

Analysis II

Georg Tamme

18. Juli 2017

Inhaltsverzeichnis

1	Topologie in metrischen Räumen	1
	Normierte und metrische Räume	1
	Folgen und Konvergenz	5
	Stetigkeit	6
	Offene und abgeschlossene Teilmengen	9
	Kompakte Teilmengen	12
2	Differenzialrechnung im $\mathbb{R}^n$	15
	Kurven im $\mathbb{R}^n$	15
	Funktionen mehrerer Variablen	16
	Totale Differenzierbarkeit und Jacobimatrix	22
	Die Kettenregel	28
	Maxima und Minima	32
	Der Satz über die Umkehrfunktion	39
	Implizite Funktionen und Extrema unter Nebenbedingungen	42
3	Ergänzungen zur Integralrechnung	48
	Anwendungen des Hauptsatzes	48
	Uneigentliche Integrale	53
	Integral und Grenzwert	57
4	Gewöhnliche Differenzialgleichungen	58
	Definition und elementare Lösungsmethoden	58
	Der Existenz- und Eindeutigkeitssatz	68
	Systeme von Differenzialgleichungen und Differenzialgleichungen höherer Ordnung	70
	Lineare Differenzialgleichungen mit konstanten Koeffizienten	74

Übersicht

Analysis I: Studium von Funktionen $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}$

- Stetigkeit
- Differenzierbarkeit, Ableitungen, Maxima, Minima
- Integrierbarkeit

Analysis II: Studium von Funktionen $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$

- Stetigkeit
- Differenzierbarkeit, (Richtungs)ableitungen, Maxima, Minima
- Einführung in Differenzialgleichungen

Literatur

- Forster, *Analysis 2*
- Fritzsche, *Grundkurs Analysis 2*
- Lang, *Calculus of Several Variables*
- Philip, *Calculus II for Statistics Students* Skript LMU

1 Topologie in metrischen Räumen

Normierte und metrische Räume

Erinnerung 1.1. Die Elemente des $\mathbb{R}^n$ sind Tupel $x = (x_1, \dots, x_n)$ mit $x_i \in \mathbb{R}$, $i = 1, \dots, n$. Der $\mathbb{R}^n$ ist ein $\mathbb{R}$ -Vektorraum bzgl. der komponentenweisen Addition und Skalarmultiplikation

$$\begin{aligned} x + y &= (x_1 + y_1, \dots, x_n + y_n) && \text{für } x, y \in \mathbb{R}^n \\ \lambda x &= (\lambda x_1, \dots, \lambda x_n) && \text{für } \lambda \in \mathbb{R}, x \in \mathbb{R}^n. \end{aligned}$$

Auf $V = \mathbb{R}^n$ haben wir das *(Standard)skalarprodukt*

$$\langle x, y \rangle = \langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = x_1 y_1 + \dots + x_n y_n.$$

Mittels des Skalarprodukts können wir die *Standard-* oder *euklidische Norm* (oder *Länge*) definieren:

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \dots + x_n^2}.$$

Diese hat folgende Eigenschaften:

- (N1) $\|x\| \geq 0$ für alle $x \in V$
- (N2) $\|x\| = 0 \Leftrightarrow x = 0$ (positiv definit)
- (N3) $\|\lambda x\| = |\lambda| \cdot \|x\|$ für alle $\lambda \in \mathbb{R}, x \in V$ (Homogenität)
- (N4) $\|x + y\| \leq \|x\| + \|y\|$ für alle $x, y \in V$ (Dreiecksungleichung)

Der Abstand zwischen zwei Punkten x und y im $\mathbb{R}^n$ ist $\|x - y\|$.

Beweis der Eigenschaften. (vgl. LA Lemma 4.5) N1 – N3 sind klar. Für N4 benutze Cauchy-Schwarzsche Ungleichung $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$:

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

□

Allgemeiner definiert man:

Definition 1.2. Ein normierter $\mathbb{R}$ -Vektorraum ist ein Paar $(V, \|\cdot\|)$ bestehend aus einem $\mathbb{R}$ -Vektorraum V und einer Abbildung (der *Norm*)

$$\|\cdot\|: V \rightarrow \mathbb{R}_{\geq 0},$$

welche die Eigenschaften (N1) bis (N4) erfüllt.

Beispiel 1.3. (a) Ist V ein $\mathbb{R}$ -Vektorraum und $\langle \cdot, \cdot \rangle$ ein Skalarprodukt auf V , so definiert $\|x\| := \sqrt{\langle x, x \rangle}$ eine Norm auf V .

(b) Betrachte die beiden Abbildungen $\|\cdot\|_1, \|\cdot\|_\infty: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\begin{aligned}\|x\|_1 &:= |x_1| + \dots + |x_n| && (1\text{-Norm}) \text{ und} \\ \|x\|_\infty &:= \max\{|x_1|, \dots, |x_n|\} && (Max\text{-Norm}).\end{aligned}$$

Auch dies sind Normen auf $\mathbb{R}^n$ (Übung). Sie kommen nicht von einem Skalarprodukt her. Sie unterscheiden sich nicht wesentlich von der euklidischen Norm, denn es gelten:

$$\begin{aligned}\|x\|_\infty &\leq \|x\| \leq \sqrt{n}\|x\|_\infty, \\ \|x\|_\infty &\leq \|x\|_1 \leq n\|x\|_\infty.\end{aligned}$$

(c) Sei X eine Menge. Dann ist

$$B(X, \mathbb{R}) = \{f: X \rightarrow \mathbb{R} : \sup_{x \in X} |f(x)| < \infty\},$$

bzgl. punktwaiser Addition und Skalarmultiplikation

$$\begin{aligned}(f+g)(x) &:= f(x) + g(x) & (f, g \in B(X, \mathbb{R}), x \in X) \\ (\lambda f)(x) &:= \lambda f(x) & (\lambda \in \mathbb{R}, f \in B(X, \mathbb{R}), x \in X)\end{aligned}$$

ein $\mathbb{R}$ -Vektorraum, der *Vektorraum der beschränkten $\mathbb{R}$ -wertigen Funktionen auf X* . Definiere

$$\|\cdot\|_\infty: B(X, \mathbb{R}) \rightarrow \mathbb{R}, \quad \|f\|_\infty := \sup\{|f(x)| : x \in X\}.$$

Dies ist eine Norm auf $B(X, \mathbb{R})$. [Für $X = \{1, \dots, n\}$ ist dies einfach $\mathbb{R}^n$ mit der Max-Norm.]

Beweis. (N1) – (N3) sind einfach. Dreiecksungleichung:

$$\begin{aligned}\|f+g\|_\infty &= \sup\{|f(x)+g(x)| : x \in X\} \\ &\leq \sup\{|f(x)| + |g(x)| : x \in X\} \\ &\leq \sup\{|f(x)| : x \in X\} + \sup\{|g(x)| : x \in X\} \\ &= \|f\|_\infty + \|g\|_\infty.\end{aligned}$$

□

(d) Seien $I = [a, b] \subseteq \mathbb{R}$ ein Intervall und

$$C(I, \mathbb{R}) = \{f: I \rightarrow \mathbb{R} : f \text{ stetig} \}$$

der Vektorraum der stetigen Funktionen auf I . Sei $p \in [1, \infty)$. Mit f ist auch die Funktion $|f|^p$ stetig und damit integrierbar. Man kann zeigen, dass

$$\|\cdot\|_p: C(I, \mathbb{R}) \rightarrow \mathbb{R}, \quad \|f\|_p := \left(\int_a^b |f(x)|^p dx \right)^{1/p}$$

eine Norm (“*p-Norm*”) auf $C(I, \mathbb{R})$ definiert.

(e) Es kann sehr unterschiedliche Normen auf demselben Vektorraum geben: Z.B. ist auf dem Vektorraum $C(I, \mathbb{R})$ aus (d) auch die sup-Norm

$$\|f\|_\infty := \sup\{|f(x)| : x \in I\}$$

eine Norm.

Erinnerung 1.4. Sei $(x_k)_{k \in \mathbb{N}}$ eine Folge in $\mathbb{R}$. Man sagt, die Folge $(x_k)_{k \in \mathbb{N}}$ *konvergiere* gegen ein $a \in \mathbb{R}$, falls für alle $\epsilon > 0$ ein $N \in \mathbb{N}$ existiert so, dass $|x_k - a| < \epsilon$ für alle $k \geq N$.

Dieser Konvergenzbegriff lässt sich sofort auf normierte Räume übertragen. Mehr noch, es wird nur der *Abstand* der Punkte x_k und a benötigt, nicht die Norm als solche. Dies führt auf den Begriff des metrischen Raumes:

Definition 1.5. Ein *metrischer Raum* ist ein Paar (X, d) bestehend aus einer Menge X und einer Abbildung $d: X \times X \rightarrow \mathbb{R}$ (der *Metrik*), welche folgende Eigenschaften erfüllt:

- (M1) $d(x, y) \geq 0$ für alle $x, y \in X$,
- (M2) $d(x, y) = 0 \Leftrightarrow x = y$,
- (M3) $d(x, y) = d(y, x)$ für alle $x, y \in X$, (Symmetrie)
- (M4) $d(x, z) \leq d(x, y) + d(y, z)$ für alle $x, y, z \in X$. (Dreiecksungleichung)

Beispiel 1.6. (a) Ist $(V, \|\cdot\|)$ ein normierter Vektorraum, so definiert $d(x, y) := \|x - y\|$ eine Metrik auf V .

(b) Auf jeder Menge X kann man die *triviale* oder *diskrete Metrik* definieren durch

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

- (c) Ist (X, d) ein metrischer Raum, und ist $Y \subseteq X$ eine Teilmenge, so erhält man durch Einschränken von d auf $Y \times Y$ eine Metrik Y .
- (d) Betrachte die Menge aller 0-1-Folgen $\{0, 1\}^{\mathbb{N}} = \text{Abb}(\mathbb{N}, \{0, 1\})$. Für zwei Folgen $(x_k)_{k \in \mathbb{N}}$ und $(y_k)_{k \in \mathbb{N}}$ sei $k_0 := \inf\{k \in \mathbb{N} : x_k \neq y_k\} \in \mathbb{N} \cup \{\infty\}$. Setze

$$d((x_k)_{k \in \mathbb{N}}, (y_k)_{k \in \mathbb{N}}) := 2^{-k_0}$$

wobei $2^{-\infty} := 0$. Dies definiert eine Metrik auf $\{0, 1\}^{\mathbb{N}}$.

- (e) Fixiere eine Primzahl p . Jede rationale Zahl $x \in \mathbb{Q} \setminus \{0\}$ lässt sich in der Form $x = p^k \cdot \frac{m}{n}$ mit $k \in \mathbb{Z}$ und $m, n \in \mathbb{Z} \setminus \{0\}$ nicht durch p teilbar schreiben. Dabei ist k eindeutig bestimmt. Setze $|x|_p := p^{-k}$ und $|0|_p := 0$. Dann definiert $d_p(x, y) := |x - y|_p$ eine Metrik auf $\mathbb{Q}$ (die *p-adische Metrik*).

Folgende Variante der Dreiecksungleichung ist oft nützlich:

Lemma 1.7 (“Inverse Dreiecksungleichung”). *Sind x, y, z Punkte im metrischen Raum (X, d) , so gilt*

$$|d(x, y) - d(x, z)| \leq d(y, z).$$

Beweis. Nach der Dreiecksungleichung gelten

$$\begin{aligned} d(x, y) &\leq d(x, z) + d(z, y) = d(x, z) + d(y, z) && \Leftrightarrow d(x, y) - d(x, z) \leq d(y, z), \\ d(x, z) &\leq d(x, y) + d(y, z) && \Leftrightarrow d(x, z) - d(x, y) \leq d(y, z). \end{aligned}$$

Zusammen ergibt das $|d(x, y) - d(x, z)| \leq d(y, z)$, wie gewünscht. $\square$

Folgen und Konvergenz

Sei (X, d) ein metrischer Raum. Wir schreiben oft einfach nur X .

Definition 1.8. Eine Folge $(x_k)_{k \in \mathbb{N}} = (x_1, x_2, \dots)$ in X *konvergiert gegen* $a \in X$: $\Leftrightarrow$ für alle $\epsilon > 0$ in $\mathbb{R}$ gibt es ein $N \in \mathbb{N}$ so, dass $d(x_k, a) < \epsilon$ für alle $k \geq N$. Schreibweise: $\lim_{k \rightarrow \infty} x_k = a$ oder $x_k \xrightarrow{k \rightarrow \infty} a$.

Eine Folge $(x_k)_{k \in \mathbb{N}}$ in X heißt *Cauchyfolge* : $\Leftrightarrow$ für alle $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$ so, dass $d(x_k, x_l) < \epsilon$ für alle $k, l \geq N$.

Bemerkung 1.9. (a) Jede konvergente Folge ist eine Cauchyfolge. Denn konvergiert $(x_k)_{k \in \mathbb{N}}$ gegen $a \in X$, so folgt aus der Dreiecksungleichung $d(x_k, x_l) \leq d(x_k, a) + d(a, x_l) < \epsilon$ für $k, l \geq N$, falls N so gewählt ist, dass $d(x_k, a) < \frac{\epsilon}{2}$ für alle $k \geq N$.

(b) Grenzwerte sind eindeutig: Konvergiert die Folge $(x_k)_{k \in \mathbb{N}}$ sowohl gegen a als auch gegen b , so gilt $a = b$. Falls nicht, wäre $\epsilon := d(a, b)/2 > 0$ und es gäbe $N_a, N_b \in \mathbb{N}$ so, dass $d(x_k, a) < \epsilon$ für alle $k \geq N_a$ und $d(x_k, b) < \epsilon$ für alle $k \geq N_b$. Für $k \geq \max\{N_a, N_b\}$ würde dann gelten:

$$d(a, b) \leq d(a, x_k) + d(x_k, b) < \epsilon + \epsilon = d(a, b),$$

Widerspruch.

Im Folgenden sei $\mathbb{R}^n$ mit der durch die euklidische Norm gegebenen Metrik (der *euklidischen Metrik*) versehen. Wie sehen dann konvergente Folgen im $\mathbb{R}^n$ aus?

Satz 1.10. Sei $(x_k)_{k \in \mathbb{N}}$ eine Folge im $\mathbb{R}^n$, wobei $x_k = (x_{k1}, \dots, x_{kn})$. Genau dann konvergiert die Folge $(x_k)_{k \in \mathbb{N}}$ gegen $a = (a_1, \dots, a_n)$, wenn für jedes $i = 1, \dots, n$ die Komponentenfolge $(x_{ki})_{k \in \mathbb{N}}$ in $\mathbb{R}$ gegen a_i konvergiert.

Beweis. “ $\Rightarrow$ ” Angenommen, die Folge $(x_k)_{k \in \mathbb{N}}$ konvergiert gegen a . Zu gegebenem $\epsilon > 0$ gibt es dann ein $N \in \mathbb{N}$ so, dass $\|x_k - a\| < \epsilon$ für alle $k \geq N$. Für jedes $i = 1, \dots, n$ gilt dann

$$|x_{ki} - a_i| \leq \|x_k - a\| < \epsilon \text{ für } k \geq N.$$

D.h. $\lim_{k \rightarrow \infty} x_{ki} = a_i$.

“ $\Leftarrow$ ” Nun gelte $\lim_{k \rightarrow \infty} x_{ki} = a_i$ für alle $i = 1, \dots, n$. Sei $\epsilon > 0$ gegeben. Dann gibt es $N_i \in \mathbb{N}$ so, dass

$$|x_{ki} - a_i| < \epsilon' := \frac{\epsilon}{\sqrt{n}} \text{ für } k \geq N_i.$$

Für $k \geq N := \max\{N_1, \dots, N_n\}$ gilt dann

$$\|x_k - a\| = \sqrt{\sum_{i=1}^n |x_{ki} - a_i|^2} \leq \sqrt{n} \epsilon' = \epsilon.$$

D.h. $\lim_{k \rightarrow \infty} x_k = a$. □

Beispiel 1.11. Wir betrachten den normierten Raum $B(\mathbb{N}, \mathbb{R})$ aus Beispiel 1.3(c). Die Elemente von $B(\mathbb{N}, \mathbb{R})$ können als beschränkte Folgen in $\mathbb{R}$ aufgefasst werden. Sei $e^i \in B(\mathbb{N}, \mathbb{R})$ dasjenige Element mit $e^i(n) = 0$ für $n \neq i$, $e^i(n) = 1$ für $i = n$. Betrachte nun die Folge $(e^k)_{k \in \mathbb{N}}$ in $B(\mathbb{N}, \mathbb{R})$. Für jedes n gilt dann $\lim_{k \rightarrow \infty} e^k(n) = 0$ aber die Folge $(e^k)_{k \in \mathbb{N}}$ konvergiert nicht in $B(\mathbb{N}, \mathbb{R})$. Begründung: $(e^k)_{k \in \mathbb{N}}$ ist nicht einmal eine Cauchyfolge, denn für $k \neq l$ ist $\|e^k - e^l\|_\infty = 1$.

Mit Hilfe des Satzes 1.10 können wir einige von $\mathbb{R}$ bekannte Aussagen direkt auf den $\mathbb{R}^n$ übertragen:

Satz 1.12. *In $\mathbb{R}^n$ konvergiert jede Cauchyfolge, d.h. $\mathbb{R}^n$ ist vollständig.*

Beweis. Sei $(x_k)_{k \in \mathbb{N}}$ mit $x_k = (x_{k1}, \dots, x_{kn})$ eine Cauchyfolge in $\mathbb{R}^n$. Da $|x_{ki} - x_{li}| \leq \|x_k - x_l\|$ für jedes $i = 1, \dots, n$ gilt, ist jede Komponentenfolge $(x_{ki})_{k \in \mathbb{N}}$ eine Cauchyfolge in $\mathbb{R}$. Da $\mathbb{R}$ vollständig ist (Ana1, Satz 2.57), konvergiert $(x_{ki})_{k \in \mathbb{N}}$ in $\mathbb{R}$ für $i = 1, \dots, n$. Nach Satz 1.10 konvergiert also $(x_k)_{k \in \mathbb{N}}$ in $\mathbb{R}^n$. $\square$

Definition 1.13. Ein metrischer Raum heißt *vollständig*, wenn in ihm jede Cauchyfolge konvergiert. Einen normierten $\mathbb{R}$ -Vektorraum, der mit der durch die Norm gegebenen Metrik vollständig ist, nennt man *Banachraum*.

Beispiel 1.14. Der Vektorraum $C([0, 1], \mathbb{R})$ mit der 1-Norm aus Bsp. 1.3(d) ist nicht vollständig: Betrachte z.B. die Folge von Funktionen $(f_k)_{k \in \mathbb{N}}$ mit

$$f_k(x) := \begin{cases} 1 & \text{für } 0 \leq x \leq \frac{1}{2} \\ -kx + 1 + \frac{k}{2} & \text{für } \frac{1}{2} \leq x \leq \frac{1}{2} + \frac{1}{k} \\ 0 & \text{für } \frac{1}{2} + \frac{1}{k} \leq x \leq 1. \end{cases}$$

Dies ist eine Cauchyfolge, die nicht konvergiert.

Stetigkeit

Es gibt viele verschiedene Charakterisierungen von Stetigkeit. Wie in Analysis 1 benutzen wir das “ ϵ - δ ”-Kriterium als Definition und beweisen dann das Folgenkriterium. Im nächsten Abschnitt werden wir noch ein weiteres Kriterium kennenlernen.

Definition 1.15. Seien X und Y metrische Räume. Eine Abbildung $f: X \rightarrow Y$ heißt *stetig im Punkt* $a \in X$, wenn gilt: Zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ so, dass

$$d(f(x), f(a)) < \epsilon \text{ für alle } x \in X \text{ mit } d(x, a) < \delta.$$

Die Abbildung f heißt *stetig*, wenn sie in jedem Punkt aus X stetig ist.

Beispiel 1.16. (a) Auf jedem metrischen Raum X ist die Identität $\text{id}_X: X \rightarrow X$ stetig.

(b) Sei X ein metrischer Raum und sei $x_0 \in X$. Dann ist die Funktion $f: X \rightarrow \mathbb{R}$, $x \mapsto d(x, x_0)$, stetig. Denn nach der inversen Dreiecksungleichung gilt für beliebige $x, a \in X$

$$|f(x) - f(a)| = |d(x, x_0) - d(a, x_0)| \leq d(x, a);$$

wir können also stets $\delta := \epsilon$ wählen.

- (c) Wir betrachten $C([0, 1], \mathbb{R})$ mit der durch die Supremumsnorm $\|\cdot\|_\infty$ gegebenen Metrik. Die Abbildungen

$$e: C([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, \quad e(f) = f\left(\frac{1}{2}\right)$$

und

$$I: C([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, \quad I(f) = \int_0^1 f(x) dx$$

sind stetig. Denn für $f, g \in C([0, 1], \mathbb{R})$ gilt $|f(x) - g(x)| < \|f - g\|_\infty$ für alle $x \in [0, 1]$ und¹

$$\begin{aligned} |I(f) - I(g)| &= \left| \int_0^1 (f(x) - g(x)) dx \right| \\ &\leq \int_0^1 |f(x) - g(x)| dx \\ &\leq 1 \cdot \|f - g\|_\infty \end{aligned}$$

Wir können also in beiden Fällen $\delta = \epsilon$ wählen.

- (d) Sei $C^1([0, 1], \mathbb{R}) \subseteq C([0, 1], \mathbb{R})$ der Vektorraum der stetigen Funktionen auf $[0, 1]$ die auf $(0, 1)$ stetig differenzierbar sind. Wir versehen $C^1([0, 1], \mathbb{R})$ wieder mit der Supremumsnorm $\|\cdot\|_\infty$. Die Abbildung

$$D: C^1([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, \quad D(f) = f'\left(\frac{1}{2}\right)$$

ist nicht stetig. Z.B. liefert $f_k(x) = \frac{1}{k} \cos(k\pi x)$ eine Folge $(f_k)_{k \in \mathbb{N}}$ in $C^1([0, 1], \mathbb{R})$, welche gegen die Nullfunktion konvergiert. Aber $D(f_k) = -\pi \sin(\frac{k\pi}{2})$, d.h. die Folge reeller Zahlen $(D(f_k))_{k \in \mathbb{N}}$ konvergiert nicht.

Satz 1.17 (Folgenkriterium für Stetigkeit). *Seien X, Y metrische Räume. Die Abbildung $f: X \rightarrow Y$ ist genau dann stetig in $a \in X$, wenn gilt: Für jede Folge $(x_k)_{k \in \mathbb{N}}$ in X mit $\lim_{k \rightarrow \infty} x_k = a$ gilt $\lim_{k \rightarrow \infty} f(x_k) = f(a)$.*

Beweis. “ $\Rightarrow$.” Sei f stetig in $a \in X$. Sei $(x_k)_{k \in \mathbb{N}}$ eine Folge in X mit $\lim_{k \rightarrow \infty} x_k = a$. Zu zeigen: $\lim_{k \rightarrow \infty} f(x_k) = f(a)$. Sei also $\epsilon > 0$ gegeben. Dann gibt es $\delta > 0$ so, dass $d(f(x), f(a)) < \epsilon$ für alle $x \in X$ mit $d(x, a) < \delta$. Wähle nun $N \in \mathbb{N}$ so groß, dass $d(x_k, a) < \delta$ für alle $k \geq N$. Dann gilt $d(f(x_k), f(a)) < \epsilon$ für alle $k \geq N$, d.h. $\lim_{k \rightarrow \infty} f(x_k) = f(a)$.

“ $\Leftarrow$.” Angenommen f ist nicht stetig in a . Wir konstruieren eine Folge $(x_k)_{k \in \mathbb{N}}$ mit $\lim_{k \rightarrow \infty} x_k = a$ so, dass $(f(x_k))_{k \in \mathbb{N}}$ nicht gegen $f(a)$ konvergiert. Da f nicht stetig in a ist, gibt es ein $\epsilon > 0$ mit folgender Eigenschaft: Für jedes $\delta > 0$ gibt es ein $x \in X$ mit $d(x, a) < \delta$ aber $d(f(x), f(a)) \geq \epsilon$. Insbesondere finden wir zu jedem $k \in \mathbb{N}$ ein $x_k \in X$ mit $d(x_k, a) < \frac{1}{k}$ aber $d(f(x_k), f(a)) \geq \epsilon$. Nach Konstruktion gilt dann $\lim_{k \rightarrow \infty} x_k = a$, aber $(f(x_k))_{k \in \mathbb{N}}$ kann nicht gegen $f(a)$ konvergieren. $\square$

¹Allgemein: Sind $a < b$ reelle Zahlen und ist $h: [a, b] \rightarrow \mathbb{R}$ integrierbar, so gilt nach Ana1, Satz 6.8: $|\int_a^b h(x) dx| \leq \int_a^b |h(x)| dx \leq \int_a^b \|h\|_\infty dx = (b - a) \|h\|_\infty$.

Wie zeigt man nun z.B., dass die Abbildung

$$\mathbb{R} \times \mathbb{R}^\times \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto \begin{pmatrix} x^3 y^5 \\ \frac{\cos(x)}{y} \end{pmatrix}$$

stetig ist?

Satz 1.18 (“Verkettung stetiger Abbildungen ist stetig”). *Seien X, Y, Z metrische Räume, und $f: X \rightarrow Y$ und $g: Y \rightarrow Z$ Abbildungen. Ist f stetig in $a \in X$ und g stetig in $f(a)$, so ist auch $g \circ f$ stetig in a .*

Beweis. Wir benutzen das Folgenkriterium. Ist $(x_k)_{k \in \mathbb{N}}$ eine Folge in X mit $\lim x_k = a$, so konvergiert wegen der Stetigkeit von f in a die Folge $(f(x_k))_{k \in \mathbb{N}}$ gegen $f(a)$. Wegen der Stetigkeit von g in $f(a)$ konvergiert dann $(g(f(x_k)))_{k \in \mathbb{N}}$ gegen $g(f(a))$. $\square$

Beachte: Eine Abbildung $f: X \rightarrow \mathbb{R}^n$ ist ein n -Tupel von Abbildungen $f = (f_1, \dots, f_n)$ mit $f_i: X \rightarrow \mathbb{R}$. Damit haben wir:

Satz 1.19. *Sei X ein metrischer Raum. Eine Abbildung $f = (f_1, \dots, f_n): X \rightarrow \mathbb{R}^n$ ist genau dann stetig (in $a \in X$), wenn alle $f_i: X \rightarrow \mathbb{R}$ (in a) stetig sind.*

Beweis. Dies folgt aus den Sätzen 1.10 und 1.17. $\square$

Will man in der Praxis die Stetigkeit einer Abbildung zeigen, so kommt man oft mit den beiden vorangehenden Sätzen und folgenden “Grundbausteinen” aus.

Satz 1.20. *Folgende Abbildungen sind stetig:*

- (a) $\text{add}: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad (x, y) \mapsto x + y,$
- (b) $\text{mult}: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad (x, y) \mapsto xy,$
- (c) $\text{inv}: \mathbb{R}^\times \rightarrow \mathbb{R}^\times, \quad x \mapsto x^{-1}.$

Beweis. Dies ist schon aus der Analysis 1 bekannt und hier nur der Vollständigkeit halber wiederholt. Man kann z.B. das Folgenkriterium mit Ana1 Sätze 2.37, 2.38 verwenden. $\square$

Korollar 1.21. *Sei X ein metrischer Raum, und seien $f, g: X \rightarrow \mathbb{R}$ stetige Abbildungen. Dann sind auch die Abbildungen $f + g: X \rightarrow \mathbb{R}, x \mapsto f(x) + g(x)$, und $fg: X \rightarrow \mathbb{R}, x \mapsto f(x)g(x)$, stetig.*

Gilt noch $g(x) \neq 0$ für alle $x \in X$, so ist auch $\frac{f}{g}: X \rightarrow \mathbb{R}, x \mapsto \frac{f(x)}{g(x)}$, stetig.

Beweis. Wir zeigen exemplarisch die letzte Aussage. Wir schreiben $\frac{f}{g}$ als Verkettung

$$X \xrightarrow{(f,g)} \mathbb{R} \times \mathbb{R}^\times \xrightarrow{\text{id} \times \text{inv}} \mathbb{R} \times \mathbb{R}^\times \xrightarrow{\text{mult}} \mathbb{R}.$$

Die einzelnen Abbildungen sind stetig nach Sätzen 1.19 und 1.20 und damit auch ihre Verkettung $\frac{f}{g}$ nach Satz 1.18. $\square$

Offene und abgeschlossene Teilmengen

Im Folgenden sei $X = (X, d)$ stets ein metrischer Raum.

Notation 1.22. Für $a \in X$ und $r > 0$ bezeichne

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

die (*offene*) *Kugel* oder den *Ball* um a vom Radius r .

Definition 1.23. (a) Eine Teilmenge $V \subseteq X$ heißt *Umgebung* des Punktes $a \in X$, falls es ein $\epsilon > 0$ mit $B_\epsilon(a) \subseteq V$ gibt.

(b) Eine Teilmenge $U \subseteq X$ heißt *offen*, falls sie Umgebung jedes ihrer Punkte ist, d.h. wenn für jedes $x \in U$ ein $\epsilon > 0$ existiert mit $B_\epsilon(x) \subseteq U$.

Beispiel 1.24. (a) Seien $a < b$ reelle Zahlen. Dann sind (a, b) , $(-\infty, a)$ und (b, ∞) offen: Für $x \in (a, b)$ setze $\epsilon := \min\{x - a, b - x\}$; dann ist $B_\epsilon(x) = (x - \epsilon, x + \epsilon) \subseteq (a, b)$ und daher (a, b) eine Umgebung von x . Die Intervalle $[a, b]$ und $[a, b]$ sind nicht offen, aber Umgebungen jedes Punktes $x \in (a, b)$.

(b) Offene Kugeln sind offen: Ist X ein metrischer Raum, $a \in X$, $r > 0$, dann ist $B_r(a)$ offen: für $x \in B_r(a)$ gilt ja $d(x, a) < r$, also $\epsilon := r - d(x, a) > 0$, und wegen der Dreiecksungleichung gilt $B_\epsilon(x) \subseteq B_r(a)$; ist nämlich $y \in B_\epsilon(x)$, so gilt $d(y, a) \leq d(y, x) + d(x, a) \leq \epsilon + d(x, a) = r$.

(c) Wir betrachten $\mathbb{Z} \subset \mathbb{Q}$ mit der p -adischen Metrik aus Beispiel 1.6(e) für eine Primzahl p . Dann sind

$$\mathbb{Z} \supset p\mathbb{Z} \supset p^2\mathbb{Z} \supset \dots$$

offene Kugeln um 0.

Satz 1.25. Sei X ein metrischer Raum.

(a) Die Teilmengen $\emptyset$ und X sind offen.

(b) Endliche Durchschnitte offener Mengen sind offen: Sind $U_1, \dots, U_n \subseteq X$ offen, so auch $U_1 \cap \dots \cap U_n$.

(c) Beliebige Vereinigungen offener Mengen sind offen: Ist I irgendeine Indexmenge und $U_i, i \in I$, eine Familie offener Teilmengen von X , so ist auch $\bigcup_{i \in I} U_i \subseteq X$ offen.

Beliebige Durchschnitte offener Mengen müssen hingegen nicht offen sein: Z.B. ist $\bigcap_{n \in \mathbb{N}} (-\frac{1}{n}, +\frac{1}{n}) = \{0\} \subseteq \mathbb{R}$ nicht offen.

Beweis. (a) Die Menge X ist Umgebung jedes $x \in X$, also offen. Die leere Menge ist offen, weil es gar keine Bedingung gibt.

(b) Seien $U_1, \dots, U_n \subseteq X$ offen, und sei $x \in U := \bigcap_{i=1}^n U_i$. Da jedes U_i offen ist, gibt es ein $\epsilon_i > 0$ mit $B_{\epsilon_i}(x) \subseteq U_i$. Für $\epsilon := \min_{i=1, \dots, n} \epsilon_i$ gilt dann $\epsilon > 0$ und $B_\epsilon(x) \subseteq U$, d.h. U ist Umgebung von x . Da $x \in U$ beliebig war, ist U offen.

