Mathematik für Ingenieure II

(für Informatiker)

Hans Grabmüller
Institut für Angewandte Mathematik

Vorlesung im Sommersemester 2000

Friedrich–Alexander–Universität Erlangen–Nürnberg

Inhaltsverzeichnis

Kapitel 6

Funktionen einer reellen Veränderlichen

6.1 Der Funktionsbegriff, Beispiele

Dieser zentrale Begriff der Analysis ist das angemessene Hilfsmittel, um die *Abhängigkeit gewisser Größen von anderen* zu beschreiben. Wir haben den Funktionsbegriff bereits in Abschnitt 1.4 auf den Mengenbegriff der *Korrespondenz* zurückgeführt. Wir wollen hier jedoch die folgende **Zuordnungsdefinition** bevorzugen (es sei zum Vergleich auch auf Abschnitt 5.1 verwiesen):

Definition 6.1 (a) *Es seien X, Y nichtleere Mengen. Eine Vorschrift f mit der Eigenschaft*

$$\boxed{\forall\, x \in X\ \exists!\, y \in Y :\ y = f(x)} \tag{A}$$

heiße **Abbildung** (*oder* **Funktion** *oder* **Zuordnung**) *von X in Y.*
(b) *Das Element $y = f(x)$ heiße* **Bild** *von x unter f, und x heiße ein* **Urbild** *von $f(x)$.*
(c) *Die Menge X heiße* **Definitionsbereich** *der Funktion f, häufig mit $D(f)$ bezeichnet. Die Menge Y heiße die* **Zielmenge** *von f. Die Menge $f(X)$ heiße* **Bildbereich** *oder* **Wertebereich** *von f, den wir in Kapitel 5 (bei den linearen Abbildungen) mit* Bild f *bezeichnet haben.*

Erst die Angaben von *Definitionsbereich, Zielmenge und Abbildungsvorschrift* legen eine Funktion f eindeutig fest. Diese drei Erfordernisse werden in der Symbolik

$$f : \begin{cases} X \to Y, \\ x \mapsto f(x), \end{cases} \qquad \textit{zum Beispiel} \quad f : \begin{cases} [0, +\infty) \to \mathbf{R}, \\ x \mapsto \sqrt{x}, \end{cases}$$

vereint zum Ausdruck gebracht. Oftmals genügt auch die vereinfachende Symbolik

$$f : X \to Y \quad \text{oder} \quad X \xrightarrow{f} Y \quad \text{oder} \quad x \mapsto f(x),\ x \in X.$$

Definition 6.2 (a) *Zwei Funktionen $f : X \to Y$ und $g : M \to N$ heißen* **gleich**, *wenn gilt:*

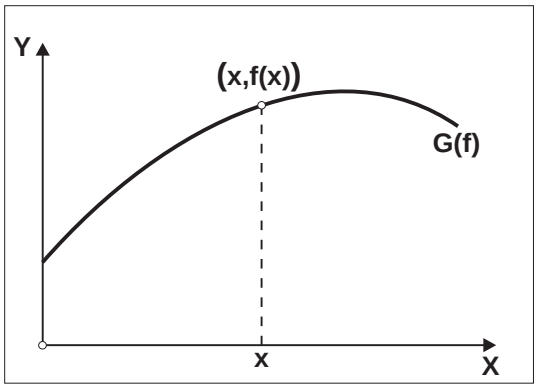
$$\text{(i) } X = M, \qquad \text{(ii) } f(x) = g(x)\ \ \forall\, x \in X.$$

Die Gleichheit der Zielmengen ist nicht unbedingt entscheidend.
(b) *Der* **Funktionenraum**

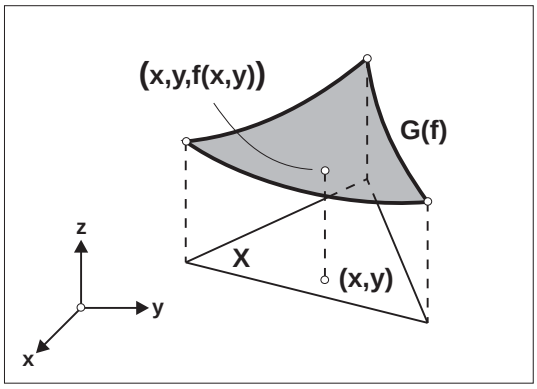
$$\boxed{\text{Abb}\,(X, Y) := \{f : X \to Y\ :\ f \textit{ ist Funktion}\}}$$

ist die Menge aller Funktionen von X in Y.

Die Funktionsbeziehung $x \mapsto f(x)$ liefert unter anderem eine Information über die Änderung des Funktionswertes $f(x)$, wenn der Wert x geändert wird. Unter diesem Gesichtspunkt heiße x die **unabhängige Veränderliche** (oder die **unabhängige Variable**), und $y = f(x)$ die **abhängige Veränderliche** (oder die **abhängige Variable**). Gelten $X, Y \subset \mathbf{R}$, so veranschaulicht man häufig diese Funktionsbeziehung durch eine *Kurve*: Man zeichnet den *Graphen* der Funktion f.



Der Graph einer Funktion der reellen Variablen x



Der Graph einer Funktion der reellen Variablen (x, y)

Definition 6.3 *Die von einer Funktion $f : X \to Y$ erzeugte Teilmenge*

$$\boxed{G(f) := \{(x, f(x)) \,:\, x \in X\} \subset X \times Y}$$

heiße der **Graph** *von f.*

Die obigen Skizzen zeigen, dass der Graph einer Funktion f im Falle $X, Y \subset \mathbf{R}$ in der Zeichenebene durch eine Kurve veranschaulicht werden kann. Das Hilfsmittel der Veranschaulichung von $G(f)$ ist auch noch dann anwendbar, wenn $X \subset \mathbf{R}^2$ und $Y \subset \mathbf{R}$ gelten.

Dem Leser wird empfohlen, sich an dieser Stelle nochmals die überaus wichtigen Begriffe der Surjektivität, Injektivität und Bijektivität für $f \in \mathrm{Abb}\,(X, Y)$ aus Abschnitt 5.1 in Erinnerung zu rufen. Darüber hinaus sollten ihm auch die folgenden Begriffe vertraut sein: Hintereinanderausführung oder Kompositum, Rechts–, Linksinverse, Inverse oder Umkehrabbildung oder Umkehrfunktion, Einschränkung oder Restriktion, Fortsetzung oder Erweiterung. Wir verweisen hier auf die Abschnitte 1.4 und 5.1.

Definition 6.4 *Eine Abbildung $f : X \to Y$ heiße* **Funktion einer reellen Veränderlichen** *x: $x \mapsto f(x)$, wenn der Definitionsbereich $X = D(f) \subset \mathbf{R}$ aus einem Intervall oder einer (nicht notwendig endlichen) Vereinigung von Intervallen besteht.*

Wir geben nun eine Reihe gebräuchlicher Funktionen einer reellen Veränderlichen x an.

BSP. (6.1.1) **Die lineare Funktion:** Das ist die Funktion

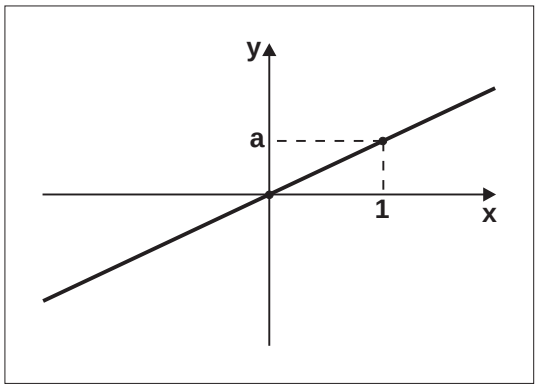
$$\boxed{f(x) := a\,x \;\; \forall\, x \in \mathbf{R}, \quad a \in \mathbf{R} \;\text{ fest.}}$$

Es gilt hier offensichtlich

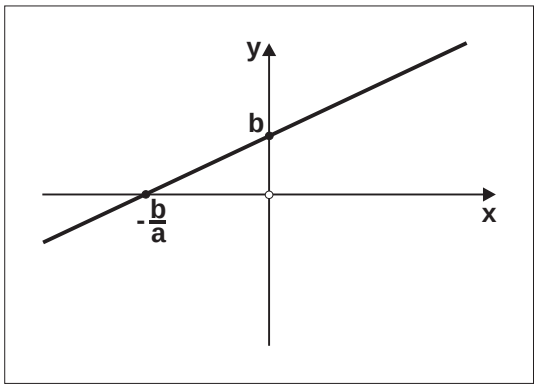
$$D(f) = \mathbf{R}, \quad \mathrm{Bild}\, f = \begin{cases} \mathbf{R} & : \; a \neq 0, \\ \{0\} & : \; a = 0. \end{cases}$$

Die lineare Funktion enthält die folgenden Spezialfälle:

(i) **Die identische Funktion** $f(x) := x$. (ii) **Die triviale Funktion** $f(x) := 0 \;\forall\, x \in \mathbf{R}$.



Die lineare Funktion $f(x) := a\,x$



Die affine Funktion $f(x) := a\,x + b$

BSP. (6.1.2) **Die affine Funktion:** Das ist die Funktion

$$\boxed{f(x) := a\,x + b \;\; \forall\, x \in \mathbf{R}, \quad a, b \in \mathbf{R} \;\text{ fest}, \;\; b \neq 0.}$$

Es gilt hier offensichtlich

$$D(f) = \mathbf{R}, \quad \text{Bild}\, f = \begin{cases} \mathbf{R} & : \; a \neq 0, \\ \{b\} & : \; a = 0. \end{cases}$$

Die affine Funktion enthält den folgenden Spezialfall:

$$\boxed{\textbf{Die konstante Funktion } f(x) := b \;\forall\, x \in \mathbf{R}.}$$

BSP. (6.1.3) **Die ganzrationale Funktion:** Das ist die Funktion

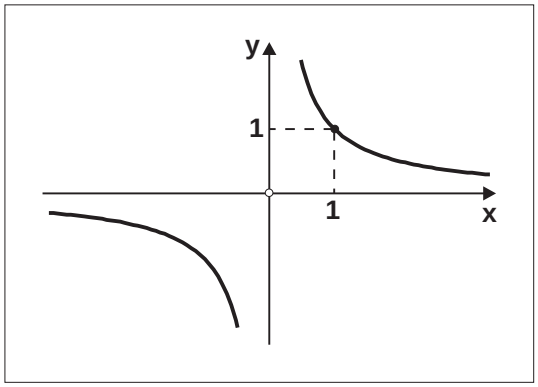
$$\boxed{f(x) := \sum_{k=0}^{n} a_k x^k \;\; \forall\, x \in \mathbf{R}, \quad a_0, a_1, \ldots, a_n \in \mathbf{R} \;\text{ fest}.}$$

Es gilt hier wiederum $D(f) = \mathbf{R}$, während sich Bild f im allgemeinen Fall nicht einfach spezifizieren lässt. Für $a_n \neq 0$ ist $f(x) = P_n(x)$ ein **Polynom** vom Grade n.

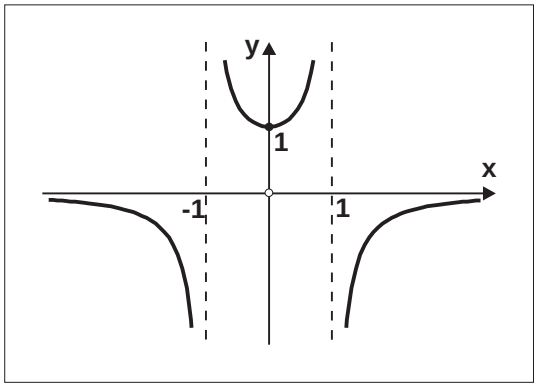
BSP. (6.1.4) **Die gebrochen rationale Funktion:** Das ist die Funktion

$$\boxed{f(x) := \frac{P_n(x)}{Q_m(x)} = \frac{a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0}{b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0}, \quad D(f) := \{x \in \mathbf{R} \,:\, Q_m(x) \neq 0\}.}$$

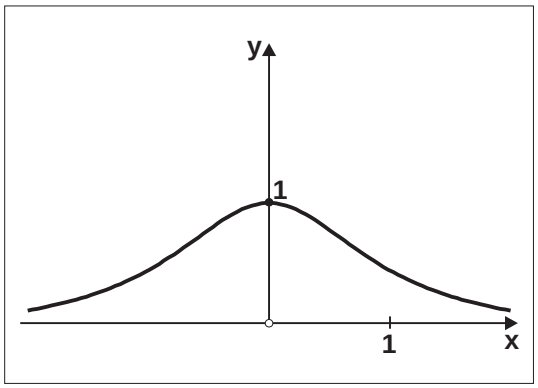
Hierin sind $a_0, a_1, \ldots, a_n \in \mathbf{R}$ und $b_0, b_1, \ldots, b_m \in \mathbf{R}$ fest vorgegeben.



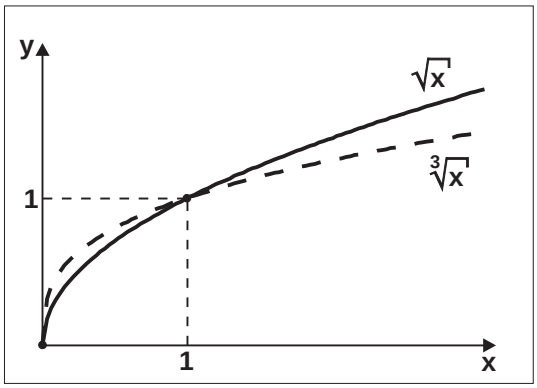
Die Funktion $f(x) := \frac{1}{x}$
mit $D(f) = \mathbf{R} \setminus \{0\} = \text{Bild}\, f$



Die Funktion $f(x) := \frac{1}{1-x^2}$ **mit**
$D(f) = \mathbf{R} \setminus \{-1, 1\}$, **Bild** $f = \mathbf{R} \setminus [0, 1)$



Die Funktion $f(x) := \frac{1}{1+x^2}$ mit $D(f) = \mathbf{R}$ und Bild $f = (0, 1]$



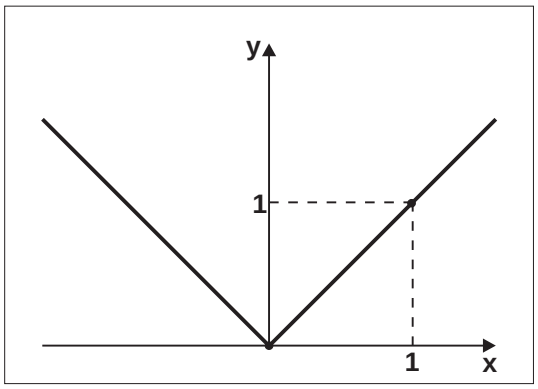
Die n–te Wurzelfunktion mit $D(f) = [0, +\infty) = \text{Bild}\, f$

BSP. (6.1.5) **Die n–te Wurzelfunktion:** Das ist die Funktion

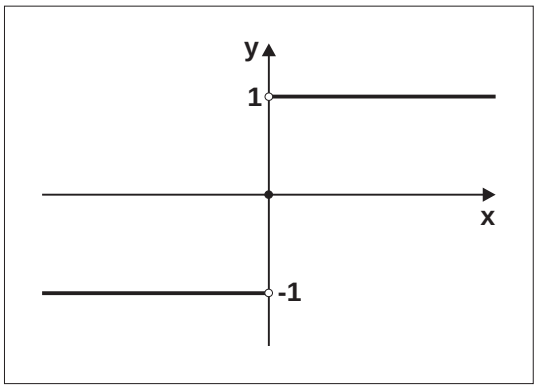
$$f(x) := \sqrt[n]{x} \quad \forall\, x \geq 0 \quad n \in \mathbf{N} \text{ fest.}$$

BSP. (6.1.6) **Der Absolutbetrag:** Das ist die Funktion

$$f(x) := |x| \quad \forall\, x \in \mathbf{R} \text{ mit } D(f) = \mathbf{R}, \quad \text{Bild}\, f = [0, +\infty).$$



Der Absolutbetrag



Die Signumsfunktion

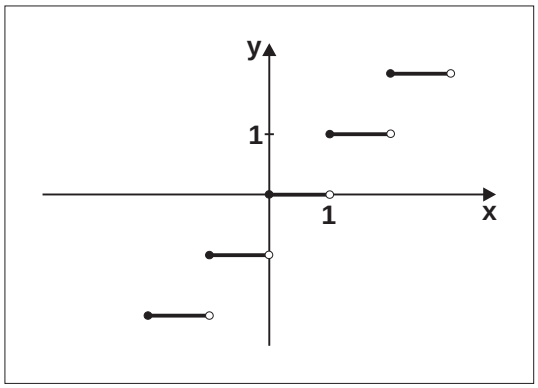
BSP. (6.1.7) **Die Signumsfunktion:** Das ist die Funktion

$$f(x) := \operatorname{sign} x = \begin{cases} 1 & : \ x > 0, \\ 0 & : \ x = 0, \\ -1 & : \ x < 0 \end{cases} \quad \text{mit } D(f) = \mathbf{R}, \quad \text{Bild}\, f = \{-1, 0, 1\}.$$

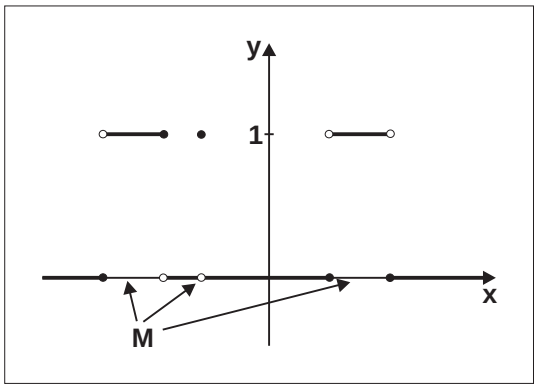
BSP. (6.1.8) **Die Entire–Funktion:** Das ist die Funktion

$$f(x) := [x] \text{ mit } D(f) = \mathbf{R}, \quad \text{Bild}\, f = \mathbf{Z}.$$

Hierbei bezeichnet $[x]$ die größte ganze Zahl $p \in \mathbf{Z}$ mit $p \leq x$.



Die Entire–Funktion



Charakteristische Funktion der Menge M

BSP. (6.1.9) **Die charakteristische Funktion** einer nichtleeren Teilmenge $M \subset \mathbf{R}$: Das ist die Funktion

$$f(x) := \chi_M(x) = \begin{cases} 1 \ : \ x \in M, \\ 0 \ : \ x \notin M \end{cases} \quad \text{mit} \ \ D(f) = \mathbf{R}, \ \ \text{Bild}\, f = \{0,1\}.$$

BSP. (6.1.10) **Die** DIRICHLET**–Funktion:** Das ist die charakteristische Funktion der Menge $\mathbf{Q}$ der rationalen Zahlen

$$f(x) := \chi_{\mathbf{Q}}(x) = \begin{cases} 1 \ : \ x \ \text{rational}, \\ 0 \ : \ x \ \text{irrational} \end{cases} \quad \text{mit} \ \ D(f) = \mathbf{R}, \ \ \text{Bild}\, f = \{0,1\}.$$

BSP. (6.1.11) **Die stückweise konstante Funktion (Treppenfunktion):**

Definition 6.5 *Es sei* $\emptyset \neq I \subset \mathbf{R}$ *ein endliches Intervall mit Randpunkten* a, b, $a < b$. *Eine Familie* $Z_n := \{I_1, I_2, \ldots, I_n\}$ *von Teilintervallen* $I_j \subset I$ *heiße eine* **endliche Zerlegung** (*kurz: Zerlegung*) *von* I, *wenn gilt:*

$$\text{(i)} \qquad a =: x_0 \leq x_1 \leq x_2 \leq \cdots \leq x_{n-1} \leq x_n := b, \tag{1.1}$$

$$\text{(ii)} \qquad I_j \ \textit{hat die Randpunkte} \ x_{j-1} \ \textit{und} \ x_j, \quad I_j \neq \emptyset \ \ \forall\, j = 1, 2, \ldots, n, \tag{1.2}$$

$$\text{(iii)} \qquad I_j \cap I_k = \emptyset, \ j \neq k, \quad \bigcup_{j=1}^{n} I_j = I. \tag{1.3}$$

Eine Funktion $f : I \to \mathbf{R}$ *heiße eine* **Treppenfunktion** *bezüglich der Zerlegung* Z_n, *wenn gilt:*

$$f(x) := \sum_{j=1}^{n} c_j \chi_{I_j}(x) \ \ \forall\, x \in I, \quad c_1, c_2, \ldots, c_n \in \mathbf{R} \ \ \textit{fest.}$$

Eine Treppenfunktion $f(x)$ hat mit anderen Worten auf dem Teilintervall I_j den konstanten Wert c_j. Es folgen $D(f) = I$ und $\text{Bild}\, f = \{c_1, c_2, \ldots, c_n\}$. Durch Hinzunahme von

$$f(x) := \begin{cases} c_{-\infty} \ \forall\, x \in (-\infty, a), \\ c_{+\infty} \ \forall\, x \in (b, +\infty), \end{cases}$$

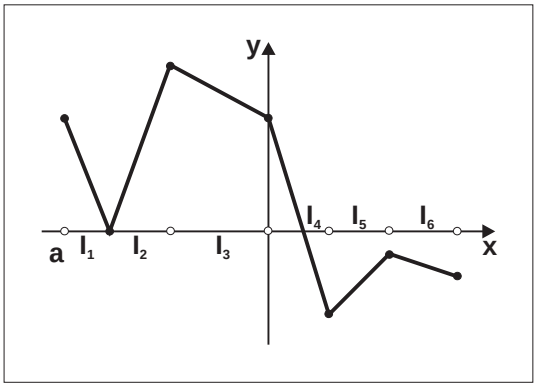
kann der Definitionsbereich $D(f)$ auf ganz $\mathbf{R}$ erweitert werden. *Zahlenbeispiel:*

$$I := (-3, 2], \quad f(x) := \begin{cases} 1.5 \ : \ x \in (-3, -1), \\ -1 \ : \ x \in [-1, \sqrt{2}), \\ \pi \ : \ x \in (\sqrt{2}, 2]. \end{cases}$$

BSP. (6.1.12) **Die stückweise affine Funktion:** Es seien I und Z_n wie in BSP. (6.1.11) erklärt. Wir setzen

$$\boxed{f(x) := \sum_{j=1}^{n} (c_j x + d_j)\, \chi_{I_j}(x) \;\; \forall\, x \in I, \quad c_j, d_j \in \mathbf{R} \;\text{ fest.}}$$

Dann ist f auf jedem Teilintervall I_j die affine Funktion $f(x) = c_j x + d_j$. Schließen sich die affinen Teilstücke in den Randpunkten der Intervalle *ohne Sprung*, so heiße f ein **Polygonzug** oder ein **linearer Spline**.



Die stückweise affine Funktion (Polygonzug)

BSP. (6.1.13) **Vektorwertige Funktionen:** Es seien

$$x_j : \begin{cases} D(x_j) \to \mathbf{R}, \\ t \mapsto x_j(t), \end{cases} \qquad j = 1, 2, \ldots, n,$$

Funktionen mit gemeinsamem Definitionsbereich $\emptyset \neq D := \bigcap_{j=1}^{n} D(x_j) \subset \mathbf{R}$. Dann ist eine **vektorwertige Funktion** $\vec{x} : D \to \mathbf{R}^n$ durch die folgende Vorschrift erklärt:

$$\boxed{\vec{x}(t) := \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \;\; \forall\, t \in D \;\text{ mit }\; D(\vec{x}) = D, \;\; \text{Bild}\, \vec{x} \subset \mathbf{R}^n.}$$

Zum Beispiel: $\vec{x}(t) := \binom{\cos t}{\sin t} = (\cos t, \sin t)^T \;\forall\, t \in [0, 2\pi]$.

BSP. (6.1.14) **Matrixwertige Funktionen:** Es seien

$$a_{jk} : \begin{cases} D(a_{jk}) \to \mathbf{R}, \\ t \mapsto a_{jk}(t), \end{cases} \qquad j = 1, 2, \ldots, m, \;\; k = 1, 2, \ldots, n,$$

Funktionen mit gemeinsamem Definitionsbereich $\emptyset \neq D := \bigcap_{\substack{j=1,\ldots,m \\ k=1,\ldots,n}} D(a_{jk}) \subset \mathbf{R}$. Dann ist eine **matrixwertige Funktion** $A : D \to \mathbf{R}^{(m,n)}$ durch die folgende Vorschrift erklärt:

$$\boxed{A(t) := \begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{m1}(t) & \cdots & a_{mn}(t) \end{bmatrix} \;\; \forall\, t \in D \;\text{ mit }\; D(A) = D, \;\; \text{Bild}\, A \subset \mathbf{R}^{(m,n)}.}$$

Zum Beispiel: $A(t) := \begin{bmatrix} t & t-\sin t \\ \frac{1}{2}t^2 & 1+\cos t \end{bmatrix} \in \mathbf{R}^{(2,2)} \ \forall\, t \in \mathbf{R}.$

Im Regelfall ist die Abbildungsvorschrift, nach welcher die Werte $f(x)$ einer Funktion f zu berechnen sind, ein *formelmäßiger* Ausdruck für $f(x)$; man vergleiche die oben angegebenen Beispiele (6.1.1)–(6.1.10).

Definition 6.6 *Der* **maximale** *oder* **natürliche Definitionsbereich** $D_{\max}(f) \subset \mathbf{R}$ *einer Funktion f der reellen Veränderlichen x ist diejenige Menge, die zu jedem ihrer Punkte $x \in D_{\max}(f)$ einen formelmäßigen Ausdruck für $f(x)$ zulässt, während $f(x)$ für $x \neq D_{\max}(f)$ nicht definierbar ist.*

BSP. (6.1.15) Die folgende Tabelle enthält einige Funktionen und ihre maximalen Definitionsbereiche:

$f(x)$	$D_{\max}(f)$
$\sqrt{\lvert x\rvert + \dfrac{1}{x}}$	$(-\infty,-1] \cup (0,+\infty)$
$\dfrac{\sqrt{x^2-1}}{x+5}$	$\mathbf{R} \setminus (\{-5\} \cup (-1,1))$
$(\sin x)^{3/2}$	$\bigcup\limits_{n\in\mathbf{Z}} I_n$ mit $I_n := [2n\pi, (2n+1)\pi]$

Ist $D(f)$ nicht angegeben, so ist stets der maximale Definitionsbereich vorauszusetzen.

Wir behandeln im folgenden ausschließlich Funktionen $f : X \to Y$ einer reellen Veränderlichen x, das heißt, es wird stets $X \subset \mathbf{R}$ angenommen.

Definition 6.7 *Es sei Y ein Vektorraum über dem Körper* $\mathbf{K}$.

(a) *Eine Zahl $x_0 \in X$ heiße* **Nullstelle** *von $f : X \to Y$, wenn $f(x_0) = 0$ gilt.*

(b) *Für $f, g \in \mathrm{Abb}\,(X,Y)$ sind die* **Summe** *$f+g$ und das* **skalare Vielfache** *λf, $\lambda \in \mathbf{K}$,* **punktweise** *erklärt durch*

$$\boxed{\text{(i) } (f+g)(x) := f(x)+g(x), \qquad \text{(ii) } (\lambda f)(x) := \lambda f(x) \ \forall\, x \in X.}$$

In diesem Sinne ist die Menge

$$\mathrm{Abb}\,(X,Y) := \{f : X \to Y \ : \ D(f) = X\}$$

ein **Vektorraum** über dem Körper $\mathbf{K}$.

Bemerkung 6.1 Wir haben in (a) den Nullvektor mit "0" bezeichnet anstatt mit "$\vec{0}$", da wir in der Regel die Vektorräume $Y = \mathbf{R}$ und $Y = \mathbf{C}$ als Zielmengen betrachten werden. □

Produkt– und Quotientenbildung von Funktionen sind nur möglich, wenn diese Operationen in der Zielmenge Y erlaubt sind, das heißt, wenn Y ein Körper ist.

Definition 6.8 *Es sei* $\mathbf{K}$ *ein Körper, und es seien $f, g \in \mathrm{Abb}\,(X, \mathbf{K})$ gegeben.*

(a) *Das* **Produkt** *fg ist* **punktweise** *erklärt durch*

$$\boxed{(fg)(x) := f(x)g(x) \ \forall\, x \in X.}$$

(b) *Der* **Quotient** f/g *ist* **außerhalb der Nullstellen von** g **punktweise** *erklärt durch*

$$\boxed{\left(\frac{f}{g}\right)(x) := \frac{f(x)}{g(x)} \ \forall\, x \in X \setminus \{x_0 \,:\, g(x_0) = 0\}.}$$

BSP. (6.1.16) Für $f(x) := \sqrt{x}$ und $g(x) := \sin x$ gelten auf der Menge $X := [0, +\infty)$:

$$(fg)(x) := \sqrt{x}\,\sin x \ \forall\, x \in X, \quad \left(\frac{f}{g}\right)(x) = \frac{\sqrt{x}}{\sin x} \ \forall\, x \in X \setminus \{n\pi \,:\, n \in \mathbf{N}_0\}.$$

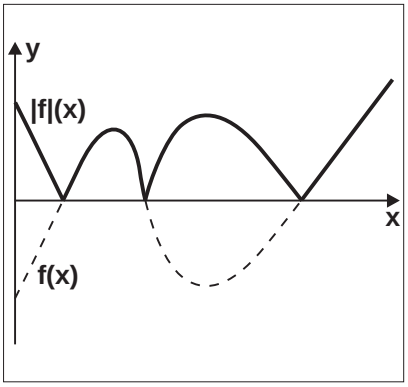
Man beachte, dass $0 \neq D(f/g)$ gilt.

Definition 6.9 *Es seien* $\mathbf{K} := \mathbf{R}$ *oder* $\mathbf{K} := \mathbf{C}$ *der Körper der reellen bzw. der komplexen Zahlen. Der* **Betrag** $|f|$ *der Funktion* $f : X \to \mathbf{K}$ *ist* **punktweise** *erklärt durch*

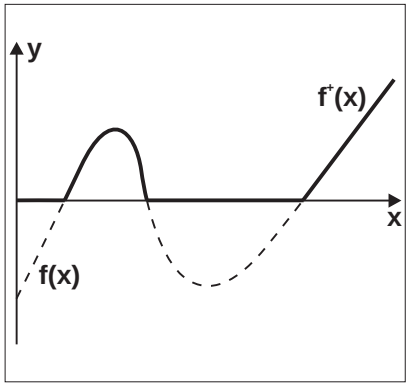
$$\boxed{|f|(x) := |f(x)| \ \forall\, x \in X.}$$

Ist speziell $\mathbf{K} := \mathbf{R}$*, so geben für* $f, g \in \mathrm{Abb}\,(X, \mathbf{K})$ *die folgenden Definitionen einen Sinn. Der* **positive Teil** f^+*, der* **negative Teil** f^-*, das* **Maximum** $\max\{f, g\}$ *und das* **Minimum** $\min\{f, g\}$ *sind* **punktweise** *erklärt durch*

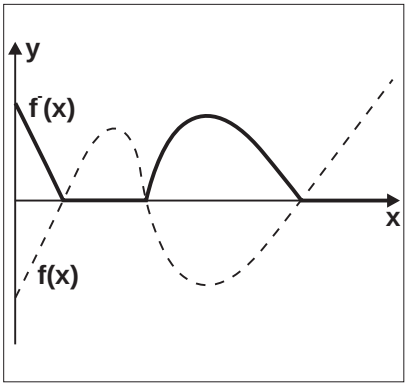
$$\boxed{\begin{array}{ll} f^+(x) := \begin{cases} f(x) & : \ f(x) \geq 0, \\ 0 & : \ f(x) < 0, \end{cases} & f^-(x) := \begin{cases} 0 & : \ f(x) \geq 0, \\ -f(x) & : \ f(x) < 0, \end{cases} \\ \multicolumn{2}{l}{(\max\{f, g\})(x) := \max\{f(x), g(x)\}, \quad (\min\{f, g\})(x) := \min\{f(x), g(x)\}.} \end{array}}$$



Der Betrag von f



Der positive Teil von f



Der negative Teil von f

Man prüft die folgenden Zusammenhänge leicht nach:

Satz 6.1 *Es seien* $f, g \in \mathrm{Abb}\,(X, \mathbf{R})$ *gegeben. Dann gelten:*

$$\boxed{f = f^+ - f^-, \quad |f| = f^+ + f^-, \quad f^+ = \frac{1}{2}(|f| + f), \quad f^- = \frac{1}{2}(|f| - f),} \tag{1.4}$$

$$\boxed{\max\{f, g\} = \frac{1}{2}(f + g + |f - g|), \quad \min\{f, g\} = \frac{1}{2}(f + g - |f - g|).} \tag{1.5}$$

BSP. (6.1.17) Für ein festes $x_0 \in \mathbf{R}$ setzen wir $f(x) := x - x_0 \; \forall \, x \in \mathbf{R}$. Es gelten

$$(x-x_0)^+ = \begin{cases} x - x_0 & : \; x \geq x_0, \\ 0 & : \; x < x_0, \end{cases} \qquad (x-x_0)^- = \begin{cases} 0 & : \; x \geq x_0, \\ x_0 - x & : \; x < x_0, \end{cases}$$

und man sieht sofort

$$(x-x_0)^+ + (x-x_0)^- = |x-x_0| = \begin{cases} x - x_0 & : \; x \geq x_0, \\ x_0 - x & : \; x < x_0. \end{cases}$$

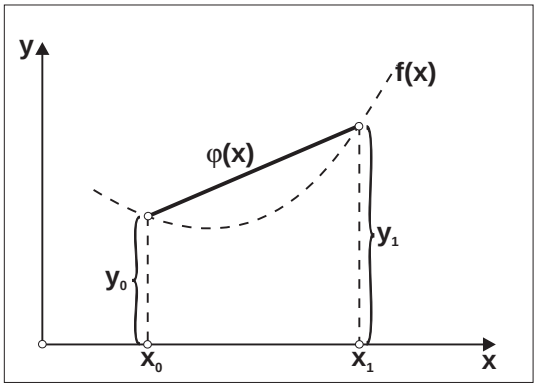
6.2 Polynominterpolation

Es sei $\mathbf{K} := \mathbf{R}$ oder $\mathbf{K} := \mathbf{C}$ der Körper der reellen oder komplexen Zahlen. Von der Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ seien einige Werte $y = f(x)$ aus Funktionstafeln oder aus Versuchsergebnissen numerisch vorgelegt. Es erhebt sich die Frage, wie mit Hilfe der schon vorhandenen Werte weitere Funktionswerte wenigstens näherungsweise bestimmt werden können. Die Behandlung dieser Aufgabe heißt

Interpolationsproblem. Gegeben seinen $n+1$ paarweise verschiedene **Stützstellen** $x_0, x_1, \ldots, x_n \in \mathbf{R}$ und dazu $n+1$ (nicht notwendig verschiedene) **Stützwerte** $y_0 := f(x_0), y_1 := f(x_1), \ldots, y_n := f(x_n)$. Bestimme eine geeignete **Interpolationsfunktion** $\varphi \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$, die die folgenden Interpolationsbedingungen erfüllt:

$$y_j = \varphi(x_j) \;\; \forall \; j = 0, 1, \ldots, n. \tag{IP}$$

Eine einfache Lösung hat man im Falle $n = 1$:



Interpolation durch eine affine Funktion

Die Funktion $f(x)$ wird im Intervall $[x_0, x_1]$ durch die **Sehne** ersetzt; diese ist ein *Polynom vom Grade 1*:

$$\varphi(x) := y_0 + \frac{y_1 - y_0}{x_1 - x_0} \cdot (x - x_0), \quad x_0 \leq x \leq x_1.$$

Beachte: Polynome bieten sich wegen ihrer einfachen Möglichkeit der Funktionsauswertung durch das Horner–Schema besonders zur Lösung des Interpolationsproblems an. Wir setzen deshalb:

$$\mathbf{K}_n[x] := \{P_n \; : \; P_n(x) := \sum_{k=0}^{n} a_k x^k, \;\; \text{Polynom mit Grad } P_n \leq n\}.$$

In $\mathbf{K}_n[x]$ ist das Interpolationsproblem eindeutig lösbar:

Satz 6.2 *Zu beliebig vorgegebenen $n+1$* **Stützpunkten** *(x_j, y_j), $j = 0, 1, \ldots, n$, mit $x_j \neq x_k$ für $j \neq k$, gibt es genau ein Polynom $P_n \in \mathbf{K}_n[x]$, welches die Interpolationsbedingungen erfüllt:*

$$\boxed{P_n(x_j) = y_j \quad \forall j = 0, 1, \ldots, n.} \tag{IP}$$

Dieses ist das LAGRANGE**–Interpolationspolynom**

$$\boxed{P_n(x) := \sum_{j=0}^{n} y_j L_j(x),} \tag{2.1}$$

worin L_j das j–te LAGRANGE**sche Polynom** *vom Grade n bezeichne:*

$$\boxed{L_j(x) := \prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{(x - x_k)}{(x_j - x_k)} = \frac{(x - x_0)\cdots(x - x_{j-1})(x - x_{j+1})\cdots(x - x_n)}{(x_j - x_0)\cdots(x_j - x_{j-1})(x_j - x_{j+1})\cdots(x_j - x_n)}.} \tag{2.2}$$

Begründungen: (a) Zur Eindeutigkeit: Wären $P_n, Q_n \in \mathbf{K}_n[x]$ zwei Polynome mit der Interpolationseigenschaft (IP), so hätte das Differenzpolynom $P(x) := P_n(x) - Q_n(x)$ mit $\operatorname{Grad} P \leq n$ mindestens die $n+1$ verschiedenen Nullstellen $x_0, x_1, \ldots, x_n$. Sofern nicht $P \equiv 0$ ist, widerspricht dies dem Fundamentalsatz der Algebra.

(b) Die Polynome L_j aus (2.2) mit $\operatorname{Grad} L_j = n$ haben offenbar die Eigenschaft

$$L_j(x_k) = \delta_{jk} := \begin{cases} 1 \text{ für } j = k, \\ 0 \text{ für } j \neq k. \end{cases} \tag{2.3}$$

Somit löst das Polynom $P_n \in \mathbf{K}_n[x]$ aus (2.1) offenkundig das Interpolationsproblem. □

BSP. (6.2.1) Man bestimme für $n = 2$ das Interpolationspolynom $P \in \mathbf{R}_2[x]$ bei Vorgabe der Stützpunkte

x_j	0	1	3
y_j	-1	1	4

Man berechne $P(2)$. *Lösung:* Aus den Formeln (2.2) resultiert

$$\begin{aligned} L_0(x) &= \frac{(x-1)(x-3)}{(0-1)(0-3)} = \frac{1}{3}(x-1)(x-3), \\ L_1(x) &= \frac{(x-0)(x-3)}{(1-0)(1-3)} = -\frac{1}{2}x(x-3), \\ L_2(x) &= \frac{(x-0)(x-1)}{(3-0)(3-1)} = \frac{1}{6}x(x-1). \end{aligned}$$

Setzt man dies in (2.1) ein, so erhält man das gesuchte Interpolationspolynom

$$P(x) = -\frac{1}{3}(x-1)(x-3) - \frac{1}{2}x(x-3) + \frac{2}{3}x(x-1), \quad P(2) = \frac{8}{3}.$$

Bemerkung 6.2 Das LAGRANGE–Interpolationspolynom (2.1) gestattet die folgende Darstellung:

$$P_n(x) = \sum_{j=0}^{n} \frac{y_j}{(x - x_j)} \cdot \underbrace{\Big\{ \prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{1}{(x_j - x_k)} \Big\}}_{=:\lambda_j} \cdot \prod_{i=0}^{n} (x - x_i). \tag{2.4}$$

Beachte: Speziell für $y_j := 1, \quad j = 0, 1, \ldots, n$, folgt aus (2.4):

$$1 = P_n(x) = \Big(\sum_{j=0}^{n} \underbrace{\frac{\lambda_j}{x - x_j}}_{=:\mu_j} \Big) \cdot \prod_{i=0}^{n} (x - x_i), \quad \text{also} \quad \prod_{i=0}^{n} (x - x_i) = 1 \Big/ \sum_{j=0}^{n} \mu_j \quad \forall\, x \in \mathbf{R}. \tag{2.5}$$

Verwenden wir (2.5) in (2.4), so erhalten wir die

Baryzentrische Formel der LAGRANGE–Interpolation: Mit den **Stützkoffizienten**

$$\boxed{\lambda_j := 1 \Big/ \prod_{\substack{k=0 \\ k \neq j}}^{n} (x_j - x_k), \quad j = 0, 1, \ldots, n,} \tag{2.6}$$

und den von der **Neustelle** x abhängigen Hilfsgrößen

$$\boxed{\mu_j := \frac{\lambda_j}{x - x_j}, \quad j = 0, 1, \ldots, n,} \tag{2.7}$$

gestattet das LAGRANGE–Interpolationspolynom P_n aus (2.1) die Darstellung □

$$\boxed{P_n(x) = \frac{\displaystyle\sum_{j=0}^{n} \mu_j y_j}{\displaystyle\sum_{j=0}^{n} \mu_j} \quad \forall\, x \in \mathbf{R}.} \tag{2.8}$$

Rechentechnik und Aufwandsoptimierung. Zu gegebenen Stützstellen $x_0, x_1 \ldots, x_n$ berechnet man zuerst die Stützkoeffizienten λ_j gemäß der Vorschrift (2.6). Für jede Neustelle x können die Gewichte μ_j gemäß (2.7) sukzessive in der Reihenfolge $j = 0, j = 1, \ldots, j = n$ berechnet werden, wobei man gleichzeitig die zwei Summen (2.8) formt. Der so entstehende *Rechenaufwand* kann bei geschickter Rechnung etwa noch halbiert werden, wenn man folgenden Zusammenhang beachtet. Angenommen, zu $x_0, x_1, \ldots, x_n$ seien die Stützkoeffizienten $\lambda_j = \lambda_j^{(n)}$ bekannt. Für eine weitere Stützstelle x_{n+1} gilt dann wegen (2.6):

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{(x_j - x_{n+1})}, \quad j = 0, 1, \ldots, n. \tag{2.9}$$

Das heißt, $\lambda_j^{(n+1)}, j = 0, 1, \ldots, n$, entsteht aus $\lambda_j^{(n)}$ durch eine einzige Division. Der fehlende Stützkoeffizient $\lambda_{n+1}^{(n+1)}$ kann auf Grund folgender Eigenschaft berechnet werden:

Satz 6.3 *Gegeben seien die paarweise verschiedenen Stützstellen $x_0, x_1, \ldots, x_n \in \mathbf{R}$. Dann gilt für die Stützkoeffizienten $\lambda_j = \lambda_j^{(n)}$ die Relation*

$$\sum_{j=0}^{n} \lambda_j^{(n)} = 0 \quad \textit{für} \quad n \geq 1. \tag{2.10}$$

Begründung: Aus (2.4) folgt offenbar die Darstellung

$$P_n(x) = \sum_{j=0}^{n} y_j \lambda_j^{(n)} \cdot \prod_{\substack{i=0 \\ i \neq j}}^{n} (x - x_i) = x^n \cdot \sum_{j=0}^{n} y_j \lambda_j^{(n)} \quad + x^{n-1} a_{n-1} + \cdots + a_0. \tag{2.11}$$

Im Sonderfall $y_j := 1 \; \forall\, j = 0, 1, \ldots, n$, resultiert wieder $P_n(x) \equiv 1$, und somit $a_j = 0 \; \forall\, j = 1, 2, \ldots, n, \; a_0 = 1$. Insbesondere gilt $a_n = \sum_{j=0}^{n} \lambda_j^{(n)} = 0$, also (2.10). □

Die Stützkoeffizienten $\lambda_j^{(k)}$ können jetzt für $k = 1, 2, \ldots, n$ aus der Rekursion (2.9) und der Beziehung (2.10) sukzessive berechnet werden, sofern ein Startwert $\lambda_0^{(0)}$ vorgegeben ist. Um einen solchen zu bestimmen, setzen wir in (2.6) $n = 1$ und die Stützstellen x_0, x_1 ein:

$$\lambda_0^{(1)} = \frac{1}{x_0 - x_1}, \qquad \lambda_1^{(1)} = \frac{1}{x_1 - x_0} = -\lambda_0^{(1)}.$$

Vergleicht man dies mit (2.9) und (2.10), so gelangt man zum Startwert $\lambda_0^{(0)} = 1$ und hiermit zu folgendem

Algorithmus zur Bestimmung der Stützkoeffizienten.

```
1:  Einlesen von x_0, x_1, ..., x_n;
2:  λ_0 := 1;
3:  für k := 1, 2, ..., n :
4:      s := 0;
5:      für j := 0, 1, ..., k − 1 :
6:          λ_j := λ_j/(x_j − x_k); s := s + λ_j;  (Ende j)
7:      λ_k := −s.  (Ende k)
```
(2.12)

Sonderfall: Äquidistante Stützstellen. Das ist der Fall $x_j := x_0 + jh$ für festes $h > 0$ und $j = 0, 1, \ldots, n$. Aus (2.6) erhalten wir jetzt:

$$\lambda_j = 1 \Big/ \prod_{\substack{k=0 \\ k \neq j}}^{n} (x_j - x_k) = \frac{(-1)^{n-j}}{h^n n!} \cdot \binom{n}{j}. \tag{2.13}$$

Da es in der Formel (2.8) offenbar *nicht* auf einen gemeinsamen Faktor in den Gewichten μ_j ankommt, kann die Konstante $(-1)^n/(h^n n!)$ in (2.13) fortgelassen werden. Man gelangt so zu **Ersatzkoeffizienten**

$$\boxed{\lambda_j^* := (-1)^j \binom{n}{j}, \quad j = 0, 1, \ldots, n,} \tag{2.14}$$

mit denen man jetzt die Interpolationsformeln (2.7) und (2.8) aufbaut. Die Zahlen λ_j^* lassen sich wiederum *rekursiv* berechnen:

$$\boxed{\lambda_0^* := 1; \quad \lambda_j^* := -\frac{n - j + 1}{j} \cdot \lambda_{j-1}^*, \quad j = 1, 2, \ldots, n.} \tag{2.15}$$

Das Lagrangesche Interpolationspolynom (2.1) hat den Nachteil, dass die Berechnung von $L_j(x)$ vollkommen neu durchgeführt werden muss, wenn sich die Zahl der Stützstellen erhöht. Dieser Nachteil wird beseitigt, wenn das (eindeutig bestimmte) Interpolationspolynom in der Newtonschen Form

$$\boxed{P_n(x) = c_0 + \sum_{k=1}^{n} c_k \prod_{i=0}^{k-1} (x - x_i)} \tag{2.16}$$

angesetzt wird. Die unbekannten Koeffizienten $c_0, c_1, \ldots, c_n$ lassen sich prinzipiell aus den Interpolationsbedingungen (IP) $y_j = P_n(x_j)\ \forall\ j = 0, 1, \ldots, n$ berechnen. Diese führen auf das folgende lineare Gleichungssystem:

$$\left.\begin{aligned}
P_n(x_0) &= c_0 &&= y_0, \\
P_n(x_1) &= c_0 + c_1(x_1 - x_0) &&= y_1, \\
P_n(x_2) &= c_0 + c_1(x_2 - x_0) + c_2(x_2 - x_0)(x_2 - x_1) &&= y_2, \\
&\ \vdots \quad \vdots && \ \vdots \\
P_n(x_n) &= c_0 + c_1(x_n - x_0) + \cdots + c_n(x_n - x_0)(x_n - x_1) \cdots (x_n - x_{n-1}) &&= y_n.
\end{aligned}\right\} \tag{2.17}$$

Aus der Dreiecksgestalt des Systems (2.17) resultieren sofort folgende **Vorteile** der NEWTON–Interpolation:

- Beginnend mit $c_0 := y_0$ können die c_j sukzessive aus (2.17) berechnet werden.
- Bei Hinzunahme einer (oder mehrerer) Stützpunkte (x_{n+1}, y_{n+1}) braucht lediglich ein neuer Koeffizient c_{n+1} berechnet zu werden (bzw. eine entsprechende Anzahl neuer Koeffizienten), während die alten Koeffizienten $c_0, c_1, \ldots, c_n$ **unverändert** bleiben.
- Das NEWTON–Interpolationspolynom (2.16) gestattet die Darstellung

  $$P_n(x) = [\cdots[[c_n(x - x_{n-1}) + c_{n-1}](x - x_{n-2}) + c_{n-2}](x - x_{n-3}) + \cdots + c_1](x - x_0) + c_0.$$

  Somit kann die Berechnung eines interpolierten Wertes $P_n(x)$ für eine Neustelle x nach einem HORNER–artigen Schema erfolgen:

	$\boxed{c_n}$		$\boxed{c_{n-1}}$		$\boxed{c_{n-2}}$	$\cdots$		$\boxed{c_1}$		$\boxed{c_0}$
	$+$		$+$		$+$	$\cdots$		$+$		$+$
	0		$(x - x_{n-1})d_{n-1}$		$(x - x_{n-2})d_{n-2}$	$\cdots$		$(x - x_1)d_1$		$(x - x_0)d_0$
x	d_{n-1}	$\nearrow$	d_{n-2}	$\nearrow$	d_{n-3}	$\cdots$	$\nearrow$	d_0	$\nearrow$	$\boxed{P_n(x)}$

Sind die Koeffizienten c_j bereitgestellt, so hat man den folgenden

Algorithmus zur Berechnung von $P_n(x)$ an einer Neustelle x:

```
1:   p := c_n;
2:   für j := n − 1, n − 2, . . . , 0 :
3:       p := c_j + (x − x_j) ∗ p.  (Ende j)
```
(2.18)

Es kommt jetzt noch darauf an, ein einfaches und effizientes Verfahren zur Berechnung der Koeffizienten c_j aufzufinden. Dazu treffen wir die folgende

Definition 6.10 *Die eindeutig durch $k+1$ Stützpunkte (x_j, y_j), $j = 0, 1, \ldots, k$, festgelegten Koeffizienten c_k in (2.17) bezeichnen wir mit*

$$\boxed{c_k =: f[x_0, x_1, \ldots, x_k], \quad k = 0, 1, \ldots, n.} \tag{2.19}$$

Sind $j_0, j_1, \ldots, j_k$ irgendwelche $k+1$ paarweise verschiedene Zahlen aus der Indexmenge $\{0, 1, \ldots, n\}$, so bezeichnen wir das **Teilinterpolationspolynom** *zu den $k+1$ Stützpunkten $(x_{j_0}, y_{j_0}), \ldots, (x_{j_k}, y_{j_k})$ mit $P_{j_0 j_1 \cdots j_k}(x)$. Es gelten also die Interpolationsbedingungen*

$$\boxed{P_{j_0 j_1 \cdots j_k}(x_{j_m}) = y_{j_m}, \quad m = 0, 1, \ldots, k.} \tag{2.20}$$

Das Teilinterpolationspolynom $P_{j_0 j_1 \cdots j_k}(x)$ hat den Grad $P_{j_0 j_1 \cdots j_k} \leq k$, und der Koeffizient der *höchsten* Potenz x^k ist gerade $f[x_{j_0}, x_{j_1}, \ldots, x_{j_k}]$. Zur rekursiven Berechnung dieses Koeffizienten zeigen wir:

Satz 6.4 *Es gelten für alle $x \in \mathbf{R}$ die folgenden Rekursionsformeln:*

$$\boxed{P_j(x) = y_j, \quad j = 0, 1, \ldots, n,} \tag{2.21}$$

$$\boxed{P_{j_0 j_1 \cdots j_k}(x) = \frac{(x - x_{j_0})P_{j_1 j_2 \cdots j_k}(x) - (x - x_{j_k})P_{j_0 j_1 \cdots j_{k-1}}(x)}{(x_{j_k} - x_{j_0})}, \quad 1 \leq k \leq n.} \tag{2.22}$$

Begründung: (2.21) stellt das Interpolationspolynom 0–ten Grades durch den Stützpunkt (x_j, y_j) dar. Das ist offenbar die Konstante $P_j(x) \equiv y_j$. Zum Beweis von (2.22) bezeichnen wir die rechte Seite von (2.22) mit $P(x)$ und zeigen, dass $P(x)$ die Eigenschaften von $P_{j_0 j_1 \cdots j_k}(x)$ besitzt. Trivialerweise gilt $\operatorname{Grad} P(x) \le k$, ferner folgt aus (2.20):

$$P(x_{j_0}) = P_{j_0 j_1 \cdots j_{k-1}}(x_{j_0}) = y_{j_0}, \quad P(x_{j_k}) = P_{j_1 j_2 \cdots j_k}(x_{j_k}) = y_{j_k},$$

sowie für $m = 1, 2, \ldots, k-1$:

$$P(x_{j_m}) = \frac{(x_{j_m} - x_{j_0})y_{j_m} - (x_{j_m} - x_{j_k})y_{j_m}}{(x_{j_k} - x_{j_0})} = y_{j_m}.$$

Wegen der Eindeutigkeit des Interpolationspolynoms resultiert daher $P(x) = P_{j_0 j_1 \cdots j_k}(x)$. □

Bemerkung 6.3 Werden nun in (2.22) auf beiden Seiten jeweils die Koeffizienten vor der (höchsten) Potenz x^k miteinander verglichen, so resultieren mit den Bezeichnungen (2.19) die gesuchten **Rekursionsformeln** für die Koeffizienten $f[x_{j_0}, x_{j_1}, \ldots, x_{j_k}]$: □

$$\boxed{f[x_{j_0}, x_{j_1}, \ldots, x_{j_k}] = \frac{f[x_{j_1}, x_{j_2}, \ldots, x_{j_k}] - f[x_{j_0}, x_{j_1}, \ldots, x_{j_{k-1}}]}{x_{j_k} - x_{j_0}}.} \tag{2.23}$$

Definition 6.11 *Die Zahl $f[x_{j_0}, x_{j_1}, \ldots, x_{j_k}]$ heiße die den Stützstellen $x_{j_0}, x_{j_1}, \ldots, x_{j_k}$ zugeordnete* k**–te dividierte Differenz**. *Die Auswertung der Rekursionsformel* (2.23) *erfolgt unter Verwendung der Startwerte $f[x_j] := y_j$, $j = 0, 1, \ldots, n$, im* **Schema der dividierten Differenzen**, *zum Beispiel für $n = 4$:*

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
x_0	$\boxed{y_0 = f[x_0]}$				
		$\boxed{f[x_0, x_1]}$			
x_1	$y_1 = f[x_1]$		$\boxed{f[x_0, x_1, x_2]}$		
		$f[x_1, x_2]$		$\boxed{f[x_0, x_1, x_2, x_3]}$	
x_2	$y_2 = f[x_2]$		$f[x_1, x_2, x_3]$		$\boxed{f[x_0, x_1, x_2, x_3, x_4]}$
		$f[x_2, x_3]$		$f[x_1, x_2, x_3, x_4]$	
x_3	$y_3 = f[x_3]$		$f[x_2, x_3, x_4]$		
		$f[x_3, x_4]$			
x_4	$y_4 = f[x_4]$				

Die Bearbeitung des Schemas der dividierten Differenzen erfolgt *spaltenweise von links nach rechts*. Die gesuchten Koeffizienten c_j des NEWTON–Polynoms (2.16) stehen in der obersten Schrägzeile als **eingerahmte Größen**. Es gelten beispielshaft die folgenden Formeln:

$$f[x_0, x_1] = \frac{f[x_1] - f[x_0]}{x_1 - x_0}, \quad f[x_3, x_4] = \frac{f[x_4] - f[x_3]}{x_4 - x_3},$$

$$f[x_1, x_2, x_3, x_4] = \frac{f[x_2, x_3, x_4] - f[x_1, x_2, x_3]}{x_4 - x_1}.$$

Bei rechnermäßiger Auswertung des obigen Schemas können die sukzessive berechneten Spalten in einem Vektor $\vec{c}$ mit $n+1$ Komponenten abgespeichert werden. Da nur der Wert der obersten

Schrägzeile relevant ist, berechnet man die Spalten jeweils von unten nach oben, so dass zum Schluss der Vektor $\vec{c}$ die Koeffizienten des NEWTON–Polynoms (2.16) enthält.

Algorithmus zur Berechnung der c_j.

```
1:  Einlesen von (x_j, y_j), j := 0, 1, ..., n;
2:  für j := 0, 1, ..., n :
3:      c_j := y_j; (Ende j)
4:  für k := 1, 2, ..., n :
5:      für j := n, n − 1, ..., k :
6:          c_j := (c_j − c_{j−1})/(x_j − x_{j−k}). (Ende j, k)
```

(2.24)

BSP. (6.2.2) Man berechne das Schema der dividierten Differenzen für die Stützstellen $x_j := 0,\ 0.5,\ 1.5,\ 2.5,\ 3.0$ und die Stützwerte $y_j := \sin x_j,\ j = 0, 1, 2, 3, 4$, die man mit Taschenrechnergenauigkeit (6 Nachkommastellen) bestimme. *Lösung:* Es gilt hier $n = 4$, und somit

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$x_0 = 0.0$	$\boxed{0.000\,000}$				
		$\boxed{0.958\,852}$			
$x_1 = 0.5$	$0.479\,426$		$\boxed{-0.293\,855}$		
		$0.518\,069$		$\boxed{-0.065\,876}$	
$x_2 = 1.5$	$0.997\,495$		$-0.458\,546$		$\boxed{0.037\,260}$
		$-0.399\,023$		$0.045\,904$	
$x_3 = 2.5$	$0.598\,472$		$-0.343\,787$		
		$-0.914\,704$			
$x_4 = 3.0$	$0.141\,120$				

Das NEWTON–Interpolationspolynom erhält damit die Form

$$P_4(x) = x\Big\{0.958\,852 + (x-0.5)\Big[-0.293\,855 + (x-1.5)\Big(-0.065\,876 + (x-2.5)0.037\,260\Big)\Big]\Big\}.$$

Zum Beispiel berechnet man seinen Wert an der Neustelle $x := \frac{\pi}{2}$ zu $P_4(x) \doteq 0.999\,993$.

Sonderfall: **Äquidistante Stützstellen.** Gilt $x_j := x_0 + jh$ für festes $h > 0$ und $j = 0, 1, \ldots, n$, so vereinfacht sich die Berechnung der k–ten dividierten Differenzen nach dem Schema (2.23) ganz erheblich.

Satz 6.5 *Zu den gegebenen Stützwerten $y_0, y_1, \ldots, y_n$ seien die* **k–ten absteigenden Differenzen** *$\Delta^k y_j$ wie folgt rekursiv definiert:*

$$\boxed{\Delta^0 y_j := y_j, \quad \Delta^k y_j := \Delta^{k-1} y_{j+1} - \Delta^{k-1} y_j, \ \ j = 0, 1, \ldots, n-k,\ k = 1, 2, \ldots, n.} \tag{2.25}$$

Hiermit gilt:

$$\boxed{f[x_j, x_{j+1}, \ldots, x_{j+k}] = \frac{1}{k!h^k} \cdot \Delta^k y_j, \ \ j = 0, 1, \ldots, n-k,\ k = 1, 2, \ldots, n.} \tag{2.26}$$

Begründung: Wir führen vollständige Induktion nach $k \le n$ durch.

Verankerung: Für $k = 0$ gilt definitionsgemäß $\Delta^0 y_j = y_j = f[x_j]\ \forall\ j = 0, 1, \ldots, n$. Für $k = 1$ folgt aus (2.23) und (2.25):

$$f[x_j, x_{j+1}] = \frac{1}{h}\Big(f[x_{j+1}] - f[x_j]\Big) = \frac{1}{h}\Big(\Delta^0 y_{j+1} - \Delta^0 y_j\Big) = \frac{1}{1!h}\,\Delta^1 y_j.$$

Vererbung: Die Formel (2.26) sei bereits wahr bis zum Index $k-1$. Unter Verwendung von (2.25) erhält man aus dieser Induktionsannahme:

$$\begin{aligned}\Delta^k y_j &= \Delta^{k-1} y_{j+1} - \Delta^{k-1} y_j \\ &= \Big(f[x_{j+1}, x_{j+2}, \ldots, x_{j+k}] - f[x_j, x_{j+1}, \ldots, x_{j+k-1}] \Big)\, (k-1)!\, h^{k-1} \\ &\overset{(2.23)}{=} f[x_j, x_{j+1}, \ldots, x_{j+k}] \cdot kh\, (k-1)!\, h^{k-1}.\end{aligned}$$

Nach Division durch $k!h^k$ resultiert nun die behauptete Relation (2.26). □

Die Formel (2.26) zeigt, dass die k–ten absteigenden Differenzen $\Delta^k y_j$ aus dem folgenden **Differenzenschema** berechnet werden können, welches genauso strukturiert ist wie das Schema der dividierten Differenzen:

$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
y_0				
	$\Delta^1 y_0$			
y_1		$\Delta^2 y_0$		
	$\Delta^1 y_1$		$\Delta^3 y_0$	
y_2		$\Delta^2 y_1$		$\Delta^4 y_0$
	$\Delta^1 y_2$		$\Delta^3 y_1$	
y_3		$\Delta^2 y_2$		
	$\Delta^1 y_3$			
y_4				

Merke: **Lückendifferenz** = unterer Wert minus oberer Wert!

Die gesuchten Koeffizienten c_k des NEWTONschen Interpolationspolynoms sind nun durch die folgenden Formeln gegeben:

$$\boxed{c_k = \frac{1}{k!h^k} \cdot \Delta^k y_0.} \tag{2.27}$$

Satz 6.6 *Das eindeutig bestimmte* NEWTON*–Interpolationspolynom vom Grade höchstens n, das in den äquidistanten Stützstellen $x_j = x_0 + jh$ die Stützwerte y_j, $j = 0, 1, \ldots, n$, annimmt, hat die Darstellung*

$$\boxed{P_n(x) = y_0 + \sum_{k=1}^{n} \frac{\Delta^k y_0}{k!h^k}(x-x_0)(x-x_1)\cdots(x-x_{k-1}).} \tag{2.28}$$

Dabei können die Koeffizienten $\Delta^k y_0$ in der obersten Schrägzeile des Differenzenschemas abgelesen werden.

Wird die **relative Distanz** der Neustelle x zu x_0 gemäß

$$t := \frac{x - x_0}{h} \tag{2.29}$$

definiert, so erhält man:

$$\frac{1}{k!h^k}(x-x_0)(x-x_1)\cdots(x-x_{k-1}) = \frac{t(t-1)\cdots(t-k+1)}{k!} =: \binom{t}{k}.$$

Deshalb resultiert aus (2.28) die Interpolationsformel von NEWTON–GREGORY:

$$\boxed{P_n(x) = y_0 + \sum_{k=1}^{n} \binom{t}{k} \Delta^k y_0 \ \text{ mit } \ t := \tfrac{1}{h}(x - x_0).} \tag{2.30}$$

Die Auswertung von (2.30) erfolgt wieder mit einem HORNER–**artigen Algorithmus**. Aus den Vorgaben x_0, x und $d_0 := y_0, d_k := \Delta^k y_0$ für $k = 1, 2, \ldots, n$ berechnet man:

```
1:  t := (x − x_0)/h; p := d_n;
2:  für j := n − 1, n − 2, ..., 0 :
3:      p := p ∗ (t − j)/(j + 1) + d_j.  (Ende j)
```
(2.31)

BSP. (6.2.3) Wir greifen nochmals das BSP. (6.2.2) mit der Funktion $y = \sin x$ auf, die wir jetzt an den **äquidistanten** Stützstellen $x_j := 0.5\, j,\ j = 0, 1, 2, 3, 4$, interpolieren wollen. Wir rechnen wieder mit Taschenrechnergenauigkeit. Es ergibt sich das folgende Differenzenschema:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
0.0	**0.000 000**				
		0.479 426			
0.5	0.479 426		**−0.117 381**		
		0.362 045		**−0.088 640**	
1.0	0.841 471		−0.206 021		**0.050 439**
		0.156 024		−0.038 201	
1.5	0.997 495		−0.244 222		
		−0.088 198			
2.0	0.909 297				

Zum Beispiel errechnet sich hieraus der interpolierte Wert an der Neustelle $x := \frac{\pi}{2}$ mit der relativen Distanz $t = \pi$ nach (2.31) zu $P_4(x) \doteq 1.000\,107$.

Von Interesse bei der Lösung des Interpolationsproblems ist noch die Beantwortung der Frage nach der Größe des **Interpolationsfehlers**

$$F_n(x) := |f(x) - P_n(x)|.$$

Eine befriedigende Antwort kann erst mit Hilfsmitteln der *Differentialrechnung* gegeben werden, und wir verschieben das Problem auf den Abschnitt 7.9. Wir können jedoch schon jetzt eine leicht nachprüfbare Bedingung für f formulieren, mit deren Hilfe der Fehler $F_n(x)$ bequem abgeschätzt werden kann.

Definition 6.12 *Eine Funktion f genügt auf dem Intervall $[a, b] \subset \mathbf{R}$ einer* LIPSCHITZ–**Bedingung**, *wenn eine* LIPSCHITZ–**Konstante** $L \geq 0$ *so bestimmt werden kann, dass gilt*

$$\boxed{|f(x) - f(y)| \leq L\,|x - y| \ \ \forall\, x, y \in [a, b].} \tag{2.32}$$

Wie in einem der folgenden Abschnitte noch zu zeigen sein wird, genügt jedes Polynom $P(x)$ auf einem abgeschlossenen Intervall $[a, b] \subset \mathbf{R}$ einer LIPSCHITZ–Bedingung. Hat zum Beispiel das NEWTON–Interpolationspolynom $P_n(x)$ auf $[a, b]$ die LIPSCHITZ–Konstante M, so gilt für jede Stützstelle x_k:

$$\begin{aligned} F_n(x) &\leq |f(x) - f(x_k)| + \underbrace{|f(x_k) - P_n(x_k)|}_{=0} + |P_n(x_k) - P_n(x)| \\ &\leq (L + M)\,|x - x_k| \ \ \forall\, x \in [a, b]. \end{aligned}$$

Bei *äquidistanten* Stützstellen x_k mit der Schrittweite $h > 0$ hat man somit die Fehlerabschätzung

$$\boxed{|f(x) - P_n(x)| \leq \frac{h}{2}(L + M) \quad \forall\, x \in [a, b].}$$

6.3 Grenzwerte von Funktionen einer reellen Veränderlichen

In diesem Abschnitt bezeichnen wir wieder mit $\mathbf{K}$ den Körper $\mathbf{R}$ der reellen bzw. den Körper $\mathbf{C}$ der komplexen Zahlen. Es sei $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{K})$ eine Funktion der reellen Veränderlichen x,

$$f : \begin{cases} D(f) \to \mathbf{K}, \\ x \mapsto f(x), \end{cases} \qquad \emptyset \neq D(f) \subseteq \mathbf{R}.$$

Wir untersuchen hier die folgende

Problemstellung: Nähert sich die Variable $x \in D(f)$ längs einer Folge $(x_n)_{n\in\mathbf{N}} \subset D(f)$ einem Grenzwert x_0 so, dass $\lim\limits_{n\to\infty} x_n = x_0$ gilt, konvergiert dann auch die Folge der Bildpunkte $\big(f(x_n)\big)_{n\in\mathbf{N}} \subset \mathbf{K}$ gegen den Bildpunkt $f(x_0)$?

Wir vermitteln in einer Reihe von Beispielen Vorinformationen über mögliche Fälle.

BSP. (6.3.1) Wir betrachten die HEAVISIDEsche Sprungfunktion

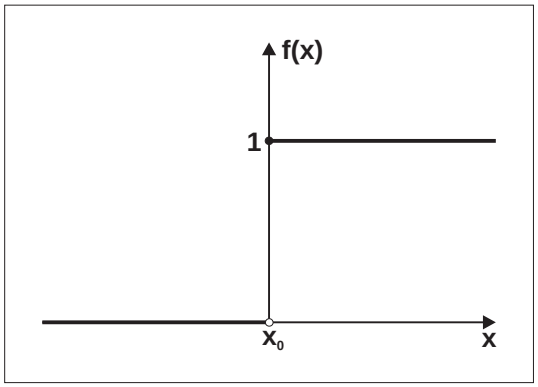
$$\boxed{f(x) := \begin{cases} 1 \;:\; x \geq 0, \\ 0 \;:\; x < 0, \end{cases} \qquad D(f) := \mathbf{R}.}$$

In dem einzig interessanten Punkt $x_0 = 0$ haben wir folgendes Verhalten:

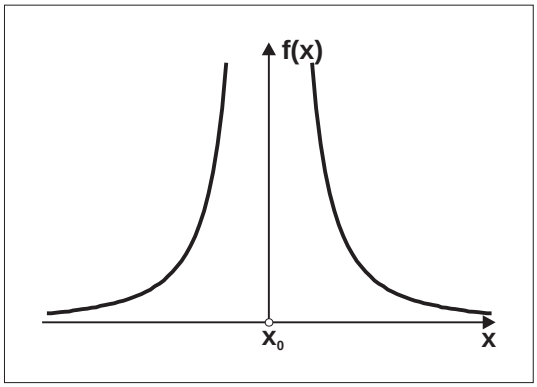
$$\text{(i)} \quad x_n := \tfrac{1}{n} > 0 \quad \Rightarrow \lim_{n\to\infty} x_n = 0 = x_0 \text{ und } \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} 1 = 1 = f(x_0).$$

$$\text{(ii)} \quad x_n := -\tfrac{1}{n} < 0 \;\Rightarrow \lim_{n\to\infty} x_n = 0 = x_0 \text{ und } \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} 0 = 0 \neq f(x_0).$$

Die Annäherung **von rechts** bzw. **von links** an den Punkt x_0 führt zu verschiedenen Grenzwerten der Bildfolge.



Die HEAVISIDEsche Sprungfunktion



Der Graph der Funktion $f(x) := \frac{1}{x^2}$

BSP. (6.3.2) Es sei

$$\boxed{f(x) := \frac{1}{x^2},\ x \in D(f) := \mathbf{R} \setminus \{0\}.}$$

Der Graph der Funktion f zeigt, dass $f(x)$ bei Annäherung an den Punkt $x_0 = 0$ unbeschränkt wächst. Dies zeigt auch die Rechnung:

$$x_n := \pm\frac{1}{n} \quad \Rightarrow \quad \lim_{n\to\infty} x_n = 0 = x_0 \ \text{ und } \ \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} n^2 = +\infty.$$

Die Folge der Bildpunkte $\Big(f(x_n)\Big)_{n\in\mathbf{N}} \subset \mathbf{R}$ konvergiert **uneigentlich gegen** $+\infty$**.** Man beachte

$$\boxed{x_0 \notin D(f).}$$

BSP. (6.3.3) Es sei

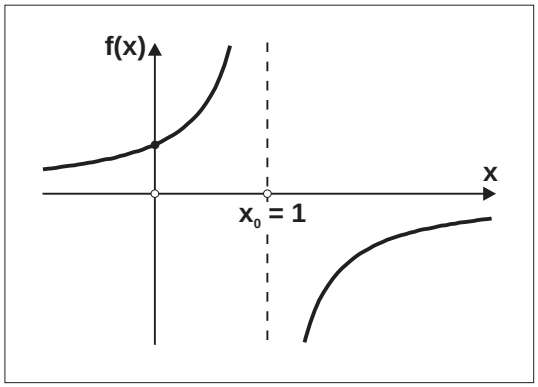
$$\boxed{f(x) := \frac{1}{1-x},\ x \in D(f) := \mathbf{R} \setminus \{1\}.}$$

Auch hier zeigt der Graph der Funktion f, dass $|f(x)|$ bei Annäherung an den Punkt $x_0 = 1$ unbeschränkt wächst. Anders als im BSP. (6.3.2) existiert aber kein (uneigentlicher) Grenzwert, denn es gilt:

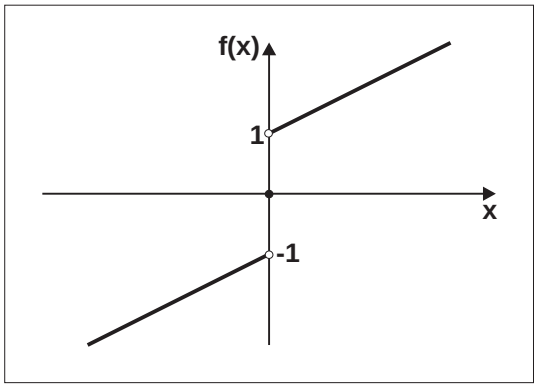
(i) $x_n := 1 - \frac{1}{n} < 1 \ \Rightarrow \ \lim_{n\to\infty} x_n = 1 = x_0$ und $\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} n = +\infty.$

(ii) $x_n := 1 + \frac{1}{n} > 1 \ \Rightarrow \ \lim_{n\to\infty} x_n = 1 = x_0$ und $\lim_{n\to\infty} f(x_n) = -\lim_{n\to\infty} n = -\infty.$

Bei Annäherung **von rechts** bzw. **von links** an den Punkt x_0 existieren **verschiedene** uneigentliche Grenzwerte der Bildfolge.



Der Graph der Funktion $f(x) := \frac{1}{1-x}$



Der Graph der Funktion $f(x) := \frac{x}{2} + \operatorname{sign} x$

BSP. (6.3.4) Es sei

$$\boxed{f(x) := \frac{x}{2} + \operatorname{sign} x,\ x \in D(f) := \mathbf{R}.}$$

In dem einzig interessanten Punkt $x_0 := 0$ haben wir $f(x_0) = 0$. Im Gegensatz dazu gilt:

$$x_n := \pm\frac{1}{n} \quad \Rightarrow \quad \lim_{n\to\infty} x_n = 0 = x_0 \ \text{ und } \ \lim_{n\to\infty} f(\tfrac{1}{n}) = +1 \neq f(x_0) \neq -1 = \lim_{n\to\infty} f(-\tfrac{1}{n}).$$

Wir präzisieren jetzt die aus der Anschauung gewonnenen Vorstellungen von der Existenz eines Funktionen–Grenzwertes (**Funktionenlimes**) in der folgenden Definition.

Definition 6.13 *Die Funktion* $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{K})$ *habe an der Stelle* $x_0 \in \mathbf{R}$ *den* **Grenzwert** $g \in \mathbf{K}$*, wenn gilt:*

$$\boxed{\begin{array}{l} \lim\limits_{n\to\infty} f(x_n) = g \text{ für } \textbf{jede } \text{Folge } (x_n)_{n\in\mathbf{N}} \subset D(f)\setminus\{x_0\} \text{ mit } \lim\limits_{n\to\infty} x_n = x_0. \\ \text{Schreibweise:} \qquad \boxed{\lim\limits_{x\to x_0} f(x) = g.} \end{array}} \tag{3.1}$$

Die Funktion f *habe an der Stelle* $x_0 \in \mathbf{R}$ *den* **rechtsseitigen Grenzwert** $g^+ \in \mathbf{K}$ *(bzw. den* **linksseitigen Grenzwert** $g^- \in \mathbf{K}$*), wenn gilt:*

$$\boxed{\begin{array}{l} \lim\limits_{n\to\infty} f(x_n) = g^+ \quad (\text{bzw. } g^-) \text{ für } \textbf{jede monoton fallende } (\text{bzw. monoton} \\ \text{wachsende}) \text{ Folge } (x_n)_{n\in\mathbf{N}} \subset D(f)\setminus\{x_0\} \text{ mit } \lim\limits_{n\to\infty} x_n = x_0. \\ \text{Schreibweise:} \qquad \boxed{\lim\limits_{x\to x_0+} f(x) = g^+ \ (\text{bzw. } \lim\limits_{x\to x_0-} f(x) = g^-).} \end{array}} \tag{3.2}$$

Bemerkung 6.4 (a) In den getroffenen Definitionen wird der Funktionswert $f(x_0)$ **nicht** benötigt. Deshalb kommt es **nicht** darauf an, ob x_0 zum Definitionsbereich $D(f)$ gehört oder nicht.

(b) In jedem Fall kommt es aber darauf an, dass die Bedingungen (3.1) und (3.2) für **alle** Folgen $(x_n)_{n\in\mathbf{N}}$ mit den dort spezifizierten Eigenschaften erfüllt sein müssen. □

BSP. (6.3.5) Es sei

$$\boxed{f(x) := \begin{cases} \sin x & : \ x \le 0, \\ \sin\frac{1}{x} & : \ x > 0, \end{cases} \qquad x \in D(f) := \mathbf{R}.}$$

In diesem Falle gilt:

$$\lim_{x\to 0-} f(x) = 0 =: g^-, \qquad \lim_{x\to+\infty} f(x) = 0,$$

während die beiden Grenzwerte $\lim\limits_{x\to 0+} f(x)$ und $\lim\limits_{x\to-\infty} f(x)$ **nicht** existieren.

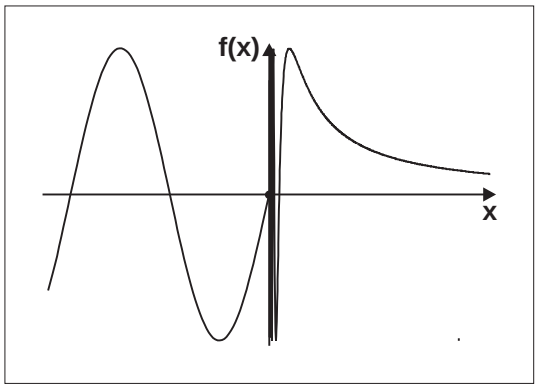
BSP. (6.3.6) Die DIRICHLET–Funktion

$$\boxed{f(x) := \chi_{\mathbf{Q}}(x) = \begin{cases} 1 & : \ x \text{ rational}, \\ 0 & : \ x \text{ irrational}, \end{cases} \qquad x \in D(f) := \mathbf{R},}$$

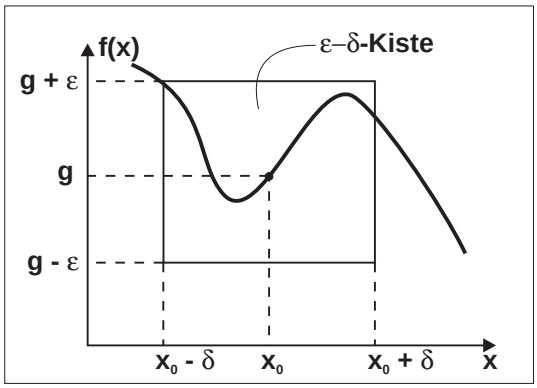
hat in **keinem** Punkt $x_0 \in \mathbf{R}$ einen Grenzwert (auch rechts– bzw. linksseitige Grenzwerte existieren nicht). Denn zu jedem Punkt $x_0 \in \mathbf{R}$ können Folgen $(x_n)_{n\in\mathbf{N}} \subset \mathbf{Q}$ als auch Folgen $(x_n^*)_{n\in\mathbf{N}} \subset \mathbf{R}\setminus\mathbf{Q}$ angegeben werden mit $\lim\limits_{n\to\infty} x_n = x_0 = \lim\limits_{n\to\infty} x_n^*$, während

$$\lim_{n\to\infty} \chi_{\mathbf{Q}}(x_n) = 1 \neq 0 = \lim_{n\to\infty} \chi_{\mathbf{Q}}(x_n^*)$$

gilt. Die Grenzwertbedingungen (3.1) und (3.2) sind verletzt, da nicht für jede Folge $x_n \to x_0$ derselbe Grenzwert $\chi_{\mathbf{Q}}(x_n) \to g$ resultiert.



$f(x) := \sin x$ **für** $x \leq 0$ **und**
$f(x) := \sin \frac{1}{x}$ **für** $x > 0$



Die $\epsilon - \delta$**–Kiste**

Die Existenz eines Funktionenlimes kann auch in der folgenden Form formuliert werden, die frei ist von der Auswahl von Folgen.

Satz 6.7 (Die $\epsilon - \delta$**–Kiste)**
Genau dann hat die Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ *an der Stelle* $x_0 \in \mathbf{R}$ *den Grenzwert* $g \in \mathbf{K}$, *wenn gilt*

$$\boxed{\forall\, \epsilon > 0\ \exists\, k \in \mathbf{N}\ :\ |f(x) - g| < \epsilon\ \ \forall\, x \in D(f)\ \ mit\ \ 0 < |x - x_0| < 10^{-k}.} \tag{3.3}$$

Begründung: Wir geben hier die formale Begründung, die an sich aber auf Grund der Konvergenzdefinition von Zahlenfolgen klar ist.

(a) Gelte zunächst die Bedingung (3.3). Man wähle $\epsilon > 0$ fest und dazu eine Zahl $k = k(\epsilon) \in \mathbf{N}$ gemäß der Vorschrift (3.3). Für jede beliebig gewählte Folge $(x_n)_{n\in\mathbf{N}} \subset D(f)\backslash\{x_0\}$ mit $\lim\limits_{n\to\infty} x_n = x_0$ existiert eine Zahl $N \in \mathbf{N}$, so dass gilt

$$0 < |x_n - x_0| < 10^{-k}\ \ \forall\, n > N.$$

Aus (3.3) erschließen wir somit $|f(x_n) - g| < \epsilon\ \forall\ n > N$, oder äquivalent $\lim\limits_{n\to\infty} f(x_n) = g$. Dies ist die behauptete Grenzwertaussage (3.1).

(b) Es gelte nun die Grenzwertaussage (3.1). Wäre (3.3) nicht erfüllt, so hätten wir im Gegenteil

$$\exists\, \epsilon_0 > 0\, \forall\, k \in \mathbf{N}\ :\ |f(x_k) - g| \geq \epsilon_0\ \text{ für ein } x_k \in D(f) \text{ mit } \ 0 < |x_k - x_0| < 10^{-k}.$$

Es gäbe somit eine konvergente Folge $(x_k)_{k\in\mathbf{N}} \subset D(f)\backslash\{x_0\}$, $\lim\limits_{k\to\infty} x_k = x_0$, mit $|f(x_k) - g| \geq \epsilon_0\ \forall\ k \in \mathbf{N}$, was im Widerspruch zu (3.1) steht. □

Bemerkung 6.5 (a) In der Bedingung (3.3) darf anstelle der Größe 10^{-k} eine beliebige Zahl $\delta > 0$ stehen. In diesem Fall ist der Ausdruck $\exists\ k \in \mathbf{N}$ zu ersetzen durch $\exists\ \delta = \delta(\epsilon) > 0$. Alles weitere bleibt unverändert.

(b) In der Definition (3.3) wird weder verlangt, dass ein Funktionswert $f(x_0)$ existiert, noch muss $f(x_0) = g$ gelten. Dies wird auch in den Grafiken auf der folgenden Seite veranschaulicht.

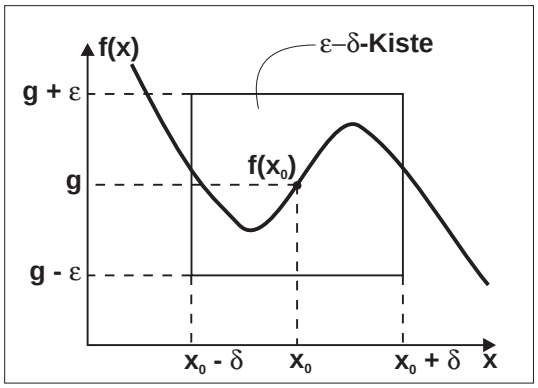
(c) Satz 6.7 gilt auch mit entsprechender Modifikation für die *einseitigen Grenzwerte* $g^{\pm}$. Die Bedingung (3.3) ist wie folgt zu ersetzen:

$$\boxed{\begin{array}{c} \lim\limits_{x\to x_0+} f(x) = g^+ \quad \Leftrightarrow \\ \forall\, \epsilon > 0\ \exists\, k \in \mathbf{N}\ :\ |f(x) - g^+| < \epsilon\ \ \forall\, x \in D(f)\ \text{ mit }\ x_0 < x < x_0 + 10^{-k}. \end{array}} \tag{3.4}$$

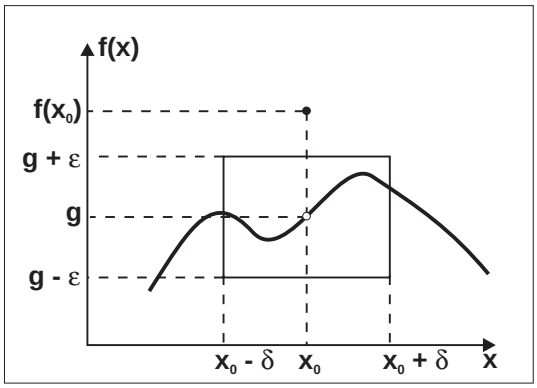
$$\boxed{\begin{array}{c}\lim\limits_{x\to x_0-} f(x) = g^- \quad \Leftrightarrow \\ \forall\,\epsilon>0\;\exists\,k\in\mathbf{N}\;:\;|f(x)-g^-|<\epsilon\;\;\forall\,x\in D(f)\;\text{ mit }\;x_0-10^{-k}<x<x_0.\end{array}} \tag{3.5}$$

(d) Gelegentlich verwenden wir für einseitige Grenzwerte auch die Symbolik □

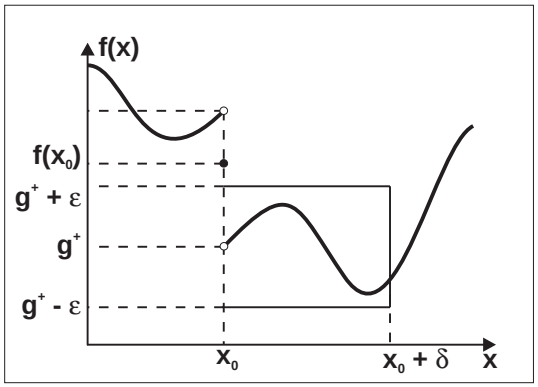
$$\lim_{x\to x_0+} f(x) = f(x_0+0), \quad \lim_{x\to x_0-} f(x) = f(x_0-0).$$



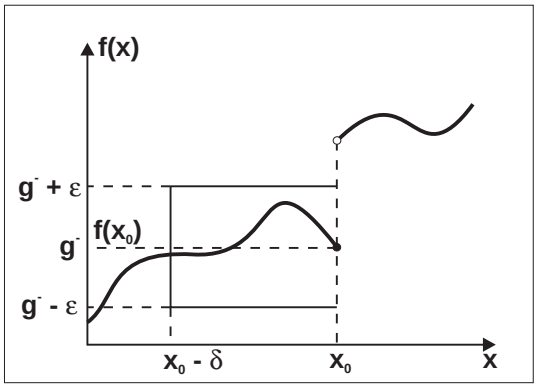
$\lim\limits_{x\to x_0} f(x) = g$, **wobei** $f(x_0)$ **nicht definiert ist**



$\lim\limits_{x\to x_0} f(x) = g$, **wobei** $f(x_0)\neq g$ **gilt**



rechtsseitiger Grenzwert $\lim\limits_{x\to x_0+} f(x) = g^+$ **mit** $g^+\neq f(x_0)$



linksseitiger Grenzwert $\lim\limits_{x\to x_0-} f(x) = g^-$ **mit** $g^- = f(x_0)$

Im Zusammenhang mit einseitigen Grenzwerten treffen wir die folgende

Definition 6.14 (a) *Existieren in einem Punkt* $x_0\in\mathbf{R}$ *voneinander verschiedene rechts– bzw. linksseitige Grenzwerte* $\lim\limits_{x\to x_0\pm} f(x) = g^\pm\in\mathbf{K}$, *so sagen wir, die Funktion* f *hat bei* x_0 *einen* **Sprung der Höhe** $|g^+-g^-|$.

(b) *Ein Punkt* $x_0\in\mathbf{R}$ *heiße* **Unbestimmtheitsstelle** *oder* **singuläre Stelle** *oder kurz* **Singularität** *von* f, *wenn wenigstens einer der Grenzwerte* g^+ *oder* g^- *in* $\mathbf{K}$ *nicht existiert.*

Singularitäten treten bei rationalen Funktionen $\frac{P(x)}{Q(x)}$ in den Nullstellen des Nennerpolynoms $Q(x)$ auf, sofern diese nicht gleichzeitig Nullstellen von $P(x)$ mindestens derselben Ordnung sind. *Zahlenbeispiel:*

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

Hier ist also $x_0 = 1$ eine Singularität.

Beim Gebrauch der Grenzwertbedingung (3.3) (und analog der Bedingungen (3.4), (3.5)) kommt es meistens auf ein geschicktes Abschätzen des Ausdrucks $|f(x)-g|$ durch einen Term der Form $|x-x_0|$ an. Die folgende Strategie muss sequentiell von links nach rechts verfolgt werden:

$$|f(x)-g| \quad \underbrace{\leq \ \cdots \ \leq \ \cdots \ \leq \ \cdots \ \leq \ \cdots \ \leq}_{\substack{\text{Die Ausdrücke } \cdots \text{ müssen für} \\ x \to x_0 \text{ nach 0 konvergieren.}}} \quad \underbrace{A(|x-x_0|) < \epsilon}_{\substack{\text{Diese Ungleichung muss nach} \\ |x-x_0| \text{ aufgelöst werden können.}}}$$

BSP. (6.3.7) Es sei

$$\boxed{f(x) := x^n,\ x \in D(f) := \mathbf{R},\ n = 0, 1, 2, \ldots}$$

Für die Grenzwertberechnung ist die folgende Ungleichung von fundamentaler Bedeutung. Wir wählen ein festes $x_0 \in \mathbf{R}$ und setzen $0 < \delta := 10^{-k} \leq 1$. Wir betrachten nur solche $x \in \mathbf{R}$ mit $0 < |x - x_0| < \delta$:

$$\begin{aligned} |x^n - x_0^n| &= |(x - x_0 + x_0)^n - x_0^n| = \left| \sum_{j=1}^{n} \binom{n}{j} (x-x_0)^j \, x_0^{n-j} \right| \\ &\leq \delta \sum_{j=0}^{n} \binom{n}{j} |x_0|^{n-j} = \delta \, (1 + |x_0|)^n. \end{aligned} \tag{3.6}$$

Wir zeigen nun unter Verwendung dieser Ungleichung $\lim\limits_{x \to x_0} f(x) = x_0^n$. In der Tat, für jedes $\epsilon > 0$ gilt

$$|x^n - x_0^n| \overset{(3.6)}{\leq} 10^{-k} \, (1 + |x_0|)^n < \epsilon \ \ \forall \, 0 < |x - x_0| < 10^{-k},$$

sofern wir $k = k(\epsilon) \in \mathbf{N}$ so groß wählen, dass $0 < \frac{1}{\epsilon} \, (1 + |x_0|)^n < 10^k$ gilt.

Beachte: Die Zahl k hängt zwar primär von ϵ ab, aber auch vom Punkt x_0 und von der Funktion f.

BSP. (6.3.8) Es sei

$$\boxed{f(x) := \frac{1}{x^n},\ x \in D(f) := \mathbf{R} \setminus \{0\},\ n = 1, 2, 3, \ldots}$$

Wir zeigen $\lim\limits_{x \to x_0} f(x) = \frac{1}{x_0^n} \ \forall \ x_0 \neq 0$. Wählen wir nämlich vorab $k \in \mathbf{N}$ so, dass $10^{-k} \leq \frac{1}{2} \, |x_0|$ gilt, so haben wir in der Tat für jedes $\epsilon > 0$ und für $0 < |x - x_0| < 10^{-k}$:

(i) $|x| = |x - x_0 + x_0| \geq |x_0| - |x - x_0| > |x_0| - 10^{-k} \geq \frac{1}{2} \, |x_0|$,

(ii) $\left| \dfrac{1}{x^n} - \dfrac{1}{x_0^n} \right| = \dfrac{|x_0^n - x^n|}{|x|^n |x_0|^n} \overset{(i)}{\leq} \dfrac{2^n |x^n - x_0^n|}{|x_0|^{2n}} \overset{(3.6)}{\leq} \dfrac{2^n \, 10^{-k}}{|x_0|^{2n}} \, (1 + |x_0|)^n < \epsilon,$

sofern wir die Zahl $k = k(\epsilon) \in \mathbf{N}$ so groß gewählt haben, dass $0 < \frac{2^n}{\epsilon |x_0|^{2n}} \, (1 + |x_0|)^n < 10^k$ gilt.

BSP. (6.3.9) Es sei

$$\boxed{f(x) := \sqrt[p]{x},\ x \in D(f) := [0, +\infty),\ p \in \mathbf{N}.}$$

Wir behaupten $\lim\limits_{x \to 0+} f(x) = 0 = f(0)$. In der Tat, für fest gewähltes $\epsilon > 0$ gilt:

$$|f(x) - f(0)| = \sqrt[p]{x} = x^{1/p} < 10^{-k/p} < \epsilon \ \ \forall \, 0 < x < 10^{-k},$$

sofern wir die Zahl $k = k(\epsilon) \in \mathbf{N}$ so groß wählen, dass $0 < \epsilon^{-p} < 10^k$ gilt.

Beachte: Es macht hier keinen Sinn, nach dem Grenzwert $\lim\limits_{x \to 0-} f(x)$ zu fragen, da $f(x)$ für $x < 0$ nicht definiert ist. Für $x_0 > 0$ zeigt man ganz analog wie im Falle $x_0 = 0$ den Grenzwert $\lim\limits_{x \to x_0} f(x) = \sqrt[p]{x_0}$.

BSP. (6.3.10) Es sei

$$\boxed{f(x) := x\sqrt{1 + \tfrac{1}{x^2}} = (\operatorname{sign} x)\sqrt{1 + x^2},\ x \in D(f) := \mathbf{R} \setminus \{0\}.}$$

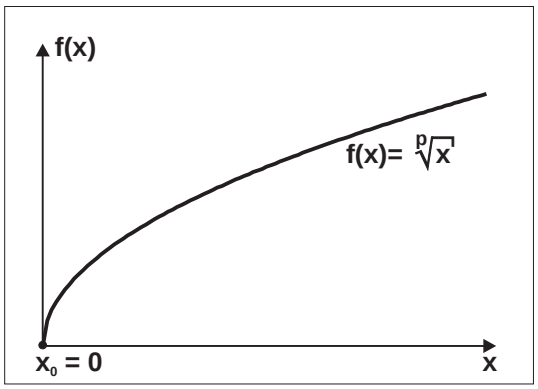
Wir behaupten

$$\lim_{x \to 0+} f(x) = +1, \quad \lim_{x \to 0-} f(x) = -1.$$

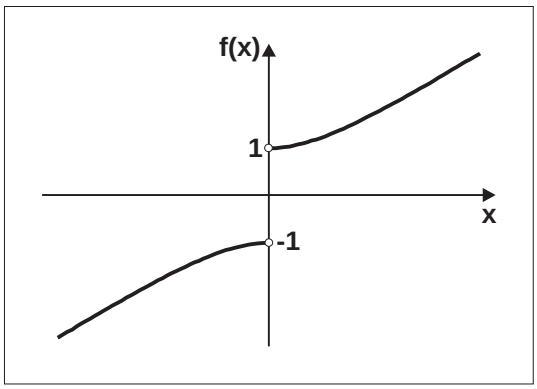
Zum Beweis der ersten Behauptung wählen wir ein beliebiges $\epsilon > 0$ fest. Dann erhalten wir für jedes $0 < x < 10^{-k}$:

$$|f(x) - 1| = |\,+\sqrt{1 + x^2} - 1| = \frac{x^2}{1 + \sqrt{1 + x^2}} \le \frac{x^2}{2} < 0.5 \cdot 10^{-2k} < \epsilon,$$

sofern wir die Zahl $k = k(\epsilon) \in \mathbf{N}$ so groß wählen, dass $\frac{1}{\sqrt{2\epsilon}} < 10^k$ gilt. Die zweite Behauptung beweist man ganz analog.



Der Graph der Funktion $f(x) = \sqrt[p]{x}$



Der Graph der Funktion $f(x) = x\sqrt{1 + 1/x^2}$

BSP. (6.3.11) Es sei

$$\boxed{f(x) := e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!},\ x \in D(f) := \mathbf{R}.}$$

Wie wir bereits in Abschnitt 3.2, BSP. (3.2.5), gezeigt haben, konvergiert diese Reihe absolut an jeder Stelle $x_0 \in \mathbf{R}$ gegen den Summenwert $e^{x_0} \in \mathbf{R}$. Wir behaupten $\lim\limits_{x \to x_0} e^x = e^{x_0}\ \forall\, x_0 \in \mathbf{R}$. Zum Beweis verwenden wir die folgende Ungleichung, die für $0 < |x - x_0| < 10^{-k} \le 1$ richtig ist:

$$|e^x - e^{x_0}| = e^{x_0}\,|e^{x-x_0} - 1| = e^{x_0}\left|\sum_{j=1}^{\infty} \frac{(x - x_0)^j}{j!}\right| \le 10^{-k}\, e^{x_0} \sum_{j=0}^{\infty} \frac{1}{j!} = 10^{-k}\, e^{1+x_0}. \qquad (3.7)$$

Nun folgt für jedes feste $\epsilon > 0$:

$$|e^x - e^{x_0}| \overset{(3.7)}{\le} 10^{-k}\, e^{1+x_0} < \epsilon\ \ \forall\, 0 < |x - x_0| < 10^{-k},$$

sofern wir die Zahl $k = k(\epsilon) \in \mathbf{N}$ so groß wählen, dass $e^{1+x_0}/\epsilon < 10^k$ gilt.

Aus der Konvergenztheorie der Zahlenfolgen (vgl. Abschnitt 3.1) ergeben sich einige allgemeingültige Grundtatsachen über das Verhalten von Funktionenlimites, die wir hier auflisten.

- Funktionenlimites sind **eindeutig**, falls sie existieren.

- Gilt $\lim\limits_{x\to x_0} f(x) = g \in \mathbf{K}$, so ist $f(x)$ in der *Umgebung* von x_0 **beschränkt**. Das heißt, sind ϵ und k gemäß der Vorschrift (3.3) gewählt, so gilt $|f(x)| < |g| + \epsilon \;\forall\; x \in D(f)$ mit $0 < |x - x_0| < 10^{-k}$.

- Die eingeführten Grenzwertbegriffe bleiben auch für Funktionen $f : D(f) \to Y$ gültig, wenn Y ein **normierter** Vektorraum über dem Körper $\mathbf{K}$ ist, insbesondere also für $Y := \mathbf{K}^n$ und $Y := \mathbf{K}^{(m,n)}$. In dem Fall hat man in (3.3)–(3.5) die Beträge bei $|f(x) - g^{(\pm)}| < \epsilon$ durch die Y–Normen zu ersetzen: $\|f(x) - g^{(\pm)}\| < \epsilon$.

- Die **Limesbildung** kann mit **algebraischen Operationen** verknüpft werden. Sofern die Grenzwerte
$$F := \lim_{x\to x_0} f(x), \quad G := \lim_{x\to x_0} g(x)$$
existieren, gelten die folgenden Regeln:

$$\boxed{\lim_{x\to x_0} [\alpha\, f(x) + \beta\, g(x)] = \alpha\, F + \beta\, G \;\; \forall\; \alpha, \beta \in \mathbf{K}. \quad (\textbf{Linearität})} \tag{3.8}$$

$$\boxed{\lim_{x\to x_0} [f(x)\, g(x)] = F \cdot G, \quad \lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{F}{G}, \quad \text{falls } G \neq 0.} \tag{3.9}$$

- Sind f, g, h **reellwertige** Funktionen, so kann die **Limesbildung** mit **Ordnungsrelationen** verknüpft werden:

$$\boxed{\text{Aus } f(x) < M \;\forall\; x \in D(f) \text{ folgt } \lim_{x\to x_0} f(x) \leq M.} \tag{3.10}$$

$$\boxed{\text{Aus } f(x) \leq g(x) \;\forall\; x \in D(f) \cap D(g) \text{ folgt } \lim_{x\to x_0} f(x) \leq \lim_{x\to x_0} g(x).} \tag{3.11}$$

Einschließungskriterium. Gelten $f(x) \leq g(x) \leq h(x)$ und $\lim\limits_{x\to x_0} f(x) = g = \lim\limits_{x\to x_0} h(x)$, so folgt:

$$\boxed{\lim_{x\to x_0} g(x) = g.} \tag{3.12}$$

BSP. (6.3.12) Es sei

$$\boxed{f(x) := \frac{x^3 + |x+1| + \operatorname{sign}(x+1)}{\operatorname{sign} x},\; x \in D(f) := \mathbf{R} \setminus \{0\}.}$$

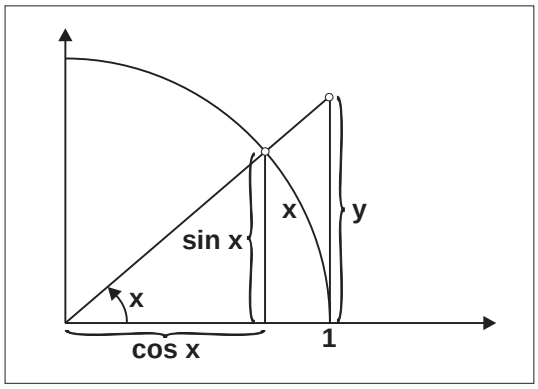
Unter Verwendung der Regeln (3.8) und (3.9) erhalten wir:

$$\begin{aligned} \lim_{x\to 0+} f(x) &= \frac{0+1+1}{1} = 2, & \lim_{x\to 0-} f(x) &= \frac{0+1+1}{-1} = -2, \\ \lim_{x\to (-1)+} f(x) &= \frac{-1+0+1}{-1} = 0, & \lim_{x\to (-1)-} f(x) &= \frac{-1+0-1}{-1} = 2. \end{aligned}$$

BSP. (6.3.13) Es sei $0 < x < \frac{\pi}{2}$. Aus dem Strahlensatz resultiert gemäß Skizze $y = \frac{\sin x}{\cos x}$, und man hat offenbar die Ungleichung

$$0 < \sin x < x < \frac{\sin x}{\cos x}.$$

Aus dieser folgt $\frac{1}{\sin x} > \frac{1}{x} > \frac{\cos x}{\sin x}$, und somit $1 > \frac{\sin x}{x} > \cos x$.



