi-)definit**, falls $\mathbf{v}^\top A\mathbf{v} > 0$ (≥ 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. ≤ 0) gilt.

Bemerkung 4.28 Wir setzen

$$\mathbb{S}^{d-1} := \{\mathbf{v} \in \mathbb{R}^d : |\mathbf{v}| = 1\}$$

(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 4.4.2). Damit existieren $\min\limits_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v}$ und $\max\limits_{\mathbf{v}\in\mathbb{S}^{d-1}} \mathbf{v}^\top A\mathbf{v}$. Außerdem folgt aus der Cauchy-Schwarzschen Ungleichung

$$|\mathbf{v}^\top A\mathbf{v}| \leq ||A|| \qquad (\mathbf{v} \in \mathbb{S}^{d-1})\ .$$

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 $\lambda_{\min} > 0$ (≥ 0) gilt und genau dann negativ (semi-)definit wenn $\lambda_{\max} < 0$ (≤ 0) gilt.

Wir zeigen folgende, für die mehrdimensionale Optimierung zentrale Erweiterung des Satzes 2.30:

Satz 4.29 *Es seien $X \subset \mathbb{R}^d$ offen, $f \in C^2(X, \mathbb{R})$ und $a \in X$ ein kritischer Punkt von f. Dann gilt*

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

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

Nach Bemerkung 4.10 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 $h \in U_\delta(a) \setminus \{0\}$, so existiert nach (4.4) wegen $\nabla f(a) = 0$ ein $\xi \in [a, a+h] \subset U_\delta(a)$ mit

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

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 4.22 mit $g''(0) = \partial_{\mathbf{v}}^2 f(a) = \mathbf{v}^\top A \mathbf{v}$. Außerdem ist 0 Minimalstelle von g und damit keine strikte Maximalstelle, also $\mathbf{v}^\top A \mathbf{v} = g''(0) \geq 0$ nach Satz 2.30.
3. Die Aussagen für Maximalstellen ergeben sich durch Betrachtung von $-f$. □

Beispiel 4.30 1. Ist $f(x, y) = x^2 + y^2$ für $x, y \in \mathbb{R}$ (vgl. Beispiel 4.26.1), 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 4.26.2) so gilt

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

Hier ist Hf stets indefinit. Also hat f nach Satz 4.29 keine lokalen Extrema. Die kritische Stelle $(0,0)$ ist ein sogenannter Sattelpunkt (siehe Abbildung 7)
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

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

Damit sind $(0,-1)$ eine strikte Maximalstelle sowie $(1,0)$ eine strikte Minimalstelle und ansonsten gibt es keine Extremstellen. In Abbildung 8 kann man die vier kritischen Stellen, von denen zwei Sattelpunkte sind, erkennen.

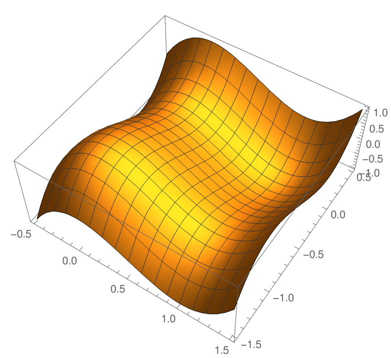

Abbildung 8: $f(x,y)$ für $|x - 1/2|, |y + 1/2| \leq 1$

5 Hauptsätze der mehrdimensionalen Analysis

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

Bemerkung und Definition 5.1 Es seien (X, d) ein metrischer Raum, $M \subset X$ und $\alpha \in [0, 1)$. Eine Abbildung $\varphi : M \to X$ heißt α-**Kontraktion**, falls

$$d(\varphi(x), \varphi(x')) \leq \alpha d(x, x') \qquad (x, x' \in M) .$$

Offensichtlich ist jede α-Kontraktion stetig. Außerdem existiert höchstens ein $x^* \in M$ mit $x^* = \varphi(x^*)$ d. h. φ hat höchstens einen Fixpunkt (ist x_* ein weiterer Fixpunkt von φ, so folgt $d(x_*, x^*) = d(\varphi(x_*), \varphi(x^*)) \leq \alpha d(x_*, x^*)$, also $d(x_*, x^*) = 0$.)