(c) Seien $U_i, i \in I$, offen, und sei $x \in U := \bigcup_{i \in I} U_i$. Dann gibt es ein $i \in I$ mit $x \in U_i$. Da $U_i \subseteq X$ offen ist, gibt es $\epsilon > 0$ mit $B_\epsilon(x) \subseteq U_i$. Also auch $B_\epsilon(x) \subseteq U$, und daher ist U offen. $\square$

Satz 1.26. *Seien X, Y metrische Räume. Eine Abbildung $f: X \rightarrow Y$ ist genau dann stetig (auf ganz X), wenn für jedes offene $U \subseteq Y$ auch das Urbild $f^{-1}(U) \subseteq X$ offen ist.*

Beweis. “ $\Rightarrow$ ”: Sei f stetig und sei $U \subseteq Y$ offen. Wir haben zu zeigen, dass $f^{-1}(U) \subseteq X$ offen ist. Sei also $a \in f^{-1}(U)$. Da U offen ist und $f(a) \in U$, gibt es ein $\epsilon > 0$ mit $B_\epsilon(f(a)) \subseteq U$. Da f stetig ist, gibt es ein $\delta > 0$ so, dass $d(f(x), f(a)) < \epsilon$ für alle $x \in X$ mit $d(x, a) < \delta$. D.h. $f(B_\delta(a)) \subseteq B_\epsilon(f(a))$. Also gilt $B_\delta(a) \subseteq f^{-1}(U)$ und daher ist $f^{-1}(U)$ offen.

“ $\Leftarrow$ ”: Nun gelte umgekehrt, dass die Urbilder offener Mengen offen sind. Sei $a \in X$. Wir zeigen, dass f in a stetig ist. Sei also $\epsilon > 0$ gegeben. Da $B_\epsilon(f(a)) \subseteq Y$ offen ist, ist nach Voraussetzung auch $f^{-1}(B_\epsilon(f(a))) \subseteq X$ offen. Insbesondere gibt es ein $\delta > 0$ mit $B_\delta(a) \subseteq f^{-1}(B_\epsilon(f(a)))$. D.h. aber $d(f(x), f(a)) < \epsilon$ für alle $x \in X$ mit $d(x, a) < \delta$. Also ist das ϵ - δ -Kriterium erfüllt und f stetig in a . $\square$

Definition 1.27. Eine Teilmenge A eines metrischen Raums X heißt *abgeschlossen*, wenn ihr Komplement $X \setminus A \subseteq X$ offen ist.

Beispiel 1.28. (a) Sind $a \leq b$ reelle Zahlen, so sind die Intervalle $[a, b]$ und auch $(-\infty, a]$, $[b, \infty)$ abgeschlossen, denn ihre Komplemente $(-\infty, a) \cup (b, \infty)$ bzw. (a, ∞) bzw. $(-\infty, b)$ sind offen nach Beispiel 1.24(a) und Satz 1.25.

(b) Punkte sind stets abgeschlossen: Ist X ein metrischer Raum und $a \in X$, so ist $\{a\} \subseteq X$ abgeschlossen. Denn ist $x \in X \setminus \{a\}$, so ist $\epsilon := d(x, a) > 0$ und $B_\epsilon(x) \subseteq X \setminus \{a\}$, also $X \setminus \{a\} \subseteq X$ offen.

(c) X und $\emptyset$ sind stets abgeschlossen, denn ihre Komplemente $\emptyset$ bzw. X sind offen nach Satz 1.25. Die Aussagen (b) und (c) des Satzes 1.25 liefern: Endliche Vereinigungen abgeschlossener Mengen sind abgeschlossen, beliebige Durchschnitte abgeschlossener Mengen sind abgeschlossen.

(d) Auch offene Kugeln können abgeschlossen sein: Z.B. in $\{0, 1\}^{\mathbb{N}}$ oder in $\mathbb{Z}$ mit der p -adischen Metrik.

Satz 1.29. *Sei X ein metrischer Raum. Eine Teilmenge $A \subseteq X$ ist genau dann abgeschlossen, wenn sie folgenabgeschlossen ist, d.h. wenn gilt: Ist $(x_k)_{k \in \mathbb{N}}$ eine Folge von Punkten $x_k \in A$, die gegen $x \in X$ konvergiert, so ist $x \in A$.*

Beweis. “ $\Rightarrow$.” Sei $A \subseteq X$ abgeschlossen und sei $(x_k)_{k \in \mathbb{N}}$ eine Folge in A mit $\lim_{k \rightarrow \infty} x_k = x$. Angenommen, $x \notin A$. Da $X \setminus A$ offen ist, gibt es ein $\epsilon > 0$ mit $B_\epsilon(x) \subseteq X \setminus A$. Da $(x_k)_{k \in \mathbb{N}}$ gegen x konvergiert, gilt $x_k \in B_\epsilon(x)$ für k groß genug, im Widerspruch zu $x_k \in A$.

“ $\Leftarrow$.” Angenommen, A ist nicht abgeschlossen. Dann ist $X \setminus A$ nicht offen, d.h. es gibt einen Punkt $x \in X \setminus A$, so dass für jede ϵ -Kugel $B_\epsilon(x)$ gilt, dass $A \cap B_\epsilon(x) \neq \emptyset$. Insbesondere finden wir für jedes $k \in \mathbb{N}$ ein $x_k \in A \cap B_{\frac{1}{k}}(x)$. Nach Konstruktion ist dann $(x_k)_{k \in \mathbb{N}}$ eine Folge von Punkten in A , die gegen $x \in X \setminus A$ konvergiert. Also ist A nicht folgenabgeschlossen. $\square$

Definition 1.30. Seien X ein metrischer Raum und $Y \subseteq X$ eine Teilmenge. Ein Punkt $x \in X$ heißt *Randpunkt von Y* , wenn in jeder Umgebung von x sowohl ein Punkt aus Y als auch ein Punkt aus $X \setminus Y$ liegt, d.h. wenn für jedes $\epsilon > 0$ gilt: $B_\epsilon(x) \cap Y \neq \emptyset$ und $B_\epsilon(x) \cap (X \setminus Y) \neq \emptyset$.

Die Menge aller Randpunkte (der *Rand*) von Y wird mit ∂Y bezeichnet.

Ein Randpunkt von Y braucht also selbst kein Punkt von Y zu sein.

Beispiel 1.31. (a) Sei $Y_1 = B_r(0) \subseteq \mathbb{R}^n$ die offene Kugel vom Radius $r > 0$ um 0. Dann gilt $\partial Y_1 = \{x \in \mathbb{R}^n : \|x\| = r\}$. Für $Y_2 = \{x \in \mathbb{R}^n : \|x\| \leq r\}$ gilt genauso $\partial Y_2 = \{x \in \mathbb{R}^n : \|x\| = r\}$.

Für beliebige metrische Räume X ist die analoge Aussage $\partial B_r(a) = \{x \in X : d(x, a) = r\}$ im Allgemeinen nicht richtig.

- (b) Für $\mathbb{Q} \subseteq \mathbb{R}$ gilt $\partial \mathbb{Q} = \mathbb{R}$, da es in jeder ϵ -Umgebung $(x - \epsilon, x + \epsilon)$ einer reellen Zahl x sowohl rationale als auch irrationale Zahlen gibt.
- (c) Die Definition des Randes ist symmetrisch in Y und dessen Komplement $X \setminus Y$. Also gilt $\partial Y = \partial(X \setminus Y)$.

Satz 1.32. Seien X ein metrischer Raum und $Y \subseteq X$ eine Teilmenge. Dann gelten:

- (a) Die Teilmenge $\overset{\circ}{Y} := Y \setminus \partial Y \subseteq X$ ist offen.
- (b) Die Teilmenge $\overline{Y} := Y \cup \partial Y \subseteq X$ ist abgeschlossen.
- (c) Der Rand $\partial Y \subseteq X$ ist abgeschlossen.

Man nennt $\overset{\circ}{Y}$ das *Innere* von Y und $\overline{Y}$ den *Abschluss* von Y .

Beweis. (a) Sei $a \in \overset{\circ}{Y}$. Da $a \in Y$ und a kein Randpunkt von Y ist, gibt es ein $\epsilon > 0$ mit $B_\epsilon(a) \subseteq Y$. Da $B_\epsilon(a)$ offen ist, ist es Umgebung jedes Punktes von $B_\epsilon(a)$. Also ist kein Punkt von $B_\epsilon(a)$ ein Randpunkt von Y , d.h. wir haben sogar $B_\epsilon(a) \subseteq \overset{\circ}{Y}$. Daher ist $\overset{\circ}{Y}$ offen.

(b) Es gilt $X \setminus \overline{Y} = (X \setminus Y) \setminus \partial Y$. Da $\partial Y = \partial(X \setminus Y)$ nach Bsp. 1.31(c), ist dies offen nach (a). Also ist $\overline{Y}$ abgeschlossen.

(c) Es gilt $\partial Y = \overline{Y} \setminus \overset{\circ}{Y} = \overline{Y} \cap (X \setminus \overset{\circ}{Y})$. Nach (a) bzw. (b) sind $X \setminus \overset{\circ}{Y}$ und $\overline{Y}$ abgeschlossen. Also ist auch ∂Y als deren Schnitt abgeschlossen. $\square$

Kompakte Teilmengen

Wir führen noch den Begriff der kompakten Teilmenge eines metrischen Raumes ein. Dieser spielt unter anderem deswegen eine wichtige Rolle, weil jede stetige $\mathbb{R}$ -wertige Funktion auf einer kompakten Teilmenge ein Maximum und ein Minimum annimmt.

Erinnerung 1.33. Der Satz von Bolzano-Weierstraß (für $\mathbb{R}$) besagt, dass jede beschränkte Folge reeller Zahlen eine konvergente Teilfolge besitzt (Ana1, Satz 2.65). Da für $a, b \in \mathbb{R}$ das Intervall $I = [a, b]$ beschränkt und folgenabgeschlossen ist, besitzt auch jede Folge $(x_k)_{k \in \mathbb{N}}$ von Elementen aus I eine konvergente Teilfolge in I , d.h. auch der Grenzwert dieser (jeder) konvergenten Teilfolge liegt in I .

Für allgemeine metrische Räume ist die analoge Aussage aber falsch. Betrachte z.B. den normierten Raum $B(\mathbb{N}, \mathbb{R})$ mit $\|\cdot\|_\infty$ und darin die Folge $(e^k)_{k \in \mathbb{N}}$ wie in Beispiel 1.11. Da $\|e^k\|_\infty = 1$ für alle $k \in \mathbb{N}$ gilt, ist die Folge beschränkt. Da aber für $k \neq l$ auch $\|e^k - e^l\|_\infty = 1$ ist, ist keine Teilfolge von $(e^k)_{k \in \mathbb{N}}$ Cauchyfolge; erst recht gibt es keine konvergenten Teilfolgen.

Eine Teilmenge eines metrischen Raumes, für die die analoge Aussage gilt, nennt man (folgen)kompakt (vgl. Ana1 Def. 3.23):

Definition 1.34. Ein metrischer Raum X heißt *folgenkompakt* oder einfach *kompakt*², wenn jede Folge in X eine konvergente Teilfolge besitzt.

Eine Teilmenge $A \subseteq X$ heißt *(folgen)kompakt*, wenn A mit der Teilraummetrik (siehe Bsp. 1.6(c)) (folgen)kompakt ist, d.h. wenn jede Folge $(x_k)_{k \in \mathbb{N}}$ von Punkten $x_k \in A$ eine Teilfolge besitzt, die gegen einen Punkt $a \in A$ konvergiert.

Zwei wichtige, direkte Konsequenzen der Definition sind die folgenden beiden Sätze:

Satz 1.35. Sei X ein kompakter metrischer Raum, und sei $f: X \rightarrow \mathbb{R}$ eine stetige Funktion. Dann nimmt f sein Maximum und Minimum an, d.h. es gibt $a, b \in X$ so, dass

$$f(a) = \sup\{f(x) : x \in X\} \quad f(b) = \inf\{f(x) : x \in X\}.$$

Insbesondere ist f beschränkt.

Beweis. Nach Definition des Supremums gibt es eine Folge in $\{f(x) : x \in X\}$ die gegen $\sup\{f(x) : x \in X\}$ konvergiert. D.h. es gibt eine Folge $(x_k)_{k \in \mathbb{N}}$ in X so, dass $\lim_{k \rightarrow \infty} f(x_k) = \sup\{f(x) : x \in X\}$ gilt. Da X kompakt ist, besitzt $(x_k)_{k \in \mathbb{N}}$ eine konvergente Teilfolge. Durch Übergang zu dieser Teilfolge können wir annehmen, dass die Folge $(x_k)_{k \in \mathbb{N}}$ selbst konvergiert. Sei $a \in X$ ihr Grenzwert. Da f stetig ist, gilt dann $f(a) = \lim_{k \rightarrow \infty} f(x_k) = \sup\{f(x) : x \in X\}$.

Wendet man das schon bewiesene auf die Funktion $-f$ an, so erhält man ein $b \in X$ mit $f(b) = \inf\{f(x) : x \in X\}$. $\square$

Definition 1.36. Seien X, Y metrische Räume und $f: X \rightarrow Y$ eine Abbildung. Man nennt f *gleichmäßig stetig*, wenn gilt: Zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ so, dass

$$d(f(x_1), f(x_2)) < \epsilon \quad \text{für alle } x_1, x_2 \in X \text{ mit } d(x_1, x_2) < \delta.$$

²Man kann die Begriffe “folgenkompakt” und “kompakt” auch in allgemeinerer Situation definieren.

Im Allgemeinen sind diese Begriffe verschieden. Für metrische Räume sind sie aber äquivalent.

Satz 1.37. Seien X, Y metrische Räume. Ist $f: X \rightarrow Y$ stetig und ist X kompakt, dann ist f gleichmäßig stetig.

Beweis. Angenommen, f ist nicht gleichmäßig stetig. Dann gibt es ein $\epsilon > 0$ mit folgender Eigenschaft: Für alle $k \in \mathbb{N}$ existieren $x_k, x'_k \in X$ mit $d(x_k, x'_k) < \frac{1}{k}$ und $d(f(x_k), f(x'_k)) \geq \epsilon$. Da X kompakt ist, besitzt $(x_k)_{k \in \mathbb{N}}$ eine konvergente Teilfolge $(x_{k_j})_{j \in \mathbb{N}}$. Sei $x = \lim_{j \rightarrow \infty} x_{k_j}$. Dann gilt auch $\lim_{j \rightarrow \infty} x'_{k_j} = x$. Es gilt

$$\epsilon \leq d(f(x_{k_j}), f(x'_{k_j})) \leq d(f(x_{k_j}), f(x)) + d(f(x), f(x'_{k_j})).$$

Da f stetig ist, konvergiert aber die rechte Seite gegen 0 für $j \rightarrow \infty$, Widerspruch. $\square$

Satz 1.38. Stetige Bilder kompakter Mengen sind kompakt: Ist $f: X \rightarrow Y$ eine stetige Abbildung metrischer Räume, und ist $A \subseteq X$ kompakt, so ist auch das Bild $f(A) \subseteq Y$ kompakt.

Beweis. Eine beliebige Folge in $f(A)$ ist von der Form $(f(x_k))_{k \in \mathbb{N}}$ mit einer Folge $(x_k)_{k \in \mathbb{N}}$ in A . Da A kompakt ist, besitzt $(x_k)_{k \in \mathbb{N}}$ eine in A konvergente Teilfolge, o.B.d.A. die Folge selbst. Sei $a \in A$ ihre Grenzwert. Da f stetig ist, gilt dann $f(a) = \lim_{k \rightarrow \infty} f(x_k)$, d.h. die Folge $(f(x_k))_{k \in \mathbb{N}}$ konvergiert. Daher ist $f(A)$ kompakt. $\square$

Kompakte Teilmengen haben also gute Eigenschaften. Kann man sie charakterisieren?

Wir nennen eine Teilmenge $A \subseteq X$ *beschränkt*, wenn es ein $a \in X$ und eine reelle Zahl $r > 0$ gibt so, dass $A \subseteq B_r(a)$.

Satz 1.39. Sei X ein metrischer Raum, und sei $A \subseteq X$ kompakt. Dann ist A beschränkt und abgeschlossen.

Wie wir gesehen haben, ist die Umkehrung im Allgemeinen falsch; für $\mathbb{R}^n$ gilt sie aber, siehe Satz 1.41 unten.

Beweis. Wähle irgendein $a \in X$. Nach Beispiel 1.16(b) ist die Funktion $f: A \rightarrow \mathbb{R}$, $x \mapsto d(x, a)$, stetig. Nach Satz 1.35 ist sie auf A beschränkt, d.h. es gibt ein $r > 0$ so, dass $d(x, a) < r$ für alle $x \in A$. D.h. $A \subseteq B_r(a)$ und A ist beschränkt.

Um zu zeigen, dass A abgeschlossen ist, benutzen wir Satz 1.29. Sei also $(x_k)_{k \in \mathbb{N}}$ eine Folge von Punkten in A , die gegen ein $x \in X$ konvergiert. Wir müssen zeigen, dass $x \in A$ gilt. Da aber A kompakt ist, besitzt $(x_k)_{k \in \mathbb{N}}$ eine Teilfolge, die gegen einen Punkt $a \in A$ konvergiert. Da aber $(x_k)_{k \in \mathbb{N}}$ selbst gegen x konvergiert, gilt dies auch für jede Teilfolge; wegen der Eindeutigkeit des Grenzwerts gilt also $x = a \in A$. $\square$

Satz 1.40. Sei X ein kompakter metrischer Raum. Dann ist eine Teilmenge $A \subseteq X$ genau dann kompakt, wenn sie abgeschlossen ist.

Beweis. “ $\Rightarrow$ ”: Nach Satz 1.39 ist $A \subseteq X$ abgeschlossen.

“ $\Leftarrow$ ”: Sei $(x_k)_{k \in \mathbb{N}}$ eine Folge in A . Da X kompakt ist, besitzt sie eine (in X) konvergente Teilfolge. Da $A \subseteq X$ abgeschlossen ist, liegt der Grenzwert dieser Teilfolge schon in A (Satz 1.29). $\square$

Satz 1.41 (Heine-Borel). *Eine Teilmenge $A \subseteq \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.*

Beweis. “ $\Rightarrow$ ” folgt aus Satz 1.39.

“ $\Leftarrow$ ”: Wir benutzen den Satz von Bolzano-Weierstraß. Sei also $A \subseteq \mathbb{R}^n$ beschränkt und abgeschlossen, und sei $(x_k)_{k \in \mathbb{N}}$ eine Folge in A mit $x_k = (x_{k1}, \dots, x_{kn})$. Dann ist jede ihrer Komponentenfolgen $(x_{ki})_{k \in \mathbb{N}}$, $i = 1, \dots, n$, beschränkt. Durch Übergang zu einer Teilfolge (und Umnummerierung) können wir also erreichen, dass $(x_{k1})_{k \in \mathbb{N}}$ konvergiert. Nach Übergang zu einer weiteren Teilfolge können wir erreichen, dass auch $(x_{k2})_{k \in \mathbb{N}}$ konvergiert. Nach n Schritten haben wir also eine Teilfolge gefunden, welche komponentenweise konvergiert. Nach Satz 1.10 konvergiert dann diese Teilfolge selbst. $\square$

Man kann kompakte metrische Räume auch durch eine Überdeckungseigenschaft charakterisieren (siehe Satz 1.44). Diese wird im Allgemeinen als Definition benutzt.

Definition 1.42. Sei A eine Teilmenge eines metrischen Raumes X . Eine *offene Überdeckung* von A ist eine Familie $(U_i)_{i \in I}$ offener Teilmengen $U_i \subseteq X$ mit

$$A \subseteq \bigcup_{i \in I} U_i.$$

Dabei ist I eine beliebige Indexmenge.

Wir sagen, die offene Überdeckung $(U_i)_{i \in I}$ von A besitze eine *endliche Teilüberdeckung*, falls es endlich viele Indizes $i_1, \dots, i_N \in I$ gibt so, dass

$$A \subseteq \bigcup_{k=1, \dots, N} U_{i_k}$$

gilt.

Beispiel 1.43. Das offene Intervall $(0, 1)$ besitzt z.B. die offenen Überdeckungen $(U_i)_{i=1,2}$ mit $U_1 = (-1, \frac{3}{4})$, $U_2 = (\frac{1}{2}, 2)$ und $(V_n)_{n \in \mathbb{N}}$ mit $V_n = (\frac{1}{n}, 1 - \frac{1}{n})$. Die Überdeckung $(U_i)_{i=1,2}$ ist selbst endlich, besitzt also trivialerweise eine endliche Teilüberdeckung. Die Überdeckung $(V_n)_{n \in \mathbb{N}}$ hingegen besitzt keine endliche Teilüberdeckung. Denn sonst müsste für geeignetes $N \in \mathbb{N}$ schon $(0, 1) \subseteq (\frac{1}{N}, 1 - \frac{1}{N})$ gelten.

Satz 1.44. *Ein metrischer Raum X ist genau dann kompakt, wenn jede offene Überdeckung von X eine endliche Teilüberdeckung besitzt.*

Beweis. Wir zeigen nur “ $\Leftarrow$ ”. Für “ $\Rightarrow$ ” siehe z.B. Fritzsche, Grundkurs Analysis 2, 1.1.11.

Wir nehmen also an, dass jede offene Überdeckung von X eine endliche Teilüberdeckung besitzt. Angenommen, es gibt eine Folge $(x_k)_{k \in \mathbb{N}}$ in X , die keine konvergente Teilfolge besitzt. Dann gibt es zu jedem Punkt $x \in X$ eine offene Menge $U_x \ni x$, die nur endlich viele Folgenglieder enthält. (Betrachte z.B. $B_{\frac{1}{k}}(x)$ für $k \in \mathbb{N}$. Lägen in jeder dieser offenen Mengen unendlich viele Folgenglieder, so könnten wir eine Teilfolge konstruieren, die gegen x konvergiert.) Nun ist $(U_x)_{x \in X}$ eine offene Überdeckung von X . Also gibt es endlich viele Punkte $x_1, \dots, x_N \in X$ so, dass

$$X \subseteq \bigcup_{i=1, \dots, N} U_{x_i}.$$

Das kann aber nicht sein, weil jedes U_{x_i} nur endlich viele Folgenglieder enthält. $\square$

2 Differenzialrechnung im $\mathbb{R}^n$

Kurven im $\mathbb{R}^n$

Definition 2.1. Eine (*parametrisierte*) *Kurve* im $\mathbb{R}^n$ ist eine stetige Abbildung

$$\alpha: I \rightarrow \mathbb{R}^n$$

für ein Intervall $I \subseteq \mathbb{R}$. Die *Bahn* oder *Spur* der Kurve ist das Bild

$$\{\alpha(t) : t \in I\} \subseteq \mathbb{R}^n.$$

Die Bahn der Kurve darf nicht mit dem Graphen Γ_α der Abbildung α verwechselt werden. Der Graph ist die Teilmenge

$$\Gamma_\alpha = \{(t, \alpha(t)) : t \in I\} \subseteq I \times \mathbb{R}^n.$$

Beispiel 2.2. (a) Seien $a \in \mathbb{R}^n$, $v \in \mathbb{R}^n \setminus \{0\}$. Dann ist $\alpha: \mathbb{R} \rightarrow \mathbb{R}^n$, $\alpha(t) = a + tv$, eine Kurve. Ihre Bahn ist eine Gerade mit Aufpunkt a und Richtung v .

(b) Die Kurve $\alpha: [0, 2\pi] \rightarrow \mathbb{R}^2$, $\alpha(t) = (\cos(t), \sin(t))$, beschreibt einen Kreis mit Radius 1, durchlaufen in mathematisch positiver Richtung, d.h. gegen den Uhrzeigersinn. Die Kurve $\beta: [0, 2\pi] \rightarrow \mathbb{R}^2$, $\beta(t) = (\cos(2t), -\sin(2t))$, durchläuft den Kreis mit Radius 1 zweimal im Uhrzeigersinn, hat aber die gleiche Bahn wie α .

(c) Die “Neilsche Parabel” wird durch die Kurve $\alpha: \mathbb{R} \rightarrow \mathbb{R}^2$, $\alpha(t) = (t^2, t^3)$, beschrieben.

(d) Sei $I \subseteq \mathbb{R}$ ein Intervall, und sei $f: I \rightarrow \mathbb{R}$ eine stetige Funktion. Dann können wir die parametrisierte Kurve $\alpha: I \rightarrow \mathbb{R}^2$, $\alpha(t) = (t, f(t))$, betrachten. Ihre Bahn ist der Graph von f .

Definition 2.3. Sei $\alpha: I \rightarrow \mathbb{R}^n$ eine Kurve mit $\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))$ für alle $t \in I$. Die Kurve α heißt *differenzierbar* bzw. *stetig differenzierbar*, wenn alle α_i , $i = 1, \dots, n$, differenzierbar bzw. stetig differenzierbar sind.

Ist α differenzierbar, so heißt für $t \in I$

$$\alpha'(t) := (\alpha'_1(t), \dots, \alpha'_n(t))$$

der *Tangentenvektor* der Kurve α zum Parameter t .

Alle Kurven aus Beispiel 2.2 sind differenzierbar.

Bemerkung 2.4. Sei $\alpha: I \rightarrow \mathbb{R}^n$ eine differenzierbare Kurve, und sei $t \in I$. Dann gilt

$$\alpha'(t) = \lim_{h \rightarrow 0} \frac{\alpha(t+h) - \alpha(t)}{h},$$

genauer: Für jede Folge reeller Zahlen $(h_k)_{k \in \mathbb{N}}$ mit $t + h_k \in I$ für alle $k \in \mathbb{N}$ und $\lim_{k \rightarrow \infty} h_k = 0$ gilt $\lim_{k \rightarrow \infty} \frac{\alpha(t+h_k) - \alpha(t)}{h_k} = \alpha'(t)$.

Beweis. Nach Definition der Ableitung reellwertiger Funktionen gilt

$$\alpha'_i(t) = \lim_{h \rightarrow 0} \frac{\alpha_i(t+h) - \alpha_i(t)}{h}$$

für jedes $i = 1, \dots, n$. Die Behauptung folgt daher aus Satz 1.10. $\square$

Beachte, dass der Tangentialvektor der Kurve α in einem Parameter $t \in I$ explizit von der Parametrisierung abhängt, nicht nur von der Bahn. Ist $\alpha'(t) \neq 0$, so ist die Gerade $\{\alpha(t) + \theta\alpha'(t) : \theta \in \mathbb{R}\}$ die Tangente an die Kurve im Punkt $\alpha(t)$.

Funktionen mehrerer Variablen

Im Gegensatz zu den vektorwertigen Funktionen einer Variablen im letzten Abschnitt wenden wir uns jetzt reellwertigen Funktionen zu, die aber von mehreren Variablen abhängen, wie z.B.

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f(x, y) = x^2 + y^2$$

oder

$$g: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad g(x, y) = x^2 - y^2.$$

Man veranschaulicht solche Funktionen, indem man den Graphen oder die Niveaumengen skizziert:

Definition 2.5. Sei $U \subseteq \mathbb{R}^n$ und $f: U \rightarrow \mathbb{R}$ eine Funktion. Der *Graph* von f ist

$$\Gamma_f := \{(x, f(x)) : x \in U\} \subseteq U \times \mathbb{R} \subseteq \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}.$$

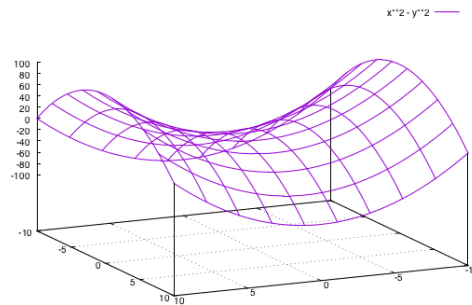


Abbildung 1: Graph der Funktion $g(x, y) = x^2 - y^2$

Definition 2.6. Die *Niveaumengen* der Funktion $f: U \rightarrow \mathbb{R}$ sind die Teilmengen

$$f^{-1}(c) = \{x \in U : f(x) = c\} \subseteq U.$$

Beispiel 2.7. (a) Betrachte die Funktion $f(x, y) = x^2 + y^2$. Für $c < 0$ ist $f^{-1}(c) = \emptyset$, $f^{-1}(0) = \{0\}$, und für $c > 0$ sind die Niveaumengen

$$f^{-1}(c) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = c\}$$

Kreise um den Ursprung.

(b) Betrachte die Funktion $g(x, y) = x^2 - y^2$. Es gilt

$$g(x, y) = 0 \Leftrightarrow y^2 = x^2 \Leftrightarrow y = \pm x,$$

die Niveaumenge $g^{-1}(0)$ ist also die Vereinigung der beiden Diagonalen in der x - y -Ebene. Für $c \neq 0$ haben wir

$$g(x, y) = c \Leftrightarrow y^2 = x^2 - c \Leftrightarrow y = \pm \sqrt{x^2 - c},$$

die Niveaumengen sind also Hyperbeln.

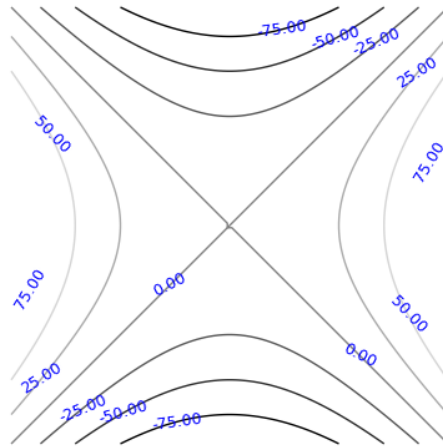


Abbildung 2: Niveaumengen der Funktion $g(x, y) = x^2 - y^2$

Nun zur Differenzierbarkeit. Wir werden im Folgenden verschiedene Differenzierbarkeitsbegriffe für Funktionen mehrerer Variablen kennenlernen und folgende Implikationen beweisen:

$$\text{stetig partiell differenzierbar} \implies (\text{total}) \text{ differenzierbar} \implies \text{partiell differenzierbar}$$

Wir beginnen mit dem Begriff der *partiellen Differenzierbarkeit*.

Wie üblich sei $e_i = (0, \dots, \overset{i}{1}, \dots, 0) \in \mathbb{R}^n$ der i -te Einheitsvektor.

Definition 2.8. Sei $U \subseteq \mathbb{R}^n$ offen. Eine Funktion $f: U \rightarrow \mathbb{R}$ heißt *partiell differenzierbar* in $x \in U$, wenn für jedes $i = 1, \dots, n$ der Grenzwert

$$D_i f(x) = \lim_{h \rightarrow 0} \frac{f(x + h e_i) - f(x)}{h}$$

existiert. Die Zahl $D_i f(x) \in \mathbb{R}$ nennt man *i-te partielle Ableitung* von f an der Stelle x . Weitere Schreibweisen:

$$D_i f(x) = \frac{\partial f}{\partial x_i}(x) = \frac{\partial}{\partial x_i} f(x)$$

Heißen die Koordinaten $(x, y) \in \mathbb{R}^2$ oder $(x, y, z) \in \mathbb{R}^3$, so schreiben wir entsprechend $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \dots$

Schreibe $x = (x_1, \dots, x_n)$. Betrachte die Funktion g mit

$$g(t) = f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n),$$

d.h. wir halten in f alle Koordinaten bis auf die i -te fest. Da $U \subseteq \mathbb{R}^n$ offen ist, ist g auf einem Intervall $(x_i - \epsilon, x_i + \epsilon) \subseteq \mathbb{R}$ mit geeignetem $\epsilon > 0$ definiert. Dann ist $D_i f(x) = g'(x_i)$ mit der gewöhnlichen Ableitung der Funktion g .

Beispiel 2.9. Für die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = x^2 + y^2$, gilt

$$\frac{\partial f}{\partial x}(x, y) = 2x, \quad \frac{\partial f}{\partial y}(x, y) = 2y.$$

Definition 2.10. Sei $U \subseteq \mathbb{R}^n$ offen. Die Funktion $f: U \rightarrow \mathbb{R}$ heißt *partiell differenzierbar*, wenn sie in jedem Punkt aus U partiell differenzierbar ist. Ist dies der Fall, so sind die partiellen Ableitungen $D_i f$ wieder Funktionen $D_i f: U \rightarrow \mathbb{R}$, und f heißt *stetig partiell differenzierbar*, wenn alle $D_i f$, $i = 1, \dots, n$, stetig sind.

Beispiel 2.11. Die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = x^2 + y^2$ ist stetig partiell differenzierbar.

Betrachte nun die Funktion

$$g: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad g(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{für } (x, y) \neq 0 \\ 0 & \text{für } (x, y) = 0. \end{cases}$$

Die Funktion ist in 0 nicht stetig. Denn $(\frac{1}{k}, \frac{1}{k})_{k \in \mathbb{N}}$ ist eine Folge in $\mathbb{R}^2$ mit Grenzwert 0, aber $g(\frac{1}{k}, \frac{1}{k}) = \frac{1}{2}$ für alle $k \in \mathbb{N}$; insbesondere $\lim_{k \rightarrow \infty} g(\frac{1}{k}, \frac{1}{k}) = \frac{1}{2} \neq 0 = g(0)$.