Zur Ungleichung $0 < \sin x < x < \frac{\sin x}{\cos x}$

Nimmt man den Grenzwert $\lim\limits_{x \to 0+} \cos x = 1$ als bewiesen an, so folgt aus dem Einschließungskriterium (3.12) der wichtige Grenzwert:

$$\boxed{\lim_{x \to 0+} \frac{\sin x}{x} = 1 = \lim_{x \to 0-} \frac{\sin x}{x}.}$$

6.4 Uneigentliche Grenzwerte von Funktionen einer reellen Veränderlichen

Ist $f(x)$ auf den unbeschränkten Intervallen $(b, +\infty)$ bzw. $(-\infty, a)$ erklärt, so definiert man Grenzwerte $\lim\limits_{x \to \pm\infty} f(x)$ in der folgenden Weise:

Definition 6.15 *Die Funktion f habe in $+\infty$ (bzw. in $-\infty$) den* **Grenzwert** *g, wenn gilt:*

$$\boxed{\forall\, \epsilon > 0\ \exists\, k \in \mathbf{N}\ :\ |f(x) - g| < \epsilon\ \ \forall\, x > 10^k > b\ \ (bzw.\ \forall\, x < -10^k < a).} \qquad (4.1)$$

Hierfür schreibt man $\lim\limits_{x \to +\infty} f(x) = g$ *(bzw.* $\lim\limits_{x \to -\infty} f(x) = g$*).*

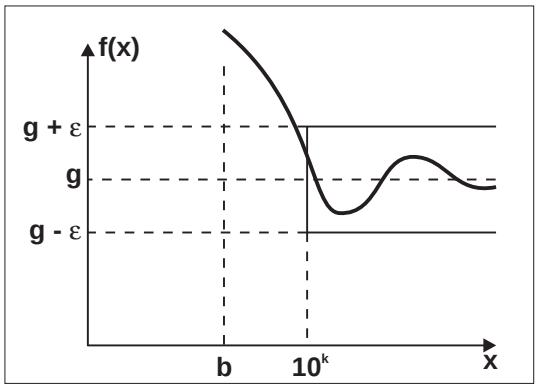
BSP. (6.4.1) Es sei

$$\boxed{f(x) := \frac{x}{x-2} = 1 + \frac{2}{x-2},\ x \in D(f) := \mathbf{R} \setminus \{2\}.}$$

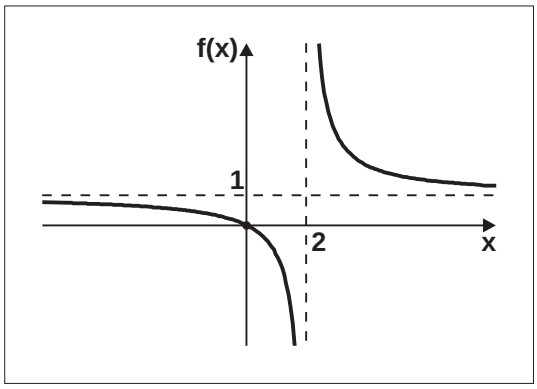
Wir behaupten $\lim\limits_{x \to +\infty} f(x) = 1$. In der Tat, für $x > 2$ haben wir

$$|f(x) - 1| = |\frac{x}{x-2} - 1| = \frac{2}{x-2} \leq \frac{2}{10^k - 2} < \epsilon\ \ \forall\, x > 10^k > 2,$$

sofern die Zahl $k = k(\epsilon) \in \mathbf{N}$ so groß gewählt wird, dass $\frac{2}{\epsilon} + 2 < 10^k$ gilt. Ganz analog zeigt man $\lim\limits_{x \to -\infty} f(x) = 1$.



Uneigentlicher Grenzwert $\lim\limits_{x\to+\infty} f(x) = g$



Der Graph der Funktion $f(x) = \frac{x}{x-2}$

Für **reellwertige Funktionen** f können auch die uneigentlichen Grenzwerte $f(x) \to \pm\infty$ erklärt werden.

Definition 6.16 *Die Funktion* $f : D(f) \to \mathbf{R}$ *habe in* $x_0 \in \mathbf{R}$ *den* **uneigentlichen Grenzwert** $+\infty$ *(bzw.* $-\infty$*), wenn gilt*

$$\boxed{\forall\, \epsilon > 0\ \exists\, k \in \mathbf{N}\ :\ f(x) > \frac{1}{\epsilon}\ \ (bzw.\ < -\frac{1}{\epsilon})\ \ \forall\, x \in D(f)\ \ mit\ \ 0 < |x - x_0| < 10^{-k}.} \qquad (4.2)$$

Hierfür schreibt man $\lim\limits_{x\to x_0} f(x) = +\infty$ *(bzw.* $\lim\limits_{x\to x_0} f(x) = -\infty$*). Analog erklärt man die* **rechts–** *bzw.* **linksseitigen** *uneigentlichen Grenzwerte* $\lim\limits_{x\to x_0+} f(x) = \pm\infty$ *bzw.* $\lim\limits_{x\to x_0-} f(x) = \pm\infty$*.*

Beachte: Mit $(\pm)\infty$ wird keine *Zahl* erklärt sondern ein bestimmtes Grenzverhalten symbolisiert.

BSP. (6.4.2) Die Funktion $f(x) := \frac{x}{x-2}$ sei wie in BSP. (6.4.1) vorgelegt. Wir behaupten $\lim\limits_{x\to 2+} f(x) = +\infty$ und $\lim\limits_{x\to 2-} f(x) = -\infty$. Dazu betrachten wir ein beliebiges, aber festes $\epsilon > 0$. Es gilt

$$f(x) = \frac{x}{x-2} \quad \begin{cases} > \dfrac{2}{10^{-k}} > \dfrac{1}{10^{-k}} > \dfrac{1}{\epsilon}\ \ \forall\, x \in \mathbf{R}\ \text{ mit }\ 2 < x < 2 + 10^{-k}, \\ = \dfrac{-x}{2-x} < -\dfrac{1}{10^{-k}} < -\dfrac{1}{\epsilon}\ \ \forall\, x \in \mathbf{R}\ \text{ mit }\ 1 < 2 - 10^{-k} < x < 2, \end{cases}$$

sofern die Zahl $k = k(\epsilon) \in \mathbf{N}$ so groß gewählt wird, dass $10^k > \frac{1}{\epsilon}$ gilt.

BSP. (6.4.3) Die folgenden uneigentlichen Grenzwerte lassen sich einfach bestimmen:

$$\lim_{x\to+\infty} x^n = \begin{cases} 1 & :\ n = 0, \\ +\infty & :\ n \in \mathbf{N}, \end{cases} \qquad \lim_{x\to-\infty} x^n = \begin{cases} 1 & :\ n = 0, \\ +\infty & :\ n = 2m \text{ gerade}, \\ -\infty & :\ n = 2m-1 \text{ ungerade}, \end{cases} \qquad m \in \mathbf{N}.$$

Für alle $m \in \mathbf{N}$ gilt:

$$\lim_{x\to\pm\infty} \frac{1}{x^m} = 0,\ \ \lim_{x\to 0} \frac{1}{x^{2m}} = +\infty,\ \ \lim_{x\to 0+} \frac{1}{x^{2m-1}} = +\infty,\ \ \lim_{x\to 0-} \frac{1}{x^{2m-1}} = -\infty.$$

Rechenregeln. (a) Für die uneigentlichen Grenzwerte $\lim\limits_{x\to\pm\infty} f(x) = g \in \mathbf{K}$ gelten wiederum die Regeln (3.8) und (3.9) aus Abschnitt 6.3 bezüglich der algebraischen Verknüpfungen "$\pm, \cdot, :$". Für reellwertige Funktionen $f : D(f) \to \mathbf{R}$ bleiben die Regeln (3.10) und (3.11) aus Abschnitt 6.3 ebenfalls richtig.

(b) Für die uneigentlichen Grenzwerte $\lim\limits_{x\to x_0(\pm)} f(x) = \pm\infty$ oder $\lim\limits_{x\to\pm\infty} f(x) = (\pm)\infty$ treten neue Regeln hinzu, die wir hier tabellarisch zusammenstellen wollen. Nachfolgend bezeichnen wir summarisch mit $\lim f(x)$ jeden der möglichen Fälle $\lim\limits_{x\to x_0\pm} f(x)$ oder $\lim\limits_{x\to\pm\infty} f(x)$. Es seien f, g, h **reellwertige** Funktionen, und wir setzen jeweils voraus, dass die folgenden uneigentlichen Grenzwerte existieren:

$$\lim f(x) = +\infty = \lim h(x), \quad \lim g(x) = g \in \mathbf{R}.$$

	Limes–Regel	Formale Rechenregel
(1)	$\lim[f(x) + \alpha\, g(x)] = +\infty \ \ \forall\, \alpha \in \mathbf{R}$	$\infty + r = \infty \ \ \forall\, r \in \mathbf{R}$
(2)	$\lim[f(x)\, g(x)] = +\infty$ falls $g > 0$	$\infty \cdot r = \infty \ \ \forall\, r > 0$
(3)	$\lim[f(x)\, h(x)] = +\infty = \lim[f(x) + h(x)]$	$\infty \dot{+} \infty = \infty$
(4)	$\lim \dfrac{g(x)}{f(x)} = 0$	$\dfrac{r}{\infty} = 0 \ \ \forall\, r \in \mathbf{R}$
(5)	$\lim \dfrac{f(x)}{g(x)} = +\infty$ falls $g > 0$	$\dfrac{\infty}{r} = \infty \ \ \forall\, r > 0$

(4.3)

Bemerkung 6.6 (a) Die Regeln (1) und (4) bleiben selbst dann noch richtig, wenn $\lim g(x)$ nicht existiert, wenn aber $|g(x)|$ beim Grenzübergang $x \to \binom{x_0}{\pm\infty}$ **beschränkt** bleibt; zum Beispiel $g(x) := \sin x$ für $x \to \pm\infty$.

(b) Es fehlen hier noch Rechenregeln für die **unbestimmten Ausdrücke**

$$\infty - \infty, \quad 0 \cdot \infty, \quad \frac{\infty}{\infty}, \quad \frac{0}{0},$$

die wir in Abschnitt 7.5 herleiten werden. Diese Rechenregeln können i.a. nicht durch algebraische Operationen aus den Grenzwerten der einzelnen Funktionen erschlossen werden. □

BSP. (6.4.4) $\displaystyle\lim_{x\to 0+} \frac{\operatorname{sign} x}{\sqrt{1+\frac{1}{x}}} = \frac{(\text{beschränkt})}{+\infty} \overset{(4)}{=} 0, \quad \lim_{x\to\pm\infty} \frac{\operatorname{sign} x}{\sqrt{1+\frac{1}{x}}} = \frac{\pm 1}{1} = \pm 1.$

BSP. (6.4.5) $\displaystyle\lim_{x\to\pm\infty} \frac{x^3 - 15x}{x^3 + 15x} \left(= \frac{\infty}{\infty}\right) = \lim_{x\to\pm\infty} \frac{1 - \frac{15}{x^2}}{1 + \frac{15}{x^2}} = \frac{1-0}{1+0} = 1.$

Allgemeiner untersuchen wir $\lim\limits_{x\to+\infty} \frac{P_n(x)}{Q_m(x)}$ für Polynome $P_n(x) = \sum\limits_{k=0}^{n} a_k x^k, a_n \neq 0$, und $Q_m(x) = \sum\limits_{k=0}^{m} b_k x^k, b_m \neq 0$. Nach Division durch x^m resultiert:

$$\lim_{x\to+\infty} \frac{P_n(x)}{Q_m(x)} \quad = \quad \lim_{x\to+\infty} \frac{a_n x^{n-m} + a_{n-1} x^{n-m-1} + \cdots + a_1 x^{1-m} + a_0 x^{-m}}{b_m + b_{m-1} x^{-1} + \cdots + b_1 x^{1-m} + b_0 x^{-m}}$$

$$= \begin{cases} 0 & : \ m > n, \\ \frac{a_n}{b_m} & : \ m = n, \\ +\infty \cdot \operatorname{sign}\left(\frac{a_n}{b_m}\right) & : \ m < n. \end{cases}$$

BSP. (6.4.6) Die Funktion $f(x) := e^x$, $x \in D(f) := \mathbf{R}$ hat die folgenden Grenzwerte:

$$\boxed{\lim_{x \to -\infty} e^x = 0, \qquad \lim_{x \to +\infty} e^x = +\infty.}$$

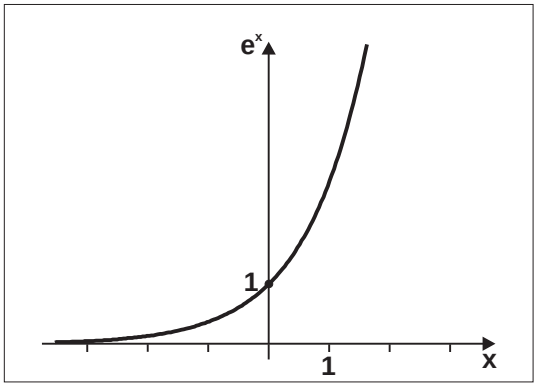
Es gelten nämlich die Ungleichungen

$$e^x > 0 \ \ \forall\, x \in \mathbf{R}, \qquad 1 + x < \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x \ \ \forall\, x > 0.$$

Aus diesen Ungleichungen erschließen wir

$$0 \leq \lim_{x \to -\infty} e^x = \lim_{y \to +\infty} e^{-y} = \lim_{y \to +\infty} \frac{1}{e^y} \leq \lim_{y \to +\infty} \frac{1}{1+y} = 0,$$

und in gleicher Weise $\lim\limits_{x \to +\infty} e^x \geq \lim\limits_{x \to +\infty} (1 + x) = +\infty$.



Der Graph der Exponentialfunktion e^x

6.5 Stetigkeit von Funktionen einer reellen Veränderlichen

Es bezeichne f wiederum eine Funktion der reellen Veränderlichen x mit Werten in dem Körper $\mathbf{K} := \mathbf{R}$ oder $\mathbf{K} := \mathbf{C}$:

$$f : \begin{cases} D(f) \to \mathbf{K}, \\ x \mapsto f(x), \end{cases} \qquad \emptyset \neq D(f) \subset \mathbf{R}.$$

Definition 6.17 *Eine Funktion* $f \in \operatorname{Abb}(\mathbf{R}, \mathbf{K})$ *heiße* **stetig im Punkte** $x_0 \in D(f)$, *falls gilt:*

$$\boxed{\lim_{x \to x_0} f(x) = f(x_0).}$$

Ist f *in* **jedem Punkte** $x_0 \in D(f)$ *stetig, so heiße* f **stetig** *(auf* $D(f)$*). Ist die Funktion* f *in einem Punkte* $x_0 \in D(f)$ *nicht stetig, so heiße sie* **unstetig** *bei* x_0*. Besitzt* f *in* $x_0 \in D(f)$ *lediglich den* **rechtsseitigen** *(bzw. den* **linksseitigen***) Grenzwert* $\lim\limits_{x \to x_0+} f(x) = f(x_0)$ *(bzw.* $\lim\limits_{x \to x_0-} f(x) = f(x_0)$*), so heiße* f *in* x_0 **rechtsseitig stetig** *(bzw.* **linksseitig stetig***).*

Bemerkung 6.7 (a) Anders als bei der Definition von Funktionenlimites **muss** der Punkt x_0 bei Stetigkeitsbetrachtungen zum Definitionsbereich $D(f)$ gehören. Das heißt, es muss ein Funktionswert $f(x_0)$ existieren.

(b) Die Stetigkeit **reellwertiger** Funktionen kann häufig durch Betrachtung des Graphen $G(f)$ geprüft werden. Kann $G(f)$ in einem Zug ohne Absetzen des Zeichenstiftes gezeichnet werden, so ist die zugeordnete Funktion f stetig. Diese Vorstellung darf aber nicht als Definition der Stetigkeit betrachtet werden. Zum Beispiel ist die Funktion $f(x) := \sin\frac{1}{x}$ auf dem Intervall $(0, +\infty)$ stetig; ihr Graph ist jedoch nicht zeichenbar. □

Unter Verwendung von Satz 6.7 kann die Stetigkeit einer Funktion f wieder durch die $\epsilon - \delta$–Kiste ausgedrückt werden. Wir wählen nun stets anstelle der Zahl $k = k(\epsilon) \in \mathbf{N}$ ein $\delta = \delta(\epsilon) \leq 10^{-k}$.

Satz 6.8 ($\epsilon - \delta$–Definition der Stetigkeit)
Eine Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{K})$ ist genau dann im Punkte $x_0 \in D(f)$ stetig, wenn gilt:

$$\boxed{\forall\, \epsilon > 0\, \exists\, \delta = \delta(\epsilon) > 0 \,:\, |f(x) - f(x_0)| < \epsilon \;\; \forall\, x \in D(f) \;\; mit \;\; 0 < |x - x_0| < \delta.} \tag{5.1}$$

Analoge Formulierungen gelten für die rechts– bzw. linksseitige Stetigkeit in x_0. **Beachte:** *Die Zahl δ hängt i.a. auch vom Punkt x_0 ab.*

BSP. (6.5.1) Wir hatten für die Funktionen $f(x) := x^n$, $\frac{1}{x^n}$, e^x in Abschnitt 6.3 gezeigt, dass $\lim\limits_{x \to x_0} f(x) = f(x_0)$ für alle $x_0 \in D(f)$ gilt, vgl. BSP. (6.3.7), BSP. (6.3.8) und BSP. (6.3.11). Die konstante Funktion $f(x) = c \in \mathbf{K}$ erfüllt wegen $|f(x) - f(x_0)| = |c - c| = 0$ trivialerweise die Bedingung (5.1).

$$\boxed{\text{Es ist stetig } f(x) := \begin{cases} x^n & \forall\, x \in \mathbf{R},\; n \in \mathbf{N}, \\ x^{-n} & \forall\, x \in \mathbf{R} \setminus \{0\},\; n \in \mathbf{N}, \\ e^x & \forall\, x \in \mathbf{R}, \\ c & \forall\, x \in \mathbf{R},\; c \in \mathbf{K}. \end{cases}}$$

BSP. (6.5.2) Die Betragsfunktion

$$\boxed{f(x) := |x|,\; x \in D(f) := \mathbf{R},}$$

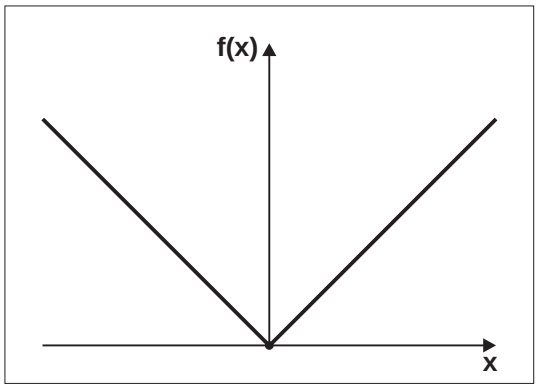
ist **stetig** auf ganz $D(f)$. Zum Nachweis der Stetigkeit verwenden wir die Dreiecksungleichung

$$\Big||x| - |x_0|\Big| \leq |x - x_0|. \tag{5.2}$$

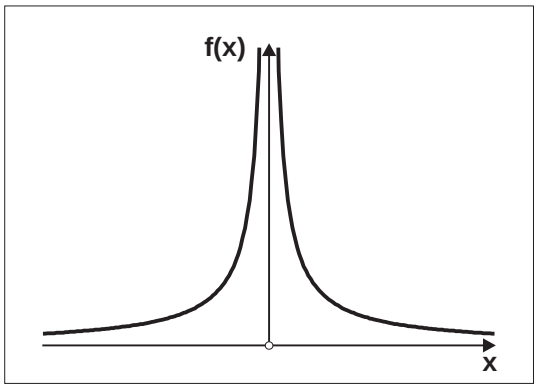
Für jedes $x_0 \in \mathbf{R}$ und für jede Zahl $\epsilon > 0$ gilt

$$|f(x) - f(x_0)| \overset{(5.2)}{\leq} |x - x_0| < \delta := \epsilon \;\; \forall\, x \in \mathbf{R} \;\; \text{mit} \;\; 0 < |x - x_0| < \delta.$$

Es ist zu beachten, dass die Zahl $\delta = \delta(\epsilon) = \epsilon$ hier **gleichmäßig** bezüglich $x_0 \in \mathbf{R}$ wählbar ist, also **nicht** von x_0 abhängt.



Der Graph der Funktion $f(x) := |x|$



Der Graph der Funktion $f(x) := \frac{1}{|x|}$

BSP. (6.5.3) Die Funktion

$$\boxed{f(x) := \frac{1}{|x|},\ x \in D(f) := \mathbf{R} \setminus \{0\}}$$

ist **stetig** auf ganz $D(f)$. Wir wählen $x_0 \neq 0$ und $\epsilon > 0$ fest. Danach definieren wir die Zahl $\delta(\epsilon) := \frac{1}{2}\ \min\{\epsilon\,|x_0|^2, |x_0|\}$. Wegen $\delta \leq |x_0|/2$ folgt zunächst $\frac{1}{2}\,|x_0| < |x| < \frac{3}{2}\,|x_0|$ für alle x mit $0 < |x - x_0| < \delta$. Hieraus ergibt sich nun unter der Einschränkung $0 < |x - x_0| < \delta$:

$$|f(x) - f(x_0)| = \frac{||x| - |x_0||}{|x||x_0|} \overset{(5.2)}{\leq} \frac{|x - x_0|}{|x||x_0|} < \frac{2|x - x_0|}{|x_0|^2} < \frac{2\epsilon\,|x_0|^2}{2\,|x_0|^2} = \epsilon.$$

Wir vermerken, dass die Zahl δ hier sowohl von ϵ als auch von x_0 abhängt.

BSP. (6.5.4) Eine Funktion $f : I \to \mathbf{K}$, die auf einem Intervall $I \subset \mathbf{R}$ der LIPSCHITZ–**Bedingung**

$$|f(x) - f(x_0)| \leq L\,|x - x_0| \quad \forall\, x, x_0 \in I, \tag{5.3}$$

genügt, ist offenbar **stetig** auf ganz I. Wählt man nämlich $\delta = \delta(\epsilon) := \epsilon/L$, so folgt die Stetigkeitsbedingung (5.1) direkt aus (5.3). Wir betrachten hier speziell die Funktionen $f(x) := \sin x$ und $f(x) := \cos x,\ x \in D(f) := \mathbf{R}$. Aus BSP. (6.3.13), Abschnitt 6.3, erhalten wir $|\sin x| \leq |x|\ \forall\, x \in \mathbf{R}$. Unter Verwendung der Additionstheoreme der trigonometrischen Funktionen resultiert:

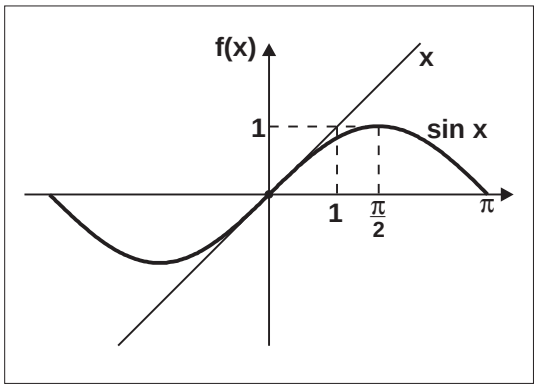
$$\sin x - \sin x_0 = 2\,\cos\frac{x + x_0}{2}\sin\frac{x - x_0}{2}, \quad \cos x - \cos x_0 = -2\,\sin\frac{x + x_0}{2}\sin\frac{x - x_0}{2}. \tag{5.4}$$

Somit folgt für alle $x, x_0 \in \mathbf{R}$

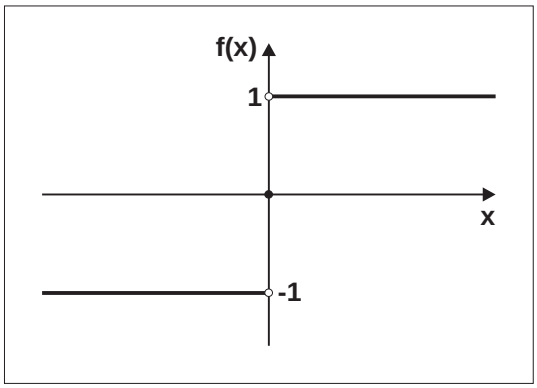
$$|\sin x - \sin x_0| \leq 2\,\left|\sin\frac{x - x_0}{2}\right| \leq |x - x_0|, \quad |\cos x - \cos x_0| \leq 2\,\left|\sin\frac{x - x_0}{2}\right| \leq |x - x_0|,$$

und das ist die LIPSCHITZ–Bedingung (5.3) mit der Konstanten $L = 1$. Wir haben zusammenfassend:

$$\boxed{\sin x \text{ und } \cos x \text{ sind auf ganz } \mathbf{R} \text{ stetige Funktionen.}}$$



Der Graph der Funktion $f(x) := \sin x$



Die Signums–Funktion

Definition 6.18 (a) *Eine Funktion* $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ *heiße auf dem Intervall* $I \subseteq D(f)$ LIPSCHITZ**–stetig**, *wenn gilt:*

$$\boxed{\exists\, L \geq 0 \,:\, |f(x) - f(x_0)| \leq L\,|x - x_0| \;\; \forall\, x, x_0 \in I.} \tag{5.5}$$

(b) f *heiße* **gleichmäßig** LIPSCHITZ**–stetig**, *wenn die Bedingung* (5.5) *auf ganz* $D(f)$ *gilt. (Das heißt, die* LIPSCHITZ*–Konstante* L *hängt nicht von* x_0 *ab.)*

Wie wir oben gesehen haben, sind $\sin x$ und $\cos x$ gleichmäßig LIPSCHITZ–stetig. Dies gilt *nicht* für die Funktion $f(x) := \sqrt{x},\ x \in D(f) := [0, +\infty)$. Denn für $x, x_0 \geq 0$ haben wir

$$|\sqrt{x} - \sqrt{x_0}|^2 = x + x_0 - 2\sqrt{x x_0} \leq x + x_0 - 2\sqrt{[\min\{x, x_0\}]^2} = |x - x_0|.$$

Es folgt $|\sqrt{x} - \sqrt{x_0}| \leq \sqrt{|x - x_0|}$, so dass $f(x)$ zwar stetig auf ganz $D(f)$ ist, nicht aber LIPSCHITZ–stetig.

Merke: LIPSCHITZ–Stetigkeit $\Rightarrow$ Stetigkeit; die umgekehrte Implikation ist i.a. falsch.

BSP. (6.5.5) Die Signums–Funktion $f(x) := \mathrm{sign}\, x,\ x \in D(f) := \mathbf{R}$, ist **stetig** $\forall\, x_0 \neq 0$, denn außerhalb des Punktes $x = 0$ ist $f(x)$ eine Konstante. Hingegen gilt im Punkt $x_0 = 0$:

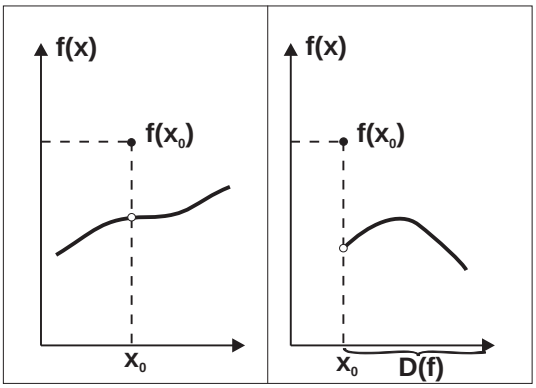
$$f(x_0 - 0) = -1 \neq f(x_0) = 0 \neq +1 = f(x_0 + 0).$$

Das heißt, f hat bei $x_0 = 0$ einen **Sprung** der Höhe 2.

Für **Unstetigkeiten** einer Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ gibt es einige *standardisierte Typenklassen*. Wir setzen nachfolgend wieder $f(x_0 \pm 0) := \lim\limits_{x \to x_0 \pm} f(x)$.

(a) **Sprung.** Die Funktion f hat bei $x_0 \in \mathbf{R}$ einen **Sprung**, wenn beide Funktionenlimites $f(x_0 \pm 0)$ (im eigentlichen Sinn) existieren, wenn jedoch $f(x_0 - 0) \neq f(x_0 + 0)$ gilt. *Beispiel:* $f(x) := \mathrm{sign}\, x$ bei $x_0 = 0$.

(b) **Hebbare Unstetigkeit.** Die Funktion f hat bei $x_0 \in \mathbf{R}$ eine **hebbare Unstetigkeit**, wenn $x_0 \in D(f)$, $\lim\limits_{x \to x_0} f(x) = g \in \mathbf{K}$ und $f(x_0) \neq g$ gelten. *Beispiel:* $f(x) := [\mathrm{sign}\, x]^2$ bei $x_0 = 0$.



Hebbare Unstetigkeiten einer Funktion

(c) **Lücke.** Die Funktion f hat bei $x_0 \in \mathbf{R}$ eine **Lücke**, wenn $x_0 \notin D(f)$, $\lim\limits_{x\to x_0} f(x) = g \in \mathbf{K}$ gelten. In diesem Fall kann f durch Hinzunahme des Wertes $f(x_0) := g$ zu einer stetigen Funktion ergänzt werden. (Die Funktion f heißt dann in x_0 **stetig ergänzbar** oder **stetig fortsetzbar** nach x_0.) *Beispiel:* Die Funktion

$$f(x) := \frac{x^2}{e^x - (1+x)},\ x \in D(f) := \mathbf{R} \setminus \{0\}$$

hat in $x_0 = 0$ den Funktionenlimes

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} x^2 \left(\sum_{k=2}^{\infty} \frac{x^k}{k!}\right)^{-1} = \lim_{x\to 0} \left(\sum_{k=2}^{\infty} \frac{x^{k-2}}{k!}\right)^{-1} = 2! = 2.$$

Das heißt, f ist in $x_0 = 0$ durch $f(0) := 2$ stetig ergänzbar.

(d) **Polstelle.** Die Funktion f hat bei $x_0 \in \mathbf{R}$ eine **Polstelle**, wenn $\lim\limits_{x\to x_0+} |f(x)| = +\infty$ und/oder $\lim\limits_{x\to x_0-} |f(x)| = +\infty$ gelten. *Beispiel:* $f(x) := 1/\sqrt{x},\ x \in D(f) := (0, +\infty)$ hat bei $x_0 = 0$ eine Polstelle.

Definition 6.19 *Eine Polstelle x_0 der Funktion $f(x)$ heiße* **Pol der Ordnung** $n \in \mathbf{N}$, *wenn der Funktionenlimes*

$$\lim_{x\to x_0} [(x - x_0)^n f(x)] = g \in \mathbf{K}$$

existiert, und wenn $g \neq 0$ gilt.

Zum Beispiel hat die Funktion $f(x) := \frac{1}{\sin x}$ bei $x_0 = 0$ einen *Pol 1.Ordnung*, denn es gilt ja $\lim\limits_{x\to 0} \frac{x}{\sin x} = 1$. Die Funktion $f(x) := \frac{1}{e^x - (1+x)}$ hat bei $x_0 = 0$ einen *Pol 2.Ordnung*, siehe oben (c).

Unstetigkeitsstellen, die nicht vom Typ (a)–(d) sind, werden i.a. nicht klassifiziert. Zu den nicht klassifizierten Beispielen zählt die Funktion $f(x) := \frac{1}{x} \sin \frac{1}{x}$, die bei $x_0 = 0$ eine *oszillierende Polstelle* hat. Hingegegen ist die Funktion $f(x) := x \sin \frac{1}{x}$ bei $x_0 = 0$ stetig ergänzbar durch $f(0) = 0$:

$$|f(x) - f(x_0)| = |x||\sin \tfrac{1}{x}| \leq |x| = |x - 0| < \epsilon \ \ \forall\, 0 < |x - 0| < \delta := \epsilon.$$

Die Stetigkeit der Funktionen f und g vererbt sich auf deren algebraische Verknüpfungen:

Satz 6.9 *Die Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ seien im Punkt $x_0 \in D(f) \cap D(g)$ stetig. Dann sind auch die folgenden Funktionen in x_0 stetig:*

(i) $f \pm g,\quad f \cdot g,\quad \lambda f\ \ \forall\, \lambda \in \mathbf{K}$;	(ii) $\frac{f}{g}$, *sofern* $g(x_0) \neq 0$ *gilt.*

Dieser Satz folgt unmittelbar aus den Rechenregeln über Grenzwerte.

Satz 6.10 *Existiert das Kompositum $g \circ f$ in $x_0 \in D(f)$, ist ferner die Funktion f stetig im Punkt x_0 und die Funktion g stetig im Punkt $f(x_0) \in D(g)$, so ist auch $g \circ f$ stetig in $x_0 \in D(f)$.*

Begründung: Aus der Stetigkeit von f folgt $\lim\limits_{x\to x_0} f(x) = f(x_0)$. Da auch g stetig ist, muss gelten

$$\lim_{x\to x_0} g[f(x)] = \lim_{f(x)\to f(x_0)} g[f(x)] = g[f(x_0)].$$

Merke: Bei stetigen Funktionen $g : D(g) \to \mathbf{K}$ ist die Verknüpfung $\lim \circ g$ auf der Menge $D(g)$ **kommutativ**:

$$\lim_{x\to x_0} g[f(x)] = g[\lim_{x\to x_0} f(x)].$$

Beispiel: Für jede im Punkte $x_0 = 1 \in D(g)$ stetige Funktion $g : D(g) \to \mathbf{K}$ gilt

$$\lim_{x\to 0} g\left(\frac{x}{e^x - 1}\right) = g\left(\lim_{x\to 0} \frac{x}{e^x - 1}\right) = g(1).$$

Ist g unstetig, so ist die Verknüpfung $\lim \circ g$ i.a. nicht kommutativ. *Beispiel:* Bezeichne g die HEAVISIDE–Funktion, $g(x) := \begin{cases} 1 & : \ x \geq 0, \\ 0 & : \ x < 0. \end{cases}$ Dann gilt

$$\lim_{x\to 0-} g(x^3) = \lim_{x\to 0-} 0 = 0, \quad \text{aber} \quad g\Big(\lim_{x\to 0-} x^3\Big) = g(0) = 1.$$

Beispiele stetiger Funktionen. Aus den Beispielen der Abschnitte 6.2–6.4 und aus den Sätzen 6.9 und 6.10 ergeben sich mühelos die folgenden Stetigkeitsaussagen.

1. Polynome $P_n(x) := \sum\limits_{k=0}^{n} a_k x^k$ sind in jedem Punkt $x_0 \in \mathbf{R}$ stetig; es gilt

$$\lim_{x\to+\infty} P_n(x) = +\infty \cdot \operatorname{sign} a_n, \quad \lim_{x\to-\infty} P_n(x) = \begin{cases} +\infty \cdot \operatorname{sign} a_n & : \ n \text{ gerade}, \\ -\infty \cdot \operatorname{sign} a_n & : \ n \text{ ungerade}. \end{cases}$$

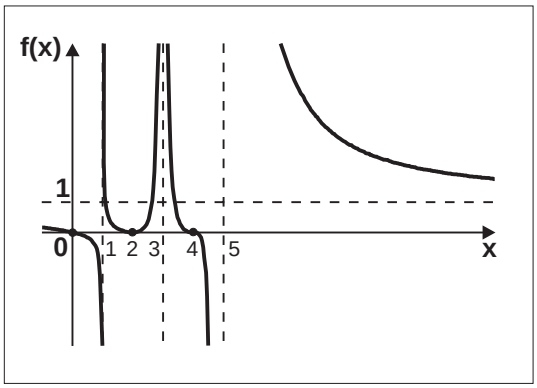
2. Die Funktionen $e^x, \sin x, \cos x$ sind stetig in ganz $\mathbf{R}$.

3. Rationale Funktionen $R(x) := \frac{P_n(x)}{Q_m(x)}$ sind in allen Punkten $x_0 \in \{x \in \mathbf{R} \ : \ Q_m(x) \neq 0\}$ stetig. Die Unstetigkeiten sind entweder Lücken oder Pole der Ordnung $k \leq m$. *Beispiel:*

$$R(x) := \frac{x(x-2)^2(x-4)^3}{(x-1)(x-3)^2(x-5)^3}, \ x \in D(R) := \mathbf{R} \setminus \{1, 3, 5\}.$$

$R(x)$ ist stetig auf ganz $D(R)$, und es gilt $\lim\limits_{x\to\pm\infty} R(x) = 1$. $R(x)$ hat ferner

$$\text{Pole bei } \begin{cases} x_0 = 1 & : \ 1. \text{ Ordnung}, \\ x_0 = 3 & : \ 2. \text{ Ordnung}, \\ x_0 = 5 & : \ 3. \text{ Ordnung}, \end{cases} \qquad \text{NS bei } \begin{cases} x_0 = 0 & : \ 1. \text{ Ordnung}, \\ x_0 = 2 & : \ 2. \text{ Ordnung}, \\ x_0 = 4 & : \ 3. \text{ Ordnung}. \end{cases}$$



Der Graph der Funktion
$R(x) := \frac{x\,(x-2)^2\,(x-4)^3}{(x-1)\,(x-3)^2\,(x-5)^3}$

4. Die Funktion $\sqrt[p]{x}$, $p \in \mathbf{N}$, ist stetig in $(0, +\infty)$ und rechtsseitig stetig bei $x_0 = 0$.

5. Die HEAVISIDE–Sprungfunktion $h(x) := \begin{cases} 1 \ : \ x \geq 0, \\ 0 \ : \ x < 0 \end{cases}$ ist stetig in jedem Punkt $x_0 \neq 0$ und rechtsseitig stetig bei $x_0 = 0$.

6. Die Signums–Funktion $\operatorname{sign} x$ ist stetig in jedem Punkt $x_0 \neq 0$ und unstetig bei $x_0 = 0$.

7. Die DIRICHLET–Funktion $\chi_{\mathbf{Q}}(x)$ ist unstetig in jedem Punkt $x_0 \in \mathbf{R}$.

8. Die Entire–Funktion $f(x) := [x]$ ist stetig in allen Punkten $x_0 \in \mathbf{R} \setminus \mathbf{Z}$ und rechtsseitig stetig bei $x_0 \in \mathbf{Z}$.

9. Die (reellwertigen) Funktionen $|f|, f^+, f^-, \max\{f,g\}, \min\{f,g\}$ sind stetig in allen Punkten $x_0 \in \mathbf{R}$, in denen f und g gemeinsam stetig sind.

10. **Stetigkeit komplexwertiger Funktionen.** Zerlegt man eine *komplexwertige* Funktion $f : D(f) \to \mathbf{C}$ mit $D(f) \subset \mathbf{R}$ in jedem Punkt $x_0 \in D(f)$ in Real– und Imaginärteil, $f(x_0) = u(x_0) + i\,v(x_0),\ u(x_0), v(x_0) \in \mathbf{R}$, so ergeben sich zwei **reellwertige** Funktionen

$$u : D(f) \to \mathbf{R}, \quad v : D(f) \to \mathbf{R}; \quad |f(x)| = \sqrt{|u(x)|^2 + |v(x)|^2} \ \ \forall\, x \in D(f).$$

Aus den Ungleichungen

$$\max\{|u|, |v|\} \leq \sqrt{|u|^2 + |v|^2} \leq \sqrt{|u|^2 + 2|u||v| + |v|^2} = |u| + |v| \tag{5.6}$$

erhält man unmittelbar:

Satz 6.11 *Eine komplexwertige Funktion $f = u + i\,v : D(f) \to \mathbf{C}$ mit $D(f) \subset \mathbf{R}$ ist genau dann im Punkt $x_0 \in D(f)$ stetig, wenn sowohl Realteil $u : D(f) \to \mathbf{R}$ als auch Imaginärteil $v : D(f) \to \mathbf{R}$ in x_0 stetig sind.*

Beispiele: (i) Die Funktion $f(x) := e^{i\,x} = \cos x + i \sin x$ ist stetig auf ganz $\mathbf{R}$, da sowohl Realteil $u(x) := \cos x$ als auch Imaginärteil $v(x) := \sin x$ in jedem Punkt $x_0 \in \mathbf{R}$ stetig sind.

(ii) Die reellwertigen Funktionen

$$u(x) := \frac{xe^x}{x-1}, \ x \in D(u) := \mathbf{R} \setminus \{1\}, \quad v(x) := \frac{\sin x}{\cos x}, \ x \in D(v) := \mathbf{R} \setminus \{(2n+1)\frac{\pi}{2} \ : \ n \in \mathbf{Z}\}$$

sind jeweils stetig auf ihren Definitionsbereichen $D(u)$ bzw. $D(v)$. Wegen Satz 6.11 ist dann auch die komplexwertige Funktion $f(x) := u(x) + i\,v(x)$ stetig in allen Punkten $x_0 \in D(u) \cap D(v) = \mathbf{R} \setminus \{1, (2n+1)\frac{\pi}{2} \ \text{ mit } \ n \in \mathbf{Z}\}$.

Stetigkeit vektor(raum)wertiger Funktionen. Ist Y ein **normierter** Vektorraum über dem Körper $\mathbf{K}$, so ist die vektorwertige Funktion $\vec{f} \in \operatorname{Abb}(\mathbf{R}, Y)$ mit Definitionsbereich $D(\vec{f}) \subset \mathbf{R}$ in einem Punkt $x_0 \in D(\vec{f})$ stetig, wenn anstelle der Bedingung (5.1) die folgende Bedingung gilt:

$$\boxed{\forall\, \epsilon > 0\ \exists\, \delta = \delta(\epsilon) > 0 \ : \ \|\vec{f}(x) - \vec{f}(x_0)\| < \epsilon \ \ \forall\, x \in D(\vec{f}) \ \text{ mit } \ 0 < |x - x_0| < \delta.} \tag{5.7}$$

BSP. (6.5.6) **Vektorwertige Funktionen** im Zahlenvektorraum $Y := \mathbf{K}^n$. Dieser Vektorraum ist ein **Praehilbertraum** mit dem Standardskalarprodukt

$$\langle \vec{f}, \vec{g} \rangle := \sum_{k=1}^{n} f_k \bar{g}_k, \quad \vec{f} = (f_1, f_2, \ldots, f_n)^T, \vec{g} = (g_1, g_2, \ldots, g_n)^T \in \mathbf{K}^n,$$

und der induzierten Norm

$$\|\vec{f}\| := \sqrt{\langle \vec{f}, \vec{f} \rangle} = \left(\sum_{k=1}^{n} |f_k|^2 \right)^{1/2}, \quad \vec{f} = (f_1, f_2, \ldots, f_n)^T \in \mathbf{K}^n. \tag{5.8}$$

Die folgende Ungleichung ist eine Verallgemeinerung der Ungleichung (5.6):

$$\max\{|f_1|, |f_2|, \ldots, |f_n|\} \leq \|\vec{f}\| \leq \sum_{k=1}^{n} |f_k|, \quad \vec{f} = (f_1, f_2, \ldots, f_n)^T \in \mathbf{K}^n. \tag{5.9}$$

Demgemäß kann die Aussage über die Stetigkeit komplexwertiger Funktionen ganz analog für die hier vorliegenden vektorwertigen Funktionen formuliert werden.

Satz 6.12 *Der Zahlenvektorraum* $\mathbf{K}^n$ *sei mit der Standardnorm* (5.8) *versehen. Genau dann ist die vektorwertige Funktion*

$$\vec{f}(x) := \Big(f_1(x), f_2(x), \ldots, f_n(x) \Big)^T, \; f_k : D(\vec{f}) \to \mathbf{K} \;\; \forall\, k = 1, 2, \ldots, n, \;\; D(\vec{f}) \subset \mathbf{R},$$

im Punkt $x_0 \in D(\vec{f})$ *stetig, wenn jede ihrer Komponentenfunktionen* f_k, $k = 1, 2, \ldots, n$, *in* x_0 *stetig ist.*

Anwendung: Parameterdarstellung von Kurven. In diesem Zusammenhang wird die unabhängige Veränderliche $x \in \mathbf{R}$ in der Regel mit t bezeichnet. t heißt dann **Kurvenparameter**.

Definition 6.20 *Im euklidischen Vektorraum* $Y := \mathbf{R}^3$ *heiße eine Punktmenge* $\Gamma := \{\vec{x}(t) \in \mathbf{R}^3 : \vec{x}(t) = \Big(x(t), y(t), z(t) \Big)^T, \; t_1 \leq t \leq t_2\}$ *eine* **stetige räumliche Kurve**, *wenn die Komponentenfunktionen*

$$x = x(t),\; y = y(t),\; z = z(t), \quad t_1 \leq t \leq t_2, \tag{5.10}$$

auf dem Intervall $[t_1, t_2]$ *stetige Funktionen des Kurvenparameters* t *sind. Die Beziehungen* (5.10) *heißen eine* **Parameterdarstellung** *von* Γ. *Liegt* Γ *ganz in einer Ebene* $E \subset \mathbf{R}^3$, *so heiße* Γ *eine* **ebene Kurve**. *Insbesondere wird durch jede vektorwertige Funktion* $\vec{x} : D(\vec{x}) \to \mathbf{R}^2$ *mit*

$$\vec{x}(t) := \Big(x(t), y(t) \Big)^T, \; t \in D(\vec{x}) := [t_1, t_2],$$

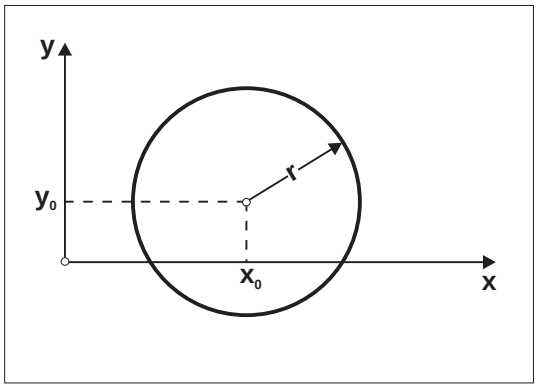
eine Parameterdarstellung einer ebenen Kurve Γ *definiert.*

Beispiele für Parameterdarstellungen stetiger ebener Kurven.

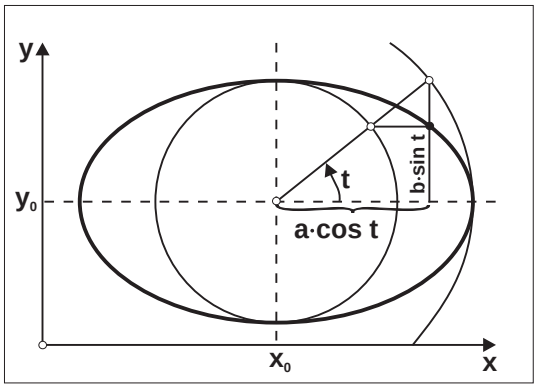
1. Die **Kreislinie** vom Radius $r > 0$ mit Mittelpunkt $M(x_0, y_0)$ gestattet eine Parameterdarstellung

$$\boxed{x(t) := x_0 + r\,\cos t, \quad y(t) := y_0 + r\,\sin t, \quad t \in [0, 2\pi).}$$

Man erhält aus dieser Darstellung wieder die bekannte Kreisgleichung $(x - x_0)^2 + (y - y_0)^2 - r^2 = 0$.



Kreislinie vom Radius $r > 0$



Ellipse mit Halbachsen $a > 0, b > 0$

2. Die **Ellipse** mit Halbachsen $a > 0, b > 0$ und Mittelpunkt $M(x_0, y_0)$ gestattet eine Parameterdarstellung

$$\boxed{x(t) := x_0 + a\,\cos t, \quad y(t) := y_0 + b\,\sin t, \quad t \in [0, 2\pi).}$$

Man erhält aus dieser Darstellung wieder die bekannte Ellipsengleichung $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} - 1 = 0$.

3. Die **Zykloide** oder *Radkurve* entsteht als Bahnkurve eines Punktes $P(x, y)$ auf der Peripherie eines Kreises vom Radius $r > 0$, wenn dieser Kreis längs der positiven x–Achse rollt. Die Bahnkurve gestattet eine Parameterdarstellung

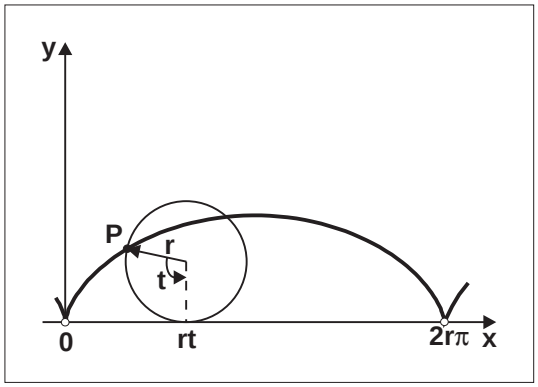
$$\boxed{x(t) := r\,(t - \sin t), \quad y(t) := r\,(1 - \cos t), \quad t \in \mathbf{R}.}$$

Eine **spezielle Parameterdarstellung** ebener Kurven erhält man durch Verwendung von **Polarkoordinaten** (vgl. Abschnitt 2.2, Polardarstellung komplexer Zahlen)

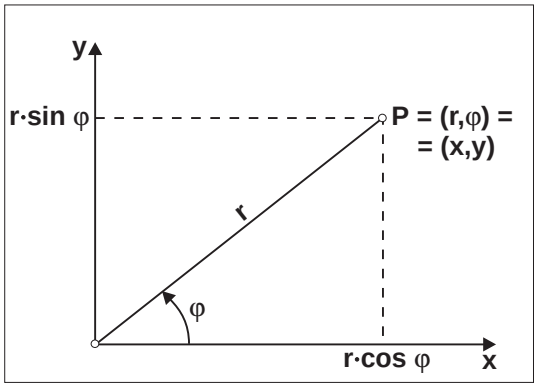
$$x = r\,\cos\varphi, \quad y = r\,\sin\varphi, \quad r \geq 0,\ 0 \leq \varphi_H < 2\pi :$$

Durch Vorgabe einer stetigen Funktion $r = r(\varphi) \geq 0$ auf einem Intervall $\varphi_1 \leq \varphi \leq \varphi_2$ wird eine ebene stetige Kurve in **Polardarstellung** definiert:

$$\boxed{x(\varphi) := r(\varphi)\,\cos\varphi, \quad y(\varphi) := r(\varphi)\,\sin\varphi, \quad \varphi \in [\varphi_1, \varphi_2].}$$



Die Zykloide



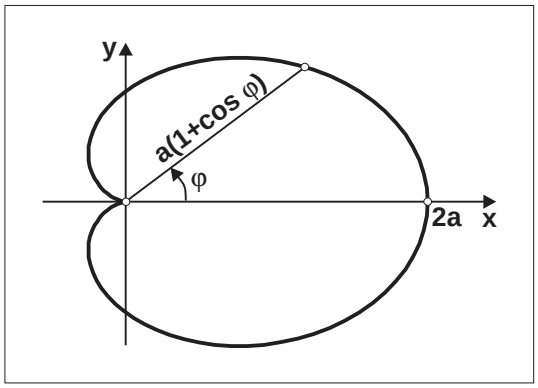
Polarkoordinaten

4. Die **Kardioide** oder Herzkurve gestattet die Polardarstellung

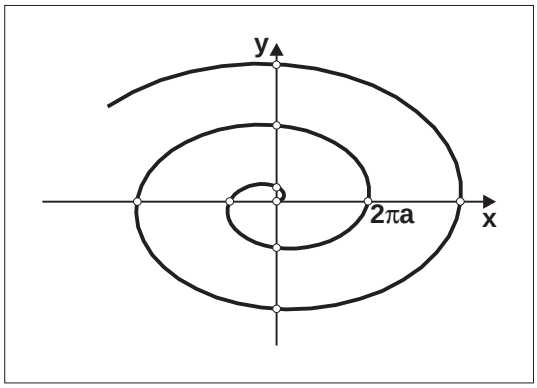
$$\boxed{x(\varphi) = r(\varphi)\,\cos\varphi, \quad y(\varphi) = r(\varphi)\,\sin\varphi, \quad r(\varphi) := a\,(1 + \cos\varphi), \ \ a > 0, \varphi \in [0, 2\pi).}$$

5. Die Archimedische **Spirale** gestattet die Polardarstellung

$$\boxed{x(\varphi) = r(\varphi)\,\cos\varphi, \quad y(\varphi) = r(\varphi)\,\sin\varphi, \quad r(\varphi) := a\,\varphi, \ \ a > 0, \varphi \geq 0.}$$



Die Kardioide



Die Archimedische Spirale

Man kann den Kurvenparameter t (oder φ) in vielen Fällen aus der Parameterdarstellung einer Kurve wieder eliminieren. Dies haben wir an den Beispielen des Kreises und der Ellipse gezeigt. Das Resultat ist eine **implizite Darstellung** $F(x,y)=0$ der ebenen Kurve in **kartesischen Koordinaten**. Für die Archimedische Spirale hat eine solche implizite Darstellung die Form

$$F(x,y)=\frac{y}{\sqrt{x^2+y^2}}-\sin\left(\frac{1}{a}\sqrt{x^2+y^2}\right)=0.$$

Merke: Für ebene Kurven in kartesischen Koordinaten hat man drei verschiedene Möglichkeiten der Darstellung:

- Die explizite Darstellung $y=f(x)$, zum Beispiel $y=x\,e^x$.
- Die implizite Darstellung $F(x,y)=0$, zum Beispiel $x^2-y^2-1=0$.
- Die Parameterdarstellung $x=x(t), y=y(t)$, zum Beispiel $x=a\cos t, y=b\sin t$.

Nicht jede ebene Kurve kann in allen drei Darstellungen beschrieben werden.

Für das folgende Beispiel benötigen wir einige vorbereitende Betrachtungen über **Matrixnormen**. Es sei $Y:=\mathbf{K}^{(n,n)}$ der Vektorraum der quadratischen $n\times n$–Matrizen. (Die Einschränkung auf quadratische Matrizen ist hier rein formaler Natur; der überwiegende Teil der folgenden Überlegungen läßt sich ohne Einschränkung auch auf dem Vektorraum $Y:=\mathbf{K}^{(m,n)}$ formulieren.)

Definition 6.21 (Matrixnorm)
Eine Abbildung $\|\cdot\|:\mathbf{K}^{(n,n)}\to\mathbf{R}$ *mit*

(N1) $\|A\|\geq 0\ \ \forall\, A\in\mathbf{K}^{(n,n)}$ *und* $\|A\|=0$ *genau für* $A=O$,

(N2) $\|\lambda\, A\|=|\lambda|\,\|A\|\ \ \forall\, A\in\mathbf{K}^{(n,n)}\ \ \forall\,\lambda\in\mathbf{K}$,

(N3) $\|A+B\|\leq\|A\|+\|B\|\ \ \forall\, A,B\in\mathbf{K}^{(n,n)}$, *(Dreiecksungleichung)*

heiße eine **Matrixnorm** *auf* $\mathbf{K}^{(n,n)}$. *Eine Matrixnorm heiße* **submultiplikativ**, *wenn zusätzlich gilt:*

(N4) $\|AB\|\leq\|A\|\,\|B\|\ \ \forall\, A,B\in\mathbf{K}^{(n,n)}$.

Die folgenden submultiplikativen Matrixnormen sind für $A=(a_{jk})\in\mathbf{K}^{(n,n)}$ gebräuchlich:

$$\begin{aligned}
\|A\|_G &:= n\cdot\max_{1\leq j,k\leq n}|a_{jk}|, && \text{(Gesamtnorm)}\\
\|A\|_Z &:= \max_{1\leq j\leq n}\sum_{k=1}^{n}|a_{jk}|, && \text{(Zeilensummennorm)}\\
\|A\|_S &:= \max_{1\leq k\leq n}\sum_{j=1}^{n}|a_{jk}|, && \text{(Spaltensummennorm)}\\
\|A\|_F &:= \Big\{\sum_{j=1}^{n}\sum_{k=1}^{n}|a_{jk}|^2\Big\}^{1/2}. && \text{(FROBENIUS–Norm)}
\end{aligned}$$

Diese Matrixnormen sind **paarweise äquivalent**; es gilt nämlich:

$$\begin{aligned}
\tfrac{1}{n}\|A\|_G &\leq \|A\|_{Z,S} \leq \|A\|_G \leq n\|A\|_{Z,G},\\
\tfrac{1}{n}\|A\|_G &\leq \|A\|_F \leq \|A\|_G \leq n\|A\|_F.
\end{aligned}\tag{5.11}$$

Sind nun $A=(a_{jk})\in\mathbf{K}^{(n,n)}$ und $\vec{x}\in\mathbf{K}^n$ gegeben, so sind die Vektoren $\vec{x}$ und $\vec{y}:=A\vec{x}\in\mathbf{K}^n$ über das Produkt $y_j=\sum_{k=1}^{n}a_{jk}x_k, j=1,2,\ldots,n$, miteinander verbunden. Ist eine Vektornorm $\|\cdot\|$ auf dem Vektorraum $\mathbf{K}^n$ gegeben, so stellt sich die Frage nach einem Zusammenhang zwischen den Größen $\|\vec{x}\|, \|\vec{y}\|$ und $\|A\|$. Die Antwort geben wir in folgender

Definition 6.22 *Eine Matrixnorm* $\|\cdot\|:\mathbf{K}^{(n,n)}\to\mathbf{R}$ *und eine Vektornorm* $\|\cdot\|:\mathbf{K}^n\to\mathbf{R}$ *heißen* **kompatibel** *oder* **verträglich**, *falls gilt:*

$$\|A\vec{x}\|\leq\|A\|\,\|\vec{x}\|\ \ \forall\,\vec{x}\in\mathbf{K}^n\ \ \forall\, A\in\mathbf{K}^{(n,n)}.\tag{5.12}$$

Auf dem Vektorraum $\mathbf{K}^n$ sind die folgende Vektornormen gebräuchlich:

$$\begin{aligned}
\|\vec{x}\|_\infty &:= \max_{1\le k\le n} |x_k|, && \text{(Maximum–Norm)}\\
\|\vec{x}\|_2 &:= \Big\{ \sum_{k=1}^{n} |x_k|^2 \Big\}^{1/2}, && \text{(Euklidische Norm)}\\
\|\vec{x}\|_1 &:= \sum_{k=1}^{n} |x_k|. && (L_1\text{–Norm})
\end{aligned}$$

Man überzeugt sich nun mit einfacher Rechnung, dass unter den hier eingeführten Matrix– und Vektornormen die folgenden kompatiblen Paare existieren:

$$\begin{array}{llll}
\|A\|_G & \text{oder} & \|A\|_Z & \text{kompatibel mit} \quad \|\vec{x}\|_\infty,\\
\|A\|_G & \text{oder} & \|A\|_S & \text{kompatibel mit} \quad \|\vec{x}\|_1,\\
\|A\|_G & \text{oder} & \|A\|_F & \text{kompatibel mit} \quad \|\vec{x}\|_2.
\end{array}$$

Die Frage, ob es zu einer gegebenen Vektornorm stets eine kompatible, submultiplikative Matrixnorm gibt, beantworten wir in dem folgenden

Satz 6.13 *Gegeben sei eine Vektornorm $\|\cdot\| : \mathbf{K}^n \to \mathbf{R}$. Dann ist die gemäß*

$$\boxed{\|A\| := \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|} = \max_{\|\vec{x}\|=1} \|A\vec{x}\|}$$

erklärte Abbildung $\|\cdot\| : \mathbf{K}^{(n,n)} \to \mathbf{R}$ eine submultiplikative Matrixnorm. Sie heißt die **natürliche Matrixnorm** *oder* **Grenznorm**, *und sie ist mit der gegebenen Vektornorm kompatibel. Unter allen mit dieser Vektornorm kompatiblen Normen ist sie die kleinste Matrixnorm.*

Bemerkung 6.8 Es ist nicht ganz einfach, die natürliche Matrixnorm zu einer gegebenen Vektornorm **explizit** zu berechnen. Die natürliche Matrixnorm zur Maximum–Norm ist beispielsweise die Zeilensummennorm: □

$$\|A\|_\infty := \max_{\|\vec{x}\|_\infty} \|A\vec{x}\|_\infty = \max_{1\le j\le n} \sum_{k=1}^{n} |a_{jk}| = \|A\|_Z.$$

BSP. (6.5.7) **Matrixwertige Funktionen.** Ganz analog zur Stetigkeitsaussage vektorwertiger Funktionen hat man:

Satz 6.14 *Der Vektorraum $\mathbf{K}^{(n,n)}$ der $n \times n$–Matrizen sei versehen mit einer Matrixnorm. Genau dann ist die matrixwertige Funktion*

$$A(t) := \begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{bmatrix}, \quad a_{jk} : D(A) \to \mathbf{K}, \;\; D(A) \subset \mathbf{R},$$

stetig im Punkt $t_0 \in D(A)$, wenn jede ihrer Komponentenfunktionen $a_{jk}(t), 1 \le j,k \le n$, in t_0 stetig ist.

6.6 Eigenschaften stetiger Funktionen

In diesem Abschnitt werden Eigenschaften stetiger Funktionen zusammengestellt, die grundlegend für die Analysis sind. Eine erste Eigenschaft stetiger Funktionen ist ihre **Beschränktheit auf abgeschlossenen Intervallen**.

Satz 6.15 *Es sei* $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{K})$ *eine auf dem abgeschlossenen Intervall* $[a,b] \subset \mathbf{R}$ *stetige Funktion. Dann ist* f **beschränkt***:*

$$\boxed{\exists\, M \,:\, |f(x)| \le M \;\; \forall\, x \in [a,b].} \tag{6.1}$$

Begründung: Wäre das Gegenteil von (6.1) wahr, nämlich

$$\forall\, n \in \mathbf{N}\; \exists\, x_n \in [a,b] \,:\, |f(x_n)| > n, \tag{6.2}$$

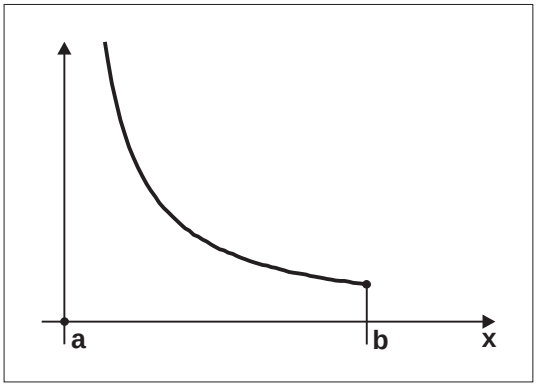
so hätte die beschränkte Folge $(x_n)_{n\in\mathbf{N}} \subset [a,b]$ nach dem Auswahlsatz von Bolzano–Weierstrass (Satz 3.9, Ing.–Math.I) mindestens einen Häufungspunkt $x_0 \in [a,b]$. Für eine Teilfolge $\mathbf{N}' \subset \mathbf{N}$ würde $\lim\limits_{j\in\mathbf{N}'} x_j = x_0$ gelten, und wegen der Stetigkeit von f auch $\lim\limits_{j\in\mathbf{N}'} f(x_j) = f(x_0)$. Diese Aussage stände im Widerspruch zu (6.2), wonach $\lim\limits_{j\in\mathbf{N}'} |f(x_j)| \ge \lim\limits_{j\in\mathbf{N}'} j = +\infty$ gilt. Also muss (6.2) falsch sein. □

Beachte: Die Aussage (6.1) wird im allgemeinen **falsch**, wenn f zwar stetig, das Definitionsintervall aber *nicht* abgeschlossen ist. *Fallbeispiel:* $f(x) := \frac{1}{x}\ \forall\ x \in (0,1]$. Dies gilt auch, wenn die Funktion $f : [a,b] \to \mathbf{K}$ *unstetig* ist. *Fallbeispiel:* $f(x) := \begin{cases} \frac{1}{x} & : \ x \in (0,1], \\ 0 & : \ x = 0. \end{cases}$

Für **reellwertige** stetige Funktionen $f : [a,b] \to \mathbf{R}$ ist also die Bildmenge $f([a,b]) \subset \mathbf{R}$ gemäß Satz 6.15 beschränkt. Aus dem Supremumsprinzip (Satz 1.16, Ing.–Math.I) folgern wir, dass die beiden Zahlen

$$\sup f([a,b]) \in \mathbf{R}, \quad \inf f([a,b]) \in \mathbf{R}$$

existieren. Diese Aussage läßt sich in der folgenden Weise noch weiter präzisieren, wie der nächste Satz zeigt.



Der Graph der Funktion
$f(0) := 0, f(x) := \frac{1}{x},\ x > 0$

Satz 6.16 (Extremalsatz)
Gegeben sei eine stetige Funktion $f : [a,b] \to \mathbf{R}$*. Dann nimmt die Funktion* f *das Maximum und das Minimum ihrer Funktionswerte jeweils in einem Punkt des Intervalls* $[a,b]$ *an:*

$$\boxed{\exists\, \underline{x}, \bar{x} \in [a,b] \,:\, f(\underline{x}) = \min_{x\in[a,b]} f(x), \quad f(\bar{x}) = \max_{x\in[a,b]} f(x).} \tag{6.3}$$

Demgemäß gilt $f(\underline{x}) \le f(x) \le f(\bar{x})\ \forall\ x \in\ [a,b]$.

Begründung: Das Supremumsprinzip sichert die Existenz der Zahl $M := \sup f([a,b])$, das heißt, es gilt

$$\forall\, \epsilon > 0\ \exists\, x \in [a,b] \,:\, M - f(x) < \epsilon, \tag{6.4}$$

(vgl. (8.1) in Abschnitt 1.8). Wir zeigen hiermit die Existenz eines Punktes $\bar{x} \in [a,b]$ mit $f(\bar{x}) = M$. Wäre nämlich $f(x) < M \ \forall\ x \in [a,b]$, so wäre die Funktion $g(x) := [M - f(x)]^{-1}$ auf ganz $[a,b]$ stetig, dort positiv und gemäß Satz 6.15 beschränkt:

$$0 < g(x) \leq L \ \ \forall\, x \in [a,b].$$

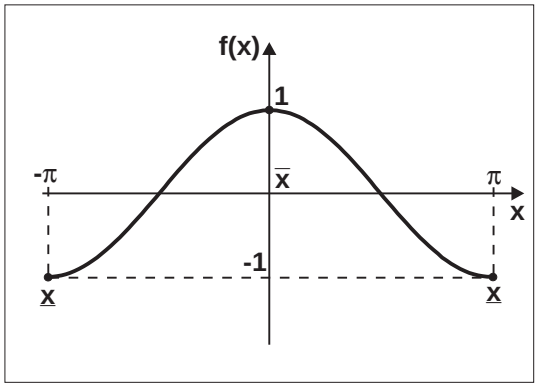
Wird jedoch in (6.4) die Zahl $\epsilon > 0$ gemäß $\epsilon := 1/2L$ gewählt, so existiert dazu ein $x \in [a,b]$ mit $2L < 1/(M - f(x)) = g(x)$, im Widerspruch zur Beschränktheit von g. Mit ähnlicher Schlussweise zeigt man auch die Existenz einer Zahl $\underline{x} \in [a,b]$, so dass $f(\underline{x}) = \inf f([a,b])$ gilt. □

Bemerkung 6.9 (a) Die Extremalstellen $\bar{x}, \underline{x} \in [a,b]$ müssen nicht eindeutig festgelegt sein. Im Satz 6.16 wird nur die Existenz **mindestens eines** solchen Paares behauptet. *Beispiel:* Die Funktion $f(x) := \cos x$ nimmt im Intervall $[-n\pi, n\pi]$ ihr Maximum in den Punkten $\bar{x}_j := 2\pi j$ an, und ihr Minimum in den Punkten $\underline{x}_j := (2j+1)\pi$.

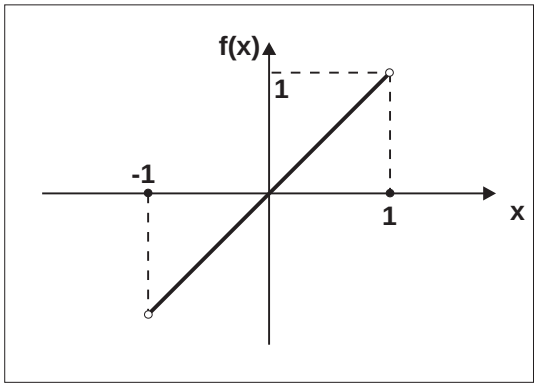
(b) Die Aussage von Satz 6.16 wird i.a. falsch, wenn f nur im offenen Intervall (a,b) stetig ist oder gar unstetig auf $[a,b]$, *zum Beispiel* (vgl. untere Skizze):

$$f(x) := \begin{cases} x & : \ x \in (-1,+1), \\ 0 & : \ x = \pm 1. \end{cases}$$

(c) Satz 6.16 gilt auch für den **Betrag** $|f|$ einer stetigen **komplexwertigen** Funktion $f : [a,b] \to \mathbf{C}$. *Beispiel:* Der Betrag der Funktion $f(x) := (x^2 - 1) + 2i\,x$ ist gegeben durch $|f(x)| = \sqrt{(x^2-1)^2 + 4x^2} = x^2 + 1$. Auf dem Intervall $[-1,+1]$ gilt deshalb $|f(\underline{x})| = |f(0)| = 1$, $|f(\bar{x})| = |f(\pm 1)| = 2$. □



$\bar{x}$ **und** $\underline{x}$ **sind i.a. nicht eindeutig**

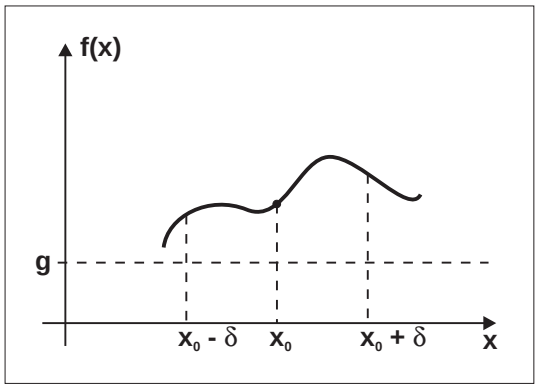


Ist f **unstetig, so existieren i.a.** $\max f(x)$ **und** $\min f(x)$ **nicht**

Eine anschaulich völlig klare Aussage wird in dem folgenden Satz formuliert:

Satz 6.17 *Es sei* $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ *eine im Punkt* $x_0 \in D(f)$ *stetige Funktion, und es gelte* $f(x_0) > g \in \mathbf{R}$. *Dann folgt:*

$$\boxed{\exists\, \delta > 0 \ : \ f(x) > g \ \ \forall\, x \in D(f) \ \ mit \ \ 0 < |x - x_0| < \delta.} \tag{6.5}$$



Zur Aussage (6.5) einer stetigen Funktion

Begründung: Wäre (6.5) falsch, so wäre im Gegensatz

$$\forall\, n \in \mathbf{N}\ \exists\, x_n \in D(f)\ :\ f(x_n) \le g \ \text{ und } \ 0 < |x_n - x_0| < \frac{1}{n}$$

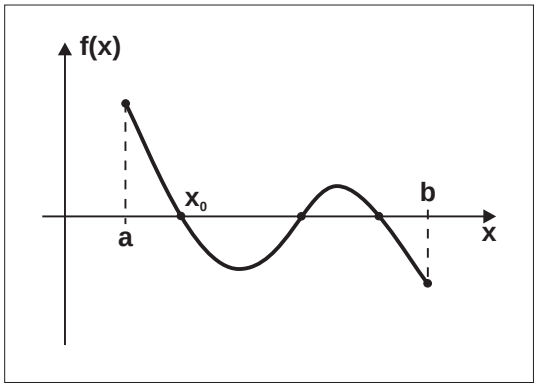
wahr. Wir hätten $\lim\limits_{n\to\infty} x_n = x_0$, und wegen der Stetigkeit von f bei x_0 folgte $g \ge \lim\limits_{n\to\infty} f(x_n) = f(x_0)$, im Widerspruch zur Voraussetzung $f(x_0) > g$. □

Bemerkung 6.10 Man kann Satz 6.17 auch in der Form $f(x) < g \ \ \forall\ 0 < |x - x_0| < \delta$ beweisen, wenn $f(x_0) < g$ vorausgesetzt wird. Wer stetig wächst und noch nicht an die Decke stößt, kann ohne Anstoßen noch ein bisschen weiterwachsen. □

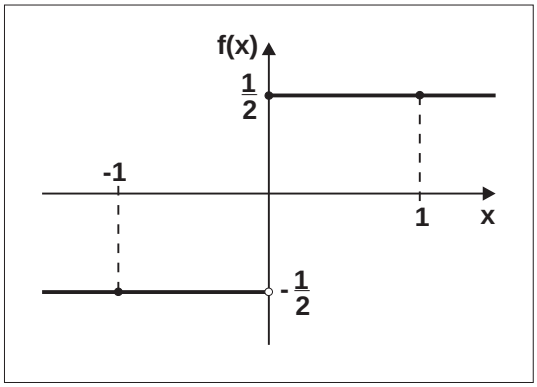
Eine unmittelbare Folgerung aus Satz 6.17 ist der fundamentale

Satz 6.18 (Nullstellensatz von Bolzano**)**
Es sei eine stetige Funktion $f : [a,b] \to \mathbf{R}$ gegeben mit $f(a)f(b) < 0$, (das heißt, entweder gilt $f(a) < 0, f(b) > 0$ oder $f(a) > 0, f(b) < 0$). Dann besitzt f im offenen Intervall (a,b) mindestens eine Nullstelle: $f(x_0) = 0$ für (mindestens) ein $x_0 \in (a,b)$.

Begründung: Wir nehmen zum Beispiel $f(a) < 0, f(b) > 0$ an und setzen $M := \{x \in [a,b] : f(x) < 0\} \subset [a,b]$. Dann ist die M beschränkt und wegen $a \in M$ nichtleer. Also existiert nach dem Supremumsprinzip (Satz 1.16) die Zahl $x_0 := \sup M \in [a,b]$. Wäre $f(x_0) < 0$, so gäbe es gemäß Satz 6.17 ein Intervall $I := (x_0, x_0 + \delta) \subset [a,b]$ mit $f(x) < 0\ \forall\, x \in I$. Dies wäre ein Widerspruch zu $x_0 = \sup M$. Also muss $f(x_0) = 0$ gelten und somit auch $a \ne x_0 \ne b$. □



Zum Nullstellensatz von Bolzano



Eine unstetige Funktion f mit $f(a)f(b) < 0$ braucht im Intervall (a,b) keine Nullstellen zu haben

Bemerkung 6.11 (a) Die Nullstelle x_0 ist i.a. **nicht eindeutig** definiert.

(b) Für unstetige Funktionen $f : [a,b] \to \mathbf{R}$ wird Satz 6.18 i.a. falsch. *Fallbeispiel:* Es sei $h(x) := \begin{cases} 1 & : x \geq 0, \\ 0 & : x < 0, \end{cases}$ die HEAVISIDE–Funktion. Die Funktion $f(x) := h(x) - \frac{1}{2}$, $x \in [-1,+1]$, erfüllt zwar die Bedingung $f(-1)f(+1) < 0$, sie hat dennoch im Intervall $(-1,+1)$ keine Nullstellen.

(c) Es ist wichtig, dass die Menge $[a,b]$ ein Teilintervall von $\mathbf{R}$ ist. Satz 6.18 gilt zum Beispiel nicht auf einer Menge $[a,b] \subset \mathbf{Q}$. *Fallbeispiel:* Die Funktion $f(x) := 2(x^2-2)$, $x \in [0,2] \cap \mathbf{Q}$, erfüllt $f(0)f(2) = -16 < 0$, während $f(x_0) = 0$ genau für $x_0 = \sqrt{2} \notin \mathbf{Q}$ gilt. Wie wir in dem Beweis zu Satz 6.18 gesehen haben, beruht der Nullstellensatz im wesentlichen auf dem **Supremumsprinzip** und somit auf der **Vollständigkeit** von $\mathbf{R}$.

(d) Als einfache Folgerung aus dem Nullstellensatz erhält man: □

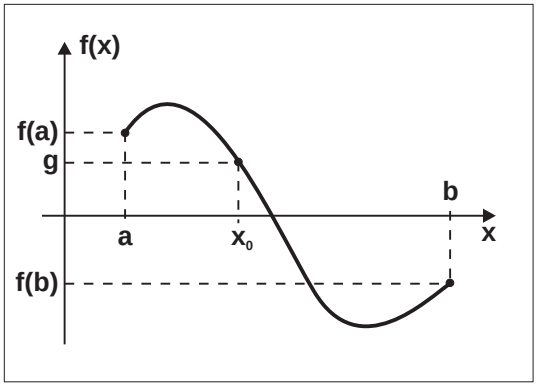
Merke: Jedes Polynom $P_n(x) = \sum_{k=0}^{n} a_k x^k$, $a_k \in \mathbf{R}, a_n \neq 0$, von **ungeradem Grade** $n = 2m+1, m \in \mathbf{N}$, besitzt mindestens eine **reelle** Nullstelle. Denn es gilt

$$\lim_{x \to \pm\infty} P_n(x) = \pm\infty \cdot \operatorname{sign} a_n.$$

Eine Verallgemeinerung des Nullstellensatzes ist der folgende

Satz 6.19 (Zwischenwertsatz von BOLZANO)
Eine stetige Funktion $f : [a,b] \to \mathbf{R}$ nimmt jeden Wert des Intervalls zwischen $f(a)$ und $f(b)$ mindenstens einmal an.

Begründung: Für $f(a) = f(b)$ ist nichts zu beweisen. Gelte also $f(a) \neq f(b)$, und sei g ein Punkt aus dem offenen Intervall zwischen $f(a)$ und $f(b)$. Dann folgt $(f(a)-g)(f(b)-g) < 0$. Das heißt, die Funktion $\varphi(x) := f(x) - g$ erfüllt die Voraussetzungen zum Nullstellensatz 6.18. Demgemäß existiert ein $x_0 \in (a,b)$ mit $\varphi(x_0) = 0 = f(x_0) - g$. □



Zum Zwischenwertsatz von Bolzano

BSP. (6.6.1) Ein Auto fährt eine Strecke von $400\,km$ ohne Stop in genau 5 Stunden (was einer Durchschnittsgeschwindigkeit von $v = 80\,km/h$ entspricht). Gibt es einen zusammenhängenden Zeitabschnitt von exakt $1\,h$, in welchem das Auto eine Strecke von genau $80\,km$ gefahren ist? Die Antwort lautet Ja, und wir begründen sie mit dem Zwischenwertsatz von BOLZANO. Dazu bezeichne $x(t)$ (in km) die Strecke, die das Auto in der Zeit $0 \leq t \leq 5$ (in Stunden) zurückgelegt hat. Die

Funktion $f(t) := x(t+1) - x(t),\ t \in [0,4]$, ist stetig und reellwertig. Wäre nun $f(t) < 80\ \forall\ t \in [0,4]$, so hätte das Auto in keinem Zeitabschnitt von $1\,h$ eine Strecke von mindestens $80\,km$ zurückgelegt. Somit kann das Auto auch nicht die Gesamtstrecke in der Zeit von $5\,h$ zurückgelegt haben. Zu einem ähnlichen Widerspruch gelangt man mit der Annahme $f(t) > 80\ \forall\ t \in [0,4]$. Also muss es Zeiten $a, b \in [0,4]$ geben mit $f(a) \geq 80$ und $f(b) \leq 80$. Aus dem Zwischenwertsatz 6.19 folgern wir jetzt: $\exists\ t_0 \in [0,4]\ :\ f(t_0) = 80$.

Bemerkung 6.12 (a) Der Extralmalsatz 6.16 sichert die Existenz von Punkten $\bar{x}, \underline{x} \in [a,b]$ mit $f(\underline{x}) \leq f(x) \leq f(\bar{x})\ \forall\ x \in [a,b]$. Identifiziert man das Intervall $[a,b]$ in Satz 6.19 mit dem abgeschlossenen Intervall zwischen $\underline{x}$ und $\bar{x}$, so nimmt $f(x)$ **jeden Wert** zwischen $f(\underline{x})$ und $f(\bar{x})$ an.

Merke: Das Bild eines abgeschlossenen Intervalls $[a,b]$ unter einer **stetigen** reellwertigen Funktion f ist das abgeschlossene Intervall

$$\Big[\min_{x\in[a,b]} f(x),\ \max_{x\in[a,b]} f(x)\Big].$$

(b) Die Stetigkeit ist lediglich eine **hinreichende** Bedingung für die Gültigkeit des Zwischenwertsatzes. Die folgende Funktion ist nur im Punkte $x_0 := \frac{1}{2}$ stetig:

$$f(x) := \begin{cases} x & :\ x \text{ rational}, \\ 1-x & :\ x \text{ irrational}, \end{cases} \quad 0 \leq x \leq 1.$$

Dennoch nimmt $f(x)$ jeden Wert zwischen dem Minimum $f(0) = 0$ und dem Maximum $f(1) = 1$ an. □

Gleichmäßige Stetigkeit

Wir hatten bereits in Abschnitt 6.5 angemerkt, dass die Zahl $\delta > 0$ in der $\epsilon - \delta$–Definition (5.1) der Stetigkeit im allgemeinen nicht nur von der Wahl der Zahl $\epsilon > 0$ abhängt, sondern auch von der Stelle $x_0 \in D(f)$, in welcher die Stetigkeit einer Funktion f nachzuweisen ist. In einigen Sonderfällen kann die Zahl $\delta > 0$ unabängig von der Stelle $x_0 \in D(f)$ gewählt werden. Solche Funktionen heißen *gleichmäßig stetig*.

Definition 6.23 *Eine Funktion* $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ *heiße* **gleichmäßig stetig** *auf* $D(f) \subset \mathbf{R}$, *wenn gilt:*

$$\forall\ \epsilon > 0\ \exists\ \delta = \delta(\epsilon) > 0\ :\ |f(x) - f(y)| < \epsilon\ \ \forall\ x, y \in D(f)\ \ \textit{mit}\ \ 0 < |x - y| < \delta. \tag{6.6}$$

BSP. (6.6.2) Die Funktion

$$f(x) := \frac{1}{1+|x|},\ x \in D(f) := \mathbf{R}$$

ist gleichmäßig stetig. Denn für festes $\epsilon > 0$ können wir $\delta(\epsilon) := \epsilon$ unabhängig von $x \in D(f)$ wählen. Es gilt nämlich für alle $x, y \in \mathbf{R}$ mit $0 < |x - y| < \delta$:

$$|f(x) - f(y)| = \left|\frac{1 + |y| - 1 - |x|}{(1+|x|)(1+|y|)}\right| \leq ||x| - |y|| \leq |x - y| < \delta = \epsilon.$$

Im Gegensatz zu diesem Beispiel ist die Funktion $f(x) := \frac{1}{x},\ x \in D(f) := (0, +\infty)$, zwar stetig (man vgl. BSP. (6.5.3) in Abschnitt 6.5), aber *nicht* gleichmäßig stetig. Andernfalls wäre (6.6) erfüllt. Für die spezielle Wahl $\epsilon := 1$ fixieren wir $\delta = \delta(\epsilon)$ gemäß (6.6) und dazu $x := \frac{1}{n},\ y := \frac{1}{n^2}$ für $1 \ll n \in \mathbf{N}$ so, dass $0 < |x-y| = \frac{1}{n}\,(1-\frac{1}{n}) < \delta$ gilt. Dann folgt $|f(x)-f(y)| = n(n-1) \gg 1 = \epsilon$, im Widerspruch zur Bedingung (6.6) der gleichmäßigen Stetigkeit.

Jede gleichmäßig stetige Funktion ist insbesondere stetig. Die Umkehrung dieser Aussage ist im allgemeinen falsch. Umso bemerkenswerter ist das folgende Resultat:

Satz 6.20 *Eine stetige Funktion $f : [a, b] \to \mathbf{R}$ ist auf dem* **abgeschlossenen Intervall** $[a, b] \subset \mathbf{R}$ *sogar gleichmäßig stetig.*

Begründung: Wir nehmen das Gegenteil der Bedingung (6.6) an:

$$\exists\, \epsilon_0 > 0\ \forall\, n \in \mathbf{N}\ \exists\, x_n, y_n \in [a, b]\ :\ |f(x_n) - f(y_n)| \geq \epsilon_0\ \text{ und }\ |x_n - y_n| < \frac{1}{n}.$$

Da die Folge $(x_n)_{n\in\mathbf{N}} \subset [a, b]$ beschränkt ist, besitzt sie nach dem Auswahlsatz von Bolzano–Weierstrass (Satz 3.9) mindestens einen Häufungspunkt. Zu einer Teilfolge $\mathbf{N}' \subset \mathbf{N}$ existiert ein Grenzwert $x_0 \in [a, b]$ mit $\lim\limits_{j\in\mathbf{N}'} x_j = x_0$. Nun gilt offenbar auch $\lim\limits_{j\in\mathbf{N}'} y_j = x_0$, und aus der Stetigkeit von f folgt im Widerspruch zur obigen Bedingung:

$$0 < \epsilon_0 \leq |\lim_{j\in\mathbf{N}'}[f(x_j) - f(y_j)]| = |f(x_0) - f(x_0)| = 0.$$

BSP. (6.6.3) Wie bereits in BSP. (6.6.2) festgestellt wurde, ist die Funktion $f(x) := \frac{1}{x},\ x \in D(f) := (0, +\infty)$, zwar stetig, nicht aber gleichmäßig stetig. Fixieren wir jedoch $a, b \in \mathbf{R}$ mit $0 < a < b < +\infty$, so gilt $[a, b] \subset D(f)$, und mit $\delta(\epsilon) := \epsilon a^2$ folgt für jedes Zahlenpaar $x, y \in [a, b], 0 < |x - y| < \delta$:

$$|f(x) - f(y)| = \frac{|x-y|}{|xy|} \leq \frac{|x-y|}{a^2} < \frac{\delta}{a^2} = \epsilon,$$

also die gleichmäßige Stetigkeit auf dem abgeschlossenen Intervall $[a, b]$.

6.7 Monotone Funktionen, Umkehrfunktionen

Da $\mathbf{R}$ ein angeordneter Körper ist, kann für **reellwertige** Funktionen ein Monotonie–Begriff eingeführt werden.

Definition 6.24 *Gegeben seien eine* **reellwertige** *Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ mit Definitionsbereich $D(f) \subset \mathbf{R}$ und ein Intervall $I \subseteq D(f)$. Die Funktion f heiße auf I (streng)* **monoton wachsend (monoton $\uparrow$)**, *wenn gilt*

$$\boxed{f(x) - f(y) \geq 0\ (bzw.\ > 0)\ \ \forall\, x, y \in I\ \ mit\ x > y.} \tag{7.1}$$

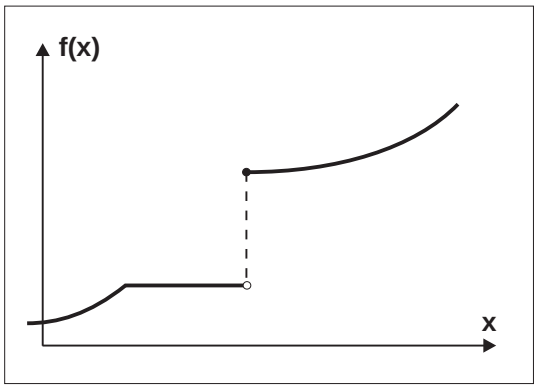
Die Funktion f heiße auf I (streng) **monoton fallend (monoton $\downarrow$)**, *wenn gilt*

$$\boxed{f(x) - f(y) \leq 0\ (bzw.\ < 0)\ \ \forall\, x, y \in I\ \ mit\ x > y.} \tag{7.2}$$

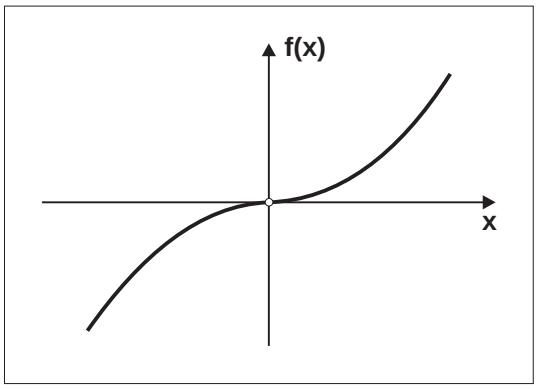
Bemerkung 6.13 Gleichwertig mit (7.1) und (7.2) sind die folgenden Bedingungen:

$$[f(x) - f(y)](x - y) \geq 0 \text{ (bzw. } > 0) \ \forall\, x, y \in I \text{ mit } x \neq y, \tag{7.1.a}$$

$$[f(x) - f(y)](x - y) \leq 0 \text{ (bzw. } < 0) \ \forall\, x, y \in I \text{ mit } x \neq y. \tag{7.2.a}$$



Graph einer (nicht streng) monoton wachsenden Funktion



Graph einer streng monoton wachsenden Funktion

BSP. (6.7.1) Die Funktion

$$\boxed{f(x) := e^x \text{ ist auf } \mathbf{R} \textbf{ streng monoton } \uparrow.}$$

Denn für jedes Zahlenpaar $x, y \in \mathbf{R}$ mit $x - y > 0$ gilt

$$e^x - e^y = e^y\,(e^{x-y} - 1) = e^y \sum_{k=1}^{\infty} \frac{(x-y)^k}{k!} > e^y \cdot (x - y) > 0.$$

Analog zeigt man, dass die Funktion $f(x) := e^{-x}$ auf $\mathbf{R}$ **streng monoton** $\downarrow$ ist.

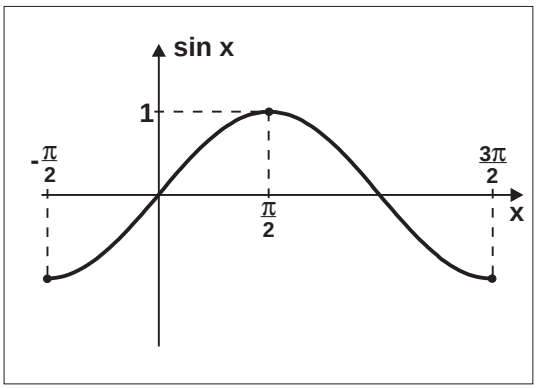
Ist $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ nicht auf dem gesamten Definitionsbereich $D(f)$ monoton, so kann $D(f)$ häufig in **Monotonieintervalle** zerlegt werden, auf denen dann f monoton ist.

BSP. (6.7.2) Wir betrachten die Funktion

$$\boxed{f(x) := \sin x,\ x \in D(f) := \mathbf{R}.}$$

(a) Wir zeigen, dass $\sin x$ auf dem Intervall $I_0 := [-\frac{\pi}{2}, +\frac{\pi}{2}]$ **streng monoton** $\uparrow$ ist. Denn für jedes Zahlenpaar $x, y \in I_0$ mit $x - y > 0$ gilt $-\frac{\pi}{2} < \frac{x+y}{2} < +\frac{\pi}{2}$ sowie $0 < \frac{x-y}{2} \leq \frac{\pi}{2}$, so dass folgt:

$$\sin x - \sin y = 2 \underbrace{\cos\left(\frac{x+y}{2}\right)}_{>0} \underbrace{\sin\left(\frac{x-y}{2}\right)}_{>0} > 0.$$



Monotonieintervalle der Funktion $\sin x$

(b) Auf dem Intervall $\tilde{I}_0 := [\frac{\pi}{2}, \frac{3\pi}{2}]$ ist $f(x) := \sin x$ **streng monoton** $\downarrow$. Denn für jedes Zahlenpaar $x, y \in \tilde{I}_0$ mit $x - y > 0$ gilt $\frac{\pi}{2} < \frac{x+y}{2} < \frac{3\pi}{2}$ sowie $0 < \frac{x-y}{2} \leq \frac{\pi}{2}$, so dass folgt:

$$\sin x - \sin y = 2 \underbrace{\cos\left(\frac{x+y}{2}\right)}_{<0} \underbrace{\sin\left(\frac{x-y}{2}\right)}_{>0} < 0.$$

Da $\sin x$ *periodisch* ist mit der Periode 2π, wiederholen sich die Monotonieintervalle I_0 und $\tilde{I}_0$ bei Verschiebung um $2\pi k, k \in \mathbf{Z}$:

$$\boxed{f(x) := \sin x \text{ ist } \begin{cases} \textbf{streng monoton} \uparrow : \ x \in I_n := [\frac{4n-1}{2}\,\pi, \frac{4n+1}{2}\,\pi], \\ \textbf{streng monoton} \downarrow : \ x \in \tilde{I}_n := [\frac{4n+1}{2}\,\pi, \frac{4n+3}{2}\,\pi], \end{cases} \forall\, n \in \mathbf{Z}.}$$

Ganz analog erhält man:

$$\boxed{f(x) := \cos x \text{ ist } \begin{cases} \textbf{streng monoton} \uparrow : \ x \in I_n := [(2n+1)\,\pi, (2n+2)\,\pi], \\ \textbf{streng monoton} \downarrow : \ x \in \tilde{I}_n := [2n\,\pi, (2n+1)\,\pi], \end{cases} \forall\, n \in \mathbf{Z}.}$$

BSP. (6.7.3) Wir betrachten die Funktion

$$\boxed{f(x) := x^n,\ x \in D(f) := \mathbf{R},\ n \in \mathbf{N}.}$$

Wir zeigen

$$\boxed{f(x) := x^n \text{ ist auf } \begin{cases} I_0 := [0, +\infty) & \textbf{streng monoton} \uparrow : \ n \in \mathbf{N}, \\ \tilde{I}_0 := (-\infty, 0] & \textbf{streng monoton} \uparrow : \ n \text{ ungerade}, \\ \tilde{I}_0 := (-\infty, 0] & \textbf{streng monoton} \downarrow : \ n \text{ gerade}. \end{cases}}$$

(a) Für jedes Zahlenpaar $x, y \in I_0$ mit $x - y > 0$ gilt nämlich:

$$x^n - y^n = (x - y + y)^n - y^n = \sum_{k=1}^{n} \binom{n}{k} \underbrace{(x-y)^k}_{>0} \underbrace{y^{n-k}}_{\geq 0} > 0.$$

Die strikte Ungleichung folgt aus $y^{n-k} = 1$ für $k = n$.

(b) Für jedes Zahlenpaar $x, y \in \tilde{I}_0$ mit $x - y > 0$ gilt $|y| - |x| > 0$ und folglich

$$(-1)^n[x^n - y^n] = -(|y|^n - |x|^n) = -\sum_{k=1}^{n} \binom{n}{k} \underbrace{(|y| - |x|)^k}_{>0} \underbrace{|x|^{n-k}}_{\geq 0} < 0.$$

Die strikte Ungleichung folgt aus $|x|^{n-k} = 1$ für $k = n$.

Ist eine Funktion f **bijektiv**, so existiert die Umkehrfunktion f^{-1}. Dieser Zusammenhang wurde bereits in Abschnitt 6.1 ausführlich erörtert. Das Nachprüfen der Bijektivität erweist sich in vielen Fällen als äußerst schwierig. Anders bei stetigen Funktionen $f : [a, b] \to \mathbf{R}$, die **streng monoton** sind. Eine solche Funktion nimmt die Extremalwerte

$$\min_{x \in [a,b]} f(x) \quad \text{und} \quad \max_{x \in [a,b]} f(x)$$

jeweils in einem der beiden Endpunkte a, b des Intervalls $[a, b]$ an. Somit wird $[a, b]$ durch die Funktion f **surjektiv** auf das Intervall mit den Endpunkten $f(a), f(b)$ abgebildet. Wir zeigen, dass f sogar **bijektiv** ist.

Satz 6.21 (Umkehrsatz für streng monotone Funktionen)
Die Funktion $f : [a,b] \to \mathbf{R}$ *sei stetig und* **streng monoton**. *Dann existiert die Umkehrfunktion* f^{-1} *auf der Bildmenge* $f([a,b])$, *und es gilt*

$$f : [a,b] \to \mathbf{R} \begin{cases} \textbf{streng monoton } \uparrow \\ \textbf{streng monoton } \downarrow \end{cases} \Rightarrow f^{-1} \begin{cases} [f(a),f(b)] \to \mathbf{R} \ \textbf{streng monoton } \uparrow, \\ [f(b),f(a)] \to \mathbf{R} \ \textbf{streng monoton } \downarrow. \end{cases}$$

Darüber hinaus ist f^{-1} *auch stetig.*

Begründung: (a) Die Funktion f sei zum Beispiel streng monoton wachsend. Dann gilt $x > y \ \Leftrightarrow f(x) > f(y)$ für jedes Zahlenpaar $x, y \in [a,b]$. Das heißt, f ist *injektiv*, und die Surjektivität hatten wir schon im Vorspann begründet.

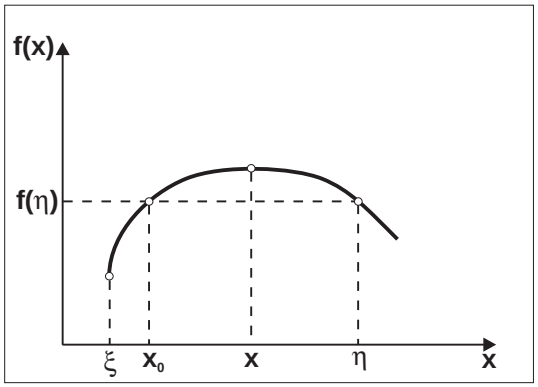
(b) Um die Stetigkeit von f^{-1} zu zeigen, sei $z_0 \in [f(a), f(b))$ fest gewählt. Wir weisen die *rechtsseitige* Stetigkeit von f^{-1} in z_0 nach. Ganz analog verfährt man mit dem Nachweis der *linksseitigen* Stetigkeit in jedem Punkt $z_0 \in (f(a), f(b)]$. Es gelte nun $f(x_0) = z_0$, und es sei $\epsilon > 0$ fest. Dann existiert eine Zahl $x_1 \in (a,b)$ mit $a \le x_0 < x_1 < x_0 + \epsilon \le b$. Wegen der Monotonie von f gibt es ein $\delta > 0$ derart, dass $z_1 := f(x_1) = z_0 + \delta$ gilt. Wir folgern

$$\underbrace{x_0 = f^{-1}(z_0) < f^{-1}(z) < f^{-1}(z_0) + \epsilon}_{\Leftrightarrow \quad 0 < f^{-1}(z) - f^{-1}(z_0) < \epsilon} \ \forall\, z \ \text{ mit } \ z_0 < z < z_0 + \delta.$$

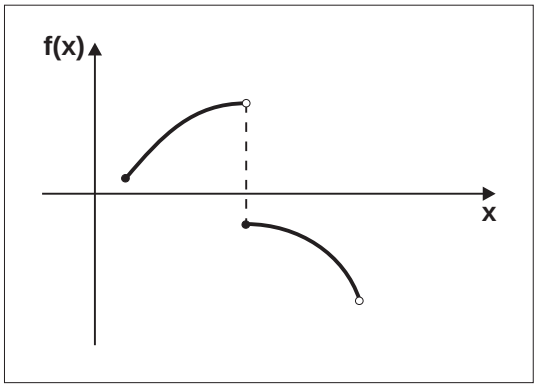
Dies ist die rechtsseitige Stetigkeit im Punkte z_0. □

Bemerkung 6.14 (a) Satz 6.21 gilt auch dann, wenn an die Stelle des abgeschlossenen Intervalls $[a,b]$ ein beliebiges Intervall $I \subseteq D(f)$ tritt.

(b) Die Existenz der Umkehrabbildung f^{-1} unter der Voraussetzung der strengen Monotonie ist auch dann noch gewährleistet, wenn auf die Stetigkeit von f verzichtet wird. Die Bildmenge $f([a,b])$ wird dann allerdings kein Intervall mehr sein sondern in eine Vereinigung paarweise disjunkter Intervalle zerfallen. Wir verweisen auf die Literatur. □



Eine stetige, nicht monotone Funktion f ist i.a. nicht injektiv



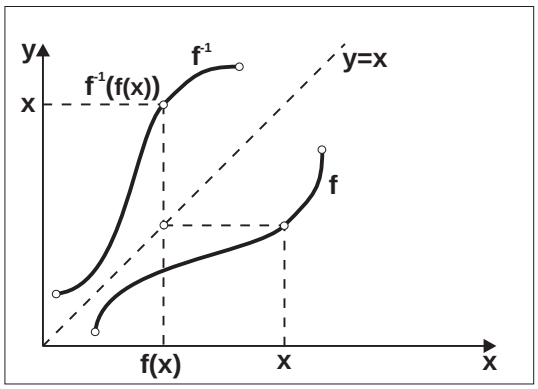
Eine nicht monotone Funktion f kann injektiv sein, wenn f unstetig ist

(c) Bei stetigen Funktionen f ist die strenge Monotonie sogar **notwendig** für die Injektivität. Zum Beweis nehmen wir an, die nicht monotone Funktion f sei injektiv. Dann gibt es Punkte $\xi < x < \eta$ mit $f(\xi) < f(x) > f(\eta) > f(\xi)$. Der Zwischenwertsatz 6.19 sichert nun die Existenz eines Punktes $x_0 \in (\xi, x)$ mit $f(x_0) = f(\eta)$, im Widerspruch zur Injektivität, wonach $x_0 = \eta$ gelten müsste. Bei unstetigen Funktionen f ist diese Schlussweise i.a. falsch.

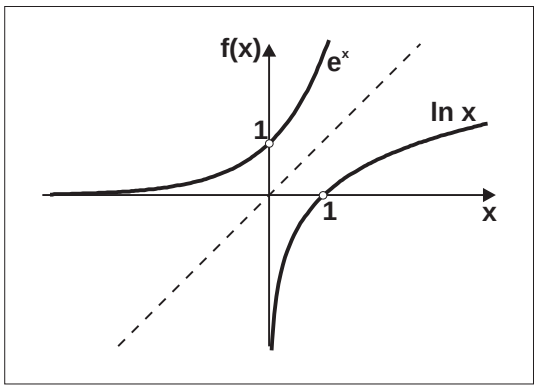
(d) Der **Graph** der Umkehrfunktion f^{-1}, nämlich

$$G(f^{-1}) = \{(y,x) \in \mathbf{R}^2 \ : \ x = f^{-1}(y), \ y \in f([a,b])\},$$

geht offensichtlich aus dem Graphen $G(f) := \{(x,y) \in \mathbf{R}^2 \,:\, y = f(x),\, x \in [a,b]\}$ der Funktion $f : [a,b] \to \mathbf{R}$ durch **Spiegelung** an der Geraden $y = x$ hervor. Dieser Sachverhalt resultiert aus der geometrischen Anschauung unter Berücksichtigung der Identität $f^{-1}[f(x)] = x \;\forall\, x \in [a,b]$.



Der Graph der Umkehrabbildung entsteht durch Spiegelung an der Winkelhalbierenden



Der Logarithmus als Umkehrfunktion der Exponentialfunktion

Anwendungen: Die Inversen der Standardfunktionen.

1.Anwendung: **Der Logarithmus**. Die Exponentialfunktion $f(x) := e^x,\; x \in D(f) := \mathbf{R}$, ist – wie in BSP. (6.7.1) gezeigt wurde – auf ganz $\mathbf{R}$ streng monoton wachsend. Da f außerdem stetig ist, existiert gemäß Satz 6.21 die Umkehrfunktion f^{-1} als stetige Funktion auf der Bildmenge $f(\mathbf{R}) = (0,+\infty)$.