Satz 5.2 (Banachscher Fixpunktsatz)
Es seien (X, d) ein vollständiger metrischer Raum, $M \subset X$ abgeschlossen und $\varphi : M \to M$ eine α-Kontraktion. Dann gilt

1. *Es existiert genau ein Fixpunkt $x^* \in M$.*

2. *Für alle $x_0 \in M$ konvergiert die Folge (x_n), definiert durch $x_{n+1} := \varphi(x_n)$ für $n \in \mathbb{N}_0$, gegen x^* mit*

$$d(x_n, x^*) \leq \alpha^n d(x_0, x^*) \leq \frac{\alpha^n}{1-\alpha} d(x_1, x_0) \quad (n \in \mathbb{N}_0) .$$

Beweis. 1. Für alle $k \in \mathbb{N}_0$ gilt

$$d(x_{k+1}, x_k) = d(\varphi(x_k), \varphi(x_{k-1})) \leq \alpha \cdot d(x_k, x_{k-1}) \leq \ldots \leq \alpha^k \cdot d(x_1, x_0) .$$

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

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

Ist $\varepsilon > 0$ gegeben, so existiert ein $R > 0$ mit

$$\frac{\alpha^n}{1-\alpha} d(x_1, x_0) < \varepsilon \qquad (n > R) ,$$

also auch $d(x_{n'}, x_n) < \varepsilon$ für $n' > n > R$. Folglich ist (x_n) eine Cauchy-Folge in X. Da X vollständig ist, existiert ein $x^* \in X$ mit $x_n \to x^*$ $(n \to \infty)$ und aufgrund der Abgeschlossenheit von M ist $x^* \in M$. Da φ insbesondere stetig ist, gilt damit

$$x^* \leftarrow x_{n+1} = \varphi(x_n) \to \varphi(x^*) \qquad (n \to \infty)$$

also $x^* = \varphi(x^*)$.
2. Es gilt

$$d(x_n, x^*) = d(\varphi(x_{n-1}), \varphi(x^*)) \leq \alpha d(x_{n-1}, x^*) \leq \ldots \leq \alpha^n d(x_0, x^*)$$

und mit (5.1) für $n = 0$ durch Grenzübergang $n' \to \infty$ auch die Abschätzung

$$d(x^*, x_0) \leq d(x_1, x_0)/(1-\alpha) .$$

□

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 $A, B \in L(V)$ die Abbildungen

$$L(V) \ni T \mapsto \begin{cases} AT + B \\ TA + B \end{cases} \tag{5.2}$$

stetig. Durch

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

ist eine Gruppe definiert, die **Automorphismengruppe** von V.

Satz 5.3 *Es sei $V \neq \{0\}$. Ist $A \in \mathrm{Aut}(V)$ und ist $B \in L(V)$ mit $||B - A|| < 1/||A^{-1}||$, so ist auch $B \in \mathrm{Aut}(V)$. Außerdem ist die Abbildung $\mathrm{Aut}(V) \ni A \mapsto A^{-1} \in \mathrm{Aut}(V)$ stetig.*

Beweis. Es seien $I := \mathrm{id}_V$ und $0 < r < 1$. Ist $C \in L(V)$ mit $||C|| \leq r/||A^{-1}||$, so gilt $||CA^{-1}|| \leq ||C|| \cdot ||A^{-1}|| \leq r$. Hieraus ergibt sich $I - CA^{-1} \in \mathrm{Aut}(V)$ und[32]