Behauptung: Die Funktion g ist auf ganz $\mathbb{R}^2$ partiell differenzierbar, aber nicht stetig partiell differenzierbar.

Beweis. Betrachte $(x, y) \neq 0$. Für festes $y \neq 0$ ist die Funktion $x \mapsto g(x, y)$ auf ganz $\mathbb{R}$ differenzierbar, und nach Produkt- und Quotientenregel erhalten wir

$$\begin{aligned} \frac{\partial g}{\partial x}(x, y) &= \frac{(x^2 + y^2)y - xy \cdot 2x}{(x^2 + y^2)^2} \\ &= \frac{y(y^2 - x^2)}{(x^2 + y^2)^2} \end{aligned}$$

Für $y = 0$ gilt $g(x, 0) = 0$ für alle $x \in \mathbb{R}$, d.h. g ist auch für $y = 0$ partiell nach x differenzierbar mit

$$\frac{\partial g}{\partial x}(x, 0) = 0.$$

Genauso existiert für alle $(x, y) \in \mathbb{R}^2$ die partielle Ableitung $\frac{\partial g}{\partial y}(x, y)$.

Die partielle Ableitung $\frac{\partial g}{\partial x}$ ist aber in 0 nicht stetig, denn

$$\frac{\partial g}{\partial x}\left(0, \frac{1}{k}\right) = k \not\rightarrow 0 = \frac{\partial g}{\partial x}(0) \text{ für } k \rightarrow \infty.$$

□

Definition 2.12. Sei $U \subseteq \mathbb{R}^n$ offen und sei $f: U \rightarrow \mathbb{R}$ differenzierbar. Für $x \in U$ heißt der Vektor

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)$$

(gesprochen ‘Nabla f von x ’) der *Gradient* von f in x .

Die Abbildung $\nabla f: U \rightarrow \mathbb{R}^n$ ordnet also jedem Punkt aus U wieder einen Vektor im $\mathbb{R}^n$ zu. Allgemein nennt man so eine Abbildung ein *Vektorfeld*.

Beispiel 2.13. Für $f(x, y) = x^2 + y^2$ gilt $\nabla f(x, y) = (2x, 2y)$.

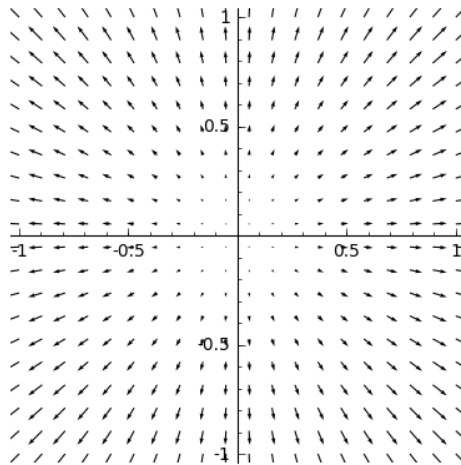


Abbildung 3: Gradientenfeld der Funktion $f(x, y) = x^2 + y^2$

Für $g(x, y) = x^2 - y^2$ gilt $\nabla g(x, y) = (2x, -2y)$.

Anschauliche Bedeutung des Gradienten: “Der Gradient von f zeigt in die Richtung des stärksten Anstiegs von f ”. Diese Aussage werden wir später präzisieren und beweisen, nachdem wir den Begriff der totalen Differenzierbarkeit eingeführt haben.

Definition 2.14. Sei $U \subseteq \mathbb{R}^n$ offen. Eine Funktion $f: U \rightarrow \mathbb{R}$ heißt *zweimal partiell differenzierbar* (bzw. *zweimal stetig partiell differenzierbar*), wenn f partiell differenzierbar ist und alle Funktionen $D_i f: U \rightarrow \mathbb{R}$, $i = 1, \dots, n$, wieder partiell differenzierbar (bzw. stetig partiell differenzierbar) sind.

Schreibe $D_j D_i f(x)$ oder auch $\frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ für die j -te partielle Ableitung von $D_i f$ an der Stelle x .

Analog definieren wir die Begriffe *k-mal partiell* (bzw. *stetig partiell*) *differenzierbar*.

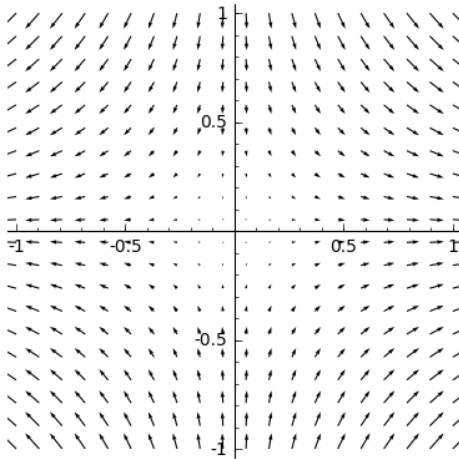


Abbildung 4: Gradientenfeld der Funktion $g(x, y) = x^2 - y^2$

Spielt die Reihenfolge der Ableitungen hier eine Rolle?

Beispiel 2.15. Für $f(x, y) = x^2 + y^2$ gilt

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial x}(x, y) &= 2, & \frac{\partial^2 f}{\partial x \partial y}(x, y) &= 0, \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) &= 0, & \frac{\partial^2 f}{\partial y \partial y}(x, y) &= 2. \end{aligned}$$

Beispiel 2.16. Betrachte nun die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$,

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

Dann gilt für $(x, y) \neq (0, 0)$:

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \frac{\partial}{\partial x} \left(\frac{x^3 y - y^3 x}{x^2 + y^2} \right) \\ &= \frac{(3x^2 y - y^3)(x^2 + y^2) - (x^3 y - y^3 x)2x}{(x^2 + y^2)^2} \\ &= \frac{x^4 y + 4x^2 y^3 - y^5}{(x^2 + y^2)^2}, \end{aligned}$$

also $\frac{\partial f}{\partial x}(0, y) = -y$ für $y \neq 0$. Da $f(x, 0) = 0$ für alle $x \in \mathbb{R}$ gilt $\frac{\partial f}{\partial x}(0, 0) = 0$. Also gilt

$$\frac{\partial f}{\partial x}(0, y) = -y \quad \text{für alle } y \in \mathbb{R}$$

und daher

$$\boxed{\frac{\partial^2 f}{\partial y \partial x}(0, 0) = -1}.$$

Da $f(x, y) = -f(y, x)$ gilt, ergibt sich für $(x, y) \neq (0, 0)$

$$\frac{\partial f}{\partial y}(x, y) = \frac{x^5 - 4x^3y^2 - xy^4}{(x^2 + y^2)^2},$$

also $\frac{\partial f}{\partial y}(x, 0) = x$ für $x \neq 0$. Da auch $f(0, y) = 0$ für alle $y \in \mathbb{R}$ gilt $\frac{\partial f}{\partial y}(0, 0) = 0$, also

$$\frac{\partial f}{\partial y}(x, 0) = x \quad \text{für alle } x \in \mathbb{R}$$

und daher

$$\boxed{\frac{\partial^2 f}{\partial x \partial y}(0, 0) = 1}.$$

In diesem Fall kommt es also beim partiellen Differenzieren auf die Reihenfolge an.

Für genügend gute Funktionen kann das aber nicht passieren:

Satz 2.17 (Schwarz). *Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}$ zweimal stetig partiell differenzierbar. Dann gilt für alle $i, j = 1, \dots, n$ und $x \in U$*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Beweis. Es genügt den Fall $n = 2$ zu betrachten. Wir nennen die Variablen x, y .

Fixiere einen Punkt $(x_0, y_0) \in U$. Wir werden $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$ zeigen.

Da U offen ist, gibt es ein $\epsilon > 0$ so, dass

$$[x_0 - \epsilon, x_0 + \epsilon] \times [y_0 - \epsilon, y_0 + \epsilon] \subseteq U$$

Dann ist für $0 < h < \epsilon$ und $0 < k < \epsilon$ die Zahl

$$F(h, k) := f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)$$

definiert.

Für festes h und k betrachten wir die Funktionen

$$\phi(x) := f(x, y_0 + k) - f(x, y_0) \quad \text{und} \quad \psi(y) := f(x_0 + h, y) - f(x_0, y).$$

Da f partiell differenzierbar ist, sind ϕ und ψ in $(x_0 - \epsilon, x_0 + \epsilon)$ bzw. $(y_0 - \epsilon, y_0 + \epsilon)$ differenzierbar mit

$$\phi'(x) = \frac{\partial f}{\partial x}(x, y_0 + k) - \frac{\partial f}{\partial x}(x, y_0) \quad \text{und} \quad \psi'(y) := \frac{\partial f}{\partial y}(x_0 + h, y) - \frac{\partial f}{\partial y}(x_0, y).$$

Nach Konstruktion ist

$$F(h, k) = \phi(x_0 + h) - \phi(x_0) = \psi(y_0 + k) - \psi(y_0)$$

Nach dem Mittelwertsatz der Differenzialrechnung (Ana1, Korollar 5.19) gibt es ein $\xi_h \in (x_0, x_0 + h)$ mit

$$\phi(x_0 + h) - \phi(x_0) = \phi'(\xi_h) \cdot h,$$

also

$$F(h, k) = \phi'(\xi_h) \cdot h = \left(\frac{\partial f}{\partial x}(\xi_h, y_0 + k) - \frac{\partial f}{\partial x}(\xi_h, y_0) \right) \cdot h.$$

Nun wenden wir den Mittelwertsatz der Differenzialrechnung auf die Funktion $y \mapsto \frac{\partial f}{\partial x}(\xi_h, y)$ an. Dies ist möglich, da f zweimal partiell differenzierbar ist. Wir erhalten ein $\eta_k \in (y_0, y_0 + k)$ mit

$$\frac{\partial f}{\partial x}(\xi_h, y_0 + k) - \frac{\partial f}{\partial x}(\xi_h, y_0) = \frac{\partial^2 f}{\partial y \partial x}(\xi_h, \eta_k) \cdot k$$

also insgesamt

$$F(h, k) = \frac{\partial^2 f}{\partial y \partial x}(\xi_h, \eta_k) \cdot h k.$$

Die gleiche Argumentation auf ψ angewendet liefert $\tilde{\xi}_h \in (x_0, x_0 + h)$ und $\tilde{\eta}_k \in (y_0, y_0 + k)$ so, dass

$$F(h, k) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{\xi}_h, \tilde{\eta}_k) \cdot h k.$$

Also gilt

$$\frac{\partial^2 f}{\partial y \partial x}(\xi_h, \eta_k) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{\xi}_h, \tilde{\eta}_k).$$

Nun benutzen wir, dass nach Voraussetzung $\frac{\partial^2 f}{\partial x \partial y}$ stetig ist: Wählen wir eine Folge von h 's und k 's, die gegen 0 konvergiert, so konvergieren die entsprechenden Folgen der $\xi_h, \tilde{\xi}_h$ gegen x_0 , die der $\eta_k, \tilde{\eta}_k$ gegen y_0 . Also $\frac{\partial^2 f}{\partial y \partial x}(\xi_h, \eta_k) \rightarrow \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$ und $\frac{\partial^2 f}{\partial x \partial y}(\tilde{\xi}_h, \tilde{\eta}_k) \rightarrow \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$. Damit folgt $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$, wie gewünscht. $\square$

Totale Differenzierbarkeit und Jacobimatrix

Wir haben gesehen, dass partiell differenzierbare Funktionen nur bedingt gute Eigenschaften haben. So brauchen sie nicht einmal stetig zu sein. Wir werden nun einen stärkeren Differenzierbarkeitsbegriff einführen, welcher die Stetigkeit und partielle Differenzierbarkeit impliziert. Etwas allgemeiner als vorher betrachten wir in diesem Abschnitt Abbildungen $U \rightarrow \mathbb{R}^m$, wobei $U \subseteq \mathbb{R}^n$ offen ist.

Erinnerung 2.18. (Siehe Ana1, Satz 5.7.) Sei $I \subseteq \mathbb{R}$ ein offenes Intervall, und sei $f: I \rightarrow \mathbb{R}$ eine Funktion. Sei $x \in I$. Dann ist f genau dann in x differenzierbar, wenn es eine Zahl $a \in \mathbb{R}$ gibt so, dass für den Fehlerterm $\phi(\xi)$ in

$$f(x + \xi) = f(x) + a \cdot \xi + \phi(\xi)$$

gilt

$$\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\xi} = 0.$$

In diesem Fall ist $a = f'(x)$. Also ist f in x differenzierbar, wenn sich f bei x , d.h. die Funktion $\xi \mapsto f(x + \xi) - f(x)$, "gut" durch die lineare Funktion $a \mapsto a\xi$ approximieren lässt. Die Bedingung $\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\xi} = 0$ bedeutet, dass $\phi(\xi)$ für $\xi \rightarrow 0$ schneller gegen 0 geht, als eine lineare Funktion. Anschaulich bedeutet dies, dass der Graph von f im Punkt $(x, f(x))$ eine Tangente besitzt.

Diese Eigenschaft werden wir als Definition der "totalen Differenzierbarkeit" benutzen.

Erinnerung 2.19. Eine Abbildung $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ heißt *linear*, wenn sie

$$\begin{aligned}\Phi(x + y) &= \Phi(x) + \Phi(y) && \text{für alle } x, y \in \mathbb{R}^n, \\ \Phi(\lambda x) &= \lambda \Phi(x) && \text{für alle } x \in \mathbb{R}^n, \lambda \in \mathbb{R}\end{aligned}$$

erfüllt (LA, Def. 2.3). Dies ist genau dann der Fall, wenn es eine $m \times n$ -Matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

gibt mit

$$\Phi(x) = Ax = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

für alle $x \in \mathbb{R}^n$ (LA, Satz 2.8).

Definition 2.20. Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}^m$ eine Abbildung. Dann heißt f in $x \in U$ *total differenzierbar* (oder einfach nur *differenzierbar*), wenn es eine Matrix $A \in \mathbb{R}^{m \times n}$ gibt so, dass für den Fehlerterm $\phi(\xi)$ in

$$f(x + \xi) = f(x) + A \cdot \xi + \phi(\xi)$$

gilt

$$\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\|\xi\|} = 0.$$

Die Abbildung ϕ ist hier in einer Umgebung von $0 \in \mathbb{R}^n$ definiert und hat Werte in $\mathbb{R}^m$. Ist $n = 2$ und $m = 1$, also f eine Funktion von zwei Variablen, so ist ihr Graph eine Fläche im $\mathbb{R}^3$. Differenzierbarkeit im Punkt (x, y) bedeutet anschaulich, dass der Graph von f im Punkt $(x, y, f(x, y))$ eine Tangentialebene besitzt.

Beispiel 2.21. (a) Sei $A \in \mathbb{R}^{m \times n}$ eine $m \times n$ -Matrix und $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ die lineare Abbildung $x \mapsto Ax$. Dann ist f in jedem $x \in \mathbb{R}^n$ differenzierbar. Denn $f(x + \xi) = A \cdot (x + \xi) = A \cdot x + A \cdot \xi = f(x) + A \cdot \xi + 0$, also $\phi(\xi) = 0$ für alle ξ .

- (b) Erinnerung: Ist $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ eine $m \times n$ -Matrix, dann ist die *transponierte Matrix* (LA, Def. 2.11) die $n \times m$ -Matrix

$$A^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}.$$

Sei nun $B = (b_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$ eine $n \times n$ -Matrix, und sei $f: \mathbb{R}^n \rightarrow \mathbb{R}$ die Funktion

$$f(x) = \langle x, Bx \rangle = x^T Bx = (x_1, \dots, x_n) \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i,j=1}^n b_{ij} x_i x_j.$$

Dann gilt

$$\begin{aligned} f(x + \xi) &= \langle x + \xi, B(x + \xi) \rangle = \langle x + \xi, Bx + B\xi \rangle \\ &= \langle x, Bx \rangle + \langle x, B\xi \rangle + \langle \xi, Bx \rangle + \langle \xi, B\xi \rangle \\ &= f(x) + \langle x, B\xi \rangle + \langle Bx, \xi \rangle + \langle \xi, B\xi \rangle. \end{aligned}$$

Es gilt

$$\langle x, B\xi \rangle + \langle Bx, \xi \rangle = x^T B\xi + (Bx)^T \xi = x^T B\xi + x^T B^T \xi = (x^T (B + B^T)) \cdot \xi.$$

Mit der Matrix $A := x^T (B + B^T)$ gilt also

$$f(x + \xi) = f(x) + A\xi + \phi(\xi)$$

mit $\phi(\xi) = \langle \xi, B\xi \rangle$. Wegen der Cauchy-Schwarzschen Ungleichung gilt

$$|\phi(\xi)| = |\langle \xi, B\xi \rangle| \leq \|\xi\| \cdot \|B\xi\|$$

und daher

$$\frac{|\phi(\xi)|}{\|\xi\|} \leq \|B\xi\|$$

Da die Multiplikation mit einer Matrix stetig ist, gilt $\lim_{\xi \rightarrow 0} \|B\xi\| = 0$ und daher auch $\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\|\xi\|} = 0$. Insgesamt sehen wir, dass f in allen $x \in \mathbb{R}^n$ differenzierbar ist. [Im Fall $n = 1$ ist B einfach eine reelle Zahl b , und f ist die Funktion $f(x) = bx^2$. Die Formel für die Matrix A vereinfacht sich dann zu $A = 2bx$, der üblichen Formel für die Ableitung $f'(x)$.]

(c) Wir betrachten nun die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

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

Man kann zeigen, dass f in 0 stetig ist. Wegen $f(x, 0) = 0$ für alle $x \in \mathbb{R}$ und $f(0, y) = 0$ für alle $y \in \mathbb{R}$ ist f in 0 auch partiell differenzierbar mit $\frac{\partial f}{\partial x}(0) = 0$, $\frac{\partial f}{\partial y}(0) = 0$.

Wir wollen anschaulich argumentieren, dass der Graph $\Gamma_f \subseteq \mathbb{R}^3$ von f in 0 keine Tangentialebene besitzt. Für jedes $\lambda \in \mathbb{R}$ gilt $f(\lambda x, \lambda y) = \lambda \cdot f(x, y)$. D.h. für beliebige $(x, y) \in \mathbb{R}^2 \setminus \{0\}$ enthält der Graph Γ_f die Gerade durch 0 mit Richtungsvektor $(x, y, f(x, y))$. Hätte Γ_f eine Tangentialebene, so müsste sie alle diese Geraden enthalten. Betrachten wir speziell $(x, y) = (0, 0), (1, 0), (0, 1), (1, 1)$, so erhalten wir, dass die Punkte $(0, 0, 0), (1, 0, 0), (0, 1, 0)$ und $(1, 1, \frac{1}{2})$ in einer Ebene liegen müssten. Dies ist aber nicht der Fall! – Die ersten drei liegen in der x - y -Ebene, der vierte nicht.

Tatsächlich kann man zeigen, dass die Funktion f nicht differenzierbar ist.

Notation 2.22. Ist ϕ eine in einer Umgebung von $0 \in \mathbb{R}^n$ definierte Funktion, so bedeutet

$$\phi(\xi) = o(\xi),$$

dass $\phi(0) = 0$ und $\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\|\xi\|} = 0$ gilt. Man nennt o *Landau-Symbol*.

Z.B. ist eine Funktion $f: U \rightarrow \mathbb{R}^m$ genau dann in x differenzierbar, falls es eine $m \times n$ -Matrix A gibt so, dass

$$f(x + \xi) = f(x) + A\xi + o(\xi).$$

Da auch $\lim_{\xi \rightarrow 0} \|\xi\| = 0$ gilt für $\phi(\xi) = o(\xi)$ insbesondere auch $\lim_{\xi \rightarrow 0} \phi(\xi) = 0$.

Bemerkung 2.23. Sei $U \subseteq \mathbb{R}^n$ offen und sei $f = (f_1, \dots, f_m): U \rightarrow \mathbb{R}^m$ eine Abbildung. Die Abbildung f ist genau dann in $x \in U$ differenzierbar, wenn jede Komponentenfunktion $f_i: U \rightarrow \mathbb{R}$ in x differenzierbar ist.

Insbesondere stimmt der neue Differenzierbarkeitsbegriff für Kurven mit dem aus Definition 2.3 überein.

Beweis. Die Gleichung

$$f(x + \xi) = f(x) + A\xi + \phi(\xi)$$

ist äquivalent zu dem System von Gleichungen

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

für $i = 1, \dots, m$. Es gilt $\phi(\xi) = o(\xi) \Leftrightarrow \phi_i(\xi) = o(\xi)$ für alle $i = 1, \dots, m$ (Satz 1.10). Ist f in x differenzierbar mit der Matrix A , so ist daher f_i in x differenzierbar mit der Matrix $(a_{i1}, \dots, a_{in})$ die nur aus der i -ten Zeile von A besteht. Sind umgekehrt alle f_i in x differenzierbar mit den $1 \times n$ -Matrizen $(a_{i1}, \dots, a_{in})$, $i = 1, \dots, m$, so ist f differenzierbar mit der Matrix A , deren Zeilen gerade die gegebenen sind. $\square$

Satz 2.24. Sei $U \subseteq \mathbb{R}^n$ offen. Ist die Abbildung $f: U \rightarrow \mathbb{R}^m$ in $x \in U$ differenzierbar, so ist sie dort auch stetig.

Beweis. Schreibe $f(x+\xi) = f(x) + A\xi + \phi(\xi)$ mit einer Matrix A und $\phi(\xi) = o(\xi)$. Wegen $\lim_{\xi \rightarrow 0} A\xi = 0$ und $\lim_{\xi \rightarrow 0} \phi(\xi) = 0$ gilt dann $\lim_{\xi \rightarrow 0} f(x+\xi) = f(x)$. Konkret bedeutet dies, dass für jede Folge $(\xi_k)_{k \in \mathbb{N}}$ mit $\lim_{k \rightarrow \infty} \xi_k = 0$ auch $\lim_{k \rightarrow \infty} f(x+\xi_k) = f(x)$ gilt. Ist $(x_k)_{k \in \mathbb{N}}$ eine Folge mit $\lim_{k \rightarrow \infty} x_k = x$, so wende obiges auf die Folge $\xi_k = x_k - x$, $k \in \mathbb{N}$, an und erhalte $\lim_{k \rightarrow \infty} f(x_k) = f(x)$. Also ist f in x stetig. $\square$

Wenn eine Abbildung f in x differenzierbar ist, kann man die Matrix A leicht berechnen:

Satz 2.25. Sei $U \subseteq \mathbb{R}^n$ offen, und $f: U \rightarrow \mathbb{R}^m$ in $x \in U$ differenzierbar. Dann ist f in x partiell differenzierbar. Ist $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ eine $m \times n$ -Matrix mit

$$(*) \quad f(x+\xi) = f(x) + A\xi + o(\xi),$$

so gilt

$$a_{ij} = \frac{\partial f_i}{\partial x_j}(x).$$

Insbesondere ist die Matrix A in $(*)$ eindeutig bestimmt. Man nennt A das *Differenzial* oder die *Jacobimatrix* von f an der Stelle x . Schreibweise: $Df(x) = (\frac{\partial f_i}{\partial x_j}(x))_{ij}$.

Die Umkehrung des Satzes ist falsch. Z.B. ist die in Beispiel 2.11 betrachtete Funktion g partiell differenzierbar, aber in 0 nicht stetig. Insbesondere kann sie wegen Satz 2.24 in 0 nicht differenzierbar sein.

Beweis von Satz 2.25. Für $i = 1, \dots, m$ gilt

$$f_i(x+\xi) = f_i(x) + \sum_{j=1}^n a_{ij}\xi_j + \phi_i(\xi)$$

mit $\phi_i(\xi) = o(\xi)$. Für $\xi = he_j$ erhalten wir

$$f_i(x+he_j) = f_i(x) + a_{ij}h + \phi_i(he_j).$$

Daraus folgt

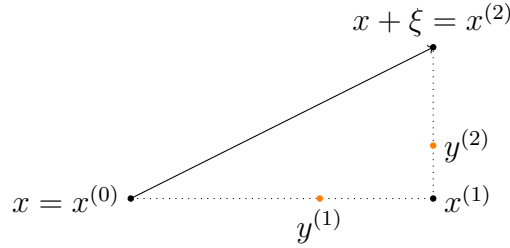
$$\lim_{h \rightarrow 0} \frac{f_i(x+he_j) - f_i(x)}{h} = a_{ij} + \lim_{h \rightarrow 0} \frac{\phi_i(he_j)}{h} = a_{ij},$$

d.h. die j -te partielle Ableitung von f_i an der Stelle x existiert und hat den Wert $\frac{\partial f_i}{\partial x_j}(x) = a_{ij}$. $\square$

Ist eine Funktion sogar stetig partiell differenzierbar, so ist sie auch differenzierbar:

Satz 2.26. Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}$ partiell differenzierbar. Sei $x \in U$. Sind alle partiellen Ableitungen $D_j f$, $j = 1, \dots, m$, in x stetig, so ist f in x differenzierbar.

Beweis. Für genügend kleines ξ schreiben wir $f(x + \xi) - f(x)$ mit Hilfe der partiellen Ableitungen um. Setze dazu $x^{(0)} := x$, $x^{(1)} := x^{(0)} + \xi_1 e_1$, $x^{(2)} := x^{(1)} + \xi_2 e_2$, usw., $x^{(n)} = x^{(n-1)} + \xi_n e_n = x + \xi$.



Da f partiell differenzierbar ist, gibt es nach dem Mittelwertsatz der Differenzialrechnung jeweils ein $\eta_j \in (0, \xi_j)$ (welches von ξ abhängt) so, dass

$$f(x^{(j)}) - f(x^{(j-1)}) = D_j f(\underbrace{x^{(j-1)} + \eta_j e_j}_{=: y^{(j)}}) \cdot \xi_j.$$

Damit erhalten wir

$$\begin{aligned} f(x + \xi) - f(x) &= (f(x^{(n)}) - f(x^{(n-1)})) + (f(x^{(n-1)}) - f(x^{(n-2)})) \cdots + (f(x^{(1)}) - f(x^{(0)})) \\ &= \sum_{j=1}^n D_j f(y^{(j)}) \cdot \xi_j \\ &= \sum_{j=1}^n D_j f(x) \cdot \xi_j + \phi(\xi) \end{aligned}$$

mit

$$\phi(\xi) = \sum_{j=1}^n (D_j f(y^{(j)}) - D_j f(x)) \cdot \xi_j.$$

Lassen wir ξ gegen 0 gehen, so konvergieren die entsprechenden $y^{(j)}$ gegen x . Da die partiellen Ableitungen $D_j f$ nach Voraussetzung in x stetig sind, gilt also $\lim_{\xi \rightarrow 0} (D_j f(y^{(j)}) - D_j f(x)) = 0$. Da $|\xi_j| \leq \|\xi\|$ für $j = 1, \dots, m$, impliziert dies

$$\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\|\xi\|} = 0,$$

wie gewünscht. □

Wir haben also nun die Implikationen

$$\text{stetig partiell differenzierbar} \implies (\text{total}) \text{ differenzierbar} \implies \text{partiell differenzierbar}$$

gezeigt. Statt *stetig partiell differenzierbar* sagen wir deshalb auch einfach *stetig differenzierbar*.

Die Kettenregel

Von fundamentaler Bedeutung für sowohl theoretische als auch praktische Anwendungen ist:

Satz 2.27 (Kettenregel). *Seien $U \subseteq \mathbb{R}^n$ und $V \subseteq \mathbb{R}^m$ offen, $f: U \rightarrow \mathbb{R}^m$ und $g: V \rightarrow \mathbb{R}^k$ Abbildungen mit $f(U) \subseteq V$. Ist f in $x \in U$ differenzierbar und g in $f(x) \in V$ differenzierbar, so ist auch die Verkettung $g \circ f$ in x differenzierbar. Ihre Jacobimatrix ist das Matrizenprodukt*

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

Bevor wir die Kettenregel beweisen, geben wir einige Anwendungen.

Korollar 2.28. *Seien $U \subseteq \mathbb{R}^n$ und $V \subseteq \mathbb{R}^m$ offen, $f = (f_1, \dots, f_m): U \rightarrow \mathbb{R}^m$ und $g: V \rightarrow \mathbb{R}$ differenzierbare Abbildungen mit $f(U) \subseteq V$. Bezeichne die Koordinaten in $\mathbb{R}^n$ mit $x_1, \dots, x_n$, die in $\mathbb{R}^m$ mit $y_1, \dots, y_m$. Dann gilt für $i = 1, \dots, n$*

$$\frac{\partial(g \circ f)}{\partial x_i}(x) = \sum_{j=1}^m \frac{\partial g}{\partial y_j}(f(x)) \cdot \frac{\partial f_j}{\partial x_i}(x).$$

Beweis. Die Jacobimatrix von $g \circ f$ an der Stelle x ist der Zeilenvektor

$$D(g \circ f)(x) = \left(\frac{\partial(g \circ f)}{\partial x_1}(x), \dots, \frac{\partial(g \circ f)}{\partial x_n}(x) \right) \in \mathbb{R}^{1 \times n}.$$

Die Jacobimatrix von g an der Stelle $y = f(x)$ ist der Zeilenvektor

$$Dg(f(x)) = \left(\frac{\partial g}{\partial y_1}(f(x)), \dots, \frac{\partial g}{\partial y_m}(f(x)) \right) \in \mathbb{R}^{1 \times m}.$$

Die Jacobimatrix von f an der Stelle x ist die $m \times n$ -Matrix

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

Nach der Kettenregel ist der i -te Eintrag von $D(g \circ f)(x)$ gerade der i -te Eintrag des Produkts $Dg(f(x)) \cdot Df(x)$. Ausmultiplizieren liefert die Behauptung. $\square$

Sei $U \subseteq \mathbb{R}^n$ offen. Ist $f: U \rightarrow \mathbb{R}$ eine differenzierbare Funktion, so ist ihre Jacobimatrix in einem Punkt $x \in U$ der *Zeilenvektor*

$$\left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right),$$

also der Gradient $\nabla f(x)$ ³. Ist $I \subseteq \mathbb{R}$ ein Intervall und $\alpha = (\alpha_1, \dots, \alpha_n): I \rightarrow \mathbb{R}^n$ eine differenzierbare Kurve, so ist ihre Jacobimatrix in einem Punkt $t \in I$ der *Spaltenvektor*

$$\begin{pmatrix} \alpha'_1(t) \\ \vdots \\ \alpha'_n(t) \end{pmatrix}.$$

Aus der Kettenregel erhalten wir:

Korollar 2.29. *Ist die Bahn von α ganz in U enthalten, so gilt*

$$(f \circ \alpha)'(t) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\alpha(t)) \cdot \alpha'_j(t) = \langle \nabla f(\alpha(t)), \alpha'(t) \rangle.$$

Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}^m$ differenzierbar. Für $x \in U$ ist $Df(x)$ eine $m \times n$ -Matrix. Ist $v \in \mathbb{R}^n$ ein beliebiger Vektor, so ist das Produkt $Df(x) \cdot v$ ein Vektor in $\mathbb{R}^m$. Wie kann man diesen interpretieren? – Dies ist die Ableitung von f an der Stelle x in Richtung v .

Wir beschränken uns wieder auf $\mathbb{R}$ -wertige Funktionen.

Definition 2.30. Sei $U \subseteq \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}$ eine Funktion, $x \in U$ und $v \neq 0$ ein Vektor in $\mathbb{R}^n$. Existiert der Grenzwert

$$D_v f(x) := \lim_{h \rightarrow 0} \frac{f(x + h \cdot v) - f(x)}{h},$$

so nennt man ihn die *Richtungsableitung* von f an der Stelle x in Richtung v .

Betrachte die Kurve $\alpha: \mathbb{R} \rightarrow \mathbb{R}^n$, $\alpha(t) = x + tv$, (eine parametrisierte Gerade durch x in Richtung v). Dann ist $D_v f(x)$ die gewöhnliche Ableitung $(f \circ \alpha)'(0)$.

Satz 2.31. *Sei $U \subseteq \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}$ differenzierbar und $x \in U$. Dann existieren in x alle Richtungsableitungen von f und für $v \in \mathbb{R}^n \setminus \{0\}$ gilt*

$$D_v f(x) = \langle \nabla f(x), v \rangle.$$

Insbesondere ist $v \mapsto D_v f(x)$ linear: $D_{\lambda v + \mu w} f(x) = \lambda \cdot D_v f(x) + \mu \cdot D_w f(x)$.