Definition 6.25 *Die Umkehrabbildung der Exponentialfunktion* $\exp : \mathbf{R} \to (0,+\infty)$ *heiße der* **natürliche Logarithmus**, *bezeichnet mit* $\ln : (0,+\infty) \to \mathbf{R}$.

Die Basiseigenschaften des Logarithmus können unmittelbar aus bekannten Eigenschaften der Exponentialfunktion abgeleitet werden. Hierzu zählen *Wachstumseigenschaften*, denen die folgende Eigenschaft der Funktion e^x zugrunde liegt:

$$\boxed{\lim_{x\to+\infty} \frac{x^n}{e^x} = 0 \;\; \forall\, n \in \mathbf{N}.} \tag{7.3}$$

Mit anderen Worten, e^x wächst für $x \to +\infty$ schneller als jede Potenz von x. In der Tat, für $x > 0$ hat man $e^x = \sum\limits_{k=0}^{\infty} \frac{x^k}{k!} > \frac{x^{n+1}}{(n+1)!}$. Hieraus folgt $0 < x^n e^{-x} < \frac{(n+1)!}{x} \to 0$ für $x \to +\infty$.

Satz 6.22 *Der natürliche Logarithmus* $\ln : (0,+\infty) \to \mathbf{R}$ *ist eine stetige, streng monoton wachsende Funktion mit folgenden Eigenschaften:*

(a) $\ln(e^x) = x \;\forall\, x \in \mathbf{R}, \qquad e^{\ln y} = y \;\forall\, y > 0.$

(b) $\ln 1 = 0, \qquad \lim\limits_{x\to 0+} \ln x = -\infty, \qquad \lim\limits_{x\to+\infty} \ln x = +\infty.$

(c) $\ln(xy) = \ln x + \ln y \;\forall\, x,y > 0$. (*Funktionalgleichung*)

(d) $\lim\limits_{x\to+\infty} \dfrac{\ln x}{x^n} = 0, \quad \lim\limits_{x\to 0+} x^n \ln x = 0 \;\; \forall\, n \in \mathbf{N}$. *Mit anderen Worten,* $\ln x$ *wächst für* $x \to +\infty$ *schwächer als jede Potenz von* x.

Begründungen: (b) Aus $e^0 = 1$ folgt sofort $0 = \ln(e^0) = \ln 1$; die restlichen Behauptungen ergeben sich aus $\lim\limits_{x\to+\infty} e^x = +\infty$ und $\lim\limits_{x\to-\infty} e^x = 0+$.

(c) Setzt man $\xi := \ln x$ und $\eta := \ln y$, so gelten $x = e^\xi,\; y = e^\eta$, und es folgt

$$\ln(xy) = \ln(e^\xi\, e^\eta) = \ln(e^{\xi+\eta}) = \xi + \eta = \ln x + \ln y.$$

(d) Wir setzen $y := \ln x$. Aus $x \to +\infty$ folgt nun $y \to +\infty$, und aus $x \to 0+$ folgt ebenso $y \to -\infty$. Hiermit resultiert unter Verwendung von (7.3):

$$\frac{\ln x}{x^n} = \frac{y}{e^{ny}} \to 0 \ (y \to +\infty), \quad x^n \ln x = y\, e^{ny} = \frac{y}{e^{-ny}} \to 0 \ (y \to -\infty).$$

Hiermit ist alles bewiesen. □

Die allgemeine Potenzfunktion und **der allgemeine Logarithmus.**
Mit Hilfe der Funktion $\ln : (0, +\infty) \to \mathbf{R}$ können neue stetige Funktionen erklärt werden.

Definition 6.26 *Für eine feste Zahl $a > 0$ sei die* **allgemeine Potenzfunktion** $f : \mathbf{R} \to (0, +\infty),\ x \mapsto f(x) := a^x$ *durch die folgende Vorschrift erklärt:*

$$\boxed{a^x := e^{x \ln a},\ x \in D(f) := \mathbf{R}.}$$

Für $a \neq 1$ existiert ihre Umkehrfunktion $f^{-1} : (0, +\infty) \to \mathbf{R}$ (siehe Satz 6.22), und diese heiße der **Logarithmus zur Basis** *a:*

$$\boxed{{}^a\log x : (0, +\infty) \to \mathbf{R},\ \ a \neq 1.}$$

Bemerkung 6.15 Häufig wird der BRIGGSsche Logarithmus $\lg x$ verwendet, das ist der Logarithmus zur Basis 10: $\lg x := {}^{10}\log x\ \forall\, x > 0$. □

Satz 6.23 *Die allgemeine Potenzfunktion und der Logarithmus zur Basis a haben folgende Eigenschaften:*

(a) *Die Funktionen $f(x) := a^x$ und $g(x) := {}^a\log x$ sind für festes $a \in (0, 1)$ streng monoton $\downarrow$ und für festes $a > 1$ streng monoton $\uparrow$ sowie stetig in beiden Fällen. Für $a = 1$ gilt $f(x) = a^x = 1\ \forall\, x \in \mathbf{R}$.*

(b) $a^{x+y} = a^x a^y\ \forall\, x, y \in \mathbf{R}, \quad {}^a\log x + {}^a\log y = {}^a\log(xy)\ \forall\, x, y > 0.$

(c) $a^0 = 1, \quad {}^a\log 1 = 0, \quad {}^a\log x = \dfrac{\ln x}{\ln a}\ \forall\, x > 0.$

(d) $$\lim_{x \to +\infty} a^x = \begin{cases} +\infty & : \ a > 1, \\ 0 & : \ 0 < a < 1, \end{cases} \qquad \lim_{x \to -\infty} a^x = \begin{cases} 0 & : \ a > 1, \\ +\infty & : \ 0 < a < 1. \end{cases}$$

(e) $$\lim_{x \to +\infty} {}^a\log x = \begin{cases} +\infty & : \ a > 1, \\ -\infty & : \ 0 < a < 1, \end{cases} \qquad \lim_{x \to 0+} {}^a\log x = \begin{cases} -\infty & : \ a > 1, \\ +\infty & : \ 0 < a < 1. \end{cases}$$

(f) $(a^x)^y = a^{xy}\ \forall\, x, y \in \mathbf{R}, \quad {}^a\log(x^y) = y \cdot {}^a\log x\ \forall\, x > 0,\ y \in \mathbf{R}.$

Begründungen. Verwendet man das Tableaux $\ln a \begin{cases} < 0 & : \ 0 < a < 1, \\ = 0 & : \ a = 1, \\ > 0 & : \ a > 1, \end{cases}$ und beachtet, dass e^x streng monoton $\uparrow$, während e^{-x} streng monoton $\downarrow$, so erhält man:

$$a^x = e^{x \ln a} = \begin{cases} e^{-x\,|\ln a|} & : \ 0 < a < 1, & \text{also streng monoton } \downarrow, \\ e^0 = 1 & : \ a = 1, & \text{also konstant}, \\ e^{x \ln a} & : \ a > 1, & \text{also streng monoton } \uparrow. \end{cases}$$

Darüber hinaus sind die Abbildungen

$$y : \begin{cases} \mathbf{R} \to \mathbf{R}, \\ x \mapsto x \ln a, \end{cases} \qquad \exp : \begin{cases} \mathbf{R} \to (0, +\infty), \\ y \mapsto e^y \end{cases}$$

beide stetig, und dies gilt auch für das Kompositum $(\exp \circ y)(x) = e^{x \ln a} = a^x$. Die restliche Behauptung folgt jetzt aus Satz 6.21.

(b) Es gilt ja $a^{x+y} = e^{(x+y)\ln a} = e^{x \ln a} \cdot e^{y \ln a} = a^x a^y\ \forall\, x, y \in \mathbf{R}$. Setzen wir hier $\xi := a^x, \eta := a^y$ oder äquivalent $x = {}^a\log \xi, y = {}^a\log \eta$, so folgt

$${}^a\log(\xi\eta) = {}^a\log(a^x a^y) = {}^a\log(a^{x+y}) = x + y = {}^a\log \xi + {}^a\log \eta\ \ \forall\, \xi, \eta > 0.$$

(c) Es ist trivialerweise $a^0 = e^{0 \ln a} = 1$. Weiterhin gilt

$$ {}^a\log x = {}^a\log(e^{\ln x}) = {}^a\log\left(e^{\frac{\ln x}{\ln a}\ln a}\right) = {}^a\log\left(a^{\frac{\ln x}{\ln a}}\right) = \frac{\ln x}{\ln a} \quad \forall\, x > 0. $$

Hieraus folgt ${}^a\log 1 = \frac{\ln 1}{\ln a} = 0$.

(d) Diese Aussage erhält man aus den Wachstumseigenschaften von $e^{\alpha x}$ für festes $\alpha \in \mathbf{R}$:

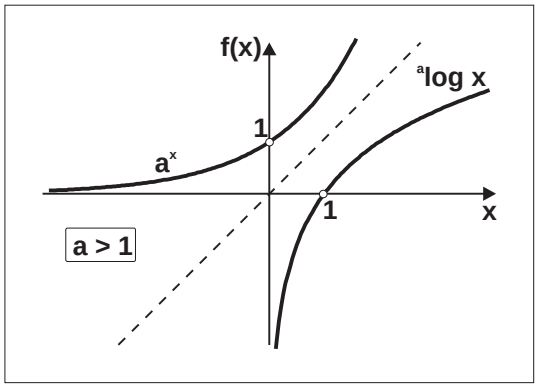
$$ \lim_{x\to+\infty} a^x = \lim_{x\to+\infty} e^{x \ln a} = \begin{cases} +\infty & : \ a > 1, \\ 0 & : \ 0 < a < 1, \end{cases} \qquad \lim_{x\to-\infty} a^x = \lim_{x\to-\infty} e^{x \ln a} = \begin{cases} 0 & : \ a > 1, \\ +\infty & : \ 0 < a < 1. \end{cases} $$

(e) Diese Behauptung folgt unmittelbar aus (d) durch Übergang zur Umkehrabbildung.

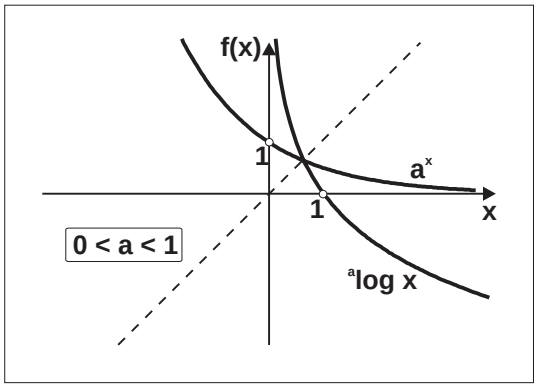
(f) Es gilt $(a^x)^y = [e^{x \ln a}]^y = e^{xy \ln a} = a^{xy} \ \forall\, x, y \in \mathbf{R}$, und schließlich

$$ {}^a\log(x^y) = {}^a\log(e^{y \ln x}) = {}^a\log\left(a^{y\frac{\ln x}{\ln a}}\right) = y\,\frac{\ln x}{\ln a} = y \cdot {}^a\log x \ \ \forall\, x > 0, y \in \mathbf{R}. $$

Hiermit ist alles bewiesen. □



Die allgemeine Potenzfunktion und der allgemeine Logarithmus für $a > 1$



Die allgemeine Potenzfunktion und der allgemeine Logarithmus für $0 < a < 1$

Bemerkung 6.16 Algebraische Verknüpfungen von allgemeinen Logarithmen zu **verschiedenen Basen** können mit Hilfe der Identität (c) aus Satz 6.23 behandelt werden. Es gilt zum Beispiel □

$$ {}^a\log x \cdot {}^x\log y = \frac{\ln x}{\ln a} \cdot \frac{\ln y}{\ln x} = {}^a\log y \ \ \forall\, y > 0, 1 \neq x > 0. $$

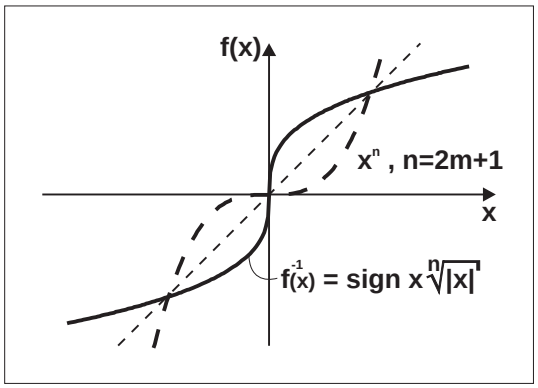
2.Anwendung: **Die Umkehrung der** x**–Potenzen**. Wir betrachten hier die Funktion $f(x) := x^n$, $x \in D(f) := \mathbf{R}, n \in \mathbf{N}$. Wir haben in BSP. (6.7.3) gesehen, dass das Monotonieverhalten von f in den beiden Fällen (i) n *ungerade* und (ii) n *gerade* verschieden ist.

(i) Es sei $n = 2m + 1, m \in \mathbf{N}_0$, eine **ungerade** Zahl. Dann ist die Funktion $f(x) = x^n$ auf ganz $\mathbf{R}$ streng monoton ↑, und somit sichert Satz 6.21 die Existenz ihrer Umkehrfunktion

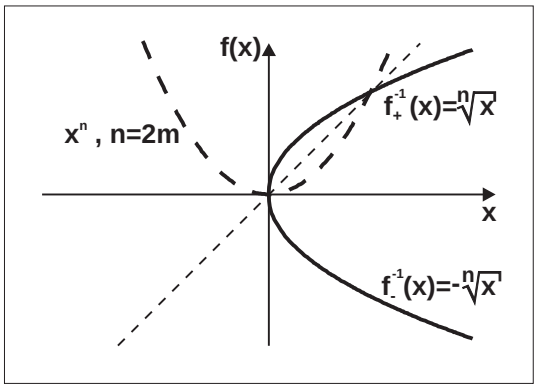
$$ \boxed{f^{-1}(x) = \operatorname{sign} x \ \sqrt[n]{|x|} \ \ \forall\, x \in \mathbf{R}.} $$

(ii) Es sei $n = 2m, m \in \mathbf{N}$, eine **gerade** Zahl. Es gibt zwei Monotonieintervalle $I_0 := [0, +\infty)$ und $\tilde{I}_0 := (-\infty, 0]$. Da $f(x) = x^n$ auf diesen Intervallen jeweils streng monoton ist, existieren Umkehrfunktionen:

$$ \boxed{\begin{array}{ll} f_+^{-1}(x) := \sqrt[n]{x} \ \forall\, x \geq 0 & : \text{ Umkehrfunktion von } f_+(x) := x^n, x \in I_0 = [0, +\infty), \\ f_-^{-1}(x) := -\sqrt[n]{x} \ \forall\, x \geq 0 & : \text{ Umkehrfunktion von } f_-(x) := x^n, x \in \tilde{I}_0 = (-\infty, 0]. \end{array}} $$



Die Umkehrfunktion von
$f(x) := x^{2m+1}, m \in \mathbf{N}_0$



Die beiden Zweige der Umkehrfunktion von $f(x) := x^{2m}, m \in \mathbf{N}$

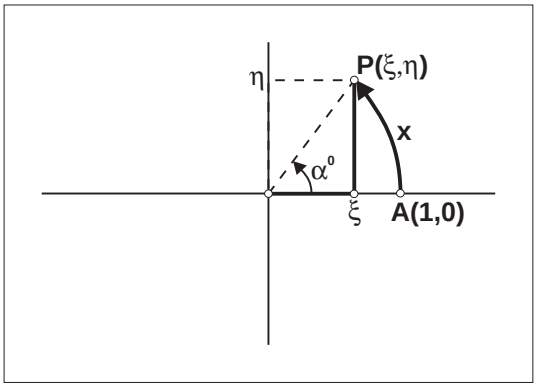
Bemerkung 6.17 Die Funktionen x^n, $\sqrt[n]{x}$ und deren algebraische Verknüpfungen durch "$\pm$","$\cdot$" und ":" sowie deren Hintereinanderausführen heißen **algebraische Funktionen**, *zum Beispiel*

$$f(x) := \sqrt{\frac{x^2-1}{x^3+\sqrt[3]{x^2+1}}} + \sqrt[5]{\frac{2+\sqrt{x}}{5x-8}}.$$

Neben der Klasse der algebraischen Funktionen gibt es die **transzendenten Funktionen**, *zum Beispiel* $f(x) := e^x, := \cos x$, usw. □

3.Anwendung: **Die Umkehrung der trigonometrischen Funktionen (zyklometrische Funktionen)**. Wir betrachten hier zunächst die beiden Funktionen

(I) **Sinus, Cosinus**.



Zur geometrischen Definition von $\sin x$ **und** $\cos x$

Das Argument x der Funktionen $\sin x, \cos x$ wird üblicherweise im **Bogenmaß** angegeben: x ist die Länge des Bogenstückes AP auf der Einheitskreislinie, wobei die Länge der Vollkreislinie vom Radius 1 zu 2π normiert ist. Zur Umrechnung auf das Gradmaß verwendet man die Relation $\alpha^o = \frac{x}{2\pi}\, 360^o$. Eine rationale Näherung für die Kreiszahl π ist gegeben durch

$$\pi \doteq 3.141\,592\,653\,589\,793\,238\,462.$$

Die **geometrische** Definition von $\sin x$ und $\cos x$ wird üblicherweise unter Heranziehung der obigen Skizze vollzogen:

$$\xi := \cos x = \cos(x + 2k\pi), \quad \eta := \sin x = \sin(x + 2k\pi),\ k \in \mathbf{Z}, \quad \cos^2 x + \sin^2 x = 1 \ \ \forall\, x \in \mathbf{R}.$$

Wir hatten bereits in Abschnitt 3.2 einen analytischen Zusammenhang zwischen der komplexen Exponentialfunktion e^{ix} und den trigonometrischen Funktionen $\cos x, \sin x$ hergestellt:

$$\boxed{\begin{array}{rcll} \cos x & = & \frac{1}{2}\left(e^{ix}+e^{-ix}\right) & = \sum\limits_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}, \\ \sin x & = & \frac{1}{2i}\left(e^{ix}-e^{-ix}\right) & = \sum\limits_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}, \end{array} \quad x \in \mathbf{R}.} \tag{7.4}$$

Diese Beziehungen nennt man häufig die **analytische** Definition von $\sin x$ und $\cos x$. Der Vollständigkeit halber listen wir noch die folgenden Relationen auf, die bereits in Abschnitt 2.2 und 2.3 gezeigt wurden.

$$\boxed{\begin{array}{l} e^{inx} = (\cos x + i\,\sin x)^n = \cos nx + i\,\sin nx \;\; \forall\, n \in \mathbf{Z}, \qquad \textbf{Formeln von } \textsc{Moivre}. \\ \left.\begin{array}{rcl} \sin(x \pm y) & = & \sin x \cos y \pm \cos x \sin y, \\ \cos(x \pm y) & = & \cos x \cos y \mp \sin x \sin y, \end{array}\right\} \;\forall\, x, y \in \mathbf{R}, \qquad \textbf{Additionstheoreme.} \end{array}} \tag{7.5}$$

Das Verhalten von $\sin x$ und $\cos x$ in der Nähe von $x = 0$ wird durch die folgenden Beziehungen charakterisiert:

$$\boxed{\lim_{x\to 0} \frac{\sin x}{x} = 1 = \lim_{x\to 0} \frac{2(1-\cos x)}{x^2}, \quad \lim_{x\to 0} \frac{1-\cos x}{x} = 0.} \tag{7.6}$$

Während die erste Limesrelation bereits mit dem Einschließungskriterium begründet wurde, zeigen wir jetzt die beiden restlichen Relationen. Es gilt

$$\frac{1-\cos x}{x^2} = -\sum_{k=1}^{\infty} \frac{(-1)^k x^{2k-2}}{(2k)!} = \frac{1}{2!} - \frac{x^2}{4!} + \frac{x^4}{6!} \mp \dots$$

Für $0 < |x| < 1$ ergibt sich deshalb aus dem LEIBNIZ–Kriterium (Satz 3.18):

$$1 - \frac{2x^2}{4!} \le \frac{2(1-\cos x)}{x^2} \le 1,$$

und dies führt mit Hilfe des Einschließungskriteriums zum behaupteten Grenzwert $\lim\limits_{x\to 0} \frac{2(1-\cos x)}{x^2} = 1$. Wegen $\frac{1-\cos x}{x} = \frac{x}{2} \cdot \frac{2(1-\cos x)}{x^2}$ folgt daraus sofort $\lim\limits_{x\to 0} \frac{1-\cos x}{x} = 0$.

Wie wir in BSP. (6.7.2) gezeigt haben, ist die Funktion $f(x) := \sin x$ auf jedem der Intervalle $[(n-\frac{1}{2})\pi, (n+\frac{1}{2})\pi]$, $n \in \mathbf{Z}$, streng monoton und stetig. Also sichert Satz 6.21 die Existenz von Umkehrfunktionen, und dies trifft ganz analog auch für die Funktion $f(x) := \cos x$ zu.

Definition 6.27 (a) *Die Umkehrfunktionen von* $\sin : [(n-\frac{1}{2})\pi, (n+\frac{1}{2})\pi] \to [-1,+1]$, $n \in \mathbf{Z}$, *heißen* **Zweige des Arcus Sinus**:

$$\boxed{\arcsin_n : [-1,+1] \to [(n-\frac{1}{2})\pi, (n+\frac{1}{2})\pi].}$$

Für $n = 0$ *liegt der* **Hauptwert des Arcus Sinus** *vor, bezeichnet mit*

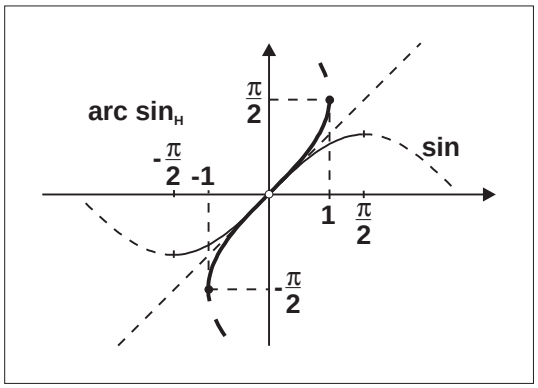
$$\boxed{\arcsin_H : [-1,+1] \to [-\frac{\pi}{2}, +\frac{\pi}{2}].}$$

(b) *Die Umkehrfunktionen von* $\cos : [n\pi, (n+1)\pi] \to [-1,+1]$, $n \in \mathbf{Z}$, *heißen* **Zweige des Arcus Cosinus**:

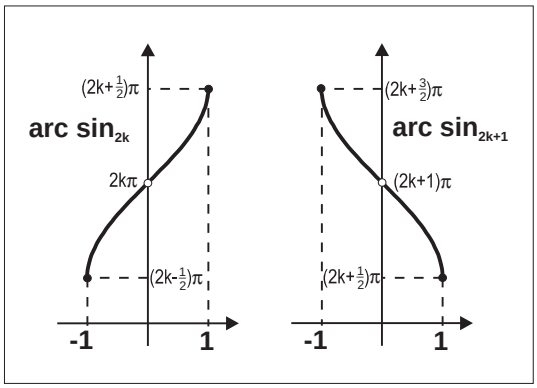
$$\boxed{\arccos_n : [-1,+1] \to [n\pi, (n+1)\pi].}$$

Für $n = 0$ *liegt der* **Hauptwert des Arcus Cosinus** *vor, bezeichnet mit*

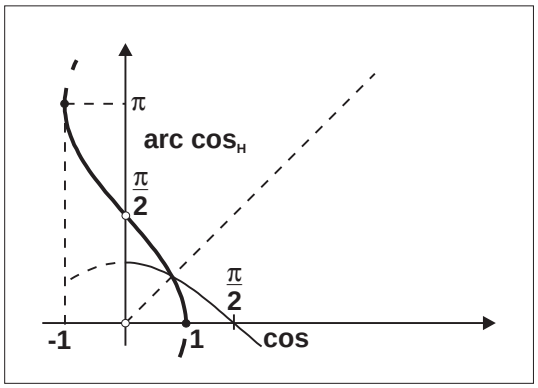
$$\boxed{\arccos_H : [-1,+1] \to [0, \pi].}$$



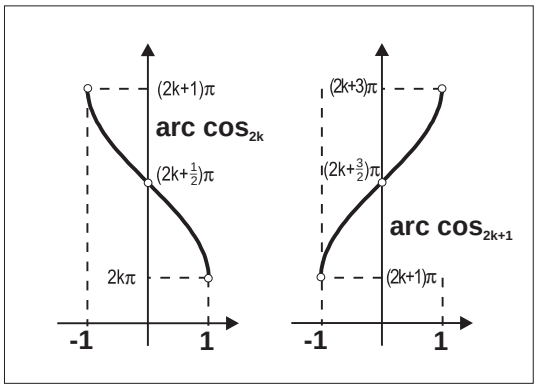
Der Hauptwert der Funktion $\arcsin x$



Zweige der Funktion $\arcsin x$



Der Hauptwert der Funktion $\arccos x$



Zweige der Funktion $\arccos x$

Wir diskutieren nun einige Eigenschaften der zyklometrischen Funktionen und beginnen mit

$$\boxed{\arccos_n y = \arcsin_{n+1} y - \frac{\pi}{2} \;\; \forall\, y \in [-1,+1] \;\forall\, n \in \mathbf{Z}.} \tag{7.7}$$

Begründung: Für $x \in [n\pi, (n+1)\pi]$ hat man $x + \frac{\pi}{2} \in [(n+\frac{1}{2})\pi, (n+\frac{3}{2})\pi]$, und somit folgt aus $y := \cos x = \sin(x + \frac{\pi}{2})$ die Relation

$$x = \arccos_n y, \quad x + \frac{\pi}{2} = \arcsin_{n+1} y.$$

Durch Elimination von x erhält man die behauptete Gleichung (7.7). □

Es ist klar, dass wegen der Beziehung (7.7) nur die Eigenschaften der Funktion Arcus Sinus diskutiert werden müssen.

$$\boxed{\begin{array}{lll} \arcsin_{2k} & : [-1,+1] \to [(2k-\frac{1}{2})\pi, (2k+\frac{1}{2})\pi] & \text{ist stetig und streng monoton } \uparrow, \\ \arcsin_{2k+1} & : [-1,+1] \to [(2k+\frac{1}{2})\pi, (2k+\frac{3}{2})\pi] & \text{ist stetig und streng monoton } \downarrow. \end{array}} \tag{7.8}$$

Dies folgt sofort aus den in BSP. (6.7.2) gezeigten Monotonieeigenschaften der Funktion $f(x) := \sin x$ sowie aus Satz 6.21.

$$\boxed{\begin{array}{lcll} \arcsin_{2k} y & = & \arcsin_H y + 2k\pi & \forall\, y \in [-1,+1] \;\forall\, k \in \mathbf{Z}, \\ \arcsin_{2k+1} y & = & -\arcsin_H y + (2k+1)\pi & \forall\, y \in [-1,+1] \;\forall\, k \in \mathbf{Z}. \end{array}} \tag{7.9}$$

Begründung: Für $x \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$ hat man $x + 2k\pi \in [(2k-\frac{1}{2})\pi, (2k+\frac{1}{2})\pi]$, und somit folgt aus $y := \sin x = \sin(x + 2k\pi)$ die Relation

$$x = \arcsin_H y, \quad x + 2k\pi = \arcsin_{2k} y.$$

Durch Elimination von x erhält man die erste der beiden Gleichungen in (7.9). Die zweite Gleichung folgt aus der Identität $\sin(-x) = \sin(x + (2k+1)\,\pi)$. □

(II) **Tangens, Cotangens**.

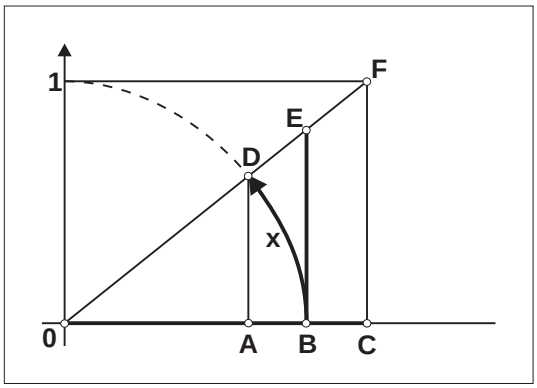
Definition 6.28 *Die Funktion* $\tan : D(\tan) \to \mathbf{R}$ *mit*

$$\boxed{\tan x := \frac{\sin x}{\cos x}, \quad x \in D(\tan) := \mathbf{R} \setminus \{(n + \frac{1}{2})\pi \,:\, n \in \mathbf{Z}\},}$$

heiße **Tangens**. *Die Funktion* $\cot : D(\cot) \to \mathbf{R}$ *mit*

$$\boxed{\cot x := \frac{\cos x}{\sin x}, \quad x \in D(\cot) := \mathbf{R} \setminus \{n\pi \,:\, n \in \mathbf{Z}\},}$$

heiße **Cotangens**. *Es gilt offenbar* $\cot x = \frac{1}{\tan x} \;\forall\, x \in D(\tan) \cap D(\cot)$.



Zur geometrischen Bedeutung der Funktionen $\tan x$ **und** $\cot x$

Die **geometrische** Bedeutung von Tangens und Cotangens lässt sich der obigen Skizze entnehmen:

$$\boxed{\tan x = \overline{EB}, \quad \cot x = \overline{OC}.} \tag{7.10}$$

Begründung: Mit den Bezeichnungen der Skizze folgern wir aus dem Strahlensatz:

$$\begin{aligned} \tan x &= \frac{\sin x}{\cos x} = \frac{\overline{AD}}{\overline{OA}} = \frac{\overline{EB}}{\overline{OB}} = \overline{EB} \;\text{ wegen }\; \overline{OB} = 1, \\ \cot x &= \frac{\cos x}{\sin x} = \frac{\overline{OA}}{\overline{AD}} = \frac{\overline{OC}}{\overline{CF}} = \overline{OC} \;\text{ wegen }\; \overline{CF} = 1. \end{aligned}$$

Hiermit ist alles gezeigt. □

Da die Quotienten stetiger Funktionen wieder stetig sind, folgern wir:

$$\boxed{\tan x \text{ und } \cot x \text{ sind stetig in jedem Punkt } x \in D(\tan) \text{ bzw. } x \in D(\cot).} \tag{7.11}$$

Es gilt ferner wegen $\tan(x + \pi) = \frac{\sin(x+\pi)}{\cos(x+\pi)} = \frac{-\sin x}{-\cos x} = \tan x$:

$$\boxed{\tan x \text{ und } \cot x \text{ sind periodisch mit der Periode } \pi.} \tag{7.12}$$

$$\boxed{\begin{aligned} &\tan x \;\text{ ist in } (-\tfrac{\pi}{2}, +\tfrac{\pi}{2}) \text{ streng monoton } \uparrow; \quad \lim_{x \to \frac{\pi}{2} - 0} \tan x = +\infty, \quad \lim_{x \to -\frac{\pi}{2} + 0} \tan x = -\infty, \\ &\cot x \;\text{ ist in } (0, \pi) \text{ streng monoton } \downarrow; \quad \lim_{x \to 0+} \cot x = +\infty, \quad \lim_{x \to \pi - 0} \cot x = -\infty. \end{aligned}} \tag{7.13}$$

Begründung: Auf dem Intervall $[0, \frac{\pi}{2})$ ist die Funktion $\cos x$ streng monoton $\downarrow$, während $\sin x$ streng monoton $\uparrow$. Der Quotient $\frac{\sin x}{\cos x}$ ist somit streng monoton $\uparrow$, und wegen $\tan(-x) = -\tan x$ gilt diese Monotonieaussage

auch auf dem Intervall $(-\frac{\pi}{2}, 0)$. Ferner erschließen wir aus $\sin x \to 1$ und $\cos x \to 0+$ für $x \to \frac{\pi}{2} - 0$ den Grenzwert $\lim\limits_{x \to \frac{\pi}{2}-0} \tan x = +\infty$, und wegen $\tan(-x) = -\tan x$ folgt hieraus $\lim\limits_{x \to -\frac{\pi}{2}+0} \tan x = -\infty$. Mit ähnlichen Argumenten zeigt man die behaupteten Eigenschaften von $\cot x$. □

$$\boxed{\begin{array}{ll} \tan x = 0 \ \forall\, x = n\pi,\ n \in \mathbf{Z}, & \tan(x + \frac{\pi}{2}) = -\cot x \ \forall\, x \neq n\pi,\ n \in \mathbf{Z}, \\ \cot x = 0 \ \forall\, x = (n + \frac{1}{2})\,\pi,\ n \in \mathbf{Z}, & \cot(x + \frac{\pi}{2}) = -\tan x \ \forall\, x \neq (n + \frac{1}{2})\,\pi,\ n \in \mathbf{Z}. \end{array}} \tag{7.14}$$

Begründung: Es gelten ja die Beziehungen $\sin x = 0, \cos x = (-1)^n$ für $x = n\pi$ sowie $\cos x = 0, \sin x = (-1)^n$ für $x = (n + \frac{1}{2})\,\pi$. Wir haben ferner

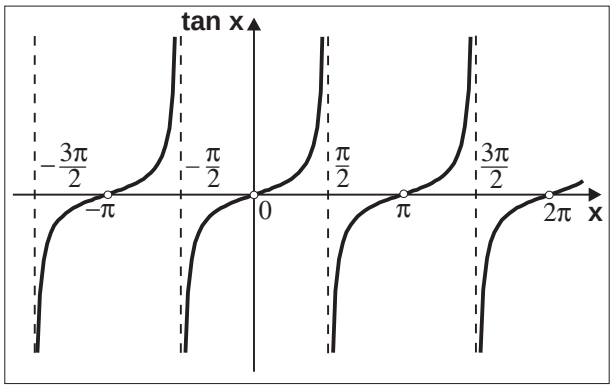
$$\tan(x + \frac{\pi}{2}) = \frac{\cos x}{-\sin x} = -\cot x, \quad \cot(x + \frac{\pi}{2}) = \frac{1}{-\cot x} = -\tan x.$$

Die Additionstheoreme von $\sin x$ und $\cos x$ liefern schließlich noch:

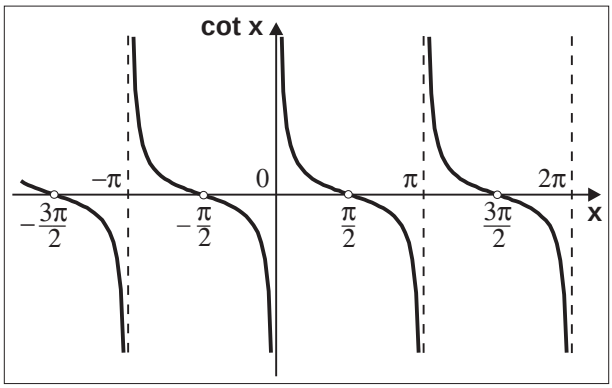
$$\tan(x + y) = \frac{\sin(x+y)}{\cos(x+y)} = \frac{\sin x \cos y + \cos x \sin y}{\cos x \cos y - \sin x \sin y} = \frac{\tan x + \tan y}{1 - \tan x \cdot \tan y}.$$

Allgemeiner folgt:

$$\boxed{\begin{array}{lcll} \tan(x \pm y) & = & \dfrac{\tan x \pm \tan y}{1 \mp \tan x \cdot \tan y} & \forall\, x, y \in \mathbf{R} \ : \ x \pm y \neq (n + \frac{1}{2})\,\pi,\ n \in \mathbf{Z}, \\ \cot(x \pm y) & = & \dfrac{\cot x \cdot \cot y \mp 1}{\cot x \pm \cot y} & \forall\, x, y \in \mathbf{R} \ : \ x \pm y \neq n\pi,\ n \in \mathbf{Z}. \end{array}} \tag{7.15}$$



Der Graph der Funktion $\tan x$



Der Graph der Funktion $\cot x$

Die folgende Tabelle nützlicher Funktionswerte von tan und cot kann häufig zu Rate gezogen werden:

x	0	$30^o \hat{=} \frac{\pi}{6}$	$45^o \hat{=} \frac{\pi}{4}$	$60^o \hat{=} \frac{\pi}{3}$	$90^o \hat{=} \frac{\pi}{2}$	$120^o \hat{=} \frac{2\pi}{3}$	$135^o \hat{=} \frac{3\pi}{4}$	$150^o \hat{=} \frac{5\pi}{6}$	$180^o \hat{=} \pi$
$\tan x$	0	$\frac{1}{3}\sqrt{3}$	1	$\sqrt{3}$	–	$-\sqrt{3}$	-1	$-\frac{1}{3}\sqrt{3}$	0
$\cot x$	–	$\sqrt{3}$	1	$\frac{1}{3}\sqrt{3}$	0	$-\frac{1}{3}\sqrt{3}$	-1	$-\sqrt{3}$	–

Aus den Monotonieeigenschaften (7.13) der Funktionen $\tan x$ und $\cot x$ erschließen wir wieder die Existenz von Umkehrfunktionen.

Definition 6.29 (a) *Die Umkehrfunktionen von* $\tan : \big((n - \frac{1}{2})\pi, (n + \frac{1}{2})\pi\big) \to \mathbf{R}$, $n \in \mathbf{Z}$, *heißen* **Zweige des Arcus Tangens***:*

$$\boxed{\operatorname{arc\,tan}_n = \mathbf{R} \to \Big((n - \frac{1}{2})\pi, (n + \frac{1}{2})\pi\Big).}$$

Für $n = 0$ *liegt der* **Hauptwert des Arcus Tangens** *vor, bezeichnet mit*

$$\boxed{\operatorname{arc\,tan}_H : \mathbf{R} \to \Big(-\frac{\pi}{2}, +\frac{\pi}{2}\Big).}$$

(b) *Die Umkehrfunktionen von* $\cot : \big(n\pi, (n+1)\pi\big) \to \mathbf{R}$, $n \in \mathbf{Z}$, *heißen* **Zweige des Arcus Cotangens**:

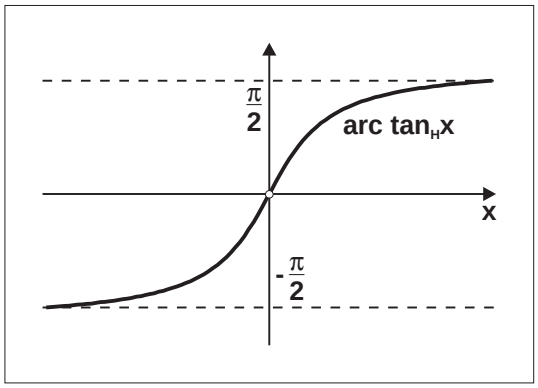
$$\boxed{\operatorname{arc\,cot}_n : \mathbf{R} \to \Big(n\pi, (n+1)\pi\Big).}$$

Für $n = 0$ *liegt der* **Hauptwert des Arcus Cotangens** *vor, bezeichnet mit*

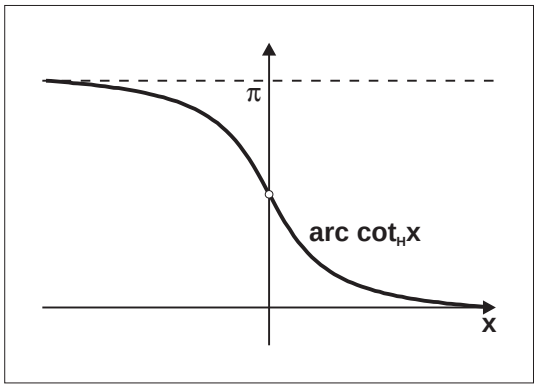
$$\boxed{\operatorname{arc\,cot}_H : \mathbf{R} \to \Big(0, \pi\Big).}$$

$$\boxed{\begin{array}{lll} \operatorname{arc\,tan}_H : \mathbf{R} \to \big(-\frac{\pi}{2}, +\frac{\pi}{2}\big) & \text{ist stetig und streng monoton } \uparrow, & \lim\limits_{x\to\pm\infty} \operatorname{arc\,tan}_H x = \pm\frac{\pi}{2}, \\ \operatorname{arc\,cot}_H : \mathbf{R} \to \big(0, \pi\big) & \text{ist stetig und streng monoton } \downarrow, & \lim\limits_{x\to\pm\infty} \operatorname{arc\,cot}_H x = \begin{cases} 0+, \\ \pi-. \end{cases} \end{array}} \tag{7.16}$$

Diese Aussagen folgen sofort aus den Eigenschaften (7.13) in Verbindung mit Satz 6.21.



Der Hauptwert von Arcus Tangens



Der Hauptwert von Arcus Cotangens

$$\boxed{\begin{array}{lcll} \operatorname{arc\,tan}_n y & = & \operatorname{arc\,tan}_H y + n\pi & \forall\, y \in \mathbf{R}\ \forall\, n \in \mathbf{Z}, \\ \operatorname{arc\,cot}_n y & = & \operatorname{arc\,cot}_H y + n\pi = -\operatorname{arc\,tan}_H y + (n+\frac{1}{2})\pi & \forall\, y \in \mathbf{R}\ \forall\, n \in \mathbf{Z}. \end{array}} \tag{7.17}$$

Begründung: Für $x \in \big(-\frac{\pi}{2}, +\frac{\pi}{2}\big)$ hat man $x + n\pi \in \big((n-\frac{1}{2})\pi, (n+\frac{1}{2})\pi\big)$, und somit folgt aus $y := \tan x = \tan(x + n\pi)$ die Relation

$$x = \operatorname{arc\,tan}_H y, \quad x + n\pi = \operatorname{arc\,tan}_n y.$$

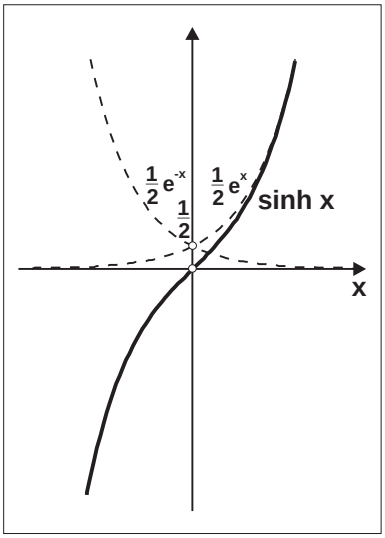
Durch Elimination von x erhält man die erste der behaupteten Gleichungen (7.17). Für $x \in (0, \pi)$ hat man $x+n\pi \in (n\pi, (n+1)\pi)$ sowie $-x-\frac{\pi}{2} \in \big(-\frac{3\pi}{2}, -\frac{\pi}{2}\big)$, und somit folgt aus $y := \cot x = \cot(x+n\pi) = \tan(-x-\frac{\pi}{2})$ die Relation

$$x = \operatorname{arc\,cot}_H y, \quad x + n\pi = \operatorname{arc\,cot}_n y, \quad -x - \frac{\pi}{2} = \operatorname{arc\,tan}_{-1} y = \operatorname{arc\,tan}_H y - \pi.$$

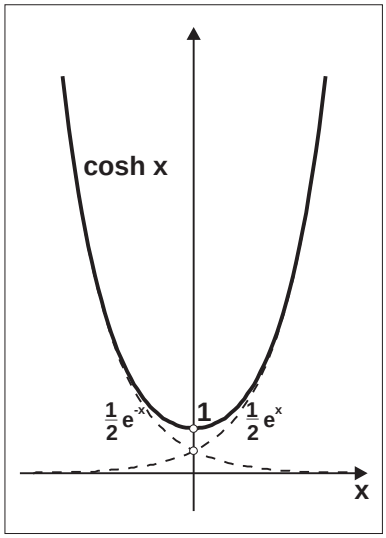
Durch Elimination von x erhält man die restlichen Gleichungen (7.17). □

Die Gleichungen (7.17) gestatten es, alle Werte von Arcus Tangens und Arcus Cotangens über den Hauptwert $\operatorname{arc\,tan}_H x$ zu berechnen.

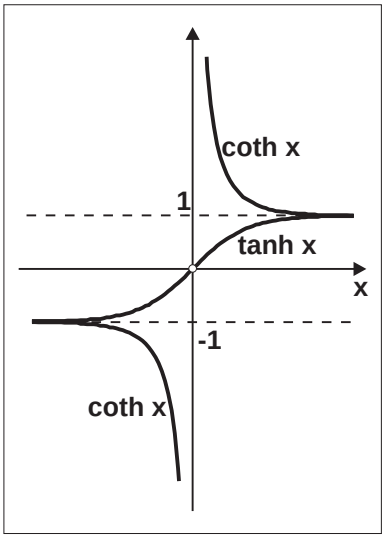
4.Anwendung: **Die Hyperbelfunktionen und ihre Umkehrfunktionen (Area–Funktionen)**. Eine bedeutende Rolle in den technischen und mathematisch–geometrischen Anwendungen spielen die sogenannten *Hyperbelfunktionen*; das sind algebraische Kombinationen der beiden Exponentialfunktionen e^x und e^{-x}. In diesem Sinne führen wir die Hyperbelfunktionen durch **analytische** Definitionen ein.



Der Graph von $\sinh x$



Der Graph von $\cosh x$



Die Graphen von $\tanh x$ **und** $\coth x$

Definition 6.30 *Die* **Hyperbelfunktionen** *seien in der folgenden Weise erklärt:*

Funktionssymbol	Definition	Def.–Bereich	Funktionsname
sinh	$\sinh x := \frac{1}{2}\,(e^x - e^{-x})$	$\mathbf{R}$	**Sinus hyperbolicus**
cosh	$\cosh x := \frac{1}{2}\,(e^x + e^{-x})$	$\mathbf{R}$	**Cosinus hyperbolicus**
tanh	$\tanh x := \dfrac{\sinh x}{\cosh x}$	$\mathbf{R}$	**Tangens hyperbolicus**
coth	$\coth x := \dfrac{\cosh x}{\sinh x}$	$\mathbf{R} \setminus \{0\}$	**Cotangens hyperbolicus**

Man leitet aus diesen Definitionen und aus dem Wert $e^0 = 1$ sofort die folgenden Symmetrie–Eigenschaften ab:

$$\boxed{\begin{array}{ll} \sinh 0 = 0 = \tanh 0, & \cosh 0 = 1, \\ \sinh(-x) = -\sinh x, & \cosh(-x) = \cosh x \ \ \forall\, x \in \mathbf{R} \\ \tanh(-x) = -\tanh x \ \ \forall\, x \in \mathbf{R}, & \coth(-x) = -\coth x \ \ \forall\, x \neq 0. \end{array}} \tag{7.18}$$

Wegen der hier gezeigten Symmetrieen ist es ausreichend, die Funktionsdiskussion auf den Bereich $x > 0$ einzuschränken.

$$\boxed{\begin{array}{lll} 0 < \sinh x < \frac{1}{2}\,e^x \ \ \forall\, x > 0, & \lim\limits_{x\to+\infty} \sinh x = +\infty, & \\ \frac{1}{2}\,e^x < \cosh x \ \ \forall\, x > 0, & \lim\limits_{x\to+\infty} \cosh x = +\infty, & \\ 0 < \tanh x < 1 \ \ \forall\, x > 0, & \lim\limits_{x\to+\infty} \tanh x = 1, & \\ 1 < \coth x \ \ \forall\, x > 0, & \lim\limits_{x\to+\infty} \coth x = 1, & \lim\limits_{x\to 0+} \coth x = +\infty. \end{array}} \tag{7.19}$$

Begründung: Für $x > 0$ gilt ja $e^x > 1$ und $0 < e^{-x} < 1$. Darüber hinaus haben wir $e^x \to +\infty, e^{-x} \to 0$ für $x \to +\infty$. Aus diesen Eigenschaften ergibt sich $0 < \frac{1}{2}\,(e^x - e^{-x}) = \sinh x < \frac{1}{2}\,e^x$ und $\frac{1}{2}\,e^x < \frac{1}{2}\,(e^x + e^{-x}) = \cosh x \to +\infty \ \ (x \to +\infty)$. Weiterhin folgern wir:

$$0 < \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{\sinh x}{\cosh x} \quad = \quad \tanh x = \frac{1 - e^{-2x}}{1 + e^{-2x}} < 1, \quad \lim_{x\to+\infty} \tanh x = 1,$$

$$1 < \frac{e^x + e^{-x}}{e^x - e^{-x}} = \frac{\cosh x}{\sinh x} \quad = \quad \coth x = \frac{1 + e^{-2x}}{1 - e^{-2x}} \to \begin{cases} 1 & : \ x \to +\infty, \\ +\infty & : \ x \to 0+. \end{cases}$$

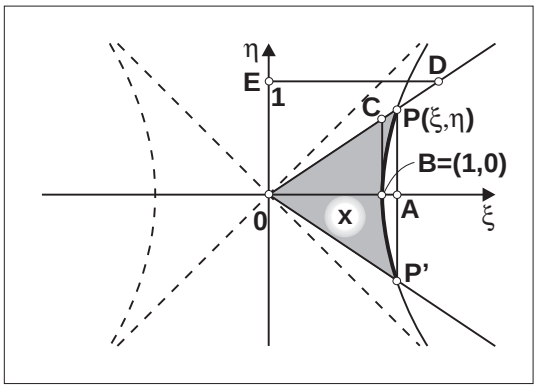
Über die Potenzreihe der Funktion e^x gelangt man sehr schnell zu einer Reihendarstellung der Funktionen $\sinh x$ und $\cosh x$:

$$\boxed{\sinh x = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}, \quad 1 \leq \cosh x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} \ \ \forall\, x \in \mathbf{R}.} \tag{7.20}$$

Begründung: Wir verwenden die Exponentialreihe $e^x = \sum\limits_{n=0}^{\infty} \frac{x^n}{n!}\ \forall\, x \in \mathbf{R}$:

$$\sinh x = \frac{1}{2} \sum_{n=0}^{\infty} \frac{x^n - (-x)^n}{n!} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{x^n}{n!} \left[1 - (-1)^n\right] \overset{n=2k+1}{=} \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$$

$$\cosh x = \frac{1}{2} \sum_{n=0}^{\infty} \frac{x^n + (-x)^n}{n!} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{x^n}{n!} \left[1 + (-1)^n\right] \overset{n=2k}{=} \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots \geq 1.$$



Zur geometrischen Deutung der Hyperbelfunktionen

Die *trigonometrischen* Funktionen konnten geometrisch am *Einheitskreis* gedeutet werden. Ganz analog gibt es eine geometrische Deutung der *Hyperbelfunktionen* an der *Einheitshyperbel* $\xi^2 - \eta^2 = 1$, man vgl. die obige Skizze. Bezeichnet x den Inhalt der Fläche OPP' unter der Hyperbel, so gelten die folgenden Relationen:

$$\sinh x \hat{=} \overline{AP}, \quad \cosh x \hat{=} \overline{OA}, \quad \tanh x \hat{=} \overline{BC}, \quad \coth x \hat{=} \overline{ED}.$$

Da der Hyperbelpunkt $P(\xi, \eta)$ die Gleichung $\xi^2 - \eta^2 = 1$ erfüllt, muss konsequenterweise gelten:

$$\boxed{\cosh^2 x - \sinh^2 x = 1 \ \ \forall\, x \in \mathbf{R}.} \tag{7.21}$$

Umkehrfunktionen der Hyperbelfunktionen. Die an den Graphen der Hyperbelfunktionen ersichtliche strenge Monotonie soll hier nicht im Einzelnen analytisch begründet werden. Wir orientieren uns an der Anschauung, welche die folgende Existenzaussage der Umkehrfunktionen motiviert:

Definition 6.31 (a) *Die Umkehrfunktion der stetigen, streng monoton wachsenden Funktion* $\sinh : \mathbf{R} \to \mathbf{R}$ *heiße* **Area Sinus hyperbolicus**, *bezeichnet mit dem Funktionssymbol*

$$\boxed{\operatorname{Ar}\sinh : \mathbf{R} \to \mathbf{R}.}$$

(b) *Auf den Monotonieintervallen der stetigen Funktion* $\cosh : \begin{cases} [0, +\infty) \to [1, +\infty), \textit{ streng monoton } \uparrow, \\ (-\infty, 0] \to [1, +\infty), \textit{ streng monoton } \downarrow, \end{cases}$

existieren Umkehrfunktionen. Diese heißen **positiver** *und* **negativer Zweig** *des* **Area Cosinus hyperbolicus**, *bezeichnet mit den Funktionssymbolen*

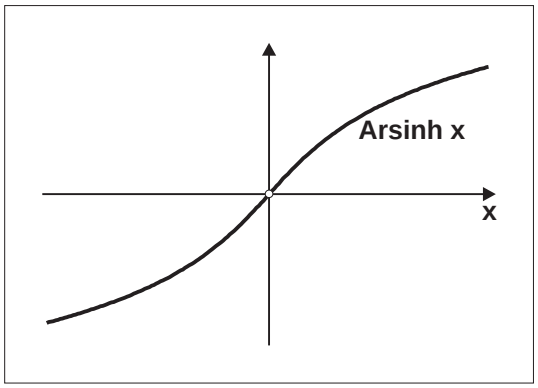
$$\boxed{\operatorname{Ar}\cosh_+ : [1, +\infty) \to [0, +\infty), \quad \operatorname{Ar}\cosh_- : [1, +\infty) \to (-\infty, 0].}$$

(c) *Die Umkehrfunktion der stetigen, streng monoton wachsenden Funktion* $\tanh : \mathbf{R} \to (-1,+1)$ *heiße* **Area Tangens hyperbolicus**, *bezeichnet mit dem Funktionssymbol*

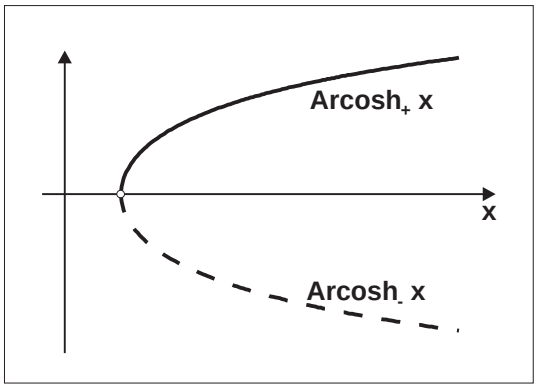
$$\boxed{\operatorname{Ar}\tanh : (-1,+1) \to \mathbf{R}.}$$

(d) *Auf den Stetigkeitsintervallen der streng monoton fallenden Funktion* $\coth : \mathbf{R} \setminus \{0\} \to (-\infty,-1) \cup (+1,+\infty)$ *existiert eine Umkehrfunktion. Diese heiße* **Area Cotangens hyperbolicus**, *bezeichnet mit dem Funktionssymbol*

$$\boxed{\operatorname{Ar}\coth : \mathbf{R} \setminus [-1,+1] \to \mathbf{R} \setminus \{0\}.}$$



Der Graph der Funktion $\operatorname{Ar}\sinh x$

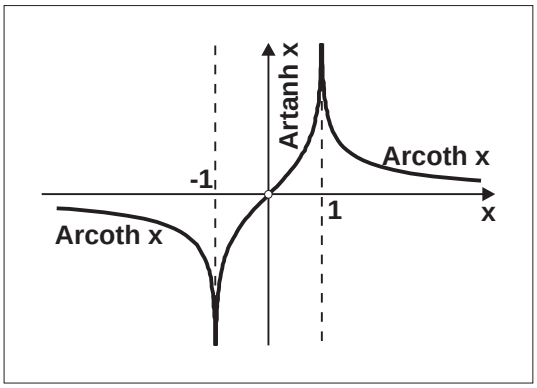


Die beiden Zweige der Funktion $\operatorname{Ar}\cosh x$

Die hier eingeführten Area–Funktionen gestatten folgende analytische Darstellungen:

$$\boxed{\begin{array}{lcll} \operatorname{Ar}\sinh x & = & \ln\left(x+\sqrt{1+x^2}\right) & \forall\, x \in \mathbf{R}, \\ \operatorname{Ar}\cosh_{\pm} x & = & \pm\ln\left(x+\sqrt{x^2-1}\right) & \forall\, x \geq 1, \\ \operatorname{Ar}\tanh x & = & \frac{1}{2}\,\ln\left(\frac{1+x}{1-x}\right) & \forall\, x \in (-1,+1), \\ \operatorname{Ar}\coth x & = & \frac{1}{2}\,\ln\left(\frac{x+1}{x-1}\right) & \forall\, x \in \mathbf{R} \setminus [-1,+1]. \end{array}} \tag{7.22}$$

Begründung: Aus der Darstellung $x := \sinh y = \frac{1}{2}\,(e^y - e^{-y})$ erhält man für e^y die quadratische Gleichung $2xe^y = e^{2y} - 1$ oder äquivalent $(e^y - x)^2 = 1 + x^2$. Die eindeutig bestimmte *positive* Lösung lautet $e^y = x + \sqrt{1+x^2} > 0$, und durch Logarithmieren resultiert die angegebene Darstellung der Funktion $\operatorname{Ar}\sinh x$. Die anderen Darstellungen erhält man in ganz analoger Weise. □



Die Graphen der Umkehrfunktionen
$\operatorname{Ar}\tanh x$ **und** $\operatorname{Ar}\coth x$

5.Anwendung: **Lösung nichtlinearer Gleichungen; Fixpunktprinzipien.** Zu den Standardaufgaben der numerischen Mathematik zählt die Lösung einer Gleichung vom Typ $F(x) = 0$. Wir gehen hier stets davon aus, dass $F \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ eine **stetige** Funktion der **reellen** Veränderlichen $x \in D(F)$ ist. Meistens gelingt es nicht, mit Hilfe endlich vieler algebraischer Operationen diese Lösung herzustellen. *Zum Beispiel* ist man schon bei der Lösung der Gleichung $F(x) := e^{2x} \sin x - 1 = 0$ auf die Verwendung von Näherungsverfahren angewiesen. Das einfachste Verfahren dieser Art ist das

(I) **Verfahren der Intervallschachtelung (Bisektionsverfahren)**. Bekannt sei ein Intervall $I_0 := [a, b] \subseteq D(F)$ mit der Eigenschaft $F(a)\,F(b) < 0$. Dann sichert der Nullstellensatz von Bolzano (Satz 6.18) die Existenz mindestens einer Nullstelle $x_0 \in (a, b)$. Man halbiert nun das Intervall I_0 durch den Punkt $x_1 := (a + b)/2$ in zwei Teilintervalle $[a, x_1]$ und $[x_1, b]$. Gilt $F(x_1) = 0$, so ist die Nullstelle bereits gefunden. Gilt hingegen $F(a)\,F(x_1) < 0$, so liegt x_0 im Inneren des Intervalls $I_1 := [a, x_1]$, und man halbiere nun I_1 durch den Punkt $x_2 := (a + x_1)/2$ usw. Falls $F(a)\,F(x_1) > 0$ gilt, so braucht man nur $I_1 := [x_1, b]$ zu setzen, und man verfährt wie vorher. Das hier beschriebene Verfahren der Intervallhalbierung liefert immer eine konvergente Folge $(x_k)_{k\geq 1}$, deren Grenzwert x_0 eine Lösung der Gleichung $F(x) = 0$ ist. Denn die Längen $L_k := (b - a)/2^k$ der Teilintervalle $I_k, k = 0, 1, \ldots$, bilden eine Nullfolge. Der Mittelpunkt x_{k+1} des Intervalls I_k ist eine Näherung für x_0, das heißt, es gilt eine **a priori– Fehlerschranke**

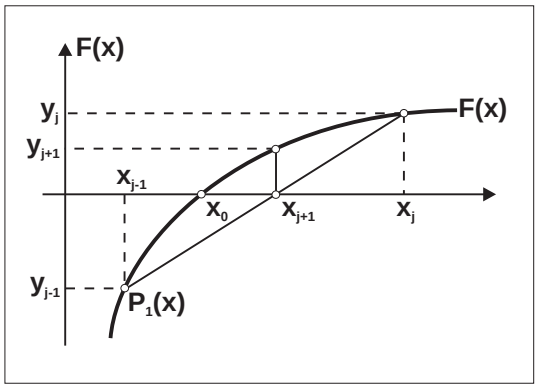
$$\boxed{|x_{k+1} - x_0| \leq \frac{b-a}{2^{k+1}}, \quad k = 0, 1, \ldots}$$

Das Verfahren wird abgebrochen, wenn die Länge L_k des Intervalls I_k unterhalb einer vorgegebenen Genauigkeit ϵ liegt.

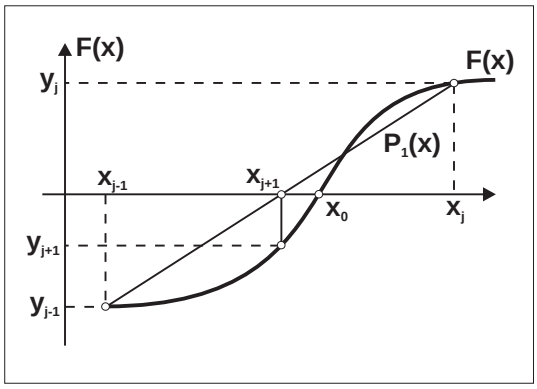
Algorithmus der Intervallschachtelung.

```
1:   Einlesen von a, b, ε;   x := 0.5 * (a + b);
2:   wiederhole:
3:      falls F(a) * F(x) < 0
4:        dann b := x                                   (7.23)
5:        sonst a := x;
6:      x := 0.5 * (a + b);
7:   bis b − a < ε.
```

Der *Nachteil* der Intervallschachtelung ist ihre langsame Konvergenz. Wird zum Beispiel $a = 0, b = 1$ angenommen, und soll eine Genauigkeit $|x_{k+1} - x_0| < 10^{-10}$ erreicht werden, so sind dafür $k > 10 \ln 10/\ln 2 - 1 \doteq 32.2$ Iterationen erforderlich.



Zur Regula falsi: Einschließung von x_0 im Teilintervall $[x_{j-1}, x_{j+1}]$



Zur Regula falsi: Einschließung von x_0 im Teilintervall $[x_{j+1}, x_j]$

(II) **Regula falsi.** (Der Name entspringt mitteralterlichem Latein: *Die Regel, vom Falschen ausgehend.*) Wie vorher gehen wir von der Kenntnis eines Intervalls $I_1 := [x_1, x_2] \subseteq D(F)$ mit $F(x_1)\,F(x_2) < 0$ aus. Dann sind x_1 und x_2 Näherungen der zu bestimmenden Lösung x_0 der Gleichung $F(x) = 0$. Nun wird anstelle der Funktion F die **interpolierende Gerade** zwischen den Stützstellen $(x_j, y_j := F(x_j)),\ j = 1, 2$, betrachtet:

$$P_1(x) := y_{j-1} + (x - x_{j-1})\,\frac{y_j - y_{j-1}}{x_j - x_{j-1}},\ x \in I_{j-1}, j = 2.$$

Ihre Nullstelle

$$\boxed{x_{j+1} := \frac{x_{j-1}y_j - x_j y_{j-1}}{y_j - y_{j-1}}, j = 2,} \tag{7.24}$$

ist ein **neuer** Näherungswert für x_0. Man verfährt nun wie bei der Intervallschachtelung. Mit dem Wert $y_{j+1} := F(x_{j+1})$ fragt man das Vorzeichen von $y_{j+1}\, y_{j-1}$ ab. Gilt $y_{j+1}\, y_{j-1} < 0$, so liegt x_0 im Inneren des Intervalls $I_j := [x_{j-1}, x_{j+1}]$, und man verfahre mit diesem Intervall wie vorher. Gilt hingegen $y_{j+1}\, y_{j-1} > 0$, so setze man $I_j := [x_{j+1}, x_j]$. Diese Situationen werden durch die obigen Skizzen beschrieben.

Liegt $F(x_{j+1}) = 0$ vor, so wird das Verfahren abgebrochen; $x_0 := x_{j+1}$ ist eine Nullstelle. Im allgemeinen wird dieser Fall jedoch nicht eintreten, sondern die oben ermittelte Folge der x_j wird jeweils nur Näherungswerte für x_0 liefern. Dass die Folge tatsächlich gegen x_0 konvergiert, soll hier nicht gezeigt werden. Es ist zweckmäßig, ein **Abbruchkriterium** für das Verfahren vorzugeben. Man iteriert solange, bis zu vorgegebener Toleranz $\epsilon > 0$ der Wert $|F(x_{j+1})| < \epsilon$ erreicht wird.

Algorithmus der Regula falsi.

```
1:  Einlesen von x_1, x_2, ε;   y_1 := F(x_1); y_2 := F(x_2);
2:  wiederhole:
3:      x := (x_1 * y_2 − x_2 * y_1)/(y_2 − y_1); y := F(x);
4:      falls y * y_1 < 0                                          (7.25)
5:        dann x_2 := x; y_2 := y
6:        sonst x_1 := x; y_1 := y;
7:  bis |y| < ε.
```

In diesem Algorithmus gibt die Variable x nach Beendigung der Iterationen die gesuchte Näherung von x_0 an.

(III) **Das Sekantenverfahren.** Dieses Verfahren kann als Modifikation der Regula falsi betrachtet werden. Bei Vorgabe der zwei Startnäherungen x_1 und x_2, die **nicht** die gesuchte Lösung x_0 der Gleichung $F(x) = 0$ einschließen müssen, wird eine verbesserte Näherung x_3 wie bei der Regula falsi mit Hilfe von (7.24) bestimmt. Anders als bei der Regula falsi fragt man jetzt **nicht** das Vorzeichen von $y_{j+1}\, y_{j-1}$ ab, sondern man bestimmt stets eine neue Näherung x_{j+1} als Nullstelle der interpolierenden Geraden durch die beiden vorangegangenen Stützstellen $(x_j, y_j = F(x_j))$ und $(x_{j-1}, y_{j-1} = F(x_{j-1}))$. Hierzu wird die Beziehung (7.24) in modifizierter Form verwendet:

$$\boxed{x_{j+1} = x_j - y_j \cdot \frac{x_j - x_{j-1}}{y_j - y_{j-1}},\ j = 2, 3, \ldots} \tag{7.26}$$

Die Konvergenz der durch (7.26) definierten Folge $(x_j)_{j\geq 1}$ gegen x_0 ist unter geeigneten Voraussetzungen an F beweisbar. Man kann ebenso zeigen, dass das Sekantenverfahren von den hier vorgestellten drei Verfahren die "besten" Konvergenzeigenschaften hat, nämlich am schnellsten konvergiert.

Algorithmus des Sekantenverfahrens.

```
1:  Einlesen von x_1, x_2, ε;   y_1 := F(x_1); y_2 := F(x_2);
2:  wiederhole:
3:      x := x_2 − y_2 * (x_2 − x_1)/(y_2 − y_1); y := F(x);       (7.27)
4:      x_1 := x_2; x_2 := x; y_1 := y_2; y_2 := y;
5:  bis |y| < ε.
```

Auch in diesem Algorithmus gibt die Variable x nach Beendigung der Iterationen die gesuchte Näherung von x_0 an.

BSP. (6.7.4) Wir testen die drei Verfahren zum Vergleich an dem Beispiel $F(x) := e^{2x} \sin x - 1$. Zu bestimmen ist die kleinste positive Lösung der Gleichung $F(x) = 0$. Als Grundintervall verwenden wir $I := [0.4, 0.5]$, denn es gilt $F(0.4) \doteq -0.133\,333$ und $F(0.5) \doteq 0.303\,214$. Die Genauigkeitsvorgabe soll $\epsilon := 5 \cdot 10^{-10}$ betragen.

Intervallschachtelung.

j	a_j	b_j	x_j	$F(x_j)$
0	0.400 000 000	0.500 000 000	0.450 000 000	0.069 842 581
1	0.400 000 000	0.450 000 000	0.425 000 000	−0.035 314 981
2	0.425 000 000	0.450 000 000	0.437 500 000	0.016 346 506
3	0.425 000 000	0.437 500 000	0.431 250 000	−0.009 710 404
4	0.431 250 000	0.437 500 000	0.434 375 000	0.003 261 118
5	0.431 250 000	0.434 375 000	0.432 812 500	−0.003 238 827
⋮	⋮	⋮	⋮	⋮
25	0.433 591 923	0.433 591 926	0.433 591 925	−0.000 000 000
26	0.433 591 925	0.433 591 926	0.433 591 925	0.000 000 003
27	0.433 591 925	0.433 591 925	0.433 591 925	0.000 000 001
28	0.433 591 925	0.433 591 925	0.433 591 925	0.000 000 000

Die Länge des letzten Intervalles ist $L_{28} \doteq 3.73 \cdot 10^{-10}$, so dass für den letzten Intervallmittelpunkt die Fehlerabschätzung $|x^{(28)} - x| \leq 1.87 \cdot 10^{-10}$ gilt.

Regula falsi.

j	x_{j-1}	x_j	x_{j+1}	$F(x_{j+1})$
1	0.400 000 000	0.500 000 000	0.430 542 750	−0.012 630 398
2	0.430 542 750	0.500 000 000	0.433 320 299	−0.001 129 516
3	0.433 320 299	0.500 000 000	0.433 567 769	−0.000 100 482
4	0.433 567 769	0.500 000 000	0.433 589 777	−0.000 008 935
5	0.433 589 777	0.500 000 000	0.433 591 734	−0.000 000 794
6	0.433 591 734	0.500 000 000	0.433 591 908	−0.000 000 071
7	0.433 591 908	0.500 000 000	0.433 591 923	−0.000 000 006
8	0.433 591 923	0.500 000 000	0.433 591 925	−0.000 000 001
9	0.433 591 925	0.500 000 000	0.433 591 925	−0.000 000 000

Die Funktion $F(x)$ ist im Intervall I **konvex**, so dass das rechte Intervallende während der Iteration fixiert bleibt.

Sekantenverfahren.

j	x_j	$F(x_j)$	$\lvert x_j - x_0 \rvert$
0	0.400 000 000	−0.133 333 541	0.033 591 925
1	0.500 000 000	0.303 213 730	0.066 408 075
2	0.430 542 750	−0.012 630 398	0.003 049 175
3	0.433 320 299	−0.001 129 516	0.000 271 625
4	0.433 593 086	0.000 004 831	0.000 001 161
5	0.433 591 924	−0.000 000 002	0.000 000 000
6	0.433 591 925	0.000 000 000	0.000 000 000

Hier wurde zur Berechnung der letzten Spalte als bestmöglicher Wert $x_0 \doteq 0.433\,591\,925$ gewählt.

Äquivalent mit der Lösung der Gleichung $F(x) = 0$ ist das Lösen einer Gleichung vom Typ $x = f(x)$ (dazu setze man zum Beispiel $f(x) := F(x) + x$).

Definition 6.32 *Ein Punkt $x \in D(f)$ heiße ein* **Fixpunkt** *der Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$, wenn $x = f(x)$ gilt.*

Es liegt nahe, Fixpunkte der Funktion f mit der einfachen Iterationsvorschrift $x_{n+1} := f(x_n)$, $n \in \mathbf{N}_0$, bei gegebenem Startwert $x_0 \in D(f)$ zu berechnen. Wir geben nachfolgend zwei Kriterien an, wann eine solche Iterationsvorschrift einen Fixpunkt liefert.

Satz 6.24 *Die stetige Funktion $f : [a,b] \to \mathbf{R}$ erfülle folgende Bedingungen:*

$$\boxed{\text{(i)}\ \ f([a,b]) \subseteq [a,b], \qquad \text{(ii)}\ \ f : [a,b] \to \mathbf{R}\ \ \textit{monoton} \uparrow.}$$

Dann konvergiert die durch die Vorschrift

$$\boxed{x_{n+1} := f(x_n),\ \ n \in \mathbf{N}_0,\ \ x_0 \in [a,b]\ \textit{beliebig},} \tag{7.28}$$

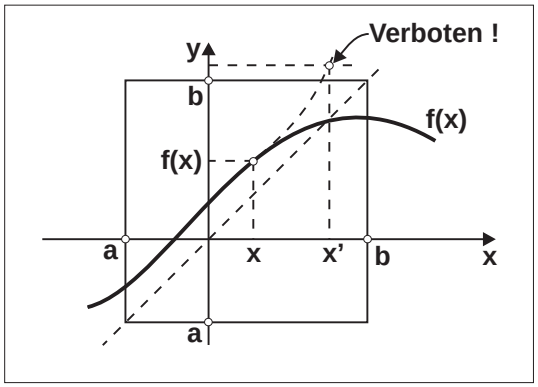
definierte Folge $(x_n)_{n\geq 0}$ monoton gegen einen Fixpunkt $x = f(x) \in [a,b]$.

Begründung: Aus der Voraussetzung (i) folgt, dass mit dem Startwert $x_0 \in [a,b]$ auch jede Iterierte $x_n, n \in \mathbf{N}$, im Intervall $[a,b]$ liegt. Wir treffen zwei Fallunterscheidungen gemäß (a) $x_1 \leq x_0$ und (b) $x_1 > x_0$. Im Falle (a) folgt aus der Monotonie $x_1 = f(x_0) \geq f(x_1) = x_2$, und somit sukzessive $x_n \geq x_{n+1}\ \forall\ n \in \mathbf{N}$. Die beschränkte Folge $(x_n)_{n\geq 0} \subset [a,b]$ ist monoton fallend und folglich konvergent. Im Falle (b) gilt $x_1 = f(x_0) \leq f(x_1) = x_2$, und somit sukzessive $x_n \leq x_{n+1}\ \forall\, n \in \mathbf{N}$. Nun ist die beschränkte Folge $(x_n)_{n\geq 0}$ monoton wachsend und somit wieder konvergent. Es sei $x \in [a,b]$ deren Grenzwert. Aus der Stetigkeit von f erschließt man $x = \lim\limits_{n\to\infty} x_{n+1} = \lim\limits_{n\to\infty} f(x_n) = f(x)$. Also ist x der behauptete Fixpunkt. □

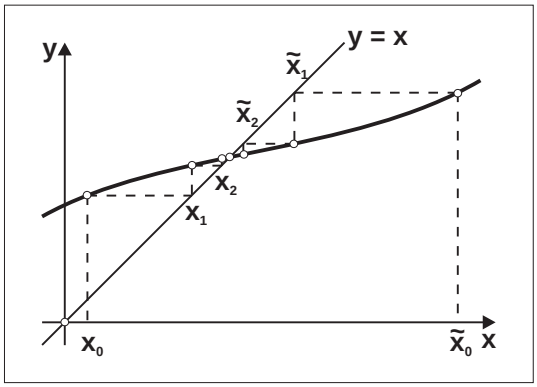
Bemerkung 6.18 (a) Die Bedingung (i) in Satz 6.24 besagt, dass der Graph $G(f)$ der Funktion f weder an der oberen noch an der unteren Kante aus dem Kasten $[a,b] \times [a,b]$ herausspringen darf.

(b) Im rechten oberen Bild weiter unten auf dieser Seite erkennt man das monotone Konvergenzverhalten der Iterationen (7.28).

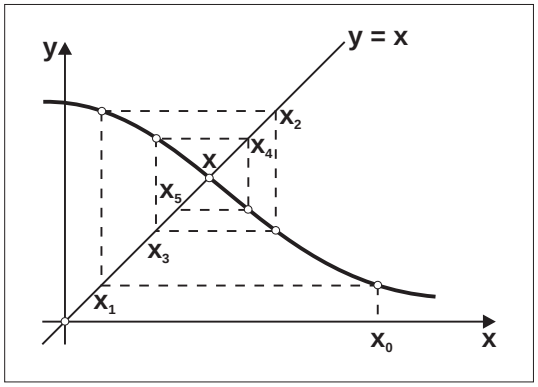
(c) Am Beispiel der Funktion $f(x) := x$ wird klar, dass die Funktion f mehrere, ja sogar beliebig viele Fixpunkte haben kann.



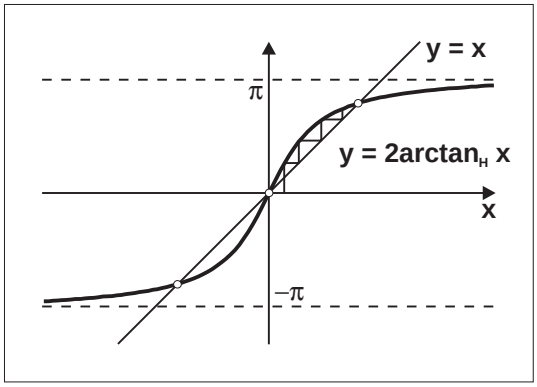
Der Graph $G(f)$ darf den Kasten $[a,b] \times [a,b]$ weder oben noch unten verlassen



Monotone Konvergenz, je nach Lage des Startwertes x_0



Fixpunktiteration bei monoton fallender Funktion



Die Funktion $f(x) := 2 \operatorname{arc\,tan}_H x$ hat genau drei Fixpunkte

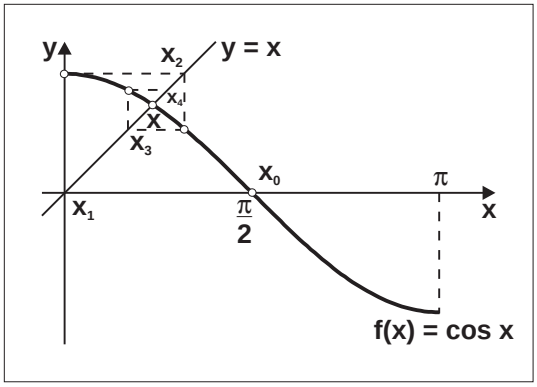
(d) Ist die Funktion $f : [a,b] \to \mathbf{R}$ **monoton fallend**, so bleibt die Folge (7.28) unter der Voraussetzung (i) auch dann noch konvergent gegen einen Fixpunkt $x = f(x)$, wenn f **nicht zu stark fällt**. Die genaue Bedingung $|f'(x)| < 1 \ \forall \ x \in [a,b]$ werden wir erst in Abschnitt 7.8 begründen können. Allerdings geht in diesem Fall die monotone Konvergenz verloren. Die Folge $(x_n)_{n\geq 0}$ alterniert um den Fixpunkt x. □

BSP. (6.7.5) Es seien $[a,b] := [-\pi, +\pi]$ und $f(x) := 2 \operatorname{arc\,tan}_H x$ gegeben. Wegen $|f(x)| < \pi$ ist die Bedingung (i) in Satz 6.24 erfüllt, und die Monotoniebedingung (ii) gilt ebenfalls. Die obigen Skizze liefert die Vorabinformation, dass f drei Fixpunkte hat; einer davon ist $x = 0$. Man entnimmt der Skizze auch folgende weitere Information. Bei Wahl eines Startwertes $x_0 > 0$ (bzw. $x_0 < 0$) konvergiert die Folge (7.28) gegen den Fixpunkt $x > 0$ (bzw. $x < 0$). Der Fixpunkt $x = 0$ kann mit keinem Startwert $x_0 \neq 0$ erreicht werden. Wir haben in der folgenden Tabelle die numerischen Werte der Iteration (7.28) zusammengefasst, die bei Wahl des Startwertes $x_0 = 2$ resultieren. Die Iteration bricht ab, wenn die Genauigkeit $|x_{n+1} - x_n| < 10^{-9}$ erreicht wird.

Fixpunktiteration für $f(x) := 2 \operatorname{arc\,tan}_H x$

n	x_n	$f(x_n)$	$\lvert x_n - x\rvert$
0	2.000 000 000	2.214 297 436	0.331 122 370
1	2.214 297 436	2.293 207 809	0.116 824 935
2	2.293 207 809	2.319 172 917	0.037 914 561
3	2.319 172 917	2.327 391 829	0.011 949 453
4	2.327 391 829	2.329 961 191	0.003 730 541
5	2.329 961 191	2.330 761 275	0.001 161 179
⋮	⋮	⋮	⋮
13	2.331 122 269	2.331 122 339	0.000 000 101
14	2.331 122 339	2.331 122 361	0.000 000 031
15	2.331 122 361	2.331 122 367	0.000 000 009
16	2.331 122 367	2.331 122 369	0.000 000 003
17	2.331 122 369	2.331 122 370	0.000 000 001
18	2.331 122 370	2.331 122 370	0.000 000 000

BSP. (6.7.6) Es seien $[a,b] := [0, \frac{\pi}{2}]$ und $f(x) := \cos x$ gegeben. Wegen $0 \leq \cos x \leq 1 \ \forall \ x \in [a,b]$ ist sicher die Bedingung (i) in Satz 6.24 erfüllt. Darüber hinaus ist $\cos x$ auf dem Intervall $[a,b]$



Zur Fixpunktiteration $x_{n+1} = \cos x_n$

monoton fallend. Die obige Skizze zeigt, dass im Intervall $(0, \frac{\pi}{2})$ genau ein Fixpunkt liegen muss, gegen den die Folge (7.28) alternierend konvergiert. Die Ergebnisse der numerischen Rechnung sind in der obigen Tabelle zusammengefasst, wobei der Startwert $x_0 = \pi/2$ gewählt wurde. Die Iteration bricht ab, wenn die Genauigkeit $|x_{n+1} - x_n| < 10^{-9}$ erreicht wird.

Fixpunktiteration für $f(x) := \cos x$

.

n	x_n	$f(x_n)$	$\lvert x_n - x\rvert$
0	1.570 796 327	0.000 000 000	0.831 711 194
1	0.000 000 000	1.000 000 000	0.739 085 133
2	1.000 000 000	0.540 302 306	0.260 914 867
3	0.540 302 306	0.857 553 216	0.198 782 827
4	0.857 553 216	0.654 289 790	0.118 468 083
5	0.654 289 790	0.793 480 359	0.084 795 343
⋮	⋮	⋮	⋮
46	0.739 085 141	0.739 085 128	0.000 000 008
47	0.739 085 128	0.739 085 137	0.000 000 005
48	0.739 085 137	0.739 085 131	0.000 000 003
49	0.739 085 131	0.739 085 135	0.000 000 002
50	0.739 085 135	0.739 085 132	0.000 000 002
51	0.739 085 132	0.739 085 134	0.000 000 001
52	0.739 085 134	0.739 085 133	0.000 000 001
53	0.739 085 133	0.739 085 134	0.000 000 001
54	0.739 085 134	0.739 085 133	0.000 000 000
55	0.739 085 133	0.739 085 133	0.000 000 000

Der *Nachteil* des Fixpunktsatzes 6.24 liegt in der Tatsache, dass weder eine Eindeutigkeitsaussage noch eine Abschätzung des Fehlers $|x_n - x|$ möglich ist. Beide Nachteile beheben wir in einem weiteren Fixpunktsatz, dem wir aber eine Verschärfung des Begriffes der LIPSCHITZ–Stetigkeit vorausschicken.

Definition 6.33 *Eine Funktion $f : D(f) \to \mathbf{K}$ heiße auf einem Intervall $I \subseteq D(f)$* **kontrahierend**, *wenn gilt*

$$\boxed{\exists\, q \in [0,1) \,:\, |f(x) - f(y)| \le q\,|x - y| \;\; \forall\, x, y \in I.} \tag{7.29}$$

Kontrahierende Funktionen sind also LIPSCHITZ–stetige Funktionen mit einer LIPSCHITZ–Konstanten $L < 1$.

BSP. (6.7.7) Es seien $f(x) := \ln x$ und $I := [a, b]$ mit $1 < a < b$ gegeben. Für $x, y \in I$ setzen wir $z := \ln \frac{x}{y}$. Dann gilt

$$\frac{|\ln x - \ln y|}{|x - y|} = \frac{|\ln \frac{x}{y}|}{|x||1 - \frac{y}{x}|} = \frac{|\ln \frac{x}{y}|}{|y||1 - \frac{x}{y}|} \le \frac{|z|}{a|1 - e^{|z|}|} \le \frac{1}{a(1 + \frac{|z|}{2!})} \le \frac{1}{a} =: q < 1.$$

Satz 6.25 (BANACHscher Fixpunktsatz)
Ist die stetige Funktion $f : [a, b] \to [a, b]$ **kontrahierend**, *so hat f im Intervall $[a, b]$ genau einen Fixpunkt $x = f(x)$. Dieser ist der Grenzwert der Folge* (7.28), *und es gelten für alle $n \in \mathbf{N}$ die Fehlerabschätzungen (FA):*

$$\boxed{|x - x_n| \le \begin{cases} \dfrac{q^n}{1-q}\,|x_1 - x_0| & : \textbf{a–priori FA}, \\ \dfrac{q}{1-q}\,|x_n - x_{n-1}| & : \textbf{a–posteriori FA}, \end{cases}} \tag{7.30}$$

mit der Kontraktionskonstanten q aus (7.29).

Begründung: Wegen $f([a,b]) \subseteq [a,b]$ gilt wiederum $x_n \in [a,b]\ \forall\, n \in \mathbf{N}$, sofern wir nur mit $x_0 \in [a,b]$ starten. Unter Verwendung der Kontraktionsbedingung (7.29) erhält man ferner:

$$|x_{n+1} - x_n| = |f(x_n) - f(x_{n-1})| \le q|x_n - x_{n-1}| \le q^2|x_{n-1} - x_{n-2}| \le \cdots \le q^n|x_1 - x_0|. \qquad (7.31)$$

Mit Hilfe der Δ–Ungleichung erschließt man hieraus für beliebiges $k \in \mathbf{N}$:

$$\begin{aligned} |x_{n+k} - x_n| &= \left|\sum_{j=1}^{k}(x_{n+j} - x_{n+j-1})\right| \le \sum_{j=1}^{k}|x_{n+j} - x_{n+j-1}| \\ &\overset{(7.31)}{\le} q^n|x_1 - x_0| \sum_{j=1}^{\infty} q^{j-1} = \frac{q^n}{1-q}\,|x_1 - x_0| \quad \text{bzw.} \\ &\overset{(7.31)}{\le} q\,|x_n - x_{n-1}| \sum_{j=1}^{\infty} q^{j-1} = \frac{q}{1-q}\,|x_n - x_{n-1}|. \end{aligned}$$

Wegen $0 \le q < 1$ resultiert aus dieser Abschätzung, dass $(x_n)_{n\ge 0} \subset [a,b]$ eine CAUCHY–Folge ist und als solche gegen einen Grenzwert $x \in [a,b]$ konvergiert. Wegen der Stetigkeit der Funktion f gilt $x = \lim\limits_{n\to\infty} x_{n+1} = \lim\limits_{n\to\infty} f(x_n) = f(x)$ sowie

$$|x - x_n| = |\lim_{k\to\infty} x_{n+k} - x_n| \le \begin{cases} \dfrac{q^n}{1-q}\,|x_1 - x_0|, \\ \dfrac{q}{1-q}\,|x_n - x_{n-1}|, \end{cases}$$

wie oben gezeigt. Wäre x nicht eindeutig bestimmt, das heißt, wäre $y = f(y) \ne x$ ein weiterer Fixpunkt, so ergäbe sich aus (7.29)

$$0 < |x - y| = |f(x) - f(y)| \le q|x - y|, \ \text{ also } \ 0 < (1-q)|x - y| \le 0.$$

Dieser Widerspruch löst sich nur für $x = y$. □

Fixpunktiteration für $f(x) := \frac{x+2}{x+1}$

n	x_n	$f(x_n)$	$\lvert x_n - x\rvert$	$\frac{q^n}{1-q}\lvert x_1 - x_0\rvert$
0	1.000 000 000	1.500 000 000	0.414 213 562	0.666 666 667
1	1.500 000 000	1.400 000 000	0.085 786 438	0.166 666 667
2	1.400 000 000	1.416 666 667	0.014 213 562	0.041 666 667
3	1.416 666 667	1.413 793 103	0.002 453 104	0.010 416 667
4	1.413 793 103	1.414 285 714	0.000 420 459	0.002 604 167
5	1.414 285 714	1.414 201 183	0.000 072 152	0.000 651 042
6	1.414 201 183	1.414 215 686	0.000 012 379	0.000 162 760
7	1.414 215 686	1.414 213 198	0.000 002 124	0.000 040 690
8	1.414 213 198	1.414 213 625	0.000 000 364	0.000 010 173
9	1.414 213 625	1.414 213 552	0.000 000 062	0.000 002 543
10	1.414 213 552	1.414 213 564	0.000 000 011	0.000 000 636
11	1.414 213 564	1.414 213 562	0.000 000 002	0.000 000 159
12	1.414 213 562	1.414 213 562	0.000 000 000	0.000 000 040
13	1.414 213 562	1.414 213 562	0.000 000 000	0.000 000 010

BSP. (6.7.8) Auf dem Intervall $I := [1,2]$ sei die Funktion $f(x) := \frac{x+2}{x+1}$ gegeben. Da f auf I streng monoton fällt, gelten $\max\limits_{x\in I} f(x) = f(1) = \frac{3}{2}$, $\min\limits_{x\in I} f(x) = f(2) = \frac{4}{3}$. Hieraus folgt $f(I) \subset I$. Um die Kontraktionseigenschaft zu zeigen, seien $x, y \in I$ fixiert:

$$|f(x) - f(y)| = \left|\frac{y - x}{(x+1)(y+1)}\right| \le \frac{|x - y|}{2 \cdot 2} = \frac{1}{4}\,|x - y|.$$

Es sind alle Voraussetzungen des Satzes 6.25 erfüllt, und dieser liefert die Existenz genau eines Fixpunktes $x = f(x) \in I$. Es liegt hier der Glücksfall vor, dass x explizit berechnet werden kann. Denn die Fixpunktgleichung $x = f(x)$ lässt sich äquivalent schreiben als $x^2 + x = x + 2$, und hieraus folgt $x = \sqrt{2}$. Die numerische Berechnung des Fixpunktes mit Hilfe der Iterationsfolge (7.28) führt auf die obigen Zahlenwerte, wenn der Startwert $x_0 = 1$ gewählt wird. In der letzten Spalte ist der aus (7.30) für $q := \frac{1}{4}$ resultierende Fehler berechnet worden; in der dritten Spalte steht der wahre Fehler $|x_n - x|$.

Kapitel 7

Differentialrechnung für Funktionen einer reellen Veränderlichen

7.1 Der Ableitungsbegriff

In Abschnitt 6.1 wurde bereits festgestellt, dass in der Euklidischen Ebene der Graph der *affinen Funktion*

$$T(x) := ax + b, \quad x \in D(T) := \mathbf{R}, \quad a, b \in \mathbf{R} \text{ fest}, \tag{1.1}$$

die **Gerade** durch die Punkte $(0, b)$ und $(-\frac{b}{a}, 0)$ ist. Wird ein *beliebiger* Punkt (x_0, y_0) der Euklidischen Ebene fixiert, so verläuft durch diesen Punkt ein ganzes **Geradenbüschel**

$$\frac{T(x) - y_0}{x - x_0} = \tan\alpha, \quad x \neq x_0, \tag{1.2}$$

mit dem Büschelparameter $\alpha \in [0, \pi]$. Natürlich sind (1.1) und (1.2) äquivalente analytische Darstellungen der affinen Funktion; es gelten nämlich $a = \tan\alpha$ und $b = y_0 - x_0 \tan\alpha$. In geometrischer Terminologie heißt a die **Steigung** der durch (1.1) beschriebenen Geraden, und b heißt der **Ordinatenabschnitt** der Geraden. Da die Steigung in jedem Punkt der Geraden dieselbe Konstante ist, resultiert für eine Gerade durch zwei vorgegebene Punkte (x_1, y_1) und (x_2, y_2) die analytische Darstellung

$$\frac{T(x) - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1} \quad (= \tan\alpha).$$

Löst man nach $T(x)$ auf, so erhält man die zu (1.1) äquivalente Form

$$T(x) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1, \quad x \in \mathbf{R}. \tag{1.3}$$

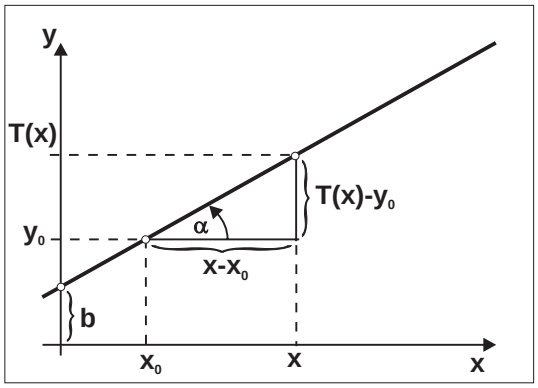
Das Leibnizsche **Tangentenproblem** (Gottfried Wilhelm Leibniz, 1646–1716) besteht in der Bestimmung derjenigen Geraden $T(x)$, die den Graph $G(f)$ einer gegebenen Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R}), D(f) \subset \mathbf{R}$, in einem Punkt $\big(x_0, y_0 = f(x_0)\big)$, $x_0 \in D(f)$, *möglichst gut approximiert*: Es soll in der *Nähe* der Stelle $x_0 \in D(f)$ eine Darstellung der Form

$$f(x) = T(x) + R(x; x_0) \text{ mit } \frac{f(x) - T(x)}{x - x_0} \to 0 \text{ für } 0 < |x - x_0| \to 0 \tag{1.4}$$

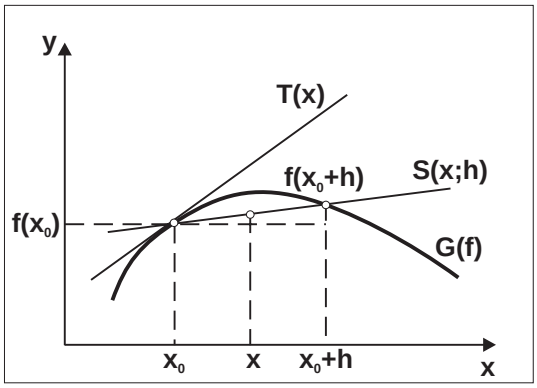
gelten. Da die gesuchte Gerade mindestens den Punkt (x_0, y_0) mit dem Graphen $G(f)$ gemeinsam haben muss, folgern wir aus (1.2) die Darstellung $T(x) = (x - x_0)\tan\alpha + y_0$, und aus (1.4) ergeben sich dann mit $y_0 = f(x_0)$ die zwei zu erfüllenden Gleichungen:

$$\boxed{f(x) - f(x_0) = (x - x_0)\tan\alpha + R(x; x_0), \quad x \in D(f),} \tag{1.5}$$

$$\boxed{\frac{f(x) - f(x_0)}{x - x_0} = \tan\alpha + \frac{R(x; x_0)}{x - x_0}, \quad 0 < |x - x_0| \to 0.} \tag{1.6}$$



Die Gerade als Graph der affinen Funktion



Die Tangente ist der Grenzwert der Sekantenfolge

Aus der Grenzwertbeziehung (1.4) erhält man $\frac{R(x;x_0)}{x-x_0} = \frac{f(x)-T(x)}{x-x_0} \to 0$ für $0 < |x - x_0| \to 0$, und dies impliziert insbesondere $R(x; x_0) \to 0$, falls $0 < |x - x_0| \to 0$. Das heißt, das Leibnizsche Tangentenproblem ist eindeutig lösbar, wenn gilt:

(i) Die Funktion $f(x)$ ist stetig in $x = x_0 \in D(f)$. Dann gilt nämlich (1.5).

(ii) Der **Differenzenquotient**

$$\boxed{\frac{f(x) - f(x_0)}{x - x_0} \equiv \frac{f(x_0 + h) - f(x_0)}{h}}$$

hat für $0 \neq x - x_0 := h \to 0$ einen Grenzwert $f'(x_0) := \tan\alpha \in \mathbf{R}$. Dann gilt (1.6).

Die gesuchte Gerade **bester Approximation** (im Sinne von (1.4)) hat somit die Form

$$\boxed{T(x) = (x - x_0)\lim_{h\to 0}\frac{f(x_0 + h) - f(x_0)}{h} + f(x_0) =: (x - x_0)\cdot f'(x_0) + f(x_0).} \tag{1.7}$$

In der obigen Skizze wird die **geometrische Bedeutung** der Aussage (1.7) veranschaulicht. Der Graph der affinen Funktion $T(x)$ ist diejenige Gerade, die im Limes $h \to 0$ aus der Familie der **Sekanten** $S(x; h)$ hervorgeht. Gemäß (1.3) hat man nämlich mit den Bezeichnungen der obigen Skizze:

$$S(x; h) = (x - x_0)\,\frac{f(x_0 + h) - f(x_0)}{h} + f(x_0), \quad x \in \mathbf{R}. \tag{1.8}$$

Definition 7.1 (a) *Die Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ mit $D(f) \subset \mathbf{R}$ heiße im Punkt $x_0 \in D(f)$* **differenzierbar**, *wenn der Grenzwert*

$$\boxed{\lim_{x\to x_0}\frac{f(x) - f(x_0)}{x - x_0} \quad \textit{oder äquivalent} \quad \lim_{h\to 0}\frac{f(x_0 + h) - f(x_0)}{h}} \tag{1.9}$$

in $\mathbf{R}$ *existiert. Dieser Limes wird mit* $f'(x_0)$ *oder* $\frac{df}{dx}(x_0)$ *bezeichnet, und er heißt die* **Ableitung** *oder der* **Differentialquotient** *von* f *in* x_0.

(b) *Die durch die Ableitung* $f'(x_0)$ *festgelegte affine Funktion*

$$\boxed{T(x) := f'(x_0)\,(x - x_0) + f(x_0), \quad x \in \mathbf{R},}$$

heiße die **Tangente** *im Punkt* $\big(x_0, f(x_0)\big)$ *an den Graph* $G(f)$ *der Funktion* f.

(c) *Ist* x_0 *ein* **Randpunkt** *von* $D(f)$, *so kann der Limes* (1.9) *nur als einseitiger Grenzwert existieren. In diesem Fall heiße*

(i) $$\boxed{\lim_{h\to 0+} \tfrac{1}{h}\,[f(x_0+h) - f(x_0)] =: \frac{d^+f}{dx}(x_0) \quad \textbf{rechtsseitige Ableitung},}$$

(ii) $$\boxed{\lim_{h\to 0-} \tfrac{1}{h}\,[f(x_0+h) - f(x_0)] =: \frac{d^-f}{dx}(x_0) \quad \textbf{linksseitige Ableitung}}$$

von f *in* x_0.

(d) *Die Funktion* f *heiße* **differenzierbar auf** $X \subseteq D(f)$, *wenn* f *in jedem Punkt* $x_0 \in X$ *differenzierbar ist.*

(e) *Die Funktion* $f' : D(f') \to \mathbf{R}$ *mit* $D(f') := \{x_0 \in D(f) : f'(x_0) \in \mathbf{R}$ *existiert*$\}$ *heiße* **Ableitung** *von* f.

Bemerkung 7.1 (a) Im allgemeinen Fall ist wohl zu unterscheiden zwischen den **einseitigen Ableitungen** $\frac{d^\pm f}{dx}(x_0)$ und den **einseitigen Funktionenlimites** $f'(x_0 \pm 0) := \lim\limits_{x\to x_0\pm} f'(x)$ in einem Punkt $x_0 \in D(f)$. Am Beispiel $f(x) := \operatorname{sign} x$, $x \in \mathbf{R}$, ist leicht zu verifizieren, dass im Punkt $x_0 = 0$ einseitige Ableitungen $\frac{d^\pm f}{dx}(0)$ nicht erklärt sind. Die Grenzwerte

$$\lim_{h\to 0\pm} \frac{f(h) - f(0)}{h} = \lim_{h\to 0\pm} \frac{\pm 1}{h}$$

existieren nicht. Hingegen gilt $f'(0 \pm 0) = \lim\limits_{x\to 0\pm} 0 = 0$.

(b) Gottfried Wilhelm Leibniz wird allgemein als der Vater der Differentialrechnung angesehen. Von ihm stammt auch die *inkrementelle* Bezeichnungsweise

$$f'(x_0) \equiv \frac{df}{dx}(x_0) = \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x} \equiv \lim_{\Delta x\to 0} \frac{f(x_0+\Delta x) - f(x_0)}{\Delta x},$$

das heißt, die Darstellung des Differentialquotienten als Grenzwert der Folge der Differenzenquotienten. Diese Bezeichnungsweise bedeutet keineswegs, dass die Grenzwerte $\lim\limits_{\Delta x\to 0} \Delta f = df$ bzw. $\lim\limits_{\Delta x\to 0} \Delta x = dx$ existieren, vielmehr ist sogar

$$\lim_{\Delta x\to 0} \Delta f = 0 = \lim_{\Delta x\to 0} \Delta x.$$

Deshalb sind df und dx nicht als Zahlen im obigen Sinn erklärt sondern nur als Symbole, deren Quotient $\frac{df}{dx} \in \mathbf{R}$ aber wohldefiniert ist.

Neben Leibniz zählt aber auch Isaac Newton (1643–1727) zu den Vätern der Differentialrechnung. Durch das Studium der Mechanik motiviert, führten Newtons Untersuchungen der zeitabhängigen Bewegung von starren Körpern zum Begriff der **Geschwindigkeit** als *Ableitung des Weges* $x(t)$ *nach der Zeit* t: □

$$\frac{x(t) - x(t_0)}{t - t_0} \;\hat{=}\; \text{mittlere Geschwindigkeit im Zeitintervall } [t_0, t],$$

$$\lim_{t\to t_0} \frac{x(t) - x(t_0)}{t - t_0} \;\hat{=}\; \text{Momentangeschwindigkeit zur Zeit } t = t_0.$$

Die Ableitung einer differenzierbaren Funktion ist stets *eindeutig bestimmt*; ferner setzt Differenzierbarkeit *notwendigerweise* Stetigkeit voraus:

Satz 7.1 (a) *Die Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ kann in einem Punkt $x_0 \in D(f)$ höchstens einen Differentialquotienten haben.*

(b) *Ist f im Punkt $x_0 \in D(f)$ differenzierbar, so ist f in x_0 auch stetig.*

Begründungen: (a) Diese Aussage folgt aus der Eindeutigkeit von Grenzwerten, Satz 3.1.

(b) Die Stetigkeit resultiert aus der Relation

$$\lim_{x\to x_0}\,[f(x)-f(x_0)] = \lim_{x\to x_0}\left[\frac{f(x)-f(x_0)}{x-x_0}\,(x-x_0)\right] = f'(x_0)\cdot 0 = 0.$$

Beachte: Die Aussage (b) in Satz 7.1 ist nicht umkehrbar. Differenzierbarkeit ist eine **stärkere** Aussage als Stetigkeit!

Beispiele differenzierbarer Funktionen.