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

mit gleichmäßiger Konvergenz auf $\{C : ||C|| \leq r/||A^{-1}||\}$[33] (Neumannsche Reihe; [Ü]). Nach Satz 1.5 ist φ stetig an 0 mit $\varphi(0) = I$. Also gilt $A^{-1}\varphi(C) \to A^{-1}I = A^{-1}$ für $||C|| \to 0$.
Ist nun $B \in \mathrm{Aut}(V)$ mit $||B - A|| \leq 1/||A^{-1}||$ und $C := A - B$, so folgt

$$B = A - C = (I - CA^{-1})A \in \mathrm{Aut}(V)$$

und $B^{-1} = A^{-1}\varphi(C) \to A^{-1}$ für $A \to B$. □

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

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

[33] Hier brauchen wir die Begriffe und Ergebnisse aus Abschnitt 1 für Funktionen mit Werten in einem Banachraum E, also im Raum $B(M, E) := \{f : M \to E : f \text{ beschränkt}\}$. Bei den Aussagen und Beweisen kann man $(\mathbb{C}, |\cdot|)$ problemlos durch $(E, |\cdot|_E)$ ersetzen.

Satz 5.4 (Hauptsatz über lokale Umkehrbarkeit)
Es seien V ein reeller Hilbertraum, $X \subset V$ offen, $f \in C^1(X, V)$ und $a \in X$ mit $f'(a) \in \mathrm{Aut}(V)$. Dann existiert eine offene Umgebung U von a mit folgenden Eigenschaften: 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. Ohne Einschränkung sei $V \neq \{0\}$. Wir setzen $A := f'(a)$. Da f' stetig ist, existiert ein $\delta > 0$ so, dass mit $U := U_\delta(a)$

$$||f'(x) - A|| < c := (2||A^{-1}||)^{-1} \qquad (x \in U)\ .$$

Insbesondere ist damit $f'(x) \in \mathrm{Aut}(V)$ für $x \in U$ nach Satz 5.3.
Für $z \in V$ sei $\varphi = \varphi_z : U \to V$ definiert durch

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

Dann ist $\varphi_z(x) = x$ genau dann, wenn $y = f(x)$ gilt. Weiter ergibt sich mit Beispiel 4.4 und der Kettenregel

$$\varphi'(x) = \mathrm{id}_V - A^{-1}f'(x) = A^{-1}(A - f'(x)) \qquad (x \in U)$$

und damit $||\varphi'(x)|| \leq ||A^{-1}|| \cdot ||A - f'(x)|| < 1/2$. Nach dem Schrankensatz (Satz 4.12) gilt

$$|\varphi(x) - \varphi(x')| \leq |x - x'|/2 \qquad (x, x' \in U).$$

Also ist φ eine 1/2-Kontraktion und damit hat φ höchstens einen Fixpunkt nach Bemerkung 5.1. Folglich ist $f|_U$ injektiv und damit existiert die Umkehrfunktion $g : f(U) \to U$ von f_U.
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 $M := x + \rho B_V \subset U$. Ist $z \in U_{c\rho}(y)$ und $\varphi = \varphi_z$, so gilt für $u \in M$

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

also $\varphi(u) \in M$ und damit $\varphi(M) \subset M$. Da $M \subset V$ abgeschlossen ist, folgt aus dem Banachschen Fixpunktsatz, dass ein $u \in M$ existiert 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 $\varphi(u) = u - A^{-1}f(u)$, so ist $\varphi(0) = 0$ und φ eine 1/2-Kontraktion. Nach der umgekehrten Dreiecksungleichung ist

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

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 $B := f'(0)$

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

für $h \to 0$. Also ist auch g differenzierbar an 0 mit $g'(0) = B^{-1} = f'(g(0))^{-1}$. Da g stetig an 0 ist und f' nach Voraussetzung sowie $T \mapsto T^{-1}$ nach Satz 5.3 stetig sind, ist g' stetig an 0. □

Bemerkung 5.5 Ist die Aussage von Satz 5.4 erfüllt, so sagt man, f sei **lokal C^1-umkehrbar** an der Stelle a. Ist V *endlich-dimensional*, so ist jede lineare Abbildung beschränkt und zudem sind Surjektivität, Injektivität und Bijektivität dann äquivalent (Dimensionsformel; siehe lineare Algebra). Damit ist $f'(a) \in \mathrm{Aut}(V)$ schon dann, wenn $f'(a)$ surjektiv (oder injektiv) ist, also a eine reguläre Stelle von f ist. Ist $V = \mathbb{R}^d$, so ist dies genau dann der Fall wenn $\det Jf(a) \neq 0$ ist.