Beweis. Die Kurve α von oben ist differenzierbar mit $\alpha'(t) = v$ für alle $t \in \mathbb{R}$. Für kleines t gilt $\alpha(t) \in U$. Nach der Kettenregel 2.27 ist dann auch $f \circ \alpha$ in 0 differenzierbar und nach Korollar 2.29 gilt

$$D_v f(x) = (f \circ \alpha)'(0) = \langle \nabla f(\alpha(0)), \alpha'(0) \rangle = \langle \nabla f(x), v \rangle.$$

□

Damit können wir auch die Aussagen über den Gradienten präzisieren:

³Wir hatten uns beim Gradienten nicht festgelegt, ob wir ihn als Zeilen- oder Spaltenvektor auffassen.

Satz 2.32. Sei $U \subseteq \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}$ eine differenzierbare Funktion und $x \in U$ ein Punkt mit $\nabla f(x) \neq 0$.

- (a) $\nabla f(x)$ zeigt in die Richtung, in der f am schnellsten wächst.
- (b) Setze $f(x) =: c$. Der Gradient $\nabla f(x)$ steht auf der Niveaumenge $f^{-1}(c)$ senkrecht. Genauer gilt: Ist $\alpha: (-\epsilon, \epsilon) \rightarrow \mathbb{R}$ eine differenzierbare Kurve mit $\alpha(0) = x$, die ganz in $f^{-1}(c)$ verläuft, so gilt $\langle \nabla f(x), \alpha'(0) \rangle = 0$.

Da α in (b) ganz in $f^{-1}(c)$ verläuft, kann man $\alpha'(0)$ als Tangentialvektor and die Niveaumenge $f^{-1}(c)$ verstehen.

Beweis. (a) Wegen $D_{\lambda v}f(x) = \lambda D_v f(x)$ können wir uns auf Vektoren v mit $\|v\| = 1$ beschränken. Zu zeigen ist, dass $D_v f(x) \in \mathbb{R}$ maximal wird, wenn $v = \frac{\nabla f(x)}{\|\nabla f(x)\|}$ ist. nach Satz 2.31 gilt

$$D_v f(x) = \langle \nabla f(x), v \rangle = \|\nabla f(x)\| \cdot \|v\| \cdot \cos(\theta)$$

wobei $\theta \in [0, \pi]$ der Winkel zwischen v und $\nabla f(x)$ ist. Dieser Ausdruck wird maximal, wenn $\theta = 0$, d.h. $v = \frac{\nabla f(x)}{\|\nabla f(x)\|}$, ist.

- (b) Da α ganz in $f^{-1}(c)$ verläuft, ist $f \circ \alpha$ konstant mit dem Wert c . Also gilt

$$0 = (f \circ \alpha)'(0) = \langle \nabla f(0), \alpha'(0) \rangle.$$

□

Mit Hilfe der Kettenregel und Kurven kann man noch weitere Aussagen der Analysis einer Variablen auf die mehrerer Variablen übertragen, zum Beispiel:

Satz 2.33 (Mittelwertsatz). Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}$ differenzierbar. Seien $a, b \in U$ derart, dass die Verbindungsstrecke $\{a + t \cdot (b - a) : t \in [0, 1]\}$ ganz in U enthalten ist. Dann gibt es ein $t \in (0, 1)$ mit

$$f(b) - f(a) = \langle \nabla f(a + t(b - a)), b - a \rangle.$$

Beweis. Setze $\alpha(t) = a + t(b - a)$ und $h(t) = f \circ \alpha(t)$. Dies ist eine auf $[0, 1]$ differenzierbare Funktion. Nach dem Mittelwertsatz aus Analysis 1 (Ana1, Korollar 5.19) gibt es ein $t \in (0, 1)$ mit

$$h(1) - h(0) = h'(t) \cdot (1 - 0) = h'(t).$$

Wegen $h(1) = f(b)$, $h(0) = f(a)$ und

$$h'(t) = \langle \nabla f(\alpha(t)), \alpha'(t) \rangle = \langle \nabla f(a + t(b - a)), b - a \rangle$$

liefert dies die Behauptung. □

Satz 2.34. Sei $U \subseteq \mathbb{R}^n$ offen und wegzusammenhängend, d.h. für je zwei Punkte $a, b \in U$ gebe es eine differenzierbare Kurve $\alpha: [0, 1] \rightarrow \mathbb{R}^n$ mit $\alpha(0) = a$, $\alpha(1) = b$, deren Bahn ganz in U liegt. Ist $f: U \rightarrow \mathbb{R}$ differenzierbar, und gilt $\nabla f(x) = 0$ für alle $x \in U$, dann ist f konstant.

Beweis. Seien $a, b \in U$ beliebig. Zu zeigen: $f(a) = f(b)$. Sei also $\alpha: [0, 1] \rightarrow \mathbb{R}^n$ wie in der Aussage. Dann gilt für $t \in [0, 1]$

$$(f \circ \alpha)'(t) = \langle \nabla f(\alpha(t)), \alpha'(t) \rangle = 0.$$

Daher ist die Funktion $f \circ \alpha$ konstant und insbesondere $f(\alpha(0)) = f(\alpha(1))$. $\square$

Für den Beweis der Kettenregel benötigen wir noch einen kurzen Einschub über die Norm von Matrizen:

Lemma 2.35. Sei $A \in \mathbb{R}^{m \times n}$ eine $m \times n$ -Matrix. Dann ist

$$\|A\| := \sup_{x \in \mathbb{R}^n: \|x\|=1} \|Ax\| < \infty$$

und für alle $x \in \mathbb{R}^n$ gilt $\|Ax\| \leq \|A\| \cdot \|x\|$.

Man kann zeigen, dass dadurch eine Norm auf dem Vektorraum $\mathbb{R}^{m \times n}$ der $m \times n$ -Matrizen definiert wird.

Beweis des Lemmas. Schreibe $S := \{x \in \mathbb{R}^n : \|x\| = 1\}$. Als Urbild der abgeschlossenen Menge $\{1\} \subseteq \mathbb{R}$ unter der stetigen Abbildung $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}$ ist $S \subseteq \mathbb{R}^n$ abgeschlossen. Da S auch beschränkt ist, ist es nach dem Satz von Heine-Borel 1.41 kompakt. Also ist die stetige Funktion $x \mapsto \|Ax\| \in \mathbb{R}$ auf S beschränkt (Satz 1.35). Daher ist $\|A\| < \infty$.

Ist $x \in \mathbb{R}^n$, $x \neq 0$, beliebig, so ist $\frac{x}{\|x\|} \in S$ und daher

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

Beweis der Kettenregel 2.27. Setze $A := Df(x)$ und $B := Dg(f(x))$. Da f in x und g in $f(x)$ differenzierbar sind, gelten

$$\begin{aligned} f(x + \xi) &= f(x) + A\xi + \phi(\xi) \\ g(f(x) + \eta) &= g(f(x)) + B\eta + \psi(\eta) \end{aligned}$$

mit $\phi(\xi) = o(\xi)$ und $\psi(\eta) = o(\eta)$. Insbesondere ergibt sich für $\eta = A\xi + \phi(\xi)$

$$\begin{aligned} (g \circ f)(x + \xi) &= g(f(x) + A\xi + \phi(\xi)) \\ &= g(f(x) + BA\xi + B\phi(\xi) + \psi(A\xi + \phi(\xi))) \\ &= (g \circ f)(x) + (BA)\xi + \chi(\xi) \end{aligned}$$

mit dem Fehlerterm

$$\chi(\xi) := B\phi(\xi) + \psi(A\xi + \phi(\xi)).$$

Es ist nun zu zeigen, dass $\chi(\xi) = o(\xi)$ gilt.

Wegen $\phi(\xi) = o(\xi)$ gilt $\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\|\xi\|} = 0$. Da die Multiplikation mit der Matrix B stetig ist, gilt auch $\lim_{\xi \rightarrow 0} \frac{B\phi(\xi)}{\|\xi\|} = 0$.

Außerdem gilt wegen $\lim_{\xi \rightarrow 0} \frac{\phi(\xi)}{\|\xi\|} = 0$ auch $\frac{\|\phi(\xi)\|}{\|\xi\|} \leq 1$ für ξ klein genug, also $\|\phi(\xi)\| \leq \|\xi\|$ für ξ klein genug. Wegen $\psi(\eta) = o(\eta)$ können wir ψ schreiben als $\psi(\eta) = \|\eta\| \cdot \tilde{\psi}(\eta)$, wobei $\lim_{\eta \rightarrow 0} \tilde{\psi}(\eta) = 0$ gilt. Also gilt für hinreichend kleines ξ

$$\|\psi(A\xi + \phi(\xi))\| = \|A\xi + \phi(\xi)\| \cdot \|\tilde{\psi}(A\xi + \phi(\xi))\| \leq (\|A\| \cdot \|\xi\| + \|\xi\|) \cdot \|\tilde{\psi}(A\xi + \phi(\xi))\|$$

und daher

$$\frac{\|\psi(A\xi + \phi(\xi))\|}{\|\xi\|} \leq (\|A\| + 1) \cdot \|\tilde{\psi}(A\xi + \phi(\xi))\|.$$

Für $\xi \rightarrow 0$ gilt $A\xi + \phi(\xi) \rightarrow 0$ und damit auch $\tilde{\psi}(A\xi + \phi(\xi)) \rightarrow 0$.

Insgesamt ergibt sich $\chi(\xi) = o(\xi)$, wie gewünscht. $\square$

Maxima und Minima

Wir wollen nun Kriterien zur Bestimmung von Maxima und Minima von differenzierbaren Funktionen $f: U \rightarrow \mathbb{R}$, $U \subseteq \mathbb{R}^n$ offen, herleiten.

Definition 2.36. Seien $U \subseteq \mathbb{R}^n$ eine Teilmenge, $f: U \rightarrow \mathbb{R}$ eine Funktion und $x \in U$. Wir sagen, f habe in x ein *(lokales) Maximum* (bzw. ein *(lokales) Minimum*), falls es eine Umgebung V von x in $\mathbb{R}^n$ gibt so, dass

$$f(y) \leq f(x) \text{ (bzw. } f(y) \geq f(x)) \text{ für alle } y \in V \cap U.$$

Die Funktion f hat in a ein *globales Maximum* (bzw. *Minimum*), wenn $f(y) \leq f(x)$ (bzw. $f(y) \geq f(x)$) für alle $y \in U$.

Wir sagen, f habe ein *(lokales) Extremum* in x , wenn es dort ein Maximum oder Minimum hat.

Satz 2.37. Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}$ differenzierbar. Hat f in $x \in U$ ein Extremum, so gilt $\nabla f(x) = 0$.

Beweis. Für $i = 1, \dots, n$ besitzt dann die differenzierbare Funktion einer Variablen $g(t) = f(x + t \cdot e_i)$ ein Extremum in $t = 0$. Also gilt

$$0 = g'(0) = \langle \nabla f(x), e_i \rangle = (\nabla f(x))_i \text{ für } i = 1, \dots, n,$$

d.h. $\nabla f(x) = 0$. $\square$

Um die Extrema einer differenzierbaren Funktion $f: U \rightarrow \mathbb{R}$ zu bestimmen, sucht man also zunächst die Punkte $x \in U$ mit $\nabla f(x) = 0$.

Definition 2.38. Ist $U \subseteq \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}$ differenzierbar und $x \in U$, so nennt man x *kritischen* oder *stationären Punkt* von f , falls $\nabla f(x) = 0$ gilt.

Man nennt x *Sattelpunkt*, falls x kritisch ist und es in jeder Umgebung von x Punkte y_1 und y_2 gibt mit

$$f(y_1) < f(x) < f(y_2).$$

Wir wollen nun hinreichende Kriterien für Extrema finden. In der eindimensionalen Analysis benutzt man hierfür die zweite Ableitung.

Angenommen, $f: U \rightarrow \mathbb{R}$ ist zweimal stetig partiell differenzierbar. Für einen beliebigen Richtungsvektor $\xi \neq 0$ sei α_ξ die Kurve $\alpha_\xi(t) = x + t\xi$. Wir wollen $(f \circ \alpha_\xi)''(0)$ berechnen. Nach der Kettenregel gilt

$$(f \circ \alpha_\xi)'(t) = \langle \nabla f(\alpha_\xi(t)), \alpha_\xi'(t) \rangle = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x + t\xi) \cdot \xi_i$$

und

$$\begin{aligned} (f \circ \alpha_\xi)''(t) &= \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \circ \alpha_\xi \right)'(t) \cdot \xi_i \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(\alpha_\xi(t)) \cdot \xi_j \right) \cdot \xi_i \\ &= \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(x + t\xi) \xi_i \xi_j \end{aligned}$$

also insbesondere

$$(f \circ \alpha_\xi)''(0) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \xi_i \xi_j.$$

Gilt $(f \circ \alpha_\xi)''(0) < 0$ für jeden Richtungsvektor $\xi \neq 0$, so liegt in x tatsächlich ein Maximum vor (s.u.). Es genügt aber nicht, für ξ nur die Basisvektoren e_i zu betrachten. Betrachte zum Beispiel

$$f(x, y) = -x^2 + 4xy - y^2.$$

Für $\xi = e_1$ erhält man $(f \circ \alpha_{e_1})''(0) = \frac{\partial^2 f}{\partial x \partial x}(0) = -2$, für $\xi = e_2$ genauso $(f \circ \alpha_{e_2})''(0) = \frac{\partial^2 f}{\partial y \partial y}(0) = -2$. Aber für $x = y$ gilt $f(x, x) = 2x^2$. Daher hat f in 0 einen Sattelpunkt.

Definition 2.39. Sei $U \subseteq \mathbb{R}^n$ offen, und $f: U \rightarrow \mathbb{R}$ zweimal stetig partiell differenzierbar. Für $x \in U$ heißt die $n \times n$ -Matrix

$$H_f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{i,j=1,\dots,n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

die *Hessematrix* von f in x .

Für die Funktion $f \circ \alpha_\xi$ von oben erhalten wir damit

$$(2.1) \quad (f \circ \alpha_\xi)''(0) = \xi^T H_f(x) \xi = \langle \xi, H_f(x) \xi \rangle.$$

Bemerkung 2.40. Die Hessematrix ist symmetrisch, da nach dem Satz von Schwarz für zweimal stetig partiell differenzierbare Funktionen $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ gilt.

Beispiel 2.41. (a) Für $f(x, y) = x^2 + y^2$ ist $H_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

(b) Für $f(x, y) = x^2 - y^2$ ist $H_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.

(c) Für $f(x, y) = -x^2 + 4xy - y^2$ ist $H_f(x, y) = \begin{pmatrix} -2 & 4 \\ 4 & -2 \end{pmatrix}$.

Um ein hinreichendes Kriterium für Extrema zu erhalten benötigen wir die Taylorformel zweiter Ordnung.

Satz 2.42 (Taylorformel zweiter Ordnung, eindimensionaler Fall). *Sei $I \subseteq \mathbb{R}$ ein Intervall, $g: I \rightarrow \mathbb{R}$ zweimal stetig differenzierbar und $t_0 \in I$. Für $t \in \mathbb{R}$ mit $t_0 + t \in I$ gibt es ein $\tau \in (0, t)$ so, dass*

$$g(t_0 + t) = g(t_0) + t \cdot g'(t_0) + \frac{t^2}{2} g''(t_0 + \tau)$$

Beweis. Eventuell später. □

Satz 2.43 (Taylorformel zweiter Ordnung). *Sei $U \subseteq \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}$ zweimal stetig differenzierbar und $x \in U$. Schreibe*

$$(*) \quad f(x + \xi) = f(x) + \langle \nabla f(x), \xi \rangle + \frac{1}{2} \xi^T H_f(x) \xi + R(\xi).$$

Dann gilt für den Fehlerterm $R(\xi)$

$$\lim_{\xi \rightarrow 0} \frac{R(\xi)}{\|\xi\|^2} = 0.$$

Mit den partiellen Ableitungen formuliert lautet Gleichung (*) konkret:

$$f(x + \xi) = f(x) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) \xi_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \xi_i \xi_j + R(\xi).$$

Beweis. Da $U \subseteq \mathbb{R}^n$ offen ist, gibt es ein $\epsilon > 0$ so, dass $B_\epsilon(x) \subseteq U$. Für $\xi \in B_\epsilon(0)$ verläuft dann die Kurve $\alpha: [-1, 1] \rightarrow \mathbb{R}^n$, $\alpha(t) = x + t\xi$, in U . Dann ist auch $g := f \circ \alpha: [-1, 1] \rightarrow \mathbb{R}$ zweimal stetig differenzierbar. Nach Satz 2.42 gibt es zu $t \in [-1, 1]$ ein $\tau \in (0, t)$ mit

$$g(t) = g(0) + t \cdot g'(0) + \frac{t^2}{2} g''(\tau).$$

Setze

$$\eta(t) := \frac{1}{2} (g''(\tau) - g''(0)).$$

Dann gilt

$$g(t) = g(0) + t \cdot g'(0) + \frac{t^2}{2} g''(0) + t^2 \eta(t).$$

Insbesondere erhalten wir für $t = 1$ (vgl. (2.1))

$$\begin{aligned} f(x + \xi) &= g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \eta(1) \\ &= f(x) + \langle \nabla f(x), \xi \rangle + \frac{1}{2}\xi^T H_f(x)\xi + R(\xi) \end{aligned}$$

wobei

$$\begin{aligned} R(\xi) &= \eta(1) = \frac{1}{2} (\xi^T H_f(x + \tau\xi)\xi - \xi^T H_f(x)\xi) \\ &= \frac{1}{2}\xi^T (H_f(x + \tau\xi) - H_f(x))\xi \\ &= \frac{1}{2} \sum_{i,j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x + \tau\xi) - \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right) \xi_i \xi_j \end{aligned}$$

mit geeignetem $\tau \in (0, 1)$, welches von ξ abhängt. Da $|\xi_i| \leq \|\xi\|$ für $i = 1, \dots, n$, gilt

$$\left| \frac{R(\xi)}{\|\xi\|^2} \right| \leq \frac{1}{2} \sum_{i,j=1}^n \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(x + \tau\xi) - \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right|.$$

Da nach Voraussetzung $\frac{\partial^2 f}{\partial x_i \partial x_j}$ stetig ist, geht die rechte Seite für $\xi \rightarrow 0$ gegen 0, also gilt $\lim_{\xi \rightarrow 0} \frac{R(\xi)}{\|\xi\|^2} = 0$, wie behauptet. $\square$

Ist x ein kritischer Punkt von f , so ergibt sich also

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

D.h. das lokale Verhalten von f bei x wird im Wesentlichen durch den Term $\frac{1}{2}\xi^T H_f(x)\xi$ beschrieben. Um diesen kümmern wir uns als nächstes.

Definition 2.44. Sei $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix, d.h. $a_{ij} = a_{ji}$ für $i, j = 1, \dots, n$. Die Matrix A heißt

- (a) *positiv definit*, falls $\xi^T A \xi = \sum_{i,j=1}^n a_{ij} \xi_i \xi_j > 0$ für alle $\xi \in \mathbb{R}^n \setminus \{0\}$,
- (b) *positiv semidefinit*, falls $\xi^T A \xi \geq 0$ für alle $\xi \in \mathbb{R}^n \setminus \{0\}$,
- (c) *negativ definit*, falls $\xi^T A \xi < 0$ für alle $\xi \in \mathbb{R}^n \setminus \{0\}$,
- (d) *negativ semidefinit*, falls $\xi^T A \xi \leq 0$, für alle $\xi \in \mathbb{R}^n \setminus \{0\}$, und
- (e) *indefinit*, falls es $\xi_1, \xi_2 \in \mathbb{R}^n$ gibt mit $\xi_1^T A \xi_1 < 0$ und $\xi_2^T A \xi_2 > 0$.

Wir werden z.B. zeigen: Ist x ein kritischer Punkt von f und ist die Hessematrix $H_f(x)$ positiv definit, so besitzt f in x ein Minimum. Um solche hinreichenden Kriterien anwenden zu können, benötigen wir aber handlichere Kriterien um die Definitheitseigenschaften symmetrischer Matrizen zu bestimmen.

Erinnerung 2.45. Sei $A \in \mathbb{R}^{n \times n}$ eine beliebige $n \times n$ -Matrix. Ein $\lambda \in \mathbb{R}$ heißt *Eigenwert* von A , wenn es einen Vektor $v \in \mathbb{R}^n \setminus \{0\}$ (ein *Eigenvektor* zum Eigenwert λ) gibt so, dass $Av = \lambda v$ gilt.

Die Eigenwerte von A sind genau die Nullstellen des charakteristischen Polynoms $\chi_A(t) = \det(A - t \cdot E_n)$.

Ist A symmetrisch, so besagt der Satz über die Hauptachsentransformation (LA Satz 6.27), dass $\mathbb{R}^n$ eine Orthonormalbasis (ONB) aus Eigenvektoren von A besitzt. D.h. es gibt Vektoren $v_1, \dots, v_n \in \mathbb{R}^n$ und $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ (nicht notwendig verschieden) mit $\langle v_i, v_j \rangle = \delta_{ij}$ und $Av_i = \lambda_i v_i$. Ist S die Matrix mit den Spalten $v_1, \dots, v_n$, dann ist S *orthogonal*, d.h. $S^T S = E_n$, also S invertierbar und $S^{-1} = S^T$ und

$$S^T A S = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

Satz 2.46. Sei $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix. Die Matrix A ist

- (a) *positiv definit* $\Leftrightarrow$ alle Eigenwerte von A sind > 0 ,
- (b) *positiv semidefinit* $\Leftrightarrow$ alle Eigenwerte von A sind ≥ 0 ,
- (c) *negativ definit* $\Leftrightarrow$ alle Eigenwerte von A sind < 0 ,
- (d) *negativ semidefinit* $\Leftrightarrow$ alle Eigenwerte von A sind ≤ 0 ,
- (e) *indefinit* $\Leftrightarrow$ es gibt einen positiven und einen negativen Eigenwert.

Beweis. Sei $v_1, \dots, v_n \in \mathbb{R}^n$ eine ONB aus Eigenvektoren von A mit zugehörigen Eigenwerten $\lambda_1, \dots, \lambda_n$. Ist $\xi \in \mathbb{R}^n$ beliebig, so gibt es eindeutige $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ mit

$$\xi = \alpha_1 v_1 + \dots + \alpha_n v_n$$

(konkret: $\alpha_i = \langle v_i, \xi \rangle$). Dann gilt

$$\begin{aligned} \xi^T A \xi &= \left(\sum_{i=1}^n \alpha_i v_i \right)^T A \left(\sum_{j=1}^n \alpha_j v_j \right) \\ &= \sum_{i,j=1}^n v_i^T A v_j \alpha_i \alpha_j = \sum_{i,j=1}^n v_i^T (\lambda_j v_j) \alpha_i \alpha_j \\ &= \sum_{i=1}^n \lambda_i \alpha_i^2. \end{aligned}$$

Daran liest man die Behauptung ab. □

Dieses Kriterium ist aber nur nützlich, wenn man die Eigenwerte der Matrix A kennt.

Beispiel 2.47. Die Matrix $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ hat den doppelten Eigenwert 2, sie ist also positiv definit. Die Matrix $\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$ hat die Eigenwerte 2, -2 , sie ist also indefinit.

Die Eigenwerte der Matrix $A = \begin{pmatrix} -2 & 4 \\ 4 & -2 \end{pmatrix}$ sind die Nullstellen von $(-2 - t)^2 - 16 = (t + 2)^2 - 16$ also 2 und -6 , d.h. A ist ebenfalls indefinit.

Allgemein gibt es noch folgendes Kriterium, das wir hier nicht beweisen wollen:

Satz 2.48. Sei $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ symmetrisch. A ist genau dann positiv definit, wenn

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix} > 0$$

für alle $k = 1, \dots, n$.

Da A genau dann negativ definit ist, wenn $-A$ positiv definit ist, und die Determinante multilinear ist, folgt auch: A ist genau dann negativ definit, wenn

$$(-1)^k \det \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix} > 0$$

für $k = 1, \dots, n$.

Wir behandeln nur den Spezialfall $n = 2$:

Satz 2.49. Sei $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ eine symmetrische 2×2 -Matrix. Dann gilt:

- (a) Ist $\det(A) < 0$, so ist A indefinit.
- (b) Ist $\det(A) > 0$ und $a > 0$, so ist A positiv definit.
- (c) Ist $\det(A) > 0$ und $a < 0$, so ist A negativ definit.

Beweis. Seien λ_1, λ_2 die beiden Eigenwerte von A . Das charakteristische Polynom ist

$$\chi_A(t) = \det \begin{pmatrix} a - t & b \\ b & d - t \end{pmatrix} = t^2 - (a + d)t + (ad - b^2).$$

Da λ_1, λ_2 die beiden Nullstellen von $\chi_A(t)$ sind, gilt

$$\lambda_1 \lambda_2 = \det(A) = ad - b^2 \quad \text{und} \quad \lambda_1 + \lambda_2 = a + d$$

Ist $\det(A) < 0$, so ist genau einer der beiden Eigenwerte > 0 , der andere < 0 .

Ist $\det(A) > 0$, so haben λ_1 und λ_2 das gleiche Vorzeichen. Außerdem ist auch $ad = \det(A) + b^2$ positiv, d.h. a und d haben ebenfalls das gleiche Vorzeichen. Diese haben dann auch das gleiche Vorzeichen wie λ_1 und λ_2 . Die Behauptung folgt nun aus Satz 2.46. $\square$

Satz 2.50. Sei $U \subseteq \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}$ zweimal stetig partiell differenzierbar. Sei $x \in U$ ein kritischer Punkt von f .

- (a) Ist $H_f(x)$ positiv definit, so besitzt f in x ein lokales Minimum.
- (b) Ist $H_f(x)$ negativ definit, so besitzt f in x ein lokales Maximum.
- (c) Ist $H_f(x)$ indefinit, so besitzt f in x einen Sattelpunkt

Beweis. Setze $q(\xi) := \xi^T H_f(x) \xi$. Da $\nabla f(x) = 0$, gilt dann nach der Taylorformel 2.43

$$f(x + \xi) = f(x) + q(\xi) + R(\xi)$$

mit $\lim_{\xi \rightarrow 0} \frac{R(\xi)}{\|\xi\|^2} = 0$.

(a) Da $H_f(x)$ positiv definit ist, gilt $q(\xi) > 0$ für $\xi \neq 0$. Außerdem ist q stetig und homogen vom Grad 2, d.h. $q(\lambda\xi) = \lambda^2 q(\xi)$ für $\xi \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$. Da $S^{n-1} := \{v \in \mathbb{R}^n : \|v\| = 1\} \subseteq \mathbb{R}^n$ abgeschlossen und beschränkt, also nach dem Satz von Heine-Borel 1.41 kompakt ist, nimmt q auf S^{n-1} ein Minimum an. D.h. es gibt ein $c > 0$ so, dass $q(\xi) \geq c$ für alle $\xi \in \mathbb{R}^n$ mit $\|\xi\| = 1$. Für beliebiges $\xi \neq 0$ erhalten wir dann

$$q(\xi) = \|\xi\|^2 q\left(\frac{\xi}{\|\xi\|}\right) \geq \|\xi\|^2 \cdot c.$$

Wegen $\lim_{\xi \rightarrow 0} \frac{R(\xi)}{\|\xi\|^2} = 0$ gibt es ein $\epsilon > 0$ so, dass

$$\left| \frac{R(\xi)}{\|\xi\|^2} \right| < \frac{c}{2} \text{ für } \xi \in B_\epsilon(0) \setminus \{0\},$$

d.h. $|R(\xi)| < \frac{c}{2} \|\xi\|^2$ für $\xi \neq 0$ mit $\|\xi\| < \epsilon$. Für diese ξ erhalten wir dann

$$f(x + \xi) > f(x) + c\|\xi\|^2 - \frac{c}{2}\|\xi\|^2 = f(x) + \frac{c}{2}\|\xi\|^2 > f(x).$$

Also ist x ein lokales Minimum. (x ist sogar ein *isoliertes* lokales Minimum, d.h. für alle $y \neq x$ einer Umgebung von x gilt $f(y) > f(x)$, nicht nur $f(y) \geq f(x)$.)

(b) Wende (a) auf $-f$ an.

(c) Ist $H_f(x)$ indefinit, so gibt es $\xi_1, \xi_2 \in \mathbb{R}^n$ mit $\xi_1^T H_f(x) \xi_1 < 0$ und $\xi_2^T H_f(x) \xi_2 > 0$. Setze $\alpha_{\xi_i}(t) = x + t\xi_i$. Dann gelten

$$(f \circ \alpha_{\xi_i})'(0) = \langle \nabla f(x), \xi_i \rangle = 0$$

und (siehe (2.1))

$$(f \circ \alpha_{\xi_1})''(0) = \xi_1^T H_f(x) \xi_1 < 0 \text{ und } (f \circ \alpha_{\xi_2})''(0) = \xi_2^T H_f(x) \xi_2 > 0.$$

Daher hat die Funktion $f \circ \alpha_{\xi_1}$ in $t = 0$ ein isoliertes Maximum, $f \circ \alpha_{\xi_2}$ ein isoliertes Minimum. Also nimmt f beliebig nahe an x sowohl Werte $< f(x)$ als auch $> f(x)$ an, und x ist ein Sattelpunkt. $\square$

Im semidefiniten Fall lässt sich keine Aussage machen:

Beispiel 2.51. Betrachte die Funktionen $f(x, y) = x^2$ und $g(x, y) = x^2 + y^3$. Für beide ist 0 ein kritischer Punkt und es gilt $H_f(0) = H_g(0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$. Diese Matrix ist positiv semidefinit. Die Funktion f hat in 0 und genauso in jedem Punkt der y -Achse ein lokales Minimum. Die Funktion g hat in 0 einen Sattelpunkt.

Beispiel 2.52. Sei $f(x, y) = x^4 - 2x^2 + 2x^2y^2 - y^2$.

Dann ist

$$\nabla f(x, y) = \begin{pmatrix} 4x^3 - 4x + 4xy^2 \\ 4x^2y - 2y \end{pmatrix} = \begin{pmatrix} 4x(x^2 - 1 + y^2) \\ 2y(2x^2 - 1) \end{pmatrix}.$$

Kritische Punkte: Angenommen, $\nabla f(x, y) = 0$.

Ist $x = 0$, dann auch $y = 0$.

Ist $y = 0$, dann gilt $x = 0$ oder $x^2 - 1 = 0$, d.h. $x = \pm 1$.

Ist $x \neq 0$ und $y \neq 0$, so gilt $x^2 - 1 + y^2 = 0$ und $2x^2 - 1 = 0$. Also $x^2 = y^2 = \frac{1}{2}$ und daher $x = \pm \frac{1}{\sqrt{2}}$, $y = \pm \frac{1}{\sqrt{2}}$.

Insgesamt erhalten wir also die kritischen Punkte $(0, 0)$, $(\pm 1, 0)$, $(\pm \frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}})$.

Die Hessematrix ergibt sich als

$$H_f(x, y) = \begin{pmatrix} 12x^2 - 4 + 4y^2 & 8xy \\ 8xy & 4x^2 - 2 \end{pmatrix},$$

also

$$H_f(0, 0) = \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix}, \quad \text{negativ definit}$$

$$H_f(\pm 1, 0) = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}, \quad \text{positiv definit}$$

$$H_f(\pm(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})) = \begin{pmatrix} 4 & 4 \\ 4 & 0 \end{pmatrix}, \quad \text{indefinit}$$

$$H_f(\pm(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})) = \begin{pmatrix} 4 & -4 \\ -4 & 0 \end{pmatrix} \quad \text{indefinit.}$$

Also: $(0, 0)$ ist Maximum, $(\pm 1, 0)$ sind Minima, die übrigen vier Punkte sind Sattelpunkte.

Der Satz über die Umkehrfunktion

Wir erinnern zunächst an den Satz über die Umkehrfunktion aus Analysis 1 (vgl. Ana1, Satz 5.13):

Satz 2.53. Sei $I \subseteq \mathbb{R}$ ein offenes Intervall, $f: I \rightarrow \mathbb{R}$ stetig differenzierbar und $x \in I$ mit $f'(x) \neq 0$. Dann gibt es ein $\epsilon > 0$ so, dass f auf $J := (x - \epsilon, x + \epsilon)$ streng monoton und daher injektiv ist. Auf $J' := f(J)$ existiert die Umkehrfunktion $f^{-1}: J' \rightarrow \mathbb{R}$. Sie ist differenzierbar und es gilt

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

Achtung: Die Umkehrfunktion f^{-1} darf nicht mit $\frac{1}{f}$ verwechselt werden!