1. Die **Betragsfunktion** $f(x) := |x|$ ist stetig in jedem Punkt $x_0 \in D(f) := \mathbf{R}$. Sie ist auch differenzierbar mit Ausnahme des Punktes $x_0 = 0$, denn dort gilt:

$$\begin{aligned}\frac{d^+f}{dx}(x_0) &= \lim_{h\to 0+}\frac{f(h)-f(0)}{h} = \lim_{h\to 0+}\frac{|h|}{h} = \lim_{h\to 0+}\frac{h}{h} = +1,\\ \frac{d^-f}{dx}(x_0) &= \lim_{h\to 0-}\frac{f(h)-f(0)}{h} = \lim_{h\to 0-}\frac{|h|}{h} = -\lim_{h\to 0-}\frac{|h|}{|h|} = -1.\end{aligned}$$

Wir haben in $x_0 = 0$ verschiedene rechts– und linksseitige Grenzwerte, und deshalb existiert $f'(0)$ *nicht*. Im übrigen gilt:

$$\boxed{f'(x) := \frac{d|x|}{dx} = \begin{cases} +1 & : \ x > 0,\\ -1 & : \ x < 0,\\ \text{nicht ex.} & : \ x = 0,\end{cases} \quad \text{oder} \quad (|x|)' = \operatorname{sign} x \ \ \forall\ x \neq 0.}$$

2. Die **konstante Funktion** $f(x) := c,\ x \in D(f) := \mathbf{R}$, erfüllt

$$\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h} = \lim_{h\to 0}\frac{c-c}{h} = 0 = f'(x_0),$$

mit anderen Worten, es gilt

$$\boxed{f'(x) := (c)' = 0 \ \ \forall\ x \in \mathbf{R}, \quad c = const.}$$

3. Die **Monome** $f(x) := x^n,\ x \in D(f) := \mathbf{R},\ n \in \mathbf{N}$, erfüllen

$$f'(x_0) = \lim_{h\to 0}\frac{(x_0+h)^n - x_0^n}{h} = \lim_{h\to 0}\sum_{k=1}^{n}\binom{n}{k}h^{k-1}x_0^{n-k} = \binom{n}{1}x_0^{n-1} = nx_0^{n-1},$$

das heißt, es gilt

$$\boxed{f'(x) := (x^n)' = nx^{n-1} \ \ \forall\ n \in \mathbf{N}\ \forall\ x \in \mathbf{R}.}$$

4. Für die **negativen Potenzen** $f(x) := x^{-n},\ x \in D(f) := \mathbf{R}\setminus\{0\},\ n \in \mathbf{N}$, folgern wir aus den Vorgaben $x_0 \in D(f)$ und $x_0 + h \in D(f)$:

$$f'(x_0) = \lim_{h\to 0}\frac{1}{h}\left[\frac{1}{(x_0+h)^n} - \frac{1}{x_0^n}\right] = \lim_{h\to 0}\frac{1}{x_0^n(x_0+h)^n}\left[\frac{x_0^n-(x_0+h)^n}{h}\right] \overset{\text{(s. BSP.3)}}{=} -n\frac{x_0^{n-1}}{x_0^{2n}} = -\frac{n}{x_0^{n+1}}.$$

Dies bedeutet

$$\boxed{f'(x) := \left(\frac{1}{x^n}\right)' = -\frac{n}{x^{n+1}} \quad \forall\, n \in \mathbf{N}\ \forall\, x \in \mathbf{R} \setminus \{0\}.}$$

5. Die **Exponentialfunktion** $f(x) := e^x,\ x \in D(f) := \mathbf{R}$, erfüllt

$$f'(x_0) = \lim_{h\to 0} \frac{e^{x_0+h} - e^{x_0}}{h} = e^{x_0} \lim_{h\to 0} \frac{e^h - 1}{h} = e^{x_0} \lim_{h\to 0} \frac{1}{h} \sum_{k=1}^{\infty} \frac{h^k}{k!} = e^{x_0}.$$

Das heißt, es gilt

$$\boxed{f'(x) := (e^x)' = e^x \quad \forall\, x \in \mathbf{R}.}$$

6. Für den **Sinus** $f(x) := \sin x,\ x \in D(f) := \mathbf{R}$, folgt aus den Additionstheoremen

$$\begin{aligned} f'(x_0) &= \lim_{h\to 0} \frac{\sin(x_0+h) - \sin x_0}{h} = \lim_{h\to 0} \frac{\sin x_0\,[\cos h - 1] + \cos x_0\ \sin h}{h} \\ &= \sin x_0 \lim_{h\to 0} \frac{\cos h - 1}{h} + \cos x_0 \lim_{h\to 0} \frac{\sin h}{h} = \cos x_0. \end{aligned}$$

Das heißt, es gilt

$$\boxed{f'(x) := (\sin x)' = \cos x \quad \forall\, x \in \mathbf{R}.}$$

7. Für den **Cosinus** $f(x) := \cos x,\ x \in D(f) := \mathbf{R}$, folgt aus den Additionstheoremen

$$\begin{aligned} f'(x_0) &= \lim_{h\to 0} \frac{\cos(x_0+h) - \cos x_0}{h} = \lim_{h\to 0} \frac{\cos x_0\,[\cos h - 1] - \sin x_0\ \sin h}{h} \\ &= \cos x_0 \lim_{h\to 0} \frac{\cos h - 1}{h} - \sin x_0 \lim_{h\to 0} \frac{\sin h}{h} = -\sin x_0. \end{aligned}$$

Das heißt, es gilt

$$\boxed{f'(x) := (\cos x)' = -\sin x \quad \forall\, x \in \mathbf{R}.}$$

7.2 Ableitungsregeln

Wir erkennen an den vorangegangenen Beispielen, dass die Berechnung der Ableitung einer Funktion in einem gegebenen Punkt $x_0 \in D(f)$ unter Verwendung der Definition 7.1 ein recht mühsamer Prozess ist. Nur in den einfachsten Fällen kann dieser Prozess elementar abgewickelt werden. Wir werden deshalb in diesem Abschnitt eine Reihe von *Differentiationsregeln* bereitstellen, mit deren Hilfe in der Praxis die komplizierte Grenzwertbestimmung vereinfacht wird.

Satz 7.2 (Summen–, Produkt–, Quotientenregel)
Die Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ seien im Punkt $x_0 \in D(f) \cap D(g) \subset \mathbf{R}$ differenzierbar. Dann sind die Funktionen $f \pm g,\ f \cdot g$, und im Falle $g(x_0) \neq 0$ auch f/g, im Punkt x_0 differenzierbar, und es gelten die folgenden Regeln:

$$\boxed{\begin{aligned} &(f \pm g)'(x_0) = f'(x_0) \pm g'(x_0), && \textbf{Summenregel} \\ &(f \cdot g)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0), && \textbf{Produktregel} \\ &\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}. && \textbf{Quotientenregel} \end{aligned}}$$

Begründungen: Man führt diese Regeln auf die Definition der Ableitung zurück und verwendet dabei die Regeln für das Rechnen mit Grenzwerten. Das Schema ist in jedem der drei Fälle das gleiche; wir beschränken uns deshalb auf den Nachweis der Quotientenregel:

$$\begin{aligned}
&\lim_{h\to 0}\frac{1}{h}\left[\frac{f}{g}(x_0+h)-\frac{f}{g}(x_0)\right]\\
&=\lim_{h\to 0}\frac{1}{g(x_0)g(x_0+h)}\left[g(x_0)\,\frac{f(x_0+h)-f(x_0)}{h}-f(x_0)\,\frac{g(x_0+h)-g(x_0)}{h}\right]\\
&=\frac{1}{g^2(x_0)}\left[f'(x_0)g(x_0)-f(x_0)g'(x_0)\right].
\end{aligned}$$

Hierbei haben wir die Stetigkeit der Funktion g im Punkt x_0 verwendet, die ja wegen Satz 7.1 gewährleistet ist. □

BSP. (7.2.1) Da die Konstante $\lambda\in\mathbf{R}$ die Ableitung 0 hat, erhält man als Sonderfall der Produktregel:

> Ist die Funktion $f: D(f)\to\mathbf{R}$ im Punkt $x_0\in D(f)$ differenzierbar, so gilt dies auch für $\lambda f,\ \lambda\in\mathbf{R}$, und es folgt
> $$(\lambda f)'(x_0)=\lambda f'(x_0)\ \ \forall\,\lambda\in\mathbf{R}.$$

Aus dieser Regel ergibt sich in Verbindung mit der Summenregel:

> Jedes Polynom $P_n(x):=\sum\limits_{k=0}^{n}a_k x^k,\ a_k\in\mathbf{R}, a_n\neq 0$, ist in $\mathbf{R}$ differenzierbar:
> $$(P_n)'(x)=\sum_{k=0}^{n}a_k\,kx^{k-1}=\sum_{k=1}^{n}a_k\,kx^{k-1}\ \ \forall\,x\in\mathbf{R}.$$

Unter Verwendung der Quotientenregel erhält man weiterhin:

> Jede rationale Funktion $R(x):=\frac{P_n(x)}{Q_m(x)},\ P_n, Q_m\in\mathbf{R}[x]$ Polynome, ist auf der Menge $D(R):=\{x\in\mathbf{R}: Q_m(x)\neq 0\}$ differenzierbar:
> $$R'(x)=\frac{P_n'(x)Q_m(x)-P_n(x)Q_m'(x)}{Q_m^2(x)}\ \ \forall\,x\in D(R).$$

Zahlenbeispiel:

$$\begin{aligned}
\left(\frac{x^5-3x^2+5x-2}{(x-2)^2(x+1)}\right)' &= \frac{(5x^4-6x+5)(x-2)^2(x+1)-(x^5-3x^2+5x-2)(x-2)3x}{(x-2)^4(x+1)^2}\\
&= \frac{2x^6-5x^5-10x^4+3x^3-4x^2+13x-10}{(x-2)^3(x+1)^2}\ \ \forall\,x\in\mathbf{R}\setminus\{-1,2\}.
\end{aligned}$$

BSP. (7.2.2) Die Ableitung der beiden trigonometrischen Funktionen $f(x):=\tan x,\ x\in D(\tan):=\{x\in\mathbf{R}: x\neq(n+\frac{1}{2})\pi,\ n\in\mathbf{Z}\}$ sowie $f(x):=\cot x,\ x\in D(\cot):=\{x\in\mathbf{R}: x\neq n\pi,\ n\in\mathbf{Z}\}$, berechnet man unter Verwendung der Quotientenregel:

$$(\tan x)'=\left(\frac{\sin x}{\cos x}\right)'=\frac{\cos^2 x+\sin^2 x}{\cos^2 x}=\frac{1}{\cos^2 x},\quad (\cot x)'=\left(\frac{\cos x}{\sin x}\right)'=\frac{-\sin^2 x-\cos^2 x}{\sin^2 x}=\frac{-1}{\sin^2 x}.$$

Wir haben also:

$$\boxed{(\tan x)'=\frac{1}{\cos^2 x}\ \ \forall\,x\neq(n+\frac{1}{2})\pi,\quad (\cot x)'=\frac{-1}{\sin^2 x}\ \ \forall\,x\neq n\pi,\ n\in\mathbf{Z}.}$$

BSP. (7.2.3) Die Ableitung der abklingenden Exponentialfunktion $f(x) := e^{-x} = \frac{1}{e^x},\ x \in D(f) := \mathbf{R}$, berechnet man ebenfalls nach der Quotientenregel:

$$\boxed{(e^{-x})' = \left(\frac{1}{e^x}\right)' = \frac{-e^x}{e^{2x}} = -e^{-x} \ \ \forall\, x \in \mathbf{R}.}$$

In Verbindung mit der Summenregel ergibt sich hieraus $2(\sinh x)' = (e^x - e^{-x})' = e^x + e^{-x} = 2\cosh x$. Ganz analog zeigt man die folgenden Relationen:

$$\boxed{\begin{aligned} &(\sinh x)' = \cosh x, \qquad (\cosh x)' = \sinh x \ \ \forall\, x \in \mathbf{R}, \\ &(\tanh x)' = \left(\frac{\sinh x}{\cosh x}\right)' = \frac{\cosh^2 x - \sinh^2 x}{\cosh^2 x} = \frac{1}{\cosh^2 x} \ \ \forall\, x \in \mathbf{R}, \\ &(\coth x)' = \left(\frac{\cosh x}{\sinh x}\right)' = \frac{\sinh^2 x - \cosh^2 x}{\sinh^2 x} = \frac{-1}{\sinh^2 x} \ \ \forall\, x \in \mathbf{R} \setminus \{0\}. \end{aligned}}$$

Die folgende *Kettenregel* ist eine der wichtigsten Differentiationsregeln:

Satz 7.3 (Kettenregel)
Für gegebene Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ sei die Hintereinanderausführung $g \circ f$ zumindest in einem offenen Intervall $X \subseteq D(f) \subset \mathbf{R}$ erklärt. Sind die Funktionen f im Punkt $x_0 \in X$ und g im Punkt $f(x_0)$ differenzierbar, so ist auch $g \circ f$ in x_0 differenzierbar, und es gilt

$$\boxed{(g \circ f)'(x_0) = g'[f(x_0)] \cdot f'(x_0). \qquad \textbf{Kettenregel}}$$

Begründung: Es sei $h \neq 0$ so bestimmt, dass $x_0 + h \in X$ gilt. Wir setzen $y_0 := f(x_0)$, ferner $y_0 + k := f(x_0 + h)$, wodurch eine Zahl $k = k(h)$ eindeutig festgelegt ist. Aus der Stetigkeit von f im Punkt x_0 erschließen wir $\lim\limits_{h\to 0} k(h) = \lim\limits_{h\to 0}[f(x_0 + h) - f(x_0)] = 0$, und somit

$$\lim_{h\to 0} \frac{(g \circ f)(x_0 + h) - (g \circ f)(x_0)}{h} = \lim_{h\to 0} \frac{g(y_0 + k) - g(y_0)}{k} \cdot \frac{k}{h}$$

$$= \lim_{k\to 0} \frac{g(y_0 + k) - g(y_0)}{k} \cdot \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} = g'(y_0) \cdot f'(x_0).$$

Bemerkung 7.2 (a) Setzt man $h(x) := (g \circ f)(x) = g[f(x)]$, so kann die Kettenregel in der folgenden einprägsamen Form geschrieben werden:

$$\boxed{\frac{dh}{dx} = \frac{dg}{dy} \cdot \frac{dy}{dx} \ \text{ mit } \ y := f(x).}$$

(b) Ist $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ speziell die *affine Funktion* $f(x) := \lambda\, x + \mu,\ \lambda, \mu \in \mathbf{R}$ fest, so hat die Kettenregel die Form

$$\boxed{\frac{dg}{dx}(\lambda\, x + \mu) = \lambda \frac{dg}{dy}(y) \ \text{ mit } \ y = \lambda\, x + \mu.}$$

Zum Beispiel: $(e^{ax+b})' = a\, e^{ax+b}$, $\quad (\sin 5x)' = 5\cos 5x$, usw. □

BSP. (7.2.4) Die folgenden Ableitungen erhält man durch Anwendung der Kettenregel:

(i) $\quad (a^x)' = (e^{x \ln a})' = \ln a\, e^{x \ln a} = a^x \ln a \ \ \forall\, x \in \mathbf{R},\ \ a > 0,$

(ii) $\quad (e^{\tan x})' = e^{\tan x} \dfrac{1}{\cos^2 x} \ \ \forall\, x \neq (n + \dfrac{1}{2})\pi,\ n \in \mathbf{Z},$

(iii) $(\sinh x^5)' = (\cosh x^5)\cdot 5x^4 \ \ \forall\, x \in \mathbf{R}$,

(iv) $(\cosh\cos x)' = (\sinh\cos x)\cdot(-\sin x) \ \ \forall\, x \in \mathbf{R}$,

(v) $\left[\sin\left(\cos e^{\cot x}\right)\right]' = \cos(\cos e^{\cot x})\cdot(-\sin e^{\cot x})\cdot\dfrac{-e^{\cot x}}{\sin^2 x} \ \ \forall\, x \neq n\pi,\ n \in \mathbf{Z}$.

Bemerkung 7.3 Wir haben hier u.a. die Tatsache benutzt, dass die Sätze 7.2 und 7.3 auch für *mehrfache Verknüpfungen* gelten, zum Beispiel in der Form □

$$\boxed{\begin{array}{rcl} (f\cdot g\cdot h)' & = & f'\cdot g\cdot h + f\cdot g'\cdot h + f\cdot g\cdot h', \\ (h\circ g\circ f)'(x) & = & \{h[g(f(x))]\}' = h'[g(f(x))]\cdot g'(f(x))\cdot f'(x), \quad \text{usw.} \end{array}}$$

Um beispielsweise die Ableitung der Funktion $f(x) := \ln x$ berechnen zu können, zeigen wir eine allgemeine Regel, nach der *Umkehrfunktionen* differenziert werden.

Satz 7.4 (Ableitung der Umkehrfunktion)
Die reelle Funktion $y = f(x)$ sei auf einem Intervall $X \subset \mathbf{R}$ stetig und streng monoton, so dass die Umkehrfunktion $f^{-1} : f(X) \to \mathbf{R}$ existiert. Ist f im Punkt $x_0 \in X$ differenzierbar mit $f'(x_0) \neq 0$, so ist auch f^{-1} im Punkt $y_0 := f(x_0)$ differenzierbar, und es gilt

$$\boxed{\left(f^{-1}\right)'(y_0) = \frac{1}{f'(x_0)} = \frac{1}{f'[f^{-1}(y_0)]}.}$$

Begründung: Für eine beliebige Nullfolge $0 \neq \epsilon_n \in \mathbf{R}$ mit $y_0+\epsilon_n \in f(X)$ setzen wir $x_n := f^{-1}(y_0+\epsilon_n)$. Da die Umkehrfunktion f^{-1} stetig ist, folgern wir $\lim\limits_{n\to\infty} x_n = f^{-1}(y_0) = x_0$. Hieraus erschließen wir:

$$\begin{array}{rcl} (f^{-1})'(y_0) & = & \lim\limits_{n\to\infty} \dfrac{f^{-1}(y_0+\epsilon_n) - f^{-1}(y_0)}{\epsilon_n} = \lim\limits_{n\to\infty} \dfrac{x_n - x_0}{f(x_n) - f(x_0)} \\ & = & \lim\limits_{n\to\infty} \left(\dfrac{f(x_n) - f(x_0)}{x_n - x_0}\right)^{-1} = \dfrac{1}{f'(x_0)}. \end{array}$$

BSP. (7.2.5) Die Ableitung des **Logarithmus**. Wie in Abschnitt 6.7 gezeigt wurde, ist der Logarithmus $\ln y,\ y > 0$, die Umkehrfunktion von $y = f(x) := e^x$. Mit $x = \ln y$ folgt also aus Satz 7.4 die Ableitung

$$(\ln y)' = \frac{1}{(e^x)'} = \frac{1}{e^x} = \frac{1}{e^{\ln y}} = \frac{1}{y} \ \ \forall y > 0.$$

Wir setzen nun anstelle der Variablen y wieder die Variable x ein. In Verbindung mit der Kettenregel erhält man folgende Ableitungen:

$$\boxed{\begin{array}{rcl} (\ln x)' & = & \dfrac{1}{x} \ \ \forall\, x > 0, \\ (x^p)' & = & (e^{p\,\ln x})' = \dfrac{p}{x}\, e^{p\,\ln x} = p\, x^{p-1} \ \ \forall\, x > 0\, \forall\, p \in \mathbf{R}, \\ (\sqrt{x})' & = & (x^{1/2})' = \dfrac{1}{2}\, x^{-1/2} = \dfrac{1}{2\sqrt{x}} \ \ \forall\, x > 0, \\ ({}^a\log x)' & = & \left(\dfrac{\ln x}{\ln a}\right)' = \dfrac{1}{x\,\ln a} \ \ \left(= \dfrac{1}{x}\,{}^a\log e\right) \ \ \forall\, x > 0\, \forall\, 0 < a \neq 1, \\ (x^x)' & = & (e^{x\,\ln x})' = (\ln x + \frac{x}{x})\, e^{x\,\ln x} = (1+\ln x)\, x^x \ \ \forall\, x > 0. \end{array}}$$

BSP. (7.2.6) Die Ableitung der **zyklometrischen Funktionen**. In Abschnitt 6.7 wurde auch gezeigt, dass die Funktion $\arcsin_H : [-1,+1] \to [-\frac{\pi}{2},+\frac{\pi}{2}]$ die Umkehrfunktion von $\sin : [-\frac{\pi}{2},+\frac{\pi}{2}] \to [-1,+1]$ ist. Es gilt $(\sin x)' = \cos x \neq 0$ nur für $x \in (-\frac{\pi}{2},+\frac{\pi}{2})$. Auf diesem Intervall erhalten wir aus Satz 7.4:

$$\Big(\arcsin_H y\Big)' = \frac{1}{\cos x} = \frac{1}{+\sqrt{1-\sin^2 x}} = \frac{1}{\sqrt{1-y^2}} \quad \forall\, y \in (-1,+1).$$

Wir setzen jetzt wieder x an die Stelle der Variablen y. In ähnlicher Weise werden die Ableitungen der anderen zyklometrischen Funktionen berechnet. Wir fassen zusammen

$$\boxed{\begin{array}{rcl}
\Big(\arcsin_H x\Big)' & = & \dfrac{1}{\sqrt{1-x^2}} \quad \forall\, x \in (-1,+1), \\
\Big(\arccos_H x\Big)' & = & \dfrac{-1}{\sqrt{1-x^2}} \quad \forall\, x \in (-1,+1), \\
\Big(\arctan_H x\Big)' & = & \dfrac{1}{1+x^2} \quad \forall\, x \in \mathbf{R}, \\
\Big(\operatorname{arc\,cot}_H x\Big)' & = & \dfrac{-1}{1+x^2} \quad \forall\, x \in \mathbf{R}.
\end{array}}$$

Die anderen Zweige der zyklometrischen Funktionen unterscheiden sich von den Hauptwerten jeweils um eine additive Konstante und eventuell um das Vorzeichen; man orientiere sich in Abschnitt 6.7. Deshalb erhält man die folgenden Ableitungsformeln:

$$\boxed{\begin{array}{rcl}
\Big(\arcsin_n x\Big)' = -\Big(\arccos_n x\Big)' & = & \dfrac{(-1)^n}{\sqrt{1-x^2}} \quad \forall\, x \in (-1,+1)\ \forall\, n \in \mathbf{Z}, \\
\Big(\arctan_n x\Big)' = -\Big(\operatorname{arc\,cot}_n x\Big)' & = & \dfrac{1}{1+x^2} \quad \forall\, x \in \mathbf{R}\ \forall\, n \in \mathbf{Z}.
\end{array}}$$

Weitere Ableitungsformeln, insbesondere für die Area–Funktionen, findet man in den gängigen Formelsammlungen aufgelistet. (Man konsultiere zum Beispiel die Formelsammlung von BRONSTEIN–SEMENDJAJEW o.ä.)

Eine Variante der Kettenregel ist die folgende Regel des *logarithmischen Differenzierens:*

Satz 7.5 (Logarithmisches Differenzieren)
Die Funktion $f \in \mathrm{Abb}(\mathbf{R},\mathbf{R})$ sei im offenen Intervall $X \subseteq D(f) \subset \mathbf{R}$ differenzierbar, und es gelte $f(x) \neq 0\ \forall\, x \in X$. Dann ist auch die Funktion $g(x) := \ln|f(x)|$ in X differenzierbar, und es gilt $g'(x) = f'(x)/f(x)\ \forall\, x \in X$. Hieraus erschließt man die Regel des logarithmischen Differenzierens

$$\boxed{f'(x) = f(x) \cdot (\ln|f(x)|)' \quad \forall\, x \in X.}$$

Begründung: Es gilt entweder $f > 0$ oder $f < 0$ überall in X. In beiden Fällen resultiert jeweils unter Verwendung der Kettenregel:

(i) $f > 0 \Rightarrow g(x) = \ln f(x)$, und somit $g'(x) = f'(x)/f(x)$,

(ii) $f < 0 \Rightarrow g(x) = \ln[-f(x)]$, und somit $g'(x) = -f'(x)/[-f(x)]$. □

BSP. (7.2.7) Auf dem offenen Intervall $X \subseteq \mathbf{R}$ seien die differenzierbaren Funktionen $f, g : X \to \mathbf{R}$ mit $f > 0$ gegeben. Für die Ableitung der Funktion $h(x) := f(x)^{g(x)} = e^{g(x)\,\ln f(x)}$ gilt gemäß Satz 7.5:

$$\frac{h'(x)}{h(x)} = \Big(\ln|h(x)|\Big)' = \Big(g(x)\ln f(x)\Big)' = g'(x)\,\ln f(x) + \frac{g(x)f'(x)}{f(x)}.$$

Durch Auflösen nach $h'(x)$ erhält man also:

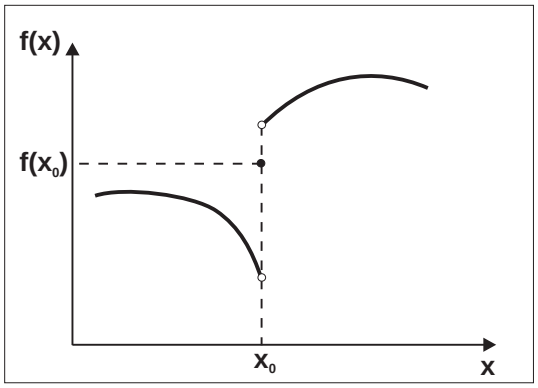
$$\boxed{\Big(f(x)^{g(x)}\Big)' = f(x)^{g(x)} \cdot [g(x)\,\ln f(x)]' = f(x)^{g(x)} \cdot \left[g'(x)\,\ln f(x) + \frac{g(x)f'(x)}{f(x)}\right].}$$

Zum Beispiel:

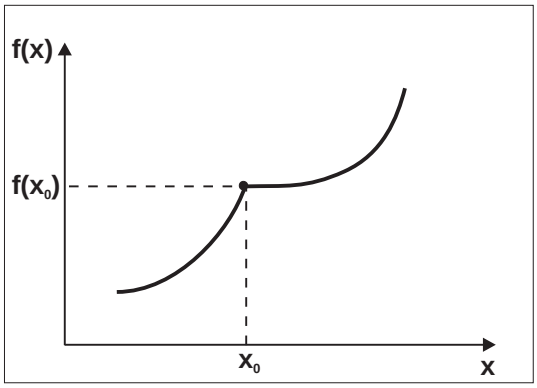
$$\left(x^{\ln x}\right)' = x^{\ln x}\,(\ln^2 x)' = x^{\ln x}\,\frac{2\,\ln x}{x} = 2\,\ln x \cdot x^{\ln x - 1}, \quad x > 0.$$

Beispiele **nicht differenzierbarer Funktionen**.

1. Ist die Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ **unstetig** im Punkt $x_0 \in D(f)$, so kann f in diesem Punkt *keine Ableitung* haben. **Warnung:** Selbst im Falle, dass die Funktion f' in x_0 gleiche rechts– und linksseitige Funktionenlimites $f'(x_0 + 0) = f'(x_0 - 0)$ besitzt, ist eine Ableitung $f'(x_0)$ im Unstetigkeitspunkt x_0 **nicht** erklärt. Dieser Fall liegt zum Beispiel bei der Funktion $f(x) := \mathrm{sign}\,x$ im Punkt $x_0 = 0$ vor. Wir haben hier zwar $f'(x_0 + 0) = 0 = f'(x_0 - 0)$, aber die einseitigen Ableitungen $\frac{d^{\pm} f}{dx}(0)$ existiert nicht.

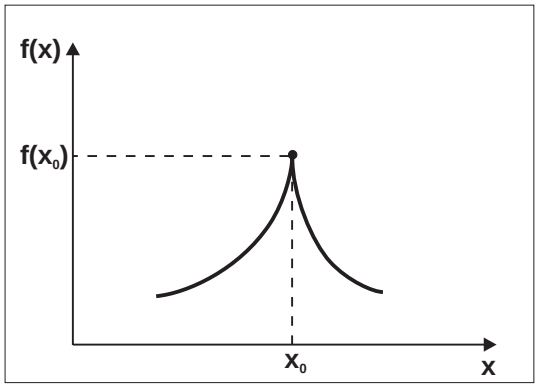


In einem Unstetigkeitspunkt x_0 existiert $f'(x_0)$ nicht



In einem Knickpunkt x_0 existiert $f'(x_0)$ nicht

2. Hat die Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ im Punkt $x_0 \in D(f)$ einen **Knick**, so ist f zwar stetig in x_0, aber es existieren *verschiedene rechts– und linksseitige Ableitungen* $\frac{d^+ f}{dx}(x_0) \neq \frac{d^- f}{dx}(x_0)$. Die Funktion f ist nicht differenzierbar in x_0. Dieser Fall liegt zum Beispiel bei der Funktion $f(x) := |x|$ im Punkt $x_0 = 0$ vor. Wir haben hier $f'(x_0 + 0) = \frac{d^+ f}{dx}(x_0) = +1 \neq -1 = \frac{d^- f}{dx}(x_0) = f'(x_0 - 0)$.



In einer Spitze x_0 existiert $f'(x_0)$ nicht

3. Hat die Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ im Punkt $x_0 \in D(f)$ eine **Spitze**, so ist f zwar stetig in x_0, es existieren aber *verschiedene uneigentliche rechts– und linksseitige Ableitungen* $\pm\infty = \frac{d^+ f}{dx}(x_0) \neq \frac{d^- f}{dx}(x_0) = \mp\infty$. Die Funktion f ist nicht differenzierbar in x_0. Dieser Fall liegt zum Beispiel bei der Funktion $f(x) := \sqrt{|x|}$ im Punkt $x_0 = 0$ vor. Wir haben hier

$$\frac{d^+ f}{dx}(0) = \lim_{h \to 0+} \frac{\sqrt{h}}{h} = +\infty, \quad \frac{d^- f}{dx}(0) = \lim_{h \to 0-} \frac{\sqrt{|h|}}{-|h|} = -\infty.$$

4. Wir betrachten die Funktionen $f_n(x) := x^n \sin\frac{1}{x},\ n := 0, 1, 2.$

(i) Die Funktion $f_0(x) = \sin\frac{1}{x}$ ist im Punkt $x_0 = 0$ **unstetig**, da der Grenzwert $\lim\limits_{x \to 0} f_0(x)$ nicht existiert. Mithin ist f_0 in $x_0 = 0$ **nicht differenzierbar**.

(ii) Die Funktion $f_1(x) = x \sin\frac{1}{x}$ ist im Punkt $x_0 = 0$ durch $f_1(0) := 0$ **stetig ergänzbar**, denn es gilt $\lim\limits_{x \to 0} x \sin\frac{1}{x} = 0$. Der Differentialquotient

$$f_1'(0) = \lim_{h\to 0} \frac{f_1(h) - f_1(0)}{h} = \lim_{h\to 0} \sin\frac{1}{h}$$

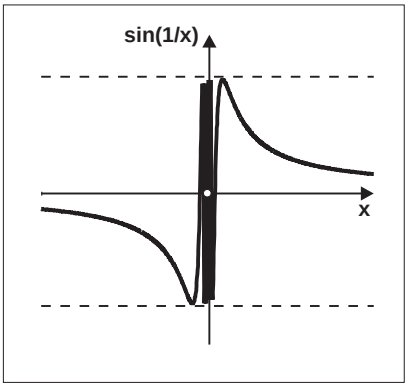
existiert jedoch nicht, so dass f_1 in $x_0 = 0$ **nicht differenzierbar** ist.

(iii) Die Funktion $f_2(x) = x^2 \sin\frac{1}{x}$ ist im Punkt $x_0 = 0$ durch $f_2(0) := 0$ stetig ergänzbar, denn es gilt wiederum $\lim\limits_{x\to 0} x^2 \sin\frac{1}{x} = 0$. Der Differentialquotient

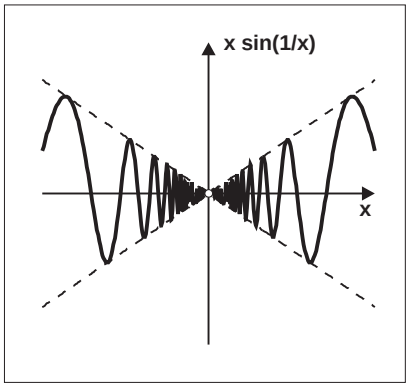
$$f_2'(0) = \lim_{h\to 0} \frac{f_2(h) - f_2(0)}{h} = \lim_{h\to 0} h \sin\frac{1}{h} = 0$$

existiert, so dass f_2 in $x_0 = 0$ **differenzierbar** ist. Hingegen existieren *keine Funktionenlimites* $f_2'(x_0 \pm 0)$. Für $x \neq 0$ berechnet man nämlich $f_2'(x) = 2x \sin\frac{1}{x} - \cos\frac{1}{x}$, also $f_2'(x_0 \pm 0) = -\lim\limits_{x\to 0\pm} \cos\frac{1}{x}$, und diese Grenzwerte existieren nicht. Deshalb ist f_2' **unstetig** bei $x_0 = 0$, und wir haben

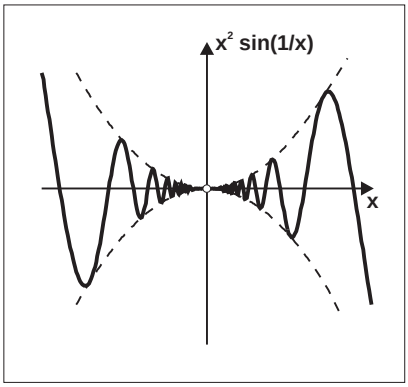
$$f_2'(x) = \begin{cases} 2x \sin\frac{1}{x} - \cos\frac{1}{x} & : \ x \neq 0, \\ 0 & : \ x = 0. \end{cases}$$



$\sin\frac{1}{x}$ **ist unstetig bei** $x_0 = 0$



$x \sin\frac{1}{x}$ **ist stetig bei** $x_0 = 0$, **aber nicht differenzierbar**



$x^2 \sin\frac{1}{x}$ **ist differenzierbar bei** $x_0 = 0$, **aber die Ableitung ist unstetig**

7.3 Ergänzungen und Erweiterungen

(I) Komplexwertige Funktionen. Die *analytische* Definition der Differenzierbarkeit lässt sich auch auf *komplexwertige* Funktionen $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{C})$ mit $D(f) \subset \mathbf{R}$ übertragen. In diesem Sinne bleibt die Definition 7.1 vollgültig, wenn man dort die Zielmenge $\mathbf{R}$ überall durch die Zielmenge $\mathbf{C}$ ersetzt. Wegen der eineindeutigen Identifikation $\mathbf{C} \rightleftharpoons \mathbf{R}^2$ durch die GAUSSsche Zahlenebene ist eine komplexwertige Funktion $f : D(f) \to \mathbf{C}$ geometrisch als Parameterdarstellung einer ebenen Kurve zu deuten. Die *Ableitung* $f'(x_0)$ im Punkt $x_0 \in D(f)$ ist dann der *Vektor der Tangentenrichtung* an die Kurve $f(x)$ in $x = x_0$.

Wird die komplexwertige Funktion f in Real– und Imaginärteil zerlegt, nämlich

$$f(x) = u(x) + i\,v(x), \quad u, v : D(f) \to \mathbf{R},$$

so ist Differenzierbarkeit von f in einem Punkt $x_0 \in D(f)$ – analog zum Fall der Stetigkeit – äquivalent mit der Differenzierbarkeit von u und v in x_0.

Eine komplexwertige Funktion $f := u + i\,v : D(f) \to \mathbf{C}$ mit $D(f) \subset \mathbf{R}$ ist im Punkt $x_0 \in D(f)$ genau dann differenzierbar, wenn beide Funktionen u und v in x_0 differenzierbar sind. Die Ableitung $f'(x_0)$ ist gegeben durch

$$\boxed{f'(x_0) = u'(x_0) + i\,v'(x_0).}$$

BSP. (7.3.1) Für die Funktion $f(x) := \frac{e^{ix}}{1+\cos x} = \frac{\cos x}{1+\cos x} + i\,\frac{\sin x}{1+\cos x} =: u(x) + i\,v(x),\ x \in D(f) := (-\pi, +\pi)$, haben wir

$$u'(x) = \left(\frac{\cos x}{1+\cos x}\right)' = \frac{-\sin x}{(1+\cos x)^2},\quad v'(x) = \left(\frac{\sin x}{1+\cos x}\right)' = \frac{(1+\cos x)\cos x + \sin^2 x}{(1+\cos x)^2} = \frac{1}{1+\cos x}.$$

Also resultiert $f'(x) = \frac{-\sin x}{(1+\cos x)^2} + i\,\frac{1}{1+\cos x}\ \ \forall\ x \in D(f)$.

Bemerkung 7.4 (a) Die *Rechenregeln* aus Satz 7.2 bleiben auch für komplexwertige Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{C})$ richtig. Das heißt, Summen–, Produkt– und Quotientenregel gelten auf der Menge $D(f) \cap D(g)$ in unveränderter Form:

$$\boxed{(f \pm g)' = f' \pm g',\quad (f \cdot g)' = f'g + fg',\quad \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}\ \text{ für }\ g(x) \neq 0.}$$

(b) Die Kettenregel kann für Funktionen $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ und $g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{C})$ formuliert werden, sofern das Kompositum $g \circ f$ auf einer Teilmenge $X \subseteq D(f)$ erklärt ist. Wir haben wie vorher, sofern alle Ableitungen existieren:

$$\boxed{[g\,(f(x))]' = g'\,(f(x)) \cdot f'(x),\quad x \in X.}$$

(c) Der Satz 7.4 von der Ableitung der Umkehrfunktion kann natürlich nicht ins Komplexe übertragen werden, da der Monotoniebegiff nur für reellwertige Funktionen erklärt wurde. □

In Erweiterung der bisherigen Ableitungsregeln gilt bei komplexen Funktionen:

Satz 7.6 *Es sei $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{C})$ differenzierbar für alle $x \in D(f) \subset \mathbf{R}$. Dann gilt:*

$$\boxed{\overline{\big(f'(x)\big)} = (\bar f)'(x)\ \ \forall\ x \in D(f),\quad \left(e^{\lambda x}\right)' = \lambda e^{\lambda x}\ \ \forall\ x \in \mathbf{R}\ \forall\ \lambda \in \mathbf{C}.}$$

Begründungen: (a) Aus der Zerlegung $f(x) = u(x) + i\,v(x)$ folgt unmittelbar die behauptete Relation

$$\overline{\big(f'(x)\big)} = \overline{\big(u'(x) + i\,v'(x)\big)} = u'(x) - i\,v'(x) = (u - i\,v)'(x) = (\bar f)'(x).$$

(b) Wir verwenden die Definition der Ableitung:

$$\left(e^{\lambda x}\right)' = \lim_{h\to 0} \frac{e^{\lambda(x+h)} - e^{\lambda x}}{h} = e^{\lambda x} \lim_{h\to 0} \frac{e^{\lambda h} - 1}{h} = \lambda e^{\lambda x} \lim_{h\to 0} \left(1 + \sum_{k=2}^{\infty} \frac{h^{k-1}\lambda^{k-1}}{k!}\right).$$

Nehmen wir bereits $|h| < 1$ an, so gilt

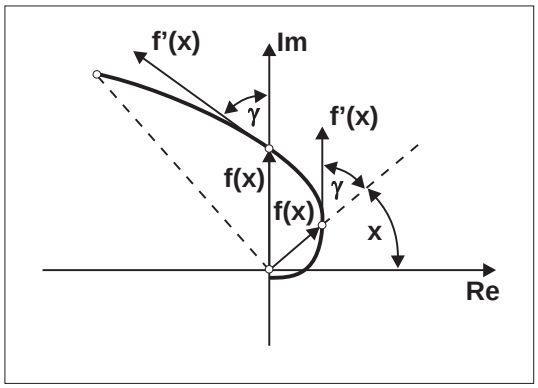
$$\Big|\sum_{k=2}^{\infty} \frac{h^{k-1}\lambda^{k-1}}{k!}\Big| \le |h| \sum_{k=2}^{\infty} \frac{|\lambda|^{k-1}}{k!} \to 0,\ h \to 0,$$

und hieraus folgt schon die behauptete Ableitungsregel. □

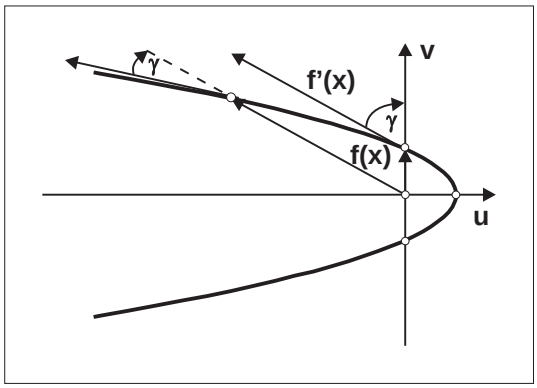
BSP. (7.3.2) Wir betrachten die komplexwertige Funktion $f(x) := e^{(1+i)x},\ x \in D(f) := \mathbf{R}$. Gemäß Satz 7.6 gilt $f'(x) = (1+i)\,f(x)\ \forall\ x \in \mathbf{R}$, und wegen $1 + i = \sqrt{2}\,e^{i\pi/4}$ kann dafür auch $f'(x) = \sqrt{2}\,e^{i\pi/4} f(x)$ geschrieben werden. Zwischen dem Ortsvektor $f(x)$ und dem Tangentenvektor $f'(x)$ besteht somit der Zusammenhang

$$|f'(x)| = \sqrt{2}|f(x)|,\quad \gamma := \sphericalangle\,(f', f) = \frac{\pi}{4} = const.$$

Die Funktion f beschreibt in $\mathbf{C}$ eine **logarithmische Spirale**.



l**Die logarithmische Spirale** $f(x) := e^{(1+i)x}$



Die Parabel $f(x) := e^{ix}/(1+\cos x)$

BSP. (7.3.3) Wir betrachten aus BSP. (7.3.1) die Funktion $f(x) := (1+\cos)^{-1}e^{ix} = (2\cos^2\frac{x}{2})^{-1}e^{ix},\ x \in (-\pi, +\pi)$. Wir berechnen nochmals $f'(x)$ unter Verwendung von Satz 7.6 und der Quotientenregel:

$$f'(x) = \frac{i\cos^2\frac{x}{2} + \sin\frac{x}{2}\cos\frac{x}{2}}{2\,\cos^4\frac{x}{2}}\,e^{ix} = \Big(i + \tan\frac{x}{2}\Big)f(x) = \frac{ie^{-ix/2}}{\cos\frac{x}{2}}\,f(x) = \frac{e^{i(\pi-x)/2}}{\cos\frac{x}{2}}\,f(x).$$

Zwischen dem Ortsvektor $f(x)$ und Tangentenvektor $f'(x)$ besteht hier der Zusammenhang

$$|f'(x)| = \frac{|f(x)|}{\cos\frac{x}{2}}, \quad \gamma := \sphericalangle\,(f', f) = \frac{\pi - x}{2}.$$

Die Funktion f beschreibt in $\mathbf{C}$ eine **Parabel** mit Scheitel im Punkt $(u, v) := (\frac{1}{2}, 0)$. Um das einzusehen, verwenden wir Polarkoordinaten $u = r(x)\cos x,\ v = r(x)\sin x$. Es ist klar, aus der Zerlegung $f(x) = u(x) + i\,v(x)$ erschließen wir $r(x) = (1+\cos x)^{-1} = (1 + \frac{u}{r})^{-1}$, und somit $r(x) = 1 - u$. Demgemäß gilt $u^2 + v^2 = r^2 = (1-u)^2$, und durch Auflösen nach v erhalten wir die Normalform der Parabelgleichung

$$v = \pm\sqrt{1-2u}, \quad u \le \frac{1}{2}.$$

(II) Höhere Ableitungen. Die höheren Ableitungen der *skalaren* Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ (mit $\mathbf{K} := \mathbf{R}$ oder $\mathbf{K} := \mathbf{C}$) werden rekursiv definiert:

Definition 7.2 *Es sei $f : D(f) \to \mathbf{K}$ gegeben. Dann gelte $f''(x) := (f')'(x)$, $f'''(x) := (f'')'(x), \ldots$, allgemein:*

$$\boxed{f^{(n+1)}(x) := \Big(f^{(n)}\Big)'(x),}$$

*sofern diese Ausdrücke existieren. Man nennt $f^{(n)}(x)$, $n \in \mathbf{N}_0$, die n–**te Ableitung** der Funktion f im Punkt $x \in D(f)$, wobei insbesondere $f^{(0)}(x) := f(x)$ gesetzt wird. Eine gleichwertige Schreibweise ist*

$$f^{(n)}(x) \equiv \frac{d^n f}{dx^n}(x).$$

BSP. (7.3.4) Die folgenden Ableitungen erhält man durch wiederholte Anwendung der bekannten Ableitungsregeln.

$$\boxed{\Big(x^m\Big)^{(n)} = \begin{cases} \dfrac{m!\,x^{m-n}}{(m-n)!} & : \ n \le m, \\ 0 & : \ n > m \end{cases} \quad \forall\, x \in \mathbf{R}\ \forall\, m, n \in \mathbf{N}.}$$

$$\boxed{\begin{aligned} \Big(e^{\lambda x}\Big)^{(n)} &= \lambda^n e^{\lambda x}, \ \forall\, x \in \mathbf{R},\ \lambda \in \mathbf{C},\ n \in \mathbf{N}_0, \\ (\sin x)^{(2n)} &= (-1)^n \sin x, \quad (\cos x)^{(2n)} = (-1)^n \cos x \ \ \forall\, x \in \mathbf{R},\ n \in \mathbf{N}_0, \\ (\sin x)^{(2n+1)} &= (-1)^n \cos x, \quad (\cos x)^{(2n+1)} = (-1)^{n+1} \sin x \ \ \forall\, x \in \mathbf{R},\ n \in \mathbf{N}_0. \end{aligned}}$$

Für Polynome $P_n(x) := \sum_{k=0}^{n} a_k x^k,\ a_k \in \mathbf{K}, a_n \neq 0$, folgt hieraus insbesondere

$$\boxed{P_n^{(n)}(x) = n!\, a_n, \quad P_n^{(n+1)}(x) = 0 \ \ \forall\, x \in \mathbf{R}.}$$

Bemerkung 7.5 (a) Ist $X \subset \mathbf{R}$ ein Intervall oder eine (nicht notwendig endliche) Vereinigung von Intervallen, so fasst man die Menge aller auf X **stetigen** Funktionen $f \in \mathrm{Abb}\,(X, \mathbf{K})$ zusammen unter dem Symbol

$$\boxed{C^0(X) := C(X) := \{f : X \to \mathbf{K} \ :\ f(x) \ \text{ist stetig} \ \ \forall\, x \in X\}.}$$

Zum Beispiel: $C([a,b])$, $C(\mathbf{R})$, $C(\overline{\mathbf{R}}_+)$ usw.

Die Menge $C(X)$ ist ein **Vektorraum** über dem Körper $\mathbf{K}$ mit folgender linearer Struktur:

$$''+'': \quad (f+g)(x) := f(x) + g(x), \qquad ''\lambda\cdot'': \quad (\lambda\, f)(x) := \lambda\, f(x) \ \ \forall\, \lambda \in \mathbf{K}.$$

(b) Ist $f \in C(X)$ auf der Menge X differenzierbar, und ist $f' : X \to \mathbf{K}$ wiederum stetig, so heiße f **stetig differenzierbar auf** X**.** Man setzt

$$\boxed{C^1(X) := \{f : X \to \mathbf{K} \ :\ f \ \text{ ist stetig differenzierbar auf } X\}.}$$

Satz 7.2 stellt sicher, dass $C^1(X)$ ebenfalls ein **Vektorraum** über dem Körper $\mathbf{K}$ ist. Die lineare Struktur ist dieselbe wie auf dem Vektorraum $C(X)$. Da differenzierbare Funktionen notwendig stetig sind, haben wir $C^1(X) \subset C^0(X)$ mit echter Inklusion.

(c) Auf dem Vektorraum $C^1(X)$ ist der **Ableitungsoperator** D wohldefiniert, der jeder Funktion $f \in C^1(X)$ die Ableitung $f' \in C^0(X)$ zuordnet:

$$D : \begin{cases} C^1(X) \to C^0(X), \\ f \mapsto Df := f' \ \text{ oder punktweise } \ f(x) \mapsto (Df)(x) := f'(x). \end{cases}$$

Mit Hilfe von Summen– und Produktregel erhält man für $\lambda, \mu \in \mathbf{K}$ und $f, g \in C^1(X)$ die Beziehung $[D(\lambda\, f + \mu\, g)](x) = (\lambda\, f + \mu\, g)'(x) = \lambda\, f'(x) + \mu\, g'(x) = \lambda\, (Df)(x) + \mu\, (Dg)(x)$. Das heißt,

$$\boxed{D : C^1(X) \to C^0(X) \text{ ist eine lineare Abbildung.}}$$

(d) Die folgenden Verallgemeinerungen liegen auf der Hand. Für $k \in \mathbf{N}$ setzt man

$$\boxed{C^k(X) := \{f : X \to \mathbf{K} \ :\ f \ \text{ ist } k\text{–mal stetig differenzierbar auf } X\}.}$$

Jeder dieser Räume ist wiederum ein Vektorraum über dem Körper $\mathbf{K}$, angeordnet in der folgenden Hierarchie:

$$\cdots \subset C^{k+1}(X) \subset C^k(X) \subset C^{k-1}(X) \subset \cdots \subset C^0(X).$$

Ebenso unmissverständlich ist die Schreibweise

$$(D^k f)(x) := f^{(k)}(x) = \frac{d^k f}{dx^k}(x), \quad f \in C^k(X).$$

Hiermit prüft man leicht nach, dass die Abbildung $D^k : C^k(X) \to C^0(X)$ wieder **linear** ist.$\square$

Der folgende Satz beinhaltet eine Verallgemeinerung der Produktregel auf n–te Ableitungen:

Satz 7.7 (Leibnizsche Differentiationsregel)
Für Funktionen $f, g \in C^n(X)$ existieren auf X die stetigen Ableitungen $D^k(fg)$, $0 \le k \le n$, und es gilt

$$\boxed{[D^k(fg)](x) = \sum_{j=0}^{k} \binom{k}{j} (D^j f)(x) \cdot (D^{k-j} g)(x) \;\; \forall\, x \in X; \quad D^0 := Id.}$$

Den Beweis erbringt man sehr einfach durch vollständige Induktion nach k. □

BSP. (7.3.5) Unter Verwendung der Ableitungen aus BSP. (7.3.4) berechnet man mit Hilfe der Leibniz–Regel:

$$\begin{aligned} D^4(\cos x \cosh x) &= \cos x \cosh x + 4(-\sin x)\sinh x + 6(-\cos x)\cosh x + 4 \sin x \sinh x + \cos x \cosh x \\ &= -4 \cos x \cosh x, \end{aligned}$$

$$D^n(x^m\, e^{\lambda x}) = e^{\lambda x} \cdot \sum_{j=0}^{\min\{m,n\}} \binom{n}{j} \frac{m!}{(m-j)!}\, x^{m-j} \lambda^{n-j}, \;\; \lambda \in \mathbf{C}; \quad n, m \in \mathbf{N}.$$

(III) Ableitungen von abstrakten (vektorwertigen) Funktionen über $\mathbf{R}$. Der für skalarwertige Funktionen $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$, $D(f) \subset \mathbf{R}$, erklärte Ableitungsbegriff lässt sich sinngemäß auf abstrakte Funktionen $f \in \mathrm{Abb}\,(\mathbf{R}, Y)$, $D(f) \subset \mathbf{R}$, übertragen, wenn wir annehmen, dass Y ein **normierter Vektorraum** ist. Es ist üblich, in diesem Fall die unabhängige Variable mit t zu bezeichnen.

Definition 7.3 *Es sei Y ein normierter Vektorraum mit einer Norm $\|\cdot\|_Y : Y \to \mathbf{R}$. Eine abstrakte Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, Y)$ heiße im Punkt $t_0 \in D(f) \subset \mathbf{R}$* **differenzierbar**, *wenn es ein Element $f'(t_0) \in Y$ gibt mit*

$$\boxed{\lim_{n\to\infty} \left\| \frac{f(t_0+\epsilon_n) - f(t_0)}{\epsilon_n} - f'(t_0) \right\|_Y = 0 \;\; \forall \text{ Nullfolgen } (\epsilon_n)_{n\ge 0} \subset \mathbf{R} \text{ mit } t_0 + \epsilon_n \in D(f).}$$

Mit dieser Definition können insbesondere Ableitungen von **vektorwertigen** Funktionen $\vec{f} : D(\vec{f}) \to \mathbf{K}^n$ und von **matrixwertigen** Funktionen $A : D(A) \to \mathbf{K}^{(m,n)}$ erklärt werden. Die folgenden Aussagen können ohne Schwierigkeit als richtig verifiziert werden:

Auf dem Vektorraum $\mathbf{K}^n$ sei eine Vektornorm $\|\cdot\| : \mathbf{K}^n \to \mathbf{R}$ gegeben, man vgl. Abschnitt 6.5, Definition 6.22 ff. Genau dann ist eine vektorwertige Funktion $\vec{f} : D(\vec{f}) \to \mathbf{K}^n$ mit

$$\vec{f}(t) := \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix}, \quad f_k : D(\vec{f}) \to \mathbf{K}, \;\; 1 \le k \le n,$$

in einem Punkt $t_0 \in D(\vec{f}) \subset \mathbf{R}$ differenzierbar, wenn jede ihrer Komponentenfunktionen f_k, $1 \le k \le n$, in t_0 differenzierbar ist.

Auf dem Vektorraum $\mathbf{K}^{(m,n)}$ der $m\times n$–Matrizen sei eine Matrixnorm $\|\cdot\| : \mathbf{K}^{(m,n)} \to \mathbf{R}$ gegeben, man vgl. Abschnitt 6.5, Definition 6.21. Genau dann ist eine matrixwertige Funktion $A : D(A) \to \mathbf{K}^{(m,n)}$ mit

$$A(t) := \begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{m1}(t) & \cdots & a_{mn}(t) \end{bmatrix}, \quad a_{jk} : D(A) \to \mathbf{K}, \ 1 \le j \le m, \ 1 \le k \le n,$$

in einem Punkt $t_0 \in D(A) \subset \mathbf{R}$ differenzierbar, wenn jede ihrer Komponentenfunktionen a_{jk}, $1 \le j \le m$, $1 \le k \le n$, in t_0 differenzierbar ist.

In beiden Fällen gilt natürlich

$$\vec{f}\,'(t) := \begin{bmatrix} f_1'(t) \\ f_2'(t) \\ \vdots \\ f_n'(t) \end{bmatrix}, \quad A'(t) := \begin{bmatrix} a_{11}'(t) & \cdots & a_{1n}'(t) \\ \vdots & \ddots & \vdots \\ a_{m1}'(t) & \cdots & a_{mn}'(t) \end{bmatrix}.$$

Für vektor– und matrixwertige Funktionen gelten verschiedene Formen der Produktregel, die sich jeweils aus der komponentenweisen Anwendung des Satzes 7.2 ergeben. Dabei setzen wir stets die Existenz aller auftretenden Ableitungen voraus.

Ableitungsregeln

(a)	$\frac{d}{dt} A(t)\vec{f}(t) = A'(t)\vec{f}(t) + A(t)\vec{f}\,'(t)$ für $A : D(A) \to \mathbf{K}^{(m,n)}$ und $\vec{f} : D(\vec{f}) \to \mathbf{K}^n$
(b)	$\frac{d}{dt} g(t)\vec{f}(t) = g'(t)\vec{f}(t) + g(t)\vec{f}\,'(t)$ für $g : D(g) \to \mathbf{K}$ und $\vec{f} : D(\vec{f}) \to \mathbf{K}^n$
(c)	$\frac{d}{dt} \langle \vec{g}(t), \vec{f}(t)\rangle = \langle \vec{g}\,'(t), \vec{f}(t)\rangle + \langle \vec{g}(t), \vec{f}\,'(t)\rangle$ für $\vec{g} : D(\vec{g}) \to \mathbf{K}^n$ und $\vec{f} : D(\vec{f}) \to \mathbf{K}^n$

(3.1)

Ableitungsregeln (Fortsetzung)

(d)	$\frac{d}{dt} \langle \vec{f}(t), \vec{f}(t)\rangle = \frac{d}{dt}\|\vec{f}(t)\|^2 = 2\,\mathrm{Re}\,\langle \vec{f}(t), \vec{f}\,'(t)\rangle$ für $\vec{f} : D(\vec{f}) \to \mathbf{K}^n$
(e)	$\frac{d}{dt} A(t)B(t) = A'(t)B(t) + A(t)B'(t)$ für $A : D(A) \to \mathbf{K}^{(m,n)}$ und $B : D(B) \to \mathbf{K}^{(n,l)}$
(f)	$\frac{d}{dt} [\vec{g}(t) \times \vec{f}(t)] = \vec{g}\,'(t) \times \vec{f}(t) + \vec{g}(t) \times \vec{f}\,'(t)$ für $\vec{g} : D(\vec{g}) \to \mathbf{R}^3$ und $\vec{f} : D(\vec{f}) \to \mathbf{R}^3$

(3.1)

Kurven im $\mathbf{R}^n$ und deren Tangentenvektoren. Wir haben in der Definition 6.20 bereits festgelegt, was wir unter einer Parameterdarstellung einer **räumlichen** Kurve verstehen. Wir fassen allgemeiner:

Definition 7.4 *Im Euklidischen Vektorraum* $Y := \mathbf{R}^n$ *heiße eine Punktmenge* $\Gamma := \left\{ \vec{f}(t) \in \mathbf{R}^n \,:\, \vec{f}(t) = \Big(f_1(t), f_2(t), \ldots, f_n(t)\Big)^T,\ t \in I \right\}$ *eine* **differenzierbare Kurve**, *wenn die Komponentenfunktionen*

$$f_1(t), f_2(t), \ldots, f_n(t), \quad t \in I,$$

auf dem Intervall $I \subset \mathbf{R}$ *differenzierbar sind. Für jedes* $t \in I$ *heiße der Vektor*

$$\boxed{\vec{f}\,'(t) := \Big(f_1'(t), f_2'(t), \ldots, f_n'(t)\Big)^T}$$

der **Tangentenvektor an** Γ **im Punkt** $\vec{f}(t)$. *Ist* $\vec{f}\,'(t_0) \neq \vec{0}$ *in einem Punkt* $t_0 \in I$, *so heiße die Gerade*

$$\boxed{T := \{\vec{x} \in \mathbf{R}^n \,:\, \vec{x} = \vec{f}(t_0) + \lambda\, \vec{f}\,'(t_0),\ \lambda \in \mathbf{R}\}}$$

die **Tangente an** Γ **im Punkt** $\vec{f}(t_0)$. *Ein Vektor* $\vec{w} \in \mathbf{R}^n$ *steht* **im Punkt** $\vec{f}(t_0)$ **senkrecht auf** Γ, *wenn gilt:*

$$\boxed{\langle \vec{w}, \vec{f}\,'(t_0) \rangle = 0.}$$

BSP. (7.3.6) Es sei in $Y := \mathbf{R}^3$ die Parameterdarstellung $\vec{f}(t) := (r\cos t, r\sin t, \frac{ht}{2\pi})^T,\ t \in I := [0, 2\pi],\ r > 0, h > 0$ fest, einer räumlichen Kurve Γ gegeben. Γ ist eine **Schraubenlinie** vom Radius r und der Ganghöhe h. Der Tangentenvektor im Punkt $t \in I$ ist durch

$$\vec{f}\,'(t) := \Big(-r\sin t, r\cos t, \frac{h}{2\pi}\Big)^T$$

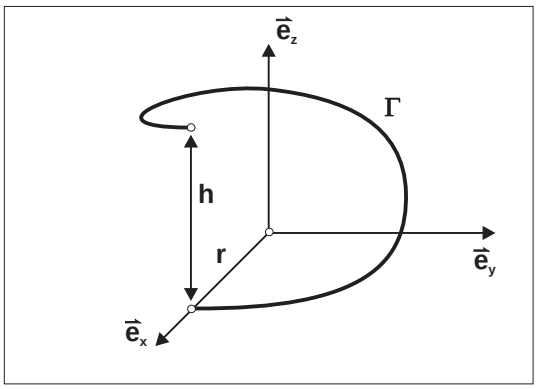
bestimmt, und es gilt demnach $\|\vec{f}\,'(t)\| = \sqrt{r^2 + \left(\frac{h}{2\pi}\right)^2} > 0$. Die Tangente an Γ im Punkt $\vec{f}(t_0)$ ist gegeben durch

$$T: \qquad \vec{x} = \begin{bmatrix} r\cos t_0 \\ r\sin t_0 \\ \frac{h\,t_0}{2\pi} \end{bmatrix} + \lambda \begin{bmatrix} -r\sin t_0 \\ r\cos t_0 \\ \frac{h}{2\pi} \end{bmatrix},\ \ \lambda \in \mathbf{R}.$$

Ein Vektor $\vec{w} = (w_1, w_2, w_3)^T \in \mathbf{R}^3$ steht im Punkt $\vec{f}(t_0)$ senkrecht auf Γ, falls

$$\langle \vec{w}, \vec{f}\,'(t_0) \rangle = r(-w_1 \sin t_0 + w_2 \cos t_0) + \frac{h w_3}{2\pi} = 0$$

gilt. Zum Beispiel erfüllt der Vektor $\vec{w} := \lambda\,(\sin t_0, -\cos t_0, \frac{2\pi r}{h})^T,\ \lambda \in \mathbf{R}$, diese Bedingung.



Die Schraubenlinie der Ganghöhe h

Bemerkung 7.6 Wird die Variable t als *physikalischer Zeitparameter* gedeutet, so beschreibt die Funktion $\vec{f}(t)$ die **Bewegung** eines Körpers auf einer Bahn Γ. In diesem Zusammenhang wählt man die Bezeichnung

$$\boxed{\frac{d}{dt} =: '' \cdot ''}$$

Man beschränkt sich auf den Raum $Y := \mathbf{R}^3$ als realen physikalischen Raum. Es heiße

$$\boxed{\begin{aligned} \vec{v}(t) &:= \dot{\vec{f}}(t) = \left(\dot{f}_1(t), \dot{f}_2(t), \dot{f}_3(t)\right)^T \quad \textbf{Geschwindigkeitsvektor}, \\ \vec{b}(t) &:= \ddot{\vec{f}}(t) = \left(\ddot{f}_1(t), \ddot{f}_2(t), \ddot{f}_3(t)\right)^T \quad \textbf{Beschleunigungsvektor} \end{aligned}}$$

des Körpers zum Zeitpunkt t im Bahnpunkt $\vec{f}(t)$. □

BSP. (7.3.7) Ein Körper auf der Bahnkurve $\vec{f}(t) := \left(t - \cos t, 3 + \sin t, t + \cos 2t\right)^T, t \geq 0$, hat zum Zeitpunkt t den folgenden Geschwindigkeitsvektor $\vec{v}(t) = \dot{\vec{f}}(t)$ und den Beschleunigungsvektor $\vec{b}(t) = \ddot{\vec{f}}(t)$:

$$\vec{v}(t) = \begin{bmatrix} 1 + \sin t \\ \cos t \\ 1 - 2\sin 2t \end{bmatrix}, \quad \vec{b}(t) = \begin{bmatrix} \cos t \\ -\sin t \\ -4\cos 2t \end{bmatrix}.$$

Die (euklidische) Länge $v(t) := \|\vec{v}(t)\|$ des Geschwindigkeitsvektors gibt die **Absolutgeschwindigkeit** im Zeitpunkt t an. Wir haben hier

$$v(t) = \sqrt{5 + 2\sin t - 4\sin 2t - 2\cos 4t}.$$

Der Geschwindigkeitsvektor $\vec{v}(t)$ fällt ganz offensichtlich mit dem Tangentenvektor $\dot{\vec{f}}(t)$ zusammen. Wir wollen feststellen, welche Richtung der Beschleunigungsvektor $\vec{b}(t)$ hat. Den Vektorraum der über dem Intervall $I \subset \mathbf{R}$ zweimal stetig differenzierbaren räumlichen Bahnkurven bezeichnen wir mit

$$\boxed{C^2(I; \mathbf{R}^3) := \{\vec{f} : I \to \mathbf{R}^3 \ : \ \vec{f} \text{ zweimal stetig differenzierbar auf } I\}.}$$

Definition 7.5 *Es sei $\vec{f} \in C^2(I; \mathbf{R}^3)$ die Parameterdarstellung einer räumlichen Bahnkurve Γ. In jedem Punkt $t \in I$ mit $\dot{\vec{f}}(t) \neq \vec{0}$ ist der Vektor*

$$\boxed{\vec{T}(t) := \frac{\dot{\vec{f}}(t)}{\|\dot{\vec{f}}(t)\|}}$$

erklärt. Dieser heiße der **Tangenteneinheitsvektor** *an Γ im Punkt $\vec{f}(t)$. In jedem Punkt $t \in I$ mit $\dot{\vec{f}}(t) \neq \vec{0} \neq \dot{\vec{T}}(t)$ ist der Vektor*

$$\boxed{\vec{N}(t) := \frac{\dot{\vec{T}}(t)}{\|\dot{\vec{T}}(t)\|}}$$

erklärt. Dieser heiße der **Normalenvektor** *an Γ im Punkt $\vec{f}(t)$.*

Satz 7.8 *Es sei $\vec{f} \in C^2(I; \mathbf{R}^3)$ die Parameterdarstellung einer räumlichen Bahnkurve Γ. Dann gilt in jedem Punkt $t \in I$ mit $\dot{\vec{f}}(t) \neq \vec{0} \neq \dot{\vec{T}}(t)$:*

$$\boxed{\|\vec{T}(t)\| = 1 = \|\vec{N}(t)\|, \quad \langle \vec{T}(t), \vec{N}(t) \rangle = 0, \quad \textit{also } \vec{N}(t) \perp \vec{T}(t).}$$

Begründung: Wir brauchen nur die Orthogonalität $\vec{N}(t) \perp \vec{T}(t)$ zu zeigen. Diese folgt aber aus der Produktregel (3.1.d):

$$0 = \frac{d}{dt} \|\vec{T}(t)\|^2 = 2 \langle \vec{T}(t), \dot{\vec{T}}(t) \rangle = 2 \langle \vec{T}(t), \vec{N}(t) \rangle \, \|\dot{\vec{T}}(t)\|.$$

Bemerkung 7.7 In jedem Punkt $t \in I$ mit $\dot{\vec{f}}(t) \neq \vec{0} \neq \dot{\vec{T}}(t)$ bilden die Vektoren $\vec{T}(t), \vec{N}(t)$ eine **ON–Basis** für den zweidimensionalen Unterraum $U(t) := \operatorname{span}\{\vec{T}(t), \vec{N}(t)\}$. Die Ebene $E(t) := \vec{f}(t) + U(t)$ enthält sowohl den Tangential– als auch den Normalenvektor der Bahnkurve Γ. Die Ebene $E(t)$ passt sich also im Punkt $\vec{f}(t)$ dem Verlauf der Bahnkurve Γ *bestmöglich* an: □

Definition 7.6 *In einem Punkt $t \in I$ mit $\dot{\vec{f}}(t) \neq \vec{0} \neq \dot{\vec{T}}(t)$ heiße die Ebene*

$$\boxed{E(t) := \{\vec{x} \in \mathbf{R}^3 \,:\, \vec{x} = \vec{f}(t) + \lambda \vec{T}(t) + \mu \vec{N}(t), \quad \lambda, \mu \in \mathbf{R}\}}$$

die **Schmiegebene** *im Punkt $\vec{f}(t)$ an die Bahnkurve Γ.*

Wir zeigen jetzt, dass der Beschleunigungsvektor $\vec{b}(t)$ in der Schmiegebene der Bahnkurve Γ liegt. Das heißt, es gilt $\vec{b}(t) = \lambda \vec{T}(t) + \mu \vec{N}(t)$. In der Tat, wir haben $\vec{v}(t) = \dot{\vec{f}}(t) = \|\vec{v}(t)\| \vec{T}(t)$, und hieraus folgt durch Differentiation unter Verwendung der Produktregel (3.1.b):

$$\begin{aligned} \vec{b}(t) &= \ddot{\vec{f}}(t) = \frac{d}{dt}\vec{v}(t) = \frac{d}{dt}\left(\|\vec{v}(t)\|\,\vec{T}(t)\right) = \left(\frac{d}{dt}\|\vec{v}(t)\|\right)\vec{T}(t) + \|\vec{v}(t)\|\,\dot{\vec{T}}(t) \\ &= \underbrace{\frac{d}{dt}\|\vec{v}(t)\|}_{=:\lambda}\,\vec{T}(t) + \underbrace{\|\vec{v}(t)\|\,\|\dot{\vec{T}}(t)\|}_{=:\mu}\,\vec{N}(t). \end{aligned}$$

Bemerkung 7.8 (a) Der Vektor $\lambda \vec{T}(t) = \left(\frac{d}{dt}\|\vec{v}(t)\|\right)\vec{T}(t)$ heißt die **Tangentialkomponente der Beschleunigung**, und der Vektor $\mu \vec{N}(t) = \|\vec{v}(t)\|\,\|\dot{\vec{T}}(t)\|\,\vec{N}(t)$ heißt die **Normalkomponente der Beschleunigung**.

(b) Wegen $\vec{T}(t) \perp \vec{N}(t)$ folgern wir $\|\vec{b}(t)\|^2 = \lambda^2 + \mu^2$, und diese Beziehung benutzt man meistens zur Berechnung von $\mu > 0$, wenn die Skalare λ und $\|\vec{b}(t)\|$ bekannt sind. □

BSP. (7.3.8) Die Bahnkurve eines Körpers sei durch folgende vektorwertige Funktion gegeben: $\vec{f}(t) := (t, \frac{1}{2}t^2, \frac{1}{3}t^3)^T$, $t \geq 0$. Zu berechnen sind in jedem Bahnpunkt die Tangential– und die Normalkomponente der Beschleunigung. *Lösung:* Der Geschwindigkeitsvektor $\vec{v}(t) = \dot{\vec{f}}(t) = (1, t, t^2)^T$ und der Beschleunigungsvektor $\vec{b}(t) = \ddot{\vec{f}}(t) = (0, 1, 2t)^T$ verschwinden in keinem Zeitpunkt $t \geq 0$. Die Absolutgeschwindigkeit beträgt $v(t) = \|\vec{v}(t)\| = \sqrt{1 + t^2 + t^4}$, und es gilt $\|\vec{b}(t)\| = \sqrt{1 + 4t^2}$. Aus diesen Daten erhalten wir $\lambda = (d/dt)v(t) = (t + 2t^3)/\sqrt{1 + t^2 + t^4}$ sowie $\mu = \sqrt{\|\vec{b}(t)\|^2 - \lambda^2} = \sqrt{1 + 4t^2 + t^4}/\sqrt{1 + t^2 + t^4}$. Somit folgt

$$\vec{T}(t) = \frac{\vec{v}(t)}{v(t)} = \frac{1}{\sqrt{1 + t^2 + t^4}}\begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}, \quad \vec{b}_{tang}(t) = \lambda \vec{T}(t) = \frac{t(1 + 2t^2)}{1 + t^2 + t^4}\begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}.$$

Schließlich erhält man noch die Normalkomponente der Beschleunigung und den Normalenvektor

$$\vec{b}_{norm}(t) = \vec{b}(t) - \vec{b}_{tang}(t) = \frac{1}{1 + t^2 + t^4}\begin{bmatrix} -t - 2t^3 \\ 1 - t^4 \\ 2t + t^3 \end{bmatrix},$$

$$\vec{N}(t) = \frac{1}{\mu}\vec{b}_{norm}(t) = \frac{1}{\sqrt{(1 + 4t^2 + t^4)(1 + t^2 + t^4)}}\begin{bmatrix} -t - 2t^3 \\ 1 - t^4 \\ 2t + t^3 \end{bmatrix}.$$

Bemerkung 7.9 (a) Aus der Produktregel (3.1.d) erhält man für jede **stetig differenzierbare Kurve** Γ $\vec{f} : I \to \mathbf{R}^n$ mit $\|\vec{f}(t)\| = const$, $t \in I$:

$$\frac{d}{dt}\|\vec{f}(t)\|^2 = 0 = 2\langle \vec{f}(t), \vec{f}\,'(t)\rangle,$$

das heißt, in diesem Fall steht der Tangentenvektor $\vec{f}\,'(t)$ an die Kurve Γ in jedem Punkt $\vec{f}(t)$ **senkrecht** auf dem Ortsvektor $\vec{f}(t)$. Speziell für die Bewegung eines Körpers auf einer Bahnkurve $\Gamma \subset \mathbf{R}^3$ folgt:

(i) Liegt die Bahnbewegung eines Körpers auf einer Kugel $\|\vec{f}(t)\| = r = const$, so ist der Geschwindigkeitsvektor $\dot{\vec{f}}(t)$ in jedem Punkt **tangential** zur Kugel.

(ii) Ist die Absolutgeschwindigkeit $\|\dot{\vec{f}}(t)\|$ der Bahnbewegung eines Körpers konstant, so hat der Beschleunigungsvektor $\vec{b}(t) = \ddot{\vec{f}}(t)$ nur eine Normalkomponente. Denn die Tangentialkomponente $\lambda \vec{T}(t)$ verschwindet wegen $\lambda = \frac{d}{dt}\|\vec{v}(t)\| = 0$.

(b) Es sei $Q : I \to \mathbf{K}^{(n,n)}$ eine stetig differenzierbare matrixwertige Funktion, und jede der Matrizen $Q(t)$, $t \in I$, sei **unitär**. Wegen $Q^*(t)Q(t) = Id = Q(t)Q^*(t)$ erhält man aus der Produktregel (3.1.e):

$$\frac{d}{dt}\,Q^*(t)Q(t) = O = \dot{Q}^*(t)Q(t) + Q^*(t)\dot{Q}(t), \quad \frac{d}{dt}\,Q(t)Q^*(t) = O = \dot{Q}(t)Q^*(t) + Q(t)\dot{Q}^*(t),$$

mit anderen Worten, die Matrizen $\dot{Q}^*(t)Q(t)$ und $\dot{Q}(t)Q^*(t)$ sind **antihermitesch**. *Zahlenbeispiel:* Die folgende matrixwertige Funktion $Q = \mathbf{R} \to \mathbf{R}^{(3,3)}$ ist für jedes $t \in \mathbf{R}$ orthogonal:

$$Q(t) := \begin{bmatrix} \cos t & 0 & -\sin t \\ 0 & 1 & 0 \\ \sin t & 0 & \cos t \end{bmatrix} \Rightarrow Q^T(t) = \begin{bmatrix} \cos t & 0 & \sin t \\ 0 & 1 & 0 \\ -\sin t & 0 & \cos t \end{bmatrix}.$$

Es folgt

$$\dot{Q}(t) = \begin{bmatrix} -\sin t & 0 & -\cos t \\ 0 & 0 & 0 \\ \cos t & 0 & -\sin t \end{bmatrix}, \quad \dot{Q}^T(t) = \begin{bmatrix} -\sin t & 0 & \cos t \\ 0 & 0 & 0 \\ -\cos t & 0 & -\sin t \end{bmatrix},$$

und somit

$$\dot{Q}(t)Q^T(t) = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad Q(t)\dot{Q}^T(t) := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

Bedeutung in der Mechanik: Die Matrix $Q(t) \in \mathbf{R}^{(3,3)}$, $t \in I$, sei **orthogonal**. Dann beschreibt die vektorwertige Funktion $\vec{f}(t) := Q(t)\vec{x}_0$, $t \in I$, $\vec{x}_0 \in \mathbf{R}^3$, eine zeitlich ablaufende **Raumdrehung**. Der Geschwindigkeitsvektor ist gegeben durch

$$\dot{\vec{f}}(t) = \dot{Q}(t)\vec{x}_0 = \dot{Q}(t)Q^T(t)[Q(t)\vec{x}_0] = \dot{Q}(t)Q^T(t)\vec{f}(t).$$

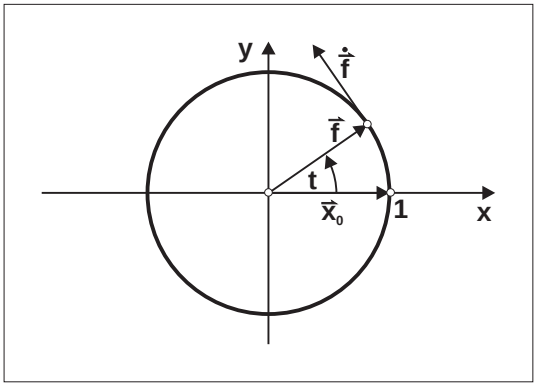
Die antisymmetrische Matrix $\dot{Q}(t)Q^T(t)$ heißt die **Spinmatrix** der Raumdrehung; sie ordnet jedem Punkt $\vec{f}(t)$ der Bewegungskurve den Geschwindigkeitsvektor $\dot{\vec{f}}(t) = \dot{Q}(t)Q^T(t)\vec{f}(t)$ zu. Aus Satz 5.22 erhalten wir die Relation $\|Q(t)\vec{x}_0\| = \|\vec{x}_0\| = const\ \forall\ t \in I$. Demzufolge gilt wiederum

$$0 = \frac{d}{dt}\,\|Q(t)\vec{x}_0\|^2 = 2\,\langle Q(t)\vec{x}_0, \dot{Q}(t)\vec{x}_0\rangle = 2\langle \vec{f}(t), \dot{\vec{f}}(t)\rangle,$$

das heißt, der Geschwindigkeitsvektor steht senkrecht auf dem Ortsvektor des Bahnpunktes: $\dot{\vec{f}}(t) \perp \vec{f}(t)\ \forall\ t \in I$. *Zahlenbeispiel:* Es seien gegeben

$$Q(t) := \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix},\ t \in \mathbf{R},\ \ \vec{x}_0 := \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \vec{f}(t) = Q(t)\vec{x}_0 = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix},\ \ \dot{Q}(t)Q^T(t) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Hier beschreibt $\vec{f}(t)$ die Bewegung eines Körpers auf einer Kreisbahn vom Radius 1. Der Geschwindigkeitsvektor $\dot{\vec{f}}(t) = (-\sin t, \cos t)^T$ steht ganz offensichtlich senkrecht auf dem Ortsvektor $\vec{f}(t)$ des Bahnpunktes $\vec{f}(t)$. □



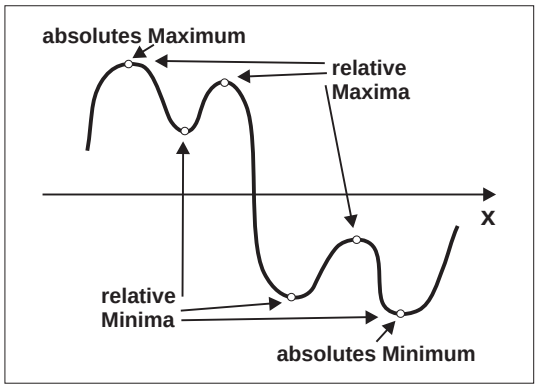
Bewegung auf einer Kreisbahn vom Radius 1

7.4 Der Mittelwertsatz der Differentialrechnung

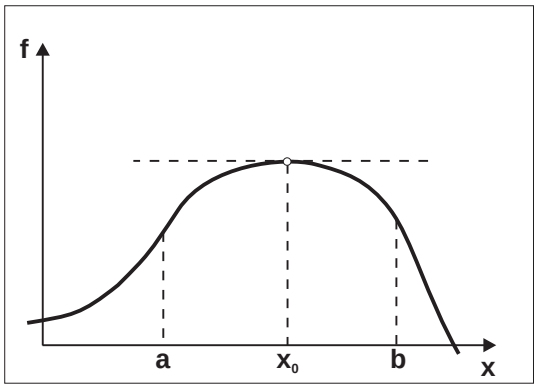
Für stetige **reellwertige** Funktionen $f : [a,b] \to \mathbf{R}$ existieren gemäß Satz 6.16 ein *absolutes* Minimum $f(\underline{x}) := \min\limits_{x\in[a,b]} f(x)$ und ein *absolutes* Maximum $f(\bar{x}) := \max\limits_{x\in[a,b]} f(x)$. Natürlich ist hiermit nichts darüber ausgesagt, wie der Graph $G(f)$ der Funktion f zwischen diesen beiden Extremwerten verläuft. Es braucht insbesondere nicht einmal Monotonie vorzuliegen.

Definition 7.7 *Für eine gegebene Funktion $f \in \mathrm{Abb}(\mathbf{R},\mathbf{R})$ heiße ein Punkt $x_0 \in D(f) \subset \mathbf{R}$ ein* **relatives Extremum** *(relatives Maximum bzw. relatives Minimum), wenn es ein Intervall $[a,b] \subseteq D(f)$ gibt, mit $x_0 \in (a,b)$ und*

$$\boxed{\begin{array}{ll} f(x) \le f(x_0) & \forall\, x \in [a,b] \ : \ \textbf{relatives Maximum}, \\ f(x) \ge f(x_0) & \forall\, x \in [a,b] \ : \ \textbf{relatives Minimum}. \end{array}}$$



Extrema einer reellwertigen Funktion



Relative Extrema sind der geometrische Ort horizontaler Tangenten

Die Ableitung $f'(x_0)$ verschwindet in einem relativen Extremum $x_0 \in D(f)$, falls f dort differenzierbar ist.

Satz 7.9 *Hat die reellwertige Funktion $f \in \mathrm{Abb}(\mathbf{R},\mathbf{R})$ in einem Punkt $x_0 \in D(f) \subset \mathbf{R}$ ein relatives Extremum, und ist f in x_0 differenzierbar, so gilt notwendig $f'(x_0) = 0$.*

Begründung: Es sei $x_0 \in D(f)$ ein relatives Maximum. Den Fall eines relativen Minimums beweist man ganz analog. Es gibt also ein Intervall $[a,b] \subseteq D(f)$ mit $x_0 \in (a,b)$ und

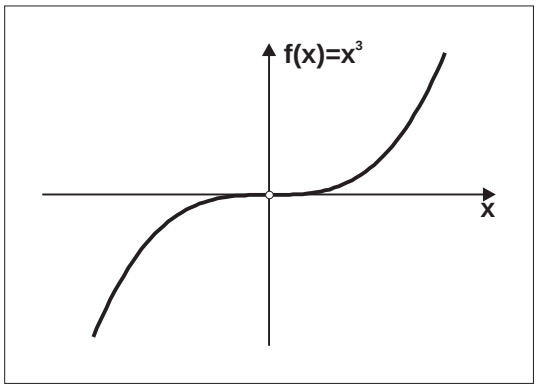
$$\frac{\Delta f}{\Delta x} := \frac{f(x)-f(x_0)}{x-x_0} \quad \begin{cases} \le 0 \ : \ x > x_0, \\ \ge 0 \ : \ x < x_0. \end{cases}$$

Da f in x_0 differenzierbar ist, existieren die Grenzwerte

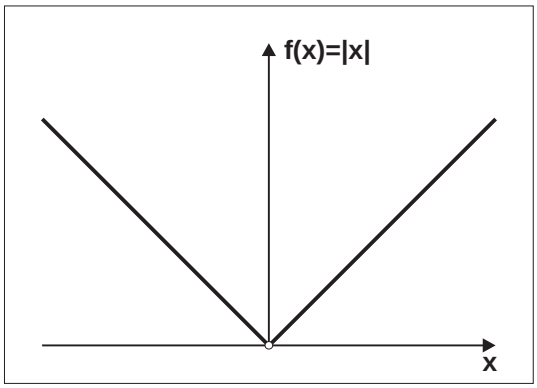
$$0 \le \lim_{x\to x_0-} \frac{\Delta f}{\Delta x} = f'(x_0) = \lim_{x\to x_0+} \frac{\Delta f}{\Delta x} \le 0.$$

Also muss $f'(x_0) = 0$ gelten. □

Bemerkung 7.10 (a) Die Bedingung $f'(x_0) = 0$ ist ein **notwendiges Kriterium** für die Existenz eines relativen Extremums bei $x_0 \in D(f)$. Es ist aber keineswegs schon *hinreichend*. *Zum Beispiel* sei $f(x) := x^3$, $x \in D(f) := \mathbf{R}$. Dann gilt $f'(x) = 3x^2$, und somit $f'(0) = 0$, obwohl im Punkt $x_0 = 0$ *kein* relatives Extremum liegt. Wir werden diese Situation noch genauer in Abschnitt 7.7 analysieren.



Für $f(x) := x^3$ gilt $f'(0) = 0$, obwohl bei $x_0 = 0$ kein Extremum vorliegt



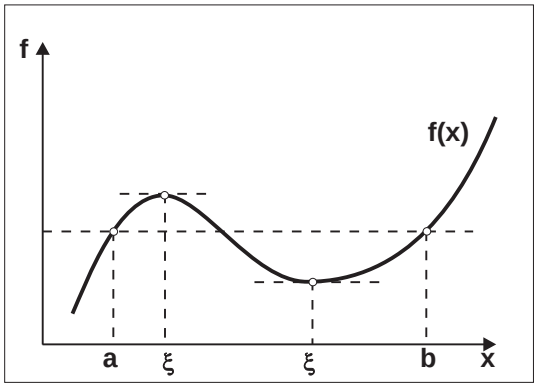
Die Funktion $f(x) := |x|$ hat bei $x_0 = 0$ ein Minimum, obwohl sie dort nicht differenzierbar ist

(b) Es können auch stetige, aber nicht differenzierbare Funktion (relative) Extrema haben, wie das *Beispiel* $f(x) := |x|,\ x \in D(f) := \mathbf{R}$, lehrt. Diese Funktion hat im Punkt $x_0 = 0$ ein absolutes Minimum, obwohl eine Ableitung $f'(x_0)$ nicht erklärt ist. □

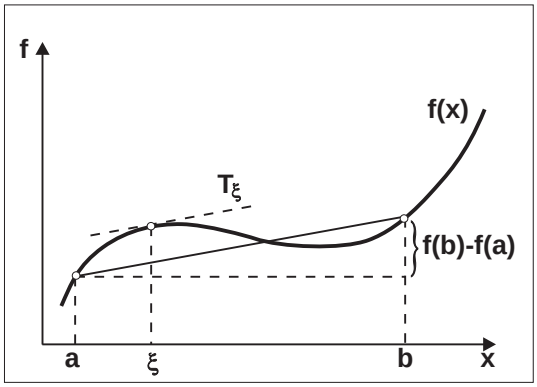
Als einfache Folgerung aus dem Satz 7.9 erhält man:

Satz 7.10 (Satz von ROLLE)
Die reellwertige Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R})$ sei in einem Intervall $[a, b] \subseteq D(f)$ stetig sowie differenzierbar in (a, b). Gelte ferner $f(a) = f(b)$. Dann gibt es mindestens eine Zwischenstelle $\xi \in (a, b)$ mit $f'(\xi) = 0$.

Begründung: Falls die Funktion f auf $[a, b]$ konstant ist, so gilt $f'(\xi) = 0$ in jedem Punkt $\xi \in (a, b)$. Ist f nicht konstant, so nimmt die Funktion f im Intervall $[a, b]$ gemäß Satz 6.16 sowohl ihr absolutes Maximum als auch ihr absolutes Minimum an, und beide Extrema sind voneinander verschieden. Wegen $f(a) = f(b)$ muss eines der beiden Extrema ein relatives Extremum in einem *inneren* Punkt $\xi \in (a, b)$ sein, und aus Satz 7.9 folgt dann $f'(\xi) = 0$. □



Zum Satz von Rolle



Zum Mittelwertsatz der Differentialrechnung

Der Satz von ROLLE hat nur einen Hilfscharakter. Man verwendet ihn nämlich zur Begründung des folgenden zentralen Satzes der Differentialrechnung:

Satz 7.11 (Mittelwertsatz der Differentialrechnung, MWS)
Die reellwertige Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R})$ sei stetig im Intervall $[a, b] \subseteq D(f)$ und differenzierbar in (a, b). Dann gilt:

$$\boxed{\exists\, \xi \in (a, b)\ :\ f'(\xi) = \frac{f(b) - f(a)}{b - a}.} \tag{4.1}$$

Begründung: Die Funktion

$$g(x) := f(x) - \frac{f(b)-f(a)}{b-a}\,(x-a)$$

hat die im Satz von ROLLE geforderten Stetigkeits– und Differenzierbarkeitseigenschaften. Ausserdem erfüllt sie $g(a) = f(a) = g(b)$, so dass ein Punkt $\xi \in (a,b)$ existiert mit

$$g'(\xi) = 0 = f'(\xi) - \frac{f(b)-f(a)}{b-a}.$$

Dies war zu zeigen. □

Bemerkung 7.11 (a) Die **geometrische** Aussage des Mittelwertsatzes ist die folgende (vgl. obige Skizze). Es gibt einen Punkt $(\xi, f(\xi)) \in G(f)$, in welchem die Tangente T_ξ an den Graph $G(f)$ **parallel** ist zur Geraden durch die beiden Punkte $(a, f(a)) \in G(f)$ und $(b, f(b)) \in G(f)$.

(b) Häufig ersetzt man in (4.1) das Intervall $[a,b]$ durch ein Intervall $[x_0, x]$ bzw. $[x, x_0]$, oder man setzt $\xi := x_0 + \theta\,(x - x_0)$ für ein $\theta \in (0,1)$:

$$\boxed{\begin{array}{rcl} f(x) - f(x_0) & = & f'(\xi)\,(x-x_0) \quad \text{für ein } \xi \in (x_0, x) \text{ (bzw. } \xi \in (x, x_0)), \\ f(x) - f(x_0) & = & f'[x_0 + \theta\,(x-x_0)]\,(x-x_0) \quad \text{für ein } \theta \in (0,1). \end{array}} \tag{4.2}$$

Der Mittelwertsatz sagt nichts darüber aus, **wo** die Zwischenstelle ξ bzw. θ liegt. Trotzdem gestattet dieser Satz eine Reihe von Anwendungen. □

BSP. (7.4.1) **Fehlerabschätzungen.** Durch Abschätzung der Ableitung $f'(\xi)$ in der Relation (4.2) kann häufig eine brauchbare Abschätzung für den Funktionswert $f(x)$ am Ende des Intervalls $[x_0, x]$ gefunden werden. *Zahlenbeispiel:* Für $0 < x < 1$ betrachten wir auf dem Intervall $[0, x]$ die Funktion $f(t) := \text{arc sin}_H\, t$. Dann existiert gemäß (4.2) ein Zwischenwert $\xi \in (0, x)$ mit

$$\text{arc sin}_H\, x = f(x) - f(0) = f'(\xi)\,(x-0) = \frac{x}{\sqrt{1-\xi^2}}.$$

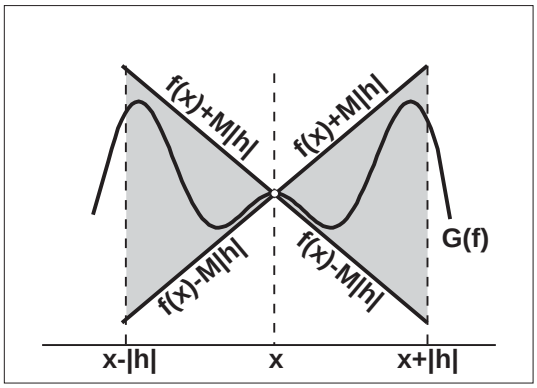
Bezieht man die Extremfälle $\xi = 0$ und $\xi = x$ mit ein, so resultiert die Abschätzung

$$x < \text{arc sin}_H\, x < \frac{x}{\sqrt{1-x^2}}.$$

Für $x := 0.01$ gelangt man zur Fehlereinschließung $0.01 < \text{arc sin}_H(0.01) < 0.010\,000\,5$. Nimmt man das arithmetische Mittel von unterer und oberer Fehlerschranke, so gelangt man zum Näherungswert

$$\text{arc sin}_H(0.01) \doteq 0.010\,000\,25 \pm \epsilon, \quad 0 < \epsilon < 2.5 \cdot 10^{-7},$$

welcher eine gute Approximation des Funktionswertes $\text{arc sin}_H(0.01) \doteq 0.010\,000\,166\,67$ darstellt.



Zur Lipschitz–Stetigkeit einer Funktion

BSP. (7.4.2) LIPSCHITZ–Stetigkeit. Ist die stetige reellwertige Funktion $f : [a,b] \to \mathbf{R}$ in jedem Punkt $x \in (a,b)$ differenzierbar und sind die Ableitungen $f'(x)$ auf $[a,b]$ beschränkt, so ist f LIPSCHITZ–stetig:

$$\boxed{\sup_{\xi\in(a,b)} |f'(\xi)| := M < +\infty \quad \Rightarrow \quad |f(x)-f(y)| \le M\,|x-y| \;\; \forall\, x,y \in [a,b].} \tag{4.3}$$

In der Tat, aus dem Mittelwertsatz in der Form (4.2) folgt nämlich $f(x)-f(y) = f'(\xi)\,(x-y),\ \xi := y+\theta\,(x-y)$ für ein $\theta \in (0,1)$ und für $x,y \in [a,b]$. Nimmt man Beträge, so folgt daraus schon (4.3). Zur Interpretation der LIPSCHITZ–Stetigkeit setzen wir in (4.3) $y = x+h$. Dann gilt:

$$f(x) - M\,|h| \le f(x+h) \le f(x) + M\,|h|,$$

das heißt, der Graph $G(f)$ verläuft in dem oben skizzierten Zwickel.

Mit Hilfe des Mittelwertsatzes gelingt es nun, die **Monotonie** einer differenzierbaren Funktion durch das Vorzeichen ihrer Ableitung zu charakterisieren.

Satz 7.12 *Die reellwertige Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ sei im Intervall $[a,b] \subseteq D(f)$ stetig und in (a,b) differenzierbar. Dann gilt:*

$$\boxed{\begin{array}{l} f'(x) > 0 \;\; \forall\, x \in (a,b) \;\Rightarrow\; f : [a,b] \to \mathbf{R} \;\; \textbf{streng monoton} \uparrow, \\ f'(x) < 0 \;\; \forall\, x \in (a,b) \;\Rightarrow\; f : [a,b] \to \mathbf{R} \;\; \textbf{streng monoton} \downarrow. \end{array}} \tag{4.4}$$

Begründung: Aus dem Mittelwertsatz folgern wir:

$$f(x) = f(x_0) + f'(\xi)\,(x-x_0) \begin{cases} > f(x_0) & \text{für } x > x_0, \text{ sofern } f'(\xi) > 0, \\ < f(x_0) & \text{für } x > x_0, \text{ sofern } f'(\xi) < 0. \end{cases}$$

BSP. (7.4.3) (i) Die Funktion $f(x) := e^x$ erfüllt $f'(x) = e^x > 0\ \forall\, x \in \mathbf{R}$. Sie ist somit auf ganz $\mathbf{R}$ streng monoton $\uparrow$.

(ii) Die Funktion $f(x) := \ln x$ erfüllt $f'(x) = \frac{1}{x} > 0\ \forall\ x > 0$. Sie ist somit auf dem Intervall $(0,+\infty)$ streng monoton $\uparrow$.

(iii) Das Polynom $P_{21}(x) := x^{21} + 5x^{17} + 3x^9 + 2x - 11$ erfüllt $P_{21}(1) = 0$. Ferner gilt $P'_{21}(x) = 21x^{20} + 85x^{16} + 27x^8 + 2 \ge 2 > 0\ \forall\, x \in \mathbf{R}$, das heißt $P_{21}(x)$ ist auf ganz $\mathbf{R}$ streng monoton $\uparrow$. Wegen $\lim\limits_{x\to\pm\infty} P_{21}(x) = \pm\infty$ hat $P_{21}(x)$ nur die eine Nullstelle $x_0 = 1$.

Typisch für Anwendungen des Mittelwertsatzes sind Problemstellungen, bei denen man aus Eigenschaften der Ableitung $f'(x)$ auf das Verhalten der Funktion $f(x)$ schließen möchte. Die einfachste dieser Anwendungen halten wir im folgenden Satz fest.

Satz 7.13 *Die skalarwertige Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{K})$ sei auf dem Intervall $[a,b] \subseteq D(f)$ stetig sowie differenzierbar in (a,b). Dann gilt*

$$\boxed{\forall\, x \in [a,b] \;:\; f(x) = const \quad \Leftrightarrow \quad \forall\, x \in (a,b) \;:\; f'(x) = 0.}$$

Begründung: Es ist klar, für $f = const$ gilt trivialerweise $f' = 0$. Nun sei umgekehrt $f'(x) = 0\ \forall\, x \in (a,b)$. Eine komplexwertige Funktion zerlegen wir gemäß $f(x) = u(x) + i\,v(x)$ in Real– und Imaginärteil. Wir haben $u'(x) = 0 = v'(x)\ \forall\, x \in (a,b)$. Zu festem $x \in (a,b]$ und passendem Zwischenwert $\xi \in (a,x)$ folgt aus dem Mittelwertsatz: $u(x) = u(a) + u'(\xi)\,(x-a) = u(a) = const$. Eine analoge Aussage erhält man für $v(x)$, so dass $f(x) = f(a) = const\ \forall\, x \in [a,b]$ resultiert. □

Bemerkung 7.12 Sind $f,g : [a,b] \to \mathbf{K}$ stetige und in (a,b) differenzierbare Funktionen mit $f'(x) = g'(x)\ \forall\, x \in (a,b)$, so folgt □

$$\boxed{f(x) = g(x) + const \;\; \forall\, x \in [a,b].}$$

BSP. (7.4.4) Es sei

$$\boxed{f(x) := x + \operatorname{arc\,tan}_H\left(\frac{1}{\tan x}\right),\ x \in D(f) := \mathbf{R} \setminus \{n\pi\,:\,n \in \mathbf{Z}\}.}$$

Man berechnet mit der Kettenregel die Ableitung

$$f'(x) = 1 + \frac{1}{1 + (1/\tan x)^2}\left(-\frac{1}{\tan^2 x}\right)\left(\frac{1}{\cos^2 x}\right) = 1 - \frac{1}{\sin^2 x + \cos^2 x} = 1 - 1 = 0\ \ \forall\, x \in D(f).$$

Wegen Satz 7.13 ist f auf den Intervallen $I_n := \big(n\pi, (n+1)\pi\big),\ n \in \mathbf{Z}$, konstant. Zur Berechnung dieser Konstanten beachten wir:

$$\lim_{x \to n\pi \pm 0} \frac{1}{\tan x} = \pm\infty \quad \Rightarrow \quad \lim_{x \to n\pi \pm 0} \operatorname{arc\,tan}_H\left(\frac{1}{\tan x}\right) = \pm\frac{\pi}{2}.$$

Somit resultiert $f(x) = (n + \frac{1}{2})\pi\ \forall\, n \in I_n,\ n \in \mathbf{Z}$.

BSP. (7.4.5) Eine Gleichung von der Form $F(x; y, y', \ldots, y^{(n)}) = 0$, in der eine gesuchte Funktion $y = f(x)$ mit ihren Ableitungen bis zur Ordnung $n \in \mathbf{N}$ auftritt, heiße **gewöhnliche Differentialgleichung n–ter Ordnung**, kurz DGl. n–ter Ordnung. *Zum Beispiel:*

(i) Die DGl. 1.Ordnung $y' = 0$ hat gemäß Satz 7.13 als differenzierbare Lösung nur die konstante Funktion

$$\boxed{y(x) = const.}$$

(ii) Die DGl. 2.Ordnung $y'' + \lambda^2 y = 0,\ 0 \neq \lambda \in \mathbf{R}$, heißt **Schwingungsdifferentialgleichung**. Man rechnet leicht nach, dass die folgenden Funktionen Lösungen sind:

$$\boxed{y_1(x) := \cos\lambda x, \quad y_2(x) := \sin\lambda x, \quad y(x) := A\, y_1(x) + B\, y_2(x),\ A, B \in \mathbf{K}.}$$

(iii) Die DGl. 2.Ordnung $y'' - \lambda^2 y = 0,\ 0 \neq \lambda \in \mathbf{R}$, heißt **Differentialgleichung der Kettenlinie**. Man rechnet leicht nach, dass die folgenden Funktionen Lösungen sind:

$$\boxed{y_1(x) := \cosh\lambda x, \quad y_2(x) := \sinh\lambda x, \quad y(x) := A\, y_1(x) + B\, y_2(x),\ A, B \in \mathbf{K}.}$$

(iv) Die DGl. 1.Ordnung $y' + \lambda\, y = 0,\ 0 \neq \lambda \in \mathbf{K}$, hat Lösungen in der Form $y(x) = C\, e^{-\lambda x},\ C \in \mathbf{K}$. Wir zeigen, dass dies die einzigen, auf ganz $\mathbf{R}$ differenzierbaren Lösungen sind.

Satz 7.14 *Alle auf ganz* $\mathbf{R}$ *stetigen und differenzierbaren Lösungen der Differentialgleichung* $y' + \lambda\, y = 0,\ 0 \neq \lambda \in \mathbf{K}$, *haben die Form*

$$\boxed{y = f(x) := C\, e^{-\lambda x}, \quad C \in \mathbf{K}.}$$

Begründung: Man verifiziert sehr einfach, dass ein solches y tatsächlich eine Lösung ist. Ist nun $y = g(x)$ eine beliebige differenzierbare Lösung, so ist g insbesondere stetig, und wir haben $g'(x) + \lambda\, g(x) = 0\ \forall\ x \in \mathbf{R}$. Daraus folgt

$$\left(\frac{g(x)}{e^{-\lambda x}}\right)' = \frac{e^{-\lambda x} g'(x) + \lambda\, e^{-\lambda x} g(x)}{e^{-2\lambda x}} = 0\ \ \forall\, x \in \mathbf{R}.$$

Gemäß Satz 7.13 gilt dann $g(x) = C\, e^{-\lambda x}$. □

Die in den Lösungen einer Differentialgleichung auftretenden Konstanten können nicht durch die DGl. selbst festgelegt werden. Ihre Bestimmung erfordert vielmehr die Vorgabe von **Nebenbedingungen**. Interpretiert man beispielsweise die Variable x als **Zeitparameter** t, so können die frei verfügbaren Konstanten in der Lösung zum Beispiel durch Vorgabe des **Anfangszustandes** der Lösung festgelegt werden. Zu einem **Anfangszeitpunkt** $t = t_0$ (meistens $t_0 = 0$) werden die **Anfangswerte**

$$y(t_0) = y_0,\ y'(t_0) = y_1, \ldots, y^{(n-1)}(t_0) = y_{n-1}, \quad y_0, y_1, \ldots, y_{n-1} \in \mathbf{K},$$

vorgeschrieben. In diesem Fall spricht man von einer **Anfangswertaufgabe** (AWA). Die beiden folgenden Beispiele behandeln solche Anfangswertaufgaben.