Beispiel 5.6 (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 5.4 lokal C^1-umkehrbar an allen Stellen (r, θ). Da $f(1, 2k\pi) = (1, 0)$ 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?
Gegeben ist eine Funktion $F : \Omega \to \mathbb{R}^m$ mit $\Omega \subset \mathbb{R}^d \times \mathbb{R}^m$ und damit Gleichungssystem

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

also mit $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ und $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$

$$\begin{aligned} F_1(x_1, \ldots, x_d,\, y_1, \ldots, y_m) &= 0 \\ &\vdots \\ F_m(x_1, \ldots, x_d,\, y_1, \ldots, y_m) &= 0 \end{aligned}$$

($d+m$ Unbekannte und m Gleichungen, also unterbestimmt). Weiter sei $(a,b) \in M$ eine Lösung, also $F(a,b)=0$. Unsere Frage ist, ob das Gleichungssystem lokal, also in einer Umgebung W von (a,b), nach y aufzulösbar ist, also die Lösungsmenge lokal der Graph einer Funktion f der Variablen x ist. Genauer gesagt lautet die Frage: Existieren Umgebungen U von a und W von (a,b) sowie eine Funktion $f : U \to \mathbb{R}^m$ so, dass

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

gilt. Für $F \in C^1(\Omega, \mathbb{R}^m)$ schreiben wir kurz

$$J_1 F := (\partial_k F_j)_{j=1,\ldots,m,\, k=1,\ldots,d} \quad \text{und} \quad J_2 F := (\partial_{d+k} F_j)_{j,k=1,\ldots,m}.$$

Beispiel 5.7 1. (Lineare Gleichungen) Es seien $A \in \mathbb{R}^{m\times d}$, $B \in \mathbb{R}^{m\times m}$ und

$$F(x,y) := Ax + By,$$

also $F(x,y) = 0$ das unterbestimmte lineare Gleichungssystem $Ax + By = 0$. Im Falle $\det J_2 F(x,y) = \det B \neq 0$ gilt $F(x,y) = 0$ genau dann, wenn $y = -B^{-1}Ax$, also

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

2. (Einheitskreis) Wir betrachten die Kreisgleichung

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

Dann ist offenbar $(a,b) = (1/\sqrt{2}, 1/\sqrt{2})$ eine Lösung der Gleichung. Für $U := (-1,1)$ und $W := U \times (0,\infty)$ gilt $F(x,y) = 0$ genau dann, wenn $y = \sqrt{1-x^2}$, also

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

Dies wird falsch etwa für die Lösung $(a,b) = (1,0)$. Hier ist die Lösungsmenge nicht mehr lokal der Graph einer Funktion von x. Man beachte dabei, dass $\det \partial_2 F(1,0) = 0$ gilt.

Satz 5.8 (Hauptsatz über implizite Funktionen)
Es seien $\Omega \subset \mathbb{R}^d \times \mathbb{R}^m$ offen und $F \in C^1(\Omega, \mathbb{R}^m)$. Weiter sei $(a,b) \in \Omega$ mit $F(a,b) = 0$ und $\det J_2 F(a,b) \neq 0$. Dann existieren offene Umgebungen U von a und W von (a,b) sowie eine Funktion $f \in C^1(U, \mathbb{R}^m)$ so, dass