Die Formel für die Ableitung der Umkehrfunktion erhält man aus der Kettenregel: Es ist $f^{-1} \circ f = \text{id}$, wobei id die Identitätsabbildung $x \mapsto x$ ist. Also folgt

$$1 = \text{id}'(x) = (f^{-1} \circ f)'(x) = (f^{-1})'(f(x)) \cdot f'(x).$$

Man kann also die Differenzierbarkeit der Umkehrfunktion zeigen und ihre Ableitung berechnen, ohne die Umkehrfunktion explizit zu kennen.

Beispiel 2.54. Die Sinusfunktion $\sin$ ist auf dem Intervall $(-\frac{\pi}{2}, \frac{\pi}{2})$ injektiv mit Bild $(-1, 1)$. Die Umkehrfunktion $\sin^{-1}: (-1, 1) \rightarrow \mathbb{R}$ heißt *Arcussinus* und wird mit $\arcsin$ bezeichnet. Nach dem Satz über die Umkehrfunktion ist $\arcsin$ differenzierbar und $\arcsin'(\sin(x)) = \frac{1}{\cos(x)}$. Setze $y = \sin(x)$. Für $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$ ist $\cos(x) > 0$ und daher $\cos(x) = \sqrt{1 - \sin(x)^2}$. Also erhalten wir

$$\arcsin'(y) = \frac{1}{\sqrt{1 - y^2}}.$$

Die Tangensfunktion $\tan$ ist auf $(-\frac{\pi}{2}, \frac{\pi}{2})$ definiert durch $\tan(x) = \frac{\sin(x)}{\cos(x)}$. Mit der Quotientenregel erhält man $\tan'(x) = \frac{1}{\cos(x)^2} > 0$. Also ist $\tan$ auf $(-\frac{\pi}{2}, \frac{\pi}{2})$ streng monoton steigend und insbesondere injektiv. Wegen $\lim_{x \rightarrow \pm \frac{\pi}{2}} \tan(x) = \pm \infty$ ist das Bild von $\tan$ ganz $\mathbb{R}$. Die Umkehrfunktion $\tan^{-1}$ nennt man *Arcustangens* $\arctan: \mathbb{R} \rightarrow \mathbb{R}$. Ihr Bild ist $(-\frac{\pi}{2}, \frac{\pi}{2})$. Für die Ableitung erhalten wir an der Stelle $y = \tan(x)$

$$\arctan'(y) = \frac{1}{\tan'(x)} = \cos(x)^2 = \frac{\cos(x)^2}{\cos(x)^2 + \sin(x)^2} = \frac{1}{1 + \tan(x)^2} = \frac{1}{1 + y^2}.$$

Sei nun $U \subset \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}^m$ differenzierbar. Falls f eine differenzierbare Umkehrabbildung $g = f^{-1}: V \rightarrow \mathbb{R}^n$ hat, also $V = f(U) \subseteq \mathbb{R}^m$ und $g \circ f = \text{id}_U$, $f \circ g = \text{id}_V$, dann liefert die Kettenregel

$$E_n = D(\text{id}_U)(x) = D(g \circ f)(x) = Dg(f(x)) \cdot Df(x) \text{ für alle } x \in U$$

und

$$E_m = D(\text{id}_V)(y) = D(f \circ g)(y) = Df(g(y)) \cdot Dg(y) \text{ für alle } y \in V.$$

Setzt man in der letzten Gleichung $y = f(x)$, so erhält man

$$E_m = Df(x) \cdot Dg(f(x)).$$

Deswegen ist die Matrix $Dg(f(x))$ invertierbar mit inverser Matrix $Df(x)$, d.h.

$$Df^{-1}(f(x)) = (Df(x))^{-1}.$$

Insbesondere gilt $m = n$.

Wie im eindimensionalen Fall sagt der Satz über die Umkehrfunktion, dass die Invertierbarkeit von $Df(x)$ auch hinreichend für die Existenz einer Umkehrabbildung ist, wenigstens lokal:

Satz 2.55 (Satz über die Umkehrfunktion). Sei $U \subseteq \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}^n$ stetig differenzierbar und $x_0 \in U$. Ist $Df(x_0)$ invertierbar, d.h. gilt $\det(Df(x_0)) \neq 0$, dann gibt es offene Umgebungen V von x_0 und $W = f(V)$ von $f(x_0)$ so, dass $f: V \rightarrow W$ bijektiv ist. Die Umkehrabbildung $f^{-1}: W \rightarrow V \subseteq \mathbb{R}^n$ ist differenzierbar und es gilt

$$Df^{-1}(f(x)) = (Df(x))^{-1}.$$

Ist $x \in U$ mit $\det(Df(x)) \neq 0$, so sagen wir auch: “ f ist lokal bei x umkehrbar” oder “invertierbar”.

Beispiel 2.56. Sei $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ die Abbildung $f(x, y) = (x^2 - y^2, 2xy)$. Wegen $f(-x, -y) = f(x, y)$ ist f nicht injektiv, also nicht global umkehrbar.

Es gilt $Df(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}$, also $\det(Df(x, y)) = 4x^2 + 4y^2$. Daher ist f überall auf $\mathbb{R}^2 \setminus \{0\}$ lokal umkehrbar.

Beispiel 2.57. Betrachte die Abbildung $f: (0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}^2$, $f(r, \phi) = (r \cos(\phi), r \sin(\phi))$. Ihr Bild ist $\mathbb{R}^2 \setminus \{0\}$. Die Abbildung f ist nicht injektiv, da $f(r, \phi) = f(r, \phi + 2\pi)$. Daher besitzt f keine globale Umkehrfunktion.

Es gilt $Df(r, \phi) = \begin{pmatrix} \cos(\phi) & -r \sin(\phi) \\ \sin(\phi) & r \cos(\phi) \end{pmatrix}$ und daher $\det(Df(r, \phi)) = r$ für alle (r, ϕ) . Daher ist f überall lokal umkehrbar. Für die Jacobimatrix der Umkehrfunktion f^{-1} gilt

$$Df^{-1}(r \cos(\phi), r \sin(\phi)) = \begin{pmatrix} \cos(\phi) & -r \sin(\phi) \\ \sin(\phi) & r \cos(\phi) \end{pmatrix}^{-1} = \frac{1}{r} \begin{pmatrix} r \cos(\phi) & r \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix}$$

Setze $(x, y) = f(r, \phi) = (r \cos(\phi), r \sin(\phi))$. Da $r > 0$, gilt $r = \sqrt{x^2 + y^2}$. Also erhalten wir

$$Df^{-1}(x, y) = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} x & y \\ -y & x \end{pmatrix} \frac{1}{\sqrt{x^2 + y^2}}.$$

Alternativ könnte man natürlich versuchen, die Umkehrfunktion explizit zu bestimmen. Z.B. gilt falls $r > 0$ und $\phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ (also $x = r \cos(\phi) > 0$)

$$\phi = \arctan(\tan(\phi)) = \arctan\left(\frac{\sin(\phi)}{\cos(\phi)}\right) = \arctan\left(\frac{y}{x}\right)$$

Also

$$f^{-1}(x, y) = (\sqrt{x^2 + y^2}, \arctan(\frac{y}{x})) \text{ für } (x, y) \in (0, \infty) \times \mathbb{R}.$$

Diese Formel funktioniert aber nicht, falls $x \leq 0$.

Für $y > 0$ können wir $\phi \in (0, \pi)$ wählen. Dann gilt $\tan(\phi - \frac{\pi}{2}) = \frac{-x}{y}$, also

$$\phi = \arctan\left(\frac{-x}{y}\right) + \frac{\pi}{2} = \frac{\pi}{2} - \arctan\left(\frac{x}{y}\right).$$

Für $(x, y) \in \mathbb{R} \times (0, \infty)$ erhalten wir also als Umkehrfunktion

$$f^{-1}(x, y) = (\sqrt{x^2 + y^2}, \frac{\pi}{2} - \arctan(\frac{x}{y})).$$

Implizite Funktionen und Extrema unter Nebenbedingungen

Wir wollen auch Extremwerte von Funktionen auf allgemeineren als nur offenen Teilmengen von $\mathbb{R}^n$ bestimmen.

Beispiel 2.58. Sei $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ die Funktion $f(x, y) = x + y$. Auf $\mathbb{R}^2$ hat sie keine kritischen Punkte, also kein Extremum. Aber wenn wir f auf $S^1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ einschränken, dann hat sie dort ein Minimum und ein Maximum (warum?). Wie können wir dieses finden? Eine naheliegende Möglichkeit ist, S^1 zu parametrisieren: Betrachte die Kurve $\alpha: \mathbb{R} \rightarrow \mathbb{R}^2$, $\alpha(t) = (\cos(t), \sin(t))$. Hat f ein Extremum in einem Punkt $\alpha(t_0) \in S^1$ mit geeignetem $t_0 \in \mathbb{R}$, so hat $f \circ \alpha$ ein Extremum in t_0 . Es gilt $f \circ \alpha(t) = \cos(t) + \sin(t)$, also $(f \circ \alpha)'(t) = \cos(t) - \sin(t)$. Dies wird 0, falls $\cos(t) = \sin(t)$, d.h. für $t = \dots, \frac{\pi}{4} - \pi, \frac{\pi}{4}, \frac{\pi}{4} + \pi, \dots$. Die Bilder dieser Punkte unter α sind $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ und $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. Offenbar ist der erste Punkt ein Maximum, der zweite ein Minimum.

Im Allgemeinen ist es aber schwierig, eine Parametrisierung zu finden. Um ein notwendiges Kriterium für ein Extremum zu finden, genügt aber die Existenz einer Parametrisierung. Diese liefert der Satz über implizite Funktionen. Er behandelt die Auflösbarkeit von Gleichungen.

Beispiel 2.59. Der Einheitskreis $S^1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ ist die Lösungsmenge der Gleichung

$$f(x, y) = 0 \quad \text{mit} \quad f(x, y) = x^2 + y^2 - 1.$$

Können wir die Gleichung $f(x, y) = 0$ nach x oder y auflösen? Genauer fragen wir: Sei (a, b) mit $f(a, b) = 0$ gegeben. Gibt es dann eine differenzierbare Funktion $g(x)$, die in einer Umgebung von a definiert ist, so, dass $g(a) = b$ und $f(x, g(x)) = 0$ für alle x , wo g definiert ist (Auflösbarkeit nach y)? Oder gibt es eine differenzierbare Funktion $h(y)$, die in einer Umgebung von b definiert ist, so, dass $h(b) = a$ und $f(h(y), y) = 0$ für alle y (Auflösbarkeit nach x)?

Für $b > 0$, also $a \in (-1, 1)$ können wir $f(x, y) = 0$ nach y auflösen: Für $x \in (-1, 1)$ ist $y = g(x) = \sqrt{1 - x^2}$ die gesuchte Funktion mit $g(a) = b$.

Genauso, falls $b < 0$: Dann ist $y = g(x) = -\sqrt{1 - x^2}$.

Für $(a, b) = (1, 0)$ (und $(-1, 0)$) ist die Gleichung aber nicht nach y auflösbar! Die Funktion $g(x) = \sqrt{1 - x^2}$ ist in $x = 1$ nicht differenzierbar. Das Problem ist, dass die Tangente an S^1 im Punkt $(1, 0)$ parallel zur y -Achse ist, oder anders gesagt, dass $\nabla f(1, 0)$ senkrecht auf der y -Achse steht.

Dafür können wir $f(x, y) = 0$ bei $(1, 0)$ nach x auflösen: $x = h(y) = \sqrt{1 - y^2}$ ist die gesuchte differenzierbare Funktion von y .

Angenommen, wir haben in (a, b) eine Auflösung nach y gefunden, also eine differenzierbare Funktion $g(x)$ mit $g(a) = b$ und $f(x, g(x)) = 0$. Dann gilt nach der Kettenregel

$$0 = \frac{\partial f}{\partial x}(x, g(x)) \cdot 1 + \frac{\partial f}{\partial y}(x, g(x)) \cdot g'(x),$$

also, falls $\frac{\partial f}{\partial y}(x, g(x)) \neq 0$,

$$g'(x) = - \left(\frac{\partial f}{\partial y}(x, g(x)) \right)^{-1} \cdot \frac{\partial f}{\partial x}(x, g(x)).$$

Die Bedingung $\frac{\partial f}{\partial y}(x, g(x)) \neq 0$ sagt gerade, dass $\nabla f(x, g(x))$ nicht senkrecht auf der y -Achse steht.

In unserem Fall erhalten wir damit

$$g'(x) = -\frac{2x}{2g(x)} = -\frac{x}{\sqrt{1-x^2}}.$$

Beispiel 2.60. Betrachte die Funktion $f(x, y) = x^3 + y^3 - 6xy$ (siehe Abb. 5). Hier gilt

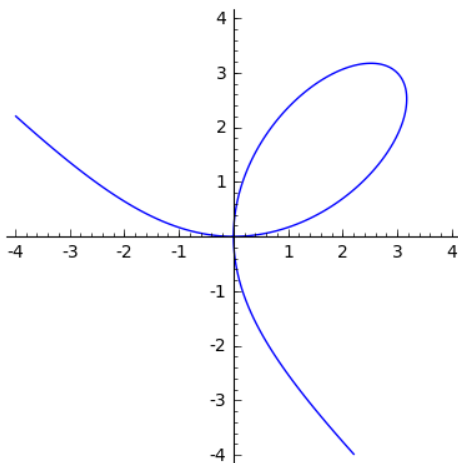


Abbildung 5: $x^3 + y^3 - 6xy = 0$

$\nabla f(0, 0) = 0$. In $(0, 0)$ ist $f(x, y) = 0$ weder nach x noch nach y auflösbar.

Im Punkt $(3, 3)$ gilt $\nabla f(3, 3) = \begin{pmatrix} 9 \\ 9 \end{pmatrix}$. Hier ist die Gleichung sowohl nach x als auch nach y auflösbar. Ist $y = g(x)$ eine Auflösung nach y mit $g(3) = 3$, so erhalten wir wieder mit der Kettenregel

$$g'(x) = -\left(\frac{\partial f}{\partial y}(x, g(x))\right)^{-1} \frac{\partial f}{\partial x}(x, g(x)) = -\frac{3x^2 - 6g(x)}{3g(x)^2 - 6x} = \frac{x^2 - 2g(x)}{2x - g(x)^2},$$

also $g'(3) = -1$. Wir können also die Ableitung von g einfach berechnen, ohne g explizit zu kennen.

Allgemeiner betrachten wir jetzt Systeme von m Gleichungen in $k + m$ Variablen. D.h. wir betrachten eine stetig differenzierbare Funktion $f: U \rightarrow \mathbb{R}^m$, wobei $U \subseteq \mathbb{R}^k \times \mathbb{R}^m$ offen ist. Wir bezeichnen die Variablen in $\mathbb{R}^k \times \mathbb{R}^m$ mit $x_1, \dots, x_k, y_1, \dots, y_m$. Die Jacobimatrix von f im Punkt $(x, y) = (x_1, \dots, x_k, y_1, \dots, y_m)$ ist dann die $m \times (k + m)$ -Matrix

$$Df(x, y) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x, y) & \cdots & \frac{\partial f_1}{\partial x_k}(x, y) & \frac{\partial f_1}{\partial y_1}(x, y) & \cdots & \frac{\partial f_1}{\partial y_m}(x, y) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x, y) & \cdots & \frac{\partial f_m}{\partial x_k}(x, y) & \frac{\partial f_m}{\partial y_1}(x, y) & \cdots & \frac{\partial f_m}{\partial y_m}(x, y) \end{pmatrix}.$$

Zur Abkürzung setzen wir

$$\frac{\partial f}{\partial x}(x, y) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x, y) & \cdots & \frac{\partial f_1}{\partial x_k}(x, y) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x, y) & \cdots & \frac{\partial f_m}{\partial x_k}(x, y) \end{pmatrix} \in \mathbb{R}^{m \times k},$$

$$\frac{\partial f}{\partial y}(x, y) := \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(x, y) & \cdots & \frac{\partial f_1}{\partial y_m}(x, y) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1}(x, y) & \cdots & \frac{\partial f_m}{\partial y_m}(x, y) \end{pmatrix} \in \mathbb{R}^{m \times m}$$

Satz 2.61 (Satz über implizite Funktionen). *Sei $U \subseteq \mathbb{R}^k \times \mathbb{R}^m$ offen, $f: U \rightarrow \mathbb{R}^m$ stetig differenzierbar und $(a, b) \in U$ ein Punkt mit $f(a, b) = 0$. Ist dann $\frac{\partial f}{\partial y}(a, b) \in \mathbb{R}^{m \times m}$ invertierbar, so gibt es Umgebungen $V \subseteq \mathbb{R}^k$ von a und $W \subseteq \mathbb{R}^m$ von b mit $V \times W \subseteq U$ und eine stetig differenzierbare Abbildung $g: V \rightarrow W$ mit folgenden Eigenschaften:*

- (a) $g(a) = b$.
- (b) Für $(x, y) \in V \times W$ gilt $f(x, y) = 0 \Leftrightarrow y = g(x)$.
- (c) $Dg(x) = - \left(\frac{\partial f}{\partial y}(x, g(x)) \right)^{-1} \cdot \frac{\partial f}{\partial x}(x, g(x)) \in \mathbb{R}^{m \times k}$ für alle $x \in V$.

Wir können also die Gleichung $f(x, y) = 0$ lokal bei (a, b) nach y auflösen.

Beweis. Setze $F: U \rightarrow \mathbb{R}^k \times \mathbb{R}^m$, $F(x, y) = (x, f(x, y))$. Dann gilt

$$DF(a, b) = \begin{pmatrix} E_k & 0 \\ \frac{\partial f}{\partial x}(a, b) & \frac{\partial f}{\partial y}(a, b) \end{pmatrix}.$$

Da nach Voraussetzung $\frac{\partial f}{\partial y}(a, b)$ invertierbar ist, ist auch $DF(a, b)$ invertierbar. Nach dem Satz über die Umkehrfunktion 2.55 ist F lokal bei (a, b) umkehrbar. D.h. es gibt eine stetig differenzierbare Abbildung G , die in einer Umgebung von $F(a, b) = (a, 0)$ definiert ist, mit $F(G(x, y)) = (x, y)$ und $G(F(x, y)) = (x, y)$ in einer Umgebung von $(a, 0)$ bzw. (a, b) . Da F die ersten k Koordinaten $x_1, \dots, x_k$ nicht ändert, gilt dies auch für die Umkehrfunktion G . D.h. G ist von der Form $G(x, y) = (x, h(x, y))$ mit einer stetig differenzierbaren Abbildung h . Setze nun $g(x) := h(x, 0)$.

(a) Wegen $F(a, b) = (a, 0)$ gilt $G(a, 0) = (a, b)$, also $g(a) = h(a, 0) = b$.

(b) Es gilt $F(x, g(x)) = F(G(x, 0)) = (x, 0)$, also $f(x, g(x)) = 0$.

Gilt umgekehrt $f(x, y) = 0$, dann ist $F(x, y) = (x, 0)$, also $(x, y) = G(x, 0)$, d.h. $y = h(x, 0) = g(x)$.

(c) Nach Voraussetzung ist $\det(\frac{\partial f}{\partial y}(a, b)) \neq 0$. Da die partiellen Ableitungen von f stetig sind, gilt dies auch in einer Umgebung von (a, b) , d.h. $\frac{\partial f}{\partial y}(x, y)$ ist für alle (x, y) in einer Umgebung von (a, b) invertierbar.

Da $f(x, g(x)) = 0$ für alle x in einer Umgebung von a , folgt

$$0 = \frac{\partial f}{\partial x}(x, g(x)) \cdot E_k + \frac{\partial f}{\partial y}(x, g(x)) \cdot Dg(x),$$

also

$$Dg(x) = - \left(\frac{\partial f}{\partial y}(x, g(x)) \right)^{-1} \cdot \frac{\partial f}{\partial x}(x, g(x)).$$

□

Bemerkung 2.62. Steckt in der Jacobimatrix $Df(a, b)$ irgendeine invertierbare $m \times m$ -Matrix (dies ist äquivalent dazu, dass der Rang⁴ von $Df(a, b)$ gleich m ist), dann können wir $f(x, y) = 0$ lokal bei (a, b) entsprechend auflösen. Betrachte z.B. $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ mit

$$f(x, y, z) = \begin{pmatrix} x^3 - 1 \\ x^3 + y^2 + z - 2 \end{pmatrix}.$$

Dann ist $f(1, 1, 0) = 0$ und

$$Df(1, 1, 0) = \begin{pmatrix} 3 & 0 & 0 \\ 3 & 2 & 1 \end{pmatrix}.$$

Die Matrix $\begin{pmatrix} 0 & 0 \\ 2 & 1 \end{pmatrix}$ ist nicht invertierbar. Aber die beiden Matrizen $\begin{pmatrix} 3 & 0 \\ 3 & 2 \end{pmatrix}$ und $\begin{pmatrix} 3 & 0 \\ 3 & 1 \end{pmatrix}$ sind invertierbar. Daher können wir die Gleichung $f(x, y, z) = 0$, d.h. das Gleichungssystem

$$\begin{aligned} x^3 &= 1 \\ x^3 + y^2 + z &= 2 \end{aligned}$$

lokal bei $(1, 1, 0)$ nach (x, y) und nach (x, z) auflösen. D.h. es gibt $\mathbb{R}^2$ -wertige stetig differenzierbare Abbildungen $g(z) = (g_1(z), g_2(z))$ und $h(y) = (h_1(y), h_2(y))$, die in einer Umgebung von $z = 0$ bzw. $y = 1$ definiert sind, mit $g(0) = (1, 1)$, $h(1) = (1, 0)$ und

$$f(g_1(z), g_2(z), z) = 0 \quad \text{bzw.} \quad f(h_1(y), h_2(y), y) = 0.$$

Definition 2.63. Sei $U \subseteq \mathbb{R}^n$ offen, $g: U \rightarrow \mathbb{R}^m$ stetig differenzierbar und sei $f: U \rightarrow \mathbb{R}$ eine Funktion. Sei $a \in U$ ein Punkt mit $g(a) = 0$ und $\text{rang}(Dg(a)) = m$. Wir sagen, f habe ein (lokales) Maximum (bzw. Minimum) unter der Nebenbedingung $g = 0$, falls es eine Umgebung V von a in U gibt so, dass

$$f(x) \leq f(a) \quad (\text{bzw.} \quad f(x) \geq f(a)) \quad \text{für alle } x \in V \text{ mit } g(x) = 0.$$

Ein notwendiges Kriterium für die Existenz eines lokalen Extremums unter einer Nebenbedingung liefert die Methode der sog. Lagrangeschen Multiplikatoren:

Satz 2.64 (Lagrangesche Multiplikatoren). Seien U, g, a wie in Definition 2.63. Schreibe $g = (g_1, \dots, g_m)$ mit $g_i: U \rightarrow \mathbb{R}$. Sei $f: U \rightarrow \mathbb{R}$ differenzierbar. Hat f in a ein lokales Extremum unter der Nebenbedingung $g = 0$, dann gibt es reelle Zahlen $\lambda_1, \dots, \lambda_m$ so, dass

$$\nabla f(a) = \lambda_1 \nabla g_1(a) + \dots + \lambda_m \nabla g_m(a).$$

⁴siehe LinAlg Def. 3.23 und 3.25

Bevor wir den Satz beweisen, betrachten wir noch einmal das Beispiel vom Beginn dieses Abschnitts: Gesucht sind Extrema von $f(x, y) = x + y$ unter der Nebenbedingung $x^2 + y^2 = 1$. Setze also $g(x, y) = x^2 + y^2 - 1$. Dann gilt $Dg(x, y) = (\nabla g(x, y))^T = (2x, 2y)$. Für (x, y) mit $g(x, y) = 0$ gilt stets $\text{rang}(Dg(x, y)) = 1$. Also können wir die Methode der Lagrangeschen Multiplikatoren anwenden:

Hat f in (x, y) ein Extremum unter der Nebenbedingung $g = 0$, so gibt es ein $\lambda \in \mathbb{R}$ mit

$$\nabla f(x, y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \cdot \nabla g(x, y) = \begin{pmatrix} 2\lambda x \\ 2\lambda y \end{pmatrix}.$$

Dies impliziert $x = y$. Mit der Nebenbedingung $x^2 + y^2 = 1$ erhalten wir also erneut die beiden Möglichkeiten $x = y = \pm \frac{1}{\sqrt{2}}$.

Beweis von Satz 2.64. Nach Voraussetzung gilt $\text{rang}(Dg(a)) = m$, d.h. in der $m \times n$ -Matrix $Dg(a)$ gibt es m linear unabhängige Spalten. Die $m \times m$ -Matrix, die aus diesen Spalten besteht, ist dann invertierbar. Setze $k := n - m$. Nach Umm Nummerierung der Koordinaten können wir annehmen, dass

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_{k+1}}(a) & \cdots & \frac{\partial g_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_{k+1}}(a) & \cdots & \frac{\partial g_m}{\partial x_n}(a) \end{pmatrix}$$

invertierbar ist. Nach dem Satz über implizite Funktionen 2.61 gibt es dann eine in einer Umgebung V von $(a_1, \dots, a_k) \in \mathbb{R}^k$ definierte stetig differenzierbare Abbildung $\phi = (\phi_1, \dots, \phi_m): V \rightarrow \mathbb{R}^m$ mit $\phi(a_1, \dots, a_k) = (a_{k+1}, \dots, a_n)$ und $g(x_1, \dots, x_k, \phi(x_1, \dots, x_k)) = 0$ für alle $x_1, \dots, x_k \in V$. Setze $\Phi: V \rightarrow \mathbb{R}^n$, $\Phi(x_1, \dots, x_k) = (x_1, \dots, x_k, \phi(x_1, \dots, x_k))$.

Da f in a ein Extremum unter der Nebenbedingung $g = 0$ hat, hat die Funktion $f \circ \Phi: V \rightarrow \mathbb{R}$ ein Extremum in $(a_1, \dots, a_k)$, also $D(f \circ \Phi)(a_1, \dots, a_k) = 0$.

Nach der Kettenregel gilt dann

$$0 = D(f \circ \Phi)(a_1, \dots, a_k) = Df(a) \cdot D\Phi(a_1, \dots, a_k),$$

d.h. $Df(a)$ ist eine Lösung des linearen Gleichungssystems (LGS)

$$v \cdot D\Phi(a_1, \dots, a_k) = 0 \quad (v \in \mathbb{R}^{1 \times n}).$$

Da $D\Phi(a_1, \dots, a_k)$ Rang k hat, hat der Lösungsraum dieses LGS die Dimension $n - k = m$.

Weiter gilt nach Konstruktion von Φ auch $g_i \circ \Phi = 0$ für $i = 1, \dots, m$, also

$$Dg_i(a) \cdot D\Phi(a_1, \dots, a_k) = 0$$

für $i = 1, \dots, m$, und die Vektoren $Dg_i(a)$, $i = 1, \dots, m$, sind ebenfalls Lösungen des LGS. Wegen $\text{rang}(Dg(a)) = m$ sind sie linear unabhängig, also eine Basis des Lösungsraums. Dies liefert die Behauptung. $\square$

Beispiel 2.65. Sei $A \in \mathbb{R}^{n \times n}$ eine symmetrische $n \times n$ -Matrix und $f: \mathbb{R}^n \rightarrow \mathbb{R}$ definiert durch $f(x) := x^T A x = \langle x, Ax \rangle$. Die Funktion f ist differenzierbar. Da $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ kompakt ist, nimmt f auf S^{n-1} ein Minimum und ein Maximum an. Diese sind Extrema von f unter der Nebenbedingung $x_1^2 + \dots + x_n^2 = 1$. Sei also $g: \mathbb{R}^n \rightarrow \mathbb{R}$ die Funktion $g(x) = x_1^2 + \dots + x_n^2 - 1$. Für alle $x \in S^{n-1}$ ist $\nabla g(x) = 2x \neq 0$, also können wir die Methode der Lagrangeschen Multiplikatoren anwenden.

Nach Beispiel 2.21(b) ist $Df(x) = x^T(A + A^T) = 2x^T A$, also $\nabla f(x) = 2Ax$. An den Extremstellen gibt es also ein $\lambda \in \mathbb{R}$ mit $2Ax = 2\lambda x$, d.h. x ist ein Eigenvektor von A mit zugehörigem Eigenwert λ . Dies zeigt (wie oben schon aus der linearen Algebra zitiert), dass jede symmetrische Matrix einen Eigenwert besitzt.

Beispiel 2.66. Wir wollen die Extremwerte der Funktion $f: \mathbb{R}^3 \rightarrow \mathbb{R}$, $f(x, y, z) = 3x + 2y + z$, unter den Nebenbedingungen $x - y + z = 1$ und $x^2 + y^2 = 1$ bestimmen.

Betrachte also $g: \mathbb{R}^3 \rightarrow \mathbb{R}^2$,

$$g(x, y, z) = \begin{pmatrix} x - y + z - 1 \\ x^2 + y^2 - 1 \end{pmatrix}.$$

Dann gilt

$$Dg(x, y, z) = \begin{pmatrix} 1 & -1 & 1 \\ 2x & 2y & 0 \end{pmatrix}.$$

Setze $N := \{(x, y, z) : g(x, y, z) = 0\}$. Für $(x, y, z) \in N$ gilt stets $(x, y) \neq (0, 0)$. Daher ist $\text{rang}(Dg(x, y, z)) = 2$ für alle $(x, y, z) \in N$, und wir können Satz 2.64 anwenden.

Es ist $\nabla f(x, y, z) = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$ für alle (x, y, z) . Hat f in (x, y, z) ein Extremum unter der Nebenbedingung $g = 0$, so gibt es λ_1, λ_2 so, dass

$$\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2x \\ 2y \\ 0 \end{pmatrix}$$

Dies ergibt $\lambda_1 = 1$, $2\lambda_2 x = 2$ und $2\lambda_2 y = 3$, also

$$x = \frac{1}{\lambda_2}, \quad y = \frac{3}{2\lambda_2}.$$

Die Nebenbedingung $x^2 + y^2 = 1$ liefert nun

$$1 = \left(\frac{1}{\lambda_2}\right)^2 + \left(\frac{3}{2\lambda_2}\right)^2 = \frac{13}{4\lambda_2^2}, \text{ also } \lambda_2 = \pm \frac{\sqrt{13}}{2}$$

Die z -Koordinate der möglichen Extrema erhalten wir aus der ersten Nebenbedingung $x - y + z = 1$. Insgesamt bekommen wir folgende möglichen Extrema:

$$\left(\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}, 1 + \frac{1}{\sqrt{13}}\right) \text{ und } \left(-\frac{2}{\sqrt{13}}, -\frac{3}{\sqrt{13}}, 1 - \frac{1}{\sqrt{13}}\right).$$

Der Wert der Funktion f an den beiden Stellen ist $1 + \sqrt{13}$ bzw. $1 - \sqrt{13}$.

Handelt es sich tatsächlich um Extrema?

Die Teilmenge $N \subseteq \mathbb{R}^3$ ist abgeschlossen und beschränkt, also kompakt. Daher nimmt f (als stetige Funktion) auf N ein Maximum und ein Minimum an. Nach dem oben gezeigten muss in $(\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}, 1 + \frac{1}{\sqrt{13}})$ ein Maximum vorliegen, in $(-\frac{2}{\sqrt{13}}, -\frac{3}{\sqrt{13}}, 1 - \frac{1}{\sqrt{13}})$ ein Minimum.

3 Ergänzungen zur Integralrechnung

Anwendungen des Hauptsatzes

Erinnerung 3.1. Seien $I \subseteq \mathbb{R}$ ein Intervall, $a \leq b \in I$, $f, g: I \rightarrow \mathbb{R}$ (Riemann-)integrierbar (Ana1, Def. 6.3, z.B. f, g stetig, Ana1, Kor. 6.7), $\lambda, \mu \in \mathbb{R}$. Dann gelten

- (a) $\lambda f + \mu g$ ist integrierbar und $\int_a^b (\lambda f + \mu g)(x) dx = \lambda \int_a^b f(x) dx + \mu \int_a^b g(x) dx$ (*Linearität*).
- (b) Ist $f(x) \leq g(x)$ für alle $x \in I$, dann ist $\int_a^b f(x) dx \leq \int_a^b g(x) dx$ (*Monotonie*).
- (c) $|f|: I \rightarrow \mathbb{R}$ ist integrierbar und $|\int_a^b f(x) dx| \leq \int_a^b |f(x)| dx$ (*Dreiecksungleichung*).
- (d) Für $a \leq b \leq c$ in I gilt $\int_a^b f(x) dx + \int_b^c f(x) dx = \int_a^c f(x) dx$ (*Additivität*).