(v) Das NEWTON**sche Abkühlungsgesetz**. Ein wärmeleitender Körper mit der Temperatur $u(t)$ zum Zeitpunkt t befinde sich in einem Medium mit konstanter Temperatur $A > 0$. Ein Wärmeaustausch findet in Richtung der niederen Temperatur statt, das heißt, es fließt ein Wärmestrom $\dot{u}(t)$ proportional zur Temperaturdifferenz $u(t) - A$:

$$\dot{u}(t) = -\kappa\,[u(t) - A], \quad \kappa > 0, \ \text{ Temperaturleitzahl.}$$

Zum Zeitpunkt $t_0 = 0$ habe der Körper die bekannte Temperatur $u(0) = u_0$. Zur Lösung der vorliegenden AWA setzen wir $y(t) := u(t) - A$ und erhalten die DGl. $\dot{y} + \kappa\, y = 0$, deren Lösung gemäß Satz 7.14 durch $y(t) = C\, e^{-\kappa t}$ bestimmt ist. Die Konstante C ist durch den Anfangswert $y(0) = C = u_0 - A$ eindeutig festgelegt. Demzufolge ist $u(t) = A + (u_0 - A)\, e^{-\kappa t}$, $t \geq 0$, die eindeutig bestimmte Lösung der AWA.

(vi) Der **Prozess des radioaktiven Kernzerfalls**. Dieser unterliegt derselben physikalischen Gesetzmäßigkeit wie das NEWTONsche Abkühlungsgesetz. Die Anzahl der unzerfallenen Teilchen einer radioaktiven Substanz sei zum Zeitpunkt t durch $N(t)$ gegeben. Zum Anfangszeitpunkt $t_0 = 0$ enthalte diese Substanz $N(0) = N_0$ Teilchen. Dann wird der Kernzerfall durch die folgende Anfangswertaufgabe beschrieben:

$$\dot{N}(t) = -\kappa\, N(t), \quad N(0) = N_0; \quad \kappa > 0, \ \text{ Zerfallsrate.}$$

Die Lösung $N(t) = N_0\, e^{-\kappa t}$, $t \geq 0$, folgt unmittelbar aus Satz 7.14. Man nennt den Zeitpunkt $\tau_0 > 0$, zu dem die radioaktive Substanz zur Hälfte zerfallen ist, die **Halbwertszeit** der Substanz. Das heißt, es gilt $0.5 = N(\tau_0)/N_0 = e^{-\kappa \tau_0}$, oder

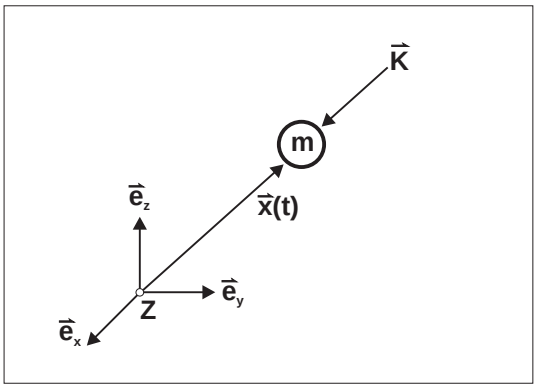
$$\tau_0 = \frac{1}{\kappa}\, \ln 2.$$

BSP. (7.4.6) Die Aussage des Satzes 7.13 darf auch komponentenweise auf **vektorwertige** Funktionen $\vec{f}\colon [a,b] \to \mathbf{K}^n$ angewendet werden:

$$\boxed{\vec{f}(x) = \vec{c} = const \ \ \forall\, x \in [a,b] \quad \Leftrightarrow \quad \vec{f}\,'(x) = \vec{0} \ \ \forall\, x \in (a,b).}$$

Wir betrachten zum Beispiel die **Bewegung einer Punktmasse** m **im Feld einer Zentralkraft**. Ein solches (zeit– und ortsabhängiges) Feld sei gegeben durch $\vec{K} := -\kappa(t, \|\vec{x}\|)\, \vec{x}(t)$. Aus dem NEWTONschen Bewegungsgesetz folgt

$$m\, \ddot{\vec{x}}(t) = \vec{K} = -\kappa(t, \|\vec{x}\|)\, \vec{x}(t).$$



Punktmasse im Feld einer Zentralkraft

Unter Verwendung des **Drehimpulses** $\vec{J} := m\, \vec{x} \times \dot{\vec{x}}$ ergibt sich dann:

$$\dot{\vec{J}}(t) = m\, \underbrace{\dot{\vec{x}}(t) \times \dot{\vec{x}}(t)}_{=\vec{0}} + \vec{x}(t) \times m\, \ddot{\vec{x}}(t) = -\kappa\, \underbrace{\vec{x}(t) \times \vec{x}(t)}_{=\vec{0}} = \vec{0}.$$

Folglich gilt $\vec{J}(t) = \vec{c} = const \in \mathbf{R}^3$. Wir können hieraus zwei Schlüsse ziehen:

(a) Es gilt $\langle \vec{J}(t), \vec{x}(t) \rangle = m\, \langle \vec{x}(t) \times \dot{\vec{x}}(t), \vec{x}(t) \rangle = m \det(\vec{x}(t), \dot{\vec{x}}(t), \vec{x}(t)) = 0 \ \forall\ t$, und somit $\vec{x}(t) \perp \vec{J}(t) = \vec{c}$. Die Bahnkurve der Punktmasse m liegt in der **Ebene** durch den Ursprung $\vec{0}$, deren Normalenvektor $\vec{c}$ ist:

$\langle \vec{x}(t), \vec{c} \rangle = 0$.

(b) Der Term $\frac{1}{2}\,\|\vec{x}(t) \times \dot{\vec{x}}(t)\| = \frac{1}{2m}\|\vec{J}(t)\|$ heißt die **Flächengeschwindigkeit** der Bahnbewegung. Aus $\vec{J}(t) = \vec{c}$ resultiert also: Unter der Wirkung einer Zentralkraft überstreicht der Strahl $\vec{x}(t)$ der Bahnkurve einer Punktmasse m in gleichen Zeiten gleiche Flächen. Dies ist das bekannte 2.KEPLERsche Gesetz der Planetenbewegung.

Zum Schluss dieses Abschnitts zeigen wir noch eine Verallgemeinerung des Mittelwertsatzes.

Satz 7.15 (Verallgemeinerter MWS der Differentialrechnung)
Die beiden reellwertigen Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ seien auf einem Intervall $[a,b] \subseteq D(f) \cap D(g)$ stetig und differenzierbar in (a,b). Gelte ferner $g'(x) \neq 0\ \forall\ x \in (a,b)$. Dann folgt:

$$\boxed{\exists\, \xi \in (a,b)\ :\ \frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(\xi)}{g'(\xi)}.} \tag{4.5}$$

Begründung: Man setze

$$\varphi(x) := [f(b)-f(a)]g(x) - [g(b)-g(a)]f(x), \quad x \in [a,b],$$

und wende auf die Funktion φ den Satz 7.10 von ROLLE an. □

Die Anwendungen des verallgemeinerten Mittelwertsatzes liegen mehr im theoretischen Bereich. Wir werden im nächsten Abschnitt bei den Regeln von L'HOSPITAL auf Beispiele stoßen.

7.5 Berechnung unbestimmter Ausdrücke: Regeln von L'Hospital

Die Berechnung des Grenzwertes $\lim\limits_{x \to x_0} \frac{f(x)}{g(x)}$ kann zu Schwierigkeiten führen, wenn beide Funktionen $f(x)$ und $g(x)$ für $x \to x_0$ gegen den Wert 0 oder gegen $\pm\infty$ streben. Wir diskutieren diesen Fall an dem folgenden

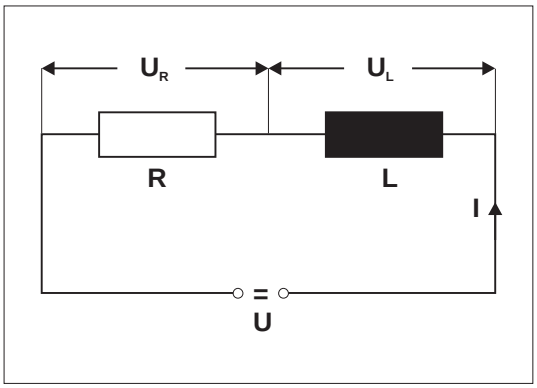
BSP. (7.5.1) **Einschaltvorgang.** Eine Induktivität L und ein OHMscher Widerstand R seien in Reihe an eine Gleichspannungsquelle U geschaltet (RL–Glied). Aus den KIRCHHOFFschen Gesetzen resultiert

$$\text{(i)}\quad U_R + U_L = U, \qquad \text{(ii)}\quad U_R = R\,I,\ \ U_L = L\frac{dI}{dt}.$$

Durch Einsetzen von (ii) in (i) erhält man die DGl. 1.Ordnung $L\frac{dI}{dt} + R\,I = U,\ t > 0$, für die zeitliche Veränderung der Stromstärke $I(t)$, wenn die Spannung U zum Zeitpunkt $t = 0$ an das RL–Glied angelegt wurde. Man darf von der empirisch gewonnenen **stationären Bedingung** $\lim\limits_{t \to +\infty} I(t) = U/R$ ausgehen. Wie man sich leicht überzeugt, wird diese Aufgabe von der folgenden Funktion gelöst:

$$I(t) = \frac{U}{R}\left(1 - e^{-\frac{R}{L}\,t}\right), \quad t \geq 0.$$

Will man hier für $L \neq 0$ den Grenzwert $\lim\limits_{R \to 0+} I(t)$ bilden, so erhält man formal einen unbestimmten Ausdruck von der Form $\frac{0}{0}$.



Einschaltvorgang bei einem RL–Glied

Auf einen solchen unbestimmten Ausdruck wird man auch in den folgenden Fällen geführt:

$$\frac{\sin x}{x} \text{ für } x \to 0, \quad (x - \frac{\pi}{2}) \tan x = \frac{\tan x}{(x - \frac{\pi}{2})^{-1}} \text{ für } x \to \frac{\pi}{2},$$

usw. Aber auch Fälle wie x^x für $x \to 0+$, oder $x^{1/x}$ für $x \to +\infty$ sowie $x^{1/x}$ für $x \to 0+$ sind interessant. Man hat allgemein die folgenden Typen von **unbestimmten Ausdrücken** vorliegen:

$$\boxed{\frac{0}{0}, \quad \frac{\infty}{\infty}, \quad \infty - \infty, \quad 0^0, \quad 1^\infty, \quad \infty^0.}$$

Nur die beiden ersten Fälle bedürfen einer mathematischen Analyse; die anderen Fälle lassen sich auf diese zurückführen.

Satz 7.16 (Regel von L'Hospital **für $\frac{0}{0}$)**
Die reellen Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ seien differenzierbar im gelochten Intervall $(a, b) \setminus \{x_0\}$, und es gelte

$$\lim_{x \to x_0} f(x) = 0 = \lim_{x \to x_0} g(x), \quad g'(x) \neq 0 \ \ \forall\, x \in (a, b) \setminus \{x_0\}.$$

Existiert der Grenzwert $c := \lim\limits_{x \to x_0} f'(x)/g'(x)$, so gilt für den unbestimmten Ausdruck

$$\boxed{\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} = c.} \tag{5.1}$$

Es kann auch $x_0 = a$ oder $x_0 = b$ gelten (einseitige Grenzwerte). Ferner sind die Fälle $x_0 = a = -\infty$ oder $x_0 = b = +\infty$ zugelassen, ebenso $c = \pm\infty$ (uneigentliche Grenzwerte).

Begründung: Es gelte zuerst $x_0 \in \mathbf{R}$. Dann sind die Funktionen f und g im Punkt x_0 stetig ergänzbar durch $f(x_0) = 0 = g(x_0)$. Aus dem verallgemeinerten Mittelwertsatz (Satz 7.15) folgt die Existenz eines Zwischenwertes $\xi := x_0 + \theta\,(x - x_0)$, $\theta \in (0, 1)$, mit:

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \frac{f'(\xi)}{g'(\xi)}.$$

Nun gilt offenkundig $\xi \to x_0$ für $x \to x_0$, so dass bereits die behauptete Relation (5.1) folgt.

Es gelte jetzt entweder $x_0 = -\infty$ oder $x_0 = +\infty$. Wir setzen $x := 1/y$ und haben nun die einseitigen Grenzwerte $y \to 0-$ bzw. $y \to 0+$ zu betrachten. Wegen

$$f'(x) = \frac{d}{dy} f\Big(\frac{1}{y}\Big) \cdot \frac{dy}{dx} = -y^2 \frac{d}{dy} f\Big(\frac{1}{y}\Big)$$

gilt hier

$$c = \lim_{x\to\pm\infty} \frac{f'(x)}{g'(x)} = \lim_{y\to 0\pm} \frac{\frac{d}{dy} f(\frac{1}{y})}{\frac{d}{dy} g(\frac{1}{y})},$$

und dieser Fall kann wie der erste behandelt werden. □

BSP. (7.5.2) (vgl. BSP(7.5.1)). Wir hatten für die zeitliche Änderung der Stromstärke $I(t)$ beim Anlegen der Spannung U an ein RL–Glied die Relation $I(t) = \frac{U}{R}\,(1 - e^{-Rt/L})$ begründet. Aus Satz 7.16 erschließen wir nun:

$$\lim_{R\to 0+} I(t) = \lim_{R\to 0+} \frac{(Ut/L)\cdot e^{-Rt/L}}{1} = \frac{U}{L}\,t,\ t \geq 0.$$

BSP. (7.5.3) Die folgenden Grenzwerte wurden bereits in Abschnitt 6.7, Formel (7.6) mit anderen Methoden berechnet. Wir verwenden hier die Regel von L'Hospital:

(i) $\lim\limits_{x\to 0} \dfrac{\sin x}{x} = \lim\limits_{x\to 0} \dfrac{\cos x}{1} = 1,$

(ii) $\lim\limits_{x\to 0} \dfrac{1-\cos x}{x} = \lim\limits_{x\to 0} \dfrac{\sin x}{1} = 0,$

(iii) $\lim\limits_{x\to 0} \dfrac{2(1-\cos x)}{x^2} = \lim\limits_{x\to 0} \dfrac{2\sin x}{2x} = 1.$

BSP. (7.5.4) Falls bei der Anwendung der Regel von L'Hospital auch der Grenzwert $\lim\limits_{x\to x_0} f'(x)/g'(x)$ auf einen unbestimmten Ausdruck $\frac{0}{0}$ führt, so lässt sich die Regel von L'Hospital ein weiteres Mal anwenden, sofern der Quotient $f'(x)/g'(x)$ die erforderlichen Voraussetzungen erfüllt. Dieser Prozess kann solange wiederholt werden, bis ein definiter Grenzwert $\lim\limits_{x\to x_0} f^{(n)}(x)/g^{(n)}(x)$ auftritt. *Zahlenbeispiele:*

(i) $\lim\limits_{x\to 0} \dfrac{x-\sin x}{x\,\sin x} \overset{\text{1.L'Hosp.}}{=} \lim\limits_{x\to 0} \dfrac{1-\cos x}{x\,\cos x + \sin x} \overset{\text{2.L'Hosp.}}{=} \lim\limits_{x\to 0} \dfrac{\sin x}{2\,\cos x - x\,\sin x} = 0,$

(ii) $\lim\limits_{x\to 0} \dfrac{1-\cos x}{1-\cos 2x} \overset{\text{1.L'Hosp.}}{=} \lim\limits_{x\to 0} \dfrac{\sin x}{2\,\sin 2x} \overset{\text{2.L'Hosp.}}{=} \lim\limits_{x\to 0} \dfrac{\cos x}{4\,\cos 2x} = \dfrac{1}{4}.$

Ganz analog zu Satz 7.16 zeigt man für unbestimmte Ausdrücke $\frac{\infty}{\infty}$:

Satz 7.17 (Regel von L'Hospital für $\frac{\infty}{\infty}$)
Die reellen Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ seien differenzierbar im gelochten Intervall $(a,b)\setminus\{x_0\}$, und es gelte

$$\lim_{x\to x_0} \frac{1}{f(x)} = 0 = \lim_{x\to x_0} \frac{1}{g(x)}, \quad g'(x) \neq 0 \ \ \forall\, x \in (a,b)\setminus\{x_0\}.$$

Existiert der Grenzwert $c := \lim\limits_{x\to x_0} f'(x)/g'(x)$, so gilt für den unbestimmten Ausdruck

$$\boxed{\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)} = c.} \tag{5.2}$$

Es kann auch $x_0 = a$ oder $x_0 = b$ gelten (einseitige Grenzwerte). Ferner sind die Fälle $x_0 = a = -\infty$ oder $x_0 = b = +\infty$ zugelassen, ebenso $c = \pm\infty$ (uneigentliche Grenzwerte).

BSP. (7.5.5) Wir verschaffen uns mit Hilfe von Satz 7.17 einen Überblick über das Wachstumsverhalten der **Exponentialfunktion** im Unendlichen. Für beliebige Zahlen $\alpha > 0$, $\beta > 0$ gilt:

$$\lim_{x\to+\infty} \frac{e^{\alpha x}}{x} \overset{\text{L'Hosp.}}{=} \lim_{x\to+\infty} \frac{\alpha\, e^{\alpha x}}{1} = +\infty, \quad \lim_{x\to+\infty} \frac{e^{\alpha x}}{x^\beta} = \lim_{x\to+\infty} \left[\frac{e^{\alpha x/\beta}}{x}\right]^\beta = \left[\lim_{x\to+\infty} \frac{e^{\alpha x/\beta}}{x}\right]^\beta = +\infty.$$

Jede noch so kleine Potenz von e^x wächst im Unendlichen schneller als jede noch so große Potenz von x. Hieraus folgert man insbesondere:

$$\boxed{\lim_{x\to+\infty} P_n(x)\, e^{-\alpha x} = 0 \ \ \forall\, \alpha > 0 \ \text{ und jedes Polynom } P_n(x).}$$

BSP. (7.5.6) Satz 7.17 liefert uns auch einen Überblick über das Wachstumsverhalten des **Logarithmus** im Unendlichen. Für beliebige Zahlen $\alpha > 0$, $\beta > 0$ gilt:

$$\lim_{x\to+\infty} \frac{\ln x}{x^\alpha} \overset{\text{L'Hosp.}}{=} \lim_{x\to+\infty} \frac{1}{\alpha\, x^\alpha} = 0, \qquad \lim_{x\to+\infty} \frac{(\ln x)^\beta}{x^\alpha} = \lim_{x\to+\infty} \left[\frac{\ln x}{x^{\alpha/\beta}}\right]^\beta = \left[\lim_{x\to+\infty} \frac{\ln x}{x^{\alpha/\beta}}\right]^\beta = 0.$$

Jede noch so große Potenz von $\ln x$ wächst im Unendlichen schwächer als jede noch so kleine Potenz von x.

$\boxed{0\cdot\infty:}$ Dieser Fall entspricht einem Grenzwert $\lim\limits_{x\to x_0} f(x)g(x)$ mit $\lim\limits_{x\to x_0} f(x) = 0$ und $\lim\limits_{x\to x_0} g(x) = \infty$. Man betrachte statt dessen einen der beiden folgenden unbestimmten Ausdrücke:

$$\boxed{\lim_{x\to x_0} \frac{f(x)}{1/g(x)} = \frac{0}{0} \quad \text{oder} \quad \lim_{x\to x_0} \frac{g(x)}{1/f(x)} = \frac{\infty}{\infty}.}$$

BSP. (7.5.7) Die folgenden Grenzwerte sind vom Typ " $0\cdot\infty$ ".

(i) Es seien wiederum $\alpha > 0, \beta > 0$ beliebige Zahlen.

$$\lim_{x\to0+} x^\alpha \ln x = \lim_{x\to0+} \frac{\ln x}{x^{-\alpha}} \overset{\text{L'Hosp.}}{=} \lim_{x\to0+} \frac{1}{-\alpha\, x^{-\alpha}} = 0,$$

$$\lim_{x\to0+} x^\alpha (\ln x)^\beta = \lim_{x\to0+} \left[x^{\alpha/\beta} \ln x\right]^\beta = \left[\lim_{x\to0+} x^{\alpha/\beta} \ln x\right]^\beta = 0.$$

(ii) $$\lim_{x\to+\infty} x \ln\left(1+\frac{1}{x}\right) = \lim_{x\to+\infty} \frac{\ln(1+\frac{1}{x})}{x^{-1}} \overset{\text{L'Hosp.}}{=} \lim_{x\to+\infty} \frac{(1+\frac{1}{x})^{-1}(-x^{-2})}{(-x^{-2})} = 1.$$

(iii) $$\lim_{x\to\frac{\pi}{2}\pm0} \left(x-\frac{\pi}{2}\right)\tan x = \lim_{x\to\frac{\pi}{2}\pm0} \frac{(x-\frac{\pi}{2})}{\cot x} \overset{\text{L'Hosp.}}{=} \lim_{x\to\frac{\pi}{2}\pm0} \frac{1}{-1/\sin^2 x} = -1.$$

$\boxed{\infty-\infty:}$ Dieser Fall entspricht einem Grenzwert $\lim\limits_{x\to x_0} \Big(f(x)-g(x)\Big)$ mit $\lim\limits_{x\to x_0} f(x) = +\infty = \lim\limits_{x\to x_0} g(x)$. Man betrachte statt dessen den unbestimmten Ausdruck

$$\boxed{\lim_{x\to x_0} f(x)\left[1-\frac{g(x)}{f(x)}\right],}$$

wobei **zwei Fälle** zu unterscheiden sind:

(i) $$\lim_{x\to x_0} \frac{g(x)}{f(x)} \neq 1 \quad\Rightarrow\quad \lim_{x\to x_0} \big(f(x)-g(x)\big) = \pm\infty.$$

(ii) $$\lim_{x\to x_0} \frac{g(x)}{f(x)} = 1 \quad\Rightarrow\quad \lim_{x\to x_0} \big(f(x)-g(x)\big) \text{ hat die Form " } 0\cdot\infty \text{ ".}$$

Fallbeispiel zu (i):

$$\lim_{x\to0+} \left[\frac{1}{x} - \frac{2}{\ln(1+x)}\right] = \lim_{x\to0+} \frac{1}{x}\Big[1 - \underbrace{\frac{2x}{\ln(1+x)}}_{\to2}\Big] = -\infty.$$

Fallbeispiel zu (ii):

$$\begin{aligned}\lim_{x\to0+} \left[\frac{1}{x} - \frac{1}{\ln(1+x)}\right] &= \lim_{x\to0+} \frac{\ln(1+x)-x}{x\,\ln(1+x)} \overset{1.\text{L'Hosp.}}{=} \lim_{x\to0+} \frac{1/(1+x)-1}{x/(1+x)+\ln(1+x)} \\ &= \lim_{x\to0+} \frac{-x}{x+(1+x)\,\ln(1+x)} \overset{2.\text{L'Hosp.}}{=} \lim_{x\to0+} \frac{-1}{1+\ln(1+x)+1} = -\frac{1}{2}.\end{aligned}$$

$\boxed{0^0:}$ Dieser Fall entspricht $\lim\limits_{x\to x_0} f(x)^{g(x)} = \lim\limits_{x\to x_0} e^{g(x)\,\ln[f(x)]} = e^{0\cdot\infty}$.

$\boxed{\infty^0:}$ Dieser Fall entspricht $\lim\limits_{x\to x_0} f(x)^{g(x)} = \lim\limits_{x\to x_0} e^{g(x)\,\ln[f(x)]} = e^{0\cdot\infty}$.

$\boxed{1^\infty:}$ Dieser Fall entspricht $\lim\limits_{x\to x_0} f(x)^{g(x)} = \lim\limits_{x\to x_0} e^{g(x)\,\ln[f(x)]} = e^{\infty\cdot 0}$.

In allen drei Fällen berechnet man also *zuerst* den Logarithmus der Grenzwerte und exponiert danach das Ergebnis:

$$\lim_{x\to x_0} f(x)^{g(x)} =: e^G \quad\Leftrightarrow\quad G := \lim_{x\to x_0} g(x)\,\ln[f(x)].$$

Fallbeispiel zu " 0^0 ":

$$\lim_{x\to+\infty}\Big(\frac{1}{x}\Big)^{3/x^2} =: e^G \quad\Leftrightarrow\quad G := \lim_{x\to+\infty}\Big(-\frac{3\,\ln x}{x^2}\Big) \overset{\text{BSP. (7.5.6)}}{=} 0 \quad\Leftrightarrow\quad e^G = 1.$$

Fallbeispiel zu " ∞^0 ":

$$\lim_{x\to\pi+0}\Big(2+3\,e^{-1/\sin x}\Big)^{x-\pi} =: e^G$$

$$\Leftrightarrow\quad G := \lim_{x\to\pi+0}(x-\pi)\ln\Big(2+3\,e^{-1/\sin x}\Big) = \lim_{x\to\pi+0}(x-\pi)\Big[-\frac{1}{\sin x} + \underbrace{\ln\Big(2\,e^{1/\sin x}+3\Big)}_{\to\ln 3}\Big]$$

$$= -\lim_{x\to\pi+0}\frac{x-\pi}{\sin x} \overset{\text{L'Hosp.}}{=} -\lim_{x\to\pi+0}\frac{1}{\cos x} = 1$$

$$\Leftrightarrow\quad e^G = e^1 = e.$$

Fallbeispiel zu " 1^∞ ":

$$\lim_{x\to1} x^{1/(x-1)} =: e^G \quad\Leftrightarrow\quad G := \lim_{x\to1}\frac{\ln x}{x-1} \overset{\text{L'Hosp.}}{=} \lim_{x\to1}\frac{1}{x} = 1 \quad\Leftrightarrow\quad e^G = e^1 = e.$$

BSP. (7.5.8) Die Regeln von L'HOSPITAL können versagen, wenn die Voraussetzungen zu den Sätzen 7.16 und 7.17 nicht strikt beachtet werden. Ein typisches Beispiel ist der Grenzwert

$$\lim_{x\to+\infty}\frac{f(x)}{g(x)} := \lim_{x\to+\infty}\frac{e^x-e^{-x}}{e^x+e^{-x}} = \frac{\infty}{\infty}.$$

Durch Ableiten von Zähler und Nenner erhält man die sich alternierend reproduzierenden Quotienten

$$\frac{e^x+e^{-x}}{e^x-e^{-x}},\quad \frac{e^x-e^{-x}}{e^x+e^{-x}},\ldots.$$

Es existiert für keine der Ableitungen $f^{(n)}(x)/g^{(n)}(x)$ ein Grenzwert. Hingegen resultiert nach Division durch e^{-x} sofort der korrekte Grenzwert

$$\lim_{x\to+\infty}\frac{e^x-e^{-x}}{e^x+e^{-x}} = \lim_{x\to+\infty}\frac{1-e^{-2x}}{1+e^{-2x}} = 1.$$

Eine ähnliche Situation liegt bei dem unbestimmten Ausdruck

$$\lim_{x\to+\infty}\frac{f(x)}{g(x)} := \lim_{x\to+\infty}\frac{x+\sin x}{x-\sin x} = \frac{\infty}{\infty}$$

vor. Die Regel von L'HOSPITAL darf hier **nicht** angewendet werden, denn der Grenzwert

$$\lim_{x\to+\infty}\frac{1+\cos x}{1-\cos x} = \lim_{x\to+\infty}\frac{f'(x)}{g'(x)}$$

existiert nicht. Es wäre ganz falsch, hier nochmals die Regel von L'HOSPITAL anwenden zu wollen. Wie man sofort verifiziert, würde dies auf einen Grenzwert -1 führen. Hingegen resultiert nach Division durch x sofort der korrekte Grenzwert

$$\lim_{x\to+\infty}\frac{x+\sin x}{x-\sin x} = \lim_{x\to+\infty}\frac{1+\frac{\sin x}{x}}{1-\frac{\sin x}{x}} = 1.$$

7.6 Der Satz von Taylor

In Abschnitt 7.1 wurde das Leibnizsche Tangentenproblem mit Hilfe der Ableitung $f'(x_0)$ einer Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ gelöst. Zur Erinnerung, es war diejenige affine Funktion $T(x) := ax + b$ zu bestimmen, die die Funktion $f(x)$ in einem Punkt $x_0 \in D(f)$ **mindestens von der Ordnung** 1 **berührt**. Darunter verstehen wir die folgende

Definition 7.8 *Zwei Funktionen $f, g \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ berühren sich im Punkt $x_0 \in D(f) \cap D(g)$* **mindestens von der Ordnung** *$n \in \mathbf{N}$, wenn für $x \in D(f) \cap D(g)$ gilt:*

$$\boxed{f(x) = g(x) + R_n(x;x_0) \quad \textit{und} \quad \lim_{x\to x_0} \frac{R_n(x;x_0)}{(x-x_0)^n} = 0.} \tag{6.1}$$

Ist die Funktion f im Punkt $x_0 \in D(f)$ differenzierbar, so berühren sich also $f(x)$ und die Tangente

$$T_1(x) := f(x_0) + f'(x_0)\,(x - x_0), \quad x \in \mathbf{R}, \tag{6.2}$$

in x_0 mindestens von der Ordnung 1. Die folgende Fragestellung ist eine naheliegende Verallgemeinerung des Leibnizschen Tangentenproblems:

> Man bestimme zu einer gegebenen Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ und zu einem Punkt $x_0 \in D(f)$ ein Polynom $T_n(x)$ vom Grad höchstens $n \in \mathbf{N}$ derart, dass sich f und T_n in x_0 mindestens von der Ordnung n berühren.

Im Fall $n = 1$ wird diese Aufgabe durch die Tangente $T_1(x)$ aus (6.2) gelöst, sofern die Funktion f im Punkt x_0 differenzierbar ist. Für den allgemeinen Fall zeigen wir in einem ersten Schritt, dass es zum obigen Problem höchstens eine Lösung $T_n(x)$ geben kann.

Satz 7.18 *Es gibt höchstens ein Polynom $T_n(x)$ vom Grad $\le n \in \mathbf{N}$, welches eine gegebene Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ in einem festen Punkt $x_0 \in D(f)$ mindestens von der Ordnung n berührt.*

Begründung: Wären $T_n(x) := \sum_{k=0}^{n} a_k (x-x_0)^k$ und $P_n(x) := \sum_{k=0}^{n} b_k (x-x_0)^k$ zwei solche Polynome, so hätten wir

$$f(x) = T_n(x) + R_n(x;x_0), \quad \lim_{x\to x_0} \frac{R_n(x;x_0)}{(x-x_0)^n} = 0,$$

$$f(x) = P_n(x) + Q_n(x;x_0), \quad \lim_{x\to x_0} \frac{Q_n(x;x_0)}{(x-x_0)^n} = 0.$$

Setzen wir $L_n(x;x_0) := R_n(x;x_0) - Q_n(x;x_0)$, so gilt dann offenbar

$$0 = \sum_{k=0}^{n} (a_k - b_k)(x-x_0)^k + L_n(x;x_0), \quad \lim_{x\to x_0} \frac{L_n(x;x_0)}{(x-x_0)^n} = 0.$$

Wäre $0 \le j \le n$ der kleinste Index mit $a_j \ne b_j$, so erhielten wir nach Division der obigen Gleichung durch $(x-x_0)^j$ und Grenzwertbildung $x \to x_0$:

$$0 = a_j - b_j + \lim_{x\to x_0} \left[\frac{L_n(x;x_0)}{(x-x_0)^n} \cdot (x-x_0)^{n-j} \right] = a_j - b_j.$$

Also muss $a_j = b_j \;\forall\, 0 \le j \le n$ gelten. □

Dass ein solches Polynom $T_n(x)$ wirklich existiert, wurde bereits von Brook Taylor (1685–1735) nachgewiesen:

Satz 7.19 (von der TAYLORschen Formel)
Für die gegebene Funktion $f \in \mathrm{Abb}(\mathbf{R},\mathbf{R})$ gebe es ein Intervall $[a,b] \subseteq D(f)$ derart, dass $f \in C^n([a,b])$ gilt. Dann existiert in jedem Punkt $x_0 \in (a,b)$ genau ein Polynom $T_n(x)$ vom Grad höchstens $n \in \mathbf{N}$ mit der Eigenschaft

$$\boxed{f(x) = T_n(x) + R_n(x;x_0) \;\; \forall\, x \in [a,b], \quad \lim_{x \to x_0} \frac{R_n(x;x_0)}{(x-x_0)^n} = 0. \quad \text{TAYLOR–Formel}} \tag{6.3}$$

Dieses ist das TAYLOR**–Polynom n–ten Grades der Funktion $f(x)$ im Entwicklungspunkt** *x_0 mit der Darstellung*

$$\boxed{T_n(x) = \sum_{k=0}^{n} \frac{1}{k!}\, f^{(k)}(x_0) \cdot (x-x_0)^k, \quad x \in \mathbf{R}.} \tag{6.4}$$

Existiert in jedem Punkt $x_0 \in (a,b)$ darüber hinaus noch die $(n+1)$–te Ableitung $f^{(n+1)}(x_0)$, so hat das **Restglied** *die Darstellung*

$$\boxed{R_n(x;x_0) = \frac{(x-x_0)^{n+1}}{(n+1)!}\, f^{(n+1)}(\xi), \quad \xi := x_0 + \theta\,(x-x_0) \;\; \textit{für ein } \theta \in (0,1).} \tag{6.5}$$

Das Restglied in der Form (6.5) *heiße* LAGRANGE**sches Restglied der** TAYLOR**–Formel.**

Begründung: Wir betrachten für festes $x_0 \in (a,b)$ und für $t \in [a,b]$ die Hilfsfunktion

$$g(t) := f(x) - \sum_{k=0}^{n-1} \frac{1}{k!}\, f^{(k)}(t)\,(x-t)^k, \;\; x \in [a,b] \;\text{ fest, }\; x \neq x_0,$$

und wir setzen

$$G(t) := g(t) - g(x_0)\,\left(\frac{x-t}{x-x_0}\right)^n.$$

Die Funktion G ist in (a,b) differenzierbar, und es gilt $g(x) = 0$, $G(x) = 0 = G(x_0)$. In dieser Situation trifft der Satz von ROLLE (Satz 7.10) zu. Es gibt eine Zwischenstelle $\xi = x_0 + \theta\,(x-x_0)$, $\theta \in (0,1)$, mit

$$0 = G'(\xi) = g'(\xi) + g(x_0)\,\frac{n(x-\xi)^{n-1}}{(x-x_0)^n}.$$

Man berechnet leicht

$$g'(\xi) = -\frac{(x-\xi)^{n-1}}{(n-1)!}\, f^{(n)}(\xi), \;\; \text{und somit } \;\; g(x_0) = \frac{(x-x_0)^n}{n!}\, f^{(n)}(\xi).$$

Wird dieser Ausdruck in die Definition der Funktion g eingesetzt, so resultiert nun

$$f(x) = \sum_{k=0}^{n-1} \frac{1}{k!}\, f^{(k)}(x_0)\,(x-x_0)^k + \frac{(x-x_0)^n}{n!}\; f^{(n)}(\xi). \tag{6.6}$$

Nach Voraussetzung ist $f^{(n)}(x)$ stetig im Punkt $x = x_0$, und somit gilt sicher

$$f^{(n)}(\xi) = f^{(n)}(x_0) + L(\xi;x_0) \;\; \text{mit} \;\; \lim_{\xi \to x_0} L(\xi;x_0) = 0.$$

Setzt man hier $R_n(x;x_0) := L(\xi;x_0)(x-x_0)^n/n!$ so ergeben sich bereits die Darstellungen (6.3) und (6.4). Schließlich ist das LAGRANGEsche Restglied (6.5) direkt an (6.6) ablesbar, wenn man die höhere Differenzierbarkeitsvoraussetzung beachtet. □

Bemerkung 7.13 (a) Neben dem LAGRANGEschen Restglied (6.5) gibt es noch andere Darstellungsformen des Restgliedes (zum Beispiel von CAUCHY, SCHLÖMILCH, Integralrestglied). Wir werden dieses Thema noch einmal in Abschnitt 8.4 behandeln.

(b) Der Satz von der TAYLORschen Formel sagt nichts darüber aus, *wie* man die Zwischenstelle ξ findet. Man kann aber wie beim Mittelwertsatz die Restgliedformel (6.5) zur Fehlerabschätzung verwenden, wenn man für den Betrag der Ableitung $|f^{(n+1)}(\xi)|$ eine obere Schranke kennt. □

BSP. (7.6.1) Wir bestimmen die TAYLOR–Polynome der Funktionen $f_1(x) := e^x, f_2(x) := \sin x, f_3(x) := \cos x$, $x \in D(f_i) := \mathbf{R}$, im Entwicklungspunkt $x_0 := 0$. Offensichtlich gelten folgende Relationen:

(i) $f_1^{(k)}(x) = e^x \Rightarrow f_1^{(k)}(0) = 1 \ \forall\, k \in \mathbf{N}_0$,

(ii) $f_2^{(2k)}(x) = (-1)^k \sin x \Rightarrow f_2^{(2k)}(0) = 0, \quad f_2^{(2k+1)}(x) = (-1)^k \cos x \Rightarrow f_2^{(2k+1)}(0) = (-1)^k \ \forall\, k \in \mathbf{N}_0$,

(iii) $f_3^{(2k)}(x) = (-1)^k \cos x \Rightarrow f_3^{(2k)}(0) = (-1)^k, \quad f_3^{(2k+1)}(x) = (-1)^{k+1} \sin x \Rightarrow f_3^{(2k+1)}(0) = 0 \ \forall\, k \in \mathbf{N}_0$.

Somit erhalten wir aus (6.4) und (6.5):

$$\boxed{\begin{aligned} e^x &= \sum_{k=0}^{n} \frac{x^k}{k!} + \frac{x^{n+1}}{(n+1)!} e^{\theta x} \ \text{ mit } \ 0 < \theta < 1, \\ \sin x &= \sum_{k=0}^{n} \frac{(-1)^k x^{2k+1}}{(2k+1)!} + \frac{(-1)^{n+1} x^{2n+3}}{(2n+3)!} \cos\theta x \ \text{ mit } \ 0 < \theta < 1, \\ \cos x &= \sum_{k=0}^{n} \frac{(-1)^k x^{2k}}{(2k)!} + \frac{(-1)^{n+1} x^{2n+2}}{(2n+2)!} \cos\theta x \ \text{ mit } \ 0 < \theta < 1. \end{aligned}}$$

Wir haben hier im Fall $f_1(x)$ das TAYLOR–Polynom vom Grade n berechnet. In den Fällen $f_2(x)$ und $f_3(x)$ sind es dagegen die TAYLOR–Polynome vom Grade $2n+1$. Wegen $|\cos\theta x| \le 1$ hat man in diesen beiden Fällen eine Fehlerabschätzung:

$$\Big|\sin x - \sum_{k=0}^{n} \frac{(-1)^k x^{2k+1}}{(2k+1)!}\Big| \le \frac{|x|^{2n+3}}{(2n+3)!}, \quad \Big|\cos x - \sum_{k=0}^{n} \frac{(-1)^k x^{2k}}{(2k)!}\Big| \le \frac{|x|^{2n+2}}{(2n+2)!}.$$

BSP. (7.6.2) Wir bestimmen das TAYLOR–Polynom vom Grad $n \in \mathbf{N}$ der Funktion $f(x) := \ln x$, $x \in D(f) := (0, +\infty)$, im Entwicklungspunkt $x_0 := 1$. Offensichtlich gelten folgende Relationen:

$$f'(x) = \frac{1}{x}, \ f''(x) = \frac{-1!}{x^2}, \ f'''(x) = \frac{(-1)^2 2!}{x^3}, \ldots, f^{(k)}(x) = \frac{(-1)^{k-1}(k-1)!}{x^k}$$

$$\Rightarrow \quad f^{(0)}(1) = 0, \ f^{(k)}(1) = (-1)^{k-1}(k-1)! \ \forall\, k \in \mathbf{N}.$$

Somit erhalten wie aus (6.4) und (6.5) für eine Zwischenstelle $\xi := 1 + \theta\,(x-1)$, $0 < \theta < 1$:

$$\boxed{\ln x = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} (x-1)^k + \frac{(-1)^n (x-1)^{n+1}}{(n+1)\xi^{n+1}}.}$$

Aufgabe: Wie groß muss $n \in \mathbf{N}$ höchstens gewählt werden, damit die Funktion $f(x) := \ln x$ im Bereich $|x-1| \le 0.5$ durch das TAYLOR–Polynom $T_n(x)$ im Entwicklungspunkt $x_0 := 1$ mit einer Genauigkeit von $\epsilon := 10^{-4}$ approximiert wird? *Lösung:* Wir brauchen n nur so groß zu machen, dass $|R_n(x;1)| \le \epsilon$ für $0.5 \le x \le 1.5$ gilt. Wir haben hier:

$$|R_n(x;1)| = \frac{|x-1|^{n+1}}{(n+1)|\xi|^{n+1}} \le \frac{(0.5)^{n+1}}{(n+1)(0.5)^{n+1}} = \frac{1}{n+1} \overset{!}{\le} 10^{-4},$$

und diese Ungleichung ist sicher für $n = 10^4 - 1 = 9999$ erfüllt.

BSP. (7.6.3) Es sei $f(x) := P_n(x)$ ein Polynom vom Grad $n \in \mathbf{N}$. Dann gilt nach Satz 7.18 notwendig $P_n(x) = T_n(x)$. In Definition 2.11, Abschnitt 2.4 hatten wir bereits erklärt, was unter der TAYLOR–Entwicklung des Polynoms $P_n(x)$ an der Stelle x_0 zu verstehen ist. Diese TAYLOR–Entwicklung ist identisch mit dem TAYLOR–Polynom $T_n(x)$ von Grad n im Entwicklungspunkt x_0, denn aus der Darstellung (4.9) in Abschnitt 2.4, nämlich aus

$$\begin{aligned} P_n(x) = \; & [\cdots[[[(x-x_0)a_n + P_1(x_0)](x-x_0) + P_2(x_0)](x-x_0) + P_3(x_0)](x-x_0) + \cdots \\ & + P_{n-1}(x_0)](x-x_0) + P_n(x_0) \\ = \; & \sum_{k=0}^{n} d_k (x-x_0)^k \;\text{ mit }\; d_k := P_{n-k}(x_0) \;\text{ und }\; d_n := P_0(x_0) = a_n, \end{aligned}$$

erhält man sofort die Beziehung $P_n^{(k)}(x_0) = k! P_{n-k}(x_0)$. Nun gilt gemäß Satz 7.19

$$P_n(x) = T_n(x) = \sum_{k=0}^{n} \frac{1}{k!} P_n^{(k)}(x_0)\,(x-x_0)^k = \sum_{k=0}^{n} P_{n-k}(x_0)\,(x-x_0)^k \;\; \forall\, x \in \mathbf{R}.$$

Somit haben wir den Zusammenhang (4.10) aus Abschnitt 2.4 zwischen den Koeffizienten des TAYLOR–Polynoms im Entwicklungspunkt x_0 und dem vollständigen HORNER–Schema hergestellt. *Zahlenbeispiel:* Man bestimme die TAYLOR–Entwicklung des Polynoms $P_4(x) := 4x^4 - 5x^3 + 6x - 30$ an der Stelle $x_0 = 2$. Wir berechnen das vollständige HORNER–Schema:

	4	−5	0	6	−30	
2	∗	8	6	12	36	
	4	3	6	18	**6**	$= P_4(2)$
2	∗	8	22	56		
	4	11	28	**74**		$= \frac{1}{1!} \cdot P_4'(2)$
2	∗	8	38			
	4	19	**66**			$= \frac{1}{2!} \cdot P_4''(2)$
2	∗	8				
	4	**27**				$= \frac{1}{3!} \cdot P_4'''(2)$
	∗					
	4					$= \frac{1}{4!} \cdot P_4^{iv}(2)$

Aus den eingerahmten Koeffizienten ergibt sich die TAYLOR–Entwicklung des Polynoms $P_4(x)$ an der Stelle $x_0 = 2$ in der Form:

$$P_4(x) = T_4(x) = 4(x-2)^4 + 27(x-2)^3 + 66(x-2)^2 + 74(x-2) + 6.$$

Das obige BSP. (7.6.3) wirft die generelle Frage auf, für welche Funktionen $f(x)$ das TAYLOR–Polynom $T_n(x)$ aus (6.4) im Grenzwert $n \to +\infty$ gegen die Funktion $f(x)$ konvergiert. Wegen $|f(x) - T_n(x)| = |R_n(x; x_0)|$ liegt die Antwort schon auf der Hand:

Satz 7.20 *Gegeben seien eine Funktion* $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ *und ein Intervall* $[a,b] \subseteq D(f)$, *so dass* $f \in C^\infty([a,b])$ *gilt. Genau dann gestattet die Funktion* $f(x)$ **an der Stelle** $x_0 \in (a,b)$ **die** TAYLOR**–Entwicklung**

$$\boxed{f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)\,(x-x_0)^k \;\; \forall\, x \in [a,b], \quad \text{TAYLOR–\textbf{Reihe}}} \tag{6.7}$$

wenn gilt:

$$\lim_{n\to\infty} R_n(x; x_0) = \lim_{n\to\infty} \frac{(x-x_0)^{n+1}}{(n+1)!} f^{(n+1)}(\xi) = 0 \;\; \forall\, x \in [a,b].$$

BSP. (7.6.4) Wie wir in Abschnitt 3.1, BSP. (3.1.10) gezeigt haben, gilt ja

$$\lim_{n\to\infty} \frac{|x|^n}{n!} = 0$$

für jedes feste $x \in \mathbf{R}$. Für die in BSP. (7.6.1) angegebenen Restglieder folgt hieraus jeweils $\lim\limits_{n\to\infty} |R_n(x;0)| = 0$ bei festgehaltenem $x \in \mathbf{R}$. Somit resultieren die folgenden TAYLOR–Reihen:

$$\boxed{e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad \sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}, \quad \cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} \ \ \forall\, x \in \mathbf{R}.}$$

BSP. (7.6.5) Für das Restglied des Logarithmus in BSP. (7.6.2) gilt die folgende Abschätzung:

$$|R_n(x;1)| = \frac{|x-1|^{n+1}}{(n+1)|\xi|^{n+1}} \le \begin{cases} \dfrac{(\frac{1}{x}-1)^{n+1}}{n+1} = \dfrac{e^{(n+1)\,\ln(1-x)/x}}{n+1} & : \ 0 < x < 1, \\ \dfrac{(x-1)^{n+1}}{n+1} = \dfrac{e^{(n+1)\,\ln(x-1)}}{n+1} & : \ x \ge 1. \end{cases}$$

Es gilt $\ln(1-x)/x \le 0$ für $0.5 \le x \le 1$ sowie $\ln(x-1) \le 0$ für $1 \le x \le 2$. Deshalb erhalten wir $\lim\limits_{n\to\infty} |R_n(x;1)| = 0\ \forall\, x \in [0.5, 2]$, und somit:

$$\boxed{\ln x = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}\,(x-1)^k \ \ \forall\, x \in (0,2].}$$

Hierbei haben wir die Tatsache benutzt, dass das Quotientenkriterium die absolute Konvergenz auch noch im Intervall $0 < x < 0.5$ sicherstellt. Setzt man insbesondere $x = 2$ ein, so erhält man nun den Summenwert der **alternierenden harmonischen** Reihe

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

BSP. (7.6.6) Wir berechnen die TAYLOR–Reihe der Funktion $f(x) := (1+x)^\alpha$, $\alpha \in \mathbf{R}$, an der Stelle $x_0 = 0$. Sofern $x \ne -1$ gilt, haben wir zunächst $f'(x) = \alpha(1+x)^{\alpha-1}$, $f''(x) = \alpha(\alpha-1)(1+x)^{\alpha-2}, \ldots$, allgemein $f^{(k)}(x) = \alpha(\alpha-1)\cdots(\alpha-k+1)(1+x)^{\alpha-k}$, und somit $\frac{1}{k!}\,f^{(k)}(0) = \binom{\alpha}{k}\ \forall\, k \in \mathbf{N}_0$. Die TAYLOR–Formel aus Satz 7.19 hat hier die Form

$$(1+x)^\alpha = \sum_{k=0}^{n} \binom{\alpha}{k} x^k + \binom{\alpha}{n+1} (1+\theta x)^{\alpha-n-1} x^{n+1}, \ \ 0 < \theta < 1.$$

Wir verschaffen uns eine grobe Restgliedabschätzung und verwenden dazu

$$\left|\binom{\alpha}{n+1}\right| = \left|\frac{\alpha(\alpha-1)\cdots(\alpha-n)}{(n+1)!}\right| = \left|\left(\frac{\alpha+1}{1}-1\right)\left(\frac{\alpha+1}{2}-1\right)\cdots\left(\frac{\alpha+1}{n+1}-1\right)\right| \le (1+|\alpha+1|)^{n+1}.$$

Hieraus erhalten wir

$$|R_n(x;0)| \le |1+\theta x|^\alpha \left|\frac{x(1+|\alpha+1|)}{1+\theta x}\right|^{n+1}.$$

Sofern $|x(1+|\alpha+1|)| < |1+\theta x|$ gilt, folgt aus dieser Abschätzung $\lim\limits_{n\to\infty} |R_n(x;0)| = 0$. Wir hatten aber bereits in Abschnitt 3.2, BSP. (3.2.6), den Bereich der absoluten Konvergenz mit $|x| < 1$ bestimmt. Also gilt hier die TAYLOR–Reihe

$$\boxed{(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \ \ \forall\, |x| < 1.}$$

BSP. (7.6.7) Die Funktion $F : \mathbf{R} \to \mathbf{R}$ mit

$$\boxed{F(x) := \frac{1}{x^2}\,(x^3 + \cos^2 x - 1), \quad x \neq 0,}$$

ist in $x_0 = 0$ stetig ergänzbar.

(a) Man bestimme $F(0)$ und verwende dazu die bekannte TAYLOR–Reihe für $\cos 2x = 2\cos^2 x - 1$.

(b) Man bestimme das TAYLOR–Polynom $T_2(x)$ 2–ten Grades der Funktion $F(x)$ im Entwicklungspunkt $x_0 = 0$. Man schätze den Fehler $|F(x) - T_2(x)|$ im Bereich $|x| \leq 0.5$ ab.

Lösung: (a) Es gilt ja

$$\cos 2x = \sum_{k=0}^{\infty} \frac{(-1)^k (2x)^{2k}}{(2k)!} = 1 - 2x^2 + \frac{2}{3}\,x^4 + \sum_{k=3}^{\infty} \frac{(-1)^k (2x)^{2k}}{(2k)!},$$

und aus dieser Relation ergibt sich sofort

$$F(x) = \frac{1}{x^2}\left(x^3 + \frac{1}{2}\,\cos 2x - \frac{1}{2}\right) = -1 + x + \frac{1}{3}\,x^2 + 2\sum_{k=3}^{\infty} \frac{(-1)^k\,(2x)^{2(k-1)}}{(2k)!}. \tag{6.8}$$

Hier erkennt man sofort den gesuchten Funktionswert $F(0) = -1$.

(b) Aus der Darstellung (6.8) erhalten wir ebenfalls:

$$T_2(x) = -1 + x + \frac{1}{3}\,x^2.$$

Es wäre falsch, die gesuchte Fehlerabschätzung über das LAGRANGEsche Restglied bestimmen zu wollen. Dazu müßten wir die Ableitung $F'''(x)$ berechnen. Dieser Aufwand ist völlig überflüssig, denn aus (6.8) ergibt sich unmittelbar:

$$|F(x) - T_2(x)| = \left|2\sum_{k=3}^{\infty} \frac{(-1)^k (2x)^{2(k-1)}}{(2k)!}\right| =: \left|\sum_{k=3}^{\infty} (-1)^k a_k\right|.$$

Die Reihenglieder $a_k := 2\,(2x)^{2(k-1)}/(2k)!$ sind positiv, und sie bilden für $|x| \leq 0.5$ eine monoton fallende Folge. Also lässt sich das LEIBNIZ–Kriterium anwenden:

$$|F(x) - T_2(x)| \leq a_3 = \frac{2}{6!}\,(2x)^4 \leq \frac{2}{6!} = \frac{1}{360} \quad \forall\, |x| \leq 0.5.$$

7.7 Extremwerte, Kurvendiskussion

Man verschafft sich am einfachsten einen Überblick über den Funktionsverlauf einer gegebenen Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ durch Zeichnen des Graphen $G(f)$. Hier sind insbesondere Kenntnisse über charakteristische Punkte der Funktion f von Wichtigkeit. Ist die Funktion f *differenzierbar*, so gibt der Satz 7.9 Auskunft über die Lage relativer Extrema. Allerdings genügt nach dem Extremalsatz 6.16 allein die *Stetigkeit* der Funktion f auf einem abgeschlossenen Intervall $[a, b]$, um dort die Existenz von Extremalwerten sicherzustellen.

Relative Extrema einer Funktion f suche man daher stets

(i) unter den Nullstellen der Ableitung f' (notwendig für *innere* Extrema, sofern f differenzierbar ist),

(ii) in den Randpunkten des Definitionsbereichs $D(f)$,

(iii) in Punkten $x = x_0$, in denen $f(x)$ nicht differenzierbar ist.

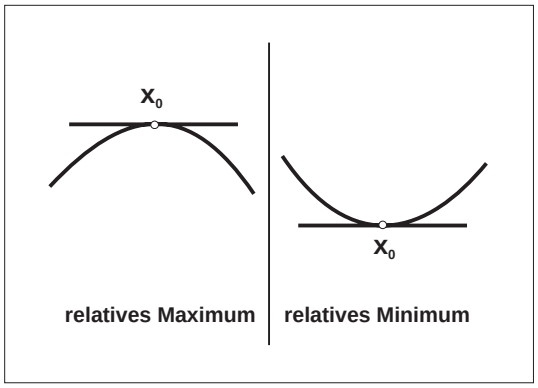
Zur weiteren Analyse des Graphen einer hinreichend oft stetig differenzierbaren Funktion definiert man:

Definition 7.9 *Ein innerer Punkt $x_0 \in D(f)$ einer gegebenen Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R})$ heiße*

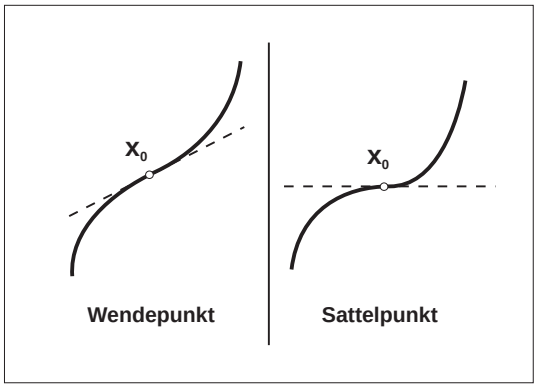
(i) **Flachpunkt** *von f, wenn $f'(x_0) = 0$ gilt,*

(ii) **Wendepunkt** *von f, wenn f' in x_0 ein relatives Extremum hat (notwendig dafür ist die Bedingung $f''(x_0) = 0$), das heißt, wenn die zweite Ableitung f'' in x_0 das Vorzeichen wechselt,*

(iii) **Sattelpunkt** *von f, wenn x_0 sowohl Wende– als auch Flachpunkt ist.*



Flachpunkte von f



Wendepunkte von f

Der folgende Satz liefert eine analytische Entscheidungshilfe bei der Diskussion der oben definierten Ausnahmepunkte.

Satz 7.21 *Gegeben sei eine Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R})$, und für $[a, b] \subseteq D(f)$ gelte $f \in C^n([a, b])$. Es gelte ferner in einem Punkt $x_0 \in (a, b)$*

$$\boxed{f'(x_0) = 0 = f''(x_0) = \cdots = f^{(n-1)}(x_0), \quad \textit{aber } f^{(n)}(x_0) \neq 0.} \tag{7.1}$$

Ist n eine **gerade Zahl**, *so liegt in x_0 ein relatives* **Extremum** *vor, und zwar für*

$$\boxed{f^{(n)}(x_0) > 0 \ \textit{ ein relatives } \textbf{Minimum}, \quad f^{(n)}(x_0) < 0 \ \textit{ ein relatives } \textbf{Maximum}.}$$

Ist n eine **ungerade Zahl**, *so liegt in x_0 ein* **Sattelpunkt** *vor.*

Begründung: Man erhält aus der TAYLORschen Formel

$$f(x) = f(x_0) + (x - x_0)^n \left[\frac{f^{(n)}(x_0)}{n!} + \frac{R_n(x; x_0)}{(x - x_0)^n}\right].$$

Da der letzte Summand im Grenzwert $x \to x_0$ verschwindet, hat der Klammerausdruck $[\cdots]$ für alle x in der Nähe von x_0 das Vorzeichen von $f^{(n)}(x_0)$. Ist n eine *gerade Zahl*, so gilt nun:

$f(x) \geq f(x_0)$, sofern $f^{(n)}(x_0) > 0$ ist; das heißt, in x_0 liegt ein *Minimum*,

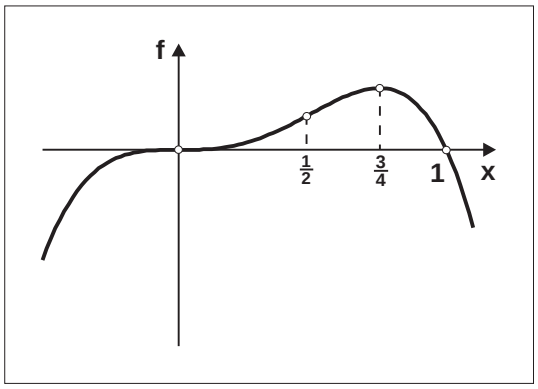
$f(x) \leq f(x_0)$, sofern $f^{(n)}(x_0) < 0$ ist; das heißt, in x_0 liegt ein *Maximum*.

Ist n eine *ungerade Zahl*, so wechselt der Term $(x - x_0)^n$ in x_0 das Vorzeichen, und dies ist typisch für die Existenz eines *Sattelpunktes* in x_0. □

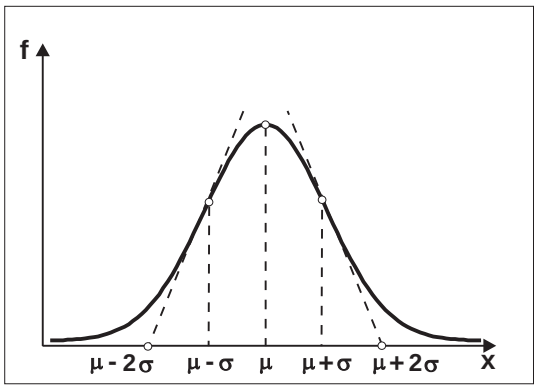
BSP. (7.7.1) Für die Funktion $f(x) := x^3 - x^4 = x^3(1 - x)$, $x \in D(f) := \mathbf{R}$, berechnet man sehr einfach

$$f'(x) = x^2(3 - 4x),\ f''(x) = 6x(1 - 2x),\ f'''(x) = 6(1 - 4x),\ f^{(4)}(x) = -24.$$

Man erkennt somit, dass **Nullstellen** von f bei $x_0 = 0$ und $x_0 = 1$ liegen. **Flachpunkte** der Funktion f liegen in $x_0 = 0$ und $x_0 = \frac{3}{4}$ vor. Wegen $f''(0) = 0$ und $f'''(0) = 6$ ist der Punkt $x_0 = 0$ ein **Sattelpunkt**. Hingegen liegt im Punkt $x_0 = \frac{3}{4}$ wegen $f''(\frac{3}{4}) = -\frac{9}{4} < 0$ ein **Maximum** der Funktion f. Die Nullstellen von f'' sind die beiden Punkte $x_0 = 0$ und $x_0 = \frac{1}{2}$. Da die Funktion f'' in beiden Punkten das Vorzeichen wechselt, liegen hier **Wendepunkte** vor.



Der Graph der Funktion $f(x) := x^3 - x^4$



Die GAUSS**–Normalverteilung**

BSP. (7.7.2) Die in der Stochastik wichtige GAUSS**sche Normalverteilung** ist die Funktion

$$\boxed{f(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \quad x \in D(f) := \mathbf{R}.}$$

Die Parameter $\sigma > 0$ und $\mu > 0$ heißen **Streuung** bzw. **Mittelwert** der Normalverteilung. Offenbar gilt $f(x) > 0$, $\lim\limits_{x\to\pm\infty} f(x) = 0$. Zur Ermittlung relativer Extrema berechnen wir

$$\begin{aligned}
f'(x) &= -\frac{x-\mu}{\sigma^3\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) = -\frac{x-\mu}{\sigma^2} f(x), \\
f''(x) &= -\frac{1}{\sigma^3\sqrt{2\pi}} \Big[1 - \frac{(x-\mu)^2}{\sigma^2}\Big] \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) = -\frac{1}{\sigma^2}\Big[1 - \frac{(x-\mu)^2}{\sigma^2}\Big] f(x), \\
f'''(x) &= \frac{x-\mu}{\sigma^5\sqrt{2\pi}} \Big[3 - \frac{(x-\mu)^2}{\sigma^2}\Big] \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) = -\frac{x-\mu}{\sigma^4}\Big[3 - \frac{(x-\mu)^2}{\sigma^2}\Big] f(x).
\end{aligned}$$

Wegen $f > 0$ hat die Gleichung $f'(x) = 0$ genau eine Lösung $x_0 = \mu$. Es gilt

$$f''(\mu) = -\frac{1}{\sigma^2} f(\mu) < 0,$$

so dass in diesem Punkt ein **Maximum** liegt. Die Gleichung $f''(x) = 0$ hat die zwei Lösungen $x_\pm = \mu \pm \sigma$. Wir berechnen $f'''(x_+) = -f'''(x_-) = \frac{2}{\sigma^3} f(x_\pm) \neq 0$, so dass die Punkte $x_\pm$ **Wendepunkte** sind. Die Tangenten in den Wendepunkten heiße **Wendetangenten**, das sind hier die beiden affinen Funktionen

$$T_1(x) := f(x_\pm)\Big(2 - \frac{x-\mu}{\sigma}\Big), \quad T_2(x) := f(x_\pm)\Big(2 + \frac{x-\mu}{\sigma}\Big), \quad x \in \mathbf{R},$$

mit den Nullstellen $x_1 := \mu + 2\sigma$ bzw. $x_2 := \mu - 2\sigma$.

Bemerkung 7.14 Die in Satz 7.21 aufgestellten Kriterien über die Existenz von Ausnahmepunkten verlieren ihre Gültigkeit, wenn f nicht mehr die erforderliche Differenzierbarkeitsstufe hat. Andererseits gibt es sogar Funktionen $f \in C^\infty(\mathbf{R})$, bei denen die Kriterien aus Satz 7.21 ebenfalls versagen. Zu diesem Typ von Funktionen gehört zum Beispiel

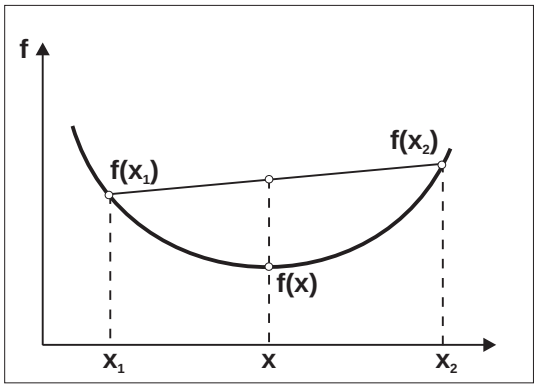
$$f(x) := \begin{cases} 0 & : \ x = 0, \\ \exp(-\frac{1}{x^2}) & : \ x \neq 0. \end{cases}$$

Diese Funktion hat die Eigenschaft $f^{(k)}(0) = 0 \;\forall\; k \in \mathbf{N}_0$. Obwohl $f(x)$ im Punkt $x_0 := 0$ ein absolutes Minimum hat, greift hier der Satz 7.21 nicht. Darüber hinaus verschwindet das TAYLOR–Polynom n–ten Grades im Entwicklungspunkt x_0 stets identisch: $T_n(x) = 0 \;\forall\; x \in \mathbf{R} \;\forall\; n \in \mathbf{N}$. Wir haben deshalb $|f(x) - T_n(x)| = |f(x)| > 0$ für $x \neq 0$, so dass das Restglied im Limes $n \to \infty$ außerhalb der Stelle $x_0 = 0$ nicht verschwindet. □

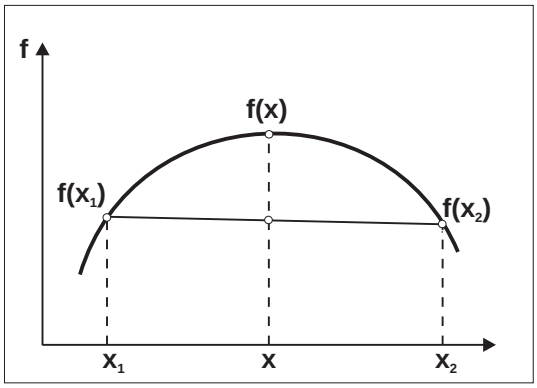
Geometrische Bedeutung von f''. Das Vorzeichen der ersten Ableitung f' einer reellwertigen Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ gibt Auskunft darüber, ob der Graph $G(f)$ steigt oder fällt. Dies haben wir in Satz 7.12 begründet. Will man noch eine zusätzliche Information darüber, ob sich der Graph $G(f)$ nach unten oder nach oben krümmt, so muss man das Vorzeichen der zweiten Ableitung f'' analysieren.

Definition 7.10 *Eine reellwertige Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ heiße* **konvex** *auf einem Intervall $[a, b] \subseteq D(f)$, wenn der Graph $G(f)$ stets* **unterhalb** *der Verbindungsgeraden zwischen zwei seiner Punkte $(x_1, f(x_1))$, $(x_2, f(x_2))$, $x_j \in [a, b]$, liegt. Liegt der Graph* **oberhalb** *der Verbindungsgeraden, so heiße die Funktion f* **konkav**. *Das heißt, für jedes Punktepaar $x_1, x_2 \in [a, b]$ gilt:*

$$f\Big((1-t)x_1 + tx_2\Big) \begin{cases} \leq (1-t)\, f(x_1) + t\, f(x_2) \;\forall\; t \in (0,1) & \Rightarrow \quad f \textbf{ konvex}, \\ \geq (1-t)\, f(x_1) + t\, f(x_2) \;\forall\; t \in (0,1) & \Rightarrow \quad f \textbf{ konkav}. \end{cases} \tag{7.2}$$



Der Graph einer konvexen Funktion



Der Graph einer konkaven Funktion

Satz 7.22 (a) *Eine reellwertige Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ ist auf einem Intervall $[a, b] \subseteq D(f)$ genau dann konvex, wenn die Funktion $-f$ dort konkav ist.*

(b) *Gilt $f \in C\Big([a, b]\Big) \cap C^2\Big((a, b)\Big)$, so ist f auf dem Intervall $[a, b]$ genau dann konvex (bzw. konkav), wenn $f''(x) \geq 0$ (bzw. $f''(x) \leq 0$) $\forall\; x \in (a, b)$ gilt.*

Begründungen: (a) Diese Behauptung folgt direkt aus der Definition (7.2) durch Multiplikation mit (-1).

(b) Wir setzen $x := (1-t)x_1 + tx_2 = x_1 + t(x_2 - x_1) = x_2 + (1-t)(x_1 - x_2)$. Für $x_1 < x < x_2$ existieren nach dem Mittelwertsatz Zwischenstellen ξ_1, ξ_2 mit $x_1 < \xi_1 < x < \xi_2 < x_2$, so dass gilt:

$$f(x) - f(x_1) = f'(\xi_1)\,(x - x_1), \quad f(x_2) - f(x) = f'(\xi_2)\,(x_2 - x).$$

(i) Sei zunächst angenommen, dass die Funktion f konvex ist. Dann folgt unter Verwendung von (7.2):

$$f'(\xi_1) = \frac{f(x) - f(x_1)}{x - x_1} \leq \frac{t[f(x_2) - f(x_1)]}{t(x_2 - x_1)} \leq \frac{(1-t)[f(x_2) - f(x)]}{(1-t)(x_2 - x)} = f'(\xi_2).$$

Also ist f' auf dem Intervall $[a,b]$ monoton wachsend, und dies ist genau dann der Fall, wenn $f''(x) \geq 0 \; \forall \, x \in (a,b)$ gilt.

(ii) Gelte umgekehrt $f''(x) \geq 0 \; \forall \, x \in (a,b)$, so ist f' auf $[a,b]$ monoton wachsend, und wir erschließen:

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

Wegen $x - x_1 = t(x_2 - x_1)$ und $x_2 - x = (1-t)(x_2 - x_1)$ ergibt sich die Bedingung (7.2). □

BSP. (7.7.3) Die Funktion $f(x) := \sin\sqrt{x}$, $x \in D(f) := [0, +\infty)$, hat die Ableitungen

$$f'(x) = \frac{\cos\sqrt{x}}{2\sqrt{x}}, \; x > 0, \quad f''(x) = -\frac{\sqrt{x}\sin\sqrt{x} + \cos\sqrt{x}}{4x\sqrt{x}}, \; x > 0.$$

Es gilt $f''(x) \leq 0$ im Intervall $0 < x \leq x_0$, worin $x_0 := \xi^2$ durch die Lösung $\xi \in (0, \pi)$ der transzendenten Gleichung $\xi = -\cot\xi$ eindeutig bestimmt ist. Der Wert von ξ kann nur numerisch berechnet werden, vgl. Abschnitt 7.8. Die Funktion $f(x)$ ist also auf dem Intervall $[0, x_0]$ *konkav*. In gleicher Weise können weitere Konkavitäts– und Konvexitätsintervalle gefunden werden.

Die Diskussion des Graphen $G(f)$ einer reellwertigen Funktion f, besonders hinsichtlich der Lage von Nullstellen, Extremwerten, Wendepunkten und der Asymptoten, nennt man **Kurvendiskussion**. Es empfiehlt sich, bei Kurvendiskussionen systematisch vorzugehen, etwa nach folgenden Gesichtspunkten.

- Man bestimme den maximalen Definitionsbereich der Funktion f. Man prüfe, ob die Funktion f Symmetrien aufweist.
- Man bestimme die Nullstellen von f, f' und f''.
- Man grenze mit diesen Nullstellen diejenigen Bereiche ab, in denen f positiv bzw. negativ ist, monoton wachsend bzw. monoton fallend, konvex bzw. konkav.
- Man bestimme die relativen Extrema von f und diskutiere, ob Maxima, Minima oder Sattelpunkte vorliegen. Man bestimme die relativen Extrema von f' (Wendepunkte). In den Wendepunkten durchsetzt die Tangente den Graphen $G(f)$ (Wendetangente). Anschaulich ändert sich der Drehsinn der Tangente.
- Ist a ein Randpunkt von $D(f)$, der eventuell nicht zu $D(f)$ gehört, so bestimme man die Funktionenlimites $\lim\limits_{x\to a} f(x)$ und $\lim\limits_{x\to a} f'(x)$.

BSP. (7.7.4) Wir diskutieren den Graphen der rationalen Funktion

$$f(x) := \frac{x^2-1}{x^2+x-2} =: \frac{P(x)}{Q(x)}, \quad x \in D(f) := \{x \in \mathbf{R} \, : \, Q(x) \neq 0\}.$$

Die Nullstellen von $Q(x) = x^2 + x - 2 = (x-1)(x+2)$ sind offenbar $x_0 = 1$ und $x_0 = -2$. Wegen $P(x) = x^2 - 1 = (x-1)(x+1)$ ist $f(x)$ im Punkt $x_0 = 1$ stetig ergänzbar zu

$$f(1) = \lim_{x\to 1} \frac{P(x)}{Q(x)} = \lim_{x\to 1} \frac{2x}{2x+1} = \frac{2}{3},$$

wobei wir die Regel von L'HOSPITAL angewendet haben. Hieraus ergibt sich der maximale Definitionsbereich $D_{\max}(f) = \mathbf{R} \setminus \{-2\}$ sowie

$$f(x) = \begin{cases} \dfrac{x+1}{x+2} & : \; x \in D_{\max}(f) \setminus \{1\}, \\ \dfrac{2}{3} & : \; x = 1. \end{cases}$$

Man erkennt sofort $f(x_0) = 0$ genau für $x_0 = -1$ sowie

$$f'(x) = \frac{(x+2)-(x+1)}{(x+2)^2} = \frac{1}{(x+2)^2} > 0 \ \forall\ x \in D_{\max}(f).$$

Somit ist f in den Intervallen $(-\infty, -2)$ und $(-2, +\infty)$ streng monoton wachsend, während im Punkt $x_p = -2$ ein Pol 1.Ordnung mit Vorzeichenwechsel vorliegt. Weiterhin erhält man:

$$f''(x) = \frac{-2}{(x+2)^3} \begin{cases} > 0 \ \forall\ x \in (-\infty, -2) \ \Rightarrow\ f \text{ ist hier konvex,} \\ < 0 \ \forall\ x \in (-2, +\infty) \ \Rightarrow\ f \text{ ist hier konkav.} \end{cases}$$

Es gilt ferner

$$f(x) \begin{cases} > 0 \ :\ x < -2, \\ < 0 \ :\ -2 < x < -1, \\ > 0 \ :\ -1 < x < +\infty. \end{cases}$$

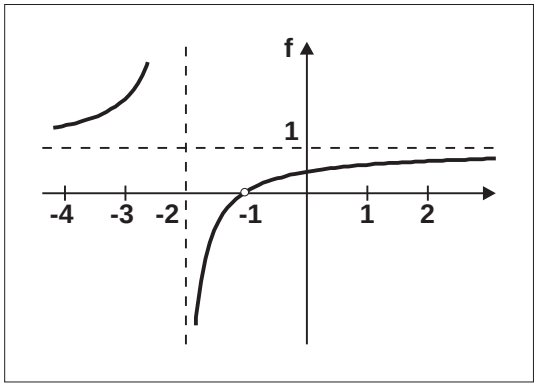
Mit den Asymptoten

$$\lim_{x \to \pm\infty} f(x) = \lim_{x \to \pm\infty} \frac{1 + 1/x}{1 + 2/x} = 1, \quad \lim_{x \to -2 \pm 0} f(x) = \mp\infty$$

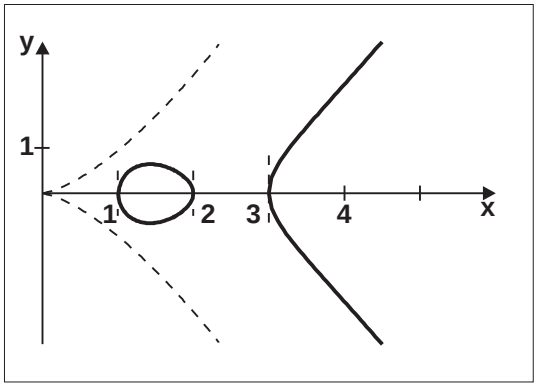
und einer Wertetabelle

x	-4	-3	0	1	2
$f(x)$	$\frac{3}{2}$	2	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$

lässt sich nun der Graph $G(f)$ sehr präzise skizzieren.



Der Graph von $f(x) := \frac{x^2-1}{x^2+x-2}$



Der Graph von $y^2 = x^3 - 6x^2 + 11x - 6$

BSP. (7.7.5) Wir betrachten die durch die Gleichung

$$\boxed{y^2 = x^3 - 6x^2 + 11x - 6 = (x-1)(x-2)(x-3)}$$

definierte algebraische Kurve. Diese zerfällt in die zwei Kurvenäste

$$f_\pm(x) = \pm\sqrt{(x-1)(x-2)(x-3)},$$

die spiegelsymmetrisch zur x–Achse liegen. Es genügt deshalb, nur den Graph $G(f_+)$ zu diskutieren.

- Maximaler Definitionsbereich: $D_{\max}(f_+) = [1,2] \cup [3, +\infty)$. Die Funktion f_+ ist auf $D_{\max}(f_+)$ stetig.
- Nullstellen: Es gilt $f_+(x_0) = 0$ für $x_0 = 1,\ x_0 = 2,\ x_0 = 3$.
- Relative Extrema: Man berechnet aus

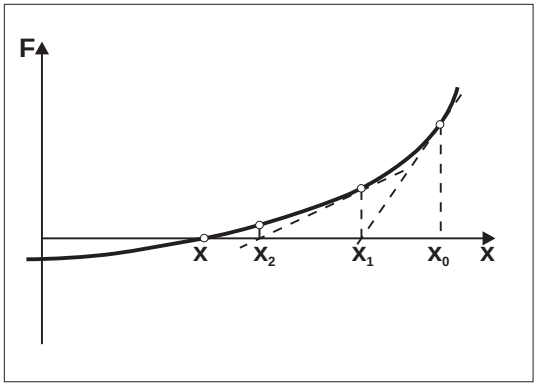
$$f'_+(x) = \frac{3x^2 - 12x + 11}{2\sqrt{(x-1)(x-2)(x-3)}}$$

die Nullstelle $f'_+(x_h) = 0 \Leftrightarrow x_h = 2 - \frac{1}{3}\sqrt{3} \in D_{\max}(f_+)$. Also ist der Punkt $(x_h, y_h := \frac{1}{3}\sqrt{2\sqrt{3}})$ der geometrische Ort einer *horizontalen* Tangente.

- Vertikale Tangenten: Es gilt $|f'_+(x_v)| = +\infty$ in den Nullstellen von f_+, so dass diese gleichzeitig geometrischer Ort vertikaler Tangenten sind.
- Asymptote: Für $x \gg 1$ kann $y^2 \sim x^3$ gesetzt werden, und dies führt auf $f_\pm(x) \sim \pm x^{3/2}$. Man nennt diese Kurve NEILLsche Parabel. Sie ist in der obige Skizze als strichpunktierte Linie eingetragen.
- Wendepunkte: Wegen
$$f''_+(x) = \frac{15x^4 - 24x^3 + 66x^2 - 72x + 23}{4[(x-1)(x-2)(x-3)]^{3/2}}$$
können Nullstellen von f''_+ nicht mehr elementar bestimmt werden.

7.8 Das Newton–Verfahren; Fixpunktsätze

Wir wenden uns hier nochmals einem Problem zu, mit dem wir uns bereits in Abschnitt 6.7 befasst haben, nämlich dem Problem der Lösung einer nichtlinearen Gleichung vom Typ $F(x) = 0$, wobei die gegebene reellwertige Funktion $F \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ mindestens *stetig* sei.



Zum NEWTON–Verfahren

Ein numerisches Verfahren zur *näherungsweisen* Bestimmung der Lösung $x \in D(F)$ ist das NEWTON–Verfahren. Dieses geht von der Annahme aus, dass die Funktion $F(x)$ für alle in Frage kommenden Werte $x \in D(F)$ *stetig differenzierbar* sei und dass die Ableitung F' einfach berechnet werden kann. Anders als die Regula falsi und das Sekantenverfahren verwendet das NEWTON–Verfahren nicht die *interpolierende Gerade* zwischen zwei aufeinanderfolgenden Stützpunkten $(x_j, F(x_j))$, $j = n-1, n$, als Näherung der Funktion F, sondern die **Tangente** im Punkt $(x_n, F(x_n))$, siehe obige Skizze.

Aus der Tangentengleichung $T(x) = F(x_n) + F'(x_n)\,(x - x_n)$ bestimmt man die Nullstelle $T(x_{n+1}) = 0 = F(x_n) + F'(x_n)\,(x_{n+1} - x_n)$, und man betrachtet x_{n+1} als Verbesserung der Näherung x_n für die gesuchte Lösung x. Ausgehend von einem passenden Startwert $x_0 \in D(F)$, gelangt man so zur Iterationsvorschrift

$$\boxed{x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}, \quad n \in \mathbf{N}_0, \;\; x_0 \in D(F) \;\; \text{geeignet.}} \tag{8.1}$$

Es ist die Frage zu beantworten, unter welchen Bedingungen die NEWTON–Folge (8.1) gegen eine Lösung x der Gleichung $F(x) = 0$ konvergiert. Wir werden eine Antwort mit Hilfe des BANACHschen Fixpunktsatzes (Satz 6.25) geben. Dazu muss die Iterationsvorschrift (8.1) in die Form (7.28), Abschnitt 6.7, nämlich $x_{n+1} = f(x_n)$, $n \in \mathbf{N}_0$, gebracht werden. Dies gelingt sehr einfach, indem wir

$$f(x) := x - \frac{F(x)}{F'(x)}, \quad x \in D(F) \setminus \{x \,:\, F'(x) = 0\} \tag{8.2}$$

setzen. Es sind nun die Voraussetzungen zum BANACHschen Fixpunktsatz zu verifizieren:

Satz 7.23 (NEWTON–Verfahren)
(a) *Die Funktion* $F \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ *erfülle auf dem Intervall* $[a,b] \subseteq D(F)$ *die Bedingungen*

$$\boxed{\text{(N1)} \quad F(a)\cdot F(b) < 0, \qquad \text{(N2)} \quad F \in C^2([a,b]), \quad F'(x) \neq 0 \ \ \forall\, x \in [a,b].}$$

Dann hat die Gleichung $F(x) = 0$ *genau eine Lösung* $x \in (a,b)$.

(b) *Unter den Bedingungen* (N1) *und* (N2) *seien für einen Startwert* $x_0 \in (a,b)$ *Zahlen* $r > 0$ *und* $0 < q < 1$ *derart bestimmt, dass mit* $I := [x_0 - r, x_0 + r] \subseteq [a,b]$ *gilt:*

$$\boxed{\text{(N3)} \quad \left|\frac{F(x)\cdot F''(x)}{F'^2(x)}\right| \leq q \ \ \forall\, x \in I, \qquad \text{(N4)} \quad \left|\frac{F(x_0)}{F'(x_0)}\right| \leq (1-q)\,r.}$$

Dann konvergiert die NEWTON*–Folge* (8.1) *mit dem Startwert* x_0 *gegen die Lösung* $x \in (a,b)$, *und es gilt die* **a posteriori–Fehlerabschätzung**

$$\boxed{|x_n - x| \leq \frac{|F(x_n)|}{m}, \quad m := \min_{t\in I} |F'(t)|.} \tag{8.3}$$

(c) *Gilt zusätzlich* $F \in C^3([a,b])$, *so konvergiert die* NEWTON*–Folge* (8.1) *unter den Bedingungen* (N1)–(N4) *sogar* **quadratisch**. *Das heißt, mit einer von* n *unabhängigen Konstanten* K *gilt*

$$\boxed{|x_{n+1} - x| \leq K\,(x_n - x)^2 \ \ \forall\, n \in \mathbf{N}_0.} \tag{8.4}$$

Begründungen: (a) Diese Behauptung folgt aus dem Nullstellensatz von BOLZANO (Satz 6.18) und der Monotonie der Funktion F (es gilt ja $F'(x) \neq 0$).

(b) Die Funktion f aus (8.2) ist stetig differenzierbar. Wegen (N3) gilt

$$|f'(x)| = \left|\frac{F(x)\cdot F''(x)}{F'^2(x)}\right| \leq q \ \ \forall\, x \in I,$$

so dass f auf dem Intervall I kontrahierend ist. Darüber hinaus haben wir wegen (N4):

$$|f(x_0) - x_0| = \left|\frac{F(x_0)}{F'(x_0)}\right| \leq (1-q)\,r < r, \tag{8.5}$$

und somit $x_1 := f(x_0) \in I$. Wir zeigen nun durch vollständige Induktion die Eigenschaft $x_n \in I\ \forall\, n \in \mathbf{N}$. Für $n = 1$ haben wir dies soeben bewiesen.
Vererbung: Mit der Ungleichung (8.5) folgern wir aus der Induktionsannahme $|x_n - x_0| \leq r$:

$$|x_{n+1} - x_0| = |f(x_n) - x_0| \leq |f(x_n) - f(x_0)| + |f(x_0) - x_0| \leq q|x_n - x_0| + (1-q)r \leq r.$$

Nun erschließen wir aus dem Fixpunktsatz 6.25 von BANACH die Behauptung (b). Die Fehlerabschätzung (8.3) resultiert aus dem Mittelwertsatz. Für eine Zwischenstelle $\xi := x_n + \theta(x - x_n),\ \theta \in (0,1)$, gilt nämlich

$$|F(x_n)| = |F(x_n) - F(x)| = |F'(\xi)|\,|x_n - x| \geq m\,|x_n - x|.$$

(c) Mit zusätzlicher Regularität $F \in C^3([a,b])$ erhalten wir aus der TAYLORschen Formel:

$$f(t) - f(x) = f'(x)\,(t-x) + \frac{1}{2!}\,f''[x + \theta(t-x)]\,(t-x)^2,\ \theta \in (0,1).$$

Setzt man hier $t := x_n$ und verwendet $f(x_n) = x_{n+1}$ sowie $f(x) = x$ und

$$f'(x) = \frac{F(x)F''(x)}{F'^2(x)} = 0,$$

so resultiert die Ungleichung (8.4), wenn wir $K := \frac{1}{2} \max\limits_{t \in I} |f''(t)|$ definieren. □

Bemerkung 7.15 (a) Neben der **a posteriori**–Fehlerabschätzung (8.3) hat man gemäß Satz 6.25 auch die **a priori**–Fehlerabschätzung

$$\boxed{|x_n - x| \leq \frac{q^n}{1-q} |x_1 - x_0| \quad \forall\, n \in \mathbf{N}.} \tag{8.6}$$

(b) Das in Abschnitt 3.1 behandelte HERON–Verfahren (babylonisches Wurzelziehen) ist nichts anderes als das NEWTON–Verfahren für die Gleichung $F(x) := x^2 - a,\ a > 0$. □

Algorithmus des NEWTON–Verfahrens.

```
1:   Einlesen von x_0, ε;   x := x_0; y := F(x);
2:   wiederhole:
3:       x := x − y/F'(x);                                   (8.7)
4:       y := F(x);
5:   bis |y| < ε.
```

Die Iterationsvorschrift (8.1) lässt sich wieder sehr einfach algorithmisch fassen, wobei die Vorgabe eines *Abbruchkriteriums* zweckmäßig ist. Die Iteration wird solange wiederholt, bis zu vorgegebener Toleranz $\epsilon > 0$ der Wert $|F(x_n)| < \epsilon$ erreicht wird. Nach Beendigung der Iteration gibt die Variable x die gesuchte Näherungslösung an.

Kritik: Wegen der quadratischen Konvergenz ist das NEWTON–Verfahren sehr beliebt bei den Anwendern. Allerdings ist es oft schwierig, wenn nicht sogar unmöglich, das Kontraktionsintervall I zu bestimmen und somit die Bedingungen (N1)–(N4) von Satz 7.23 zu verifizieren. In der Praxis wird man sich nicht der mühevollen Analyse der Bedingungen (N1)–(N4) unterwerfen, sondern pragmatisch vorgehen. Hat man eine ungefähre Vorstellung von der Lage der gesuchten Lösung x, so wird man das NEWTON–Verfahren mit einem entsprechenden Startwert x_0 initialisieren und seine Konvergenzeigenschaften beobachten. Abgesehen von diesem Problem der Lokalisierung eines Kontraktionsintervalles – welches in gleicher Weise beim Sekantenverfahren auftritt – beachte man, dass in jedem NEWTON–Schritt *zwei* Funktionsauswertungen, nämlich $F(x_n)$ und $F'(x_n)$, vorgenommen werden müssen. Geht man dabei von vergleichbarem Aufwand aus, könnten statt dessen *zwei* Schritte mit dem Sekantenverfahren ausgeführt werden, für welche gemäß (7.26) aus Abschnitt 6.7 pro Iterationsschritt nur *eine* Funktionsauswertung erforderlich ist. Ein Doppelschritt des Sekantenverfahrens hat darüber hinaus die Konvergenzordnung $(\sqrt{5}+3)/2 \doteq 2.618$ (was hier nicht begründet werden kann), und diese ist größer als die Konvergenzordnung 2 des NEWTON–Verfahrens. Das Sekantenverfahren ist somit in jeder Hinsicht effizienter als das NEWTON–Verfahren, wenn man noch hinzunimmt, dass die analytische Vorgabe der Ableitung $F'(x)$ wegfällt.

BSP. (7.8.1) Wir greifen hier nochmals das BSP. (6.7.4) der Funktion $F(x) := e^{2x} \cdot \sin x - 1$ aus Abschnitt 6.7 auf. Auf dem Intervall $[a, b] := [0.4, 0.5]$ sind wenigstens die Bedingungen (N1) und (N2) des Satzes 7.23 erfüllt. Es gilt $F'(x) = e^{2x}(2\sin x + \cos x)$. In der folgenden Tabelle sind die numerischen Resultate des NEWTON–Verfahrens für dieses Beispiel zusammengestellt. Als Abbruchtoleranz wurde $\epsilon := 10^{-10}$ vorgegeben.

n	x_n	$F(x_n)$	$F'(x_n)$	$F(x_n)/F'(x_n)$
0	0.400 000 000	−0.133 333 541	3.783 191 858	−0.035 243 664
1	0.435 243 664	0.006 887 033	4.179 201 870	0.001 647 930
2	0.433 595 733	0.000 015 844	4.159 985 179	0.000 003 809
3	0.433 591 925	0.000 000 000	4.159 940 848	0.000 000 000

BSP. (7.8.2) Die m–te Wurzel einer positiven Zahl a ist die reelle Lösung der Gleichung $F(x) := x^m - a = 0,\ \ a > 0$. Das NEWTON–Verfahren (8.1) hat in diesem Fall die Form

$$\boxed{x_{n+1} = \frac{1}{m} \cdot \Big(\frac{a}{(x_n)^{m-1}} + (m-1)x_n \Big),\ \ k \in \mathbf{N}_0.} \tag{8.8}$$

Für $m = 2$ erhält man daraus das HERON–Verfahren aus Abschnitt 3.1. Für $m = -1$ ergibt sich hingegen ein **divisionsfreier Algorithmus** zur Berechnung von $\frac{1}{a}$, nämlich

$$x_{n+1} = \Big(2 - a \cdot x_n \Big)\, x_n,\ \ n \in \mathbf{N}_0.$$

Vereinfachtes NEWTON–Verfahren. In (8.1) wird die Ableitung $F'(x)$ nur einmal für den (guten!) Startwert x_0 berechnet. Man betrachtet jetzt anstelle von (8.1) das weniger aufwendige Iterationsverfahren

$$\boxed{x_{n+1} = x_n - \frac{F(x_n)}{F'(x_0)},\ \ k \in \mathbf{N}_0.} \tag{8.9}$$

Die Konvergenz ist nun nicht mehr quadratisch, jedoch kann das Verfahren (8.9) bei geeignetem Startwert x_0 oft sehr schnell konvergieren.

BSP. (7.8.3) Wie in BSP. (7.8.1) sei $F(x) := e^{2x} \cdot \sin x - 1$. Wir wählen die sehr gute Startnäherung $x_0 = 0.43$. Die folgende Tabelle zeigt die numerischen Ergebnisse zum vereinfachten NEWTON–Verfahren:

n	x_n	$F(x_n)$	$F'(x_n)$	$F(x_n)/F'(x_n)$
0	0.430 000 000	−0.014 867 305	4.118 297 521	−0.003 610 061
1	0.433 610 061	0.000 075 448		0.000 018 320
2	0.433 591 741	−0.000 000 765		−0.000 000 186
3	0.433 591 927	0.000 000 008		0.000 000 002
4	0.433 591 925	−0.000 000 000		−0.000 000 000

Der Nachteil des NEWTON–Verfahrens gegenüber der einfachen Fixpunktiteration

$$\boxed{x_{n+1} = f(x_n),\ n \in \mathbf{N}_0,\ x_0 \in D(f),} \tag{8.10}$$

liegt möglicherweise in der sehr aufwendig werdenden Berechnung der Ableitung $F'(x_n)$. Dafür hat das NEWTON–Verfahren erheblich bessere Konvergenzeigenschaften als die einfache Fixpunktiteration (8.10). Wir wollen uns trotzdem nochmals mit der Fixpunktiteration (8.10) auseinandersetzen, und zwar unter dem Aspekt der **Differenzierbarkeit** der Funktion f. Anstelle des Fixpunktsatzes 6.25 von BANACH haben wir:

Satz 7.24 (Kontraktionsfixpunktsatz)
Die Funktion $f \in C^1([a,b])$ habe die Eigenschaften

$$\boxed{\text{(F1)}\quad f : [a,b] \to [a,b], \qquad \text{(F2)}\quad |f'(x)| \le q < 1\ \ \forall\, x \in [a,b].}$$

Dann hat f im Intervall $[a,b]$ genau einen Fixpunkt $x = f(x)$. Dieser ist der Grenzwert der Folge (8.10), wobei der Startwert $x_0 \in [a,b]$ beliebig wählbar ist. Es gilt die a priori–Fehlerabschätzung

$$\boxed{|x - x_n| \le \frac{q^n}{1-q}\, |x_1 - x_0|.} \tag{8.11}$$

Begründung: Da die Funktion f auf dem Intervall $[a,b]$ stetig differenzierbar ist, gilt nach dem Mittelwertsatz und der Voraussetzung (F2):

$$|f(x)-f(y)| \le \max_{\xi\in[a,b]} |f'(\xi)|\,|x-y| \le q\,|x-y| \quad \forall\, x,y\in[a,b].$$

Die Funktion f ist kontrahierend, und somit gilt der BANACHsche Fixpunktsatz 6.25. □

Ergänzungen. (a) Es ist nicht immer ganz einfach, das Intervall $[a,b]$ so zu bestimmen, dass die Funktion f die Selbstabbildungseigenschaft (F1) erfüllt. Ist jedoch bereits bekannt, dass f im Intervall $[a,b]$ einen Fixpunkt x hat (zum Beispiel durch Nachprüfen der Bedingung $\big(f(a)-a\big)\cdot\big(f(b)-b\big)<0$), so bleibt die Aussage des Fixpunktsatzes 7.24 richtig, wenn (F1) und (F2) ersetzt werden durch die Bedingung

$$\boxed{\text{(F3)}\quad |f'(x)| \le q < 1 \ \ \forall\, x\in I := (2a-b,2b-a) \supset [a,b].}$$

Begründung: Wir zeigen durch vollständige Induktion, dass die Folge (8.10) das Intervall I nicht verlässt, wenn man mit $x_0\in[a,b]$ startet. Es gilt nämlich wegen (F3) und wegen $x\in[a,b]$:

$$|x_1-x| = |f(x_0)-f(x)| \le q\,|x_0-x| < b-a,$$

und somit $a+x-b<x_1<b+x-a$. Wegen $a\le x\le b$ folgt daraus $2a-b<x_1<2b-a$, also $x_1\in I$. *Vererbung:* Es sei nun schon $x_1,x_2,\ldots,x_n\in I$ nachgewiesen. Dann folgt wie vorher:

$$|x_{n+1}-x| = |f(x_n)-f(x)| \le q|x_n-x| \le \cdots \le q^{n+1}|x_0-x| < b-a.$$

Wir erhalten wiederum $2a-b<x_{n+1}<2b-a$ sowie $\lim\limits_{n\to\infty}|x_{n+1}-x|=0$. □

BSP. (7.8.4) Die Funktion $f(x):=e^{-x}$ erfüllt auf dem Intervall $[a,b]:=[e^{-1},1]$ sicher die Bedingung (F1). Da f streng monoton fällt, gilt nämlich $f([a,b])=[e^{-1},e^{-1/e}]\subset[a,b]$. Es ist ferner (F2) für $q:=e^{-1/e}\doteq 0.6922$ erfüllt. Also ist Satz 7.24 anwendbar.

Fixpunktiteration für $f(x)=e^{-x}$

n	x_n	$f(x_n)$	$\lvert x_n-x\rvert$	$\frac{q^n}{1-q}\lvert x_1-x_0\rvert$
0	1.000 000 000	0.367 879 441	0.432 856 710	2.053 677 218
1	0.367 879 441	0.692 200 628	0.199 263 849	1.421 556 659
2	0.692 200 628	0.500 473 501	0.125 057 337	0.984 002 411
3	0.500 473 501	0.606 243 535	0.066 669 790	0.681 127 087
4	0.606 243 535	0.545 395 786	0.039 100 245	0.471 476 597
⋮	⋮	⋮	⋮	⋮
33	0.567 143 288	0.567 143 292	0.000 000 002	0.000 010 970
34	0.567 143 292	0.567 143 290	0.000 000 002	0.000 007 593
35	0.567 143 290	0.567 143 291	0.000 000 001	0.000 005 256
36	0.567 143 291	0.567 143 290	0.000 000 001	0.000 003 638
37	**0.567 143 290**	0.567 143 291	0.000 000 000	0.000 002 518

Wir haben hier mit einer Fehlertoleranz $\epsilon=10^{-9}$ gerechnet. Die Fixpunktiteration (8.10) konvergiert recht langsam.

BSP. (7.8.5) Wir betrachten die Funktion $f(x):=\cos\frac{x}{2}$ auf dem Intervall $[a,b]:=[0,\pi]$. Man überlegt sich graphisch, dass in diesem Intervall ein Fixpunkt x von f liegen muss. Es gilt hier $I=(2a-b,2b-a)=(-\pi,2\pi)$, und wegen

$$|f'(x)| = \frac{1}{2}\,|\sin\frac{x}{2}| \le 0.5 =: q < 1 \ \ \forall\, x\in I$$

ist die Bedingung (F3) erfüllt. Die Fixpunktiteration (8.10) konvergiert für jeden Startwert $x_0\in[0,\pi]$.

Fixpunktiteration für $f(x) = \cos\frac{x}{2}$

n	x_n	$f(x_n)$	$\lvert x_n - x\rvert$	$\frac{q^n}{1-q}\lvert x_1 - x_0\rvert$
0	0.000 000 000	1.000 000 000	0.900 367 223	2.000 000 000
1	1.000 000 000	0.877 582 562	0.099 632 777	1.000 000 000
2	0.877 582 562	0.905 265 843	0.022 784 661	0.500 000 000
3	0.905 265 843	0.899 298 752	0.004 898 621	0.250 000 000
4	0.899 298 752	0.900 599 556	0.001 068 470	0.125 000 000
⋮	⋮	⋮	⋮	⋮
11	0.900 367 247	0.900 367 217	0.000 000 025	0.000 976 562
12	0.900 367 217	0.900 367 224	0.000 000 005	0.000 488 281
13	0.900 367 224	0.900 367 222	0.000 000 001	0.000 244 141
14	0.900 367 222	0.900 367 223	0.000 000 000	0.000 122 070
15	0.900 367 223	0.900 367 223	0.000 000 000	0.000 061 035

(b) Man kann in der Bedingung (F3) auf die Einführung des Intervalls $I := (2a - b, 2b - a)$ verzichten, wenn die Funktion f auf dem Intervall $[a, b]$ **monoton wächst**. Wie vorher sei jedoch bereits bekannt, dass f im Intervall $[a, b]$ einen Fixpunkt x hat. Dann bleibt die Aussage des Fixpunktsatzes 7.24 richtig, wenn (F1) und (F2) ersetzt werden durch die Bedingung

$$\boxed{\text{(F4)} \quad 0 \le f'(x) \le q < 1 \ \ \forall\, x \in [a, b].}$$

Begründung: Wir zeigen abermals durch vollständige Induktion, dass die Folge (8.10) das Intervall $[a, b]$ nicht verlässt, wenn man mit $x_0 \in [a, b]$ startet (*Verankerung*).
Vererbung: Es sei schon $x_1, x_2, \ldots, x_n \in [a, b]$ nachgewiesen. Dann folgt aus dem Mittelwertsatz für eine Zwischenstelle $\xi = x + \theta(x_n - x),\ \theta \in (0, 1)$:

$$x_{n+1} - x = f(x_n) - f(x) = f'(\xi)\,(x_n - x).$$

Wegen $0 \le f'(\xi) < 1$ liegt x_{n+1} zwischen x_n und x, also im Intervall $[a, b]$. Dies zeigt gleichzeitig die Monotonie der Folge $(x_n)_{n\ge 0}$ auf. Die Konvergenz gegen den Fixpunkt x zeigt man wie in (a). □

Fixpunktiteration für $f(x) = \sqrt{2 + \ln x}$

n	x_n	$f(x_n)$	$\lvert x_n - x\rvert$	$\frac{q^n}{1-q}\lvert x_1 - x_0\rvert$
0	1.500 000 000	1.550 956 192	0.064 462 259	0.078 825 058
1	1.550 956 192	1.561 688 714	0.013 506 068	0.027 868 867
2	1.561 688 714	1.563 895 055	0.002 773 545	0.009 853 132
3	1.563 895 055	1.564 346 362	0.000 567 204	0.003 483 608
4	1.564 346 362	1.564 438 582	0.000 115 897	0.001 231 642
⋮	⋮	⋮	⋮	⋮
9	1.564 462 218	1.564 462 251	0.000 000 041	0.000 006 804
10	1.564 462 251	1.564 462 258	0.000 000 008	0.000 002 406
11	1.564 462 258	1.564 462 259	0.000 000 002	0.000 000 850
12	1.564 462 259	1.564 462 259	0.000 000 000	0.000 000 301
13	1.564 462 259	1.564 462 259	0.000 000 000	0.000 000 106

BSP. (7.8.6) Gesucht ist die Lösung der Gleichung

$$\boxed{F(x) := x^2 - \ln x - 2 = 0.}$$

Wegen $F(1) = -1$ und $F(2) = 2 - \ln 2 > 0$ hat man sicher eine Lösung x im Intervall $[1,2]$. Für die Bestimmung von x schreiben wir $F(x) = 0$ in der Form $x^2 = 2 + \ln x$ und gewinnen daraus die Fixpunktgleichung $x = f(x) := \sqrt{2 + \ln x}$. Nun gilt

$$0 \le f'(x) = \frac{1}{2x\sqrt{2+\ln x}} \le f'(1) = \frac{1}{4}\sqrt{2} =: q < 1 \quad \forall\, x \in [1,2].$$

Also ist die Bedingung (F4) erfüllt, und die Fixpunktiteration (8.10) konvergiert für jeden Startwert $x_0 \in [1,2]$.

(c) Ist anstelle von (F3) die Bedingung $|f'(x)| \ge 1 \;\forall\, x \in [a,b]$ erfüllt, so kann die Folge (8.10) nicht gegen einen Fixpunkt $x \in [a,b]$ konvergieren. Denn entweder tritt ein erster Index n auf mit $x_n \ne [a,b]$, oder es gilt $x_{n+1} = f(x_n) \in [a,b]$. Aus dem Mittelwertsatz erhält man dann

$$|x_{n+1} - x| = |f'(\xi)|\,|x_n - x| \ge |x_n - x|.$$

Das heißt, x_{n+1} und alle weiteren Folgenglieder $x_k \in [a,b]$ liegen nicht näher bei x als x_n.