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

Dabei gilt

$$Jf(x) = -(J_2 F(x, f(x))^{-1} J_1 F(x, f(x)) \qquad (x \in U) .$$

Beweis. Wir betrachten $G : \Omega \to \mathbb{R}^d \times \mathbb{R}^m$ mit

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

Dann gilt $G(a,b) = (a,0)$ und $G \in C^1(\Omega, \mathbb{R}^d \times \mathbb{R}^m)$. Bezeichnet E_d die d-dimensionale Einheitsmatrix, so ist

$$JG = \begin{pmatrix} E_d & 0 \\ J_1F & J_2F \end{pmatrix}$$

und damit $\det JG(a,b) = \det J_2F(a,b) \neq 0$. Nach Satz 5.4 ist G lokal C^1-umkehrbar an der Stelle (a,b). Damit existieren ein $\delta > 0$ und eine offene Umgebung W von (a,b) so, dass $G_W : W \to U_\delta(a) \times U_\delta(0)$ bijektiv ist mit $\det JG(x,y) \neq 0$ für $(x,y) \in W$ und stetig differenzierbarer Umkehrfunktion.
Wir setzen $U := U_\delta(a)$ und schreiben im Weiteren kurz G statt G_W. Aus der Definition von G ergibt sich, dass G^{-1} mit einer geeigneten Funktion $H \in C^1(U \times U_\delta(0), \mathbb{R}^m)$ von der Form

$$G^{-1}(x,z) = (x, H(x,z)) \qquad ((x,z) \in U \times U_\delta(0))$$

ist. Definiert man $\pi : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^m$ durch $\pi(x,y) := y$, so gilt

$$\pi(G(x,y)) = \pi(x, F(x,y)) = F(x,y) \qquad ((x,y) \in W)$$

und folglich

$$F(x, H(x,z)) = \pi(G(x, H(x,z))) = \pi(x,z) = z \quad ((x,z) \in U \times U_\delta(0))\ .$$

Nun definieren wir $f : U \to \mathbb{R}^m$ durch

$$f(x) := H(x,0) \qquad (x \in U).$$

Aus $H \in C^1(U \times U_\delta(0), \mathbb{R}^m)$ ergibt sich $f \in C^1(U, \mathbb{R}^m)$. Außerdem gilt

$$F(x, f(x)) = 0 \qquad (x \in U)$$

und für $(x,y) \in W$ mit $F(x,y) = 0$ folgt

$$G(x,y) = (x,0) = G(G^{-1}(x,0)) = G(x, f(x))$$

und damit auch $y = f(x)$. Schließlich ergibt sich mit $h(x) := F(x, f(x)) = 0$ für $x \in U$ aus der Kettenregel

$$0 = Jh(x) = JF(x, f(x)) \cdot \begin{pmatrix} E_d \\ Jf(x) \end{pmatrix} = J_1F(x, f(x)) + J_2F(x, f(x))Jf(x).$$

Wegen $\det J_2F(x,y) = \det JG(x,y) \neq 0$ für alle $(x,y) \in W$ ergibt sich die Zusatzbehauptung durch Auflösen der letzten Gleichung nach $Jf(x)$. $\square$

Beispiel 5.9 (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 5.8 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 θ so, dass $\cos(2\theta) > 0$.

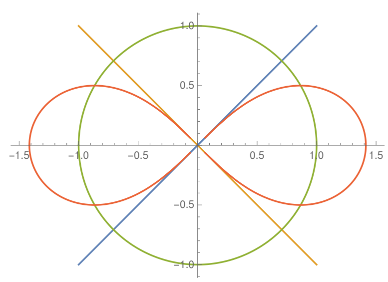

Abbildung 9: Lemniskate mit Einheitskreis und Winkelhalbierenden

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

Definition 5.10 Es seien (X,d) ein metrischer Raum, E Vektorraum und $f : X \to \mathbb{R}$ sowie $g : X \to E$. Ist

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

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

Satz 5.11 (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.[34] *Ist a eine Extremstelle von f unter der Nebenbedingung $g(x) = 0$, so existiert ein $\lambda \in \mathbb{R}^m$ mit*

$$\nabla f(a) = (Jg)^\top(a)\lambda = \sum_{j=1}^{m} \lambda_j \, \nabla g_j(a) \, .$$