Außerdem setzt man $\int_b^a f(x) dx := -\int_a^b f(x) dx$.

Der Hauptsatz der Differenzial- und Integralrechnung (Ana1, Satz 6.11) besagt: Sei $I \subseteq \mathbb{R}$ ein Intervall, $a \in I$ und $f: I \rightarrow \mathbb{R}$ stetig. Für $x \in I$ setze $F(x) := \int_a^x f(x) dx$. Dann ist F differenzierbar und es gilt $F'(x) = f(x)$.

Eine *Stammfunktion* von $f: I \rightarrow \mathbb{R}$ ist eine differenzierbare Funktion $F: I \rightarrow \mathbb{R}$ mit $F' = f$.

Sind F, G zwei Stammfunktionen von f , so ist $F - G$ konstant.

Mittels Stammfunktionen kann man Integrale berechnen:

Satz 3.2. Sei $f: I \rightarrow \mathbb{R}$ stetig, und sei F eine Stammfunktion von f . Dann gilt für $a, b \in I$

$$\int_a^b f(x) dx = F(b) - F(a).$$

Beweis. Setze $G(x) = \int_a^x f(x) dx$. Nach dem Hauptsatz ist G eine Stammfunktion. Also ist $F - G$ konstant. Deswegen gilt

$$F(b) - F(a) = G(b) - G(a) = \int_a^b f(x) dx - \int_a^a f(x) dx = \int_a^b f(x) dx.$$

□

Notation 3.3. Für eine Funktion f schreiben wir manchmal $f(x)|_a := f(a)$ und $f(x)|_a^b := f(b) - f(a)$. Also z.B. $\int_a^b f(x) dx = F(x)|_a^b$.

Beispiel 3.4.

$$\int_0^1 \frac{1}{1+x^2} dx = \arctan(x)|_0^1 = \frac{\pi}{4}$$

(siehe Beispiel 2.54).

Beispiel 3.5. Für $x > 0$ gilt $\frac{d}{dx} \log(x) = \frac{1}{x}$. Für $x < 0$ gilt $\frac{d}{dx} \log(-x) = \frac{-1}{-x} = \frac{1}{x}$. Insgesamt also $\frac{d}{dx} \log(|x|) = \frac{1}{x}$ für $x \neq 0$. Daher erhalten wir für $a < b$ mit $0 \notin [a, b]$:

$$\int_a^b \frac{1}{x} dx = \log(|b|) - \log(|a|)$$

Beispiel 3.6. Um $\int_a^b \frac{1}{1-x^2} dx$, wobei $\pm 1 \notin [a, b]$, zu berechnen, kann man die Methode der Partialbruchzerlegung benutzen: Es gilt $1 - x^2 = (1+x)(1-x)$. Ansatz:

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

Dies liefert $\alpha = \beta = \frac{1}{2}$. Also $\frac{1}{1-x^2} = \frac{1}{2}(\frac{1}{1+x} + \frac{1}{1-x}) = \frac{1}{2}(\frac{1}{1+x} - \frac{1}{x-1})$. Daher gilt

$$\begin{aligned} \int_a^b \frac{1}{1-x^2} dx &= \frac{1}{2} \int_a^b \frac{1}{1+x} dx - \frac{1}{2} \int_a^b \frac{1}{x-1} dx \\ &= \frac{1}{2} \log(|1+x|)|_a^b - \frac{1}{2} \log(|x-1|)|_a^b \\ &= \frac{1}{2} \log\left(\frac{|x+1|}{|x-1|}\right)|_a^b. \end{aligned}$$

Satz 3.7 (Substitutionsregel). Sei $f: I \rightarrow \mathbb{R}$ stetig und sei $\phi: [a, b] \rightarrow \mathbb{R}$ stetig differenzierbar mit $\phi([a, b]) \subseteq I$. Dann gilt

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

Beweis. Sei $F: I \rightarrow \mathbb{R}$ eine Stammfunktion von f . Dann gilt $(F \circ \phi)'(t) = f(\phi(t)) \cdot \phi'(t)$, d.h. $F \circ \phi$ ist eine Stammfunktion von $t \mapsto f(\phi(t))\phi'(t)$. Also gilt

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

□

Beispiel 3.8. Sei $\phi: [a, b] \rightarrow \mathbb{R}$ stetig differenzierbar, $\phi(t) \neq 0$ für alle $t \in [a, b]$. Dann gilt

$$\int_a^b \frac{\phi'(t)}{\phi(t)} dt = \log(|\phi(t)|)|_a^b,$$

Substitutionsformel mit $f(x) = \frac{1}{x}$ und Beispiel 3.5.

Beispiel 3.9. Der Flächeninhalt der halben Kreisscheibe $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, y \geq 0\}$ ist

$$\int_{-1}^1 \sqrt{1-x^2} dx.$$

Für $-1 < a < b < 1$ gilt mit Substitution $x = \sin(t)$:

$$\int_a^b \sqrt{1-x^2} dx = \int_{\arcsin(a)}^{\arcsin(b)} \sqrt{1-\sin(t)^2} \cos(t) dt = \int_u^v \cos(t)^2 dt,$$

wobei $u := \arcsin(a)$, $v := \arcsin(b)$. Mit $\cos(t) = \frac{1}{2}(e^{it} + e^{-it})$ erhalten wir

$$\cos(t)^2 = \frac{1}{4}(e^{2it} + 2 + e^{-2it}) = \frac{1}{2} \cos(2t) + \frac{1}{2}.$$

Also

$$\int_a^b \sqrt{1-x^2} dx = \frac{1}{2} \int_u^v (\cos(2t) + 1) dt = \frac{1}{4} \sin(2t)|_u^v + \frac{1}{2} t|_u^v.$$

Es gilt $\sin(2t) = 2 \sin(t) \cos(t) = 2 \sin(t) \sqrt{1-\sin(t)^2}$ und daher

$$\frac{1}{4} \sin(2t)|_u^v = \frac{1}{2} x \sqrt{1-x^2} |_a^b.$$

Damit erhalten wir insgesamt

$$\int_a^b \sqrt{1-x^2} dx = \frac{1}{2} x \sqrt{1-x^2} |_a^b + \frac{1}{2} \arcsin(x) |_a^b.$$

Mit $a \rightarrow -1$, $b \rightarrow 1$, folgt

$$\int_{-1}^1 \sqrt{1-x^2} dx = \frac{1}{2} (\arcsin(1) - \arcsin(-1)) = \frac{\pi}{2}.$$

Satz 3.10 (Partielle Integration). *Seien $f, g: [a, b] \rightarrow \mathbb{R}$ stetig differenzierbar. Dann gilt*

$$\int_a^b f(x) g'(x) dx = f(x) g(x) |_a^b - \int_a^b f'(x) g(x) dx.$$

Beweis. Für $F = fg$ gilt nach der Produktregel (Ana1, Satz 5.10) $F'(x) = f'(x)g(x) + f(x)g'(x)$. Also

$$f(x)g(x)|_a^b = \int_a^b F'(x) dx = \int_a^b f'(x)g(x) dx + \int_a^b f(x)g'(x) dx.$$

Daraus folgt die Behauptung. □

Beispiel 3.11.

$$\begin{aligned} \int_a^b \log(x) dx &= \int_a^b \log(x) \cdot 1 dx = \log(x)x |_a^b - \int_a^b \log'(x)x dx = x \log(x) |_a^b - \int_a^b 1 dx \\ &= x(\log(x) - 1) |_a^b \end{aligned}$$

Als Anwendung der Substitutionsregel diskutieren wir die Bogenlänge von Kurven.

Definition 3.12. Sei $\alpha: [a, b] \rightarrow \mathbb{R}^n$ eine stetig differenzierbare Kurve. Die *Länge* oder *Bogenlänge* von α ist die Zahl

$$L(\alpha) := \int_a^b \|\alpha'(t)\| dt.$$

Eine heuristische Erklärung dieser Definition ist folgende: Wir stellen uns $[a, b]$ in sehr kleine Teilintervalle zerlegt vor: $a = t_0, t_1 < \dots < t_N = b$. Ist t_i nah genug an t_{i-1} , so sollte die Länge des Bogens zwischen $\alpha(t_{i-1})$ und $\alpha(t_i)$ etwa gleich $\|\alpha(t_i) - \alpha(t_{i-1})\|$ sein, und dies wiederum etwa $(t_i - t_{i-1})\|\alpha'(t_{i-1})\|$. Also sollte $\sum_{i=1}^N (t_i - t_{i-1})\|\alpha'(t_{i-1})\|$ eine gute Approximation an die Bogenlänge sein. Macht man die Zerlegung immer feiner, so wird aus der letzten Summe das Integral $\int_a^b \|\alpha'(t)\| dt$.

Dies kann man in der Tat präzise machen, siehe z.B. Forster, Analysis 2, §4.

Beispiel 3.13. Der Einheitskreis wird parametrisiert durch $\alpha: [0, 2\pi] \rightarrow \mathbb{R}^2$, $\alpha(t) = (\cos(t), \sin(t))$. Es gilt $\alpha'(t) = (-\sin(t), \cos(t))$, also $\|\alpha'(t)\| = 1$ für alle $t \in [0, 2\pi]$. Daher erhalten wir $L(\alpha) = 2\pi$, wie erwartet.

Den oberen Halbkreis können wir auch durch $\beta: [-1, 1] \rightarrow \mathbb{R}^2$, $\beta(t) = (t, \sqrt{1-t^2})$, parametrisieren. (Wir vernachlässigen im Folgenden, dass β an den Randpunkten nicht differenzierbar ist; vgl. aber die Argumentation in Beispiel 3.9.) Dann gilt $\beta'(t) = (1, \frac{-t}{\sqrt{1-t^2}})$, also

$$\|\beta'(t)\| = \sqrt{1 + \frac{t^2}{1-t^2}} = \sqrt{\frac{1}{1-t^2}} = \frac{1}{\sqrt{1-t^2}}.$$

Daher gilt (vgl. Beispiel 2.54)

$$\int_{-1}^1 \|\beta'(t)\| dt = \int_{-1}^1 \frac{1}{\sqrt{1-t^2}} dt = \arcsin(t)|_{-1}^1 = \frac{\pi}{2} - (-\frac{\pi}{2}) = \pi,$$

ebenfalls wie erwartet.

Allgemeiner ist die Bogenlänge unabhängig von der Parametrisierung in folgendem Sinn:

Definition 3.14. Seien $[a, b]$ und $[a', b']$ Intervalle, und sei $\phi: [a, b] \rightarrow [a', b']$ bijektiv. Sind ϕ und ϕ^{-1} stetig differenzierbar, so nennen wir ϕ eine *Parametertransformation*. Eine Parametertransformation ϕ heißt *orientierungserhaltend*, falls $\phi'(t) > 0$ für alle $t \in [a, b]$ (also ϕ streng monoton wachsend), und *orientierungsumkehrend*, falls $\phi'(t) < 0$ für alle $t \in [a, b]$ (also ϕ streng monoton fallend).

Satz 3.15. Sei $\alpha: [a', b'] \rightarrow \mathbb{R}^n$ eine stetig differenzierbare Kurve, und sei $\phi: [a, b] \rightarrow [a', b']$ eine Parametertransformation. Dann gilt $L(\alpha \circ \phi) = L(\alpha)$.

Beweis. Wir nehmen an, dass ϕ orientierungsumkehrend ist. Dann gilt $\phi(a) = b'$ und $\phi(b) = a'$. Der andere Fall geht genauso.

$$\begin{aligned}
L(\alpha \circ \phi) &= \int_a^b \|(\alpha \circ \phi)'(t)\| dt \\
&= \int_a^b \|\alpha'(\phi(t))\phi'(t)\| dt \\
&= \int_a^b \|\alpha'(\phi(t))\|(-\phi'(t)) dt && (\text{da } \phi'(t) < 0) \\
&= - \int_{b'}^{a'} \|\alpha'(s)\| ds && (\text{Substitutionsregel 3.7}) \\
&= \int_{a'}^{b'} \|\alpha'(s)\| ds \\
&= L(\alpha).
\end{aligned}$$

□

Als Anwendung der partiellen Integration beweisen wir die eindimensionale Taylorformel zweiter Ordnung 2.42:

Satz 3.16. Sei $I \subseteq \mathbb{R}$ ein Intervall, $g: I \rightarrow \mathbb{R}$ zweimal stetig differenzierbar und $x_0 \in I$. Für $x \in I$ gilt dann

$$g(x) = g(x_0) + (x - x_0) \cdot g'(x_0) + R(x)$$

mit $R(x) = - \int_{x_0}^x (t - x) \cdot g''(t) dt$.

Beweis. Nach dem Hauptsatz gilt

$$\begin{aligned}
g(x) - g(x_0) &= \int_{x_0}^x g'(t) \cdot 1 dt \\
&= g'(t)(t - x)|_{x_0}^x - \int_{x_0}^x g''(t)(t - x) dt && (\text{partielle Integration}) \\
&= -g'(x_0)(x_0 - x) - \int_{x_0}^x (t - x)g''(t) dt.
\end{aligned}$$

Umstellen liefert die Behauptung.

□

Korollar 3.17. Sei $I \subseteq \mathbb{R}$ ein Intervall, $g: I \rightarrow \mathbb{R}$ zweimal stetig differenzierbar und $x_0 \in I$. Für $x \in I$ gibt es ein $\xi \in (x_0, x)$ so, dass

$$g(x) = g(x_0) + (x - x_0) \cdot g'(x_0) + \frac{(x - x_0)^2}{2} g''(\xi)$$

Dies liefert Satz 2.42 durch die Substitution $t_0 = x_0$, $t = x - x_0$, $\tau = \xi - x_0$.

Beweis. Nach Satz 3.16 haben wir

$$g(x) = g(x_0) + (x - x_0) \cdot g'(x_0) + \int_{x_0}^x (x - t) \cdot g''(t) dt.$$

Nach dem Mittelwertsatz der Integralrechnung (Ana1, Satz 6.10) gibt es ein $\xi \in (x_0, x)$ so, dass

$$\int_{x_0}^x (x - t) \cdot g''(t) dt = g''(\xi) \cdot \int_{x_0}^x (x - t) dt = g''(\xi) \frac{-(x - t)^2}{2} \Big|_{x_0}^x = \frac{(x - x_0)^2}{2} g''(\xi).$$

□

Uneigentliche Integrale

Definition 3.18 (uneigentliches Integral über $[a, \infty)$). Sei $a \in \mathbb{R}$ und $f: [a, \infty) \rightarrow \mathbb{R}$ eine Funktion, die auf jedem Intervall $[a, b]$ mit $a < b \in \mathbb{R}$ integrierbar ist (z.B. f stetig). Wir sagen, dass das *uneigentliche Integral* $\int_a^\infty f(x) dx$ *konvergiert*, falls der Grenzwert

$$\int_a^\infty f(x) dx := \lim_{b \rightarrow \infty} \int_a^b f(x) dx$$

existiert. Andernfalls sagen wir, dass das uneigentliche Integral *divergiert*. Analog definiert man das uneigentliche Integral $\int_{-\infty}^b f(x) dx$ für eine Funktion $f: (-\infty, b] \rightarrow \mathbb{R}$.

Beispiel 3.19. Wir betrachten die Funktion $f: [1, \infty) \rightarrow \mathbb{R}$, $f(x) = x^\alpha$ für $\alpha \in \mathbb{R}$.

- (a) Für $\alpha = -1$ gilt $x^\alpha = \frac{1}{x}$ und $\int_1^b \frac{1}{x} dx = \log(b) - \log(1) \rightarrow \infty$ für $b \rightarrow \infty$, d.h. $\int_1^\infty \frac{1}{x} dx$ divergiert.
- (b) Für $\alpha \neq -1$ ist $\frac{1}{\alpha+1} x^{\alpha+1}$ eine Stammfunktion von x^α . Also $\int_1^b x^\alpha dx = \frac{1}{\alpha+1} (b^{\alpha+1} - 1)$. Für $\alpha > -1$ divergiert daher $\int_1^\infty x^\alpha dx$. Für $\alpha < -1$ konvergiert es und hat den Wert $\int_1^\infty x^\alpha dx = -\frac{1}{\alpha+1}$.

Definition 3.20 (uneigentliches Integral über $[a, b)$). Seien $a, b \in \mathbb{R}$ und $f: [a, b) \rightarrow \mathbb{R}$ auf jedem Intervall $[a, c]$ mit $a \leq c < b$ integrierbar. Wir sagen, dass das *uneigentliche Integral* $\int_a^b f(x) dx$ *konvergiert*, falls der Grenzwert

$$\int_a^b f(x) dx := \lim_{c \nearrow b} \int_a^c f(x) dx$$

existiert. Andernfalls sagen wir, dass das uneigentliche Integral *divergiert*. Analog definiert man das uneigentliche Integral $\int_a^b f(x) dx$ für eine Funktion $f: (a, b] \rightarrow \mathbb{R}$.

Beispiel 3.21. Wir betrachten die Funktion $f(x) = x^\alpha$ auf $(0, 1]$. Für $\alpha < 0$ ist diese in 0 nicht definiert.

- (a) $\alpha = -1$: $\int_c^1 \frac{1}{x} dx = -\log(c) \nearrow \infty$ für $c \searrow 0$. D.h. $\int_0^1 \frac{1}{x} dx$ divergiert.

- (b) $\alpha > -1$: $\int_c^1 x^\alpha dx = \frac{1}{\alpha+1} x^{\alpha+1} \Big|_c^1 \rightarrow \frac{1}{\alpha+1}$ für $c \searrow 0$. D.h. für $\alpha > -1$ konvergiert $\int_0^1 x^\alpha dx$.
- (c) $\alpha < -1$: $\int_c^1 x^\alpha dx = \frac{1}{\alpha+1} x^{\alpha+1} \Big|_c^1 \nearrow \infty$ für $c \searrow 0$, d.h. das uneigentliche Integral divergiert.

Beispiel 3.22. Wir betrachten den Logarithmus $\log$ auf $(0, 1]$. Für $c \in (0, 1]$ gilt nach Beispiel 3.11

$$\int_c^1 \log(x) dx = x(\log(x) - 1) \Big|_c^1 = -1 - c \log(c) + c \xrightarrow{c \searrow 0} -1,$$

d.h. das uneigentliche Integral $\int_0^1 \log(x) dx$ konvergiert und $\int_0^1 \log(x) dx = -1$.

Definition 3.23 (uneigentliches Integral über (a, b) , $a = -\infty, b = \infty$ erlaubt). Sei $f: (a, b) \rightarrow \mathbb{R}$ auf jedem kompakten Teilintervall von (a, b) integrierbar. Wähle ein $c \in (a, b)$. Wir sagen, dass das *uneigentliche Integral* $\int_a^b f(x) dx$ *konvergiert*, falls beide uneigentlichen Integrale $\int_a^c f(x) dx$ und $\int_c^b f(x) dx$ konvergieren. In diesem Fall setzen wir

$$\int_a^b f(x) dx := \int_a^c f(x) dx + \int_c^b f(x) dx.$$

Diese Definition ist unabhängig von der Wahl von c .

Beispiel 3.24. (a) Das uneigentliche Integral $\int_0^\infty x^\alpha dx$ konvergiert für kein $\alpha \in \mathbb{R}$ (wähle $c = 1$ und benutze die Beispiele 3.19 und 3.21).

(b) Für $a > 0$ gilt $\int_{-a}^a x dx = 0$. Also existiert der Grenzwert $\lim_{a \rightarrow \infty} \int_{-a}^a x dx = 0$. Das uneigentliche Integral $\int_{-\infty}^\infty x dx$ konvergiert aber nicht, da $\int_0^\infty x dx$ nicht konvergiert.

(c) Das uneigentliche Integral $\int_{-\infty}^\infty \frac{1}{1+x^2} dx$ konvergiert: $\int_0^b \frac{1}{1+x^2} dx = \arctan(x) \Big|_0^b = \arctan(b) \rightarrow \frac{\pi}{2}$ für $b \rightarrow \infty$ und $\int_a^0 \frac{1}{1+x^2} dx = -\arctan(a) \rightarrow \frac{\pi}{2}$ für $a \rightarrow -\infty$, also

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

(d) Das uneigentliche Integral $\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx$ konvergiert ebenfalls: Auf $(-1, 1)$ ist $\arcsin$ eine Stammfunktion, also $\int_0^b \frac{1}{\sqrt{1-x^2}} dx = \arcsin(b) \xrightarrow{b \nearrow 1} \frac{\pi}{2}$, $\int_a^0 \frac{1}{\sqrt{1-x^2}} dx = -\arcsin(a) \xrightarrow{a \searrow -1} \frac{\pi}{2}$ und insgesamt

$$\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx = \pi.$$

Manchmal kann man wenigstens die Konvergenz uneigentlicher Integrale zeigen, ohne sie explizit zu berechnen, also ohne Kenntnis einer Stammfunktion.

Definition 3.25. In der Situationen von Definition 3.18 (und analog bei Def. 3.20 und 3.23) sagen wir, dass das uneigentliche Integral $\int_a^\infty f(x)dx$ *absolut konvergiert*, falls $\int_a^\infty |f(x)|dx$ konvergiert.

Die folgenden beiden Sätze sind nur für den Fall $\int_a^\infty$ aus Definition 3.18 formuliert, gelten aber analog in den anderen beiden Fällen.

Satz 3.26. *Konvergiert $\int_a^\infty f(x)dx$ absolut, so konvergiert es, und es gilt*

$$\left| \int_a^\infty f(x)dx \right| \leq \int_a^\infty |f(x)|dx.$$

Beweis. Für $x \in [a, \infty)$ gilt $0 \leq |f(x)| - f(x) \leq 2|f(x)|$. Wegen der Monotonie des Integrals folgt für alle $b \in [a, \infty)$

$$\int_a^b (|f(x)| - f(x))dx \leq 2 \int_a^b |f(x)|dx \leq 2 \int_a^\infty |f(x)|dx.$$

Die zweite Ungleichung gilt, da nach Voraussetzung $\int_a^\infty |f(x)|dx$ konvergiert und der Integrand nicht negativ ist. Da auch $|f(x)| - f(x) \geq 0$, wächst $\int_a^b (|f(x)| - f(x))dx$ monoton für $b \nearrow \infty$. Da es außerdem beschränkt ist, existiert der Grenzwert $\lim_{b \rightarrow \infty} \int_a^b (|f(x)| - f(x))dx = \int_a^\infty (|f(x)| - f(x))dx$ (Ana1, Satz 2.62). Es folgt, dass auch $\int_a^\infty f(x)dx$ konvergiert.

Aus der Dreiecksungleichung $\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx$ folgt mit $b \rightarrow \infty$ die zweite Behauptung. $\square$

Satz 3.27 (Majorantenkriterium). *Seien $f, g: [a, \infty) \rightarrow \mathbb{R}$ auf jedem Intervall $[a, b] \subseteq [a, \infty)$ integrierbar. Konvergiert $\int_a^\infty g(x)dx$ und gilt $|f(x)| \leq g(x)$ für alle $x \in [a, \infty)$, dann konvergiert $\int_a^\infty f(x)dx$ absolut.*

Man nennt g eine *Majorante* für f .

Beweis. Für alle $b \in [a, \infty)$ gilt wieder wegen der Monotonie des Integrals

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

Wie im Beweis von Satz 3.26 folgt, dass das uneigentliche Integral $\int_a^\infty |f(x)|dx$ konvergiert. $\square$

Beispiel 3.28. (a) $\int_1^\infty \frac{\cos(x)}{x^2}dx$ konvergiert absolut, da $\int_1^\infty \frac{1}{x^2}dx$ konvergiert (siehe Beispiel 3.19). Das uneigentliche Integral $\int_0^1 \frac{\cos(x)}{\sqrt{x}}dx$ konvergiert absolut, da $\int_0^1 \frac{1}{\sqrt{x}}dx$ konvergiert (Beispiel 3.21).

(b) Das uneigentliche Integral $\int_0^\infty \frac{\sin(x)}{x}dx$ konvergiert, aber es konvergiert nicht absolut.

Beweis. Zunächst gilt $\lim_{x \searrow 0} \frac{\sin(x)}{x} = \lim_{x \searrow 0} \frac{\cos(x)}{1} = 1$ nach l'Hospital (Ana 1, Satz 5.23), die untere Integrationsgrenze ist also unproblematisch.

Wir zeigen erst, dass $\int_0^\infty \frac{\sin(x)}{x} dx$ nicht absolut konvergiert. Es gilt:

$$\begin{aligned} \int_0^{k\pi} \left| \frac{\sin(x)}{x} \right| dx &= \sum_{\ell=1}^k \int_{(\ell-1)\pi}^{\ell\pi} \left| \frac{\sin(x)}{x} \right| dx \\ &\geq \sum_{\ell=1}^k \frac{1}{\ell\pi} \int_{(\ell-1)\pi}^{\ell\pi} |\sin(x)| dx \\ &= \frac{2}{\pi} \sum_{\ell=1}^k \frac{1}{\ell} \end{aligned}$$

Aber $\sum_{\ell=1}^k \frac{1}{\ell} \rightarrow \infty$ für $k \rightarrow \infty$. Also divergiert $\int_0^\infty \left| \frac{\sin(x)}{x} \right| dx$.

Um zu zeigen, dass $\int_0^\infty \frac{\sin(x)}{x} dx$ konvergiert, reicht es zu zeigen, dass $\int_{\frac{\pi}{2}}^\infty \frac{\sin(x)}{x} dx$ konvergiert. Für $b \geq \frac{\pi}{2}$ gilt mit partieller Integration

$$\int_{\frac{\pi}{2}}^b \frac{1}{x} \sin(x) dx = \left. \frac{-\cos(x)}{x} \right|_{\frac{\pi}{2}}^b - \int_{\frac{\pi}{2}}^b \frac{\cos(x)}{x^2} dx = -\frac{\cos(b)}{b} - \int_{\frac{\pi}{2}}^b \frac{\cos(x)}{x^2} dx.$$

Der erste Term geht für $b \rightarrow \infty$ gegen 0; der zweite konvergiert nach dem Majorantenkriterium. $\square$

Satz 3.29 (Integralvergleichskriterium). *Sei $f: [1, \infty) \rightarrow \mathbb{R}$ auf jedem Teilintervall $[1, b]$ integrierbar, monoton fallend und $f(x) \geq 0$ für alle $x \in [1, \infty)$. Dann gilt:*

$$\sum_{k=1}^{\infty} f(k) \text{ konvergiert} \iff \int_1^{\infty} f(x) dx \text{ konvergiert.}$$

Beweis. Für $x \in [k, k+1]$ gilt $f(k) \geq f(x) \geq f(k+1)$. Wegen der Monotonie des Integrals folgt

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

und damit

$$\sum_{k=1}^N f(k) \geq \int_1^{N+1} f(x) dx \geq \sum_{k=2}^{N+1} f(k).$$

Daraus folgt die Behauptung. $\square$

Beispiel 3.30. Mit der Funktion $f(x) = \frac{1}{x^\alpha} = x^{-\alpha}$ und Beispiel 3.19 erhalten wir:

$$\sum_{k=1}^{\infty} \frac{1}{k^\alpha} \text{ konvergiert} \iff \alpha > 1.$$

Integral und Grenzwert

Erinnerung 3.31. Sei X ein metrischer Raum (z.B. ein Intervall in $\mathbb{R}$) und sei $(f_n)_{n \in \mathbb{N}}$ eine Folge von Funktionen $f_n: X \rightarrow \mathbb{R}$. Die Folge $(f_n)_{n \in \mathbb{N}}$ *konvergiert gleichmäßig* gegen $f: X \rightarrow \mathbb{R}$, wenn es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ gibt so, dass

$$|f(x) - f_n(x)| < \epsilon \text{ für alle } n \geq N \text{ und } x \in X.$$

Dies ist gleichbedeutend mit

$$\|f - f_n\|_\infty \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Konvergiert die Folge gleichmäßig und sind alle f_n stetig, dann ist auch f stetig (Ana1, Satz 3.32).

Gilt nur

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \text{ für alle } x \in X,$$

so sagen wir, $(f_n)_{n \in \mathbb{N}}$ konvergiere *punktweise* gegen f .

Beispiel 3.32. Betrachte die Folge von Funktionen $f_n: [0, 1] \rightarrow \mathbb{R}$, $f_n(x) = x^n$. Dann konvergiert $(f_n)_{n \in \mathbb{N}}$ punktweise gegen die Funktion $f: [0, 1] \rightarrow \mathbb{R}$, $f(x) = 0$ für $x < 1$, $f(1) = 1$. Sie konvergiert aber nicht gleichmäßig.

Satz 3.33. Sei $I = [a, b] \subseteq \mathbb{R}$ ein Intervall, und sei $(f_n)_{n \in \mathbb{N}}$ eine Folge stetiger Funktionen $f_n: I \rightarrow \mathbb{R}$, welche gleichmäßig gegen die Funktion $f: I \rightarrow \mathbb{R}$ konvergiert. (Also ist f stetig und insbesondere integrierbar). Dann gilt

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b f(x) dx.$$

Beweis. Sei $\epsilon > 0$ gegeben. Wähle N so groß, dass $\|f - f_n\|_\infty \leq \frac{\epsilon}{b-a}$ für alle $n \geq N$ gilt. Dann gilt für diese n

$$\begin{aligned} \left| \int_a^b f(x) dx - \int_a^b f_n(x) dx \right| &= \left| \int_a^b (f(x) - f_n(x)) dx \right| \\ &\leq \int_a^b |f(x) - f_n(x)| dx && \text{(Dreiecksungleichung)} \\ &\leq \int_a^b \frac{\epsilon}{b-a} dx \\ &= \epsilon \end{aligned}$$

Daraus folgt die Behauptung. □

Beispiel 3.34. (a) Betrachte die Funktionen $f_n: [0, 1] \rightarrow \mathbb{R}$,

$$f_n(x) = \begin{cases} 2n^2 x & \text{für } x \leq \frac{1}{2n}, \\ 2n - 2n^2 x & \text{für } \frac{1}{2n} \leq x \leq \frac{1}{n}, \\ 0 & \text{für } \frac{1}{n} \leq x \leq 1. \end{cases}$$

Die Funktionen f_n sind stetig und konvergieren für $n \rightarrow \infty$ punktweise gegen die Nullfunktion (da $f_n(x) = 0$ sobald $n \geq \frac{1}{x}$). Sie konvergiert aber nicht gleichmäßig, da $\|f_n - 0\|_\infty = \|f_n\|_\infty = n$. Für die Integrale gilt $\int_0^1 f_n(x) dx = \frac{1}{2}$ für alle n . Also konvergieren sie, aber nicht gegen das Integral der Grenzfunktion.

- (b) Setze $f_n(x) = \sum_{k=0}^n \frac{x^k}{k!}$. Die Folge von Funktionen $(f_n)_{n \in \mathbb{N}}$ konvergiert auf ganz $\mathbb{R}$ gleichmäßig gegen die Exponentialfunktion $\exp$. Daher gilt

$$\begin{aligned} \int_a^b \exp(x) dx &= \int_a^b \sum_{k=0}^{\infty} \frac{x^k}{k!} dx = \sum_{k=0}^{\infty} \int_a^b \frac{x^k}{k!} dx = \sum_{k=0}^{\infty} \frac{x^{k+1}}{(k+1)!} \Big|_a^b = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Big|_a^b \\ &= \exp(x) \Big|_a^b, \end{aligned}$$

wie wir natürlich schon wussten.

Als Anwendung erhalten wir auch ein Kriterium, wann man Ableitung und Grenzwert vertauschen kann:

Satz 3.35. Sei $I \subseteq \mathbb{R}$ ein Intervall, und sei $(f_n)_{n \in \mathbb{N}}$ eine Folge stetig differenzierbarer Funktionen $f_n: I \rightarrow \mathbb{R}$, die punktweise gegen die Funktion $f: I \rightarrow \mathbb{R}$ konvergiere. Konvergiert die Folge der Ableitungen $(f'_n)_{n \in \mathbb{N}}$ gleichmäßig gegen die Funktion $g: I \rightarrow \mathbb{R}$, dann ist f stetig differenzierbar und es gilt $f' = g$.

D.h. also die Ableitung der Grenzfunktion ist die Grenzfunktion der Ableitungen.