(d) Gilt jedoch auf einem Intervall I_1, welches den Fixpunkt x der Funktion $f \in C^1(I_1)$ enthalte, die Bedingung

$$\boxed{\text{(F5)} \quad |f'(x)| \ge M > 1 \;\; \forall\, x \in I_1,}$$

so ist die Funktion f auf I_1 streng monoton. Es existiert die Umkehrfunktion $\varphi(y) := f^{-1}(y)$, und für deren Ableitung folgt nach Satz 7.4 und (F5):

$$|\varphi'(y)| = \frac{1}{|f(x)|} \le \frac{1}{M} =: q < 1 \quad \forall\, y \in f(I_1).$$

Das heißt, wird im Fall (F5) anstelle der Fixpunktgleichung $x = f(x)$ die **inverse Gleichung** $\varphi(x) := f^{-1}(x) = x$ betrachtet, so können wegen $|\varphi'(y)| \le q < 1$ die Resultate aus Satz 7.24 bzw. aus den obigen Ergänzungen (a) und (b) auf die inverse Gleichung angewendet werden.

BSP. (7.8.7) Gesucht werden die Nullstellen $x \in \mathbf{R}$ des Polynoms $G(x) := x^3 + x - 5$.

(a) Man zeige, dass es genau eine Lösung $x \in \mathbf{R}$ der Gleichung $G(x) = 0$ gibt. Man bestimme ein Intervall $[N, N+1]$, $N \in \mathbf{N}$, welches die Lösung x enthält.

(b) Zur Gleichung $G(x) = 0$ sind die beiden Fixpunktdarstellungen

$$x = 5 - x^3 =: f_1(x), \qquad x = (5-x)^{1/3} =: f_2(x)$$

äquivalent. Welche der beiden Funktionen $f_j, j = 1,2$, ermöglicht die Berechnung von x mit Hilfe der Fixpunktiteration $x_{n+1} = f_j(x_n)$, $n \in \mathbf{N}_0$?

(c) Man schätze die Zahl k der Iterationsschritte ab, die man für die Fixpunktiteration in (b) höchstens benötigt, damit der Fehler $|x_k - x| < \epsilon$ wird. Man wähle als Startwert $x_0 := N + 1$ sowie $\epsilon = 10^{-4}$.

Lösungen: (a) Die Funktion G ist wegen $G'(x) = 3x^2 + 1 \ge 1 > 0$ auf $\mathbf{R}$ streng monoton wachsend, und es gilt $\lim\limits_{x \to \pm\infty} G(x) = \pm\infty$. Also hat die Gleichung $G(x) = 0$ genau eine Lösung $x \in \mathbf{R}$. Es gilt ferner $G(1) \cdot G(2) = (-1) \cdot 5 < 0$. Somit folgt $x \in [1,2]$ aus dem Nullstellensatz von Bolzano, und wir haben $N = 1$.

(b) Es ist $|f_1'(x)| = 3x^2 \ge 3 > 1 \;\forall\, x \in I_1 := [1,2]$, und somit gilt die Bedingung (F5). Da die Funktion f_2 die Umkehrfunktion von f_1 ist, eignet sich f_2 zur iterativen Berechnung des Fixpunktes x, nicht aber f_1.

(c) Wegen (b) haben wir $|f_1'(x)| \ge M := 3 \;\forall\, x \in I_1$, und somit $|f_2'(y)| \le 1/3 =: q < 1 \;\forall\, y \in f_1(I_1)$. Da die Funktion f_1 monoton wächst, folgern wir $f_1(I_1) = [f_1(1), f_1(2)] = [-1,5] \supset I_1$. Mithin gilt $f_2(I_1) \subset I_1$, und die Funktion f_2 erfüllt somit alle Voraussetzungen zum Fixpunktsatz 7.24. Aus (8.11) resultiert für $x_0 = 2$ die Fehlerabschätzung

$$|x - x_k| \le \frac{q^k}{1-q}\,|x_1 - x_0| = \frac{q^k}{1-q}\,(2 - 3^{1/3}) < \epsilon = 10^{-4}.$$

Mit $q = 1/3$ ergibt sich also

$$q^k < \frac{2 \cdot 10^{-4}}{3\,(2 - 3^{1/3})} =: A \quad \Rightarrow \quad k > \frac{\ln A}{\ln q} \doteq 8.22.$$

Fixpunktiteration für $f_2(x) := (5 - x)^{1/3}$

n	x_n	$f(x_n)$	$\lvert x_n - x\rvert$	$\frac{q^n}{1-q}\,\lvert x_1 - x_0\rvert$
0	2.000 000 000	1.442 249 570	0.484 019 772	0.836 625 645
1	1.442 249 570	1.526 599 655	0.073 730 657	0.278 875 215
2	1.526 599 655	1.514 438 405	0.010 619 427	0.092 958 405
3	1.514 438 405	1.516 203 823	0.001 541 823	0.030 986 135
4	1.516 203 823	1.515 947 796	0.000 223 595	0.010 328 712
⋮	⋮	⋮	⋮	⋮
8	1.515 980 327	1.515 980 213	0.000 000 099	0.000 127 515
9	1.515 980 213	1.515 980 230	0.000 000 014	0.000 042 505
10	1.515 980 230	1.515 980 227	0.000 000 002	0.000 014 168
11	1.515 980 227	1.515 980 228	0.000 000 000	0.000 004 723
12	1.515 980 228	1.515 980 228	0.000 000 000	0.000 001 574

Die obige Rechnung wurde für eine Fehlertoleranz $\epsilon = 10^{-9}$ durchgeführt. Die Fehlertoleranz $\epsilon = 10^{-4}$ wird nach dem 6.Iterationsschritt unterschritten.

7.9 Der Interpolationsfehler

Wir verfügen jetzt über die Hilfsmittel der Differentialrechnung, um noch einmal das Problem des **Interpolationsfehlers** zu untersuchen, welches wir bereits in Abschnitt 6.2 im Rahmen der Polynominterpolation andiskutiert hatten. Nachfolgend bezeichnen wir mit $P_n(x)$ das eindeutig bestimmte Interpolationspolynom (in der Form von LAGRANGE oder von NEWTON), welches eine gegebene Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ in den $n+1$ Stützpunkten $(x_j, y_j = f(x_j))$, $j = 0, 1, \ldots, n$, interpoliert. Eine qualitative Aussage über den Interpolationsfehler

$$\boxed{F_n(x) := |f(x) - P_n(x)|}$$

treffen wir im folgenden Satz.

Satz 7.25 *Gegeben seien die reelle Funktion $f \in C^{n+1}([a,b])$ und dazu Funktionswerte $y = f(x)$ in den $n+1$ Stützpunkten (x_j, y_j), $j = 0, 1, \ldots, n$, mit $x_j \neq x_k \;\forall\; j \neq k$. Gelte $a = \min\limits_{0 \le j \le n} x_j$ sowie $b = \max\limits_{0 \le j \le n} x_j$, und sei $P_n(x)$ das zugeordnete Interpolationspolynom vom Grad n. Dann existiert für jede Zahl $x \in [a,b]$ eine Zwischenstelle $\xi = \xi(x) \in (a,b)$ mit*

$$\boxed{f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot \prod_{j=0}^{n} (x - x_j).} \tag{9.1}$$

Begründung: Ist $x = x_k$, so ist $F_n(x_k) = 0$, und die Aussage (9.1) ist trivial. Sei also $x \neq x_k$. Dann ist die Funktion

$$G(t) := f(t) - P_n(t) - \frac{\prod\limits_{j=0}^{n} (t - x_j)}{\prod\limits_{j=0}^{n} (x - x_j)} \cdot [f(x) - P_n(x)]$$

im Intervall $[a,b]$ $(n+1)$–mal stetig differenzierbar. Ausserdem hat $G(t)$ mindestens die $n+2$ Nullstellen x, x_0, $x_1,\ldots,x_n$. Wiederholte Anwendung des Satzes von ROLLE (Satz 7.10) liefert, dass $G^{(n)}(t)$ noch mindestens zwei Nullstellen haben muss. Abermals unter Verwendung des Satzes von ROLLE erhält man die Existenz einer Zwischenstelle $\xi=\xi(x)\in(a,b)$ mit

$$G^{(n+1)}(\xi)=0=f^{(n+1)}(\xi)-\frac{(n+1)!}{\prod\limits_{j=0}^{n}(x-x_j)}\cdot[f(x)-P_n(x)].$$

Auflösen nach $f(x)-P_n(x)$ liefert die behauptete Relation (9.1). □

Bemerkung 7.16 (a) Es gilt nun die folgende Abschätzung für den Interpolationsfehler $F_n(x)=|f(x)-P_n(x)|$:

$$\boxed{F_n(x)\le\frac{M_{n+1}}{(n+1)!}\cdot\prod_{j=0}^{n}|x-x_j|\le\frac{(b-a)^{n+1}}{(n+1)!}\cdot M_{n+1},\quad M_{n+1}:=\max_{\xi\in[a,b]}|f^{(n+1)}(\xi)|.}\tag{9.2}$$

(b) Wir betrachten **äquidistante Stützstellen** sowie die Fälle $n=1,2,3$. Das Maximum der Funktion $|\varphi(x)|:=\prod\limits_{j=0}^{n}|x-x_j|$ kann im Intervall $[a,b]$ durch *Kurvendiskussion* explizit bestimmt werden. *Zum Beispiel* hat man im Fall $n=1$ nur $x_1=x_0+h$ zu berücksichtigen. Hier gilt:

$$\begin{aligned}\varphi(x)&=(x-x_0)(x-x_1)=(x-x_0)^2-h(x-x_0)<0\ \ \forall\ x\in[x_0,x_1],\\ \varphi'(x)&=2(x-x_0)-h\stackrel{!}{=}0\Leftrightarrow x-x_0=\tfrac{h}{2},\\ \varphi''(x)&=2>0.\end{aligned}$$

Das heißt, $\varphi(x)$ hat ein absolutes Minimum bei $x=x_0+\frac{h}{2}$, so dass $|\varphi(x)|\le|\varphi(x_0+\frac{h}{2})|=\frac{h^2}{4}$ gilt. Anstelle von (9.2) hat man für die

Lineare Interpolation $(n=1)$**:**

$$\boxed{|f(x)-P_1(x)|\le\frac{h^2}{8}\cdot M_2\ \ \forall\ x\in[x_0,x_1];\quad x_1:=x_0+h.}\tag{9.3}$$

Analoge Rechnungen liefern für die

Quadratische Interpolation $(n=2)$**:**

$$\boxed{|f(x)-P_2(x)|\le\frac{\sqrt{3}h^3}{27}\cdot M_3\ \ \forall\ x\in[x_0,x_2];\quad x_2:=x_0+2h.}\tag{9.4}$$

Kubische Interpolation $(n=3)$**:**

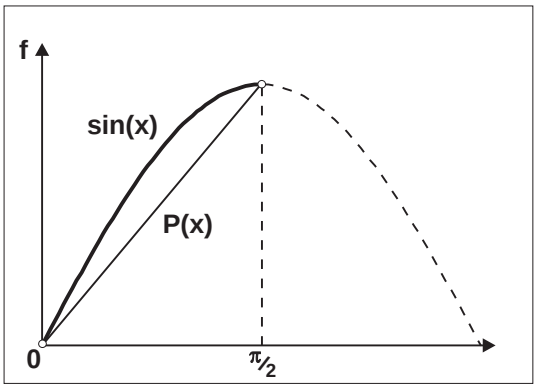
$$\boxed{|f(x)-P_3(x)|\le\begin{cases}\frac{3h^4}{128}\cdot M_4\ \ \forall\ x\in[x_1,x_2],\\ \frac{h^4}{24}\cdot M_4\quad\forall\ x\in[x_0,x_1]\cup[x_2,x_3],\end{cases}\qquad x_j:=x_0+jh.}\tag{9.5}$$

BSP. (7.9.1) Die Funktion $f(x):=\sin x$ soll auf dem Intervall $[0,\frac{\pi}{2}]$ *linear* $(n=1)$ interpoliert werden. Das heißt, es ist die interpolierende Gerade $P_1(x)$ durch die beiden Stützpunkte $(0,0)$ und $(\frac{\pi}{2},1)$ zu bestimmen. Wir haben $x_0=0$, $x_1=\frac{\pi}{2}$ und somit offensichtlich $P_1(x)=2x/\pi$. Für den Interpolationsfehler resultiert aus (9.3):

$$|\sin x-\frac{2x}{\pi}|\le\frac{\pi^2}{32}\max_{\xi\in[0,\pi/2]}|\sin\xi|=\frac{\pi^2}{32}\doteq0.308.$$

Bei *quadratischer* ($n = 2$) äquidistanter Interpolation auf dem Intervall $[0, \pi]$ erhält man das Interpolationspolynom $P_2(x) = 4x(\pi - x)/\pi^2$. Aus (9.4) resultiert mit $h := \pi/2$ der Interpolationsfehler

$$|\sin x - \frac{4x}{\pi^2}(\pi - x)| \leq \frac{\sqrt{3}\pi^3}{8 \cdot 27} \max_{\xi \in [0,\pi]} |-\cos \xi| = \frac{\sqrt{3}\pi^3}{216} \doteq 0.248.$$

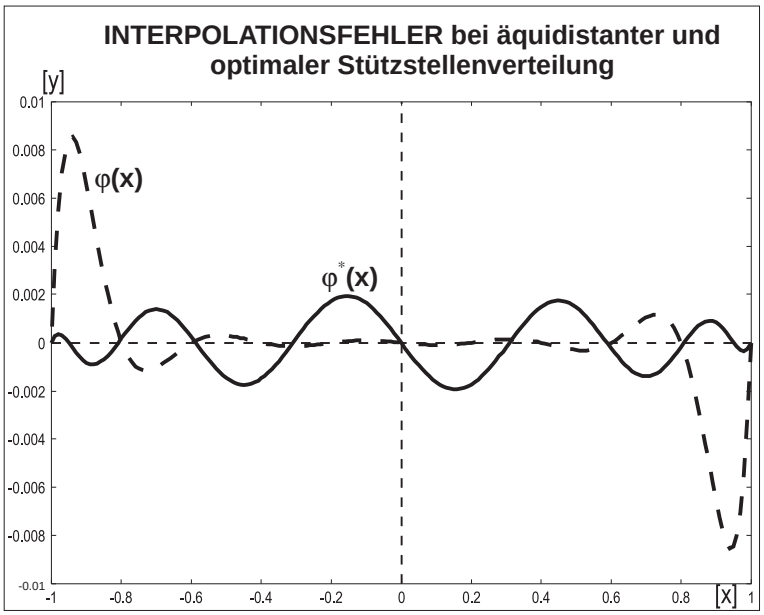


Lineare Interpolation von $\sin x$

Kritik: Zur Polynominterpolation ist festzustellen, dass die Güte der Approximation sehr empfindlich von der Verteilung der Stützstellen $x_0, x_1, \ldots, x_n$ auf dem Intervall $[a, b]$ ahhängt. Wegen (9.1) wird die Fehlerfunktion $f(x) - P_n(x)$ weitestgehend durch die Funktion

$$\varphi(x) := \prod_{j=0}^{n} (x - x_j) \tag{9.6}$$

bestimmt. Für *äquidistante Stützstellen* $x_j := x_0 + jh$ und größere n *oszilliert* $\varphi(x)$ *sehr stark an den Intervallenden* des Intervalls $[a, b] := [x_0, x_n]$. Im Falle $n = 10$ belegt dies die folgende Grafik:



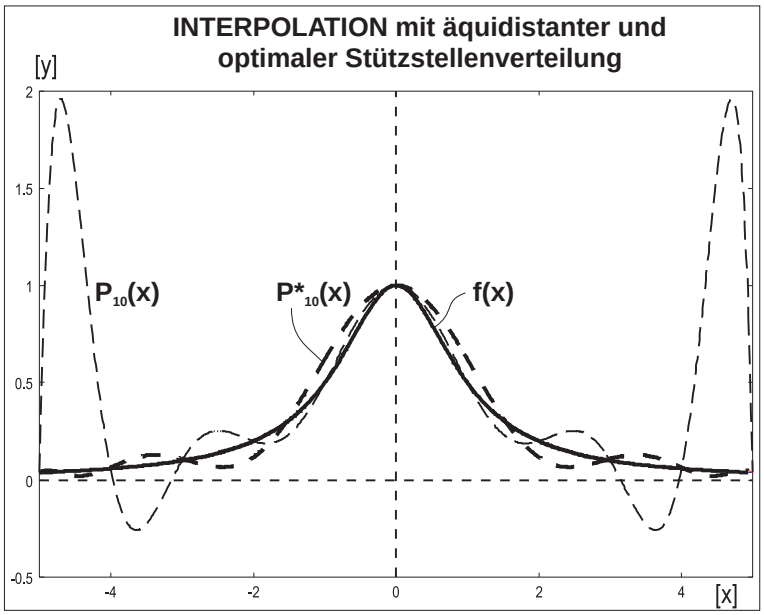
Einfluss der Stützstellenverteilung auf den Interpolationsfehler

Eine Verbesserung tritt grundsätzlich ein, wenn die Stützstellen an den Intervallenden *dichter* gelegt werden als in der Intervallmitte. Man kann begründen ($\rightarrow$ Spezialliteratur), dass die Stützstellen im Interpolationsintervall $[-1, +1]$ genau dann (unter den Nebenbedingungen $x_0 := -1$, $x_n := +1$) optimal verteilt sind, wenn sie die Extremalwerte des n–ten TSCHEBYSCHEFF–Polynoms $T_n(x)$ sind. Bei Transformation auf ein allgemeines Intervall $[a, b]$ heißt dies konkret, dass man

$$\boxed{x_k^* := \frac{a+b}{2} + \frac{b-a}{2} \cdot \cos(\tfrac{n-k}{n} \cdot \pi), \quad k = 0, 1, \ldots, n,} \tag{9.7}$$

zu setzen hat. Die Funktion $\varphi^*(x) := \prod_{j=0}^{n}(x - x_j^*)$ nivelliert die Oszillationen stärker. Man kann zeigen, dass stets $\max\limits_{x\in[a,b]} |\varphi^*(x)| \leq \max\limits_{x\in[a,b]} |\varphi(x)|$ gilt. Man vergleiche dazu auch die obige Grafik.

BSP. (7.9.2) Ein *klassisches Beispiel* für diesen Sachverhalt stammt von C.RUNGE: Für $f(x) := 1/(1+x^2)$ wird $P_n(x)$ auf dem Intervall $[-5, +5]$ zu äquidistanten Stützstellen $x_k := -5 + 10\,k/n$ berechnet sowie $P_n^*(x)$ auf $[-5, +5]$ zu nichtäquidistanten Stützstellen x_k^* aus (9.7). Während für $P_n(x)$ der Interpolationsfehler mit n zunimmt, konvergiert $P_n^*(x)$ gegen $f(x)$. In der folgenden Grafik ist die Situation bei Wahl von $n = 10$ dargestellt.



Interpolation von $f(x) = 1/(1 + x^2)$ mit äquidistanter und optimaler Stützstellenverteilung

---

7.10 Numerische Differentiation und Extrapolation

Die empirische Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ sei in den $n+1$ Stützpunkten $(x_j, y_j := f(x_j))$, $j = 0, 1, \ldots, n$, mit $x_j \neq x_k \;\forall\; j \neq k$ gegeben. Man bestimme aus diesen Vorgaben eine geeignete Näherung der n–ten Ableitung $f^{(n)}(x)$. Es ist sicher nicht abwegig, das LAGRANGE–Interpolationspolynom $P_n(x)$ zu den vorgegebenen Stützpunkten zu bilden und n–mal nach x zu differenzieren. Die so erhaltene Ableitung $P_n^{(n)}(x)$ verwenden wir als Approximation von $f^{(n)}(x)$. Aus der Darstellung (2.12) in Abschnitt 6.2 resultiert zunächst:

$$\boxed{P_n^{(n)}(x) = n! \cdot \sum_{j=0}^{n} y_j \lambda_j^{(n)} \overset{!}{\approx} f^{(n)}(x).} \tag{10.1}$$

Dieser Ausdruck ist offenkundig unabhängig von x. Deshalb bleibt die Frage offen, für welche Punkte x durch (10.1) eine brauchbare Approximation geliefert wird. Eine Antwort geben wir in dem folgenden Satz:

Satz 7.26 *Die reellwertige Funktion $y = f(x)$ sei in den $n+1$ Stützpunkten (x_j, y_j), $j = 0, 1, \ldots, n$, mit $x_j \neq x_k \;\forall\; j \neq k$ gegeben. Es gelte $a := \min\limits_{0\leq j\leq n} x_j$, $b := \max\limits_{0\leq j\leq n} x_j$ sowie $f \in$*

$C^n([a,b])$. *Dann existiert eine Zwischenstelle* $\xi \in (a,b)$ *mit*

$$\boxed{f^{(n)}(\xi) = P_n^{(n)}(\xi) = n! \cdot \sum_{j=0}^{n} y_j \lambda_j^{(n)},} \tag{10.2}$$

worin $P_n(x)$ *das Interpolationspolynom* (2.1) *aus Abschnitt* 6.2 *zu den gegebenen Stützpunkten bezeichnet.*

Begründung: Man setze $g(x) := f(x) - P_n(x)$. Dann gilt $g(x_j) = 0 \ \forall \ j = 0, 1, \ldots, n$. Zwischen je zwei aufeinanderfolgenden Nullstellen kann der Satz von ROLLE (Satz 7.10) angewendet werden mit dem Ergebnis, dass $g'(x)$ in (a,b) mindestens n Nullstellen haben muss. Wiederholte Anwendung des Satzes von ROLLE zeigt, dass $g''(x)$ mindestens $n-1$ Nullstellen in (a,b) haben muss usf., bis schließlich $g^{(n)}(x)$ mindestens eine Nullstelle $\xi \in (a,b)$ haben muss: $g^{(n)}(\xi) = f^{(n)}(\xi) - P_n^{(n)}(\xi) = 0$. Also gilt (10.2). □

Bemerkung 7.17 Die Relation (10.1) heißt **Regel der numerischen Differentiation** . Im allgemeinen wird sie nur für **äquidistante Stützstellen** $x_j := x_0 + jh, \ j = 0, 1, \ldots, n, \ h > 0$, verwendet. Unter Berücksichtigung von (2.13) aus Abschnitt 6.2 resultiert in diesem Fall anstelle von (10.1): □

$$\boxed{f^{(n)}(x) \approx \frac{1}{h^n} \cdot \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} y_j.} \tag{10.3}$$

Definition 7.11 *Der Ausdruck* (10.3) *heißt* n**–ter Differenzenquotient** *der* $n+1$ *Stützwerte* y_j. *Insbesondere heißen*

$$\boxed{\begin{array}{lll} f'(x) \approx \dfrac{y_1 - y_0}{h} & & \textbf{1.Differenzenquotient}, \\ f''(x) \approx \dfrac{y_2 - 2y_1 + y_0}{h^2} & & \textbf{2.Differenzenquotient}, \\ f'''(x) \approx \dfrac{y_3 - 3y_2 + 3y_1 - y_0}{h^3} & & \textbf{3.Differenzenquotient}. \end{array}} \tag{10.4}$$

Erfahrungswert: Die Stelle ξ, für die gemäß Satz 7.26 die Gleichheit (10.2) gilt, liegt häufig in der Nähe des Mittelpunktes $x_M := \frac{1}{2}(x_0 + x_n)$.

Es ist nicht zwingend, wie in (10.3) geschehen, die Ableitung $f^{(n)}(x)$ durch die n–te Ableitung des Interpolationspolynoms $P_n(x)$ zu approximieren. Im allgemeinen Fall kann die p–te Ableitung $f^{(p)}(x)$ durch $P_n^{(p)}(x)$ approximiert werden. Da $P_n^{(p)}(x) \neq const$ für $p < n$ gelten wird, ist jetzt die Approximation $f^{(p)}(x) \approx P_n^{(p)}(x)$ von der Stelle x abhängig.

BSP. (7.10.1) Man approximiere die 1.Ableitung $f'(x)$ unter Verwendung des Interpolationspolynoms 2.Grades:

$$\begin{aligned} P_2(x) &= \frac{1}{2h^2}\Big(y_0(x-x_1)(x-x_2) - 2y_1(x-x_0)(x-x_2) + y_2(x-x_0)(x-x_1)\Big), \\ P_2'(x) &= \frac{1}{2h^2}\Big(y_0(2x-x_1-x_2) - 2y_1(2x-x_0-x_2) + y_2(2x-x_0-x_1)\Big). \end{aligned}$$

Setzt man $f'(x) \approx P_2'(x)$ so resultiert insbesondere im Punkt $x := x_1$:

$$\boxed{f'(x_1) \approx \frac{1}{2h}\,(-y_0 + y_2).} \tag{10.5}$$

Definition 7.12 *Der Ausdruck* (10.5) *heiße* **zentraler Differenzenquotient.**

Um einen Vergleich zwischen den beiden Approximationen (10.4) und (10.5) zu bekommen, setzen wir

$$\begin{aligned} T_1(h) &:= \frac{1}{h}\,(y_1 - y_0) = \frac{1}{h}\,\big[f(x_0+h) - f(x_0)\big], \\ T_2(h) &:= \frac{1}{2h}\,(-y_0 + y_2) = \frac{1}{2h}\,\big[f(x_1+h) - f(x_1-h)\big]. \end{aligned}$$

Durch TAYLOR–Entwicklung von $f(x \pm h)$ an der Stelle $h = 0$ resultiert dann:

$$\begin{aligned} T_1(h) &= \frac{1}{h}\left\{f(x_0) - f(x_0) + h\cdot f'(x_0) + \frac{h^2}{2}\cdot f''(\xi)\right\} = f'(x_0) + \frac{h}{2}\cdot f''(\xi), \\ T_2(h) &= \frac{1}{2h}\left\{f(x_1) - f(x_1) + h\cdot\big[f'(x_1) + f'(x_1)\big] + \frac{h^2}{2}\cdot\big[f''(x_1) - f''(x_1)\big] + \frac{h^3}{6}\cdot\big[f'''(\xi) + f'''(\xi)\big]\right\} \\ &= f'(x_1) + \frac{h^2}{6}\cdot f'''(\xi). \end{aligned}$$

Hieraus ergeben sich die Fehlerabschätzungen

$$\boxed{|f'(x_0) - T_1(h)| \le \frac{h}{2}\,M_2, \qquad |f'(x_1) - T_2(h)| \le \frac{h^2}{6}\,M_3} \tag{10.6}$$

mit $M_2 := \max\limits_{\xi\in[x_0,x_0+h]} |f''(\xi)|$ und $M_3 := \max\limits_{\xi\in[x_1-h,x_1+h]} |f'''(\xi)|$. Der zentrale Differenzenquotient (10.5) approximiert die 1.Ableitung von $f(x)$ *besser* als der Differenzenquotient (10.4).

BSP. (7.10.2) Man approximiere die 1.Ableitung $f'(x)$ unter Verwendung des LAGRANGE–Interpolationspolynoms $P_3(x)$ vom Grade 3. Eine zum vorstehenden BSP. (7.10.1) analoge Rechnung zeigt:

$$\boxed{\begin{aligned} f'(x_0) &\approx \frac{1}{6h}\Big(-11y_0 + 18y_1 - 9y_2 + 2y_3\Big), \\ f'(x_1) &\approx \frac{1}{6h}\Big(-2y_0 - 3y_1 + 6y_2 - y_3\Big), \\ f'(x_M) &\approx \frac{1}{24h}\Big(y_0 - 27y_1 + 27y_2 - y_3\Big), \quad x_M := (x_0+x_3)/2, \\ f''(x_M) &\approx \frac{1}{2h^2}\Big(y_0 - y_1 - y_2 + y_3\Big), \quad x_M := (x_0+x_3)/2. \end{aligned}} \tag{10.7}$$

Bemerkung 7.18 Für *äquidistante Stützstellen* erhält man immer Formeln vom Typ

$$\boxed{f'(x) \approx \frac{1}{h}\cdot\sum_k c_k y_k \quad \text{mit} \quad \sum_k c_k = 0.}$$

Die letzte Bedingung stellt sicher, dass die konstante Funktion $f \equiv const$ exakt differenziert wird. □

BSP. (7.10.3) Für die Funktion $f(x) := e^{2x}\cdot\sin x$ soll unter Verwendung der Formel (10.4) die zweite Ableitung $f''(x) = e^{2x}\cdot(4\cos x + 3\sin x)$ an der Stelle $x_1 := 0.5$ näherungsweise berechnet werden. Es gilt bei analytischer Rechnung $f''(0.5) \doteq 13.451\,708\,113$. In der folgenden Tabelle sind für verschiedene Schrittweiten $h > 0$ die numerischen Werte von y_0, y_2 und der aus (10.4) resultierende Näherungswert für $f''(0.5)$ aufgelistet.

h	y_0	y_2	$f''(x_1) \approx$	abs.Fehler
0.1	0.866 666 459	1.874 679 031	13.491 803 094	0.040 094 981
0.01	1.253 962 085	1.353 810 585	13.452 109 124	0.000 401 011
0.001	1.298 228 507	1.308 212 405	13.451 704 945	0.000 003 168
0.0001	1.302 714 603	1.303 712 991	13.451 062 841	0.000 645 272
0.00001	1.303 163 811	1.303 263 650	13.387 762 010	0.063 946 103
0.000001	1.303 208 738	1.303 218 722	5.456 968 211	7.994 739 902

Mit kleiner werdender Schrittweite h tritt keinesfalls – wie erwartet werden sollte – Konvergenz gegen den exakten Wert auf. Vielmehr entstehen durch immer katastrophaler werdende Stellenauslöschungen ganz falsche Näherungswerte.

Merke: Numerische Differentiation ist i.a. ein gefährlicher Prozess. Der Limes $h \to 0$ ist aus numerischen Gründen wegen wachsender Stellenauslöschung nicht vollziehbar.

Wir zeigen, dass mit Hilfe der **Extrapolation** ein Ausweg aus dieser Situation gefunden werden kann. Zur Motivation erinnern wir nochmals an die TAYLOR–Entwicklungen $T_j(h)$, mit deren Hilfe die Fehlerabschätzungen (10.6) hergeleitet wurden. Diese Entwicklungen waren von der Form

$$T_j(h) = f'(x) + a_1 h + a_2 h^2 + \cdots + a_n h^n + \cdots .$$

Das heißt, eine **berechenbare Größe** $T(h)$, die von einem Parameter h abhängt, approximiert einen gesuchten Wert A (hier $f'(x)$) mit einem Fehler, der als Potenzreihe in h darstellbar ist:

$$\boxed{T(h) = A + a_1 h + a_2 h^2 + \cdots + a_n h^n + \cdots, \quad a_j \text{ fest.}} \tag{10.8}$$

Wie das vorangegangene BSP. (7.10.3) zeigt, können numerische Gründe verhindern, die Größe $T(h)$ für hinreichend kleine Werte von h zu bestimmen, um so durch $T(h)$ eine hinreichend genaue Approximation für den gesuchten Wert A zu erhalten. Ein sehr wirksames Gegenmittel ist das folgende

Prinzip der Extrapolation.

Man bestimme zunächst Funktionswerte $T(h_k)$ für einige Stützstellen $h_0 > h_1 > \cdots > h_n > 0$. Danach werte man das zugeordnete Interpolationspolynom $P_n(h)$ an der Neustelle $h = 0$ aus. Für wachsendes n sollte nun $P_n(0)$ bessere Näherungen des gesuchten Wertes $A = T(0)$ liefern. Da wir das Interpolationspolynom $P_n(h)$ nur an der einzigen Stelle $h = 0$ auszuwerten haben, ist es vom Rechenaufwand her nicht vertretbar, den in Abschnitt 6.2 aufgezeigten Weg über die Berechnung der Stützkoeffizienten zu gehen. Es ist vielmehr sinnvoller, das Interpolationspolynom in der NEWTONschen Form (2.16), Abschnitt 6.2, zu verwenden und die beiden Algorithmen (2.18) zur Berechnung von $P_n(x)$ an einer Neustelle x sowie (2.24) zur Berechnung der Koeffizienten c_j aus Abschnitt 6.2 in einem einzigen Algorithmus zu vereinigen.

Ein solcher Schritt wird durch den NEVILLE–**Algorithmus** bewerkstelligt. Dieser berechnet aus der Vorgabe der Stützpunkte (x_j, y_j), $j = 0, 1, \ldots, n$, direkt den Wert des Interpolationspolynoms $P_n(x)$ an einer Neustelle x, und zwar unter Verwendung der Rekursionsformeln (2.21) und (2.22) für die Teilinterpolationspolynome $P_{j_0 j_1 \cdots j_k}(x)$ aus Abschnitt 6.2. Die Berechnung des gesuchten Funktionswertes $P_n(x) \equiv P_{012\ldots n}(x)$ erfolgt nun ganz analog dem Schema der dividierten Differenzen, vgl. Definition 6.11. Im *Fallbeispiel* $n = 4$ hat man dabei das folgende Rechenschema spaltenweise aufzubauen:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
x_0	$\boxed{y_0 = P_0(x)}$				
x_1	$y_1 = P_1(x)$	$\boxed{P_{01}(x)}$			
x_2	$y_2 = P_2(x)$	$P_{12}(x)$	$\boxed{P_{012}(x)}$		
x_3	$y_3 = P_3(x)$	$P_{23}(x)$	$P_{123}(x)$	$\boxed{P_{0123}(x)}$	
x_4	$y_4 = P_4(x)$	$P_{34}(x)$	$P_{234}(x)$	$P_{1234}(x)$	$\boxed{P_{01234}(x)}$

(10.9)

Die eingerahmten Werte $P_{012\ldots k}(x)$ sind die Funktionswerte der Teilinterpolationspolynome $P_k(x)$ zu den Stützstellen $x_0, x_1, \ldots, x_k$. Zur Implementierung des NEVILLE–Algorithmus auf einem Rechner wird nur ein einziger Vektor $\vec{p}$ mit $n+1$ Komponenten benötigt, der sukzessive die Werte der k–ten Spalte aufnimmt und dessen Komponenten nach Ablauf der Rechnung die eingerahmten Werte des Schemas (10.9) enthalten. Jede Spalte wird von unten nach oben abgearbeitet.

Rechenvorschrift für den NEVILLE–Algorithmus:

```
1:  Einlesen von x; (x_j, y_j), j := 0, 1, ..., n;
2:  für j := 0, 1, ..., n :
3:      p_j := y_j; (Ende j)
4:  für k := 1, 2, ..., n :
5:      für j := n, n-1, ..., k :
6:          p_j := p_j + (x - x_j) * (p_j - p_{j-1})/(x_j - x_{j-k}). (Ende j, k)
```

(10.10)

Nach Beendigung der Rechnung sind die gesuchten Funktionswerte $P_k(x)$ gleich den Werten von p_k.

Der NEVILLE–Algorithmus ist sehr gut geeignet, die oben dargelegte Extrapolationsaufgabe auf den Wert $h = 0$ durchzuführen, da hier genau ein einziger extrapolierter Wert zu berechnen ist. Setzen wir $p_j^{(k)} := P_{j-k,j-k+1,\ldots j}(h)$, so resultieren aus dem NEVILLE–Schema (10.10) bei Extrapolation auf $h = 0$ die folgenden Rekursionsformeln:

$$\boxed{p_j^{(k)} = p_j^{(k-1)} - \frac{h_j}{h_j - h_{j-k}} \cdot [p_j^{(k-1)} - p_{j-1}^{(k-1)}].} \tag{10.11}$$

Wir gehen davon aus, dass die Stützstellen $h_0 > h_1 > \cdots > h_n > 0$ eine monotone Folge bilden. Es ist jetzt zweckmäßig, die folgenden, zu (10.11) äquivalenten Rekursionsformeln zu benutzen:

$$\left.\begin{aligned} p_j^{(k)} &= p_j^{(k-1)} + \frac{h_j}{h_{j-k} - h_j} \cdot [p_j^{(k-1)} - p_{j-1}^{(k-1)}] \\ &= p_j^{(k-1)} + \frac{1}{\frac{h_{j-k}}{h_j} - 1} \cdot [p_j^{(k-1)} - p_{j-1}^{(k-1)}], \end{aligned}\right\} \quad \begin{array}{l} j = k, k+1, \ldots, n; \\ k = 1, 2, \ldots, n. \end{array} \tag{10.12}$$

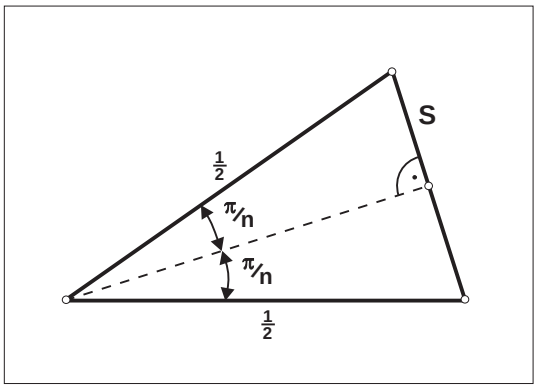
Die erste Formel verwendet man, wenn die h_j eine unregelmäßige Folge bilden. Die zweite Formel verwendet man bei speziellen Parameterfolgen, zum Beispiel bei äquidistanten Stützstellen. Anstelle von (10.10) haben wir jetzt den folgenden

Algorithmus zur Extrapolation auf Null:

```
1:  Einlesen von h_j, T(h_j), j := 0, 1, ..., n;
2:  für j := 0, 1, ..., n :
3:      p_j := T(h_j); (Ende j)
4:  für k := 1, 2, ..., n :
5:      für j := n, n-1, ..., k :
6:          p_j := p_j + h_j * (p_j - p_{j-1})/(h_{j-k} - h_j). (Ende j, k)
```

(10.13)

Nach Beendigung der Rechnung (10.13) liefern die Komponenten p_k den gesuchten Wert $P_k(0)$ des Interpolationspolynoms von Grade k an der Stelle $h = 0$.



Zur Berechnung der Kreiszahl π

BSP. (7.10.4) Die **Kreiszahl** π soll mit Hilfe von einbeschriebenen regulären n–Ecken näherungsweise berechnet werden. Wir müssen zunächst den Fehler $T(h)$ nach der Vorschrift (10.8) analytisch in Form einer Potenzreihe darstellen. Gemäß Skizze beträgt die Länge S der Sehne in dem regelmäßigen n–Eck:

$$S = \sin \frac{\pi}{n}.$$

Mithin resultiert der Umfang

$$\boxed{U_n = n \cdot S = n \cdot \sin \tfrac{\pi}{n}.}$$

Unter Verwendung der Potenzreihenentwicklung von $\sin x$ folgt daraus:

$$U_n = \pi - \frac{\pi^3}{3!} \cdot n^{-2} + \frac{\pi^5}{5!} \cdot n^{-4} - \frac{\pi^7}{7!} \cdot n^{-6} \pm \cdots \equiv T\big[(\frac{1}{n})^2\big]. \tag{10.14}$$

Setzen wir also $h := (1/n)^2$, so ist $T(h) = U_n$ die berechenbare Größe, die den gesuchten Wert $T(0) = A := \pi$ approximiert. Die Extrapolation auf $h = 0$ sollte dann eine brauchbare Näherung für π liefern. Die Ermittlung von Stützpunkten $(h_k, T(h_k))$ darf nicht unter Verwendung trigonometrischer Funktionen erfolgen, da diese mit der gesuchten Zahl π arbeiten. Es gibt jedoch einige elementar berechenbare Umfänge U_n, die in der folgenden Tabelle aufgelistet sind:

$n =$	2	3	4	6	8
$U_n =$	2	$\frac{3}{2}\sqrt{3}$	$2\sqrt{2}$	3	$4\sqrt{2-\sqrt{2}}$

Mit den Stützstellen $h_0 := \frac{1}{4}$, $h_1 := \frac{1}{9}$, $h_2 := \frac{1}{16}$, $h_3 := \frac{1}{36}$, $h_4 := \frac{1}{64}$ liefert der NEVILLE–Algorithmus zur Extrapolation auf Null bei zehnstelliger Rechnung:

h	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
0.250 000 000	2.000 000 000				
0.111 111 111	2.598 076 211	3.076 537 180			
0.062 500 000	2.828 427 125	3.124 592 585	3.140 611 053		
0.027 777 778	3.000 000 000	3.137 258 300	3.141 480 205	3.141 588 849	
0.015 625 000	3.061 467 459	3.140 497 049	3.141 576 632	3.141 592 411	3.141 592 648

Das Resultat ist verblüffend genau im Vergleich zu den exakten zehn Stellen von $\pi \doteq 3.141\,592\,653\,5$.

Wie in (10.14) tritt in den Anwendungen sehr häufig der Fall auf, dass in der Potenzreihenentwicklung einer berechenbaren Größe $T(h)$ nur **gerade Potenzen** h^{2k} vorkommen. Aus praktischen Gründen lässt man dann am besten die Stützstellen $h_0 > h_1 > \cdots > h_n > 0$ eine **geometrische Folge** mit Quotienten $q := \frac{1}{2}$ bilden. Die Parameterwerte h_j^2 bilden somit eine geometrische Folge mit Quotienten $q^2 = \frac{1}{4}$, so dass die Rekursionsformeln (10.12) in der speziellen Form

$$\boxed{p_j^{(k)} = p_j^{(k-1)} + \frac{1}{4^k - 1} \cdot [p_j^{(k-1)} - p_{j-1}^{(k-1)}]} \tag{10.15}$$

vorliegen. Hier ist der Faktor $1/(4^k - 1)$ in der k–ten Spalte konstant. Man nennt diesen Spezialfall des NEVILLE–Schemas zur Extrapolation auch das ROMBERG–**Schema**.

BSP. (7.10.5) Für die Funktion $f(x) := e^{2x} \cdot \sin x$ soll durch numerische Differentiation die Ableitung $f'(x_1)$ bei $x_1 := 0.5$ unter Verwendung des zentralen Differenzenquotienten $T(h) := \frac{1}{2h} \cdot (y_2 - y_0) = \frac{1}{2h} \cdot [f(x_1 + h) - f(x_1 - h)]$ berechnet werden. Gemäß (10.6) treten in der TAYLOR–Entwicklung von $T(h)$ nur gerade Potenzen von h auf, so dass wir die Ableitung durch Extrapolation nach Null mit Hilfe des ROMBERG–Schemas ermitteln können. Dazu nehmen wir die geometrische Folge $h_0 := 0.5$, $h_1 := 0.25$, $h_2 := 0.125$, $h_3 := 0.0625$, $h_4 := 0.03125$. Die resultierenden Werte des ROMBERG–Schemas lauten dann:

j	h_j	$T(h_j)$	$p_j^{(1)}$	$p_j^{(2)}$	$p_j^{(3)}$	$p_j^{(4)}$
0	0.500 00	6.217 676 312				
1	0.250 00	5.293 985 621	4.986 088 724			
2	0.125 00	5.067 164 784	4.991 557 838	4.991 922 445		
3	0.062 50	5.010 730 996	4.991 919 734	4.991 943 860	4.991 944 200	
4	0.031 25	4.996 639 742	4.991 942 657	4.991 944 185	4.991 944 190	4.991 944 190

Man vergleiche dazu den analytisch ermittelten Wert $f'(0.5) = e^1 \left(2 \sin(0.5) + \cos(0.5)\right) \doteq 4.991\,944\,190\,30$.

Kapitel 8

Integration von Funktionen einer reellen Veränderlichen

8.1 Stammfunktionen und Integration

Die Notwendigkeit einer **Integralrechnung** ist hinreichend motiviert durch die beiden folgenden Fragestellungen, die allerdings zwei Seiten derselben Sache betreffen:

(A) Wie lässt sich der Prozess der Differentiation umkehren, das heißt, wie löst man die Gleichung $F'(x) = f(x)$ bei gegebener Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ nach $F(x)$ auf?

(B) Wie bestimmt man den Flächeninhalt krummlinig berandeter ebener Flächenstücke?

Wir stellen die Diskussion des Problems (A) an den Anfang. *Zum Beispiel* hat die Aufgabe $F'(x) = 1/\cosh^2 x$ eine Lösung $F_0(x) := \tanh x$. Darüber hinaus ist aber auch $F_C(x) := \tanh x + C$ für jede Konstante $C \in \mathbf{R}$ eine Lösung. Diese Feststellung zeigt bereits, dass das Problem (A) in seiner allgemeinen Form **nicht eindeutig** lösbar ist.

Definition 8.1 *Gegeben sei die reellwertige Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$. Eine Funktion $F \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ heiße auf einem Intervall $I \subseteq D(f) \cap D(F)$ eine* **Stammfunktion** *von f, wenn $F'(x) = f(x)\ \forall\, x \in I$ gilt.*

Beispiel: Die Funktion $f(x) := \sin x$ hat auf $I := \mathbf{R}$ eine Stammfunktion $F(x) := -\cos x$.

Die Frage, zu welchen Funktionen f eine Stammfunktion **existiert**, werden wir in befriedigender Weise in Abschnitt 8.3 beantworten. Dort werden wir zeigen, dass zumindest jede *stetige reellwertige* Funktion $f \in C(I)$ eine Stammfunktion besitzt, so dass die Reichhaltigkeit der Klasse der Stammfunktionen gewährleistet ist. Über die **Eindeutigkeit** von Stammfunktionen treffen wir die folgende Aussage:

Satz 8.1 *Auf dem Intervall $I \subseteq \mathbf{R}$ seien F_1, F_2 Stammfunktionen einer gegebenen Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$. Dann gilt $F_2(x) - F_1(x) = C = const\ \forall\, x \in I$.*

Begründung: Wir haben $F_2'(x) - F_1'(x) = f(x) - f(x) = 0\ \forall\, x \in I$, und somit folgt die Behauptung aus Satz 7.13. □

Definition 8.2 *Ist $F(x)$ auf einem Intervall $I \subseteq D(f)$ Stammfunktion der gegebenen Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$, so heiße $F(x)$ auch ein* **unbestimmtes Integral von f auf I**. *Dieses wird nach* G.W. Leibniz *mit dem Symbol*

$$\boxed{F(x) = \int f(x)\,dx, \quad \text{äquivalent} \quad \int f\,dx \ \text{ oder } \int^x f(t)\,dt, \ \ x \in I,} \tag{1.1}$$

bezeichnet. Man nennt die Variable t die **Integrationsvariable**; *sie darf durch jeden anderen Buchstaben ersetzt werden. Gemäß Satz* 8.1 *ist durch*

$$\boxed{\int f(x)\,dx = F_0(x) + C,\ \ C \in \mathbf{R},} \tag{1.2}$$

die Gesamtheit aller unbestimmten Integrale von f auf I festgelegt, wenn $F_0(x)$ nur eine Stammfunktion von f ist. Man nennt (1.2) *auch* **das unbestimmte Integral von f auf I schlechthin**.

Bemerkung 8.1 (a) Die unbestimmte Integration ist also die Umkehroperation der Differentiation:

$$\boxed{\frac{d}{dx}\Big(\int f(x)\,dx\Big) = f(x), \quad \int f'(x)\,dx = f(x) + C \ \ \forall\, x \in I.}$$

(b) Man verifiziert die Richtigkeit einer Stammfunktion $F(x) = \int f(x)\,dx$ immer durch Differentiation, das heißt, man prüft die Relation $F'(x) = f(x)\ \forall\ x \in I$ nach. Aus dieser Beziehung ergibt sich sofort: □

Jede Ableitungsformel liefert eine Integrationsformel.

BSP. (8.1.1) Aus den Ableitungsformeln der Elementarfunktionen erhält man ohne Schwierigkeiten die folgende Zusammenstellung von Grundintegralen:

	unbestimmtes Integral	Definitionsbereich
(a)	$\int \lambda\,dx = \lambda\,x + C$	$x \in \mathbf{R},\ \ \lambda \in \mathbf{R}$
(b)	$\int x^p\,dx = \dfrac{x^{p+1}}{p+1} + C$	$\begin{cases} x \in \mathbf{R} & : \ p \in \mathbf{N}, \\ x \in (-\infty, 0) \ \text{ oder } \ x \in (0, +\infty) & : \ p = -2, -3, -4, \ldots, \\ x \in (0, +\infty) & : \ p \in \mathbf{R} \setminus \{-1\} \ \text{ sonst} \end{cases}$
(c)	$\int \dfrac{dx}{x} = \ln\lvert x\rvert + C$	$x \in (-\infty, 0) \ \text{ oder } \ x \in (0, +\infty)$

Bemerkung 8.2 Spezialfälle der Formel (b), jeweils auf geeigneten Intervallen, sind:

$$\boxed{\int \frac{dx}{x^2} = -\frac{1}{x} + C, \qquad \int \sqrt{x}\,dx = \frac{2}{3}\,x^{3/2} + C, \qquad \int \frac{dx}{\sqrt{x}} = 2\sqrt{x} + C.}$$

In der Formel (c) darf der Betrag beim Logarithmus **nicht** vergessen werden. Denn für $x < 0$ gilt ja nach der Kettenregel: □

$$\big[\ln|x|\big]' = \big[\ln(-x)\big]' = \frac{-1}{-x} = \frac{1}{x}.$$

	unbestimmtes Integral	Definitionsbereich
(d)	$\int e^x\,dx = e^x + C$	$x \in \mathbf{R}$
(e)	$\int \cos x\,dx = \sin x + C, \quad \int \sin x\,dx = -\cos x + C$	$x \in \mathbf{R}$
(f)	$\int \cosh x\,dx = \sinh x + C, \quad \int \sinh x\,dx = \cosh x + C$	$x \in \mathbf{R}$

Aus den Ableitungsformeln der zyklometrischen Funktionen und der Area–Funktionen erhält man weiterhin:

	unbestimmtes Integral	Definitionsbereich
(g)	$\int \frac{dx}{1+x^2} = \operatorname{arc\,tan}_H x + C$	$x \in \mathbf{R}$
(h)	$\int \frac{dx}{1-x^2} = \frac{1}{2} \ln \left\lvert \frac{1+x}{1-x} \right\rvert + C = \begin{cases} \operatorname{Ar\,coth} x + C, \\ \operatorname{Ar\,tanh} x + C, \end{cases}$	$x \in (-\infty, -1)$ oder $x \in (1, +\infty)$ $x \in (-1, +1)$
(i)	$\int \frac{dx}{\sqrt{1+x^2}} = \operatorname{Ar\,sinh} x + C$	$x \in \mathbf{R}$
(k)	$\int \frac{dx}{\sqrt{1-x^2}} = \operatorname{arc\,sin}_H x + C$	$x \in (-1, +1)$
(l)	$\int \frac{dx}{\sqrt{x^2-1}} = \begin{cases} \operatorname{Ar\,cosh} x + C, \\ -\operatorname{Ar\,cosh}(-x) + C, \end{cases}$	$x \in (1, +\infty)$ $x \in (-\infty, -1)$

Falls die Funktion $f(x)$ für alle $x \in I$ differenzierbar ist und falls $f(x) \neq 0 \; \forall \; x \in I$ gilt, so ist ja $\Big(\ln |f(x)| \Big)' = f'(x)/f(x) \; \forall \; x \in I$. Hieraus erhält man

$$\boxed{\int \frac{f'(x)}{f(x)}\, dx = \ln |f(x)| + C, \;\; x \in I.} \tag{1.3}$$

Mit Hilfe dieser Integrationsregel berechnet man die folgenden unbestimmten Integrale:

	unbestimmtes Integral	Definitionsbereich
(m)	$\int \tan x \, dx = \int \frac{\sin x}{\cos x}\, dx = -\ln |\cos x| + C$	$x \neq (n + \frac{1}{2})\pi, \; n \in \mathbf{Z}$
(n)	$\int \cot x \, dx = \int \frac{\cos x}{\sin x}\, dx = \ln |\sin x| + C$	$x \neq n\pi, \; n \in \mathbf{Z}$
(p)	$\int \tanh x \, dx = \int \frac{\sinh x}{\cosh x}\, dx = \ln \cosh x + C$	$x \in \mathbf{R}$
(q)	$\int \coth x \, dx = \int \frac{\cosh x}{\sinh x}\, dx = \ln |\sinh x| + C$	$x \neq 0$

Verwendet man die Ableitungsformeln von $\tan x$, $\cot x$, $\tanh x$ und $\coth x$, so erhält man die folgenden unbestimmten Integrale:

	unbestimmtes Integral	Definitionsbereich
(r)	$\int \frac{1}{\cos^2 x}\, dx = \tan x + C$	$x \neq (n + \frac{1}{2})\pi, \; n \in \mathbf{Z}$
(s)	$\int \frac{1}{\sin^2 x}\, dx = -\cot x + C$	$x \neq n\pi, \; n \in \mathbf{Z}$

	unbestimmtes Integral	Definitionsbereich
(t)	$\int \frac{1}{\cosh^2 x}\, dx = \tanh x + C$	$x \in \mathbf{R}$
(u)	$\int \frac{1}{\sinh^2 x}\, dx = -\coth x + C$	$x \neq 0$

Verwendet man schließlich noch die Identität

$$\sin x = 2 \sin \frac{x}{2} \cos \frac{x}{2} = 2 \tan \frac{x}{2} \cos^2 \frac{x}{2} =: \frac{f(x)}{f'(x)}, \;\; f(x) := \tan \frac{x}{2},$$

und eine analoge Identität für $\sinh x$, so erhält man aus der Regel (1.3) die folgenden unbestimmten Integrale:

	unbestimmtes Integral	Definitionsbereich
(v)	$\int \frac{1}{\sin x}\,dx = \ln\lvert\tan\frac{x}{2}\rvert + C$	$x \neq n\pi,\ n \in \mathbf{Z}$
(w)	$\int \frac{1}{\sinh x}\,dx = \ln\lvert\tanh\frac{x}{2}\rvert + C$	$x \neq 0$

Weitere unbestimmte Integrale findet man in den gängigen Tafelwerken und Formelsammlungen.

Bemerkung 8.3 (a) Die Differentiation von Elementarfunktionen führt stets wieder auf eine Elementarfunktion. Hingegen gibt es Elementarfunktionen, deren Stammfunktion **keine** Elementarfunktion ist. Dies trifft zum Beispiel auf die folgenden unbestimmten Integrale zu, die nicht mehr auf Elementarfunktionen zurückgeführt werden können:

$$\int \frac{e^x}{x}\,dx, \quad \int \frac{\sin x}{x}\,dx, \quad \int e^{\lambda x^2}\,dx \ \text{ für } \ \lambda \neq 0.$$

(b) Die Stammfunktionen nicht aller algebraischen Verknüpfungen von Elementarfunktionen sind bereits formelmäßig erfasst. Deshalb verschafft man sich **Integrationstechniken**, mit deren Hilfe zahlreiche Integrale auf solche Integrale zurückgeführt werden, die bereits formelmäßig bekannt sind. Darüber hinaus gewinnt man eine Einsicht, wie die gängigen Integraltafeln entstanden sind. Wir werden solche Techniken in Abschnitt 8.2 zusammenstellen. □

Ist $F(x)$ auf dem Intervall I eine Stammfunktion der gegebenen Funktion f, so muss $F(x)$ wegen Satz 7.1 notwendig in jedem Punkt $x \in I$ **stetig** sein. Mit dieser Feststellung treffen wir die folgende

Definition 8.3 *Es sei $F(x)$ auf dem Intervall I eine Stammfunktion der gegebenen Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$, und es gelte $[a,b] \subseteq I$. Dann heiße*

$$\boxed{\int_a^b f(x)\,dx := F(b) - F(a) =: \Big[F(x)\Big]_a^b =: F(x)\Big|_a^b} \tag{1.4}$$

das **bestimmte Integral von f über $[a,b]$**. *Die Punkte a und b heißen* **untere** *bzw.* **obere Integrationsgrenze**. *Die Funktion f heiße der* **Integrand**.

Beispiel: Es gilt $\int_{-1}^{2} x^3\,dx = \frac{1}{4}\,x^4\Big|_{-1}^{2} = \frac{1}{4}\,(16-1) = \frac{15}{4}$.

Bemerkung 8.4 (a) Die Definition 8.3 des bestimmten Integrals ist *unabhängig* von der speziellen Wahl der Stammfunktion $F(x)$. Ist $F_0(x)$ auf dem Intervall I eine andere Stammfunktion der Funktion f, so muss wegen Satz 8.1 $F_0(x) = F(x) + C$ gelten. Dies führt auf $F_0(b) - F_0(a) = F(b) - F(a)$, unabhängig von der Konstanten C.

(b) Es ist wichtig, dass das Intervall $[a,b]$ nicht über I hinausgeht, wenn $F'(x) = f(x)\ \forall\, x \in I$ gilt. Die folgenden Beziehungen sind wegen Nichtbeachtung dieser Regeln **falsch**:

(i) $\int_{-1}^{2} \frac{dx}{\cos^2 x} = \tan x\Big|_{-1}^{2} = \tan 2 - \tan 1$. Dies ist **falsch**, weil die Funktion $\tan x$ an der Stelle $x = \frac{\pi}{2} \in [-1,2]$ **unstetig** ist.

(ii) $\int\limits_{-1}^{1} \frac{\operatorname{sign} x}{\sqrt{|x|}}\,dx = 2\sqrt{|x|}\Big|_{-1}^{1} = 2-2 = 0$. Dies ist **falsch**, weil die Funktion $\sqrt{|x|}$ an der Stelle $x = 0 \in [-1,+1]$ zwar stetig ist, dort aber eine Spitze hat, so dass sie bei $x = 0$ **nicht differenzierbar** ist. □

Aus der Beziehung (1.4) resultieren einige offensichtliche Eigenschaften des bestimmten Integrals:

Satz 8.2 *Es sei $F(x)$ auf dem Intervall I eine Stammfunktion der gegebenen Funktion $f \in$* Abb $(\mathbf{R}, \mathbf{R})$.

(a) *Es gelten die folgenden Regeln:*

$$\boxed{\begin{aligned} &\int_a^a f(x)\,dx = 0 \ \ \forall\, a \in I, \quad \int_a^b f(x)\,dx = -\int_b^a f(x)\,dx \ \ \forall\, a,b \in I, \\ &\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx \ \ \forall\, a,b,c \in I, \\ &\frac{d}{dx}\int_a^x f(t)\,dt = f(x) \ \ \forall\, a,x \in I, \quad \frac{d}{dx}\int_x^b f(t)\,dt = -f(x) \ \ \forall\, x,b \in I. \end{aligned}} \tag{1.5}$$

(b) *Die* **Anfangswertaufgabe:** *Finde eine Funktion $y : I \to \mathbf{R}$ mit $y' = f(x)$ für $x \in I$ und $y(a) = y_0 \in \mathbf{R}$, hat unter der Bedingung $a \in I$ die Lösung*

$$y(x) = y_0 + \int_a^x f(t)\,dt, \quad x \in I.$$

(c) *Es sei $I := [-x_0, +x_0]$ ein* **symmetrisches** *Intervall. Dann gelten die folgenden Implikationen*

$$\boxed{\begin{aligned} f(-x) = -f(x) \quad &\Rightarrow \quad \int_{-a}^{a} f(x)\,dx = 0, \\ f(-x) = f(x) \quad &\Rightarrow \quad \int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx, \end{aligned} \quad \forall\, a \in I.} \tag{1.6}$$

Begründung: Wir zeigen nur die Eigenschaft (1.6). In der Tat, wir haben $\Big(F(-x)\Big)' = -F'(-x) = -f(-x) = f(x) = F'(x)$. Also ist auch $F(-x)$ eine Stammfunktion von f, und es muss deshalb $F(-x) = F(x) + C \ \forall\ x \in I$ gelten. Speziell für $0 \in I$ folgt daraus $F(0) = F(0) + C$, und somit $F(-x) = F(x)$. Hiermit gilt

$$\int_{-a}^{a} f(x)\,dx = F(a) - F(-a) = 0.$$

Die zweite Relation zeigt man ganz analog. □

8.2 Integrationsregeln

Die Integrationsformeln in Abschnitt 8.1 wurden aus bekannten Ableitungsformeln gewonnen. Wir zeigen in diesem Abschnitt, wie die allgemeinen Ableitungsregeln auf korrespondierende Integrationsregeln führen.

Satz 8.3 (a) **Linearität des Integrals:** *Haben die Funktionen f und g auf dem Intervall $I \subset \mathbf{R}$ Stammfunktionen $F(x)$ bzw. $G(x)$, so ist die Funktion $\lambda F(x) + \mu G(x)$ auf I eine Stammfunktion von $\lambda f(x) + \mu g(x)$, $\lambda, \mu \in \mathbf{R}$:*

$$\boxed{\int \Big(\lambda f(x) + \mu g(x)\Big)\, dx = \lambda \int f(x)\, dx + \mu \int g(x)\, dx \;\; \forall\, \lambda, \mu \in \mathbf{R}.} \tag{2.1}$$

Eine entsprechende Formel gilt für bestimmte Integrale $\int_a^b \cdots dx$, $a, b \in I$.

(b) **Partielle Integration:** *Sind f und g im Intervall I differenzierbar und hat die Funktion $f'g$ eine Stammfunktion $H(x)$, so ist $fg - H$ eine Stammfunktion von fg':*

$$\boxed{\begin{aligned} &\int f(x)g'(x)\, dx = f(x)g(x) - \int f'(x)g(x)\, dx, \\ &\int_a^b f(x)g'(x)\, dx = f(x)g(x)\Big|_a^b - \int_a^b f'(x)g(x)\, dx \;\; \forall\, a, b \in I. \end{aligned}} \tag{2.2}$$

(c) **Substitutionsregel:** *Hat die Funktion f auf dem Intervall $I \subseteq D(f)$ eine Stammfunktion $F(x)$ und ist die Funktion $g : I_0 \to I$ differenzierbar, so ist $F \circ g$ auf dem Intervall I_0 eine Stammfunktion von $(f \circ g) \cdot g'$:*

$$\boxed{\begin{aligned} &\int_{g(a)}^{g(x)} f(u)\, du = \int_a^x f[g(t)]\, g'(t)\, dt \;\; \forall\, x \in I_0, \\ &\int_{g(a)}^{g(b)} f(u)\, du = \int_a^b f[g(t)]\, g'(t)\, dt \;\; \forall\, a, b \in I_0. \end{aligned}} \tag{2.3}$$

Gilt darüber hinaus $g \in C^1(I_0)$ sowie $g'(t) \neq 0 \;\forall\, t \in I_0$, so besitzt die Funktion g eine Inverse, und es gilt

$$\boxed{\int_a^b f(x)\, dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f[g(t)]\, g'(t)\, dt \;\; \forall\, a, b \in I.} \tag{2.4}$$

Begründungen: (a) Die Linearität des Integrals ergibt sich sofort aus der Summenregel der Differentiation (Satz 7.2).

(b) Mit der Produktregel aus Satz 7.2 findet man:

$$(fg - H)'(x) = f'(x)g(x) + f(x)g'(x) - f'(x)g(x) = f(x)g'(x).$$

(c) Aus der Kettenregel der Differentiation (Satz 7.3) folgt:

$$\Big(F \circ g\Big)'(x) = F'\Big(g(x)\Big) \cdot g'(x) = f\Big(g(x)\Big) \cdot g'(x) = \Big(f \circ g\Big)(x) \cdot g'(x).$$

BSP. (8.2.1) **Anwendung der Linearität.** Die folgenden Integrale erhält man aus der Linearitätsaussage des Satzes 8.3.a:

$$\begin{aligned} \int \cos^2 x\, dx &= \int \frac{1}{2}\,(1 + \cos 2x)\, dx = \frac{x}{2} + \frac{1}{4}\sin 2x + C \;\; \forall\, x \in \mathbf{R}, \\ \int \sin^2 x\, dx &= \int \frac{1}{2}\,(1 - \cos 2x)\, dx = \frac{x}{2} - \frac{1}{4}\sin 2x + C \;\; \forall\, x \in \mathbf{R}, \\ \int \Big(\sum_{k=0}^{n} a_k x^k\Big)\, dx &= \sum_{k=0}^{n} a_k \frac{x^{k+1}}{k+1} + C \;\; \forall\, x \in \mathbf{R}. \end{aligned}$$

BSP. (8.2.2) **Anwendung der partiellen Integration.** Die Formel (2.2) der partiellen Integration nimmt im Sonderfall $g(x) := x$ die folgende Form an:

$$\boxed{\int f(x)\,dx = x\,f(x) - \int x\,f'(x)\,dx.} \tag{2.5}$$

Man verwendet (2.5), wenn entweder das Integral $\int x\,f'(x)\,dx$ bekannt ist oder das Integral $\int f(x)\,dx$. In den folgenden Beispielen wird die jeweilige Wahl so getroffen, dass links das zu berechnende Integral steht.

$$\int \underbrace{\ln x}_{=:f(x)}\,dx = x\,\ln x - \int x\frac{1}{x}\,dx = x\,\ln x - x + C \;\;\forall\, x > 0,$$

$$\int \underbrace{x\,e^x}_{=:x\,f'(x)}\,dx = x\,e^x - \int e^x\,dx = (x-1)\,e^x + C \;\;\forall\, x \in \mathbf{R},$$

$$\int \underbrace{\operatorname{arc\,tan}_H x}_{=:f(x)}\,dx = x \operatorname{arc\,tan}_H x - \frac{1}{2}\int \underbrace{\frac{2x}{1+x^2}}_{=:h'(x)/h(x)}\,,dx = x \operatorname{arc\,tan}_H x - \frac{1}{2}\,\ln(1+x^2) + C \;\;\forall\, x \in \mathbf{R},$$

$$\int \underbrace{\operatorname{arc\,sin}_H x}_{=:f(x)}\,dx = x \operatorname{arc\,sin}_H x - \int \underbrace{\frac{x}{\sqrt{1-x^2}}}_{=-(\sqrt{1-x^2})'}\,,dx = x \operatorname{arc\,sin}_H x + \sqrt{1-x^2} + C \;\;\forall\, x \in (-1,+1).$$

BSP. (8.2.3) **Rekursionsformeln.** Wir verwenden nochmals die Regel (2.5) zur Berechnung des folgenden Integrals:

$$I_n(x) := \int \underbrace{\Big(\ln x\Big)^n}_{=:f(x)}\,dx = x\,\Big(\ln x\Big)^n - n\int \Big(\ln x\Big)^{n-1}\,dx = x\,\Big(\ln x\Big)^n - n\cdot I_{n-1}(x), \;\; x > 0,\; n \geq 2.$$

Wir haben eine **Rekursionsformel** gefunden, die es erlaubt, das Integral $I_n(x)$, $n \geq 2$, sukzessive auf das bereits bekannte Integral $I_1(x) := x\,(\ln x - 1) + C$ zurückzuführen. Zum Beispiel erhält man für $n = 3$:

$$I_3(x) = \int \Big(\ln x\Big)^3\,dx = x\,\Big(\ln x\Big)^3 - 3x\,\Big(\ln x\Big)^2 + 6x\,\ln x - 6x + C.$$

Wie in diesem Beispiel, gelingt es häufig auch in anderen Fällen, eine **Rekursionsformel** durch **ein– oder mehrfache** partielle Integration zu erstellen. Dabei können verschiedene Fälle auftreten, die wir hier schematisch andeuten wollen. Es sei zum Beispiel das Integral $I_n(x) := \int f_n(x)\,dx$, $n \in \mathbf{N}$, auszuwerten. Durch p–fache partielle Integration können dabei folgende Resultate entstehen:

$$\boxed{I_n(x) \overset{\text{part.Int.}}{\longrightarrow} I_{n-p}(x)} \quad \text{oder} \quad \boxed{I_n(x) \overset{\text{part.Int.}}{\longrightarrow} I_{n+p}(x)} \quad \text{oder} \quad \boxed{I_n(x) \overset{\text{part.Int.}}{\longrightarrow} I_n(x).}$$

Im mittleren Fall löst man nach $I_{n+p}(x)$ auf, im letzten Fall nach $I_n(x)$, sofern dies möglich ist. Wir wenden zum Beispiel die Regel (2.2) der partiellen Integration einmal auf den Integranden $f(x)\cdot g'(x) := x^n \cdot e^{\alpha x}$, $\alpha \neq 0$, an:

$$I_n(x) := \int x^n \cdot e^{\alpha x}\,dx = \frac{x^n}{\alpha}\,e^{\alpha x} - \frac{n}{\alpha}\int x^{n-1}\cdot e^{\alpha x}\,dx = \frac{x^n}{\alpha}\,e^{\alpha x} - \frac{n}{\alpha}\,I_{n-1}(x).$$

Es gilt insbesondere $I_0(x) = \frac{1}{\alpha}\,e^{\alpha x} + C$, und mit obiger Rekursionsformel zeigt man durch vollständige Induktion nach n:

$$\boxed{\int x^n\,e^{\alpha x}\,dx = \frac{1}{\alpha}\,e^{\alpha x}\,\sum_{k=0}^{n}(-1)^k\,\frac{n!}{(n-k)!}\,\frac{1}{\alpha^k}\,x^{n-k} + C, \;\; x \in \mathbf{R},\; n \in \mathbf{N}_0,\; \alpha \neq 0.}$$

Mit Hilfe dieses Resultats gewinnt man auch formelmäßige Ausdrücke für die Integrale $\int x^n \sinh \alpha x\,dx$ und $\int x^n \cosh \alpha x\,dx$. Dabei sind die Relationen

$$\sinh \alpha x = \frac{1}{2}\left(e^{\alpha x} - e^{-\alpha x}\right), \quad \cosh \alpha x = \frac{1}{2}\left(e^{\alpha x} + e^{-\alpha x}\right)$$

anzuwenden. Mit den beiden folgenden Beispielen schließen wir die Technik der Rekursionsformeln ab:

BSP. (8.2.4) Beim folgenden Integral verwenden wir die Regel (2.5) der partiellen Integration:

$$I_n(x) := \int \underbrace{(1+x^2)^{-n}}_{=:f(x)}\,dx = \frac{x}{(1+x^2)^n} + 2n\int \frac{x^2+1-1}{(1+x^2)^{n+1}}\,dx = \frac{x}{(1+x^2)^n} + 2n\,I_n(x) - 2n\,I_{n+1}(x).$$

Hier löst man nach $I_{n+1}(x)$ auf und erhält so die folgende Rekursionsformel:

$$\boxed{\begin{array}{lcl} I_{n+1}(x) & := & \displaystyle\int \frac{dx}{(1+x^2)^{n+1}} = \frac{1}{2n}\frac{x}{(1+x^2)^n} + \frac{2n-1}{2n}\,I_n(x), \quad x\in\mathbf{R},\ n\in\mathbf{N}, \\ I_1(x) & := & \displaystyle\int \frac{dx}{1+x^2} = \operatorname{arc\,tan}_H\, x + C, \quad x\in\mathbf{R}. \end{array}}$$

BSP. (8.2.5) In dem folgenden Integral $F(x)$ wenden wir die Regel (2.2) der partiellen Integration zweimal an, und zwar setzen wir jeweils $g'(x) := e^{ax}$. Nach der zweiten partiellen Integration tritt $F(x)$ wieder auf:

$$F(x) := \int e^{ax}\,\sin bx\,dx \overset{\text{part.Int.}}{=} \frac{e^{ax}}{a}\,\sin bx - \frac{b}{a}\int e^{ax}\,\cos bx\,dx \overset{\text{part.Int.}}{=} \frac{e^{ax}}{a}\,\sin bx - \frac{be^{ax}}{a^2}\,\cos bx - \Big(\frac{b}{a}\Big)^2 F(x).$$

Durch Auflösen nach $F(x)$ erhält man die gesuchte Integralformel. Ganz analog verfährt man, wenn anstelle von $\sin bx$ die Funktion $\cos bx$ im Integranden steht:

$$\boxed{\begin{array}{lcl} \displaystyle\int e^{ax}\,\sin bx\,dx & = & \displaystyle\frac{a\,\sin bx - b\,\cos bx}{a^2+b^2}\,e^{ax} + C, \quad x\in\mathbf{R},\ a^2+b^2\neq 0, \\ \displaystyle\int e^{ax}\,\cos bx\,dx & = & \displaystyle\frac{a\,\cos bx + b\,\sin bx}{a^2+b^2}\,e^{ax} + C, \quad x\in\mathbf{R},\ a^2+b^2\neq 0. \end{array}}$$

BSP. (8.2.6) **Anwendung der Substitutionsregel.** Hat man insbesondere die Funktion $f(u) := u^p$ vorliegen, so resultiert aus (2.3) der folgende Sonderfall der Substitutionsregel:

$$\boxed{\begin{array}{lcl} \displaystyle\int^x \Big(g(t)\Big)^p g'(t)\,dt & = & \displaystyle\int^{g(x)} u^p\,du = \frac{1}{1+p}\Big(g(x)\Big)^{p+1} + C, \quad p\neq -1, \\ \displaystyle\int^x \frac{g'(t)}{g(t)}\,dt & = & \displaystyle\int^{g(x)} \frac{du}{u} = \ln|g(x)| + C. \end{array}} \tag{2.6}$$

Hierbei muss die Funktion g natürlich die Voraussetzungen zum Satz 8.3.c erfüllen. In dem folgenden Beispiel setzen wir $g(t) := \operatorname{arc\,tan}_H\, x$:

$$\int^x \Big(\operatorname{arc\,tan}_H\, t\Big)^p \frac{dt}{1+t^2} = \begin{cases} \dfrac{1}{p+1}\Big(\operatorname{arc\,tan}_H\, x\Big)^{p+1} + C & : \ p\neq -1, \\ \ln|\operatorname{arc\,tan}_H\, x| + C & : \ p=-1, \end{cases} \qquad x\in\mathbf{R}.$$

Ganz analog erhält man für $g(x) := x^3 - 3x^2 + 5x + a$ auf jedem Intervall $I\subset\mathbf{R}$, welches keine Nullstelle der Funktion g enthält:

$$\int \frac{3x^2-6x+5}{(x^3-3x^2+5x+a)^p}\,dx = \begin{cases} \dfrac{1}{1-p}\,\dfrac{1}{(x^3-3x^2+5x+a)^{p-1}} + C & : \ p\neq 1, \\ \ln|x^3-3x^2+5x+a| + C & : \ p=1, \end{cases} \qquad x\in I.$$

In dem folgenden Beispiel sind $g(x) := \sin x$ und $p := -\frac{1}{2}$ gewählt worden:

$$\int \frac{\cos x}{\sqrt{\sin x}}\,dx = 2\sqrt{\sin x} + C, \quad x\in(0,\pi).$$

Für die richtige Anwendung der Substitutionsregel bedarf es oft einer gewissen Erfahrung und einer Portion Fingerspitzengefühl. Generell kann gesagt werden, dass man das unbestimmte Integral $\int f(x)\,dx$ mit einer solchen Substitution $u = g(x)$ berechnet, deren Ableitung $g'(x)$ als Faktor von $f(x)$ auftritt. Werden Substitutionen vom Typ $x = g^{-1}(u)$ verwendet, so ist sorgfältig auf die Bijektivität der Abbildung $u = g(x)$ zu achten. Für $g \in C^1(I_0)$ muss deshalb $g'(x) \neq 0\ \forall\ x \in I_0$ gelten.

BSP. (8.2.7) Im folgenden Integral ist die Substitution $u = g(x) := \frac{b}{a}\,x,\ du = \frac{b}{a}\,dx$, erfolgreich:

$$\int \frac{dx}{a^2 + b^2x^2} = \frac{1}{a^2} \int \frac{dx}{1 + (bx/a)^2} = \frac{1}{ab} \int^{bx/a} \frac{du}{1+u^2} = \frac{1}{ab}\,\text{arc tan}_H\,\big(\frac{bx}{a}\big) + C,\quad x \in \mathbf{R},\ ab \neq 0.$$

BSP. (8.2.8) Im folgenden Integral verwenden wir die Substitution $x = g^{-1}(u) := \sin u,\ dx = \cos u\ du,\ x \in [-1,+1]$. Für $u \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$ ist die erforderliche Bijektivität gewährleistet, und es gilt $u = \text{arc sin}_H\,x$:

$$\begin{aligned}\int \sqrt{1-x^2}\,dx &= \int^{\text{arc sin}_H\,x} \cos^2 u\,du \overset{\text{BSP. (8.2.1)}}{=} \frac{u}{2} + \frac{1}{2}\sin u\cos u + C \\ &= \frac{1}{2}\,\text{arc sin}_H\,x + \frac{x}{2}\sqrt{1-x^2} + C,\quad x \in [-1,+1].\end{aligned}$$

BSP. (8.2.9) Im folgenden Integral ist die Substitution $u = g(x) := x^2,\ du = 2x\,dx$, erfolgreich:

$$\int x^5 e^{-x^2}\,dx = \frac{1}{2}\int^{x^2} u^2 e^{-u}\,du \overset{\text{BSP. (8.2.3)}}{=} -\frac{1}{2}\,e^{-u}\left(u^2+2u+2\right) + C = -\frac{1}{2}\,e^{-x^2}\left(x^4+2x^2+2\right) + C,\quad x \in \mathbf{R}.$$

BSP. (8.2.10) Im folgenden Integral verwenden wir die Zerlegung $\cos x = \cos^2(x/2) - \sin^2(x/2) = \cos^2(x/2)\left(1 - \tan^2(x/2)\right)$. Nun führt die Substitution $u = g(x) := \tan(x/2),\ du = dx/\left(2\cos^2(x/2)\right),\ x \in (-\frac{\pi}{2}, +\frac{\pi}{2})$, zum Ziel:

$$\begin{aligned}\int \frac{dx}{\cos x} &= \int^{\tan(x/2)} \frac{2\,du}{1-u^2} = \int^{\tan(x/2)} \Big(\frac{1}{1+u} + \frac{1}{1-u}\Big)\,du = \ln\Big|\frac{1+u}{1-u}\Big| + C \\ &= \ln\Big|\frac{1+\tan(x/2)}{1-\tan(x/2)}\Big| + C,\quad x \in \Big(-\frac{\pi}{2}, +\frac{\pi}{2}\Big).\end{aligned}$$

Der sorgfältige Umgang mit der Substitutionsregel trifft in gleicher Weise auf bestimmte Integrale zu, wie das folgende Beispiel lehrt:

BSP. (8.2.11) Man berechne das bestimmte Integral

$$F(x) := \int_0^{\sin^2 x} \text{arc sin}_H\,\sqrt{t}\,dt.$$

Bei der naheliegenden Substitution $t = g^{-1}(u) := \sin^2 u,\ dt = 2\,\sin u\cos u\,du = \sin 2u\,du$ ist wiederum auf die Bijektiviät zu achten. Die Transformation $u = g(t) = \text{arc sin}_H\,\sqrt{t}$ bildet offensichtlich das Intervall $I_0 := [0,1]$ auf das Intervall $I := [0, \frac{\pi}{2}]$ ab. Also darf die Variable x nur das Intervall I durchlaufen. Wir erhalten durch partielle Integration:

$$F(x) = \int_0^x u\,\sin 2u\,du = -\frac{u}{2}\,\cos 2u\,\Big|_0^x + \frac{1}{2}\int_0^x \cos 2u\,du = -\frac{x}{2}\,\cos 2x + \frac{1}{4}\,\sin 2x,\quad x \in I.$$

Wegen $\sin^2(x+\pi) = \sin^2 x,\ x \in \mathbf{R}$, und $\sin^2(\pi - x) = \sin^2 x,\ x \in [\frac{\pi}{2}, \pi]$, resultiert somit für alle Werte $x \in \mathbf{R}$:

$$F(x) = \begin{cases} \frac{1}{4}\sin 2x - \frac{x}{2}\cos 2x & : \; x \in [0, \frac{\pi}{2}], \\ F(\pi - x) = \frac{x-\pi}{2}\cos 2x - \frac{1}{4}\sin 2x & : \; x \in [\frac{\pi}{2}, \pi], \\ F(x+\pi) & : \text{ periodische Fortsetzung über } [0,\pi] \text{ hinaus.} \end{cases}$$

Wir nehmen an, die Funktion $y = f(x)$ besitze eine differenzierbare Umkehrfunktion $x = f^{-1}(y)$. Setzt man diese anstelle von f in die Formel (2.5) ein, so ergibt sich

$$\int f^{-1}(y)\,dy = y\,f^{-1}(y) - \int y\,(f^{-1})'(y)\,dy.$$

Ist $f'(x) \neq 0$, so gilt wegen Satz 7.4 $(f^{-1})'(y) = 1/f'(x)$, und die Substitution $y = f(x),\ dy = f'(x)\,dx$ im Integral auf der rechten Seite führt zu folgendem Ergebnis:

Satz 8.4 (Integration der Umkehrfunktion)
Die Funktion $f : I \to \mathbf{R}$ sei differenzierbar, und es gelte $f'(x) \neq 0\ \forall\ x \in I$. Hat f auf dem Intervall I eine Stammfunktion $F(x)$, so hat auch f^{-1} auf dem Intervall $f(I)$ eine Stammfunktion, und es gilt

$$\boxed{\int f^{-1}(y)\,dy = y\,f^{-1}(y) - \int^{f^{-1}(y)} f(x)\,dx, \quad y \in f(I).} \tag{2.7}$$

BSP. (8.2.12) Die Funktion $f(x) := \cos x,\ x \in I := (0,\pi)$, erfüllt die Voraussetzungen des Satzes 8.4. Die Umkehrfunktion $f^{-1}(y) = \text{arc cos}_H\, y$ hat also auf dem Intervall $f(I) = (-1,+1)$ eine Stammfunktion, und es gilt gemäß (2.7)

$$\begin{aligned} \int \text{arc cos}_H\, y\,dy &= y\,\text{arc cos}_H\, y - \int^{\text{arc cos}_H\, y} \cos x\,dx = y\,\text{arc cos}_H\, y - \sin(\text{arc cos}_H\, y) + C \\ &= y\,\text{arc cos}_H\, y - \sqrt{1-\cos^2 x}\big|_{x=\text{arc cos}_H\, y} + C \\ &= y\,\text{arc cos}_H\, y - \sqrt{1-y^2} + C, \quad y \in (-1,+1). \end{aligned}$$

Integration komplexwertiger Funktionen. Zerlegt man eine *komplexwertige* Funktion $f \in \text{Abb}\,(\mathbf{R},\mathbf{C})$ gemäß $f(x) = u(x) + i\,v(x)$ in ihren Real– und Imaginärteil, und haben die Funktionen u, v auf dem Intervall I Stammfunktionen $U(x)$ bzw. $V(x)$, so ist offenbar durch $F(x) := U(x) + i\,V(x)$ auf I eine Stammfunktion von f definiert. Das heißt, es gilt stets

$$\boxed{f(x) := u(x) + i\,v(x) \Rightarrow \int f(x)\,dx = \int u(x)\,dx + i\int v(x)\,dx,} \tag{2.8}$$

sofern die Integrale über u und v existieren. Aus dieser Beziehung resultieren:

$$\boxed{\text{Re}\int f(x)\,dx = \int \text{Re}\,(f(x))\,dx, \quad \text{Im}\int f(x)\,dx = \int \text{Im}\,(f(x))\,dx.} \tag{2.9}$$

BSP. (8.2.13) Für $a \in \mathbf{R}$ gilt $e^{iax} = \cos ax + i\,\sin ax = i\,(\sin ax - i\,\cos ax)$. Hiermit folgt sofort:

$$\int e^{iax}\,dx = \int \cos ax\,dx + i\int \sin ax\,dx = \frac{1}{a}\,(\sin ax - i\,\cos ax) + C = \frac{1}{ia}\,e^{iax} + C, \quad x \in \mathbf{R},\ a \neq 0.$$

Allgemeiner gilt

$$\boxed{\int e^{\lambda x}\,dx = \frac{1}{\lambda}\,e^{\lambda x} + C \quad \forall\, x \in \mathbf{R},\ 0 \neq \lambda \in \mathbf{C}.}$$

Mit Hilfe der Substitution $u = g(x) := \ln x,\ du = dx/x,\ x > 0$, kann nun das folgende komplexe Integral berechnet werden:

$$\boxed{\int x^{\lambda}\,dx = \int e^{(\lambda+1)\ln x}\,\frac{dx}{x} = \int^{\ln x} e^{(\lambda+1)u}\,du = \frac{1}{\lambda+1}\,x^{\lambda+1} + C \quad \forall\, x > 0,\ -1 \neq \lambda \in \mathbf{C}.}$$

BSP. (8.2.14) Manchmal wird durch den Umweg über das Komplexe die Berechnung reeller Integrale vereinfacht. In dem folgenden Beispiel sei $\lambda := a + ib,\ a, b \in \mathbf{R},\ a^2 + b^2 > 0$, gesetzt. Wegen $e^{ax}\sin bx = \operatorname{Im} e^{\lambda x}$ erhalten wir unter Verwendung von (2.9)

$$\int e^{ax}\sin bx\,dx = \operatorname{Im}\int e^{\lambda x}\,dx = \frac{1}{|\lambda|^2}\operatorname{Im}\left(\bar{\lambda}\,e^{\lambda x}\right) + C = \frac{a\,\sin bx - b\,\cos bx}{a^2+b^2}\,e^{ax} + C.$$

Partialbruchzerlegung und Integration rationaler Funktionen

Wir werden im folgenden zeigen, dass die Klasse der **rationalen Funktionen** elementar integrierbar ist, das heißt, dass sich das unbestimmte Integral

$$\int \frac{P(x)}{Q(x)}\,dx = \int \frac{a_0 + a_1 x + \cdots + a_m x^m}{b_0 + b_1 x + \cdots + b_n x^n}\,dx$$

stets elementar berechnen lässt. Wir gehen davon aus, dass der Integrand eine **echt gebrochen** rationale Funktion $R(x)$ ist. Es gelte ferner stets $a_k, b_k \in \mathbf{K}$ mit $\mathbf{K} := \mathbf{R}$ oder $\mathbf{K} := \mathbf{C}$.

Voraussetzung: Es sei $R(x) := P(x)/Q(x)$ eine echt gebrochen rationale Funktion:

$$m = \operatorname{Grad} P < \operatorname{Grad} Q = n.$$

Anderfalls muss in einem **Vorbereitungsschritt** mit Hilfe des Euklidischen Teileralgorithmus ein Polynom $T(x)$ so abgespalten werden, dass gilt:

$$\frac{P(x)}{Q(x)} = T(x) + \frac{\tilde{P}(x)}{Q(x)}, \quad \operatorname{Grad}\tilde{P} < \operatorname{Grad} Q.$$

Der Fundamentalsatz der Algebra (Satz 2.10) stellt sicher, dass das Polynom $Q(x)$ vom Grad $n \geq 1$ genau n Nullstellen in $\mathbf{C}$ hat; jede Nullstelle wird ihrer Vielfachheit entsprechend oft gezählt. Mit Hilfe dieses Resultats konnten wir die *Linearfaktorzerlegung* eines Polynoms beweisen (Satz 2.11): Sind $z_1, z_2, \ldots, z_p,\ p \leq n$, die paarweise verschiedenenen (komplexen) Nullstellen des Polynoms $Q(x) := \sum_{k=0}^{n} b_k x^k,\ b_n \neq 0$, und sind $k_1, k_2, \ldots, k_p$ ihre Vielfachheiten, so gilt $k_1 + k_2 + \cdots + k_p = n$, und $Q(x)$ gestattet die Linearfaktorzerlegung

$$Q(x) = b_n(x - z_1)^{k_1}(x - z_2)^{k_2}\cdots(x - z_p)^{k_p} \quad \forall\, x \in \mathbf{R}. \tag{2.10}$$

Mit Hilfe dieser Linearfaktorzerlegung zeigt man:

Satz 8.5 *Gegeben sei die echt gebrochen rationale Funktion* $R(x) := P(x)/Q(x)$, $D(R) := \{x \in \mathbf{R} : Q(x) \neq 0\}$. *Es sei* $z_0 \in \mathbf{C}$ *eine* k*–fache Nullstelle des Nennerpolynoms* $Q(x)$, *so dass gilt:* $Q(x) = (x - z_0)^k \tilde{Q}(x)$ *mit* $\tilde{Q}(z_0) \neq 0$. *Dann existieren ein eindeutig bestimmtes Polynom* $\tilde{P}(x)$ *und eine Konstante* $A \in \mathbf{C}$ *mit*

$$\boxed{\frac{P(x)}{Q(x)} = \frac{A}{(x-z_0)^k} - \frac{\tilde{P}(x)}{(x-z_0)^{k-1}\tilde{Q}(x)} \quad \forall\, x \in D(R).} \tag{2.11}$$

Hierbei ist A *durch die Vorschrift* $A = P(z_0)/\tilde{Q}(z_0)$ *festgelegt.*

Begründung: Im Sinne einer Analyse betrachten wir (2.11) als Ansatz. Setzen wir $Q(x) = (x - z_0)^k \tilde{Q}(x)$ in (2.11) ein, so folgt

$$\lim_{x\to z_0} \frac{(x-z_0)^k\, P(x)}{Q(x)} = \frac{P(z_0)}{\tilde{Q}(z_0)} \overset{(2.11)}{=} A - \lim_{x\to z_0} \frac{(x-z_0)\,\tilde{P}(x)}{\tilde{Q}(x)} = A.$$

Das heißt, die Konstante A ist durch den Ansatz (2.11) in der angegebenen Weise eindeutig festgelegt. Mit diesem Wert von A gilt $\lim\limits_{x\to z_0} \Big(A\tilde{Q}(x) - P(x)\Big) = 0$, so dass das Polynom $A\tilde{Q}(x) - P(x)$ die Nullstelle $x = z_0$ besitzt. Wegen Satz 2.9 gibt es dann ein eindeutig bestimmtes Polynom $\tilde{P}(x)$ mit $A\tilde{Q}(x) - P(x) = (x - z_0)\tilde{P}(x)\ \forall\, x \in \mathbf{R}$. Für $x \in D(R)$ kann diese Gleichung durch $Q(x)$ dividiert werden, und es resultiert (2.11). □

Da die Funktion $\tilde{R}(x) := \tilde{P}(x)/\Big((x-z_0)^{k-1}\tilde{Q}(x)\Big)$ wiederum echt gebrochen rational ist, kann das Abspaltungsverfahren (2.11) erneut auf $\tilde{R}(x)$ angewendet werden. Das Nennerpolynom hat nun bei $z_0 \in \mathbf{C}$ eine Nullstelle der Ordnung $k - 1$. Nach insgesamt k Schritten gelangt man auf diese Weise zu einer echt gebrochen rationalen Funktion, deren Nennerpolynom nun keine Nullstelle z_0 besitzt. Man führt das Verfahren mit der nächsten Nullstelle von $Q(x)$ fort. Nach insgesamt n Schritten hat man das folgende Resultat vorliegen:

Satz 8.6 (Partialbruchzerlegung im Komplexen, PBZ)
Gegeben sei die echt gebrochen rationale Funktion $R(x) := P(x)/Q(x)$, $D(R) := \{x \in \mathbf{R} : Q(x) \neq 0\}$. *Das Nennerpolynom* $Q(x)$ *habe die Linearfaktorzerlegung* (2.10) *mit den paarweise verschiedenen Nullstellen* $z_1, z_2, \ldots, z_p$ *der Vielfachheiten* $k_1, k_2, \ldots, k_p$. *Dann gibt es eindeutig bestimmte Koeffizienten* $A_{jk} \in \mathbf{C}$ *mit*

$$\boxed{\begin{aligned} R(x) &= \sum_{j=1}^{p}\sum_{k=1}^{k_j} \frac{A_{jk}}{(x-z_j)^k} \\ &= \frac{A_{11}}{x-z_1} + \frac{A_{12}}{(x-z_1)^2} + \cdots + \frac{A_{1k_1}}{(x-z_1)^{k_1}} \\ &+ \frac{A_{21}}{x-z_2} + \frac{A_{22}}{(x-z_2)^2} + \cdots + \frac{A_{2k_2}}{(x-z_2)^{k_2}} \\ &\vdots \\ &+ \frac{A_{p1}}{x-z_p} + \frac{A_{p2}}{(x-z_p)^2} + \cdots + \frac{A_{pk_p}}{(x-z_p)^{k_p}}. \end{aligned}} \tag{2.12}$$

Die Darstellung (2.12) *heiße die* **komplexe Partialbruchzerlegung** *der rationalen Funktion* $R(x)$.

Mit diesem Ergebnis ist die Integration $\int R(x)\,dx$ einer allgemeinen rationalen Funktion $R(x) = P(x)/Q(x) =: T(x) + \tilde{P}(x)/Q(x) = T(x) + \tilde{R}(x)$ zurückgeführt auf die Integration eines Polynoms $T(x)$ und die Integration von Partialbrüchen der Form $(x-a)^{-k},\ a \in \mathbf{C},\ k \in \mathbf{N}$, die wir aus der Partialbruchzerlegung (2.12) der echt gebrochen rationalen Funktion $\tilde{R}(x)$ gewinnen. Das unbestimmte Integral $\int R(x)\,dx$ ist also eine Linearkombination von unbestimmten Integralen der Form

$$\boxed{\begin{aligned} I_0(x) &:= \int \sum_{j=0}^{r} c_j x^j\,dx = \sum_{j=0}^{r} \frac{c_j}{j+1}\,x^{j+1} + C, \\ I_1(x) &:= \int \frac{dx}{(x-a)^k} = \begin{cases} -\dfrac{1}{k-1}\,\dfrac{1}{(x-a)^{k-1}} + C & : \ k > 1, \\ \ln|x-a| + C & : \ k = 1. \end{cases} \end{aligned}}$$

Die Hauptarbeit ist somit bei der Berechnung der Nullstellen z_k des Nennerpolynoms und der Koeffizienten A_{jk} der Partialbruchzerlegung (2.12) zu leisten. Für die *Handrechnung* bedient man sich zweier Verfahren zur Bestimmung der A_{jk}. Dabei wird stets die Kenntnis aller Nullstellen des Nennerpolynoms vorausgesetzt.

(A) **Methode des Koeffizientenvergleichs.** Diese Methode ist aufwendig. Die Partialbruchzerlegung (2.12) wird mit unbestimmten Koeffizienten A_{jk} angesetzt. Danach bringt man die Partialbrüche auf den gemeinsamen Nenner (dieser ist das Nennerpolynom $Q(x)$). Im Zähler führe man Koeffizientenvergleich mit dem gegebenen Zählerpolynom $P(x)$ durch. Es resultiert ein lineares Gleichungssystem für die unbekannten Koeffizienten A_{jk}.

BSP. (8.2.15) Die Partialbruchzerlegung der echt gebrochen rationalen Funktion $R(x) := (3+x)/(x^4-x^2)$ ist zu bestimmen. Wegen $Q(x) := x^4 - x^2 = x^2(x+1)(x-1)$ erfordert die PBZ (2.12) den folgenden Ansatz:

$$R(x) = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x+1} + \frac{D}{x-1} = \frac{(A+C+D)x^3 + (B-C+D)x^2 - Ax - B}{x^2(x^2-1)} \stackrel{!}{=} \frac{3+x}{x^4-x^2}.$$

Durch Koeffizientenvergleich der beiden Zählerpolynome erhält man das folgende lineare Gleichungssytem:

$$\begin{array}{lrcrcrcrcl} [x^0]: & & -B & = & 3 & \Rightarrow & B & = & -3, & \\ [x^1]: & & -A & = & 1 & \Rightarrow & A & = & -1, & \\ [x^2]: & B-C+D & = & 0 & \Rightarrow & D-C & = & 3 & \Rightarrow & C=-1, \\ [x^3]: & A+C+D & = & 0 & \Rightarrow & D+C & = & 1 & \Rightarrow & D=2. \end{array}$$

Hieraus resultiert die gesuchte PBZ

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

(B) **Grenzwertmethode.** Diese Methode ist besonders effektiv, wenn alle Nullstellen des Nennerpolynoms $Q(x)$ **einfach** sind. Die Partialbruchzerlegung (2.12) wird wiederum mit unbestimmten Koeffizienten A_{jk} angesetzt. Danach multipliziert man beide Seiten in der Zerlegung (2.12) mit dem Binom $(x-z_j)^{k_j}$ und bildet den Limes $x \to z_j$. Dieser Grenzwert liefert direkt den Koeffizienten A_{jk_j}. Nun wird auf beiden Seiten der Zerlegung (2.12) der schon bekannte Ausdruck $A_{jk_j}/(x-z_j)^{k_j}$ subtrahiert. Multiplikation mit $(x-z_j)^{k_j-1}$ und Grenzwertbildung $x \to z_j$ liefert den Koeffizienten $A_{jk_{j-1}}$ usf.

BSP. (8.2.16) Wir betrachten hier nochmals die gebrochen rationale Funktion $R(x)$ aus BSP. (8.2.15). Der Ansatz

$$R(x) = \frac{3+x}{x^2(x+1)(x-1)} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x+1} + \frac{D}{x-1}$$

führt vermöge der Grenzwertmethode sofort auf die Konstanten B, C und D:

$$\begin{aligned} x^2R(x)\Big|_{x=0} &= \frac{3+x}{(x+1)(x-1)}\Big|_{x=0} = -3 = B, \\ (x+1)\,R(x)\Big|_{x=-1} &= \frac{3+x}{x^2(x-1)}\Big|_{x=-1} = -1 = C, \\ (x-1)\,R(x)\Big|_{x=1} &= \frac{3+x}{x^2(x+1)}\Big|_{x=1} = 2 = D. \end{aligned}$$

Die Berechnung der Konstanten A kann nach dem oben geschilderten Verfahren unter Verwendung der Regel von L'HOSPITAL vorgenommen werden:

$$A = \lim_{x\to 0} x\Big(R(x) + \frac{3}{x^2}\Big) = \lim_{x\to 0} \frac{1}{x} \underbrace{\Big(\frac{3+x}{(x^2-1)} + 3\Big)}_{=:g(x);\ g(0)=0} = \lim_{x\to 0} g'(x) = -1.$$

Weit weniger aufwendig ist allerdings die Bestimmung von A durch **Einsetzen eines speziellen Wertes** x_0, wobei x_0 keine Nullstelle des Nennerpolynoms sein darf. Zum Beispiel gilt für $x_0 = 2$:

$$R(2) = \frac{5}{12} = \Big(\frac{A}{x} - \frac{3}{x^2} - \frac{1}{x+1} + \frac{2}{x-1}\Big)\Big|_{x=2} = \frac{A}{2} + \frac{11}{12}.$$

Auch hier ergibt sich wieder $A = -1$.