Beweis. Wir fassen $\mathbb{R}^d$ als $\mathbb{R}^{d-m} \times \mathbb{R}^m$ auf und schreiben $x = (u, y)$ mit $u \in \mathbb{R}^{d-m}$, $y \in \mathbb{R}^m$ sowie $a = (c, b)$ mit $a \in \mathbb{R}^{d-m}$, $c \in \mathbb{R}^m$. Da $Jg(a) = (J_1 g(c, b), J_2 g(c, b))$ vollen Rang m hat, können wir nach geeigneter Permutation der Variablen annehmen, dass

$$\det J_2 g(a) \neq 0$$

gilt (Lineare Algebra). Nach dem Hauptsatz über implizite Funktionen ist dann die Gleichung

$$g(x) = g(u, y) = 0$$

an (c, b) lokal nach y auflösbar. Insbesondere existieren eine offene Umgebung U von c sowie eine Funktion $\varphi \in C^1(U, \mathbb{R}^m)$ so, dass $\varphi(c) = b$ und

$$g(u, \varphi(u)) = 0 \qquad (u \in U)$$

gilt. Da die Ableitung von $u \mapsto g(u, \varphi(u))$ verschwindet, gilt nach der Kettenregel damit insbesondere

$$0 = Jg(c, b) \begin{pmatrix} E_{d-m} \\ J\varphi(c) \end{pmatrix} = J_1 g(a) + J_2 g(a) J\varphi(c).$$

Ist $h \in C^1(U, \mathbb{R})$ definiert durch

$$h(u) := f(u, \varphi(u)) \qquad (u \in U) \, ,$$

so ist c eine Extremstelle von h (ohne Nebenbedingung). Also gilt nach Satz 4.25

$$Jh(c) = (\nabla h)^\top(c) = 0$$

und mit der Kettenregel wie oben

$$0 = Jh(c) = J_1 f(a) + J_2 f(a) J\varphi(c) \, .$$

Wegen $\det J_2 g(a) \neq 0$ hat das lineare Gleichungssystem

$$\lambda^\top \cdot J_2 g(a) = J_2 f(a)$$

genau eine Lösung $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$. Hieraus ergibt sich wiederum

$$J_1 f(a) = -J_2 f(a) J\varphi(c) = -\lambda^\top J_2 g(a) J\varphi(c) = \lambda^\top J_1 g(a)$$

und damit insgesamt $(\nabla f)^\top(a) = (J_1 f(a), J_2 f(a)) = \lambda^\top (J_1 g(a), J_2 g(a)) = \lambda^\top \cdot Jg(a)$. $\square$

[34] Regulär bedeutet hier, dass $Jg(a) \in \mathbb{R}^{m \times d}$ vollen Rang ($= m$) hat.

Bemerkung 5.12 Der Satz von Lagrange besagt, dass (unter der Regularitätsbedingung) an Extremstellen von f unter der Nebenbedingung $g(x) = 0$ der Gradient von f eine Linearkombination der Gradienten von $g_1, \dots, g_m$ ist. Ähnlich wie im Falle von Satz 4.25 liefert der Satz lediglich eine notwendige Bedingung. Um die entsprechenden Punkte a zu bestimmen, hat man die Gleichungen

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

und

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

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

Beispiel 5.13 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

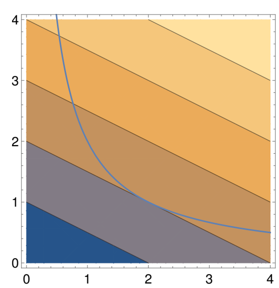

Abbildung 10: Niveaulinie $P(x, y) = \sqrt{2}$ und Niveaulinien zu $K(x, y) = x + 2y$

einem vorgegebenen Produktionsniveau $c > 0$, d. h. wir wollen $K(x, y)$ unter der Nebenbedingung $P(x, y) - c = 0$ minimieren. Nach Satz 5.11 ist wegen $\nabla P \neq 0$ 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, λ. 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 10). Man sieht, dass an der Minimalstelle die Höhenlinien von P und K tangential liegen. Dies liegt daran, dass die Gradienten linear abhängig sind und beide jeweils senkrecht auf den entsprechenden Höhenlinien stehen ([Ü]).

Zum Abschluss bewiesen wir ein Ergebnis über die Glattheit von Parameterintegralen.