Beweis. Fixiere ein $x_0 \in I$. Für beliebige $x \in I$ gilt dann

$$\begin{aligned} f(x) - f(x_0) &= \lim_{n \rightarrow \infty} f_n(x) - f_n(x_0) \\ &= \lim_{n \rightarrow \infty} \int_{x_0}^x f'_n(t) dt \\ &= \int_{x_0}^x g(t) dt \quad (\text{Satz 3.33 und Voraussetzung}), \end{aligned}$$

d.h. $f(x) = f(x_0) + \int_{x_0}^x g(t) dt$. Nach dem Hauptsatz der Differential- und Integralrechnung ist daher f differenzierbar mit Ableitung $f' = g$. $\square$

Beispiel 3.36. Betrachte $f_n: \mathbb{R} \rightarrow \mathbb{R}$, $f_n(x) = \frac{1}{n} \sin(nx)$. Die Folge $(f_n)_{n \in \mathbb{N}}$ konvergiert gleichmäßig gegen die Nullfunktion. Es gilt $f'_n(x) = \cos(nx)$. Da z.B. $f'_n(\pi) = \cos(n\pi) = (-1)^n$ gilt, konvergiert die Folge der Ableitungen $(f'_n)_{n \in \mathbb{N}}$ nicht punktweise, also erst recht nicht gleichmäßig. Der Satz ist also nicht anwendbar.

4 Gewöhnliche Differenzialgleichungen

Definition und elementare Lösungsmethoden

Eine Differenzialgleichung (DGL) ist eine Gleichung für die Ableitung einer Funktion in Termen der Funktion selbst. Beispiel: Die DGL

$$y' = 2y$$

bedeutet, dass wir nach einer Funktion ϕ suchen, für die $\phi' = 2\phi$, d.h. $\phi'(x) = 2\phi(x)$ für alle x im Definitionsbereich von ϕ gilt. Eine Lösung wäre z.B. die Funktion $\phi(x) = e^{2x}$. Aber auch $\phi(x) = ce^{2x}$ für beliebiges $c \in \mathbb{R}$ ist eine Lösung der DGL. Die Lösung einer DGL ist also nicht eindeutig.

Gibt man zusätzlich eine Bedingung wie $y(0) = 1$ vor, so spricht man von einem Anfangswertproblem (AWP). So eines ist oft eindeutig lösbar. In unserem Beispiel ist $\phi(x) = e^{2x}$ die eindeutige Lösung des AWP $y' = 2y$, $y(0) = 1$.

Die rechte Seite der DGL darf aber auch noch von x abhängen, z.B.

$$y' = xy.$$

Eine Lösung dieser Gleichung ist z.B. die Funktion $\phi(x) = e^{\frac{x^2}{2}}$.

Das wollen wir jetzt formalisieren:

Definition 4.1. Sei $U \subseteq \mathbb{R}^2$ eine Teilmenge und $f: U \rightarrow \mathbb{R}$ eine Funktion. Dann nennt man

$$y' = f(x, y)$$

eine (*gewöhnliche*) *Differenzialgleichung erster Ordnung*. Eine *Lösung* der DGL ist eine auf einem Intervall $I \subseteq \mathbb{R}$ definierte differenzierbare Funktion $\phi: I \rightarrow \mathbb{R}$ mit folgenden Eigenschaften:

(a) Der Graph von ϕ liegt ganz in U .

(b) Für alle $x \in I$ gilt

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

Ist zusätzlich $(x_0, y_0) \in U$ und die Bedingung $y(x_0) = y_0$ gegeben, so spricht man von einem *Anfangswertproblem* (AWP). Eine Lösung des AWP ist eine Lösung $\phi: I \rightarrow \mathbb{R}$ der Differenzialgleichung mit $x_0 \in I$ und $\phi(x_0) = y_0$.

Unter der *allgemeinen Lösung* der DGL $y' = f(x, y)$ versteht man die Angabe aller Lösungen zu beliebigen Anfangsbedingungen $y(x_0) = y_0$, $(x_0, y_0) \in U$.

Im ersten Beispiel oben ist f die Funktion $f(x, y) = 2y$, im zweiten $f(x, y) = xy$. Die allgemeine Lösung der DGL $y' = 2y$ ist $\phi(x) = ce^{2x}$, $c \in \mathbb{R}$. Die allgemeine Lösung der DGL $y' = xy$ ist $\phi(x) = ce^{\frac{x^2}{2}}$, $c \in \mathbb{R}$.

Bemerkung 4.2. Graphische Interpretation einer DGL: Die Funktion f liefert ein Vektorfeld auf $U \subseteq \mathbb{R}^2$ durch $(x, y) \mapsto (1, f(x, y))$. Für eine Lösung ϕ der DGL ist dieses Vektorfeld 'tangential' an den Graph von ϕ , siehe Abbildungen 6 und 7.

In diesem Abschnitt werden wir einige elementare Methoden zur Bestimmung von Lösungen einer DGL einführen.

Beispiel 4.3. Eine Lösung der DGL $y' = xy$ kann man vielleicht erraten. Etwas systematischer kann man hier (etwas unpräzise) so vorgehen:

- Teile durch y und erhalte $\frac{y'}{y} = x$.

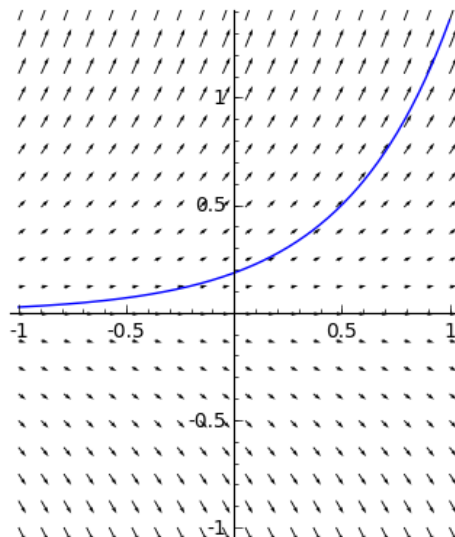


Abbildung 6: DGL $y' = 2y$ mit Lösung $\frac{1}{2}e^{2x-1}$

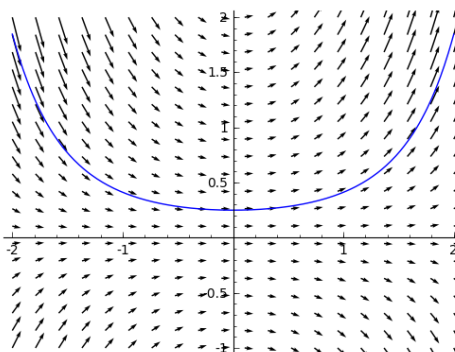


Abbildung 7: DGL $y' = xy$ mit Lösung $\frac{1}{4}e^{x^2/2}$

- Die linke Seite ist $\log(|y|)'$, also $\log(|y|)' = x$.
- Integrieren liefert $\log(|y|) = \frac{x^2}{2} + c$ mit einer Konstanten c , also $y = \pm e^{\frac{x^2}{2} + c}$.

Hierbei haben wir uns nicht darum gekümmert, ob y Null werden könnte, das Teilen durch y also überhaupt erlaubt war.

Noch etwas nachlässiger, aber vielleicht eingängiger, gehen wir so vor:

$$\frac{dy}{dx} = xy \rightsquigarrow \frac{dy}{y} = x dx \rightsquigarrow \int \frac{dy}{y} = \int x dx \rightsquigarrow \log(|y|) = \frac{x^2}{2} + c \rightsquigarrow y = \pm e^{\frac{x^2}{2} + c}$$

Dass dieses Vorgehen funktioniert, liegt daran, dass es sich um eine so genannte DGL mit getrennten Variablen handelt:

Definition 4.4. Eine Differenzialgleichung *mit getrennten Variablen* ist eine DGL der Form

$$y' = f(x)g(y).$$

Hierbei sind $I, J \subseteq \mathbb{R}$ offene Intervalle und $f: I \rightarrow \mathbb{R}$, $g: J \rightarrow \mathbb{R}$, stetige Funktionen mit $g(y) \neq 0$ für alle $y \in J$.

Für diese DGL haben wir:

Satz 4.5 (DGL mit getrennten Variablen). *Fixiere $(x_0, y_0) \in I \times J$ und definiere $F: I \rightarrow \mathbb{R}$, $G: J \rightarrow \mathbb{R}$ durch*

$$F(x) := \int_{x_0}^x f(t)dt, \quad G(y) := \int_{y_0}^y \frac{1}{g(s)}ds.$$

Sei $I' \subseteq I$ ein Intervall mit $x_0 \in I$ und $F(I') \subseteq G(J)$. Dann gibt es genau eine Lösung $\phi: I' \rightarrow \mathbb{R}$ des AWP

$$y' = f(x)g(y), \quad y(x_0) = y_0.$$

Die Funktion ϕ erfüllt die Gleichung

$$(4.1) \quad G(\phi(x)) = F(x) \quad \text{für alle } x \in I'.$$

Da $G'(y) = \frac{1}{g(y)} \neq 0$ für alle $y \in J$, ist die Funktion G streng monoton, besitzt also eine stetig differenzierbare Umkehrfunktion $H: G(J) \rightarrow J$. Also ist ϕ durch (4.1) eindeutig bestimmt:

$$(4.2) \quad \phi(x) = H(F(x)) \quad \text{für alle } x \in I'.$$

In der Praxis muss man H explizit kennen und ϕ zu berechnen.

Beweis. (Eindeutigkeit) Wir zeigen, dass jede Lösung $\phi: I' \rightarrow \mathbb{R}$ des AWP Gleichung (4.1) erfüllt. Dann folgt die Eindeutigkeit aus der Vorbemerkung. Gelte also

$$\phi'(x) = f(x)g(\phi(x)), \quad \phi(x_0) = y_0.$$

Dann folgt

$$\begin{aligned} F(x) &= \int_{x_0}^x f(t)dt = \int_{x_0}^x \frac{\phi'(t)}{g(\phi(t))}dt \\ &= \int_{y_0}^{\phi(x)} \frac{1}{g(s)}ds && \text{(Substitutionsformel)} \\ &= G(\phi(x)). \end{aligned}$$

(Existenz) Wir definieren die Funktion $\phi: I' \rightarrow \mathbb{R}$ durch Gleichung (4.2). Dann gilt $\phi(x_0) = H(F(x_0)) = H(0) = y_0$, da $G(y_0) = 0$.

Außerdem gilt $G(\phi(x)) = F(x)$ für alle $x \in I'$, also

$$\frac{\phi'(x)}{g(\phi(x))} = G'(\phi(x))\phi'(x) = F'(x) = f(x),$$

also $\phi'(x) = f(x)g(\phi(x))$. Daher ist ϕ eine Lösung des AWP. □

Beispiel 4.6. Wir betrachten die DGL

$$y' = -\frac{y}{x} \text{ auf } I \times J := (0, \infty) \times (0, \infty).$$

Mit den Notationen von oben ist also $f(x) = -\frac{1}{x}$, $g(y) = y$. Wir suchen eine Lösung ϕ der DGL mit der Anfangsbedingung $\phi(1) = c > 0$. Also:

$$F(x) = - \int_1^x \frac{1}{t} dt = -\log(x)$$

und

$$G(y) = \int_c^y \frac{1}{s} ds = \log(y) - \log(c) = \log\left(\frac{y}{c}\right).$$

Es gilt $G((0, \infty)) = \mathbb{R}$. Wir können also $I' = I = (0, \infty)$ wählen. Nach Satz 4.5 gilt für die Lösung ϕ

$$\log\left(\frac{\phi(x)}{c}\right) = -\log(x),$$

also

$$\phi(x) = \frac{c}{x}, \quad x \in (0, \infty).$$

Hätten wir die gleiche DGL auf $I \times J = (-\infty, 0) \times (0, \infty)$ mit der Anfangsbedingung $\phi(-1) = c > 0$ betrachtet, so hätte sich

$$F(x) = - \int_{-1}^x \frac{1}{t} dt = -\log(-x)$$

und als Lösung

$$\phi(x) = \frac{-c}{x}$$

ergeben.

Etwas informeller können wir auch so vorgehen:

- $y' = -\frac{y}{x}$, $\rightsquigarrow \frac{y'}{y} = \frac{-1}{x} \rightsquigarrow \log(|y|)' = \frac{-1}{x} \rightsquigarrow \log(|y|) = -\log(|x|) + d$ mit einer Konstante $d \rightsquigarrow |y| = \frac{e^d}{|x|}$
- Bestimme jetzt die Konstante d , das Vorzeichen und den Definitionsbereich aus den Anfangsbedingungen: Geben wir wieder die Anfangsbedingung $\phi(1) = c > 0$ vor, so ergibt sich $d = \log(c)$ und $\phi(x) = \frac{c}{x}$. Das maximale Definitionsintervall dieser Funktion ϕ , welches 1 enthält, ist $(0, \infty)$.
- Man prüft direkt nach, dass ϕ auch tatsächlich eine Lösung der DGL ist.

Beispiel 4.7. Betrachte jetzt die DGL

$$y' = \frac{x}{y} \text{ auf } I \times J := \mathbb{R} \times (0, \infty)$$

Also:

$$\frac{dy}{dx} = \frac{x}{y} \rightsquigarrow ydy = xdx \rightsquigarrow \int ydy = \int xdx \rightsquigarrow \frac{y^2}{2} = \frac{x^2}{2} + \text{const.} \rightsquigarrow y^2 = x^2 + \text{const.}$$

Als Kandidaten für eine Lösung erhalten wir also $\phi(x) = \pm\sqrt{x^2 + c}$ mit einer Konstanten c . Da wir die DGL nur auf $\mathbb{R} \times (0, \infty)$ betrachten, kommt $-\sqrt{x^2 + c}$ als Lösung nicht in Frage. Geben wir die Anfangsbedingung $\phi(1) = d$ mit $d > 0$ vor, so erhalten wir also als mögliche Lösung $\phi(x) = \sqrt{x^2 - 1 + d^2}$. Für $d \leq 1$ ist das maximale Definitionsintervall von ϕ das Intervall $(\sqrt{1 - d^2}, \infty)$. Ist $d > 1$, so ist das maximale Definitionsintervall ganz $\mathbb{R}$. Man prüft wieder direkt nach, dass es sich tatsächlich um Lösungen der DGL handelt.

Definition 4.8. Sei $I \subseteq \mathbb{R}$ ein Intervall. Eine DGL der Form

$$y' = a(x)y + b(x)$$

mit stetigen Funktionen $a, b: I \rightarrow \mathbb{R}$ heißt *lineare Differenzialgleichung erster Ordnung*. Sie heißt *homogen*, falls $b = 0$, und sonst *inhomogen*.

Z.B. sind die oben betrachteten DGL $y' = 2y$, $y' = xy$ und $y' = -\frac{y}{x}$ homogene lineare DGL, $y' = \frac{x}{y}$ ist hingegen nicht linear.

Bemerkung 4.9 (Superpositionsprinzip). Sind $\phi_1, \phi_2: I \rightarrow \mathbb{R}$ zwei Lösungen der *homogenen* linearen DGL $y' = a(x)y$, so ist auch jede Linearkombination $\lambda_1\phi_1 + \lambda_2\phi_2$ mit $\lambda_1, \lambda_2 \in \mathbb{R}$ eine Lösung, denn

$$(\lambda_1\phi_1 + \lambda_2\phi_2)'(x) = \lambda_1\phi_1'(x) + \lambda_2\phi_2'(x) = \lambda_1a(x)\phi_1(x) + \lambda_2a(x)\phi_2(x) = a(x)(\lambda_1\phi_1 + \lambda_2\phi_2)(x).$$

Ist $\phi_s: I \rightarrow \mathbb{R}$ eine Lösung der inhomogenen linearen DGL $y' = a(x)y + b(x)$ und $\phi_1: I \rightarrow \mathbb{R}$ eine Lösung der homogenen linearen DGL $y' = a(x)y$, dann ist auch $\phi_1 + \phi_s$ eine Lösung der inhomogenen DGL:

$$(\phi_1 + \phi_s)'(x) = \phi_1'(x) + \phi_s'(x) = a(x)\phi_1(x) + a(x)\phi_s(x) + b(x) = a(x)(\phi_1 + \phi_s)(x) + b(x).$$

Genauso zeigt man auch: Sind zwei Lösungen der inhomogenen linearen DGL $y' = a(x)y + b(x)$ gegeben, so ist ihre Differenz eine Lösung der homogenen linearen DGL $y' = a(x)y$.

Satz 4.10 (homogene lineare DGL). Sei $I \subseteq \mathbb{R}$ ein Intervall und sei $a: I \rightarrow \mathbb{R}$ stetig. Sei $x_0 \in I$ und $y_0 \in \mathbb{R}$. Dann hat die homogene lineare DGL

$$y' = a(x)y$$

eine eindeutige Lösung $\phi: I \rightarrow \mathbb{R}$, welche die Anfangsbedingung $\phi(x_0) = y_0$ erfüllt. Für sie gilt

$$(4.3) \quad \phi(x) = y_0 \exp \left(\int_{x_0}^x a(t) dt \right).$$

Die allgemeine Lösung der homogenen linearen DGL ist also $\phi(x) = c \exp(A(x))$, $c \in \mathbb{R}$, wobei $A: I \rightarrow \mathbb{R}$ eine fixierte Stammfunktion von a ist.

Beweis. (Existenz) Durch (4.3) wird eine differenzierbare Funktion $\phi: I \rightarrow \mathbb{R}$ definiert. Für sie gilt $\phi(x_0) = y_0$ und nach Kettenregel und Hauptsatz $\phi'(x) = \phi(x) \cdot a(x)$, d.h. ϕ ist die gesuchte Lösung.

(Eindeutigkeit) Sei $\psi: I \rightarrow \mathbb{R}$ irgendeine Lösung der DGL. Setze $\tilde{\psi}(x) := \exp\left(-\int_{x_0}^x a(t)dt\right)$ und $\xi(x) := \psi(x)\tilde{\psi}(x)$. Es gilt $\xi(x_0) = y_0$ und es ist zu zeigen, dass ξ konstant ist. Nach obigem ist $\tilde{\psi}'(x) = -a(x)\tilde{\psi}(x)$, also

$$\begin{aligned}\xi'(x) &= \psi'(x)\tilde{\psi}(x) + \psi(x)\tilde{\psi}'(x) && \text{(Produktregel)} \\ &= a(x)\psi(x)\tilde{\psi}(x) - a(x)\psi(x)\tilde{\psi}(x) && (\psi \text{ Lösung der DGL}) \\ &= 0.\end{aligned}$$

Also ist ξ konstant. □

Die homogene lineare DGL erster Ordnung ist auch eine DGL mit getrennten Variablen. Wir hätten also auch einfach Satz 4.5 anwenden können.

Beispiel 4.11. (a) Sei $\alpha \in \mathbb{R}$. Dann hat die homogene lineare DGL $y' = \alpha y$ mit der Anfangsbedingung $y(x_0) = y_0$ ($x_0, y_0 \in \mathbb{R}$) die Lösung $\phi(x) = y_0 e^{\alpha(x-x_0)}$.

(b) Die DGL $y' = x^2 y$ mit der Anfangsbedingung $y(x_0) = y_0$ hat die Lösung $\phi(x) = y_0 e^{\frac{1}{3}(x^3-x_0^3)}$.

Die *inhomogene* lineare DGL erster Ordnung ist hingegen keine DGL mit getrennten Variablen. Man löst sie mit der Methode der *Variation der Konstanten*: Sei also $I \subseteq \mathbb{R}$ ein Intervall, $x_0 \in I$ und $a, b: I \rightarrow \mathbb{R}$ stetig. Wir wollen die DGL

$$(4.4) \quad y' = a(x)y + b(x)$$

mit der Anfangsbedingung $y(x_0) = y_0$ lösen. Bestimme zunächst die Lösung $\phi_0(x)$ der zugehörigen homogenen DGL

$$y' = a(x)y$$

mit der Anfangsbedingung $y(x_0) = 1$. Nach Satz 4.10 also

$$\phi_0(x) = \exp\left(\int_{x_0}^x a(t)dt\right).$$

Nun “variieren wir die Konstante”, d.h. wir machen den Ansatz

$$\phi(x) = c(x)\phi_0(x).$$

für eine Funktion $c: I \rightarrow \mathbb{R}$. Dann gelten

$$\begin{aligned}\phi'(x) &= c'(x)\phi_0(x) + c(x)\phi_0'(x) = c'(x)\phi_0(x) + c(x)a(x)\phi_0(x) \text{ und} \\ a(x)\phi(x) + b(x) &= a(x)c(x)\phi_0(x) + b(x).\end{aligned}$$

Daher ist ϕ genau dann eine Lösung von (4.4), falls

$$c'(x)\phi_0(x) = b(x) \text{ für alle } x \in I$$

gilt. Da ϕ_0 nirgends verschwindet, können wir diese Gleichung nach c' auflösen und erhalten

$$c(x) = c(x_0) + \int_{x_0}^x \phi_0(t)^{-1} b(t) dt.$$

Wegen $\phi(x_0) = c(x_0)$ ist die Anfangsbedingung $\phi(x_0) = y_0$ ist genau dann erfüllt, wenn $c(x_0) = y_0$. Also ist

$$\phi(x) = \left(y_0 + \int_{x_0}^x \phi_0(t)^{-1} b(t) dt \right) \cdot \phi_0(x)$$

die gesuchte Lösung.

Beispiel 4.12. Gegeben sei die inhomogene lineare DGL

$$y' = 2xy + x^3.$$

Eine Lösung der homogenen linearen DGL $y' = 2xy$ zur Anfangsbedingung $y(0) = 1$ ist $\phi_0(x) = e^{x^2}$. Daher lautet die Lösung der inhomogenen linearen DGL mit Anfangsbedingung $y(0) = y_0$

$$\phi(x) = \left(y_0 + \int_0^x e^{-t^2} t^3 dt \right) e^{x^2}.$$

Berechne das Integral:

$$\begin{aligned} \int_0^x e^{-t^2} t^3 dt &= \frac{1}{2} \int_0^{x^2} e^{-u} u du && \text{Substitution: } u = t^2, du = 2t dt \\ &= \frac{1}{2} \left(-e^{-u} u \Big|_0^{x^2} + \int_0^{x^2} e^{-u} du \right) && \text{partielle Integration} \\ &= -\frac{1}{2} e^{-x^2} x^2 - \frac{1}{2} e^{-x^2} + \frac{1}{2}. \end{aligned}$$

Also

$$\phi(x) = \left(y_0 + \frac{1}{2} \right) e^{x^2} - \frac{1}{2} (x^2 + 1).$$

Allgemein haben wir:

Satz 4.13 (Inhomogene lineare DGL erster Ordnung). *Seien $I \subseteq \mathbb{R}$ ein Intervall, $x_0 \in I$, $y_0 \in \mathbb{R}$ und $a, b: I \rightarrow \mathbb{R}$ stetig. Dann hat das AWP*

$$y' = a(x)y + b(x), \quad y(x_0) = y_0,$$

genau eine Lösung $\phi: I \rightarrow \mathbb{R}$. Diese lautet

$$\phi(x) = \left(y_0 + \int_{x_0}^x \phi_0(t)^{-1} b(t) dt \right) \cdot \phi_0(x),$$

wobei $\phi_0(x) := \exp \left(\int_{x_0}^x a(t) dt \right)$.

Die allgemeine Lösung ϕ_{inhom} der inhomogenen linearen DGL $y' = a(x)y + b(x)$ erhält man, indem man zur allgemeinen Lösung ϕ_{hom} der zugehörigen homogenen linearen DGL eine spezielle Lösung ϕ_s der inhomogenen linearen DGL addiert:

$$\phi_{\text{inhom}} = \phi_{\text{hom}} + \phi_s.$$

Beweis. Existenz: Nach obigem ist ϕ eine Lösung. Eindeutigkeit: Sei $\psi: I \rightarrow \mathbb{R}$ eine weitere Lösung. Nach dem Superpositionsprinzip 4.9 ist dann $\phi - \psi$ eine Lösung der homogenen linearen DGL $y' = a(x)y$ zur Anfangsbedingung $(\phi - \psi)(x_0) = y_0 - y_0 = 0$. Nach Satz 4.10 hat das AWP dieser homogenen linearen DGL aber als eindeutige Lösung die Nullfunktion. Also ist $\phi = \psi$. $\square$

Wir besprechen nun noch die Methode der *Substitution*, mit der man manchmal eine gegebene DGL in eine einfachere Form überführen kann.

Beispiel 4.14. Wir betrachten die DGL

$$y' = (x + y)^2 \quad (x, y \in \mathbb{R}).$$

Sie ist weder linear noch mit getrennten Variablen. Wir substituieren nun

$$z := x + y.$$

Dann gilt $z' = 1 + y'$, die DGL geht also über in

$$z' = 1 + z^2,$$

eine DGL mit getrennten Variablen. Diese können wir lösen:

$$\frac{dz}{dx} = 1 + z^2 \rightsquigarrow \frac{dz}{1 + z^2} = dx \rightsquigarrow \int \frac{1}{1 + z^2} dz = \int dx \rightsquigarrow \arctan(z) = x + c$$

mit einer Konstante c . Damit erhalten wir

$$z = \tan(x + c)$$

und wegen $z = x + y$

$$y = \tan(x + c) - x.$$

Geben wir z.B. die Anfangsbedingung $y(0) = 0$ vor, so erhalten wir als Kandidat für eine Lösung die Funktion

$$\phi(x) = \tan(x) - x$$

mit maximalem Definitionsintervall $(-\pi, \pi)$. Man kann wieder direkt überprüfen, dass ϕ tatsächlich eine Lösung der ursprünglichen DGL ist:

$$\phi'(x) = \tan'(x) - 1 = \frac{1}{\cos(x)^2} - 1 = \frac{1 - \cos(x)^2}{\cos(x)^2} = \tan(x)^2 = (x + \phi(x))^2.$$

Beispiel 4.15. Wir betrachten die DGL

$$y' = 1 + \frac{y}{x} + \left(\frac{y}{x}\right)^2 \quad (x > 0, y \in \mathbb{R}).$$

Substitution $z = \frac{y}{x}$ liefert $z' = \frac{y'}{x} - \frac{y}{x^2}$. Die DGL geht über in

$$z' = \frac{1}{x}(1 + z^2).$$

Diese können wir wieder mit Trennung der Variablen lösen:

$$\frac{dz}{1+z^2} = \frac{dx}{x} \rightsquigarrow \arctan(z) = \log(x) + c \rightsquigarrow z = \tan(\log(x) + c)$$

mit einer Konstanten c . Wegen $z = \frac{y}{x}$ hat die ursprüngliche DGL also eine Lösung der Form

$$\phi(x) = x \tan(\log(x) + c).$$

Geben wir eine Anfangsbedingung $y(x_0) = y_0$ vor, so erhalten wir

$$c = \arctan\left(\frac{y_0}{x_0}\right) - \log(x_0) \quad \text{und} \quad \phi(x) = x \tan\left(\log\left(\frac{x}{x_0}\right) + \arctan\left(\frac{y_0}{x_0}\right)\right).$$

Die DGL im letzten Beispiel ist eine so genannte *homogene DGL*. Für solche funktioniert die Substitution $z = \frac{y}{x}$ immer.

Definition 4.16. Eine *homogene Differenzialgleichung* ist eine DGL der Form

$$y' = f\left(\frac{y}{x}\right),$$

wobei $f: J \rightarrow \mathbb{R}$ eine stetige Funktion ist. Ihr Definitionsbereich ist $U = \{(x, y) \in \mathbb{R}^* \times \mathbb{R} : \frac{y}{x} \in J\}$.

Satz 4.17 (homogene DGL). *Seien J, f, U wie in Definition 4.16. Seien $I \subseteq \mathbb{R}^*$ ein Intervall und $(x_0, y_0) \in U$ mit $x_0 \in I$. Dann ist eine Funktion $\phi: I \rightarrow \mathbb{R}$ genau dann eine Lösung des AWP*

$$(4.5) \quad y' = f\left(\frac{y}{x}\right), \quad y(x_0) = y_0,$$

wenn die Funktion $\psi: I \rightarrow \mathbb{R}$, $\psi(x) := \frac{\phi(x)}{x}$, eine Lösung des AWP

$$(4.6) \quad z' = \frac{1}{x}(f(z) - z) \quad ((x, z) \in \mathbb{R}^* \times J), \quad z(x_0) = \frac{y_0}{x_0}$$

ist.

Beachte, dass (4.6) stets eine DGL mit getrennten Variablen ist.

Beweis. Man sieht direkt, dass der Graph von ϕ genau dann in U liegt, wenn der Graph von ψ in $\mathbb{R}^* \times J$ liegt.

“ $\implies$ ”: Sei ϕ eine Lösung des AWP (4.5). Dann gelten $\psi(x_0) = \frac{\phi(x_0)}{x_0} = \frac{y_0}{x_0}$ und

$$\begin{aligned}\psi'(x) &= \frac{\phi'(x)}{x} - \frac{\phi(x)}{x^2} = \frac{1}{x}(\phi'(x) - \frac{\phi(x)}{x}) = \frac{1}{x}(f(\frac{\phi(x)}{x}) - \frac{\phi(x)}{x}) \\ &= \frac{1}{x}(f(\psi(x)) - \psi(x)).\end{aligned}$$

“ $\impliedby$ ”: Sei ψ eine Lösung von (4.6). Dann gelten $\phi(x_0) = x_0 \cdot \psi(x_0) = x_0 \cdot \frac{y_0}{x_0} = y_0$ und

$$\begin{aligned}\phi'(x) &= \frac{d}{dx}(x\psi(x)) = \psi(x) + x\psi'(x) = \psi(x) + x \cdot \frac{1}{x}(f(\psi(x)) - \psi(x)) = f(\psi(x)) \\ &= f(\frac{\phi(x)}{x}).\end{aligned}$$

□

Der Existenz- und Eindeigkeitsatz

Bisher hatten wir für spezielle Typen von DGL gezeigt, dass sie zu vorgegebenen Anfangsbedingungen eindeutige Lösungen besitzen. Wie sieht es aber im Allgemeinen aus?

Beispiel 4.18. Wir betrachten die DGL

$$y' = y^{\frac{2}{3}} = \sqrt[3]{y^2} \quad ((x, y) \in \mathbb{R}^2).$$

Offensichtlich ist $\phi_0: \mathbb{R} \rightarrow \mathbb{R}$, $\phi_0(x) = 0$, eine Lösung der DGL zur Anfangsbedingung $\phi_0(0) = 0$.

Hierbei handelt es sich nicht um eine DGL mit getrennten Variablen im Sinne von Definition 4.4. Daher ist Satz 4.5 nicht anwendbar.

Gehen wir trotzdem wie bei der Methode der Trennung der Variablen vor, erhalten wir

$$\frac{dy}{y^{2/3}} = dx \rightsquigarrow 3y^{\frac{1}{3}} = x + c \rightsquigarrow y = \frac{1}{27}(x + c)^3$$

mit einer Konstanten c . Direktes Nachrechnen bestätigt, dass $\psi_c(x) = \frac{1}{27}(x + c)^3$ für jedes $c \in \mathbb{R}$ eine Lösung der DGL $y' = y^{\frac{2}{3}}$ ist. Insbesondere ist $\psi_0: \mathbb{R} \rightarrow \mathbb{R}$ eine zweite Lösung der DGL mit der Anfangsbedingung $\psi_0(0) = 0$.

Es ist sogar noch schlimmer: Für $a \leq 0 \leq b$ ist auch die Funktion

$$\psi_{a,b}(x) = \begin{cases} \frac{1}{27}(x - a)^3 & \text{für } x \leq a, \\ 0 & \text{für } a \leq x \leq b, \\ \frac{1}{27}(x - b)^3 & \text{für } b \leq x, \end{cases}$$

eine Lösung der DGL mit $\psi_{a,b}(0) = 0$.

Wir werden nun ein relativ allgemeines Kriterium angeben, das die Existenz und Eindeutigkeit von Lösungen eines AWP garantiert.

Definition 4.19. Seien $U \subseteq \mathbb{R} \times \mathbb{R}$ eine Teilmenge und $f: U \rightarrow \mathbb{R}$ eine Funktion. Wir sagen, f genüge in U *lokal einer Lipschitz-Bedingung*, wenn es zu jedem Punkt $(a, b) \in U$ eine Umgebung V und eine Konstante $L \geq 0$ gibt mit

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|$$

für alle $(x, y_1), (x, y_2) \in V$.

Bemerkung 4.20. Genügt $f: U \rightarrow \mathbb{R}$ lokal einer Lipschitzbedingung, so ist bei festem x insbesondere die Funktion $y \mapsto f(x, y)$ stetig (falls $L > 0$, wähle ' $\delta = \frac{\epsilon}{L}$ '). Die Umkehrung gilt aber nicht. Z.B. erfüllt die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = y^{\frac{2}{3}}$ nicht lokal eine Lipschitz-Bedingung.