BSP. (8.2.17) Man bestimme die Partialbruchzerlegung der echt gebrochen rationalen Funktion $R(x) := (x+1)/(x^4 - x^3 + x^2 - x)$. Das Nennerpolynom $Q(x) := x(x^3 - x^2 + x - 1)$ hat ganz offenkundig die einfache Nullstelle $z_1 = 0$. Eine weitere Nullstelle $z_2 = 1$ kann leicht erraten werden. Wir spalten den Linearfaktor $(x-1)$ unter Verwendung des HORNER–Schemas ab:

	1	−1	1	−1
$z_2 = 1$	*	1	0	1
	1	0	1	0

Somit erhalten wir die Linearfaktorzerlegung $Q(x) = x(x-1)(x^2+1) = x(x-1)(x+i)(x-i)$, die den folgenden Ansatz der PBZ erfordert:

$$R(x) = \frac{x+1}{x(x-1)(x+i)(x-i)} = \frac{A}{x} + \frac{B}{x-1} + \frac{C}{x+i} + \frac{D}{x-i}.$$

Wir bestimmen die Koeffizienten A, B, C, D mit der Grenzwertmethode:

$$\begin{aligned} xR(x)\Big|_{x=0} &= \frac{x+1}{(x-1)(x^2+1)}\Big|_{x=0} = -1 = A, \\ (x-1)\,R(x)\Big|_{x=1} &= \frac{x+1}{x(x^2+1)}\Big|_{x=1} = 1 = B, \\ (x+i)\,R(x)\Big|_{x=-i} &= \frac{x+1}{x(x-1)(x-i)}\Big|_{x=-i} = -i/2 = C, \\ (x-i)\,R(x)\Big|_{x=i} &= \frac{x+1}{x(x-1)(x+i)}\Big|_{x=i} = i/2 = D. \end{aligned}$$

Hieraus resultiert die gesuchte PBZ

$$R(x) = \frac{x+1}{x^4 - x^3 + x^2 - x} = -\frac{1}{x} + \frac{1}{x-1} + \frac{i}{2}\Big(\frac{1}{x-i} - \frac{1}{x+i}\Big).$$

Beachte: Fasst man die beiden letzten Summanden zusammen, so ergibt sich ein **reeller** Partialbruch mit quadratischem Nennerpolynom, nämlich

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

Hiermit gelangt man zur folgenden **reellen Partialbruchzerlegung** der rationalen Funktion $R(x)$:

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

Hat das Nennerpolynom $Q(x)$ der rationalen Funktion $R(x) = P(x)/Q(x)$ ausschließlich **reelle Koeffizienten**, so können gemäß Satz 2.14 komplexe Nullstellen nur als **konjugiert komplexe Paare** auftreten. Mit $z_0 := x_0 + iy_0$ ist auch $\bar{z}_0 = x_0 - iy_0$ eine Nullstelle von $Q(x)$. Hat auch das Zählerpolynom $P(x)$ ausschließlich reelle Koeffizienten, so können Partialbrüche

$$\frac{C}{x-z_0} + \frac{D}{x-\bar{z}_0} = \frac{(C+D)x - (C\bar{z}_0 + Dz_0)}{x^2 - 2x_0x + |z_0|^2} =: \frac{Ax+B}{x^2+\alpha x+\beta}, \quad A,B,\alpha,\beta \in \mathbf{R},$$

stets in **reeller Form** zusammengefasst werden. Sind z_0 und $\bar{z}_0$ jeweils k–fache Nullstellen, so treten in der Partialbruchzerlegung (2.12) Partialbrüche von der Form

$$\frac{A_1x+B_1}{x^2+\alpha x+\beta} + \frac{A_2x+B_2}{(x^2+\alpha x+\beta)^2} + \cdots + \frac{A_kx+B_k}{(x^2+\alpha x+\beta)^k} \tag{2.13}$$

auf. Sämtliche Koeffizienten sind reell. Die quadratischen Faktoren $q(x) := x^2 + \alpha x + \beta$ sind über $\mathbf{R}$ irreduzibel, vgl. Definition 2.15. Wegen

$$q(x) = 0 \quad \Leftrightarrow \quad x = z_\pm := -\frac{\alpha}{2} \pm \frac{i}{2}\sqrt{4\beta - \alpha^2}$$

treten die Terme (2.13) genau dann auf, wenn $\boxed{4\beta > \alpha^2}$ gilt. In diesem Fall spricht man von einer **Partialbruchzerlegung rationaler Funktionen im Reellen**. Mit der Zerlegung (2.13) kann das unbestimmte Integral $\int R(x)\,dx$ im Reellen berechnet werden. Dazu sind die folgenden Teilintegrale auszuwerten:

$$\begin{array}{|lcl|}
\hline
I_2^{(1)}(x) & := & \displaystyle\int \frac{dx}{x^2+\alpha x+\beta} \\
& = & \displaystyle\frac{2}{\sqrt{4\beta-\alpha^2}} \operatorname{arc\,tan}_H \left(\frac{2x+\alpha}{\sqrt{4\beta-\alpha^2}}\right) + C, \\
\hline
I_2^{(k)}(x) & := & \displaystyle\int \frac{dx}{(x^2+\alpha x+\beta)^k} \\
& = & \displaystyle\frac{2x+\alpha}{(k-1)(4\beta-\alpha^2)(x^2+\alpha x+\beta)^{k-1}} + \frac{2(2k-3)}{(k-1)(4\beta-\alpha^2)} I_2^{(k-1)}(x), \quad k \geq 2, \\
\hline
I_3^{(1)}(x) & := & \displaystyle\int \frac{Ax+B}{x^2+\alpha x+\beta}\,dx \\
& = & \displaystyle\frac{A}{2} \ln|x^2+\alpha x+\beta| + \left(B - \frac{A\alpha}{2}\right) I_2^{(1)}(x), \\
\hline
I_3^{(k)}(x) & := & \displaystyle\int \frac{Ax+B}{(x^2+\alpha x+\beta)^k}\,dx \\
& = & \displaystyle\frac{-A}{2(k-1)(x^2+\alpha x+\beta)^{k-1}} + \left(B - \frac{A\alpha}{2}\right) I_2^{(k)}(x), \quad k \geq 2. \\
\hline
\end{array}$$

BSP. (8.2.18) Man berechne im Reellen das unbestimmte Integral $\int R(x)\,dx$ der gebrochen rationalen Funktion $R(x) := (-2x^4 + x^3 - 3x^2 - 4)/(x^5 - x^4 + 2x^3 - 2x^2 + x - 1)$. *Lösung:* Das Nennerpolynom $Q(x) :=$

$x^5 - x^4 + 2x^3 - 2x^2 + x - 1$ hat die leicht zu erratende Nullstelle $z_1 = 1$. Wir spalten den Linearfaktor $(x-1)$ unter Verwendung des HORNER–Schemas ab:

	1	-1	2	-2	1	-1
$z_1 = 1$	$*$	1	0	2	0	1
	1	0	2	0	1	$\boxed{0}$

Somit erhalten wir die Faktorzerlegung $Q(x) = (x-1)(x^4+2x^2+1) = (x-1)(x^2+1)^2 = (x-1)(x+i)^2(x-i)^2$, die im Reellen den folgenden Ansatz der PBZ erfordert:

$$R(x) = \frac{-2x^4 + x^3 - 3x^2 - 4}{x^5 - x^4 + 2x^3 - 2x^2 + x - 1} = \frac{A}{x-1} + \frac{Bx+C}{1+x^2} + \frac{Dx+E}{(1+x^2)^2}.$$

Den Koeffizienten A erhält man mit der Grenzwertmethode:

$$(x-1)\,R(x)\Big|_{x=1} = \frac{-2x^4 + x^3 - 3x^2 - 4}{(1+x^2)^2}\Big|_{x=1} = -2 = A.$$

Zur Bestimmung der Koeffizienten B, C, D, E kann die Grenzwertmethode **nicht** mehr verwendet werden. Es ist zweckmäßig, durch Einsetzen spezieller Werte x_j, $j = 1,2,3,4$, in den obigen Ansatz ein lineares Gleichungssytem aufzubauen und dieses mit dem GAUSS–Algorithmus zu lösen:

$$\begin{array}{lcrcrrrr}
x=0 & : & 4 & = & 2 & +C & +E & \\
x=-1 & : & \frac{5}{4} & = & 1 & +\frac{1}{2}C & +\frac{1}{4}E & -\frac{1}{2}B \quad -\frac{1}{4}D \\
x=2 & : & -\frac{40}{25} & = & -2 & +\frac{1}{5}C & +\frac{1}{25}E & +\frac{2}{5}B \quad +\frac{2}{25}D \\
x=-2 & : & \frac{56}{75} & = & \frac{2}{3} & +\frac{1}{5}C & +\frac{1}{25}E & -\frac{2}{5}B \quad -\frac{2}{25}D
\end{array}
\quad \Rightarrow \quad
\begin{array}{|rrrrr|}
\hline
B & C & D & E & 1 \\ \hline
0 & 1 & 0 & 1 & 2 \\
2 & -2 & 1 & -1 & -1 \\
10 & 5 & 2 & 1 & 10 \\
10 & -5 & 2 & -1 & -2 \\ \hline
\end{array}$$

Dieses lineare Gleichungssystem hat die eindeutig bestimmte Lösung $B = 0, C = E = 1, D = 2$, und es resultiert die folgende PBZ:

$$R(x) = \frac{-2x^4 + x^3 - 3x^2 - 4}{x^5 - x^4 + 2x^3 - 2x^2 + x - 1} = -\frac{2}{x-1} + \frac{1}{1+x^2} + \frac{2x+1}{(1+x^2)^2}.$$

Unter Verwendung der obigen Integralformeln erhält man nun das unbestimmte Integral

$$\int R(x)\,dx = -2\,\ln|x-1| + \operatorname{arc\,tan}_H x + \frac{x-2}{2(1+x^2)} + \frac{1}{2}\operatorname{arc\,tan}_H x + C.$$

BSP. (8.2.19) Man berechne im Reellen das unbestimmte Integral $\int R(x)\,dx$ der rationalen Funktion $R(x) := (x^5 + x^2 + x)/(x^4+1)$. *Lösung:* Da $R(x)$ nicht echt gebrochen rational ist, verwenden wir zunächst den Euklidischen Teileralgorithmus zur Abspaltung des polynomialen Anteils $T(x)$:

$$\begin{array}{rrr|l}
(x^5 & +x^2 & +x) & : \overbrace{(x^4+1)}^{=Q(x)} = \overbrace{x}^{=:T(x)} \\
x^5 & & +x & \\ \hline
 & x^2 & & =: P(x)
\end{array}$$

Für die Partialbruchzerlegung der echt gebrochen rationalen Funktion $\tilde{R}(x) = P(x)/Q(x) = x^2/(x^4+1)$ berechnen wir zunächst die Nullstellen des Nennerpolynoms $Q(x) := x^4+1$. Diese sind die komplexen Wurzeln

$$\sqrt[4]{-1} = \left\{ e^{i(\pi+2k\pi)/4} \,:\, k = 0,1,2,3 \right\} = \left\{ \frac{1}{2}\sqrt{2}(1+i),\, -\frac{1}{2}\sqrt{2}(1-i),\, -\frac{1}{2}\sqrt{2}(1+i),\, \frac{1}{2}\sqrt{2}(1-i) \right\}.$$

Demgemäß erhält man die Faktorisierung $Q(x) = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1)$, und die Partialbruchzerlegung (2.13) erfordert den Ansatz

$$\tilde{R}(x) = \frac{x^2}{x^4+1} = \frac{Ax+B}{x^2 - \sqrt{2}x + 1} + \frac{Cx+D}{x^2 + \sqrt{2}x + 1}.$$

Wir berechnen die Koeffizienten A, B, C, D nach der Methode des Koeffizientenvergleichs. Es muss gelten:

$$(A+C)x^3 + (B+D+\sqrt{2}(A-C))x^2 + (A+C+\sqrt{2}(B-D))x + B + D \stackrel{!}{=} x^2.$$

Durch Vergleich der x–Potenzen erhält man

A	B	C	D	1
1	0	1	0	0
$\sqrt{2}$	1	$-\sqrt{2}$	1	1
1	$\sqrt{2}$	1	$-\sqrt{2}$	0
0	1	0	1	0

und dieses lineare Gleichungssystem hat die eindeutig bestimmte Lösung $B = D = 0, A = -C = \sqrt{2}/4$. Es resultiert die folgende PBZ:

$$R(x) = \frac{x^5 + x^2 + x}{x^4 + 1} = x + \frac{\sqrt{2}}{4}\Big(\frac{x}{x^2 - \sqrt{2}x + 1} - \frac{x}{x^2 + \sqrt{2}x + 1}\Big).$$

Unter Verwendung der Integralformel für $I_3^{(1)}(x)$ hat man schließlich

$$\int R(x)\,dx = \frac{x^2}{2} + \frac{\sqrt{2}}{8}\ln\left|\frac{x^2 - \sqrt{2}x + 1}{x^2 + \sqrt{2}x + 1}\right| + \frac{\sqrt{2}}{4}\Big(\operatorname{arc\,tan}_H(\sqrt{2}x - 1) + \operatorname{arc\,tan}_H(\sqrt{2}x + 1)\Big) + C.$$

Rationalisierung durch Substitution. Häufig gelingt es, den Integranden durch eine geeignete Substitution in eine **rationale Funktion** zu transformieren. Dabei sind selbstverständlich die Substitutionsregeln aus Satz 8.3 zu beachten. Wir diskutieren hier drei Klassen von Integranden, bei denen mit Standardsubstitutionen die Rationalisierung erreicht wird.

(I) Rationale Funktionen von e^x. Es bezeichne R eine rationale Funktion.

$$\boxed{\int R(e^x)\,dx \quad : \text{Substituiere } u = g(x) := e^x,\ du = u\,dx,\ x \in \mathbf{R}.}$$

Da die hyperbolischen Funktionen rationale Funktionen von $u = e^x$ sind, nämlich

$$\sinh x = \frac{u^2 - 1}{2u}, \quad \cosh x = \frac{u^2 + 1}{2u}, \quad \tanh x = \frac{u^2 - 1}{u^2 + 1}, \quad \coth x = \frac{u^2 + 1}{u^2 - 1},$$

gilt in gleicher Weise

$$\boxed{\int R(e^x, \sinh x, \cosh x, \tanh x, \coth x)\,dx \quad : \text{Substituiere } u = g(x) := e^x.}$$

BSP. (8.2.20) Für die rationale Funktion $R(e^x) := (e^{2x} - 7e^x)/(e^{2x} - 2e^x - 3)$ gilt nach der Substitution $u := e^x$:

$$R(u) = \frac{u - 7}{u^2 - 2u - 3}\,u = \frac{u - 7}{(u+1)(u-3)}\,u = \Big(\frac{2}{u+1} - \frac{1}{u-3}\Big)\,u.$$

Der Faktor u kürzt sich bei der Substitution von $dx = du/u$ heraus, so dass wir schon die Partialbruchzerlegung des Integranden vorliegen haben. Wir erschließen hieraus:

$$\int R(e^x)\,dx = \int^{e^x} \Big(\frac{2}{u+1} - \frac{1}{u-3}\Big)\,du = \ln\frac{(e^x+1)^2}{|e^x - 3|} + C.$$

(II) Rationale Funktionen von $\sin x, \cos x$. Es bezeichne R wieder eine rationale Funktion.

$\int R(\sin x, \cos x)\, dx$: Substituiere $u = g(x) := \tan \frac{x}{2},\ x \in (-\pi, +\pi)$, und verwende die folgenden Transformationsformeln:

$$\begin{aligned} x &= 2 \operatorname{arc\,tan}_H u, \quad dx = \frac{2\,du}{1+u^2}, \\ \sin x &= 2 \sin \frac{x}{2} \cos \frac{x}{2} = \frac{2 \tan \frac{x}{2}}{1 + \tan^2 \frac{x}{2}} = \frac{2u}{1+u^2}, \\ \cos x &= \cos^2 \frac{x}{2} - \sin^2 \frac{x}{2} = \frac{1 - \tan^2 \frac{x}{2}}{1 + \tan^2 \frac{x}{2}} = \frac{1-u^2}{1+u^2}. \end{aligned}$$

Die folgenden Spezialfälle lassen sich einfacher behandeln:

$\int R(\cos x) \cdot \sin x\, dx$: Substituiere $u = g(x) := \cos x,\ du = -\sin x\, dx,$

$\int R(\sin x) \cdot \cos x\, dx$: Substituiere $u = g(x) := \sin x,\ du = \cos x\, dx.$

BSP. (8.2.21) Für die rationale Funktion $R(\cos x) := 1/\cos^3 x$ erhalten wir nach obiger Vorschrift

$$\int R(\cos x)\, dx = \int^{\tan(x/2)} \Big(\frac{1+u^2}{1-u^2}\Big)^3 \frac{2\,du}{1+u^2} = -2 \int^{\tan(x/2)} \frac{(u^2+1)^2}{(u-1)^3(u+1)^3}\, du.$$

Die Partialbruchzerlegung des rationalen Integranden erfordert den folgenden Ansatz:

$$\tilde{R}(u) := \frac{(u^2+1)^2}{(u-1)^3(u+1)^3} = \frac{A_1}{u-1} + \frac{B_1}{(u-1)^2} + \frac{C_1}{(u-1)^3} + \frac{A_2}{u+1} + \frac{B_2}{(u+1)^2} + \frac{C_2}{(u+1)^3}.$$

Hier können die Koeffizienten C_j mit der Grenzwertmethode berechnet werden:

$$C_1 = (u-1)^3 \tilde{R}(u)\Big|_{u=1} = \frac{1}{2}, \quad C_2 = (u+1)^3 \tilde{R}(u)\Big|_{u=-1} = -\frac{1}{2}.$$

Für die Koeffizienten B_j folgt nun ebenfalls nach der Grenzwertmethode unter Verwendung der Regel von L'HOSPITAL:

$$\begin{aligned} B_1 &= \lim_{u\to 1} \Big((u-1)^2 \tilde{R}(u) - \frac{1}{2(u-1)}\Big) = \lim_{u\to 1} \frac{1}{u-1}\Big(\frac{(u^2+1)^2}{(u+1)^3} - \frac{1}{2}\Big) \\ &\overset{\text{L'Hosp.}}{=} \frac{1}{8} \lim_{u\to 1} \Big(4u(u^2+1) - \frac{3}{2}(u+1)^2\Big) = \frac{1}{4}, \\ B_2 &= \lim_{u\to -1} \Big((u+1)^2 \tilde{R}(u) + \frac{1}{2(u+1)}\Big) = \lim_{u\to -1} \frac{1}{u+1}\Big(\frac{(u^2+1)^2}{(u-1)^3} + \frac{1}{2}\Big) \\ &\overset{\text{L'Hosp.}}{=} -\frac{1}{8} \lim_{u\to -1} \Big(4u(u^2+1) + \frac{3}{2}(u-1)^2\Big) = \frac{1}{4}. \end{aligned}$$

Die Berechnung der verbleibenden Koeffizienten A_j erfolgt am einfachsten durch Einsetzen spezieller Werte u in den obigen Ansatz, zum Beispiel:

$$\begin{aligned} u = 0: \quad \tilde{R}(0) = -1 = -A_1 + \tfrac{1}{4} - \tfrac{1}{2} + A_2 + \tfrac{1}{4} + \tfrac{1}{2} \quad &\Rightarrow \quad -A_1 \quad +A_2 \quad = \quad -\tfrac{1}{2}, \\ u = 2: \quad \tilde{R}(2) = \tfrac{25}{27} = A_1 + \tfrac{1}{4} + \tfrac{1}{2} + \tfrac{1}{3} A_2 + \tfrac{1}{36} - \tfrac{1}{54} \quad &\Rightarrow \quad A_1 \quad +\tfrac{1}{3} A_2 \quad = \quad \tfrac{1}{6}. \end{aligned}$$

Dieses lineare Gleichungssystem hat die eindeutige Lösung $A_1 = 1/4 = -A_2$, so dass die vollständige Partialbruchzerlegung in folgender Form vorgelegt ist:

$$\tilde{R}(u) := \frac{1}{4}\Big(\frac{1}{u-1} + \frac{1}{(u-1)^2} + \frac{2}{(u-1)^3} - \frac{1}{u+1} + \frac{1}{(u+1)^2} - \frac{2}{(u+1)^3}\Big).$$

Wir sind jetzt in der Lage, das unbestimmte Integral $\int R(\cos x)\,dx$ zu berechnen:

$$\begin{aligned}
\int \frac{dx}{\cos x} &= -2 \int^{\tan(x/2)} \tilde{R}(u)\,du \\
&= \frac{1}{2}\,\ln\left|\frac{\tan(x/2)+1}{\tan(x/2)-1}\right| + \frac{1}{2}\Big(\frac{1}{\tan(x/2)-1} + \frac{1}{\tan(x/2)+1}\Big) \\
&+ \frac{1}{2}\Big(\frac{1}{(\tan(x/2)-1)^2} - \frac{1}{(\tan(x/2)+1)^2}\Big) + C \\
&= \frac{1}{2}\,\ln\left|\frac{\tan(x/2)+1}{\tan(x/2)-1}\right| + \frac{\sin x}{2\cos^2 x} + C.
\end{aligned}$$

(III) Rationale Funktionen von x und Wurzelfunktionen. Es bezeichne R auch hier eine rationale Funktion.

$$\begin{aligned}
&\int R(x;\sqrt{x^2+a^2})\,dx \quad : \text{Substituiere } x = g^{-1}(u) := a\,\sinh u,\ dx = a\,\cosh u\,du,\ \surd = a\,\cosh u, \\
&\int R(x;\sqrt{x^2-a^2})\,dx \quad : \text{Substituiere } x = g^{-1}(u) := a\,\cosh u,\ dx = a\,\sinh u\,du,\ \surd = a\,\sinh u, \\
&\int R(x;\sqrt{a^2-x^2})\,dx \quad : \text{Substituiere } x = g^{-1}(u) := a\,\sin u,\ dx = a\,\cos u\,du,\ \surd = a\,\cos u.
\end{aligned}$$

BSP. (8.2.22) Die Funktion $f(x) := (4x - x^2)^{-3/2} = \Big(4 - (x-2)^2\Big)^{-3/2}$ geht mit der Substitution $t := x - 2$ in einen Integranden vom obigen Typ, nämlich $R(t) := (4-t^2)^{-3/2}$, über. Die Standardsubstitution $t = g^{-1}(u) := 2\,\sin u$, $dt = 2\,\cos u\,du$, führt nun zu folgendem Resultat:

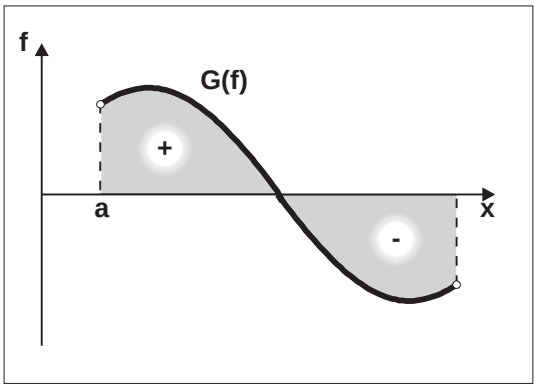
$$\int_2^3 f(x)\,dx = \int_0^1 R(t)\,dt = \int_0^{\pi/6} \frac{2\cos u\,du}{(2\cos u)^3} = \frac{1}{4}\int_0^{\pi/6} \frac{du}{\cos^2 u} = \frac{1}{4}\tan\frac{\pi}{6} = \frac{1}{12}\surd 3.$$

8.3 Das Riemann–Integral

Wir behandeln in diesem Abschnitt die Fragestellung (B), die wir am Anfang von Abschnitt 8.1 formuliert hatten. Es sei auf einem Intervall $[a,b] \subset \mathbf{R}$ eine reellwertige Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ gegeben. Wir fragen nach dem elementargeometrischen Flächeninhalt

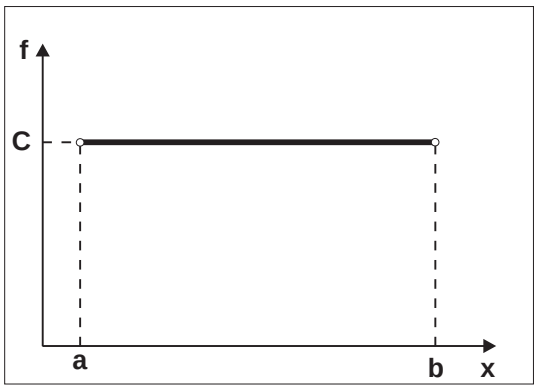
$$A := \int_a^b f(x)\,dx \tag{3.1}$$

der ebenen Fläche **unter dem Graphen** $G(f)$. Darunter verstehen wir das Flächenstück zwischen der x–Achse und $G(f)$. Ganz wesentlich für das Folgende wird die Vereinbarung einer **Orientierung von ebenen Flächenstücken** sein: Flächenstücke **oberhalb** der x–Achse werden stets mit **positivem** Inhalt gezählt; Flächenstücke **unterhalb** der x–Achse hingegen mit **negativem** Inhalt.

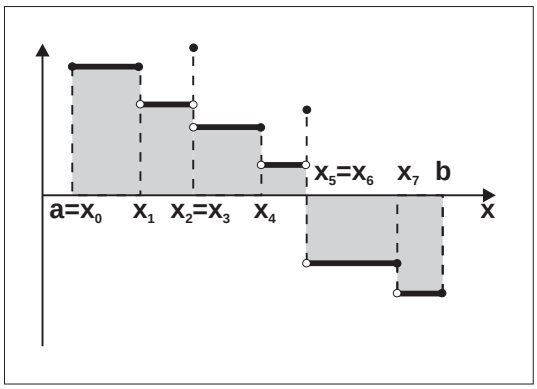


Zur Orientierung von Flächenstücken

Wir müssen noch erläutern, warum die Definition (3.1) über ihren symbolischen Charakter hinaus tatsächlich den Inhalt eines orientierten Flächenstückes bestimmt. Dazu muss sichergestellt sein, dass die Definition (3.1) verträglich ist mit den elementargeometrischen Eigenschaften der Flächenmessung. Dies ist *zum Beispiel* der Fall bei der Funktion $f(x) := C = const$, deren Graph im Intervall $[a, b]$ das Rechteck vom Inhalt $A = C(b-a)$ einschließt. Wir gelangen mit $\int_a^b f(x)\,dx = C(b-a)$ zum gleichen Resultat. Also gilt in diesem Fall (3.1).



Der Rechteckinhalt als Integral über die Funktion $f(x) := C = const$



Die Fläche unter einer Treppenfunktion

Eine Verallgemeinerung dieser Elementareigenschaft auf endlich viele solcher Rechtecke lässt sich am einfachsten mit Hilfe von *Treppenfunktionen* (vgl. BSP. (6.1.11), Abschnitt 6.1) formulieren. Wir hatten in Definition 6.5 den Begriff der **endlichen Zerlegung** eines Intervalls $[a, b] \subset \mathbf{R}$ eingeführt, den wir zur Voraussetzung für das Folgende machen:

Voraussetzung: Es sei eine endliche Zerlegung $Z_n := \{I_1, I_2, \ldots, I_n\}$ des endlichen Intervalls $I := [a, b] \subset \mathbf{R}$ gegeben, so dass die folgenden Eigenschaften gelten:

(Z1) $\quad a =: x_0 \leq x_1 \leq x_2 \leq \cdots \leq x_n := b,$

(Z2) $\quad I_j$ hat die Randpunkte x_{j-1} und x_j, $\; I_j \neq \emptyset, \; \forall\, j = 1, 2, \ldots, n,$

(Z3) $\quad I_j \cap I_k = \emptyset, \; j \neq k, \; \bigcup_{j=1}^{n} I_j = I.$

Ist $|I_j| := x_j - x_{j-1}$ die Länge des Intervalls I_j, so bezeichne die Zahl

$$\boxed{|Z_n| := \max_{1 \leq j \leq n} |I_j|}$$

das **Feinheitsmaß** der Zerlegung Z_n.

Eine Treppenfunktion $T_n : [a,b] \to \mathbf{R}$ bezüglich der Zerlegung Z_n ist nun eine Funktion in der Form

$$T_n(x) := \sum_{j=1}^{n} y_j \chi_{I_j}(x) \;\; \forall\, x \in [a,b] \;\; \text{mit} \;\; \chi_{I_j}(x) := \begin{cases} 1 \;:\; x \in I_j, \\ 0 \;:\; x \notin I_j. \end{cases} \tag{3.2}$$

Die Interpretation der Definition (3.1), die wir die RIEMANN–**Integrierbarkeit** der Funktion $f(x)$ nennen, soll die beiden folgenden Forderungen einschließen:

Forderung 1: Jede Treppenfunktion $T_n : [a,b] \to \mathbf{R}$ ist (RIEMANN–) integrierbar, und es gilt:

$$A = \int_a^b T_n(x)\,dx = \sum_{j=1}^{n} y_j (x_j - x_{j-1}) = \sum_{j=1}^{n} y_j\,|I_j|.$$

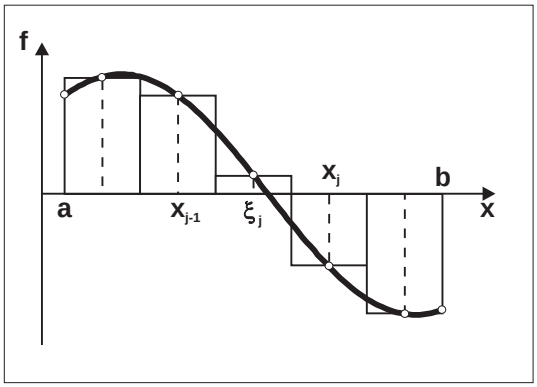
Der Begriff der (RIEMANN–) Integrierbarkeit muss den Begriff des bestimmten Integrals umfassen, wenn die Funktion f auf dem Intervall $[a,b]$ eine Stammfunktion $F(x)$ hat:

Forderung 2: Ist $F(x)$ auf dem Intervall $I := [a,b]$ eine Stammfunktion der gegebenen Funktion $f : I \to \mathbf{R}$, und ist f (RIEMANN–) integrierbar, so gilt:

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

Wir betonen hier, dass mit diesen Forderungen keineswegs eine Präzisierung des Begriffs der (RIEMANN–) Integrierbarkeit vorweggenommen werden soll. Die genaue Definition wird weiter unten erfolgen. Wir versuchen hier lediglich, einen neuen Begriff in einen bereits bekannten Rahmen einzupassen. Während durch die Forderung 1 der *elementargeometrische Aspekt* des Integralbegriffs unterstrichen wird, kommt in der Forderung 2 die in Abschnitt 8.1 diskutierte *Umkehroperation* der Differentiation zum Vorschein. Wir zeigen zunächst den folgenden Zusammenhang:

Ist $f : I \to \mathbf{R}$ *stetig*, so gilt die Implikation **Forderung 1** $\Rightarrow$ **Forderung 2**.



Zum RIEMANN–**Integral einer Funktion** f

In der Tat, ist $F(x)$ auf dem Intervall $I := [a,b]$ eine Stammfunktion von f, so gilt also $F'(x) = f(x)\;\forall\, x \in I$. Es sei Z_n eine endliche Zerlegung von I. Dann gilt auf jedem Teilintervall $I_j \in Z_n$ der Mittelwertsatz (Satz 7.11):

$$F(x_j) - F(x_{j-1}) = f(\xi_j)\,(x_j - x_{j-1}), \quad x_{j-1} < \xi_j < x_j. \tag{3.3}$$

Gilt $x_{j-1} = x_j$, so setzen wir vereinbarungsgemäß $\xi_j := x_j$. Durch Summation von (3.3) über alle $j = 1, 2, \ldots, n$ erhält man

$$F(b) - F(a) = \sum_{j=1}^{n} f(\xi_j)(x_j - x_{j-1}) = \sum_{j=1}^{n} f(\xi_j)\,|I_j|. \tag{3.4}$$

Auf der rechten Seite dieser Gleichung steht der Inhalt der von der Treppenfunktion $T_n(x) := \sum_{j=1}^{n} f(\xi_j)\chi_{I_j}(x)$ begrenzten Fläche. Auf der linken Gleichungsseite steht das bestimmte Integral der Funktion f:

$$\int_a^b f(x)\,dx = \int_a^b T_n(x)\,dx = \sum_{j=1}^{n} f(\xi_j) \int_{x_{j-1}}^{x_j} dx = \sum_{j=1}^{n} f(\xi_j)\,|I_j|. \tag{3.5}$$

Hier ist die linke Gleichungsseite unabhängig von der speziellen Wahl der Zerlegung Z_n, also insbesondere unabhängig von der Zahl n der Teilintervalle I_j. Deshalb darf auch die rechte Gleichungsseite nicht von n abhängen: Die konstante Folge $\Big(\sum_{j=1}^{n} f(\xi_j)\,|I_j|\Big)_{n\geq 1}$ hat im Limes $|Z_n| \to 0$ den Grenzwert $F(b) - F(a) = \int_a^b f(x)\,dx$. Dieser ist gemäß der *geometrischen* Bedeutung der Summe $\sum_{j=1}^{n} f(\xi_j)\,|I_j|$ der Flächeninhalt A unter dem Graphen $G(f)$. Das Problem besteht nun im Auffinden der Zwischenstellen $\xi_j \in I_j$ in der Beziehung (3.3). Man darf erwarten, dass die Folge der Summen $\sum_{j=1}^{n} f(\tau_j)\,|I_j|$ für **jede Wahl** einer Zwischenstelle $\tau_j \in I_j$ im Limes $|Z_n| \to 0$ gegen den obigen Grenzwert $F(b) - F(a)$ konvergiert. Um dies zu zeigen, verwenden wir den Satz 6.20 von der **gleichmäßigen** Stetigkeit der Funktion $f : I \to \mathbf{R}$:

$$\forall\, \epsilon > 0\ \exists\, \delta = \delta(\epsilon)\ :\ |f(x) - f(y)| < \epsilon\ \ \forall\, x, y \in I\ \text{ mit }\ 0 < |x - y| < \delta. \tag{3.6}$$

Zu $\epsilon > 0$ sei nun ein solches $\delta(\epsilon)$ gewählt. Falls schon $|Z_n| < \delta$ gilt, so folgt für die in (3.3) fixierte Zwischenstelle $\xi_j \in I_j$:

$$|f(\xi_j) - f(\tau_j)| < \epsilon\ \ \forall\, \tau_j \in I_j.$$

Unter Verwendung der Gleichung (3.5) resultiert nun

$$\Big|\int_a^b f(x)\,dx - \sum_{j=1}^{n} f(\tau_j)\,|I_j|\Big| = \Big|\sum_{j=1}^{n} \Big[f(\xi_j) - f(\tau_j)\Big]\,|I_j|\Big| \leq \epsilon \sum_{j=1}^{n} |I_j| = \epsilon(b - a).$$

Das heißt, wir haben

$$F(b) - F(a) = \int_a^b f(x)\,dx = \lim_{|Z_n|\to 0} \sum_{j=1}^{n} f(\tau_j)\,(x_j - x_{j-1}).$$

Motiviert durch dieses Resultat, wird die folgende Definition sinnvoll:

Definition 8.4 *Die Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R})$ heiße auf dem Intervall $\mathbf{R} \supset [a, b]$* **Riemann–integrierbar** *(kurz: R–integrierbar), wenn die Folge der* **Riemann–Summen**

$$\boxed{S_{Z_n} := \sum_{j=1}^{n} f(\xi_j)\,(x_j - x_{j-1})} \tag{3.7}$$

für jede Wahl von Zerlegungen $Z_n = \{I_1, I_2, \ldots, I_n\}$ des Intervalls $[a, b]$ und jede Wahl der Zwischenstellen $\xi_j \in I_j$ im Limes $|Z_n| \to 0$ demselben Grenzwert S zustrebt. In diesem Fall heiße S das **Riemann–Integral** *(kurz: R–Integral) von f über $[a, b]$:*

$$\boxed{S = \int_a^b f(x)\,dx = \lim_{|Z_n|\to 0} S_{Z_n} = \lim_{|Z_n|\to 0} \sum_{j=1}^{n} f(\xi_j)\,(x_j - x_{j-1}).}$$

Die Berechnung des R–Integrals einer Funktion f mit Hilfe der RIEMANN–Summen ist natürlich mit einem nicht zu vertretenden Aufwand verbunden. Aus der im Vorspann gegebenen Herleitung resultiert jedoch ein einfaches Verfahren zur Berechnung von $\int_a^b f(x)\,dx$:

Satz 8.7 (Erster Hauptsatz der Differential– und Integralrechnung)
Die Funktion $F \in \mathrm{Abb}(\mathbf{R},\mathbf{R})$ *habe eine auf dem Intervall* $[a,b] \subset \mathbf{R}$ *stetige oder auch nur R–integrierbare Ableitung* $F'(x) = f(x)$, $x \in [a,b]$. *Dann gilt*

$$\boxed{\int_a^b F'(x)\,dx = \int_a^b f(x)\,dx = F(b) - F(a) =: F(x)\Big|_a^b.} \tag{3.8}$$

Begründung: Ist f stetig, so folgt die Behauptung aus den Betrachtungen im Vorspann. Ist f R–integrierbar, so ist $F(x)$ gemäß Vorgabe eine Stammfunktion von f, denn es gilt ja $F'(x) = f(x)$ auf dem Intervall $[a,b]$. Da die Funktion $F(x)$ notwendig stetig sein muss, gilt der Mittelwertsatz (Satz 7.11). Mit jeder endlichen Zerlegung Z_n des Intervalls $[a,b]$ gilt für geeignete Zwischenwerte $\xi_j \in I_j \in Z_n$:

$$F(b) - F(a) = \sum_{j=1}^{n} \Big(F(x_j) - F(x_{j-1})\Big) = \sum_{j=1}^{n} f(\xi_j)\,(x_j - x_{j-1}) \overset{|Z_n|\to 0}{=} \int_a^b f(x)\,dx.$$

Bemerkung 8.5 (a) Mit der getroffenen Definition sind also die Forderungen 1 und 2 erfüllt.

(b) Man kann das R–Integral mit Hilfe von Stammfunktionen berechnen, und dies ist die gängigste und praktisch brauchbarste Methode. Man halte jedoch sorgfältig auseinander: Die Existenz einer Stammfunktion ist nicht dasselbe wie die Existenz des R–Integrals. Die Frage, ob es genügend viele R–integrierbare Funktionen gibt, werden wir später in Satz 8.12 beantworten. Vorerst genügt uns die Aussage von Satz 8.7: Hat die stetige Funktion $f \in \mathrm{Abb}(\mathbf{R},\mathbf{R})$ auf dem Intervall $[a,b]$ eine Stammfunktion $F(x)$, so ist sie dort R–integrierbar. □

BSP. (8.3.1) Bei der Berechnung der folgenden R–Integrale werden jeweils die Techniken zur Bestimmung einer Stammfunktion verwendet. R–Integral und bestimmtes Integral sind in diesen Fällen gleich.

$$\int_a^b \lambda\,dx = \lambda x\Big|_a^b = \lambda\,(b-a) \quad \forall\,\lambda \in \mathbf{R},$$

$$\int_1^e \ln x \frac{dx}{x} = \int_1^e f(x)\,f'(x)\,dx = \frac{1}{2}\Big(\ln x\Big)^2\Big|_1^e = \frac{1}{2},$$

$$\int_{-1}^{1} x e^{-x^2}\,dx = -\frac{1}{2}\int_{-1}^{1} e^{-x^2}(-2x)\,dx = -\frac{1}{2}\int_{-1}^{1} e^{g(x)}\,g'(x)\,dx = -\frac{1}{2}e^{-x^2}\Big|_{-1}^{+1} = 0,$$

$$\int_0^{\pi} \cos^2 x\,dx = \frac{1}{2}\int_0^{\pi}(1+\cos 2x)\,dx = \frac{1}{2}\Big(x + \frac{1}{2}\sin 2x\Big)\Big|_0^{\pi} = \frac{\pi}{2}.$$

BSP. (8.3.2) In den folgenden Integralen werden für Zahlen $n, m \in \mathbf{N}$ die Identitäten

$$\cos mx \cdot \sin nx = \frac{1}{2}\Big(\sin(m+n)x - \sin(m-n)x\Big), \quad \cos mx \cdot \cos nx = \frac{1}{2}\Big(\cos(m-n)x + \cos(m+n)x\Big)$$

verwendet. Es gilt:

$$\int_0^{2\pi} \cos mx \cdot \sin nx\,dx = \begin{cases} -\frac{1}{4n}\cos 2nx\Big|_0^{2\pi} = 0 & : \ n = m, \\ -\frac{1}{2}\Big(\frac{1}{m+n}\cos(m+n)x - \frac{1}{m-n}\cos(m-n)x\Big)\Big|_0^{2\pi} = 0 & : \ n \neq m, \end{cases}$$

$$\int_0^{2\pi} \cos mx \cdot \cos nx\,dx = \begin{cases} \Big(\frac{x}{2} + \frac{1}{4n}\sin 2nx\Big)\Big|_0^{2\pi} = \pi & : \ n = m, \\ \frac{1}{2}\Big(\frac{1}{m-n}\sin(m-n)x + \frac{1}{m+n}\sin(m+n)x\Big)\Big|_0^{2\pi} = 0 & : \ n \neq m. \end{cases}$$

Eigenschaften, die wir bereits für das durch Stammfunktionen erklärte bestimmte Integral $\int_a^b f(x)\,dx$ abgeleitet hatten, sollen auch für das R–Integral ihre Geltung behalten. Dies begründet die folgende

Definition 8.5 *Für jede reellwertige Funktion f mit $a \in D(f)$ gelte*

$$\boxed{\int_a^a f(x)\,dx := 0.} \tag{3.9}$$

Für $a < b$ und für jede R–integrierbare Funktion $f : [a,b] \to \mathbf{R}$ gelte

$$\boxed{\int_a^a f(x)\,dx := -\int_a^b f(x)\,dx.} \tag{3.10}$$

Satz 8.8 *Die reellwertigen Funktionen f, g seien auf dem Intervall $[a,b]$ R–integrierbar. Dann gilt:*

(a) *Die Funktion $\lambda f(x) + \mu g(x)$, $\lambda, \mu \in \mathbf{R}$ ist auf $[a,b]$ R–integrierbar mit*

$$\boxed{\int_a^b \Big(\lambda f(x) + \mu g(x)\Big)\,dx = \lambda \int_a^b f(x)\,dx + \mu \int_a^b g(x)\,dx. \quad \textbf{Linearität}} \tag{3.11}$$

(b) *Die Funktion $(f \cdot g)(x)$ ist auf $[a,b]$ R–integrierbar.*

(c) *Gilt $g(x) \leq f(x)\ \forall\ x \in [a,b]$, so folgt*

$$\boxed{\int_a^b g(x)\,dx \leq \int_a^b f(x)\,dx, \quad \textit{insbesondere } 0 \leq \int_a^b f(x)\,dx, \ \textit{falls } g = 0.} \tag{3.12}$$

(d) *Die Funktion $|f(x)|$ ist auf $[a,b]$ R–integrierbar mit*

$$\boxed{\Big|\int_a^b f(x)\,dx\Big| \leq \int_a^b |f(x)|\,dx.} \tag{3.13}$$

Diese Eigenschaften können direkt aus der Definition des R–Integrals abgeleitet werden. Die Linearitätsaussage (3.11) besagt, dass die Klasse der auf $[a,b]$ R–integrierbaren Funktionen einen **Vektorraum über dem Körper R** bilden.

Satz 8.9 *Es seien $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ und $c \in [a,b]$ gegeben. Genau dann ist die Funktion f auf dem Intervall $[a,b]$ R–integrierbar, wenn f auf jedem der beiden Teilintervalle $[a,c]$ und $[c,b]$ R–integrierbar ist. In diesem Fall gilt:*

$$\boxed{\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.} \tag{3.14}$$

Begründung: Sind Z_m und Z_n endliche Zerlegungen der Teilintervalle $[a,c]$ bzw. $[c,b]$, so ist $Z_{m+n} := Z_m \cup Z_n$ eine endliche Zerlegung des Intervalls $[a,b]$. Es gilt $|Z_{m+n}| \to 0$ genau, wenn $|Z_m| \to 0$ und $|Z_n| \to 0$ streben. Ferner gilt für die zugeordneten RIEMANN–Summen:

$$S_{Z_{m+n}} = S_{Z_m} + S_{Z_n}.$$

Daraus folgt im Limes $|Z_{m+n}| \to 0$ die Behauptung. □

Setzt man in Satz 8.9 $x := c$, so erhält man zu jeder R–integrierbaren Funktion $f : [a,b] \to \mathbf{R}$ die Existenz der Funktion

$$F(x) := \int_a^x f(t)\,dt \;\; \forall\, x \in [a,b].$$

Wir folgern mit (3.10) und (3.14):

$$F(x) - F(y) = \int_a^x f(t)\,dt - \int_a^y f(t)\,dt = \int_y^a f(t)\,dt + \int_a^x f(t)\,dt = \int_y^x f(t)\,dt \;\; \forall\, x,y \in [a,b].$$

Ist die Funktion f auf $[a,b]$ *beschränkt*, so erhalten wir unter Verwendung von (3.13):

$$|F(x) - F(y)| \le \sup_{t\in[a,b]} |f(t)| \cdot |x-y| \;\; \forall\, x,y \in [a,b],$$

das heißt, die Funktion $F(x)$ ist auf dem Intervall $[a,b]$ LIPSCHITZ–stetig. Wir ergänzen:

Satz 8.10 *Die Funktion $f : [a,b] \to \mathbf{R}$ sei R–integrierbar. Dann ist f auf dem Intervall $[a,b]$ auch beschränkt, und die Funktion*

$$\boxed{F(x) := \int_a^x f(t)\,dt, \quad x \in [a,b],} \tag{3.15}$$

ist auf $[a,b]$ LIPSCHITZ–stetig mit einer LIPSCHITZ–Konstanten $L := \sup\limits_{t\in[a,b]} |f(t)|$.

Begründung: Wir brauchen offensichtlich nur noch die Beschränktheit der Funktion f zu zeigen. Wäre f nämlich unbeschränkt, so gäbe es einen Punkt $x_0 \in [a,b]$, und zu jeder Zahl $R > 0$ ein Intervall I_R mit Randpunkt x_0 derart, dass $|f(x)| \ge R \;\forall\, x \in I_R$ folgte. Wegen $\sup\limits_{x\in I_R} |f(x)| = +\infty$ gäbe es eine Zwischenstelle $\xi_R \in I_R$ mit

$$|f(\xi_R)| > \frac{1}{|I_R|}\Big(R + \int_a^b |f(x)|\,dx\Big).$$

Für jede Zerlegung Z_n des Intervalls $[a,b]$, die das Teilintervall I_R enthält, folgte dann

$$S_{Z_n} - \int_a^b |f(x)|\,dx \ge |f(\xi_R)|\,|I_R| - \int_a^b |f(x)|\,dx > R.$$

Dies widerspräche der Definition des R–Integrals von $|f(x)|$, nach der die Konvergenz $S_{Z_n} \to \int_a^b |f(x)|\,dx$ erfolgt. □

Die Klasse der auf einem Intervall $[a,b]$ R–integrierbaren Funktionen bildet sicher keinen Unterraum des Vektorraums $C([a,b])$, da ja gemäß Forderung 1 auch die *unstetigen* Treppenfunktionen R–integrierbar sind. Es liegt jedoch nahe, auf der Basis der Relation (3.15) die Existenz von Stammfunktionen mit Hilfe des RIEMANN–Integral zu begründen. Wir werden dieses Ziel in zwei Schritten erreichen.

Satz 8.11 (Erster Mittelwertsatz der Integralrechnung)
Die Funktion $f : [a,b] \to \mathbf{R}$ sei R–integrierbar.

(a) *Gibt es Schranken m, M mit $m \le f(x) \le M \;\forall\; x \in [a,b]$, so folgt*

$$\boxed{m(b-a) \le \int_a^b f(x)\,dx \le M(b-a).} \tag{3.16}$$

(b) *Ist f auf $[a,b]$ stetig, so existiert eine Zwischenstelle $\xi \in [a,b]$ mit*

$$\boxed{\int_a^b f(x)\,dx = f(\xi)\,(b-a).} \tag{3.17}$$

Begründungen: (a) Wegen $m \le f(x) \le M$ erhalten wir aus (3.12):

$$m \int_a^b 1\,dx \le \int_a^b f(x)\,dx \le M \int_a^b 1\,dx.$$

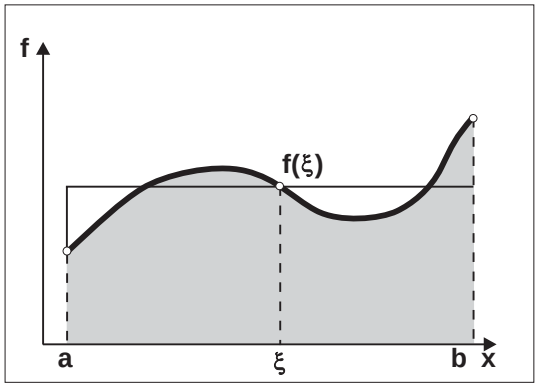
(b) Wegen (a) liegt die Zahl $\eta := \frac{1}{b-a} \int_a^b f(x)\,dx$ im Intervall $[m,M]$. Nach dem Zwischenwertsatz von BOLZANO (Satz 6.19) nimmt $f(x)$ jeden Wert zwischen K $m := \min f(t)$ und $M := \max f(t)$ an. Also existiert eine Zwischenstelle $\xi \in [a,b]$ mit $f(\xi) = \eta$. □

Die Relation (3.17) besagt, dass die Rechteckfläche $f(\xi)\,(b-a)$ für eine Zwischenstelle $\xi \in [a,b]$ flächengleich ist mit der Fläche der Funktion f unterhalb des Graphen $G(f)$. Die Stelle ξ ist im allgemeinen nicht eindeutig bestimmt.

Definition 8.6 *Die Zahl*

$$\boxed{\bar{f} := f(\xi) = \frac{1}{b-a} \int_a^b f(x)\,dx}$$

*heißt der (Integral–)***Mittelwert von f auf** $[a,b]$.



Der Integralmittelwert einer Funktion f auf $[a,b]$

Wir beweisen nun mit Hilfe des Mittelwertsatzes das folgende zentrale Resultat der Infinitesimalrechnung:

Satz 8.12 (Zweiter Hauptsatz der Differential– und Integralrechnung)
Für jede stetige Funktion $f \in C([a,b])$ gilt:

(a) *Das* RIEMANN*–Integral* $\int_a^b f(x)\,dx$ *existiert.*

(b) *Die Funktion*

$$\boxed{F(x) := \int_a^x f(t)\,dt, \quad x \in [a,b],}$$

ist auf dem Intervall $[a,b]$ eine Stammfunktion von f. Das heißt, es gilt $F'(x) = f(x)\ \forall\ x \in [a,b]$.

Begründungen: (a) Es sei $Z_n = \{I_1, I_2, \ldots, I_n\},\ n \in \mathbf{N}$, eine Folge endlicher Zerlegungen des Intervalls $[a,b]$ mit $|Z_n| \to 0$, und es seien $Z_{n_j} = \{I_{j1}, I_{j2}, \ldots, I_{jn_j}\}$ endliche Zerlegungen der Teilintervalle $I_j,\ j = 1,2,\ldots,n$. Dann ist $Z := \bigcup_{j=1}^{n} Z_{n_j}$ eine *Verfeinerung* der Zerlegung Z_n. Da die Funktion f auf dem Intervall $[a,b]$ gleichmäßig stetig ist, gilt die Beziehung

$$|f(x) - f(y)| < \epsilon \ \ \forall\ x,y \in [a,b] \ \text{ mit } \ 0 < |x-y| < \delta(\epsilon).$$

Die Zahl $\delta = \delta(\epsilon)$ hängt dabei nur von der beliebig wählbaren Zahl $\epsilon > 0$ ab. Wir wählen zu festem $\epsilon > 0$ die Zerlegung Z_n so, dass $|Z_n| < \delta$ gilt. Dann gilt für Zwischenstellen $\xi_j \in I_j$ und $\tau_{jk} \in I_{jk}$:

$$\begin{aligned} |S_{Z_n} - S_Z| &= \Big| \sum_{j=1}^{n} f(\xi_j)\,(x_j - x_{j-1}) - \sum_{j=1}^{n}\sum_{k=1}^{n_j} f(\tau_{jk})\,(x_{jk} - x_{j,k-1}) \Big| \\ &\le \sum_{j=1}^{n}\sum_{k=1}^{n_j} \underbrace{|f(\xi_j) - f(\tau_{jk})|}_{<\epsilon}\,(x_{jk} - x_{j,k-1}) < \epsilon(b-a). \end{aligned}$$

Da ϵ beliebig klein werden darf, ist $\left(S_{Z_n}\right)_{n\in\mathbf{N}} \subset \mathbf{R}$ eine CAUCHY–Folge und somit konvergent.

(b) Wir wählen $x \in [a,b]$ fest und dazu $h \neq 0$ so, dass $x+h \in [a,b]$ gilt. Dann folgt unter Verwendung des Mittelwertsatzes (3.17) für eine Zwischenstelle $\xi = x + \theta h,\ \theta \in (0,1)$:

$$F'(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(t)\,dt = \lim_{h\to 0} f(\xi) = f(x).$$

Bemerkung 8.6 (a) Wir haben gezeigt, dass jede stetige Funktion f R–integrierbar ist. Somit ist $C([a,b])$ Unterraum des Vektorraums der R–integrierbaren Funktionen.

(b) Zu jeder Funktion $f \in C([a,b])$ existiert eine Stammfunktion. Flächenmessung kann wegen der Forderung 2 mit Hilfe einer Stammfunktion erfolgen. Umgekehrt kann die Berechnung des R–Integrals $\int_a^b f(x)\,dx$ näherungsweise durch Flächenmessung erfolgen. Auf dieser Tatsache beruhen die Verfahren der **numerischen Integration**, mit denen wir uns in Abschnitt 8.5 auseinandersetzen werden.

(c) Mit der Berechnung von $F(x) := \int_a^x f(t)\,dt$ können aus stetigen Funktionen stetig differenzierbare Funktionen gewonnen werden. □

BSP. (8.3.3) Wir bestimmen Elementarfunktionen durch Integration gebrochen rationaler Funktionen:

$$\boxed{\ln x = \int_1^x \frac{dt}{t} \ \forall\, x > 0, \quad \operatorname{arc\,tan}_H x = \int_0^x \frac{dt}{1+t^2} \ \forall\, x \in \mathbf{R}.}$$

Wie wir in Abschnitt 8.2 gesehen haben, lassen sich weitere transzendente Funktionen nicht durch Integration **rationaler** Funktionen gewinnen.

BSP. (8.3.4) Wir bestimmen nicht elementare Funktionen durch Integration von Elementarfunktionen, zum Beispiel:

(i) **Das** GAUSS**sche Fehlerintegral** (error function) ist die Funktion

$$\boxed{\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \ \forall\, x \geq 0.}$$

Der **Normierungsfaktor** $2/\sqrt{\pi}$ gewährleistet hier den Grenzwert $\lim\limits_{x \to +\infty} \operatorname{erf}(x) = 1$.

(ii) **Die** FRESNEL**schen Integrale** sind die beiden Funktionen

$$\boxed{C(x) := \int_0^x \cos\left(\frac{\pi t^2}{2}\right) dt, \quad S(x) := \int_0^x \sin\left(\frac{\pi t^2}{2}\right) dt \ \forall\, x \in \mathbf{R}.}$$

Mit Hilfe der Substitution $u = g(t) := \pi t^2/2,\ du/\sqrt{u} = \sqrt{2\pi}\, dt$, erhält man

$$\int_0^x \frac{\cos u}{\sqrt{u}}\, du = \sqrt{2\pi}\, C\left(\sqrt{\frac{2x}{\pi}}\right), \quad \int_0^x \frac{\sin u}{\sqrt{u}}\, du = \sqrt{2\pi}\, S\left(\sqrt{\frac{2x}{\pi}}\right) \ \forall\, x \geq 0.$$

(iii) **Die elliptischen Integrale** sind folgende Funktionen:

$$\boxed{\begin{aligned}
&\text{1. Gattung:} && F(\varphi,k) && := \int_0^\varphi \frac{dt}{\sqrt{1-k^2\sin^2 t}}, && 0 \leq \varphi \leq \frac{\pi}{2},\ k^2 \leq 1,\\
&\text{2. Gattung:} && E(\varphi,k) && := \int_0^\varphi \sqrt{1-k^2\sin^2 t}\, dt, && 0 \leq \varphi \leq \frac{\pi}{2},\ k^2 \leq 1,\\
&\text{3. Gattung:} && \Pi(\varphi,n,k) && := \int_0^\varphi \frac{dt}{(1+n\sin^2 t)\sqrt{1-k^2\sin^2 t}}, && 0 \leq \varphi \leq \frac{\pi}{2},\ k^2 \leq 1,\ n \in \mathbf{N}.
\end{aligned}}$$

Für diese Funktionen existieren wie für die Elementarfunktionen Wertetafeln.

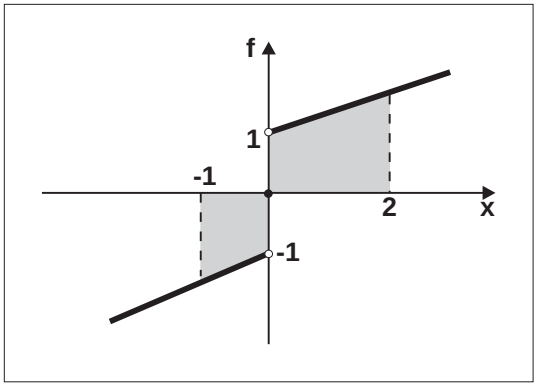
Die Klasse der Treppenfunktionen auf dem Intervall $[a,b]$ bilden zusammen mit dem Vektorraum $C([a,b])$ der stetigen Funktionen bereits einen reichhaltigen Fundus R–integrierbarer Funktionen. Damit ist aber noch längst nicht die Gesamtheit der R–integrierbaren Funktionen ausgeschöpft. Die folgenden Beispiele sollen einerseits die Erweiterungsmöglichkeiten aufzeigen, andererseits aber auch die Grenzen des RIEMANNschen Integralbegriffs verdeutlichen.

BSP. (8.3.5) Die Funktion $f(x) := \frac{x}{2} + \operatorname{sign} x,\ x \in D(f) := [-1,2]$, ist im Punkt $x_0 = 0$ unstetig. Den Inhalt der Fläche unter dem Graphen $G(f)$ kann man jedoch elementargeometrisch mit Hilfe der Formel für den Trapezinhalt angeben:

$$A = 2\left(1 + 0.5(2-1)\right) - 1\left(1 + 0.5(1.5-1)\right) = \frac{7}{4}.$$

Dieses Ergebnis steht im Einklang mit der R–Integration:

$$A = \int_{-1}^{2} f(x)\,dx = \int_{-1}^{0} \Big(\frac{x}{2} - 1\Big)\,dx + \int_{0}^{2} \Big(\frac{x}{2} + 1\Big)\,dx = \Big(\frac{x^2}{4} - x\Big)\Big|_{-1}^{0} + \Big(\frac{x^2}{4} + x\Big)\Big|_{0}^{2} = \frac{7}{4}.$$



RIEMANN–**Integration einer stückweise stetigen Funktion**

In Verallgemeinerung dieses Beispiels führen wir den folgenden Begriff ein:

Definition 8.7 *Eine Funktion* $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ *heiße auf dem Intervall* $[a, b]$ **stückweise stetig**, *wenn* $f(x)$ *stetig ist in jedem Punkt* $x \in [a, b]$ *mit Ausnahme von höchstens endlich vielen Sprungstellen* $a \le x_0 < x_1 < \cdots < x_n \le b$, *in denen die Funktion* f *Sprünge von endlicher Höhe haben darf:* $|f(x_j - 0) - f(x_j + 0)| =: K_j < +\infty,\ j = 0, 1, \ldots, n.$

Satz 8.13 *Jede auf dem Intervall* $[a, b]$ *beschränkte und stückweise stetige Funktion* f *ist R–integrierbar. Sind die Sprungstellen von* f *in der Reihenfolge* $a \le x_0 < x_1 < \cdots < x_n \le b$ *angeordnet, so gilt*

$$\boxed{\int_a^b f(x)\,dx = \int_a^{x_0} f(x)\,dx + \sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} f(x)\,dx + \int_{x_n}^{b} f(x)\,dx.} \tag{3.18}$$

Die Bedingung der stückweisen Stetigkeit in dem obigen Satz kann weiter abgeschwächt werden. Eine Funktion f ist auch dann noch auf dem Intervall $[a, b]$ R–integrierbar, wenn sie dort beschränkt und bis auf endlich viele Ausnahmestellen stetig ist. Dazu zeigen wir in einem ersten Schritt, dass eine auf dem endlichen **offenen** Intervall (a, b) stetige Funktion f R–integrierbar ist, wenn sie dort beschränkt ist:

Satz 8.14 *Die Funktion* $f : (a, b) \to \mathbf{R}$ *sei stetig und beschränkt:* $\sup\limits_{x \in (a,b)} |f(x)| \le M < +\infty.$ *Dann ist* f *auf dem endlichen Intervall* $[a, b] \subset \mathbf{R}$ *auch R–integrierbar, und es gilt:*

$$\boxed{\int_a^b f(x)\,dx = \lim_{\epsilon \to 0+} \int_{a+\epsilon}^{b-\epsilon} f(x)\,dx.} \tag{3.19}$$

Begründung: Für jedes $\epsilon > 0$ ist $f : [a + \epsilon, b - \epsilon] \to \mathbf{R}$ stetig und somit R–integrierbar. Es seien nun $Z_{kn_k} = \{I_{k1}, I_{k2}, \ldots, I_{kn_k}\},\ k = 1, 2, 3,$ endliche Zerlegungen der Intervalle $[a, a + \epsilon], [a + \epsilon, b - \epsilon]$ und

$[b-\epsilon, b]$. Dann ist $Z := Z_{1n_1} \cup Z_{2n_2} \cup Z_{3n_3}$ eine endliche Zerlegung des Intervalls $[a,b]$, und es gilt

$$\begin{aligned} \Big| \lim_{|Z|\to 0} S_Z - \int_{a+\epsilon}^{b-\epsilon} f(x)\,dx \Big| &\le \limsup_{|Z|\to 0} \Big\{ \sum_{j=1}^{n_1} |f(\tau_j)|\,|I_{1j}| + \sum_{j=1}^{n_3} |f(\sigma_j)|\,|I_{3j}| \Big\} \\ &\le M\Big((a+\epsilon-a) + (b-b+\epsilon) \Big) = 2M\epsilon. \end{aligned}$$

Hierin bezeichnen $a < \tau_j \in I_{1j}$ und $b > \sigma_j \in I_{3j}$ die in den RIEMANN–Summen auftretenden Zwischenstellen. Da $\epsilon > 0$ beliebig klein gewählt werden darf, folgt dann aus der obigen Ungleichung die Behauptung. □

Indem man Satz 8.14 auf endlich viele offene Intervalle $(a, x_0), (x_{j-1}, x_j),\ j = 1, 2, \ldots, n,\ (x_n, b)$ anwendet, erhält man:

Satz 8.15 *Die Funktion $f : (a,b) \to \mathbf{R}$ sei beschränkt, und sie sei stetig mit Ausnahme von höchstens endlich vielen Unstetigkeitsstellen $a \le x_0 < x_1 < \cdots < x_n \le b$. Dann ist f auf dem endlichen Intervall $[a,b] \subset \mathbf{R}$ R–integrierbar, und es gilt die Beziehung* (3.18).

Bemerkung 8.7 (a) Die Regeln der partiellen Integration und die Substitutionsregeln aus Satz 8.3 gelten bei entsprechender Modifikation der Voraussetzungen nun auch für R–integrierbare Funktionen.

(b) Zum Beispiel ist die Funktion $f(x) := \cos x \cdot \operatorname{sign}(\sin x)$ auf jedem Teilintervall $[a,b] \subset \mathbf{R}$ stückweise stetig und somit gemäß Satz 8.13 R–integrierbar.

(c) Die Funktion $f(x) := \tan x$ ist auf dem Intervall $[0, \frac{\pi}{2}]$ **nicht** R–integrierbar. Sie ist zwar auf dem offenen Intervall $(0, \frac{\pi}{2})$ stetig, dort aber nicht mehr beschränkt.

(d) Die Funktion $f(x) := \sin \frac{1}{x}$ ist auf dem Intervall $[-1,1]$ R–integrierbar, denn sie ist dort beschränkt mit einer Unstetigkeitsstelle $x_0 = 0$. Da die einseitigen Funktionenlimites $f(x_0 \pm 0)$ **nicht** existieren, ist die Stelle x_0 keine Sprungstelle. Wir können in diesem Fall jedoch Satz 8.15 anwenden. Überdies kann in dem vorliegenden Beispiel auch die Substitutionsregel angewendet werden. Für $x \neq 0$ setze man $u = g(x) := 1/x,\ du = -u^2 dx$. Dann folgt

$$\int_{-1}^{1} \sin\frac{1}{x}\,dx = \int_{-1}^{0-} \sin\frac{1}{x}\,dx + \int_{0+}^{1} \sin\frac{1}{x}\,dx = \int_{-\infty}^{-1} \frac{\sin u}{u^2}\,du + \int_{1}^{+\infty} \frac{\sin u}{u^2}\,du.$$

Die beiden letzten Integrale existieren aber nicht im *eigentlichen* Sinn des R–Integrals, da kein *endliches* Integrationsintervall vorliegt. Solche *uneigentlichen* R–Integrale werden wir in Abschnitt 8.4 behandeln. □

BSP. (8.3.6) Wir konstruieren hier das R–Integral einer Funktion mit *abzählbar unendlich vielen Sprungstellen*. Dazu zerlegen wir das Intervall $[0,1]$ in der folgenden Weise:

$$I_0 := [0,0], \quad I_1 := [2^{-1}, 1], \quad I_2 := [2^{-2}, 2^{-1}), \ \text{ allgemein: } \ I_j := [2^{-j}, 2^{-j+1}),\ j \ge 2.$$

Dann gelten $|I_0| = 0$ und $|I_j| = 2^{-j}\ \forall\ j \ge 1$ sowie

$$\bigcup_{j=0}^{\infty} I_j = [0,1], \quad I_j \cap I_k = \emptyset,\ j \neq k, \quad \sum_{j=0}^{\infty} |I_j| = \sum_{j=1}^{\infty} \Big(\frac{1}{2}\Big)^j = 1.$$

Für eine beschränkte Zahlenfolge $(y_j)_{j\ge 0} \subset \mathbf{R}$ definieren wir die Funktion $f : [0,1] \to \mathbf{R}$ gemäß

$$f(x) := \sum_{j=0}^{\infty} y_j \chi_{I_j}(x) \ \ \forall\ x \in [0,1].$$

Dies ist eine Treppenfunktion, die auf dem Intervall $[0,1]$ abzählbar unendlich viele Sprungstellen hat, und zwar in den Stellen $x_j := 2^{-j},\ j \in \mathbf{N}$. Das R–Integral von f über dem Intervall $[0,1]$ ist elementargeometrisch bestimmt durch

$$\int_0^1 f(x)\,dx = \sum_{j=0}^{\infty} y_j |I_j| = \sum_{j=1}^{\infty} \frac{y_j}{2^j}.$$

Diese Reihe ist konvergent, denn aus der Beschränktheit $|y_j| \le K < +\infty$ der Folge $(y_j)_{j\ge 0}$ erschließen wir

$$\Big|\sum_{j=1}^{\infty} \frac{y_j}{2^j}\Big| \le K \sum_{j=1}^{\infty} \frac{1}{2^j} = K.$$

BSP. (8.3.7) Ist die Funktion $f : [a,b] \to \mathbf{R}$ beschränkt, und ist sie stetig mit Ausnahme einer *monotonen* Folge $(x_j)_{j\in\mathbf{N}} \subset [a,b]$ von Unstetigkeitsstellen mit dem Häufungspunkt $s \in [a,b]$, so ist f auf dem Intervall $[a,b]$ R–integrierbar. Nehmen wir zum Beispiel an, es gelte $a \le x_0 < x_1 < \cdots < x_j < x_{j+1} < \cdots < s \le b$. Dann existieren wegen Satz 8.15 die Integrale $R_j := \int_{x_{j-1}}^{x_j} f(x)\,dx,\ j \in \mathbf{N}$. Wir erweitern (3.18) in der folgenden Weise:

$$\int_a^b f(x)\,dx = \int_a^{x_0} f(x)\,dx + \sum_{j=1}^{\infty} R_j + \int_s^b f(x)\,dx.$$

Gilt $|f(x)| \le M\ \forall\ x \in [a,b]$, so haben wir wegen

$$\Big|\sum_{j=1}^{\infty} R_j\Big| \le \sum_{j=1}^{\infty} \int_{x_{j-1}}^{x_j} |f(x)|\,dx \le M(s - x_0)$$

Konvergenz der obigen Reihe vorliegen.

BSP. (8.3.8) Im Gegensatz zu BSP. (8.3.7) ist die DIRICHLET–Funktion auf dem Intervall $[a,b]$

$$f(x) := \begin{cases} 1 \ : \ x \in \mathbf{Q} \cap [a,b], \\ 0 \ : \ x \in [a,b],\ \text{ irrational} \end{cases}$$

nicht R–integrierbar, obwohl f beschränkt ist und auf der abzählbar unendlichen Menge $\mathbf{Q} \cap [a,b]$ Unstetigkeitsstellen besitzt. Die Funktion f ist aber in keinem Punkt $x \in [a,b]$ stetig, siehe BSP. (6.3.6), Abschnitt 6.3. Wir geben eine Folge endlicher Zerlegungen $Z_n = \{I_1, I_2, \ldots, I_n\}$ des Intervalls $[a,b]$ mit $|Z_n| \to 0$ vor. In der RIEMANN–Summe

$$S_{Z_n} = \sum_{j=1}^{n} f(\xi_j)\,|I_j|, \quad \xi_j \in I_j,$$

kann die Zwischenstelle ξ_j entweder stets in einem rationalen Punkt des Intervalls I_j gewählt werden; dann gilt $S_{Z_n} = b - a > 0$. Oder $\xi_j \in I_j$ wird in einen irrationalen Punkt gelegt; dann gilt $S_{Z_n} = 0$. Die Folge der RIEMANN–Summen hat zwei Häufungspunkte, sie kann deshalb nicht konvergieren.

Bemerkung 8.8 Die DIRICHLET–Funktion ist auf keinem Intervall $[a,b]$ R–integrierbar. Dieses Phänomen zählt zu den Merkwürdigkeiten der Menge $\mathbf{Q}$ der rationalen Zahlen: Die Menge $\mathbf{Q}$ ist zwar unendlich, aber nur abzählbar, vgl. Definition 5.3. In diesem Sinne ist $\mathbf{Q}$ eine **kleine** Menge. Sie ist andererseits auch eine **große** Menge, denn $\mathbf{Q}$ liegt ja **dicht** in $\mathbf{R}$, vgl. Satz 1.18. Dort hatten wir gezeigt, dass es zu jeder Zahl $x \in \mathbf{R}$ und jedem $\epsilon > 0$ stets eine rationale Zahl $r \in \mathbf{Q}$ gibt mit $|x - r| < \epsilon$. Wir beleuchten einen dritten Aspekt, der wiederum zeigt, dass $\mathbf{Q}$ eine **kleine** Menge ist. Dazu nehmen wir an, die abzählbare Menge $\mathbf{Q}$ sei in der Form $\mathbf{Q} = \{r_1, r_2, r_3 \ldots\}$ durchnumeriert. Zu $\epsilon > 0$ werde

$$I_k := \Big(r_k - \frac{\epsilon}{2^{k+1}}, r_k + \frac{\epsilon}{2^{k+1}}\Big), \quad k = 1, 2, \ldots$$

gesetzt, so dass $r_k \in I_k$ und mithin $\mathbf{Q} \subset \bigcup_{k=1}^{\infty} I_k$ gelten. Die Menge $\mathbf{Q}$ wird von den Intervallen I_k der Länge $|I_k| = \epsilon/2^k$ **überdeckt**, und deren Gesamtlänge beträgt

$$\sum_{k=1}^{\infty} |I_k| = \epsilon \sum_{k=1}^{\infty} 2^{-k} = \epsilon.$$

Das heißt, die Menge $\mathbf{Q}$ wird von abzählbar vielen Intervallen mit beliebig kleiner Gesamtlänge vollständig überdeckt. Mengen mit dieser Eigenschaft heißen **Nullmengen**. □

Definition 8.8 *Eine Teilmenge $M \subset \mathbf{R}$ heiße* **Nullmenge** *oder* **Menge vom Maße** 0*, wenn es zu jedem $\epsilon > 0$ höchstens abzählbar viele (abgeschlossene oder offene) Intervalle $I_1, I_2, \ldots$ gibt mit $M \subset \bigcup\limits_k I_k$ und $\sum\limits_k |I_k| \leq \epsilon$.*

Mit der in Bemerkung 8.8 vorgenommenen Konstruktion zeigt man, dass jede **endliche** oder **abzählbare** Teilmenge von $\mathbf{R}$ eine Nullmenge ist. Ebenso einsichtig ist die Aussage, dass jede Teilmenge einer Nullmenge wieder Nullmenge ist.

Satz 8.16 (a) *Jede* **endliche** *oder* **abzählbare** *Teilmenge von $\mathbf{R}$ ist eine Nullmenge.*

(b) *Jede Teilmenge einer Nullmenge ist eine Nullmenge.*

(c) *Die Vereinigung höchstens abzählbar vieler Nullmengen ist eine Nullmenge.*

(d) *Eine* **abgeschlossene beschränkte** *Teilmenge $M \subset \mathbf{R}$ ist genau dann Nullmenge, wenn es zu jedem $\epsilon > 0$ stets* **endlich viele** *(abgeschlossene oder offene) Intervalle $I_1, I_2, \ldots, I_N$, $N = N(\epsilon)$, gibt mit $M \subset \bigcup\limits_{k=1}^{N} I_k$ und $\sum\limits_{k=1}^{N} |I_k| \leq \epsilon$.*

Begründungen: (c) Es sei $J \subset \mathbf{N}$ eine (endliche oder unendliche) Indexmenge, und es seien $M_j \subset \mathbf{R}$, $j \in J$, Nullmengen. Dann existieren zu jedem $\epsilon > 0$ und zu M_j Intervalle $I_{j1}, I_{j2}, \ldots$ mit $M_j \subset \bigcup\limits_k I_{jk}$ und $\sum\limits_k |I_{jk}| \leq \epsilon/2^j$, $j \in J$. Hieraus folgern wir:

$$\bigcup_{j\in J} M_j \subset \bigcup_{j\in J} \Big(\bigcup_k I_{jk} \Big), \qquad \sum_{j\in J} \sum_k |I_{jk}| \leq \epsilon \sum_{j=1}^{\infty} 2^{-j} = \epsilon.$$

Also ist $\bigcup\limits_{j\in J} M_j$ eine Nullmenge.

(d) Gilt die im Satz angegebene endliche Überdeckungseigenschaft, so ist M trivialerweise eine Nullmenge. Sei andererseits M als Nullmenge vorgegeben. Dann gibt es zu jedem $\epsilon > 0$ Intervalle $I_1, I_2, \ldots$ mit $M \subset \bigcup\limits_k I_k$. Die abgeschlossene beschränkte Teilmenge M unterliegt dem Überdeckungssatz von Heine–Borel (siehe Ing.–Math. III): Es reichen bereits **endlich** viele der I_k zur Überdeckung von M aus. Deren Gesamtlänge ist erst recht $\leq \epsilon$. □

Bemerkung 8.9 Die Aussagen von Satz 8.16 implizieren keineswegs, dass Nullmengen ausschließlich abzählbar sind. Es gibt sehr wohl überabzählbare Nullmengen. □

Grundlegend für die weitere Analysis von Funktionen $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ sind Eigenschaften, die **fast überall** auf $D(f) \subset \mathbf{R}$ gelten, wie *zum Beispiel* Stetigkeitseigenschaften oder Differenzierbarkeit.

Definition 8.9 *Eine Eigenschaft* (E) *gelte für eine Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$* **fast überall auf einer Menge** *$X \subseteq D(f)$, wenn es eine Nullmenge $M \subset X$ gibt, so dass $f(x)$ die Eigenschaft* (E) *in* **allen** *Punkten $x \in X \setminus M$ besitzt.*

BSP. (8.3.9) (a) Die Funktionen $f(x)$ aus den BSPn (8.3.6) und (8.3.7) sind auf den Intervallen $X := [0, 1]$ bzw. $X := [a, b]$ jeweils fast überall stetig. Denn die Teilmenge $M \subset X$ der Unstetigkeitspunkte ist in beiden Fällen abzählbar, also gemäß Satz 8.16 eine Nullmenge.

(b) Die Dirichlet–Funktion aus BSP. (8.3.8) ist in **keinem** Punkt stetig. Eine zu (a) analoge Aussage gilt hier nicht.

Wir können nach diesen Vorbereitungen einen Satz formulieren, der die Klasse der R–integrierbaren Funktionen genau charakterisiert, und der eine Erweiterung des Satzes 8.15 darstellt.

Satz 8.17 (Integrabilitätskriterium von Lebesgue)
Eine Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{K})$ ist genau dann auf dem endlichen Intervall $[a, b]$ R–integrierbar, wenn sie dort beschränkt und fast überall stetig ist.

Für eine Begründung verweisen wir auf die Standardliteratur, z.B. H. HEUSER, Lehrbuch der Analysis, Teil 1, B.G. Teubner Verlag, Stuttgart 1980, S.471.

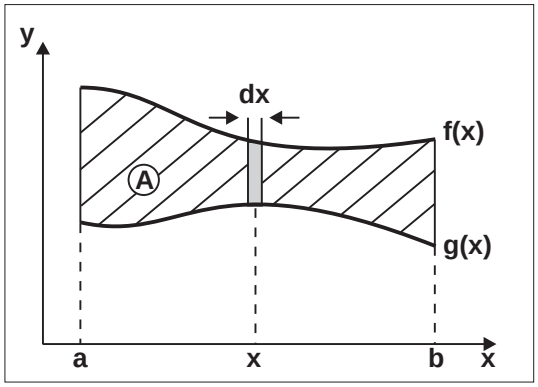
Bemerkung 8.10 (a) Es gibt andere Integraldefinitionen als die des R–Integrals, so *zum Beispiel* das LEBESGUE–**Integral**, mit welchem wir uns in Abschnitt 8.6 befassen werden. Solche Definitionen umfassen in der Regel eine größere Klasse integrierbarer Funktionen als dies die Klasse der R–integrierbaren Funktionen tut. Dabei sind aber stets die am Anfang formulierten Forderungen 1 und 2 zu erfüllen. Das heißt, Funktionen, die für verschiedene Integraldefinitionen integrierbar sind, müssen bei jedem Verfahren denselben Integralwert haben. In der Theorie des LEBESGUE–Integrals werden wir zeigen, dass die obige DIRICHLET–Funktion LEBESGUE–integrierbar ist zum Integralwert 0. Im LEBESGUEschen Sinn ändert nämlich das Integral $\int_a^b f(x)\,dx$ seinen Integralwert *nicht*, wenn der Integrand $f(x)$ auf einer Teilmenge $M \subset [a,b]$ **vom Maße Null** abgeändert wird. Da die Menge $M := \mathbf{Q} \cap [a,b]$ diese Eigenschaft besitzt, haben die beiden Funktionen $f(x)$ und 0 auf $[a,b]$ denselben (LEBESGUEschen) Integralwert. Dieser ist der (RIEMANNsche) Integralwert 0 der R–integrierbaren Funktion 0.

(b) Auch das RIEMANN–Integral $\int_a^b f(x)\,dx$ ändert seinen Wert nicht, wenn der Integrand $f(x)$ auf einer Teilmenge $M \subset [a,b]$ vom Maße Null so abgeändert wird, dass f insgesamt beschränkt und außerhalb M stetig bleibt. Dies ist eine Folgerung aus Satz 8.17. □

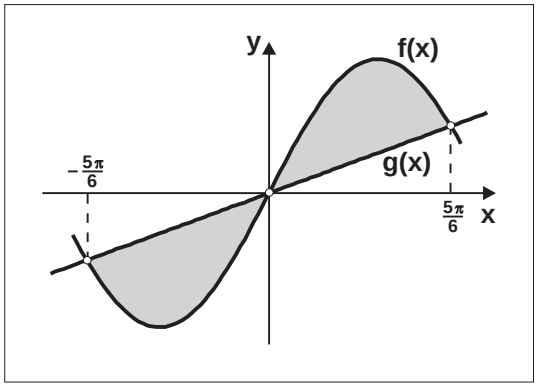
8.4 Anwendungen der Integralrechnung

(I) Flächeninhalte. Sind zwei R–integrierbare Funktionen $f, g : [a,b] \to \mathbf{R}$ gegeben, so ist der **geometrische Flächeninhalt** A zwischen den Graphen $G(f)$ und $G(g)$ in der folgenden Weise definiert:

$$\boxed{A = \int_a^b \Big(f(x) - g(x)\Big)\,dx, \quad \text{sofern} \quad f(x) \geq g(x)\ \forall\, x \in [a,b].}$$



Der Flächeninhalt zwischen zwei Kurven $G(f)$ und $G(g)$



Der Flächeninhalt zwischen $\sin x$ und $3x/5\pi$

BSP. (8.4.1) Gesucht ist der Inhalt A des (endlichen) Flächenstückes zwischen den Graphen der Funktionen $f(x) := \sin x$ und $g(x) := 3x/5\pi$. *Lösung:* Man überlegt sich graphisch, dass $f(x) = g(x)$ genau für $x_{-1} := -5\pi/6$, $x_0 := 0$, $x_1 := 5\pi/6$ erfüllt ist, und dies bestätigt man leicht mit einer Wertetabelle.

x	$-\frac{5\pi}{6}$	0	$\frac{5\pi}{6}$	2π
$f(x)$	$-\frac{1}{2}$	0	$\frac{1}{2}$	0
$g(x)$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{6}{5} > 1$

Für $x \geq 2\pi$ gilt $g(x) \geq 6/5 > 1$, so dass keine weiteren Nullstellen der Funktion $f - g$ auftreten. Es folgt aus dieser Vorüberlegung:

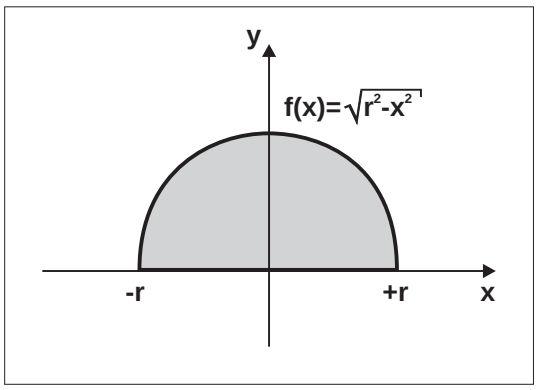
$$A = \int\limits_{-5\pi/6}^{0} \Big(\frac{3x}{5\pi} - \sin x\Big)\, dx + \int\limits_{0}^{5\pi/6} \Big(\sin x - \frac{3x}{5\pi}\Big)\, dx = \Big[\frac{3x^2}{10\pi} + \cos x\Big]_{-5\pi/6}^{0} - \Big[\cos x + \frac{3x^2}{10\pi}\Big]_{0}^{5\pi/6} = 2 + \sqrt{3} - \frac{5\pi}{12}.$$

BSP. (8.4.2) Der Flächeninhalt eines Halbkreises vom Radius $r > 0$ ist mit den Mitteln der Integralrechnung zu bestimmen. *Lösung:* Aus der Gleichung $x^2 + y^2 = r^2$ der Kreislinie vom Radius $r > 0$ erhält man für den oberen Halbkreisbogen die explizite Darstellung $y = f(x) = \sqrt{r^2 - x^2}$, $-r \leq x \leq r$. Das Integral

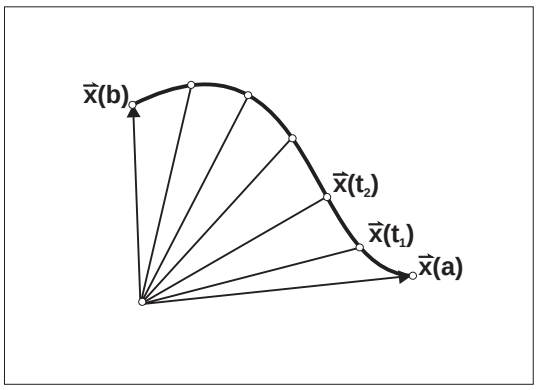
$$A = \int\limits_{-r}^{r} f(x)\, dx = 2r \int\limits_{0}^{r} \sqrt{1 - (x/r)^2}\, dx$$

berechnet man mit Hilfe der Substitution $x = r\, \sin u,\ dx = r\, \cos u\, du$:

$$A = 2r^2 \int\limits_{0}^{\pi/2} \cos^2 u\, du = r^2 \int\limits_{0}^{\pi/2} (1 + \cos 2u)\, du = r^2 \Big[u + \frac{1}{2}\, \sin 2u\Big]_{0}^{\pi/2} = \frac{1}{2}\, r^2 \pi.$$



Flächeninhalt eines Halbkreises



Zum Inhalt einer Fläche, die von einer ebenen Kurve $\vec{x}(t)$ berandet wird

BSP. (8.4.3) Zu bestimmen ist der Inhalt A derjenigen Fläche, die von der ebenen, stetig differenzierbaren Kurve $\vec{x} : [a, b] \to \mathbf{R}^2$ berandet wird, und die vom Ortsvektor $\vec{x}(t)$ im Intervall $a \leq t \leq b$ überstrichen wird. In der folgenden Überlegung wird die Tatsache verwendet, dass der Flächeninhalt des von zwei Vektoren $\vec{x}_1,\ \vec{x}_2 \in \mathbf{R}^3$ aufgespannten Dreiecks durch $\frac{1}{2}\, \|\vec{x}_1 \times \vec{x}_2\|$ bestimmt ist. Es sei nun für ein $n \in \mathbf{N}$ eine *äquidistante Zerlegung* Z_n des Intervalls $[a, b]$ in der folgenden Weise induziert:

$$h := \frac{b - a}{n}, \quad t_j := a + jh,\ j = 0, 1, \ldots, n.$$

Eine Näherungssumme für den gesuchten Flächeninhalt A ist nun offenbar (vgl. obige Skizze):

$$S_{Z_n} = \frac{1}{2} \sum_{j=1}^{n} \Big\|\vec{x}\big[a + (j-1)h\big] \times \vec{x}(a + jh)\Big\| = \frac{h}{2} \sum_{j=1}^{n} \Big\|\vec{x}\big[a + (j-1)h\big] \times \frac{1}{h}\Big(\vec{x}(a + jh) - \vec{x}\big[a + (j-1)h\big]\Big)\Big\|,$$

wobei wir die allgemeingültige Relation $\vec{y} \times \vec{y} = \vec{0}$ verwendet haben. Da die Vektorfunktion $\vec{x}(t)$ stetig differenzierbar ist, erhält man im Limes $h \to 0$ aus dieser Folge von RIEMANN–Summen das RIEMANN–Integral

$$\boxed{A = \frac{1}{2} \int\limits_{a}^{b} \|\vec{x}(t) \times \frac{d}{dt}\vec{x}(t)\|\, dt.} \tag{4.1}$$

(a) Bei **allgemeiner Parameterdarstellung** $\vec{x}(t) := \Big(x(t), y(t)\Big)^T$ der ebenen Randkurve mit $x, y \in C^1([a,b])$ resultiert

$$\vec{x}(t) \times \frac{d}{dt}\vec{x}(t) = \begin{vmatrix} \vec{e}_x & x(t) & \dot{x}(t) \\ \vec{e}_y & y(t) & \dot{y}(t) \\ \vec{e}_z & 0 & 0 \end{vmatrix} = \Big(x(t)\dot{y}(t) - y(t)\dot{x}(t)\Big)\,\vec{e}_z,$$

und somit

$$\boxed{A = \frac{1}{2}\int\limits_a^b \Big|x(t)\dot{y}(t) - y(t)\dot{x}(t)\Big|\,dt.} \tag{4.2}$$

BSP. (8.4.4) **Flächeninhalt der Ellipse.** Die Ellipse mit den Halbachsen $a, b > 0$ hat die Parameterdarstellung $x(t) = a\cos t,\ y(t) = b\sin t,\ t \in [0, 2\pi]$. Hieraus erhält man $x(t)\dot{y}(t) - y(t)\dot{x}(t) = ab\,(\cos^2 t + \sin^2 t) = ab$, also

$$\boxed{A_{Ell.} = \frac{1}{2}\int\limits_0^{2\pi} ab\,dt = \pi ab.}$$

Im Grenzfall $a = b = r$ resultiert der Flächeninhalt $A_{Kr} = \pi r^2$ eines Kreises vom Radius $r > 0$.

(b) In **Polarkoordinaten** $x(\varphi) = r(\varphi)\,\cos\varphi,\ y(\varphi) = r(\varphi)\,\sin\varphi$ mit $r \in C^1([\varphi_a, \varphi_b])$ gelten die Beziehungen

$$\dot{x}(\varphi) = \dot{r}(\varphi)\,\cos\varphi - r(\varphi)\,\sin\varphi, \quad \dot{y}(\varphi) = \dot{r}(\varphi)\,\sin\varphi + r(\varphi)\,\cos\varphi,$$

aus denen wir $x(\varphi)\dot{y}(\varphi) - y(\varphi)\dot{x}(\varphi) = r^2(\varphi)$ erhalten. Deshalb gilt nun

$$\boxed{A = \frac{1}{2}\int\limits_{\varphi_a}^{\varphi_b} r^2(\varphi)\,d\varphi.} \tag{4.3}$$

BSP. (8.4.5) **Flächeninhalt der Kardioide.** Die Kardioide ist in Polarkoordinaten durch die Gleichung $r(\varphi) := a(1 + \cos\varphi),\ \varphi \in [0, 2\pi]$, bestimmt. Es gilt somit

$$A_{Kard.} = 2 \cdot \frac{1}{2}\int\limits_0^{\pi} a^2(1+\cos\varphi)^2\,d\varphi = a^2\int\limits_0^{\pi}\Big(1 + 2\cos\varphi + \frac{1}{2} + \frac{1}{2}\,\cos 2\varphi\Big)\,d\varphi = \frac{3\pi}{2}\,a^2.$$

(II) Flächenmomente und Schwerpunkte. Aus der Definition

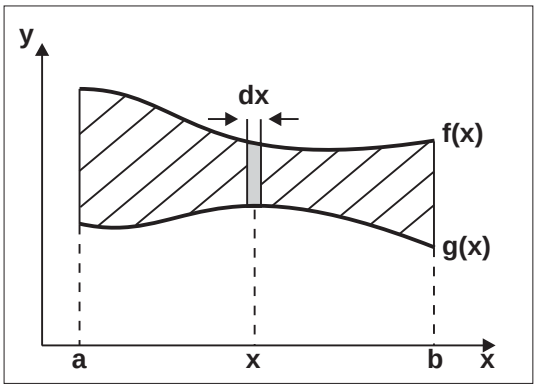
$$\boxed{\textbf{Flächenmoment} := \textbf{Fläche} \times \textbf{Hebelarm}}$$

ergeben sich die beiden folgenden Momente, wenn wir wiederum $f(x) \geq g(x)\ \forall\ x \in [a,b]$ voraussetzen (siehe nachfolgende Skizze):

$$\boxed{\begin{aligned} M_x &:= \int\limits_a^b x\Big(f(x) - g(x)\Big)\,dx, \\ M_y &:= \int\limits_a^b \frac{1}{2}\,[f(x) + g(x)]\,[f(x) - g(x)]\,dx = \frac{1}{2}\int\limits_a^b \Big(f^2(x) - g^2(x)\Big)\,dx. \end{aligned}}$$

Im **Schwerpunkt** $\vec{x}_S := (x_S, y_S)^T$ einer Fläche vom Inhalt A gelten die Relationen $A \cdot x_S = M_x,\ A \cdot y_S = M_y$, und daraus erhält man die Schwerpunktskoordinaten

$$\boxed{x_S = \frac{M_x}{A}, \quad y_S = \frac{M_y}{A}.}$$



Zur Definition der Flächenmomente

BSP. (8.4.6) Es sind die Flächenmomente und der Schwerpunkt des oberen Halbkreises mit Radius $r > 0$ und Mittelpunkt $(0,0)$ zu bestimmen. *Lösung:* Wir haben hier $f(x) := \sqrt{r^2 - x^2}$, $-r \le x \le r$, und $g(x) := 0$. Da der Integrand von M_x eine ungerade Funktion ist, resultiert:

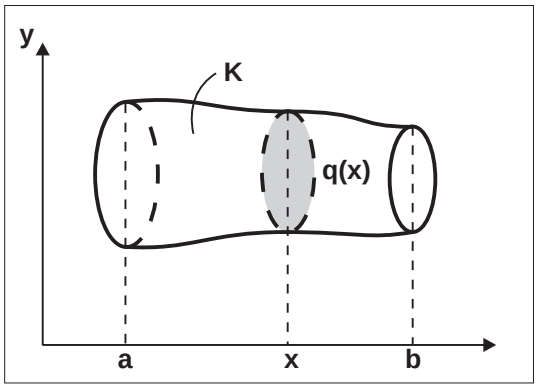
$$M_x = \int_{-r}^{r} x\sqrt{r^2 - x^2}\, dx = 0,$$

$$M_y = \int_{-r}^{r} \frac{1}{2}\,(r^2 - x^2)\, dx = \int_{0}^{r} (r^2 - x^2)\, dx = \Big[r^2 x - \frac{1}{3}\, x^3\Big]_0^r = \frac{2}{3}\, r^3.$$

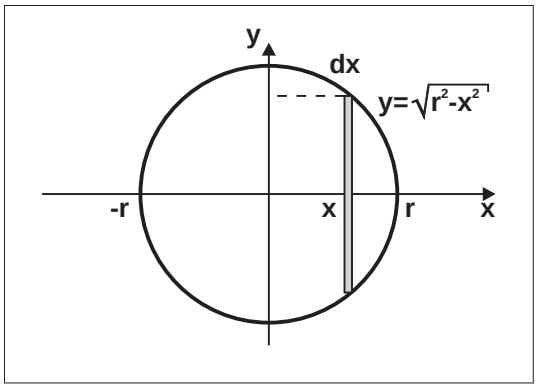
Der Schwerpunkt liegt also im Punkt $\vec{x}_S = (0, \frac{4r}{3\pi})^T$.

(III) Volumenbestimmung. Hat die Schnittfläche eines Körpers K mit der Ebene $x = const$ den Flächeninhalt $q(x)$, so berechnet sich das Gesamtvolumen von K nach dem CAVALIERischen Prinzip:

$$\boxed{V = \int_a^b q(x)\, dx.}$$



Das CAVALIERIsche Prinzip



Zur Volumenberechnung einer Kugel

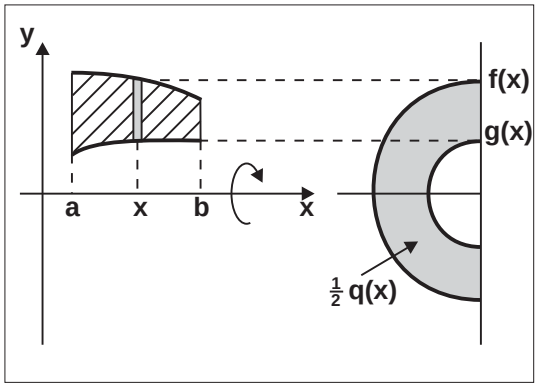
BSP. (8.4.7) Es ist das Volumen einer Kugel vom Radius $r > 0$ nach dem CAVALIERIschen Prinzip zu bestimmen. *Lösung:* Die Schnittfläche der Kugel um den Mittelpunkt $(0,0)$ mit der Ebene $x = const$ hat den Flächeninhalt $q(x) = \pi y^2 = \pi(r^2 - x^2)$. Hieraus folgt

$$\boxed{V_{Kugel} = \pi \int_{-r}^{r} (r^2 - x^2)\, dx = \frac{4}{3}\,\pi r^3.}$$

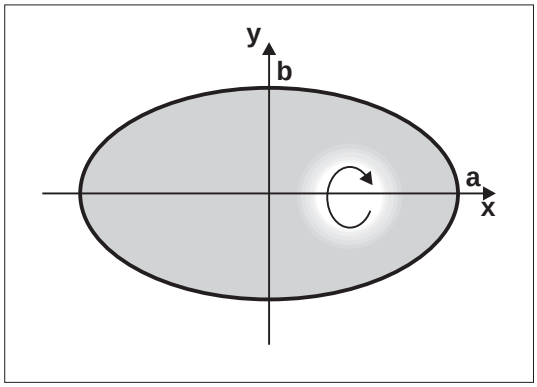
Die Kugel ist der Spezialfall eines **Rotationskörpers**. Für die Volumina von Rotationskörpern gelten folgende Vereinfachungen.

(a) Bei **Rotation um die x-Achse**: Es gelte $f(x) \geq g(x)\ \forall\ x \in [a,b]$. Dann hat die Schnittfläche den Flächeninhalt $q(x) = \pi\Big(f^2(x) - g^2(x)\Big)$, und somit folgt

$$\boxed{V_x = \pi \int_a^b \Big(f^2(x) - g^2(x)\Big)\, dx.}$$



Volumen eines Rotationskörpers bei Rotation um die x–Achse

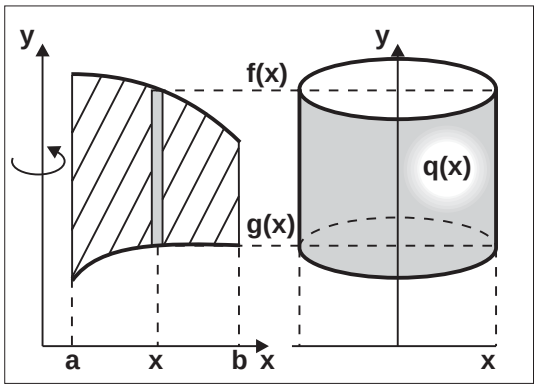


Volumen eines Rotationsellipsoids

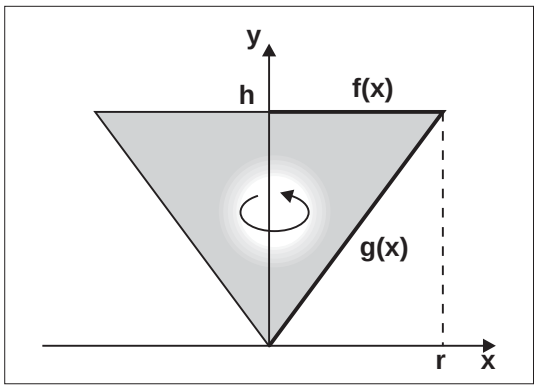
BSP. (8.4.8) Man bestimme das Volumen desjenigen Rotationskörpers, der durch Rotation der Ellipse $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$ um die x–Achse entsteht. *Lösung:* Wir haben hier $f(x) := \sqrt{b^2(1 - x^2/a^2)}$, $g(x) := 0$, $-a \leq x \leq a$, zu setzen. Es folgt

$$\boxed{V_{Ell.} = \pi \int_{-a}^{a} b^2\Big(1 - \frac{x^2}{a^2}\Big)\, dx = \frac{4}{3}\,\pi b^2 a.}$$

Hier ist auch der Sonderfall der Kugelvolumens mit $a = b = r$ enthalten.



Volumen eines Rotationskörpers bei Rotation um die y–Achse



Volumen eines geraden Kreiskegels

(b) Bei **Rotation um die y–Achse**: Es gelte $f(x) \geq g(x)\ \forall\ x \in [a,b]$. Dann ist $q(x)$ ein Zylindermantel mit dem Flächeninhalt $q(x) = 2\pi x\Big(f(x) - g(x)\Big)$, und somit folgt

$$\boxed{V_y = 2\pi \int_a^b x\Big(f(x) - g(x)\Big)\, dx.}$$

BSP. (8.4.9) Man bestimme das Volumen eines geraden Kreiskegels der Höhe h und Basiskreisradius $r > 0$. *Lösung:* Wir haben hier $f(x) := h$, $g(x) := hx/r$, $0 \le x \le r$, zu setzen. Es gilt nun $x\Big(f(x) - g(x)\Big) = h(x - x^2/r)$, und somit

$$\boxed{V_{Kegel} = 2\pi h \int_0^r \Big(x - \frac{x^2}{r}\Big)\, dx = \frac{1}{3}\,\pi r^2 h.}$$

Bemerkung 8.11 Vergleicht man die Volumina V_x, V_y mit den Formeln (II) für die Flächenmomente M_x, M_y, so erhält man:

$$\boxed{V_x = 2\pi \cdot M_y = 2\pi \cdot A \cdot y_S, \quad V_y = 2\pi \cdot M_x = 2\pi \cdot A \cdot x_S.}$$

Dieses Ergebnis heißt die GULDINsche **Regel** für Rotationskörper: □

$$\boxed{\textbf{Volumen des Rotationskörpers} = \textbf{Flächeninhalt} \times \textbf{Schwerpunktweg der Fläche.}}$$

(VI) Volumenmomente, Schwerpunkt und Trägheitsmomente. Aus der Definition

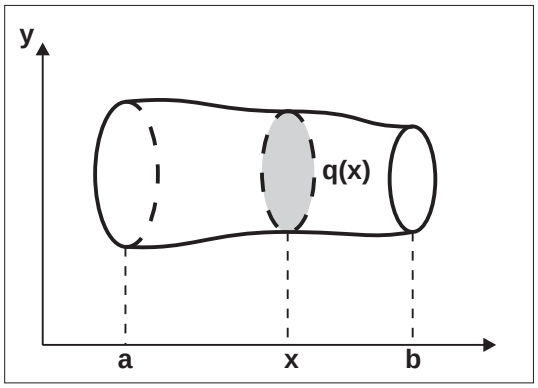
$$\boxed{\textbf{Volumenmoment} := \textbf{Volumen} \times \textbf{Hebelarm}}$$

ergeben sich in der Standardbasis des $\mathbf{R}^3$ drei Volumenmomente M_x, M_y, M_z. Verwenden wir wiederum das CAVALIERIsche Prinzip, so ergibt sich *zum Beispiel*

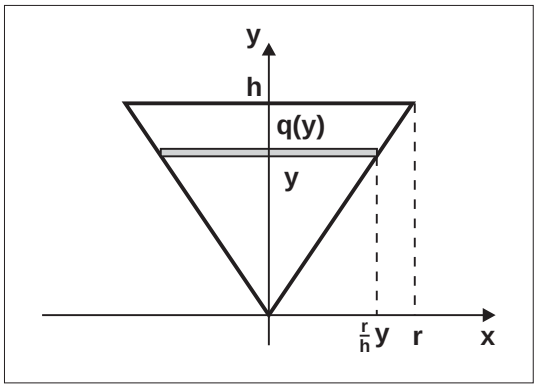
$$\boxed{M_x := \int_a^b xq(x)\, dx.}$$

Im **Schwerpunkt** $\vec{x}_S := (x_S, y_S, z_S)^T$ eines Volumens V gelten die Relationen $V \cdot x_S = M_x$, $V \cdot y_S = M_y$, $V \cdot z_S = M_z$, und daraus erhält man die Schwerpunktskoordinaten

$$\boxed{x_S = \frac{M_x}{V}, \quad y_S = \frac{M_y}{V}, \quad z_S = \frac{M_z}{V}.}$$



Zur Definition des Volumen–momentes M_x

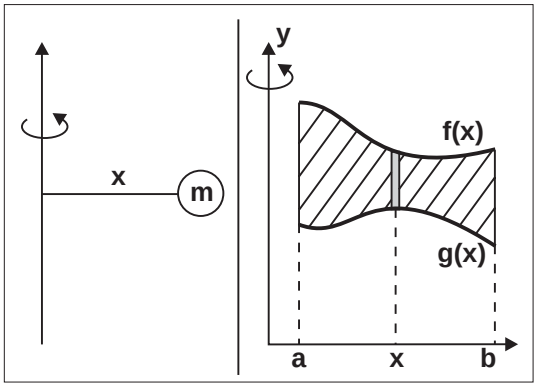


Volumenmoment M_y des geraden Kreiskegels

Bei Rotationskörpern liegt der Schwerpunkt aus Symmetriegründen **stets auf der Rotationsachse.**

BSP. (8.4.10) Man berechne den Schwerpunkt eines geraden Kreiskegels der Höhe h mit Spitze im Ursprung und Basiskreisradius $r > 0$. *Lösung:* Verwenden wir die Bezeichnungen der obigen Skizze, so muss der Schwerpunkt auf der y–Achse liegen. Wir haben hier $q(y) := \pi r^2 y^2/h^2$, und somit

$$M_y = \int_0^h y\, q(y)\, dy = \frac{\pi r^2}{h^2} \int_0^h y^3\, dy = \frac{\pi}{4}\, r^2 h^2, \quad y_S = \frac{M_y}{V_{Kegel}} = \frac{3}{4}\, h,\ x_S = z_S = 0.$$



Zur Definition von Trägheitsmomenten bei Rotationskörpern

Trägheitsmomente entstehen, wenn eine Masse m im Abstand x um eine feste Achse rotiert:

$$\boxed{\Theta := mx^2.}$$

Betrachtet man einen **Rotationskörper** mit Rotationsachse y oder x (siehe obige Skizze), der eine in der Schnittfläche $x = const$ konstante Dichteverteilung $\rho = \rho(x)$ hat, so erhält man die folgenden Trägheitsmomente.

(a) Bei **Rotation um die y–Achse:**

$$\boxed{\Theta = 2\pi \int_a^b \rho(x) x^3 \Big(f(x) - g(x)\Big)\, dx.}$$

(b) Bei **Rotation um die x–Achse:**

$$\boxed{\Theta = \frac{\pi}{2} \int_a^b \rho(x) \Big(f^2(x) - g^2(x)\Big)\Big(f^2(x) + g^2(x)\Big)\, dx = \frac{\pi}{2} \int_a^b \rho(x) \Big(f^4(x) - g^4(x)\Big)\, dx.}$$

BSP. (8.4.11) Die Gesamtmasse und das Trägheitsmoment des obigen Kreiskegels sind bei homogener Dichteverteilung $\rho = const$ zu bestimmen. *Lösung:* Die Gesamtmasse beträgt $m = \rho \cdot V_{Kegel} = \frac{1}{3}\pi\rho h r^2$. Für das Trägheitsmoment hingegen folgt

$$\Theta = 2\pi\rho \int_0^r h x^3 \Big(1 - \frac{x}{r}\Big)\, dx = \pi\rho h \frac{r^4}{10} = \frac{3}{10}\, r^2\, m.$$

(V) Integralmittelwerte.

Definition 8.10 *Für eine gegebene Funktion* $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$ *heiße die Zahl*

$$\boxed{\bar{f} := \frac{1}{b-a}\int\limits_a^b f(x)\,dx}$$

*der (Integral–)***Mittelwert** *von* f *auf dem Intervall* $[a,b]$*, sofern das R–Integral existiert. Das* **quadratische Mittel** *von* f *auf* $[a,b]$ *ist die Zahl*

$$\boxed{\|f\|_2 := \Big(\frac{1}{b-a}\int\limits_a^b f^2(x)\,dx\Big)^{1/2},}$$

ebenfalls unter der Voraussetzung der Existenz dieses Integrals.