Satz 5.14 (**Parameterintegrale**)
Es seien $X \subset \mathbb{R}^d$, $[a, b] \subset \mathbb{R}$ 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 gilt:

1. *Ist X offen oder abgeschlossen, so ist Φ stetig.*
2. *Ist X offen und ist $\nabla_1 \varphi := (\partial_1 \varphi, \ldots, \partial_d \varphi)^\top : X \times [a, b] \to \mathbb{C}^d$ stetig, so ist Φ differenzierbar mit*

$$\Phi'(x)h = \int_a^b (\nabla_1 \varphi)^\top(x, t) \cdot h\, dt \qquad (x \in X,\ h \in \mathbb{R}^d).$$

Beweis. Es sei $x \in X$ fest. Ist X offen, so wählen wir $r > 0$ so, dass $x + rB \subset X$, und ist X abgeschlossen, so wählen wir $r > 0$ beliebig. Dann ist $K := (x + rB) \cap X$ kompakt.
1. Da $K \times [a, b]$ kompakt ist, ist φ 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 Φ stetig an der Stelle x.
2. Wir setzen für $t \in [a, b]$

$$\psi_t(h) := \varphi(x + h, t) - (\nabla_1 \varphi)^\top(x, t) \cdot h \qquad (|h| < r).$$

Dann ist ψ_t stetig differenzierbar auf $U_r(0)$ mit

$$\nabla \psi_t(h) = \nabla_1 \varphi(x + h, t) - \nabla_1 \varphi(x, t).$$

Nun sei wieder $\varepsilon > 0$ gegeben. Da $(u,t) \mapsto \nabla\psi_t(u)$ stetig auf $K \times [a,b]$ ist und $\nabla\psi_t(0) = 0$ gilt, existiert wie in 1. ein $\delta > 0$ so, dass $||\psi_t'(u)|| \le |\nabla\psi_t(u)| < \varepsilon$ 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

$$|\varphi(x+h,t) - \varphi(x,t) - (\nabla_1\varphi)^\top(x,t)\cdot h| = |\psi_t(h) - \psi_t(0)| \le |\nabla\psi_t(\xi)|\cdot|h| < \varepsilon|h|\,.$$

Setzt man $Ah := \int_a^b (\nabla_1\varphi)^\top(x,t)\cdot h\,dt$, so ist $A : \mathbb{R}^d \to \mathbb{C}$ linear (und beschränkt) mit

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

für $|h| < \delta$. Da $\varepsilon > 0$ beliebig war, ist $|\cdot|^{-1}(\tau_x\Phi - A)$ abklingend an 0 und damit Φ differenzierbar an x mit $\Phi'(x) = A$. □

Als Anwendung des Satzes 5.14 beweisen wir

Satz 5.15 *Es gilt*

$$\int_0^\infty e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}.$$

Beweis. Wir betrachten die Funktion $\Phi : [0,\infty) \to [0,\infty)$, definiert durch

$$\Phi(x) := \int_0^1 e^{-x^2(1+t^2)}\frac{dt}{1+t^2} \qquad (x \ge 0).$$

Mit $f(t) := e^{-t^2}$ für $t \in \mathbb{R}$ und $Vf = V_0 f =$ gilt nach Satz 5.14, der Substitutionsregel, der Kettenregel und dem Hauptsatz über Integralfunktionen

$$\Phi'(x)1 = -2e^{-x^2}\int_0^1 xe^{-(xt)^2}\,dt = -2e^{-x^2}\int_0^x e^{-s^2}\,ds = -((Vf)^2)'(x) \quad (x \ge 0)\,.$$

Also ist $(Vf)^2 + \Phi$ konstant auf $[0,\infty)$. 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 $e^{-x^2t^2} \le 1$ für $t \in [0,1]$ ergibt sich $\Phi(x) \le e^{-x^2} \to 0$ für $x \to \infty$, also $(Vf)^2(x) \to \pi/4$ für $x \to \infty$ und damit

$$\int_0^\infty e^{-t^2}\,dt = Vf\big|_0^\infty = \sqrt{\pi}/2.$$

□

Ende

Index