In der Praxis genügen aber viele Funktionen lokal einer Lipschitz-Bedingung:

Satz 4.21. Sei $U \subseteq \mathbb{R}^2$ offen, und sei $f: U \rightarrow \mathbb{R}$ stetig partiell nach y differenzierbar. Dann genügt f lokal einer Lipschitz-Bedingung.

Beweis. Sei $(a, b) \in U$. Sei V eine kompakte Umgebung von (a, b) die ganz in U liegt (gibt's, da U offen ist). Nach Voraussetzung ist die Funktion $\frac{\partial f}{\partial y}: U \rightarrow \mathbb{R}$ stetig. Da V kompakt ist, ist sie auf V beschränkt, d.h. es gibt ein $L \geq 0$ mit

$$|\frac{\partial f}{\partial y}(x, y)| \leq L \text{ für alle } (x, y) \in V.$$

Dann gilt aber nach dem Mittelwertsatz für $(x, y_1), (x, y_2) \in V$

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|.$$

□

Satz 4.22 (Picard-Lindelöf). Sei $U \subseteq \mathbb{R}^2$ offen, und sei $f: U \rightarrow \mathbb{R}$ eine stetige Funktion, die lokal einer Lipschitz-Bedingung genügt. Dann gibt es zu jedem $(x_0, y_0) \in U$ eine eindeutige Lösung des AWP

$$y' = f(x, y), \quad y(x_0) = y_0,$$

d.h. also: Es gibt ein offenes Intervall $I \subseteq \mathbb{R}$, welches x_0 enthält, und eine Lösung $\phi: I \rightarrow \mathbb{R}$ des AWP (Existenz). Sind $\phi, \psi: I \rightarrow \mathbb{R}$ zwei Lösungen des AWP, so gilt $\phi = \psi$ (Eindeutigkeit).

Systeme von Differenzialgleichungen und Differenzialgleichungen höherer Ordnung

Bisher haben wir nur DGL untersucht, in denen die ‘Zeit’ x , die Funktion y und ihre Ableitung y' auftauchen. Ihre Lösungen waren reellwertige Funktionen. Allgemeiner begegnet man aber oft DGL, bei denen auch die zweite oder höhere Ableitungen auftreten, oder deren Lösungen vektorwertige Funktionen sind. Klassisches Beispiel aus der Physik: Beschreibt $y(x) \in \mathbb{R}^3$ den Ort eines Teilchens zur Zeit $x \in \mathbb{R}$, so ist $y'(x)$ die Geschwindigkeit und $y''(x)$ die Beschleunigung des Teilchens zur Zeit x . Wirkt auf das Teilchen eine Kraft F , die nur von der Zeit, dem Ort und der Geschwindigkeit des Teilchens abhängt, also $F(x, y, y') \in \mathbb{R}^3$, so gilt nach Newton mit der Masse m des Teilchens:

$$y''(x) = mF(x, y(x), y'(x)),$$

d.h. y ist eine Lösung des Systems von DGL zweiter Ordnung

$$\begin{aligned}y_1'' &= mF_1(x, y, y'), \\y_2'' &= mF_2(x, y, y'), \\y_3'' &= mF_3(x, y, y').\end{aligned}$$

Wir behandeln zunächst Systeme von DGL erster Ordnung.

Definition 4.23. Sei $U \subseteq \mathbb{R} \times \mathbb{R}^n$ eine Teilmenge. Wir bezeichnen die Koordinaten in $\mathbb{R} \times \mathbb{R}^n$ mit $(x, y) = (x, y_1, \dots, y_n)$. Sei $f: U \rightarrow \mathbb{R}^n$, $(x, y) \mapsto f(x, y)$, eine Funktion. Dann nennt man

$$(4.7) \quad y' = f(x, y)$$

ein *System* von n Differenzialgleichungen erster Ordnung. Eine *Lösung* des DGL-Systems ist eine auf einem Intervall $I \subseteq \mathbb{R}$ definierte differenzierbare vektorwertige Funktion $\phi: I \rightarrow \mathbb{R}^n$ (also eine differenzierbare Kurve) mit folgenden Eigenschaften:

(a) Der Graph von ϕ liegt ganz in U .

(b) Für alle $x \in I$ gilt

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

Hängt die Funktion f nicht von x ab, so spricht man von einem *autonomen System*.

Analog zu Definition 4.1 ist der Begriff des Anfangswertproblems definiert.

Ausgeschrieben ist (4.7) das System

$$\begin{aligned}y_1' &= f_1(x, y_1, \dots, y_n), \\y_2' &= f_2(x, y_1, \dots, y_n), \\&\vdots \\y_n' &= f_n(x, y_1, \dots, y_n).\end{aligned}$$

Beispiel 4.24. Betrachte das autonome System

$$\begin{aligned}y_1' &= -y_2, \\ y_2' &= y_1.\end{aligned}$$

Wir können dieses wieder im $\mathbb{R}^2$ veranschaulichen (siehe Abbildung 8): Dem Punkt $(y_1, y_2) \in \mathbb{R}^2$ wird der Vektor $f(y_1, y_2) = (-y_2, y_1) \in \mathbb{R}^2$ zugeordnet. Beachte, dass wir im Gegensatz zu den Bilder im letzten Abschnitt nun die ‘Zeitkoordinate’ x weglassen haben. Eine Lösung des autonomen Systems ist eine Kurve $\phi: I \rightarrow \mathbb{R}^2$, deren Tangentialvektor $\phi'(x)$ zur Zeit x die Bedingung $\phi'(x) = f(\phi(x))$ erfüllt.

Anhand des Bildes können wir raten, dass $\phi(x) = (\cos(x), \sin(x))$ eine Lösung ist.

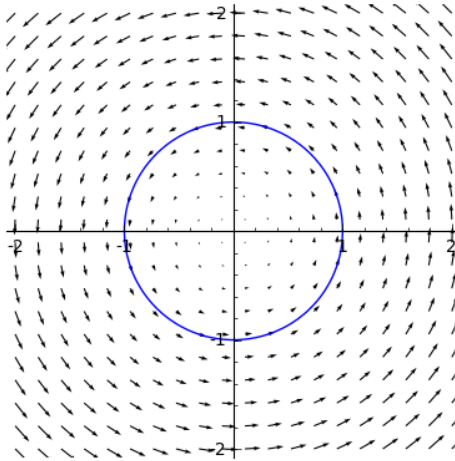


Abbildung 8: autonomes System mit Lösung

Manchmal kann man Systeme von DGL sukzessive lösen:

Beispiel 4.25. Betrachte das DGL-System

$$\begin{aligned}y_1' &= y_1 y_2, \\ y_2' &= x.\end{aligned}$$

Dann liefert die zweite DGL $y_2(x) = \frac{x^2}{2} + c$ mit einer Konstanten c . Eingesetzt in die erste DGL ergibt dies

$$y_1' = y_1 \left(\frac{x^2}{2} + c \right)$$

und diese können wir mit Trennung der Variablen lösen.

Bemerkung 4.26. Der Satz von Picard-Lindelöf 4.22 (Existenz- und Eindeutigkeit von Lösungen) gilt für DGL-Systeme ganz analog:

In der Definition 4.19 (lokale Lipschitz-Bedingung) ersetzt man $U \subseteq \mathbb{R} \times \mathbb{R}$ durch $U \subseteq \mathbb{R} \times \mathbb{R}^n$ und die Bedingung durch

$$\|f(x, y) - f(x, \tilde{y})\| \leq L \cdot \|y - \tilde{y}\|.$$

Ist $U \subseteq \mathbb{R} \times \mathbb{R}^n$ offen, $f: U \rightarrow \mathbb{R}^n$ stetig partiell nach $y_1, \dots, y_n$ differenzierbar, dann erfüllt f lokal eine Lipschitzbedingung.

Ist $U \subseteq \mathbb{R} \times \mathbb{R}^n$ und ist $f: U \rightarrow \mathbb{R}^n$ eine stetige Abbildung, die lokal einer Lipschitz-Bedingung genügt, dann ist das DGL-System

$$y' = f(x, y)$$

zu jeder vorgegebenen Anfangsbedingung eindeutig lösbar.

Als nächstes führen wir DGL höherer Ordnung ein.

Beispiel 4.27. Beispiel einer DGL zweiter Ordnung ist

$$y'' = y.$$

Diese besitzt als Lösung die Exponentialfunktion $\phi(x) = e^x$, da $\phi'(x) = \phi(x)$ und $\phi''(x) = \phi(x)$. Aber auch $\psi(x) = e^{-x}$ ist eine Lösung, da $\psi'(x) = -e^{-x}$, $\psi''(x) = e^{-x} = \psi(x)$. Genauso ist jede Linearkombination von ϕ und ψ eine Lösung der DGL.

Definition 4.28. Sei $U \subseteq \mathbb{R} \times \mathbb{R}^n$ eine Teilmenge und $f: U \rightarrow \mathbb{R}$ eine Funktion. Dann heißt

$$(4.8) \quad y^{(n)} = f(x, y, y', y'', \dots, y^{(n-1)})$$

eine (*gewöhnliche*) *Differenzialgleichung n-ter Ordnung*. Eine *Lösung* der DGL (4.8) ist eine auf einem Intervall $I \subseteq \mathbb{R}$ definierte, n -mal differenzierbare Funktion $\phi: I \rightarrow \mathbb{R}$ mit folgenden Eigenschaften:

(a) Die Menge

$$\{(x, \phi(x), \phi'(x), \dots, \phi^{(n-1)}(x)) \in \mathbb{R} \times \mathbb{R}^n : x \in I\}$$

liegt ganz in U .

(b) Für alle $x \in I$ gilt

$$\phi^{(n)}(x) = f(x, \phi(x), \phi'(x), \dots, \phi^{(n-1)}(x)).$$

DGL höherer Ordnung sind aber nur scheinbar allgemeiner als DGL(-Systeme) erster Ordnung:

Bemerkung 4.29 (Reduktion auf System von DGL 1. Ordnung). Gegeben sei die DGL n -ter Ordnung

$$(4.9) \quad y^{(n)} = f(x, y, y', \dots, y^{(n-1)}).$$

Wir führen nun neue Variable $y_0, \dots, y_{n-1}$ ein und ersetzen (4.9) durch das System

$$(4.10) \quad \begin{aligned} y_0' &= y_1, \\ y_1' &= y_2, \\ &\vdots \\ y_{n-2}' &= y_{n-1}, \\ y_{n-1}' &= f(x, y_0, y_1, \dots, y_{n-1}). \end{aligned}$$

Ist dann $\phi: I \rightarrow \mathbb{R}$ eine Lösung von (4.9), so ist

$$\Phi: I \rightarrow \mathbb{R}^n, \Phi(x) := \begin{pmatrix} \phi(x) \\ \phi'(x) \\ \vdots \\ \phi^{(n-1)}(x) \end{pmatrix},$$

eine Lösung von (4.10).

Ist Umgekehrt

$$\Psi: I \rightarrow \mathbb{R}^n, \Psi(x) = \begin{pmatrix} \psi_0(x) \\ \psi_1(x) \\ \vdots \\ \psi_{n-1}(x) \end{pmatrix}$$

eine Lösung von (4.10), so ist $\psi_0: I \rightarrow \mathbb{R}$ eine Lösung von (4.9), denn: Es gilt ja $\psi_0'(x) = \psi_1(x)$. Da ψ_1 differenzierbar ist, ist also ψ_0 sogar zweimal differenzierbar. Als nächstes gilt $\psi_1' = \psi_2$, also $\psi_0'' = \psi_2$ usw. Insgesamt erhalten wir, dass ψ_0 sogar n -mal differenzierbar ist und $\psi^{(n)}(x) = f(x, \psi(x), \psi'(x), \dots, \psi^{(n-1)}(x))$ gilt.

Beispiel 4.30. Betrachte die DGL 2. Ordnung

$$y'' = -y.$$

Zurückführen auf ein System von zwei DGL erster Ordnung liefert:

$$\begin{aligned} y_0' &= y_1 \\ y_1' &= -y_0. \end{aligned}$$

Dieses hat z.B. die Lösungen $\Phi(x) = \begin{pmatrix} \sin(x) \\ \cos(x) \end{pmatrix}$ und $\Psi(x) = \begin{pmatrix} \cos(x) \\ -\sin(x) \end{pmatrix}$. Die ursprüngliche DGL 2. Ordnung hat also z.B. die Lösungen $\phi(x) = \sin(x)$ und $\psi(x) = \cos(x)$.

Die Reduktion auf ein System von DGL 1. Ordnung liefert uns sofort den Existenz- und Eindeigkeitssatz für DGL höherer Ordnung:

Satz 4.31. *Sei $U \subseteq \mathbb{R} \times \mathbb{R}^n$ offen, und sei $f: U \rightarrow \mathbb{R}$ eine stetige Funktion, die lokal einer Lipschitz-Bedingung genügt. Dann gibt es zu jedem $(x_0, c_0, c_1, \dots, c_{n-1}) \in U$ ein offenes Intervall $I \ni x_0$ und eine eindeutige Lösung $\phi: I \rightarrow \mathbb{R}$ der DGL*

$$y^{(n)} = f(x, y, y', \dots, y^{(n-1)}),$$

welche die Anfangsbedingung

$$\phi(x_0) = c_0, \phi'(x_0) = c_1, \dots, \phi^{(n-1)}(x_0) = c_{n-1}$$

erfüllt.

Beispiel 4.32. Betrachte die DGL

$$y'' = y.$$

Wir hatten schon gesehen (Beispiel 4.27), dass $\phi_1: \mathbb{R} \rightarrow \mathbb{R}$, $\phi_1(x) = e^x$ und $\phi_{-1}: \mathbb{R} \rightarrow \mathbb{R}$, $\phi_{-1}(x) = e^{-x}$, und jede Linearkombination der beiden Lösungen der DGL sind. Seien $c_0, c_1 \in \mathbb{R}$ gegeben. Bestimme die Lösung ϕ der DGL zur Anfangsbedingung $y(0) = c_0$, $y'(0) = c_1$: Wegen

$$\phi_1(0) = \phi_{-1}(0) = 1$$

und

$$\phi_1'(0) = 1 \text{ und } \phi_{-1}'(0) = -1$$

erfüllt

$$\phi(x) = \frac{c_0 + c_1}{2}e^x + \frac{c_0 - c_1}{2}e^{-x}$$

das Verlangte. Mit dem Eindeutigkeitssatz folgt hieraus, dass die allgemeine Lösung der DGL $y'' = y$ durch

$$\phi_{a,b}: \mathbb{R} \rightarrow \mathbb{R}, \phi_{a,b}(x) = ae^x + be^{-x}, \quad a, b \in \mathbb{R}$$

gegeben ist.

Lineare Differenzialgleichungen mit konstanten Koeffizienten

Wir beschäftigen uns jetzt noch mit linearen Differenzialgleichungen n -ter Ordnung mit konstanten Koeffizienten. Diese kann man (mehr oder weniger) explizit lösen.

Es bietet sich an, hier auch komplexwertige Funktionen als Lösungen zu betrachten.

Definition 4.33. Eine *lineare DGL n -ter Ordnung mit konstanten Koeffizienten* ist eine DGL der Form

$$(4.11) \quad y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1y' + a_0y = b(x)$$

mit $a_0, \dots, a_{n-1} \in \mathbb{R}$ (oder $\mathbb{C}$) und einer stetigen Funktion $b: J \rightarrow \mathbb{R}$, $J \subseteq \mathbb{R}$ ein Intervall.

Sie heißt *homogen*, wenn $b = 0$, sonst *inhomogen*.

Eine Lösung von (4.11) auf einem Intervall $I \subseteq J$ ist eine n -mal differenzierbare⁵ Funktion $\phi: I \rightarrow \mathbb{C}$ mit

$$\phi^{(n)}(x) + a_{n-1}\phi^{(n-1)}(x) + \cdots + a_0\phi(x) = b(x) \quad \text{für alle } x \in I.$$

Das *charakteristische Polynom* der DGL (4.11) ist das Polynom

$$p(T) = T^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0 \in \mathbb{C}[T].$$

⁵wir identifizieren $\mathbb{C}$ mit $\mathbb{R}^2$; dann sollen die einzelnen Komponenten n -mal differenzierbar sein.

Bemerkung 4.34. Schreiben wir (4.11) in der Form

$$y^{(n)} = f(x, y, y', \dots, y^{(n-1)}),$$

so ist $f: J \times \mathbb{R}^n \rightarrow \mathbb{R}$ die Funktion

$$f(x, y_0, \dots, y_{n-1}) = b(x) - a_0 y_0 - a_1 y_1 - \dots - a_{n-1} y_{n-1}.$$

Diese erfüllt offenbar lokal (sogar global) eine Lipschitz-Bedingung (wähle $L = \max\{|a_i| : i = 0, \dots, n-1\}$). Daher ist (4.11) zu vorgegebener Anfangsbedingung

$$y(x_0) = c_0, y'(x_0) = c_1, \dots, y^{(n-1)}(x_0) = c_{n-1}$$

stets eindeutig lösbar.

Wir betrachten im Folgenden zunächst homogene lineare Differenzialgleichungen n -ter Ordnung der Form

$$(4.12) \quad y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_1y' + a_0y = 0$$

und fixieren ein offenes Intervall $I \subseteq \mathbb{R}$.

Bemerkung 4.35. Sind $\phi_1, \phi_2: I \rightarrow \mathbb{C}$ Lösungen von (4.12), so auch jede Linearkombination $\lambda_1\phi_1 + \lambda_2\phi_2$ mit $\lambda_1, \lambda_2 \in \mathbb{C}$ (vgl. Bemerkung 4.9). Das heißt: Für ein festes Intervall $I \subseteq \mathbb{R}$ ist

$$\mathcal{L}(I) := \{\phi: I \rightarrow \mathbb{C} : \phi \text{ ist Lösung von (4.12)}\}$$

ein $\mathbb{C}$ -Vektorraum. Man nennt ihn auch den *Lösungsraum* von (4.12).

Satz 4.36. Für den Lösungsraum $\mathcal{L}(I)$ von (4.12) gilt

$$\dim_{\mathbb{C}} \mathcal{L}(I) \leq n.$$

Beweis. Fixiere ein $x_0 \in I$. Die Abbildung

$$A: \mathcal{L}(I) \rightarrow \mathbb{C}^n, \quad A(\phi) = (\phi(x_0), \phi'(x_0), \dots, \phi^{(n-1)}(x_0)),$$

ist linear. Da eine Lösung ϕ durch die Anfangsbedingungen $(\phi(x_0), \phi'(x_0), \dots, \phi^{(n-1)}(x_0))$ eindeutig bestimmt ist (vgl. Bemerkung 4.34), ist A injektiv. Nach der Dimensionsformel (LA 3.22) ist also $\dim_{\mathbb{C}} \mathcal{L}(I) \leq \dim_{\mathbb{C}} \mathbb{C}^n = n$. $\square$

Definition 4.37. Unter einem *Fundamentalsystem* von Lösungen der Differenzialgleichung (4.12) auf dem Intervall I versteht man eine Basis des Lösungsraumes $\mathcal{L}(I)$.

Wir werden zeigen, dass in der Tat $\dim_{\mathbb{C}} \mathcal{L}(I) = n$ gilt, d.h. dass es n linear unabhängige Lösungen auf I gibt.

Satz 4.38. Sei $p(T) = T^n + a_{n-1}T^{n-1} + \dots + a_1T + a_0$ das charakteristische Polynom von (4.12). Seien $\lambda_1, \dots, \lambda_r \in \mathbb{C}$ die Nullstellen von $p(T)$ mit Vielfachheiten $k_1, \dots, k_r$, d.h.

$$p(T) = (T - \lambda_1)^{k_1} \dots (T - \lambda_r)^{k_r}.$$

Dann bilden die Funktionen

$$\phi_{ij}: I \rightarrow \mathbb{C}, \phi_{ij}(x) = x^i e^{\lambda_j x}, \quad j = 1, \dots, r, \quad i = 0, \dots, k_j - 1,$$

ein Fundamentalsystem von Lösungen der DGL (4.12).

Inbesondere gilt

$$\dim_{\mathbb{C}} \mathcal{L}(I) = k_1 + \dots + k_r = n.$$

Für den Beweis des Satzes ist es nützlich mit so genannten Differentialoperatoren zu arbeiten:

Setze $D := \frac{d}{dx}$. Dann ist D eine Abbildung, die einer Funktion $f: I \rightarrow \mathbb{C}$ ihre Ableitung $D(f) = f'$ zuordnet. Sie ist linear. Für $q \in \mathbb{N}$ setzen wir $D^q = D \circ \dots \circ D$ (q mal), also $D^q(f) = f^{(q)}$, und $D^0 := \text{id}$, d.h. $D^0(f) = f$.

Ist $p(T) = a_n T^n + a_{n-1} T^{n-1} + \dots + a_1 T + a_0$ ein Polynom, so setzen wir

$$p(D) = a_n \cdot D^n + \dots + a_1 \cdot D + a_0 \cdot \text{id}.$$

Dann ist $p(D)$ eine Abbildung, die einer (n -mal differenzierbaren) Funktion $f: I \rightarrow \mathbb{C}$ die Funktion

$$p(D)(f) = a_n f^{(n)} + \dots + a_1 f' + a_0 f$$

zuordnet. Man nennt $p(D)$ einen (linearen) Differenzialoperator mit konstanten Koeffizienten.

Beispiel 4.39. Die DGL (4.11) schreibt sich damit als

$$p(D)y = b(x)$$

und (4.12) als

$$p(D)y = 0,$$

wobei p jeweils das charakteristische Polynom der DGL ist.

Beispiel 4.40.

- (a) Seien $p(T)$ ein Polynom und $\lambda \in \mathbb{C}$. Wegen $D(e^{\lambda x}) = \lambda e^{\lambda x}$ und $D^q(e^{\lambda x}) = \lambda^q e^{\lambda x}$ ergibt sich

$$p(D)(e^{\lambda x}) = p(\lambda)e^{\lambda x}.$$

- (b) Ist $f: I \rightarrow \mathbb{C}$ eine differenzierbare Funktion, so gilt

$$(D - \lambda \cdot \text{id})(f(x)e^{\lambda x}) = f'(x)e^{\lambda x} + \lambda \cdot f(x)e^{\lambda x} - \lambda \cdot f(x)e^{\lambda x} = f'(x)e^{\lambda x},$$

und allgemeiner

$$(D - \lambda \cdot \text{id})^q(f(x)e^{\lambda x}) = f^{(q)}(x)e^{\lambda x}.$$

(c) Ist f ein Polynom vom Grad $\deg(f) < q$, so folgt insbesondere

$$(D - \lambda \cdot \text{id})^q(f(x)e^{\lambda x}) = 0.$$

Für den Beweis des Satzes benötigen wir:

Lemma 4.41. *Sind p_1, p_2 zwei Polynome, so gilt*

$$(p_1 \cdot p_2)(D) = p_1(D) \circ p_2(D) = p_2(D) \circ p_1(D).$$

Beweis. Sei $p_2(T) = b_m T^m + \dots + b_1 T + b_0$. Da sich über $\mathbb{C}$ jedes Polynom in Linearfaktoren zerlegen lässt, genügt es, die Behauptung für den Fall $p_1(T) = T - a$ mit $a \in \mathbb{C}$ zu zeigen.

Dann gilt $(p_1 \cdot p_2)(T) = b_m T^{m+1} + (b_{m-1} - ab_m)T^m + \dots + (b_0 - ab_1)T - a$, also

$$(p_1 \cdot p_2)(D)(f) = b_m f^{(m+1)} + (b_{m-1} - ab_m) f^{(m)} + \dots + (b_0 - ab_1) f' - af.$$

Andrerseits gilt

$$\begin{aligned} p_1(D) \circ p_2(D)(f) &= p_1(D) \left(\sum_{i=0}^m b_i f^{(i)} \right) = \left(\sum_{i=0}^m b_i f^{(i)} \right)' - a \cdot \left(\sum_{i=0}^m b_i f^{(i)} \right) \\ &= \left(\sum_{i=0}^m b_i f^{(i+1)} \right) - a \cdot \left(\sum_{i=0}^m b_i f^{(i)} \right) \\ &= b_m f^{(m+1)} + (b_{m-1} - ab_m) f^{(m)} + \dots + (b_0 - ab_1) f' - af \\ &= (p_1 \cdot p_2)(D)(f). \end{aligned}$$

□

Beweis von Satz 4.38. Wir zeigen als erstes, dass die angegebenen Funktionen $\phi_{ij}(x) = x^i e^{\lambda_j x}$, $i = 0, \dots, k_j - 1$, tatsächlich Lösungen der DGL (4.12) sind. Da λ_j Nullstelle des charakteristischen Polynoms $p(T)$ mit Vielfachheit k_j ist, gibt es ein Polynom $q(T)$ so, dass $p(T) = q(T) \cdot (T - \lambda_j)^{k_j}$ gilt. Nach Lemma 4.41 gilt also

$$p(D) = q(D) \circ (D - \lambda_j \cdot \text{id})^{k_j}.$$

Nach Beispiel 4.40(c) gilt aber

$$(D - \lambda_j \cdot \text{id})^{k_j}(x^i e^{\lambda_j x}) = 0,$$

also folgt auch $p(D)(x^i e^{\lambda_j x}) = 0$, d.h. ϕ_{ij} ist eine Lösung der DGL.

Nun zeigen wir die lineare Unabhängigkeit. Angenommen, $\sum_{j=1}^r \sum_{i=0}^{k_j-1} c_{ij} \phi_{ij}(x) = 0$ für alle $x \in I$. Setze $f_j(x) = \sum_{i=0}^{k_j-1} c_{ij} x^i$. Dann sind die f_j Polynome und es gilt

$$f_1(x)e^{\lambda_1 x} + \dots + f_r(x)e^{\lambda_r x} = 0 \text{ für alle } x \in I.$$

Das folgende Lemma 4.42 liefert dann $f_1 = \dots = f_r = 0$, also $c_{ij} = 0$ für alle i, j . □

Lemma 4.42. Seien $\lambda_1, \dots, \lambda_r \in \mathbb{C}$ paarweise verschieden. Sind $f_1, \dots, f_r \in \mathbb{C}[x]$ Polynome mit der Eigenschaft, dass

$$f_1(x)e^{\lambda_1 x} + \dots + f_r(x)e^{\lambda_r x} = 0 \text{ für alle } x \in I$$

gilt, dann gilt schon $f_1 = \dots = f_r = 0$.

Beweis. Induktion über r .

Induktionsanfang $r = 1$: Sei $f_1(x)e^{\lambda_1 x} = 0$ für alle $x \in I$. Da $e^{\lambda_1 x} \neq 0$, folgt $f_1(x) = 0$ für alle $x \in I$. Da ein Polynom nur endlich viele Nullstellen hat, folgt $f_1 = 0$.

Induktionsschritt $r - 1 \rightsquigarrow r$: Sei die Behauptung für $r - 1$ Polynome bereits bewiesen (Induktionsannahme) und gelte

$$f_1(x)e^{\lambda_1 x} + \dots + f_r(x)e^{\lambda_r x} = 0 \text{ für alle } x \in I.$$

Dann folgt

$$(4.13) \quad f_1(x) + f_2(x)e^{(\lambda_2 - \lambda_1)x} + \dots + f_r(x)e^{(\lambda_r - \lambda_1)x} = 0 \text{ für alle } x \in I.$$

Sei d der Grad des Polynoms f_1 . Leiten wir Gleichung (4.13) $d + 1$ -mal nach x ab, so erhalten wir eine Gleichung der Form

$$F_2(x)e^{(\lambda_2 - \lambda_1)x} + \dots + F_r(x)e^{(\lambda_r - \lambda_1)x} = 0 \text{ für alle } x \in I$$

mit Polynomen $F_2, \dots, F_r$. Nach Induktionsannahme folgt also $F_2 = \dots = F_r = 0$.

Andrerseits ergibt sich für den führenden Koeffizienten von F_2 nach der Produktregel

$$(\text{führender Koeff. von } F_2) = (\lambda_2 - \lambda_1)^{d+1} \cdot (\text{führender Koeff. von } f_2)$$

und analog für die anderen F_j . Da die λ_j paarweise verschieden sind, folgt $\deg(F_j) = \deg(f_j)$ für $j = 2, \dots, r$. Also folgt aus $F_j = 0$ auch $f_j = 0$ für $j = 2, \dots, r$. $\square$

In der Praxis ist man meist an reellen Lösungen interessiert.

Bemerkung 4.43. Seien die Koeffizienten $a_0, \dots, a_{n-1}$ in der DGL (4.12) reell. Ist $\lambda = \mu + \nu i \in \mathbb{C}$ eine Nullstelle des charakteristischen Polynoms der DGL, so ist auch $\bar{\lambda} = \mu - \nu i$ eine Nullstelle. Dann können wir im Fundamentalsystem aus Satz 4.38 die beiden $\mathbb{C}$ -wertigen Lösungen $x^i e^{\lambda x}$ und $x^i e^{\bar{\lambda} x}$ durch die beiden reellen Lösungen

$$\begin{aligned} \operatorname{Re}(x^i e^{\lambda x}) &= x^i e^{\mu x} \cos(\nu x) = \frac{1}{2} \left(x^i e^{\lambda x} + x^i e^{\bar{\lambda} x} \right) \text{ und} \\ \operatorname{Im}(x^i e^{\lambda x}) &= x^i e^{\mu x} \sin(\nu x) = \frac{1}{2i} \left(x^i e^{\lambda x} - x^i e^{\bar{\lambda} x} \right) \end{aligned}$$

ersetzen.

Beispiel 4.44. Die DGL

$$y'' = -y$$

lässt sich schreiben als $y'' + y = 0$, hat also das charakteristische Polynom $T^2 + 1$ mit den beiden Nullstellen $\pm i$. Ein Fundamentalsystem von komplexen Lösungen ist also

$$\phi_1(x) = e^{ix}, \quad \phi_2(x) = e^{-ix},$$

ein Fundamentalsystem von reellen Lösungen ist

$$\psi_1(x) = \cos(x), \quad \psi_2(x) = \sin(x).$$

Beispiel 4.45. Die DGL

$$y''' - y'' - 2y' = 0$$

hat das charakteristische Polynom $p(T) = T^3 - T^2 - 2T$. Dieses hat die Nullstellen $\lambda_1 = 0$, $\lambda_2 = 2$, $\lambda_3 = -1$. Also erhalten wir als Fundamentalsystem von Lösungen

$$\phi_1(x) = 1, \quad \phi_2(x) = e^{2x}, \quad \phi_3(x) = e^{-x} \quad (x \in \mathbb{R}).$$

Die allgemeine (reelle) Lösung der DGL hat also die Form

$$\phi(x) = c_1 + c_2 e^{2x} + c_3 e^{-x}, \quad c_1, c_2, c_3 \in \mathbb{R}.$$

Suchen wir die Lösung zu einer gegebenen Anfangsbedingung, z.B. $\phi(0) = 2$, $\phi'(0) = 3$, $\phi''(0) = 3$, so gehen wir wie folgt vor: Für die allgemeine Lösung gilt

$$\phi'(x) = 2c_2 e^{2x} - c_3 e^{-x}$$

$$\phi''(x) = 4c_2 e^{2x} + c_3 e^{-x}.$$

Die Anfangsbedingung liefert also das Gleichungssystem

$$c_1 + c_2 + c_3 = 2$$

$$2c_2 - c_3 = 3$$

$$4c_2 + c_3 = 3$$

Dieses besitzt die Lösung $c_1 = 2$, $c_2 = 1$, $c_3 = -1$. Die gesuchte Lösung der DGL ist also

$$2 + e^{2x} - e^{-x}.$$

Beispiel 4.46. Die DGL

$$y''' - 5y'' + 8y' - 6y = 0$$

hat das charakteristische Polynom $p(T) = T^3 - 5T^2 + 8T - 6$ durch Probieren findet man zunächst die Nullstelle $\lambda_1 = 3$. Polynomdivision liefert $p(T) = (T - 3) \cdot (T^2 - 2T + 2)$. Damit ergeben sich die beiden weiteren Nullstellen $\lambda_2 = 1 + i$ und $\lambda_3 = 1 - i$. Ein Fundamentalsystem komplexer Lösungen ist daher

$$e^{3x}, e^{(1+i)x}, e^{(1-i)x},$$

ein Fundamentalsystem reeller Lösungen ist

$$e^{3x}, e^x \cos(x), e^x \sin(x).$$