BSP. (8.4.12) Man berechne die Mittelwerte eines Wechselstroms $I(t) := I_1\,\sin\omega_1 t + I_2\,\sin\omega_2 t$. Die Kreisfrequenzen ω_j sollen rational teilbar sein: $\omega_1/\omega_2 = r > 0$ mit $r \in \mathbf{Q}$. In diesem Fall haben die Teilströme $I_j(t) := I_j\,\sin\omega_j t$, $j = 1,2$, eine kleinste gemeinsame Periode T. Ist nämlich $T_j := 2\pi/\omega_j$, $j = 1,2$, die Periode von $I_j(t)$, so muss es für eine kleinste gemeinsame Periode T Zahlen $p, q \in \mathbf{N}$ geben mit

$$T = pT_1 = qT_2, \quad \text{also } \omega_1 T = 2\pi p,\ \omega_2 T = 2\pi q, \quad \text{oder } \frac{\omega_1}{\omega_2} = \frac{p}{q} = r.$$

Hieraus erschließen wir:

$$\begin{aligned}
\bar{I} &= \frac{1}{T}\int\limits_0^T (I_1\,\sin\omega_1 t + I_2\,\sin\omega_2 t)\,dt = -\Big[\frac{I_1}{T\omega_1}\,\cos\omega_1 t + \frac{I_2}{T\omega_2}\,\cos\omega_2 t\Big]_0^T \\
&= \frac{I_1}{T\omega_1}[1-\cos 2\pi p] + \frac{I_2}{T\omega_2}[1-\cos 2\pi q] = 0, \\
\|I\|_2^2 &= \frac{1}{T}\int\limits_0^T \Big(I_1^2\,\sin^2\omega_1 t + I_2^2\,\sin^2\omega_2 t + 2I_1 I_2\,\sin\omega_1 t \sin\omega_2 t\Big)\,dt \\
&= \frac{1}{T}\int\limits_0^T \Big(\frac{I_1^2+I_2^2}{2} - \frac{1}{2}I_1^2\,\cos 2\omega_1 t - \frac{1}{2}\,I_2^2\,\cos 2\omega_2 t + I_1 I_2[\cos(\omega_1-\omega_2)t - \cos(\omega_1+\omega_2)t]\Big)\,dt \\
&= \frac{1}{2}\,(I_1^2 + I_2^2).
\end{aligned}$$

Hieraus erhält man die **effektive Stromstärke** $\|I\|_2 = \frac{1}{2}\,\sqrt{2(I_1^2+I_2^2)}$.

Zwischen den beiden Mittelwerten einer Funktion f auf einem Intervall $[a,b]$ besteht die folgende Relation:

$$\boxed{|\bar{f}| = \Big|\frac{1}{b-a}\int\limits_a^b f(x)\,dx\Big| \le \Big(\frac{1}{b-a}\int\limits_a^b f^2(x)\,dx\Big)^{1/2} = \|f\|_2.}$$

Diese Ungleichung ergibt sich als Spezialfall $g(x) := 1/(b-a)$ aus der folgenden

Satz 8.18 (Ungleichung von Schwarz**)**
Für je zwei R–integrierbare Funktionen $f, g : [a,b] \to \mathbf{R}$ *gilt:*

$$\boxed{\Big|\int\limits_a^b f(x)g(x)\,dx\Big| \le \int\limits_a^b |f(x)g(x)|\,dx \le \Big(\int\limits_a^b f^2(x)\,dx\Big)^{1/2}\Big(\int\limits_a^b g^2(x)\,dx\Big)^{1/2}.}$$

Begründung: Wir brauchen nur den hinteren Teil der Ungleichung zu zeigen. Dazu sei eine Folge Z_n endlicher Zerlegungen des Intervalls $[a,b]$ gegeben mit $|Z_n| \to 0$. Unter Verwendung der Ungleichung von CAUCHY–SCHWARZ (Satz 1.13) erhält man für jede Zwischenstelle $\xi_j \in I_j \in Z_n$:

$$\sum_{j=1}^{n} |f(\xi_j)g(\xi_j)|\,|I_j| \le \Big(\sum_{j=1}^{n} f^2(\xi_j)\,|I_j|\Big)^{1/2}\Big(\sum_{j=1}^{n} g^2(\xi_j)\,|I_j|\Big)^{1/2}.$$

Im Limes $|Z_n| \to 0$ erhält man daraus die behauptete Ungleichung. □

(VI) Integral–Restglied der TAYLOR–Formel und erweiterter Mittelwertsatz.
Der Mittelwertsatz der Integralrechnung (Satz 8.11) gestattet die folgende Verallgemeinerung:

Satz 8.19 (Zweiter oder erweiterter MWS der Integralrechnung)
Gegeben seien eine stetige Funktion $f \in C([a,b])$ und eine R–integrierbare Funktion $g : [a,b] \to \mathbf{R}$. Falls überall auf $[a,b]$ entweder $g \le 0$ oder $g \ge 0$ gilt, so existiert eine Zwischenstelle $\xi \in [a,b]$ mit

$$\boxed{\int_a^b f(x)g(x)\,dx = f(\xi)\int_a^b g(x)\,dx.}$$

Begründung: Für $g = 0$ ist nichts zu zeigen. Gelte also *zum Beispiel* $g \ge 0$ und $\int_a^b g(x)\,dx > 0$. Wir setzen $m = \min_{x\in[a,b]} f(x)$ und $M = \max_{x\in[a,b]} f(x)$, so dass folgt:

$$m\int_a^b g(x)\,dx \le \int_a^b f(x)g(x)\,dx \le M\int_a^b g(x)\,dx.$$

Das heißt, die Zahl $\eta := \int_a^b f(x)g(x)\,dx \Big/ \int_a^b g(x)\,dx$ liegt im Intervall $[m,M]$. Da die Funktion f auf Grund des Zwischenwertsatzes von BOLZANO jeden Wert in diesem Intervall erreicht, gibt es eine Zwischenstelle $\xi \in [a,b]$ mit $f(\xi) = \eta$. □

Satz 8.20 (TAYLOR–Formel und Integral–Restglied)
Gegeben sei eine Funktion $f \in C^{n+1}([a,b])$. Dann gilt in jedem festen Punkt $x_0 \in [a,b]$ und für jedes $x \in [a,b]$:

$$\boxed{f(x) = \sum_{k=0}^{n} \frac{1}{k!}\,f^{(k)}(x_0)\,(x-x_0)^k + \frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,dt.}$$

Begründung: Wir wenden Satz 8.19 auf das obige Integral–Restglied an. Es existiert eine Zwischenstelle $\xi := x_0 + \theta(x-x_0)$, $\theta \in [0,1]$ mit

$$\frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,dt = f^{(n+1)}(\xi)\,\frac{1}{n!}\int_{x_0}^{x}(x-t)^n\,dt = \frac{(x-x_0)^{n+1}}{(n+1)!}\,f^{(n+1)}(\xi).$$

Rechts steht gerade das Restglied von LAGRANGE, vgl. Satz 7.19. □

Bemerkung 8.12 Durch partielle Integration des Integral–Restglieds erhält man das triviale Ergebnis

$$\frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,dt = f(x) - \sum_{k=0}^{n}\frac{1}{k!}\,f^{(k)}(x_0)\,(x-x_0)^k.$$

Deshalb kann das Integral–Restglied keinesfalls zur Berechnung des exakten Fehlers zwischen der Funktion $f(x)$ und dem TAYLOR–Polynom $T_n(x)$ vom Grad n verwendet werden. Wie das LAGRANGE–Restglied, ist das Integral–Restglied aber sehr gut für eine Fehlerabschätzung geeignet. □

BSP. (8.4.13) Die Funktion $f(x) := \ln(1+x)$ hat die Ableitungen $f^{(k)}(x) = (-1)^{k+1}(k-1)!/(1+x)^k,\ k \geq 1$. Daraus berechnet man an der Stelle $x_0 = 0$:

$$\ln(1+x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}x^k}{k} + \frac{1}{n!}\int_0^x (-1)^n n! \frac{(x-t)^n}{(1+t)^{n+1}}\,dt.$$

Für $x > 0$ schätzt man sehr einfach ab:

$$\Big|\int_0^x (-1)^n \frac{(x-t)^n}{(1+t)^{n+1}}\,dt\Big| \leq \int_0^x (x-t)^n\,dt = \frac{x^{n+1}}{n+1}.$$

Die gleiche Abschätzung gilt in diesem Bereich auch für das LAGRANGE–Restglied:

$$|R_n(x;x_0)| = \Big|\frac{n!x^{n+1}}{(n+1)!(1+\xi)^{n+1}}\Big| \leq \frac{x^{n+1}}{n+1},\quad x > 0.$$

(VII) Uneigentliche Integrale. In unseren bisherigen Überlegungen zum RIEMANN–Integral haben wir stets vorausgesetzt, dass das Integrationsintervall $[a,b] \subset \mathbf{R}$ **endlich** ist und dass der Integrand $f(x)$ auf $[a,b]$ **beschränkt** ist. Wir werden zeigen, dass zu beiden Voraussetzungen Ausnahmen erlaubt sind, die wir hier unter dem Begriff des **uneigentlichen** RIEMANN–**Integrals** zusammenfassen. Die beiden folgenden Beispiele charakterisieren bereits die wichtigsten Typen uneigentlicher Integrale.

BSP. (8.4.14) Für eine feste Zahl $a \in \mathbf{R}$ und für $b > a$ betrachten wir

$$I := \lim_{b\to+\infty} \int_a^b e^{-x}\,dx = \lim_{b\to+\infty} \Big[-e^{-x}\Big]_a^b = e^{-a} - \lim_{b\to+\infty} e^{-b} = e^{-a}.$$

Wir schreiben formal $I = \int_0^{+\infty} e^{-x}\,dx = e^{-a}$. Das Integrationsintervall $[a,+\infty)$ ist also **unbeschränkt**; es liegt hier kein R–Integral *im eigentlichen Sinn* vor.

BSP. (8.4.15) Wir betrachten nun die auf dem Intervall $[0,1)$ stetige Funktion $f(x) := 1/\sqrt{1-x^2}$, die wegen $\lim_{x\to 1-0} f(x) = +\infty$ auf $[0,1]$ **unbeschränkt** ist. Das R–Integral darf auf $[0,1]$ *im eigentlichen Sinn* nicht existieren. Wir haben jedoch für $0 < \epsilon < 1$:

$$I := \lim_{\epsilon\to 0+} \int_0^{1-\epsilon} \frac{dx}{\sqrt{1-x^2}} = \lim_{\epsilon\to 0+} \text{arc sin}_H(1-\epsilon) = \frac{\pi}{2}.$$

Definition 8.11 *Sind im Integralausdruck*

$$\boxed{\int_a^b f(x)\,dx} \tag{4.4}$$

nicht *beide Integrationsgrenzen a und b* **endlich**, *oder ist der Integrand $f(x)$* **nicht** *an beiden (endlichen) Intervallenden a und b* **beschränkt**, *so dass* $\lim_{x\to b-0} |f(x)| = +\infty$ *und/oder*

$\lim\limits_{x\to a+0}|f(x)| = +\infty$ *gilt, so heiße das Integral* (4.4) *ein* **uneigentliches (RIEMANN–)Integral**. *Existieren die Grenzwerte*

$$\boxed{I := \lim_{b\to+\infty}\int_a^b f(x)\,dx \quad bzw. \quad \lim_{a\to-\infty}\int_a^b f(x)\,dx \quad bzw. \quad \lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_a^b f(x)\,dx,}$$

oder

$$\boxed{I := \lim_{\epsilon\to0+}\int_a^{b-\epsilon} f(x)\,dx \quad bzw. \quad \lim_{\epsilon\to0+}\int_{a+\epsilon}^b f(x)\,dx \quad bzw. \quad \lim_{\epsilon\to0+}\int_{a+\epsilon}^{b-\epsilon} f(x)\,dx,}$$

so heiße das uneigentliche Integral (4.4) **konvergent zum Integralwert** I. *Existiert auch das uneigentliche Integral* $\int\limits_a^b |f(x)|\,dx$ *in dem oben präzisierten Sinn, so heiße das uneigentliche Integral* (4.4) **absolut konvergent.**

Bemerkung 8.13 Mit Hilfe der Ungleichung $\left|\int\limits_a^b f(x)\,dx\right| \le \int\limits_a^b |f(x)|\,dx$ zeigt man, dass absolute Konvergenz stets auch einfache Konvergenz impliziert. Die Umkehrung gilt im allgemeinen nicht. Ein bekanntes Beispiel für diesen Sachverhalt ist das einfach konvergente Integral

$$I := \int_0^{+\infty}\frac{\sin x}{x}\,dx,$$

welches **nicht** absolut konvergent ist. □

BSP. (8.4.16) Wir betrachten das **doppelt uneigentliche Integral**

$$I := \int_0^{+\infty}\frac{dx}{x^p} = \int_0^1\frac{dx}{x^p} + \int_1^{+\infty}\frac{dx}{x^p} = \lim_{\epsilon\to0+}\int_\epsilon^1\frac{dx}{x^p} + \lim_{b\to+\infty}\int_1^b\frac{dx}{x^p} =: I_1 + I_2.$$

Es gilt hier:

$$I_1 = \lim_{\epsilon\to0+}\int_\epsilon^1\frac{dx}{x^p} = \begin{cases}\dfrac{1}{1-p} & : p<1,\\ +\infty & : p\ge1,\end{cases} \qquad I_2 = \lim_{b\to+\infty}\int_1^b\frac{dx}{x^p} = \begin{cases}\dfrac{1}{p-1} & : p>1,\\ +\infty & : p\le1.\end{cases}$$

Das heißt, das uneigentliche Integral I existiert für keinen Exponenten $p\in\mathbf{R}$. Hingegen existieren die Teilintegrale I_1 und I_2 auf verschiedenen Teilbereichen $p\in\mathbf{R}$.

Das Hauptproblem besteht im Nachweis der Konvergenz des uneigentlichen Integrals (4.4), **ohne** auf eine Stammfunktion des Integranden f zurückgreifen zu müssen. Mit dem folgenden zentralen Satz kann häufig diese Konvergenzfrage geklärt werden.

Satz 8.21 (Vergleichskriterium)
Die gegebenen Funktionen f, g *seien auf jedem endlichen Teilintervall* $[\alpha,\beta]\subset(a,b)$ *R–integrierbar, und es gelte* $0\le f(x)\le g(x)\ \forall\ x\in(a,b)$.

(a) *Konvergiert das uneigentliche Integral*

$$\boxed{\int_a^b g(x)\,dx := \lim_{\substack{\alpha\to a+0\\ \beta\to b-0}}\int_\alpha^\beta g(x)\,dx,}$$

so ist auch $\int_a^b f(x)\,dx$ *konvergent.*

(b) *Divergiert das uneigentliche Integral* $\int_a^b f(x)\,dx$*, so ist auch* $\int_a^b g(x)\,dx$ *divergent.*

Begründungen: Wir setzen

$$\beta_n := \begin{cases} n & : \ b = +\infty, \\ b - 1/n & : \ b < +\infty, \end{cases} \qquad \alpha_n := \begin{cases} -n & : \ a = -\infty, \\ a + 1/n & : \ a > -\infty. \end{cases}$$

Hierin sei $n \in \mathbf{N}$ eine hinreichend große Zahl. Wegen $f \geq 0$ ist die Zahlenfolge $I_n := \int_{\alpha_n}^{\beta_n} f(x)\,dx$ monoton steigend und nach oben durch $\int_a^b g(x)\,dx$ beschränkt. Daher folgt aus dem Hauptsatz über monotone Folgen (Satz 3.3) die Konvergenz. □

Das Vergleichskriterium ist besonders geeignet für den Nachweis der **absoluten Konvergenz** uneigentlicher Integrale. Man vergleicht $|f(x)|$ mit einer Funktion $g(x) \geq 0$, für die das Konvergenzverhalten des Integrals $\int_a^b g(x)\,dx$ bekannt ist. Häufig wird die Vergleichsfunktion $g(x) := C/x^p$ verwendet und die Ergebnisse von BSP. (8.4.16) genutzt.

BSP. (8.4.17) (i) Für die Funktion $f(x) := \frac{\sin x}{x^p} \cdot \operatorname{sign}(\cos x)$ gilt auf dem Intervall $[1, +\infty)$ die Ungleichung $|f(x)| \leq 1/x^p =: g(x)$. Aus BSP. (8.4.16) erhalten wir $\int_1^{+\infty} g(x)\,dx < +\infty$ genau für $p > 1$. Deshalb ist das uneigentliche Integral

$$I := \int_1^{+\infty} \frac{\sin x}{x^p} \cdot \operatorname{sign}(\cos x)\,dx$$

für $p > 1$ absolut konvergent.

(ii) Wir betrachten auf dem Intervall $(0, 1]$ den Integranden $f(x) := \frac{\sinh x}{x^p}$. Wegen $(\sinh x)/x \geq 1$ erhalten wir $f(x) \geq 1/x^{p-1} =: g(x)$. Wiederum aus BSP. (8.4.16) resultiert $\int_0^1 g(x)\,dx = +\infty$ für $p \geq 2$, so dass das uneigentliche Integral

$$I := \int_0^1 \frac{\sinh x}{x^p}\,dx$$

für jedes $p \geq 2$ divergent ist. Hingegen folgt aus der Stetigkeit der Funktion $(\sinh x)/x$ im Punkte $x = 0$ die Existenz von $M := \max_{x \in [0,1]} \frac{\sinh x}{x}$. Wir haben deshalb $0 \leq f(x) \leq M/x^{p-1} =: g(x)$, und aus BSP. (8.4.16) folgt $\int_0^1 g(x)\,dx < +\infty$ für $p < 2$. Das heißt, das uneigentliche Integral I ist konvergent für jedes $p < 2$.

Das Vergleichskriterium kann also unter Verwendung der Vergleichsfunktionaus BSP. (8.4.16) in folgenden Regeln zusammengefasst werden, wobei wir nur die obere Integrationsgrenze als uneigentlich betrachten. Analoge Regeln für die untere Integrationsgrenze lassen sich leicht ergänzen.

Konvergenz uneigentlicher Integrale. Es seien $a \in \mathbf{R}$ und $b \leq +\infty$ gegeben, und die Funktion $f(x)$ sei stetig auf jedem Teilintervall $[a, \beta] \subset [a, b)$. Dann gelten folgende Implikationen:

(a) $x^p|f(x)| \leq C < +\infty \ \forall\, x \geq R > a$ mit $p > 1$ $\quad \Rightarrow \quad \int_a^{+\infty} f(x)\,dx$ **absolut konvergent**

(b) $x^p|f(x)| \geq C > 0 \ \forall\, x \geq R > a$ mit $p \leq 1$ $\quad \Rightarrow \quad \int_a^{+\infty} f(x)\,dx$ **divergent**

(c) $0 \leq (b - x)^p|f(x)| \leq C < +\infty \ \forall\, x \in [a, b]$ mit $p < 1$ $\quad \Rightarrow \quad \int_a^b f(x)\,dx$ **absolut konvergent**

(d) $(b - x)^p|f(x)| \geq C > 0 \ \forall\, x \in [a, b]$ mit $p \geq 1$ $\quad \Rightarrow \quad \int_a^b f(x)\,dx$ **divergent**.

BSP. (8.4.18) Wir untersuchen das uneigentliche Integral $I := \int_0^1 (1-x^2)^{-p}\,dx$. Wir setzen $f(x) := (1-x^2)^{-p} = (1+x)^{-p}(1-x)^{-p}$. Nun gilt:

$$(1-x)^p|f(x)| = (1+x)^{-p} \le 1 \ \ \forall\, x \in [0,1] \quad \overset{(c)}{\Rightarrow} \quad I \ \text{ist konvergent für } p<1,$$

$$(1-x)^p|f(x)| = (1+x)^{-p} \ge \frac{1}{2^p} \ \ \forall\, x \in [0,1] \quad \overset{(d)}{\Rightarrow} \quad I \ \text{ist divergent für } p \ge 1.$$

Den Sonderfall $p = 1/2$ hatten wir bereits in BSP. (8.4.15) behandelt.

BSP. (8.4.19) Wir untersuchen das uneigentliche Integral $I := \int_0^{+\infty} (1+x^2)^{-p}\,dx$. Wir setzen $f(x) := (1+x^2)^{-p}$. Dann gilt:

$$x^{2p}|f(x)| = (1+\frac{1}{x^2})^{-p} \le 1 \ \ \forall\, x \ge 1 \quad \overset{(a)}{\Rightarrow} \quad I \ \text{ist konvergent für } p>1/2,$$

$$x^{2p}|f(x)| = (1+\frac{1}{x^2})^{-p} \ge \frac{1}{2^p} \ \ \forall\, x \ge 1 \quad \overset{(b)}{\Rightarrow} \quad I \ \text{ist divergent für } p \le 1/2.$$

Im Sonderfall $p = 1$ haben wir

$$\int_0^{+\infty} \frac{dx}{1+x^2} = \lim_{b\to+\infty} \int_0^b \frac{dx}{1+x^2} = \lim_{b\to+\infty} \operatorname{arc\,tan}_H b = \frac{\pi}{2}.$$

Das Vergleichskriterium Satz 8.21 ist ein *hinreichendes* Kriterium für die **absolute Konvergenz** uneigentlicher Integrale. Es kann beispielsweise nicht verwendet werden bei dem Integral

$$\int_0^{+\infty} \frac{\sin \alpha x}{x}\,dx, \ \ \alpha > 0,$$

dessen **einfache Konvergenz** zum Integralwert $\pi/2$ bekannt ist, während absolute Konvergenz nicht vorliegt. Dieses Beispiel passt aber in den Rahmen des folgenden Konvergenzsatzes.

Satz 8.22 *Die Funktion $f : [a,+\infty) \to \mathbf{R}$, $a > 0$, sei stetig, und es sei $F(x) := \int_a^x f(t)\,dt$ beschränkt:*

$$\boxed{\exists\, C > 0 \ : \ |F(x)| \le C < +\infty \ \ \forall\, x \ge a.}$$

Dann ist das uneigentliche Integral

$$\boxed{\int_a^{+\infty} \frac{f(x)}{x^p}\,dx}$$

für alle $p > 0$ konvergent.

Begründung: Durch partielle Integration folgt für jede Zahl $b > a$:

$$\int_a^b \frac{f(x)}{x^p}\,dx = \frac{F(x)}{x^p}\Big|_a^b + p\int_a^b \frac{F(x)}{x^{1+p}}\,dx = \frac{F(b)}{b^p} + p\int_a^b \frac{F(x)}{x^{1+p}}\,dx.$$

Wegen $|F(x)| \le C$ ergibt sich hieraus im Limes $b \to +\infty$:

$$\Big|\int_a^{+\infty} \frac{f(x)}{x^p}\,dx\Big| \le Cp\int_a^{+\infty} \frac{dx}{x^{1+p}} = \frac{C}{a^p} < +\infty.$$

BSP. (8.4.20) Das FRESNEL–Integral $I := \int_0^{+\infty} \sin x^2\,dx$ lässt sich mit der Substitution $u = g(x) := x^2$, $du = 2\sqrt{u}\,dx$ in das folgende Integral transformieren:

$$I = \frac{1}{2}\int_0^{+\infty} \frac{\sin u}{\sqrt{u}}\,du = \frac{1}{2}\int_0^{\pi/2} \frac{\sin u}{\sqrt{u}}\,du + \frac{1}{2}\int_{\pi/2}^{+\infty} \frac{\sin u}{\sqrt{u}}\,du =: I_1 + I_2.$$

Die Funktion $(\sin u)/\sqrt{u}$ ist auf dem Intervall $[0, \pi/2]$ stetig, und deshalb existiert das Integral I_1. Im Integral I_2 setzen wir $f(u) := \sin u$. Dann sind mit $a := \pi/2$ und

$$F(x) := \int_{\pi/2}^{x} \sin u\,du = -\cos x, \quad |F(x)| \le 1, \quad x \ge a,$$

die Voraussetzungen von Satz 8.22 erfüllt. Dieser liefert die Konvergenz des uneigentlichen Integrals I_2.

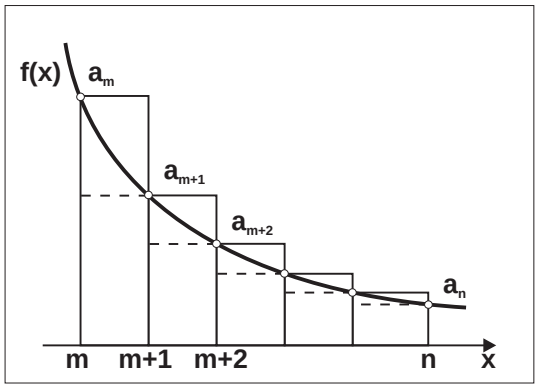
(VIII) Das Integralvergleichskriterium von CAUCHY. Mit Hilfe von uneigentlichen Integralen können wir jetzt ein Konvergenzkriterium für unendliche Zahlenreihen formulieren.

Satz 8.23 (Integralvergleichskriterium)
Für festes $m \in \mathbf{N}_0$ sei $f : [m, +\infty) \to \mathbf{R}$ eine stetige, monoton fallende, positive Funktion, und es gelte $a_k := f(k)\ \forall\ k = m, m+1, \ldots$ Dann haben die unendliche Reihe $\sum_{k=m}^{\infty} a_k$ und das uneigentliche Integral $\int_m^{+\infty} f(x)\,dx$ dasselbe Konvergenzverhalten.

Begründung: Für $n > m$ entnimmt man der folgenden Skizze die Abschätzung

$$\sum_{k=m}^{n} a_k - a_m < \int_m^n f(x)\,dx < \sum_{k=m}^{n-1} a_k.$$

Hieraus erschließt man die Konvergenz der Reihe, sofern das uneigentliche Integral konvergiert sowie Divergenz der Reihe, wenn das uneigentliche Integral divergiert. □



Zum Integralvergleichskriterium

Aus dem obigen Beweisgang erhält man unter den Voraussetzungen des Satzes 8.23 eine **Fehlerabschätzung** für den Reihenrest $\sum_{k=N}^{\infty} a_k$. Für jede Zahl $N \ge m$ und für $n > N$ gilt ja

$$\sum_{k=N}^{n} a_k - a_N < \int_N^n f(x)\,dx < \sum_{k=N}^{n-1} a_k.$$

Im Limes $n \to +\infty$ resultiert daraus die Fehlereinschließung

$$\boxed{\int_N^{+\infty} f(x)\,dx \le \sum_{k=N}^{\infty} a_k \le a_N + \int_N^{+\infty} f(x)\,dx.} \tag{4.5}$$

BSP. (8.4.21) Wir setzen $f(x) := 1/x^p$ für $x \ge 1$. Für $p > 0$ und $x \ge 1$ gilt $f'(x) = -px^{-p-1} < 0$, so dass die Funktion f monoton fallend, stetig und positiv ist. Wir können das Integralvergleichskriterium anwenden. Das uneigentliche Integral

$$\int_1^{+\infty} \frac{dx}{x^p} = \begin{cases} 1/(p-1) < +\infty & : \ p > 1, \\ +\infty & : \ p \le 1, \end{cases}$$

führt zu folgender Konvergenzaussage:

$$\boxed{\sum_{k=1}^{\infty} \frac{1}{k^p} \begin{cases} p > 1 \ : \ \textbf{konvergent}, \\ p \le 1 \ : \ \textbf{divergent}. \end{cases}}$$

BSP. (8.4.22) Für $x \ge 2$ setzen wir $f(x) := (\ln x)/x^2$. Dann gilt $f'(x) = (1 - 2\ln x)/x^3 < 0 \ \forall \ x \ge 2$ (man beachte $2\ln 2 = \ln 4 \doteq 1.386 > 1$). Somit ist die Funktion f monoton fallend, stetig und positiv. Wir können wiederum das Integralvergleichskriterium anwenden. Das uneigentliche Integral

$$\int_2^{+\infty} \frac{\ln x\,dx}{x^2} \overset{\text{part. Int.}}{=} -\frac{\ln x}{x}\Big|_2^{+\infty} + \int_2^{+\infty} \frac{dx}{x^2} = \frac{1}{2}(\ln 2 + 1) < +\infty$$

ist konvergent, und somit konvergiert auch die unendliche Reihe

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

Problem: Wie groß ist die Zahl N zu wählen, damit der Summenwert der obigen Reihe durch die Partialsumme $\sum_{k=2}^{N-1} \frac{\ln k}{k^2}$ mit einer Genauigkeit $\epsilon := 10^{-5}$ approximiert wird? *Lösung:* Wir verwenden die Fehlerabschätzung (4.5) aus dem Integralvergleichskriterium. Es folgt

$$F := \sum_{k=N}^{\infty} \frac{\ln k}{k^2} \le \frac{\ln N}{N^2} + \int_N^{+\infty} f(x)\,dx = \frac{\ln N}{N^2} + \frac{1}{N}(\ln N + 1) \overset{!}{\le} 10^{-5}.$$

Man ermittelt mit dem Taschenrechner $\int_N^{+\infty} f(x)dx \doteq 1.4816 \cdot 10^{-5}$ für $N = 10^6$ sowie $a_N + \int_N^{+\infty} f(x)dx \doteq 1.7118 \cdot 10^{-6}$ für $N = 10^7$, so dass die gesuchte Zahl N zwischen 10^6 und 10^7 liegen muss.

8.5 Numerische Integration

Ziel dieses Abschnitts ist es, einige wichtige Verfahren zur näherungsweisen Berechnung eines bestimmten Integrals

$$I := \int_a^b f(x)\,dx$$

bereitzustellen. Wir gehen in der Regel von der Vorgabe einer reellen, R–integrierbaren Funktion $f : [a, b] \to \mathbf{R}$ aus. Das Problem besteht in der Bestimmung einer Näherung Q von I, so dass der Fehler $R = I - Q$ der Bedingung $|R| < \epsilon$ mit gegebener Toleranz $\epsilon > 0$ genügt.

Definition 8.12 *Eine endliche Rechenvorschrift für die Berechnung der Zahl Q, so dass die Relation*

$$\boxed{I := \int_a^b f(x)\,dx = Q + R}$$

mit $|R| < \epsilon$ bei vorgegebener Toleranz $\epsilon > 0$ gilt, heißt eine **Quadraturformel**.

Ein einfaches Verfahren zur Konstruktion von Quadraturformeln besteht in der Interpolation des gegebenen Integranden f durch ein geeignetes Polynom, dessen R–Integral exakt berechnet werden kann und die gesuchte Näherung Q liefert. Quadraturformeln auf der Basis dieser Idee heißen **interpolatorische Quadraturformeln**. Sie gehen aus von der Vorgabe einer R–integrierbaren Funktion $f : [a, b] \to \mathbf{R}$ in den $n + 1$ *Stützstellen*

$$a \le x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n \le b, \tag{5.1}$$

mit den *Stützwerten* $f_k := f(x_k) \;\; \forall\; k = 0, 1, \ldots, n$. Die Konstruktion von Quadraturformeln besteht nun darin, die Funktion $f(x)$ in den *Knotenpunkten* (x_k, f_k) durch ein einfaches Polynom zu interpolieren und danach die Interpolierende exakt zu integieren. Wir werden uns hier ausschließlich mit den

$$\boxed{\text{Newton–Côtes–}\textbf{Quadraturformeln}}$$

befassen. Diese Formeln basieren auf der Verwendung des Lagrange–Interpolationspolynoms $P_n(x)$ in den Knotenpunkten $(x_k, f_k),\; k = 0, 1, \ldots, n$. Gemäß Satz 6.2 ist das Polynom $P_n(x)$ eindeutig bestimmt, und es kann in der folgenden Form dargestellt werden, vgl. Abschnitt 6.2:

$$P_n(x) = \sum_{k=0}^{n} f_k L_k(x) \quad \text{mit} \quad L_k(x) := \prod_{\substack{j=0 \\ j \ne k}}^{n} \frac{(x - x_j)}{(x_k - x_j)} \;\; \forall k = 0, 1, 2, \ldots, n. \tag{5.2}$$

Durch Integration von $P_n(x)$ erhält man jetzt eine Quadraturformel

$$\boxed{Q_n := \int_a^b P_n(x)\,dx = \sum_{k=0}^{n} f_k \int_a^b L_k(x)\,dx =: (b-a)\sum_{k=0}^{n} w_k f(x_k).} \tag{5.3}$$

Definition 8.13 *Die nur von den Stützstellen $x_0, x_1, \ldots, x_n$ und von der Differenz $b - a$ abhängigen Größen*

$$\boxed{w_k := \frac{1}{b-a} \int_a^b L_k(x)\,dx \;\; \forall\; k = 0, 1, \ldots, n,} \tag{5.4}$$

heißen **Integrationsgewichte der Quadraturformel** (5.3) **zu den Stützstellen** x_k.

Bemerkung 8.14 Ist $f(x)$ ein *Polynom* mit Grad $f \le n$, so ist es klar, dass die Quadraturformel (5.3) nach Konstruktion den exakten Wert I liefert. Im allgemeinen Fall erhält man durch Q_n eine Näherung des Integralwertes I mit einem Fehler

$$R_n[f] := I - Q_n = \int_a^b f(x)\,dx - (b-a)\sum_{k=0}^{n} w_k f_k. \tag{5.5}$$

Wir werden Beispiele finden, in denen die Quadraturformel Q_n auch noch einen exakten Integralwert liefert für Polynome vom Grade $> n$. □

Definition 8.14 *Eine Quadraturformel Q_n vom Typ* (5.3) *besitzt den* **Genauigkeitsgrad** $m \in \mathbf{N}$*, wenn Q_n alle Polynome bis zum Grad m exakt integriert, wobei m maximal sei.*

Bemerkung 8.15 Die Quadraturformel Q_n hat den Genauigkeitsgrad $m \in \mathbf{N}$ genau dann, wenn gilt:

$$R_n[x^j] = 0 \ \ \forall\ j = 0, 1, \ldots, m \quad \text{und} \quad R_n[x^{m+1}] \neq 0. \tag{5.6}$$

Klar, denn dies ist eine Konsequenz aus der Linearität des Fehlerfunktionals $R_n[f]$ bezüglich f. □

Zusammenfassend hat man den folgenden

Satz 8.24 *Gegeben seien $n+1$ Stützstellen x_k in der Anordnung* (5.1)*. Dann gibt es genau eine interpolatorische Quadraturformel*

$$\boxed{Q_n = (b-a)\sum_{k=0}^{n} w_k f(x_k) \quad \textit{mit} \quad w_k := \frac{1}{b-a}\int_a^b L_k(x)\,dx \ \ \forall\ \ k = 0, 1, \ldots, n,} \tag{5.7}$$

deren Genauigkeitsgrad mindestens gleich n ist.

Sonderfälle: Wir betrachten **äquidistante Stützstellen**

$$h := \frac{b-a}{n};\ \ x_j := a + jh,\ f_j := f(x_j)\ \ \forall\ j = 0, 1, \ldots, n. \tag{5.8}$$

Es ist in diesem Falle zweckmäßig, die Quadraturformel (5.3) in der folgenden Form zu schreiben:

$$\boxed{Q_n = h\sum_{k=0}^{n} \alpha_k f(x_k) \quad \text{mit} \quad \alpha_k := \int_0^n \prod_{\substack{j=0 \\ j\neq k}}^{n} \frac{t-j}{k-j}\,dt \ \ \forall\, k = 0, 1, \ldots, n.} \tag{5.9}$$

Diese Darstellung ergibt sich aus (5.7) durch die Substitution $x = a + ht$.

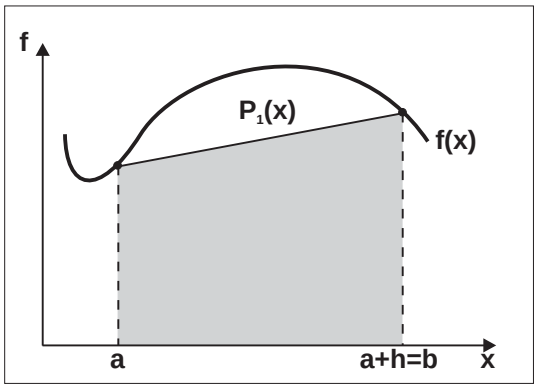
Definition 8.15 *Die Formeln* (5.9) *heißen* **geschlossene** NEWTON–CÔTES**–Quadraturformeln***, da sie sowohl Anfangs– als auch Endpunkt des Intervalls $[a, b]$ als Stützstelle verwenden.*

Bemerkung 8.16 Da das Polynom $P_0(x) \equiv 1$ von jeder Quadraturformel Q_n exakt integriert wird, haben wir wegen (5.8) und (5.9) stets □

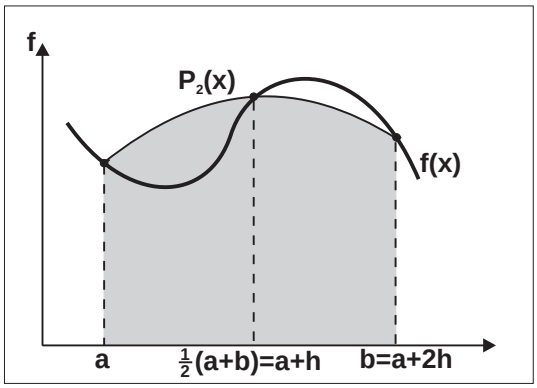
$$n = \sum_{k=0}^{n} \alpha_k. \tag{5.10}$$

BSP. (8.5.1) Die **Trapez–Regel** resultiert aus (5.9) im Sonderfall $\boxed{n = 1.}$ Die Interpolierende $P_1(x)$ ist eine Gerade durch die Knotenpunkte $(a, f(a))$, $(b, f(b))$. Man erhält mit einfacher Rechnung

$$\boxed{T(h) := Q_1 = \frac{b-a}{2}\,[f(a) + f(b)] = \frac{h}{2}\,[f_0 + f_1].} \tag{5.11}$$



Die Trapez–Regel



Die SIMPSON–Regel

BSP. (8.5.2) Die SIMPSON–**Regel** (oder KEPLERsche Faßregel) erhält man aus (5.9) im Sonderfall $n = 2$. Die Interpolierende $P_2(x)$ ist eine quadratische Parabel durch die Knotenpunkte

$$(a, f(a)),\ \left(\frac{a+b}{2}, f\left(\frac{a+b}{2}\right)\right),\ (b, f(b)).$$

Man findet in diesem Fall:

$$\boxed{S(h) := Q_2 = \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] = \frac{h}{3}\left[f_0 + 4f_1 + f_2\right].} \tag{5.12}$$

Beachte: Es gilt für das Fehlerfunktional

$$R_2[x^3] = \int_a^b x^3\,dx - \frac{b-a}{6}\left[a^3 + 4\left(\frac{a+b}{2}\right)^3 + b^3\right] = 0, \quad R_2[x^4] \neq 0.$$

Das heißt, die SIMPSON–Regel hat den Genauigkeitsgrad $m = 3$, obwohl zu ihrer Herleitung nur ein Polynom vom Grade 2 verwendet wurde.

Die Beobachtung bei der SIMPSON–Regel, dass der Genauigkeitsgrad m größer ist als der Grad des Interpolationspolynoms, trifft allgemein zu, falls n eine **gerade** Zahl ist. Es gilt dann $m = n + 1$. Ist n hingegen **ungerade**, so gilt stets auch $n = m$. Wir teilen diesen Sachverhalt ohne Beweis mit.

Über die Größe des Quadraturfehlers $R_n[f]$ haben wir folgende Information. Gemäß (9.1), Abschnitt 7.9, leistet die Polynom–Interpolation eine Genauigkeit

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\,\varphi_n(x) \quad \text{mit} \quad \varphi_n(x) := \prod_{j=0}^{n}(x - x_j).$$

Dabei ist $\xi = \xi(x) \in (a, b)$ unbekannt. Mit dieser Fehlerrelation gelangt man durch Integration zu folgender Aussage:

$$R_n[f] = \frac{1}{(n+1)!}\int_a^b f^{(n+1)}[\xi(x)]\,\varphi_n(x)\,dx.$$

Die Diskussion dieses Integrals erfordert wegen der Vorzeichenwechsel von $\varphi_n(x)$ in jeder Stützstelle $x_k \in [a, b]$ umfangreiche Hilfsmittel. Wir referieren hier nur das Resultat.

Satz 8.25 *Gegeben sei die Funktion $f \in C^{(n+2)}([a,b])$, und es sei n eine* **gerade** *Zahl. Dann beträgt der Quadraturfehler $R_n[f]$ der* NEWTON–CÔTES*–Quadraturformel (5.9):*

$$\boxed{R_n[f] = \int_a^b f(x)\,dx - Q_n = \frac{K_n}{(n+2)!}\, f^{(n+2)}(\xi), \quad a < \xi < b; \quad K_n := \int_a^b x\phi_n(x)\,dx.} \tag{5.13}$$

Insbesondere werden Polynome $P_m(x)$ vom Grade $m = n+1$ exakt integriert.

Satz 8.26 *Gegeben sei die Funktion $f \in C^{(n+1)}([a,b])$, und es sei n eine* **ungerade** *Zahl. Dann beträgt der Quadraturfehler $R_n[f]$ der* NEWTON–CÔTES*–Quadraturformel (5.9):*

$$\boxed{R_n[f] = \int_a^b f(x)\,dx - Q_n = \frac{K_n}{(n+1)!}\, f^{(n+1)}(\xi), \quad a < \xi < b; \quad K_n := \int_a^b \phi_n(x)\,dx.} \tag{5.14}$$

BSP. (8.5.3) (a) Für $n = 1$ erhalten wir aus Satz 8.26

$$K_1 = \int_a^b (x-a)(x-b)\,dx = -\frac{1}{6}\,(b-a)^3 = -\frac{1}{6}\,h^3.$$

Mit dieser Beziehung resultiert für die **Trapez–Regel** ein Quadraturfehler

$$\boxed{R_1[f] = -h^3 \cdot \frac{1}{12}\, f^{(2)}(\xi), \quad a < \xi < b.} \tag{5.15}$$

(b) Für $n = 2$ ergibt sich aus Satz 8.25

$$K_2 = \int_a^b x(x-a)\left(x - \frac{a+b}{2}\right)(x-b)\,dx = -\frac{1}{120}\,(b-a)^5 = -\frac{4}{15}\,h^5.$$

Das heißt, die SIMPSON–**Regel** integriert mit einem Quadraturfehler von

$$\boxed{R_2[f] = -h^5 \cdot \frac{1}{90}\, f^{(4)}(\xi), \quad a < \xi < b.} \tag{5.16}$$

Wir stellen die NEWTON–CÔTES–Quadraturformeln bis zur Ordnung $n = 6$ in der folgenden Tabelle zusammen. Dabei wählen wir die Zahl $s \in \mathbf{N}$ so, dass $\sigma_j := s \cdot \alpha_j$, $j = 0, 1, \ldots, n$, ganzzahlig wird. In diesem Falle haben wir:

$$\boxed{Q_n = \int_a^b P_n(x)\,dx = h \sum_{k=0}^{n} \alpha_k f_k = \frac{b-a}{n \cdot s} \sum_{k=0}^{n} \sigma_k f_k.}$$

n	σ_k	$n \cdot s$	$R_n[f]$	Name
1	1 1	2	$-h^3 \cdot \frac{1}{12}\, f^{(2)}(\xi)$	**Trapez–Regel**
2	1 4 1	6	$-h^5 \cdot \frac{1}{90}\, f^{(4)}(\xi)$	SIMPSON**–Regel**
3	1 3 3 1	8	$-h^5 \cdot \frac{3}{80}\, f^{(4)}(\xi)$	$\frac{3}{8}$–Regel von NEWTON
4	7 32 12 32 7	90	$-h^7 \cdot \frac{8}{945}\, f^{(6)}(\xi)$	MILNE**–Regel**
5	19 75 50 50 75 19	288	$-h^7 \cdot \frac{275}{12096}\, f^{(6)}(\xi)$	———
6	41 216 27 272 27 216 41	840	$-h^9 \cdot \frac{9}{1400}\, f^{(8)}(\xi)$	WEDDLE**–Regel**

Für $n \geq 8$ treten negative Gewichte σ_k auf, und die NEWTON–CÔTES–Formeln werden numerisch unbrauchbar. Eine Begründung für diesen Sachverhalt ist gegeben durch den folgenden Satz, der eine Aussage über die Konvergenz des Quadraturfehlers $R_n[f]$ trifft.

Satz 8.27 (W.A. STEKLOW)
Es sei eine Folge von interpolatorischen Quadraturformeln gegeben:

$$Q_n := (b-a)\sum_{k=0}^{n} w_k^{(n)} f(x_k^{(n)}), \quad x_k^{(n)} := a + k\,\frac{b-a}{n}, k = 0, 1, \ldots, n; \quad n \in \mathbf{N}.$$

Gibt es eine Zahl K, so dass

$$\sum_{k=0}^{n} |w_k^{(n)}| \leq K \;\; \forall\, n \in \mathbf{N}$$

gilt, so konvergiert die Folge Q_n für jede stetige Funktion $f(x)$ gegen das Integral $\int_a^b f(x)\,dx$.

Begründung: Der Quadraturfehler

$$R_n[f] := \int_a^b f(x)\,dx - Q_n$$

erfüllt für jedes Polynom $P_n(x)$ vom Grade $\leq n$ die Relation $R_n[f] = R_n[f - P_n]$, weil wir Interpolationsquadraturen betrachten. Also wird

$$\begin{aligned} |R_n[f]| &= |R_n[f-P_n]| \leq \int_a^b |f(x) - P_n(x)|\,dx + (b-a)\sum_{k=0}^{n} |w_k^{(n)}| \cdot |f(x_k^{(n)}) - P_n(x_k^{(n)})| \\ &\leq (b-a)(1+K)\max_{a\leq x\leq b} |f(x) - P_n(x)|. \end{aligned}$$

Da jede stetige Funktion $f(x)$ in dem abgeschlossenen Intervall $[a,b]$ beliebig genau durch Polynome approximiert werden kann (Approximationssatz von K. WEIERSTRASS), braucht man nur n hinreichend groß zu wählen, um den letzten Ausdruck beliebig klein zu machen. Dies führt auf die behauptete Konvergenz. □

Bemerkung 8.17 Der häufig vorliegende Fall $w_k^{(n)} \geq 0,\ k = 0, 1, \ldots, n$, führt auf

$$\sum_{k=0}^{n} |w_k^{(n)}| = \sum_{k=0}^{n} w_k^{(n)} = \frac{1}{b-a}\int_a^b dx = 1.$$

Also gilt hier $K = 1$. Bei den NEWTON–CÔTES–Formeln sind die Gewichte für $n \geq 8$ nicht mehr alle positiv, und auch die Betragssumme der Gewichte lässt sich nicht mehr unabhängig von n durch eine Konstante K abschätzen. Aus diesem Grunde werden die NEWTON–CÔTES–Formeln für große n numerisch unbrauchbar. □

BSP. (8.5.4) Wir berechnen die beiden Integrale

$$I_1 := \int_0^1 \frac{1}{1+x^2}\,dx = \frac{\pi}{4} \doteq 7.853\,981\,633\,974\,48\mathrm{E}^{-01}, \quad I_2 := \int_0^1 e^x\,dx = e - 1 \doteq 1.718\,281\,828\,459\,05\mathrm{E}^{+00}$$

näherungsweise mit den NEWTON–CÔTES–Formeln der Ordnung $n = 1$ bis $n = 6$:

Integralwert $I_1 := \int_a^b f(x)\,dx$**:**

h	Q_n	$R_n[f]$
1/1	$7.500\,000\,000\,000\mathrm{E}^{-01}$	$3.539\,816\,339\,744\mathrm{E}^{-02}$
1/2	$7.833\,333\,333\,334\mathrm{E}^{-01}$	$2.064\,830\,063\,993\mathrm{E}^{-03}$
1/3	$7.846\,153\,846\,152\mathrm{E}^{-01}$	$7.827\,787\,822\,263\mathrm{E}^{-04}$
1/4	$7.855\,294\,117\,648\mathrm{E}^{-01}$	$-1.312\,483\,674\,172\mathrm{E}^{-04}$
1/5	$7.854\,696\,044\,812\mathrm{E}^{-01}$	$-7.144\,108\,377\,634\mathrm{E}^{-05}$
1/6	$7.853\,927\,139\,175\mathrm{E}^{-01}$	$5.449\,479\,921\,458\mathrm{E}^{-06}$

Integralwert $I_2 := \int_a^b f(x)\,dx$**:**

Q_n	$R_n[f]$
$1.859\,140\,914\,230\mathrm{E}^{+00}$	$-1.408\,590\,857\,709\mathrm{E}^{-01}$
$1.718\,861\,151\,876\mathrm{E}^{+00}$	$-5.793\,234\,174\,600\mathrm{E}^{-04}$
$1.718\,540\,153\,361\mathrm{E}^{+00}$	$-2.583\,249\,020\,731\mathrm{E}^{-04}$
$1.718\,282\,687\,924\mathrm{E}^{+00}$	$-8.594\,658\,298\,274\mathrm{E}^{-07}$
$1.718\,282\,312\,990\mathrm{E}^{+00}$	$-4.845\,313\,292\,351\mathrm{E}^{-07}$
$1.718\,281\,829\,517\mathrm{E}^{+00}$	$-1.058\,454\,559\,134\mathrm{E}^{-09}$

Insbesondere am Beispiel des Integral I_2 erkennt man, dass die oben angegebene Fehlergesetzmäßigkeit sehr genau erfüllt wird. Es wird ferner die Aussage des Satzes 8.25 bestätigt, dass die geschlossenen NEWTON–CÔTES–Formeln *gerader* Ordnung n wegen ihrer höheren Genauigkeit stets vorteilhafter sind als diejenigen ungerader Ordnung.

Summierte NEWTON–CÔTES–Formeln.

Ist $[a,b]$ ein "großes" Intervall, das heißt, gilt $b-a \gg 1$, so wird die Schrittweite h für kleine Werte n eine relativ große Zahl werden. Die Fehler $R_n[f]$ in den obigen Formeln werden somit nicht mehr vernachlässigbar sein. Abhilfe schafft eine Zerlegung des Intervalls $[a,b]$ in äquidistante Teilintervalle I_j. Auf jedem Teilintervall kann nun eine NEWTON–CÔTES–Formel Q_n angewendet werden.

BSP. (8.5.5) **Verallgemeinerte (oder summierte) Trapez–Regel:** Man setzt

$$h := \frac{b-a}{N}; \quad x_j := a + jh \;\forall\, j = 0,1,\ldots,N; \quad I_j := [x_{j-1}, x_j] \;\forall\, j = 1,\ldots,N.$$

Auf jedem Teilintervall I_j wenden wir die **Trapez**–Regel (5.11) an:

$$T_j(h) := Q_{1j} := \frac{h}{2}\,[f_{j-1} + f_j].$$

Summation über $j = 1,2,\ldots,N$ liefert:

$$\boxed{T(h) = \sum_{j=1}^{N} T_j(h) = \frac{h}{2}\left[f_0 + f_N + 2\sum_{k=1}^{N-1} f_k\right]; \quad f_k := f(x_k).} \tag{5.17}$$

Summiert man die Quadraturfehler (5.15) der Teilintegrale bezüglich I_j, so erhält man eine Fehlerabschätzung

$$\boxed{R_1[f] = \int_a^b f(x)\,dx - T(h) = -\frac{(b-a)}{12}h^2 f^{(2)}(\xi), \;\; a < \xi < b.} \tag{5.18}$$

BSP. (8.5.6) **Verallgemeinerte (oder summierte) SIMPSON–Regel:** Man setzt

$$h := \frac{b-a}{2N}; \quad x_j := a + jh \;\forall\, j = 0,1,\ldots,2N; \quad I_j := \left[x_{2(j-1)}, x_{2j}\right] \;\forall\, j = 1,2,\ldots,N.$$

Auf jedem Teilintervall I_j wenden wir jetzt die SIMPSON–Regel (5.12) an:

$$S_j(h) := Q_{2j} := \frac{h}{3}\,[f_{2j-2} + 4f_{2j-1} + f_{2j}].$$

Summation über $j = 1, 2, \ldots, N$ liefert:

$$\boxed{S(h) = \sum_{j=1}^{N} S_j(h) = \frac{h}{3}\left[f_0 + 4f_1 + f_{2N} + 2\sum_{k=1}^{N-1}(f_{2k} + 2f_{2k+1})\right]; \quad f_k := f(x_k).} \tag{5.19}$$

Der Quadraturfehler ist gemäß (5.16) bestimmt durch

$$\boxed{R_2[f] = \int_a^b f(x)\,dx - S(h) = -\frac{(b-a)}{180}\,h^4 f^{(4)}(\xi), \quad a < \xi < b.}$$

BSP. (8.5.7) **Verallgemeinerte (oder summierte)** MILNE–**Regel:** Man setzt

$$h := \frac{b-a}{4N}; \quad x_j := a + jh\ \forall\, j = 0, 1, \ldots, 4N; \quad I_j := [x_{4(j-1)}, x_{4j}]\ \forall\, j = 1, 2, \ldots, N.$$

Auf jedem Teilintervall I_j wenden wir die MILNE–Regel an:

$$M_j(h) := Q_{4j} := \frac{4h}{90}\left[7f_{4(j-1)} + 32f_{4j-3} + 12f_{4j-2} + 32f_{4j-1} + 7f_{4j}\right].$$

Summation über $j = 1, 2, \ldots, N$ liefert:

$$\boxed{\begin{aligned} M(h) \quad &= \sum_{j=1}^{N} M_j(h) = \frac{2h}{45}\Big[7(f_0 + f_{4N}) + 32(f_1 + f_3) + 12f_2 \\ &\quad + \sum_{k=1}^{N-1}(14f_{4k} + 32f_{4k+1} + 32f_{4k+3} + 12f_{4k+2})\Big]; \quad f_k := f(x_k). \end{aligned}} \tag{5.20}$$

Der Quadraturfehler beträgt hier

$$\boxed{R_4[f] = \int_a^b f(x)\,dx - M(h) = -\frac{2(b-a)}{945}\,h^6 f^{(6)}(\xi), \quad a < \xi < b.}$$

BSP. (8.5.8) Zur genäherten Berechnung von

$$I = \int_1^2 \frac{dx}{x} = \ln 2 \doteq 6.931\,471\,805\,599\,45\mathrm{E}^{-01}$$

verwenden wir die summierten Quadraturformeln $S(h)$ und $M(h)$ aus (5.19) bzw. (5.20) für $N = 2, 3, \ldots, 10$. Die numerischen Werte und die Quadraturfehler sind in der folgenden Tabelle aufgelistet:

Berechnung von $I := \int_a^b f(x)\,dx$ mit summierter SIMPSON– **und** MILNE–**Regel:**

N	$S(h)$	$R_2[f]$	$M(h)$	$R_4[f]$
2	$6.932\,539\,682\mathrm{E}^{-01}$	$-1.067\,876\,940\mathrm{E}^{-04}$	$6.931\,479\,014\mathrm{E}^{-01}$	$-7.209\,214\,250\mathrm{E}^{-07}$
3	$6.931\,697\,931\mathrm{E}^{-01}$	$-2.261\,260\,990\mathrm{E}^{-05}$	$6.931\,472\,534\mathrm{E}^{-01}$	$-7.291\,843\,178\mathrm{E}^{-08}$
4	$6.931\,545\,306\mathrm{E}^{-01}$	$-7.350\,094\,716\mathrm{E}^{-06}$	$6.931\,471\,942\mathrm{E}^{-01}$	$-1.373\,715\,510\mathrm{E}^{-08}$
5	$6.931\,502\,306\mathrm{E}^{-01}$	$-3.050\,129\,140\mathrm{E}^{-06}$	$6.931\,471\,842\mathrm{E}^{-01}$	$-3.703\,731\,832\mathrm{E}^{-09}$
6	$6.931\,486\,622\mathrm{E}^{-01}$	$-1.481\,649\,148\mathrm{E}^{-06}$	$6.931\,471\,818\mathrm{E}^{-01}$	$-1.260\,095\,404\mathrm{E}^{-09}$
7	$6.931\,479\,838\mathrm{E}^{-01}$	$-8.033\,151\,789\mathrm{E}^{-07}$	$6.931\,471\,810\mathrm{E}^{-01}$	$-5.045\,637\,196\mathrm{E}^{-10}$
8	$6.931\,476\,528\mathrm{E}^{-01}$	$-4.722\,595\,027\mathrm{E}^{-07}$	$6.931\,471\,807\mathrm{E}^{-01}$	$-2.279\,730\,271\mathrm{E}^{-10}$
9	$6.931\,474\,759\mathrm{E}^{-01}$	$-2.954\,210\,961\mathrm{E}^{-07}$	$6.931\,471\,806\mathrm{E}^{-01}$	$-1.129\,003\,328\mathrm{E}^{-10}$
10	$6.931\,473\,746\mathrm{E}^{-01}$	$-1.941\,053\,198\mathrm{E}^{-07}$	$6.931\,471\,806\mathrm{E}^{-01}$	$-6.026\,776\,213\mathrm{E}^{-11}$

Die obigen Überlegungen zeigen, dass der Quadraturfehler $R_n[f]$ der NEWTON–CÔTES–Formeln $Q_n(h)$ bei einer Schrittweite $h > 0$ der Asymptotik $|R_n[f]| = O(h^p)$ für $h \to 0+$ unterworfen ist. Die Zahl p heißt *Konvergenzordnung* der Quadraturformel $Q_n(h)$. Hierbei ist die Güte von $Q_n(h)$ sicher entscheidend von der Größe der Potenz p abhängig. Zum Beispiel hat man $p_T = 2$, $p_S = 4$ und $p_M = 6$ für die summierte TRAPEZ–, SIMPSON– bzw. MILNE–Regel. Ist von einer Quadraturformel $Q_n(h)$ die Konvergenzordnung unbekannt, so kann sie näherungsweise wie folgt berechnet werden. Man wendet $Q_n(h)$ auf das Integral

$$I := \int_a^b f(x)\,dx$$

an, dessen exakter Wert I bekannt sei. Es gilt dann

$$I - Q_n(h) = h^p \cdot C_f.$$

Falls $f(x)$ durch $Q_n(h)$ nicht exakt integriert wird, so ist die von h unabhängige Konstante C_f nicht Null. In diesem Falle erhält man durch Halbierung der Schrittweite h:

$$\frac{I - Q_n(h)}{I - Q_n(h/2)} \approx 2^p.$$

Aus dieser Relation folgt näherungsweise die Konvergenzordnung

$$\boxed{p \approx \frac{1}{\ln 2}\,\ln\left(\left|\frac{I - Q_n(h)}{I - Q_n(h/2)}\right|\right).}$$

Es ist zweckmäßig, sich mehrere Werte für p durch wiederholte Halbierung von h zu verschaffen.

BSP. (8.5.9) Das Integral

$$I := \int_1^4 \frac{dx}{x} = \ln 4 \doteq 1.386\,294\,361\,119\,89$$

wird mit Hilfe der summierten Quadraturformeln $S(h)$ und $M(h)$ für $N = 1, 2, 4, 8, 16, 32, 64$ numerisch berechnet. In der folgenden Tabelle sind die nach der obigen Vorschrift bestimmten Konvergenzordnungen p_S und p_M aufgelistet. Man erkennt, dass die theoretischen Werte $p_S = 4$ und $p_M = 6$ sehr genau erreicht werden.

Berechnung von $I := \int_a^b f(x)\,dx$ mit summierter SIMPSON– und MILNE–Regel:

N	$S(h)$	$R_2[f]$	p_S	$M(h)$	$R_4[f]$	p_M
1	$1.425\,000\,000\mathrm{E}^{+00}$	$-3.870\,563\mathrm{E}^{-02}$	$\cdots$	$1.389\,395\,604\mathrm{E}^{+00}$	$-3.101\,243\mathrm{E}^{-03}$	$\cdots$
2	$1.391\,620\,879\mathrm{E}^{+00}$	$-5.326\,518\mathrm{E}^{-03}$	2.86	$1.386\,483\,705\mathrm{E}^{+00}$	$-1.893\,445\mathrm{E}^{-04}$	4.03
4	$1.386\,804\,779\mathrm{E}^{+00}$	$-5.104\,179\mathrm{E}^{-04}$	3.38	$1.386\,300\,890\mathrm{E}^{+00}$	$-6.529\,620\mathrm{E}^{-06}$	4.86
8	$1.386\,332\,383\mathrm{E}^{+00}$	$-3.802\,263\mathrm{E}^{-05}$	3.75	$1.386\,294\,506\mathrm{E}^{+00}$	$-1.456\,451\mathrm{E}^{-07}$	5.49
16	$1.386\,296\,874\mathrm{E}^{+00}$	$-2.512\,957\mathrm{E}^{-06}$	3.92	$1.386\,294\,363\mathrm{E}^{+00}$	$-2.571\,181\mathrm{E}^{-09}$	5.82
32	$1.386\,294\,520\mathrm{E}^{+00}$	$-1.594\,703\mathrm{E}^{-07}$	3.98	$1.386\,294\,361\mathrm{E}^{+00}$	$-4.166\,142\mathrm{E}^{-11}$	5.95
64	$1.386\,294\,371\mathrm{E}^{+00}$	$-1.000\,595\mathrm{E}^{-08}$	3.99	$1.386\,294\,361\mathrm{E}^{+00}$	$-7.066\,919\mathrm{E}^{-13}$	5.88

Offene NEWTON–CÔTES–Formeln.

Das Intervall $[a, b]$ wird wiederum äquidistant zerlegt:

$$h := \frac{b-a}{n}; \quad x_j := a + jh,\ j = 0, 1, \ldots, n.$$

Im Gegensatz zu den geschlossenen NEWTON–CÔTES–Formeln wird die zu integrierende Funktion $f(x)$ nur in den **inneren** Intervallpunkten $x_1, x_2, \ldots, x_{n-1}$ interpoliert und danach die Interpolierende (also das LAGRANGE–Interpolationspolynom) integriert. Offenbar gibt es nur im Fall $n \geq 2$ innere Punkte. Wir behandeln hier lediglich die beiden Spezialfälle $n = 2$ und $n = 3$.

BSP. (8.5.10) Die **Mittelpunkt–Regel** oder **Tangententrapez–Regel** erhält man für $n = 2$. Die Interpolierende durch den inneren Punkt $x_1 = \frac{a+b}{2}$ ist die Gerade $P_0(x) := f(x_1)$. Somit resultiert

$$\boxed{Q_0^o := (b-a) f(x_1); \quad x_1 = \frac{a+b}{2}.} \tag{5.21}$$

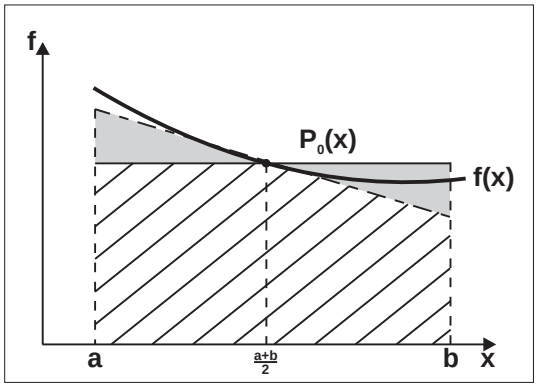
Offenbar kann der Rechteckinhalt Q_0^o auch als Fläche des Trapezes interpretiert werden, welches durch die Tangente an $f(x)$ im Punkte $(x_1, f(x_1))$ gebildet wird (siehe Skizze). Dies rechtfertigt den Namen Tangententrapezregel. Der Genauigkeitsgrad ist offenkundig $m = 1$. Der Quadraturfehler beträgt

$$\boxed{R_0^o[f] := \frac{1}{24} (b-a)^3 f''(\xi), \quad a < \xi < b;} \tag{5.22}$$

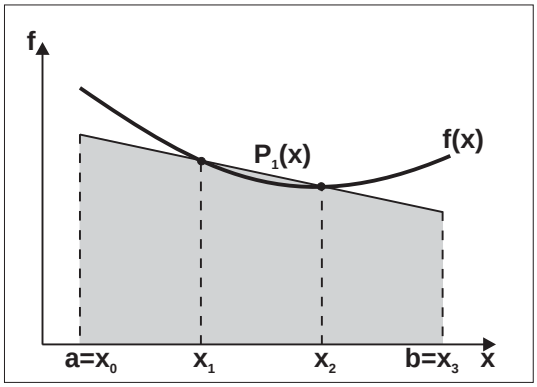
er ist gegenüber dem Quadraturfehler der Trapezregel etwa halb so groß und trägt entgegengesetztes Vorzeichen. Es gelten folglich die Einschließungen:

$$f(x) \text{ konvex } (\Rightarrow f''(x) \geq 0) \quad \Rightarrow \quad Q_0^o \leq \int_a^b f(x)\, dx \leq Q_1,$$

$$f(x) \text{ konkav } (\Rightarrow f''(x) \leq 0) \quad \Rightarrow \quad Q_1 \leq \int_a^b f(x)\, dx \leq Q_0^o.$$



Die Mittelpunkt–Regel



Offene NC–Regel für $n = 3$

BSP. (8.5.11) **Offene NEWTON–CÔTES–Quadraturformel für $n = 3$.** Die Interpolierende $P_1(x)$ ist die Gerade durch die Punkte $(x_1, f(x_1))$, $(x_2, f(x_2))$, siehe obige Skizze. Man erhält mit einfacher Rechnung:

$$\boxed{Q_1^o = \frac{3h}{2} [f_1 + f_2]; \quad h := \frac{b-a}{3}; \quad f_1 := f(a+h), \; f_2 := f(b-h).} \tag{5.23}$$

Der Genauigkeitsgrad von Q_1^o beträgt wiederum $m = 1$, und der Quadraturfehler ist gegeben durch

$$\boxed{R_1^o[f] = \frac{1}{108} (b-a)^3 f''(\xi), \quad a < \xi < b;} \tag{5.24}$$

er ist damit um den Faktor 9 kleiner als bei der Trapezregel, die denselben Aufwand erfordert wie die Auswertung von (5.23).

8.6 Das Lebesgue–Integral

Zur Motivation der folgenden Überlegungen greifen wir nochmals das BSP. (8.3.8) der DIRICHLET–Funktion über einem Intervall $[a,b]$ auf:

$$f(x) := \begin{cases} 1 \ : \ x \in \mathbf{Q} \cap [a,b], \\ 0 \ : \ x \in [a,b], \text{ irrational.} \end{cases} \tag{6.1}$$

Wie wir in Abschnitt 8.3 gezeigt haben, ist $f(x)$ **nicht** R–integrierbar. Ist andererseits die Menge der rationalen Zahlen im Intervall $[a,b]$ in irgendeiner Numerierung durch $\{r_1, r_2, \ldots\}$ gegeben, so definieren wir eine Funktionenfolge

$$f_n(x) := \begin{cases} 1 \ : \ x \in \{r_1, r_2, \ldots, r_n\}, \\ 0 \ : \ \text{sonst.} \end{cases} \tag{6.2}$$

Ganz offenbar gelten die folgenden Eigenschaften: (i) $\lim\limits_{n\to\infty} f_n(x) = f(x)$ punktweise auf $[a,b]$, und (ii) $f_n(x) \le f_{n+1}(x) \ \forall\, x \in [a,b] \ \forall\, n \in \mathbf{N}$.

Definition 8.16 *Eine Folge von Funktionen $f_n \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$* **konvergiert monoton** *(fast überall) auf $[a,b]$ gegen eine Grenzfunktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$, wenn gilt:*

(i) $\lim\limits_{n\to\infty} f_n(x) = f(x)$ *punktweise (fast überall) auf $[a,b]$,*

(ii) $f_n(x) \le f_{n+1}(x)$ *(oder $f_n(x) \ge f_{n+1}(x)$)* $\forall\, x \in [a,b] \ \forall\, n \in \mathbf{N}$.

Die Treppenfunktionen (6.2) sind gemäß Satz 8.15 R–integrierbar mit $\int_a^b f_n(x)\,dx = 0$ für alle $n \in \mathbf{N}$. Es liegt nun auf der Hand, der Grenzfunktion f durch die Festsetzung

$$\int_a^b f(x)\,dx := \lim_{n\to\infty} \int_a^b f_n(x)\,dx \quad (=0)$$

einen Integralwert zuzuordnen. Dabei steht auf der linken Seite **kein** R–Integral, denn dieses existiert ja nicht.

Der hier aufgezeigte Zugang zu einem neuen Integralbegriff soll nun systematisch aufgebaut werden. Ausgangsbasis ist ein (nicht notwendig beschränktes) Intervall $I \subset \mathbf{R}$ mit Randpunkten $a < b$, wobei $a = -\infty$ und $b = +\infty$ zugelassen sind. Ist $I := [a,b]$ ein **endliches** Intervall, so können **endliche Zerlegungen** Z_n von I erklärt werden, und zu jeder Zerlegung Z_n Treppenfunktionen $t_n : I \to \mathbf{R}$, vgl. Abschnitt 8.3. Falls I **unbeschränkt** ist, so definieren wir eine Verallgemeinerung der Treppenfunktionen:

Definition 8.17 *Eine Funktion $\varphi \in \mathrm{Abb}\,(I,\mathbf{R})$ heiße eine* **Treppenfunktion** *auf I, wenn ein* **endliches** *Teilintervall $I_\varphi \subseteq I$ existiert und eine Treppenfunktion $t_n : I_\varphi \to \mathbf{R}$ bezüglich einer endlichen Zerlegung Z_n von I_φ mit*

$$\varphi(x) := \begin{cases} t_n(x) \ : \ x \in I_\varphi, \\ 0 \qquad : \ x \in I \setminus I_\varphi. \end{cases}$$

Es bezeichne $T(I)$ den Vektorraum aller so definierten Treppenfunktionen auf I.

Eine Darstellung von $t_n(x)$ findet man in (3.2). Da $\varphi \in T(I)$ außerhalb eines gewissen Intervalls $I_\varphi := [a_0, b_0]$ verschwindet, folgt aus der Forderung 1, Abschnitt 8.3, stets die Existenz des R–Integrals

$$\int_a^b \varphi(x)\,dx = \int_{a_0}^{b_0} t_n(x)\,dx = \sum_{j=1}^{n} y_j |I_j|.$$

Als Grundmenge für einen neuen Integralbegriff betrachten wir die Menge $L^+(I) = L^+(a,b)$, die wir wie folgt definieren:

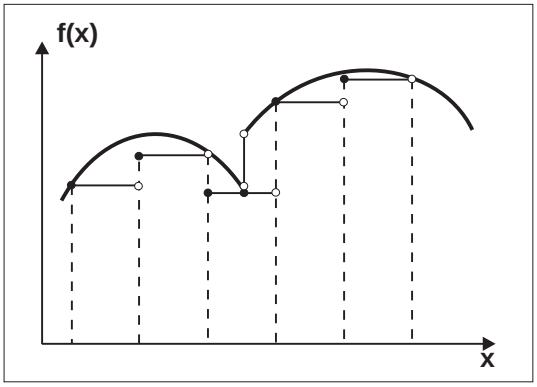
Definition 8.18 *Genau dann gelte $f \in L^+(I)$, wenn es eine Folge von Treppenfunktionen $\varphi_n \in T(I)$ gibt mit den Eigenschaften*

(a) $\left(\varphi_n\right)$ *konvergiert auf I monoton wachsend fast überall gegen f,*

(b) *die Integralfolge $\left(\int_a^b \varphi_n(x)\,dx\right)$ ist (nach oben) beschränkt, also auch konvergent.*

Bemerkung 8.18 Es ist klar, dass die DIRICHLET–Funktion (6.1) in der Menge $L^+(a,b)$ liegt. Wir zeigen weiter: Jede Funktion $f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{R})$, die auf dem endlichen Intervall $[a,b]$ R–integrierbar ist, erfüllt $f \in L^+(a,b)$. Dazu definieren wir eine Folge endlicher Zerlegungen $Z_n := \{I_{n0}, I_{n1}, \ldots, I_{np}\}$, $p := 2^n - 1$, von $[a,b]$ gemäß

$$x_{nk} := a + k\,\frac{b-a}{2^n}, \quad k = 0,1,\ldots,p+1, \qquad I_{nk} := [x_{nk}, x_{n,k+1}), \quad k = 0,1,\ldots,p. \tag{6.3}$$



Approximation durch Treppenfunktionen

Für jedes feste $n \in \mathbf{N}$ sei die Treppenfunktion $\varphi_n : [a,b] \to \mathbf{R}$ bezüglich der oben eingeführten Zerlegung Z_n in der folgenden Weise erklärt:

$$\varphi_n(x) := \begin{cases} m_{nk} & : \; x \in I_{nk}, \\ f(b) & : \; x = b, \end{cases} \qquad m_{nk} := \inf_{x \in I_{nk}} f(x), \quad k = 0,1,\ldots,p.$$

Nach Konstruktion gilt $\varphi_n(x) \le f(x)$ für alle $x \in [a,b]$ und alle $n \in \mathbf{N}$, und die Folge $\left(\varphi_n\right)$ ist monoton wachsend. Ist $\xi \in (a,b)$ ein **Stetigkeitspunkt** von f, so ist es intuitiv einleuchtend, dass $\lim\limits_{n\to\infty} \varphi_n(\xi) = f(\xi)$ gelten muss. Gemäß Satz 8.17 bilden die Unstetigkeitspunkte von f eine Nullmenge, und daran ändert sich nichts, wenn die oben unberücksichtigt gebliebenen Endpunkte a, b hinzugefügt werden. Insgesamt erfüllt somit die Folge $\left(\varphi_n\right)$ die Forderung (a) der Definition 8.18. Die Forderung (b) ergibt sich offenkundig aus den Eigenschaften $\varphi_n(x) \le f(x)$ und aus der Monotonie der Folge $\left(\varphi_n\right)$:

$$\int_a^b \varphi_n(x)\,dx \le \int_a^b f(x)\,dx < +\infty.$$

Die monoton wachsende und nach oben beschränkte Integralfolge $\left(\int_a^b \varphi_n(x)\,dx\right)$ ist also konvergent. Wir zeigen

$$\lim_{n\to\infty} \int_a^b \varphi_n(x)\,dx = \int_a^b f(x)\,dx. \tag{6.4}$$

In der Tat, es sei $K := \sup_{x\in I} |f(x)|$ gesetzt. Dann gibt es zu jedem festen $\epsilon > 0$ offene Teilintervalle $I_1, I_2, \ldots$ von I der Gesamtlänge $\leq \epsilon/2K$, die die Nullmenge der Unstetigkeitspunkte von f überdecken. Dabei darf getrost angenommen werden, dass die Endpunkte von I_k zur Menge der Punkte x_{nk} aus (6.3) gehören. Setzt man $M := \bigcup_k I_k,\ |M| \leq \epsilon/2K$, so gilt

$$\int_a^b \varphi_n(x)\,dx = \sum_{I_{nk}\in Z_n} m_{nk}|I_{nk}| = \sum_{I_{nk}\cap M=\emptyset} m_{nk}|I_{nk}| + \sum_{I_{nk}\cap M\neq\emptyset} m_{nk}|I_{nk}|.$$

Da die Funktion f auf jedem abgeschlossenen Teilintervall $\overline{I_{nk}}$ mit $I_{nk}\cap M = \emptyset$ stetig ist, existiert ein Punkt $\xi_{nk} \in \overline{I_{nk}}$ mit $m_{nk} = \inf_{x\in I_{nk}} f(x) = f(\xi_{nk})$. Darüber hinaus folgt aus der Konstruktion:

$$\limsup_{n\to\infty} \Big| \sum_{I_{nk}\cap M\neq\emptyset} m_{nk}\,|I_{nk}| \Big| \leq K \sum_k |I_k| \leq \frac{\epsilon}{2},$$

und für die RIEMANN–Summe $\sum_{I_{nk}\cap M=\emptyset} m_{nk}|I_{nk}|$ gilt:

$$\lim_{n\to\infty} \sum_{I_{nk}\cap M=\emptyset} m_{nk}|I_{nk}| = \lim_{n\to\infty} \sum_{I_{nk}\cap M=\emptyset} f(\xi_{nk})|I_{nk}| = \int_{[a,b]\setminus M} f(x)\,dx.$$

Schließlich haben wir

$$\limsup_{n\to\infty} \left| \int_a^b \varphi_n(x)\,dx - \int_a^b f(x)\,dx \right| \leq \limsup_{n\to\infty} \Big| \sum_{I_{nk}\cap M\neq\emptyset} m_{nk}|I_{nk}| \Big| + \int_M |f(x)|\,dx \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Da $\epsilon > 0$ beliebig wählbar war, resultiert daraus die behauptete Konvergenz (6.4). □

Die obige Bemerkung motiviert uns nun zu folgender Definition eines Integralbegriffs für Funktionen $f \in L^+(I) = L^+(a,b)$:

Definition 8.19 *Es sei $f \in L^+(I)$ gegeben, und es sei $\left(\varphi_n\right) \subset T(I)$ die der Funktion f definitionsgemäß zugeordnete Folge von Treppenfunktionen. Dann heiße die Zahl*

$$\boxed{\int_a^b f(x)\,dx := \lim_{n\to\infty} \int_a^b \varphi_n(x)\,dx} \tag{6.5}$$

das LEBESGUE*–Integral (kurz: L–Integral) von f über I.*

Diese Definition bekommt erst dann einen Sinn, wenn die Unabhängigkeit des Ausdrucks $\int_a^b f(x)\,dx$ von der speziellen Wahl der Folge $\left(\varphi_n\right)$ gezeigt ist. Dazu verwenden wir den folgenden Satz, den wir hier ohne Beweis zitieren, man vgl. etwa H. HEUSER, Lehrbuch der Analysis, Teil 2. B.G. Teubner Verlag, Stuttgart 1981, dort S. 85:

Satz 8.28 *Gegeben seien $f, g \in L^+(I)$ mit zugeordneten Folgen $\left(\varphi_n\right), \left(\psi_n\right) \subset T(I)$, und es gelte $f \leq g$ fast überall auf I. Dann folgt*

$$\lim_{n\to\infty} \int_a^b \varphi_n(x)\,dx \leq \lim_{n\to\infty} \int_a^b \psi_n(x)\,dx.$$

Gilt nun $f = g$ fast überall auf I, so ergibt sich sowohl $\lim\limits_{n\to\infty} \int_a^b \varphi_n(x)\,dx \le \lim\limits_{n\to\infty} \int_a^b \psi_n(x)\,dx$ als auch $\lim\limits_{n\to\infty} \int_a^b \psi_n(x)\,dx \le \lim\limits_{n\to\infty} \int_a^b \varphi_n(x)\,dx$; die beiden Grenzwerte stimmen also überein.

Ganz offenkundig ist $L^+(I)$ **kein** Vektorraum und somit das bisher erklärte LEBESGUE–Integral kein lineares Funktional. Denn aus $f, g \in L^+(I)$ kann i.a. **nicht** auf $f - g \in L^+(I)$ geschlossen werden. Hingegen gelten sicher $f + g \in L^+(I)$ sowie $\lambda f \in L^+(I)$ für jede Zahl $\lambda \ge 0$. Daraus erhält man unter Verwendung der Linearität des R–Integrals in Verbindung mit (6.5):

$$\int_a^b \Big(f(x) + g(x)\Big)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx, \quad \int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx \;\forall\; \lambda \ge 0. \tag{6.6}$$

Definition 8.20 *Eine Funktion $f \in \mathrm{Abb}(\mathbf{R}, \mathbf{R})$ heiße auf I* LEBESGUE**–integrierbar** *(kurz: L–integrierbar) genau dann, wenn es $g, h \in L^+(I)$ gibt mit $f = g - h$. Das* LEBESGUE*–Integral von f ist erklärt durch*

$$\boxed{\int_a^b f(x)\,dx := \int_a^b g(x)\,dx - \int_a^b h(x)\,dx.} \tag{6.7}$$

Es bezeichne $L(I)$ die Menge aller so definierten, auf I L–integrierbaren Funktionen.

Bemerkung 8.19 (a) Die Definition (6.7) hängt nicht von der speziellen Wahl der Funktionen $g, h \in L^+(I)$ ab. Gilt nämlich auch noch $f = g_1 - h_1$ mit $g_1, h_1 \in L^+(I)$, so ist $g - h = g_1 - h_1$ also $g + h_1 = g_1 + h$. Somit folgt aus (6.6)

$$\begin{aligned} \int_a^b g(x)\,dx + \int_a^b h_1(x)\,dx &= \int_a^b g_1(x)\,dx + \int_a^b h(x)\,dx, \text{ also } \int_a^b g(x)\,dx - \int_a^b h(x)\,dx \\ &= \int_a^b g_1(x)\,dx - \int_a^b h_1(x)\,dx. \end{aligned}$$

(b) Es gilt $L^+(I) \subset L(I)$. Klar, wegen $0 \in L^+(I)$ haben wir $f = f - 0 \in L^+(I)$ für jedes $f \in L^+(I)$ und somit auch $f \in L(I)$. Ebenso folgt für endliche Intervalle $I := [a, b]$, dass $L(I)$ auch die Menge aller auf I R–integrierbaren Funktionen enthält. □

Die konkrete Berechnung des L–Integrals einer Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ gemäß der Definition (6.7) ist offenbar sehr unpraktisch. Genau wie beim R–Integral wird man deshalb Rechenregeln für das L–Integral herleiten, mit deren Hilfe die Berechnung des L–Integrals von f auf möglicherweise bekannte Integrale zurückgeführt werden kann.

Wir beginnen zunächst mit der wichtigsten Eigenschaft des L–Integrals, nämlich seiner Invarianz gegenüber Änderungen des Integranden auf einer Nullmenge.

Satz 8.29 *Für $f_1, f_2 \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ gelte $f_1 \in L(I)$ sowie $f_1 = f_2$ fast überall auf I. Dann folgt $f_2 \in L(I)$ und*

$$\boxed{\int_a^b f_1(x)\,dx = \int_a^b f_2(x)\,dx.}$$

Begründung: Zu $f_1 \in L(I)$ existieren definitionsgemäß $g_1, h_1 \in L^+(I)$ mit $f_1 = g_1 - h_1$. Das L–Integral von f_1 hat die Darstellung (6.7) mit g, h ersetzt durch g_1, h_1. Setzt man $h_2 := h_1 + (f_1 - f_2)$, so gilt voraussetzungsgemäß $h_2 = h_1$ fast überall auf I. Ist der Funktion h_1 die Folge $\big(\varphi_n\big) \subset T(I)$ von Treppenfunktionen zugeordnet, so folgt aus Definition 8.18, dass dieselbe Folge auch der Funktion

h_2 zugeordnet ist. Somit gilt $h_2 \in L^+(I)$, und wegen $f_2 = g_1 - h_2$ dann auch $f_2 \in L(I)$. Ferner habe wir

$$\int_a^b f_2(x)\,dx = \int_a^b g_1(x)\,dx - \int_a^b h_2(x)\,dx = \int_a^b g_1(x)\,dx - \int_a^b h_1(x)\,dx = \int_a^b f_1(x)\,dx,$$

und dies war noch zu zeigen. □

Die folgenden Eigenschaften erhält man unmittelbar aus der Definition des L–Integrals; ihre Begründung liegt nahezu auf der Hand.

Satz 8.30 (a) *$L(I)$ ist ein Vektorraum über dem Körper $\mathbf{R}$, und das L–Integral ist linear: Für $f, g \in L(I)$ und $\lambda \in \mathbf{R}$ gelten:*

$$\boxed{\int_a^b \Big(f(x) + g(x)\Big)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx, \qquad \int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx.}$$

(b) *Das L–Integral ist ordnungserhaltend: Aus $f, g \in L(I)$ und $f \geq g$ fast überall auf I folgt*

$$\boxed{\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx, \quad \textit{speziell für } g = 0\textit{:} \quad \int_a^b f(x)\,dx \geq 0.}$$

(c) *Es sei $a < c < b$. Genau dann ist $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ auf (a, b) L–integrierbar, wenn dies auf (a, c) und (c, b) zutrifft. In diesem Fall gilt*

$$\boxed{\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.}$$

Wie bei R–Integralen setzt man der Vollständigkeit halber noch

$$\boxed{\int_a^a f(x)\,dx := 0, \quad \int_b^a f(x)\,dx := -\int_a^b f(x)\,dx, \text{ falls } a < b.}$$

Im Gegensatz zur RIEMANNschen Integrationstheorie tritt in der Theorie des L–Integrals der Begriff eines **uneigentlichen Integrals** nicht auf, da die Konstruktion des L–Integrals von vorneherein auch unbeschränkte Integrationsintervalle zulässt. Unterschiede zwischen R– und L–Integral können daher im Bereich der uneigentlichen Integration auftreten. Wir hatten bereits in Abschnitt 8.4 vermerkt, dass das uneigentliche R–Integral $\int_0^\infty \frac{\sin x}{x}\,dx$ zum Integralwert $\pi/2$ *einfach* konvergiert, während *absolute* Konvergenz nicht vorliegt. Ein solches unterschiedliches Konvergenzverhalten von uneigentlichen R–Integralen kann bei L–Integralen nicht auftreten; wir haben hier:

Satz 8.31 *Für $f \in L(I)$ gilt stets auch $|f| \in L(I)$, und es folgt:*

$$\boxed{\Big|\int_a^b f(x)\,dx\Big| \leq \int_a^b |f(x)|\,dx.}$$

Begründung: Wir zeigen zuerst $|f| \in L(I)$. Dazu seien $f_1, f_2 \in L^+(I)$ mit zugeordneten Folgen $\big(\varphi_n\big), \big(\psi_n\big) \subset T(I)$ und $f = f_1 - f_2$ vorgelegt. Wegen $|f| = \max\{f_1, f_2\} - \min\{f_1, f_2\}$ müssen wir $\max\{f_1, f_2\}, \min\{f_1, f_2\} \in L^+(I)$ beweisen. Nun konvergieren aber fast überall auf I $\max\{\varphi_n, \psi_n\}$ und $\min\{\varphi_n, \psi_n\}$ monoton wachsend gegen $\max\{f_1, f_2\}$ bzw. $\min\{f_1, f_2\}$. Darüber hinaus darf $\varphi_n \geq 0, \psi_n \geq 0$ angenommen werden. Andernfalls ersetze man $f_1, f_2, \varphi_n, \psi_n$ durch $f_1 - h, f_2 - h, \varphi_n - h$ bzw. $\psi_n - h$, wobei $h := \min\{\varphi_1, \psi_1\}$ gesetzt sei. Wir erschließen hieraus noch

$$\begin{aligned} \int_a^b \min\{\varphi_n(x), \psi_n(x)\}\, dx &\leq \int_a^b \max\{\varphi_n(x), \psi_n(x)\}\, dx \leq \int_a^b \Big(\varphi_n(x) + \psi_n(x)\Big)\, dx \\ &\leq \int_a^b f_1(x)\, dx + \int_a^b f_2(x)\, dx. \end{aligned}$$

Nun gelten also $\max\{f_1, f_2\}, \min\{f_1, f_2\} \in L^+(I)$, und somit $|f| \in L(I)$. Wegen $f \leq |f|$ und $-f \leq |f|$ erhalten wir mit Satz 8.30(b) schließlich auch

$$\int_a^b f(x)\, dx \leq \int_a^b |f(x)|\, dx, -\int_a^b f(x)\, dx \leq \int_a^b |f(x)|\, dx, \text{ also } \Big|\int_a^b f(x)\, dx\Big| \leq \int_a^b |f(x)|\, dx.$$

Als eine Folgerung aus diesem Satz stellen wir fest, dass die Funktion $\frac{\sin x}{x}$ auf $[0, +\infty)$ nicht L–integrierbar ist.

Eines der wichtigsten Ergebnisse aus der Theorie der L–Integrale sind Aussagen über die Vertauschbarkeit von Limes–Prozessen und Integration. Dabei geht es um die Beziehung zwischen der punktweisen Konvergenz fast überall einer Folge $\big(f_n\big) \subset L(I)$ gegen eine Grenzfunktion f und der Konvergenz der Integralfolge $\big(\int_a^b f_n(x)\, dx\big)$ gegen das L–Integral $\int_a^b f(x)\, dx$, und zwar ohne die Voraussetzung $f \in L(I)$. Wir werden eine entsprechende Beziehung in mehreren Schritten herleiten. Wir beginnen zunächst mit einem Resultat, welches mehr oder minder Hilfscharakter hat.

Satz 8.32 *Gegeben sei eine monoton wachsende Folge von Treppenfunktionen $\big(\varphi_n\big) \subset T(I)$ mit beschränkter Integralfolge $\big(\int_a^b \varphi_n(x)\, dx\big)$. Dann existiert $\lim\limits_{n\to\infty} \varphi_n =: f$ fast überall auf I, und es gilt $f \in L^+(I)$.*

Begründung: (a) Wir zeigen die Existenz eines Grenzwertes $\lim\limits_{n\to\infty} \varphi_n = f$ fast überall auf I. Ohne Einschränkung sei $\varphi_n \geq 0$ vorausgesetzt. Andernfalls ersetze man φ_n durch $\varphi_n - \varphi_1$. Ferner darf angenommen werden, dass jedes φ_n in seinen Unstetigkeitsstellen links– oder rechtsseitig stetig ist: $\varphi_n(x) = \varphi_n(x+0)$ oder $= \varphi_n(x-0)$ für alle $x \in I$. Es sei $K \geq 0$ mit $\int_a^b \varphi_n(x)\, dx \leq K \ \forall\, n \in \mathbf{N}$ gewählt. Zu beliebigem $\epsilon > 0$ setze man $M_{\epsilon n} := \{x \in I : \varphi_n(x) \geq K/\epsilon\}$. Da φ_n ja eine Treppenfunktion ist, besteht $M_{\epsilon n}$ im Falle $M_{\epsilon n} \neq \emptyset$ aus einer endlichen Anzahl disjunkter Intervalle $I_1, I_2, \ldots, I_m$, und es gilt

$$\frac{K}{\epsilon} \sum_{k=1}^m |I_k| \leq \sum_{k=1}^m \int_{I_k} \varphi_n(x)\, dx \leq \int_a^b \varphi_n(x)\, dx \leq K.$$

Hieraus erhält man die Gesamtlänge $|M_{\epsilon n}| = \sum\limits_{k=1}^m |I_k| \leq \epsilon$. Wegen $\varphi_n \leq \varphi_{n+1}$ ergibt sich $M_{\epsilon n} \subset M_{\epsilon,n+1}$, so dass die Differenzmenge $M_{\epsilon,n+1} \setminus M_{\epsilon n}$ (sofern sie nichtleer ist) als Vereinigung endlich vieler disjunkter Intervalle darstellbar ist. Demgemäß ist die Menge $M_\epsilon := \bigcup\limits_{n=1}^\infty M_{\epsilon n}$ als Vereinigung von höchstens abzählbar vielen disjunkten Intervallen $J_1, J_2, \ldots$ darstellbar: Man nimmt zuerst alle Intervalle der Menge $M_{\epsilon 1}$, danach alle Intervalle der Menge $M_{\epsilon 2} \setminus M_{\epsilon 1}$, danach alle Intervalle von $M_{\epsilon 3} \setminus M_{\epsilon 2}$, usf. Endlich viele Intervalle $J_1, J_2, \ldots, J_n$ sind somit in einer gewissen Menge $M_{\epsilon m}$ enthalten,

und demzufolge gilt $\sum_{k=1}^{n} |J_k| \le |M_{\epsilon m}| \le \epsilon$, und zwar für jedes $n \in \mathbf{N}$. Also ist das Intervallsystem $J_1, J_2, \ldots$ längenbeschränkt durch ϵ. Nun ist die Menge N aller Punkte $x \in I$, für die die Zahlenfolge $\Big(\varphi_n(x)\Big)$ divergiert, sicher in M_ϵ enthalten, also eine Nullmenge (wegen der Monotonie muss $\varphi_n(x)$ im Divergenzfall gegen $+\infty$ streben).

(b) Setzen wir schließlich

$$f(x) := \begin{cases} \lim\limits_{n\to\infty} \varphi_n(x) & : \ x \in I \setminus N, \\ 0 & : \ x \in N, \end{cases}$$

so erfüllt f die Bedingungen (a), (b) aus Definition 8.18. Somit gilt $f \in L^+(I)$. □

Mit diesem Hilfsresultat ist es nun möglich, eine erste Aussage über die Vertauschbarkeit von Limes–Bildung und Integration zu begründen.

Satz 8.33 *Gegeben sei eine monoton wachsende Folge von Funktionen $\Big(f_n\Big) \subset L^+(I)$ mit beschränkter Integralfolge $\Big(\int_a^b f_n(x)\,dx\Big)$. Dann existiert $\lim\limits_{n\to\infty} f_n =: f \in L^+(I)$ fast überall auf I, und es gilt:*

$$\boxed{\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.} \tag{6.8}$$

Begründung: Es existiert ein $K \ge 0$ mit $\int_a^b f_n(x)\,dx \le K$ für alle $n \in \mathbf{N}$. Für festes $n \in \mathbf{N}$ sei $\Big(\varphi_{nk}\Big) \subset T(I)$ die der Funktion $f_n \in L^+(I)$ zugeordnete Folge von Treppenfunktionen. Die Folge $\varphi_n := \max\{\varphi_{jk} \ : \ j,k = 1,2,\ldots,n\}$ besteht dann ebenfalls aus Treppenfunktionen, und sie ist offenbar monoton wachsend. Wir haben fast überall auf I

$$\varphi_{jk} \le f_j \le f_n \text{ für } j \le n, \quad \text{und somit } \varphi_n \le f_n, \quad \int_a^b \varphi_n(x)\,dx \le \int_a^b f_n(x)\,dx \le K. \tag{6.9}$$

Satz 8.32 trifft auf die Folge $\Big(\varphi_n\Big)$ zu: Fast überall auf I existiert $\lim\limits_{n\to\infty} \varphi_n =: f \in L^+(I)$, und definitionsgemäß gilt

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b \varphi_n(x)\,dx. \tag{6.10}$$

Um die Konvergenz $\lim\limits_{n\to\infty} f_n = f$ fast überall auf I zu zeigen, beachten wir, dass für $j \le n$ nach Konstruktion ja $\varphi_{jn} \le \varphi_n$ gilt. Im Limes $n \to \infty$ folgt daraus $f_j \le f$ auf $I \setminus N_j$, N_j eine Nullmenge. Setzen wir $N := \bigcup\limits_{j\in\mathbf{N}} N_j$, so ist N wieder eine Nullmenge, und es gilt wegen (6.9)

$$\varphi_n \le f_n \le f \quad \text{für alle } n \in \mathbf{N} \text{ auf } I \setminus N.$$

Daraus folgt die behauptete Konvergenz. Integriert man diese Ungleichung über I, so folgt unter Beachtung von (6.10) schließlich auch noch die Relation (6.8). □

Unser nächstes Ziel besteht in der Übertragung der Aussagen von Satz 8.33 auf den Funktionenraum $L(I)$. Dazu benötigen wir ein weiteres Hilfsergebnis:

Satz 8.34 *Sei $f \in L(I)$. Dann existieren zu jedem $\epsilon > 0$ Funktionen $g, h \in L^+(I)$ mit*

$$f = g - h, \quad h \ge 0, \quad \int_a^b h(x)\,dx \le \epsilon.$$

Begründung: Definitionsgemäß existiert eine Darstellung $f = g_1 - h_1$ mit $g_1, h_1 \in L^+(I)$. Sei $\left(\varphi_n\right) \subset T(I)$ die der Funktion h_1 zugeordnete Folge von Treppenfunktionen, und sei $m \in \mathbf{N}$ so fixiert, dass gilt:

$$0 \le \int_a^b h_1(x)\,dx - \int_a^b \varphi_m(x)\,dx \le \epsilon.$$

Setzen wir $h_2 := h_1 - \varphi_m$, so gilt offenbar $h_2 \in L^+(I)$, $h_2 \ge 0$ fast überall auf I und $\int_a^b h_2(x)\,dx \le \epsilon$. Der nichtnegative Anteil $h := h_2^+$ unterscheidet sich also nur auf einer Nullmenge von h_2. Deshalb gilt gemäß Satz 8.29 auch $h \in L^+(I)$ sowie $\int_a^b h(x)\,dx \le \epsilon$. Des weiteren erfüllt $g := (g_1 - \varphi_m) + (h - h_2)$ die Relation $g - h = (g_1 - \varphi_m) - (h_1 - \varphi_m) = f$. Weil $h - h_2 = 0$ fast überall auf I gilt, folgt aus $g_1 - \varphi_m \in L^+(I)$ schließlich auch noch $g \in L^+(I)$. □

Wir sind jetzt in der Lage, eines der zentralen Resultate der LEBESGUEschen Integrationstheorie zu beweisen, nämlich den folgenden

Satz 8.35 (Konvergenzsatz von BEPPO LEVI)
Gegeben sei eine monotone Folge von Funktionen $(f_n) \subset L(I)$ mit beschränkter Integralfolge $\left(\int_a^b f_n(x)\,dx\right)$. Dann existiert $\lim\limits_{n\to\infty} f_n =: f \in L(I)$ fast überall auf I, und es gilt:

$$\boxed{\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.}$$

Begründung: Wir nehmen ohne Beschränkung der Allgemeinheit an, dass die Folge $\left(f_n\right)$ wächst (andernfalls betrachten wir $\left(-f_n\right)$) und dass $f_n \ge 0$ für alle $n \in \mathbf{N}$ gilt (andernfalls ersetzen wir f_n durch $f_n - f_1$). Unter Hinzunahme von $f_0 := 0$ gilt für alle $n \in \mathbf{N}$ die Darstellung

$$f_n = \sum_{k=1}^{n} (f_k - f_{k-1}) \quad \text{mit } f_k - f_{k-1} \ge 0.$$

Gemäß Satz 8.34 existieren für jedes k Funktionen $g_k, h_k \in L^+(I)$ mit

$$f_k - f_{k-1} = g_k - h_k, \quad h_k \ge 0, \quad \int_a^b h_k(x)\,dx \le \frac{1}{2^k},$$

und es resultiert $g_k = f_k - f_{k-1} + h_k \ge 0$. Die Folgen $s_n := \sum\limits_{k=1}^{n} g_k$ und $t_n := \sum\limits_{k=1}^{n} h_k$ wachsen monoton; es gilt nach Konstruktion

$$f_n = s_n - t_n \quad \text{sowie} \quad s_n, t_n \in L^+(I). \tag{6.11}$$

Wegen

$$\int_a^b t_n(x)\,dx = \sum_{k=1}^{n} \int_a^b h_k(x)\,dx \le \sum_{k=1}^{n} \frac{1}{2^k} < 1$$

ist die Integralfolge $\left(\int_a^b t_n(x)\,dx\right)$ beschränkt, und dies trifft wegen

$$\int_a^b s_n(x)\,dx = \int_a^b f_n(x)\,dx + \int_a^b t_n(x)\,dx$$

auch auf die Integralfolge $\left(\int_a^b s_n(x)\,dx\right)$ zu. Gemäß Satz 8.33 existieren Grenzwerte $\lim\limits_{n\to\infty} s_n =: s \in L^+(I)$ und $\lim\limits_{n\to\infty} t_n =: t \in L^+(I)$ fast überall auf I mit

$$\lim_{n\to\infty} \int_a^b s_n(x)\,dx = \int_a^b s(x)\,dx \quad \text{und} \quad \lim_{n\to\infty} \int_a^b t_n(x)\,dx = \int_a^b t(x)\,dx.$$

Nun ist $f := s - t \in L(I)$, und wegen (6.11) folgt

$$\lim_{n\to\infty} f_n = f \text{ fast überall auf } I, \quad \lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

Wir haben alles bewiesen. □

Der Konvergenzsatz von Beppo Levi gibt Anlass zu einigen wichtigen Folgerungen, die wir in dem folgenden Satz zusammenfassen.

Satz 8.36 (a) *Konvergiert die Reihe*

$$\sum_{k=1}^{\infty} \int_a^b |f_k(x)|\,dx \quad \textit{mit} \quad f_k \in L(I),$$

so konvergiert die Reihe $\sum\limits_{k=1}^{\infty} f_k$ *fast überall auf* I *gegen eine Grenzfunktion* $f \in L(I)$*, und man darf Integration und Summation vertauschen:*

$$\boxed{\sum_{k=1}^{\infty} \int_a^b f_k(x)\,dx = \int_a^b \sum_{k=1}^{\infty} f_k(x)\,dx = \int_a^b f(x)\,dx.} \tag{6.12}$$

(b) *Genau dann verschwindet* $f \in L(I)$ *fast überall auf* I*, wenn* $\int_a^b |f(x)|\,dx = 0$ *gilt.*

(c) *Sind* $I_1, I_2, \ldots$ *Teilintervalle von* I *mit aufsteigender Folge* $I_1 \subset I_2 \subset \cdots$ *und* $I = \bigcup\limits_{n=1}^{\infty} I_n$*, und ist* $f \in \operatorname{Abb}(I, \mathbf{R})$ *auf jedem* I_n *L–integrierbar mit beschränkter Integralfolge* $\left(\int_{I_n} |f(x)|\,dx\right)$*, so gilt* $f \in L(I)$ *sowie*

$$\boxed{\int_I f(x)\,dx = \lim_{n\to\infty} \int_{I_n} f(x)\,dx.} \tag{6.13}$$

Begründungen: (a) Es bezeichne f_k^+ bzw. f_k^- den positiven bzw. negativen Anteil von f_k. Setzt man

$$s_n := \sum_{k=1}^{n} f_k^+ \quad \text{und} \quad t_n := \sum_{k=1}^{n} f_k^-,$$

so sind $\left(s_n\right), \left(t_n\right)$ monoton wachsende Folgen aus $L(I)$. Wegen

$$\int_a^b f_k^+(x)\,dx \le \int_a^b |f_k(x)|\,dx, \qquad \int_a^b f_k^-(x)\,dx \le \int_a^b |f_k(x)|\,dx$$

sind deren Integralfolgen voraussetzungsgemäß beschränkt, so dass jeweils der Konvergenzsatz von B. Levi zutrifft. Da $s_n - t_n = \sum\limits_{k=1}^{n} f_k$ gilt, erhält man aus Satz 8.35 die oben behaupteten Aussagen.

(b) Gilt $f = 0$ fast überall auf I, so trifft dies auch auf die Funktion $|f|$ zu. Deshalb gilt gemäß Satz 8.29 $\int_a^b |f(x)|\,dx = 0$. Ist umgekehrt $\int_a^b |f(x)|\,dx = 0$, so folgt trivialerweise $\sum\limits_{k=1}^{\infty} \int_a^b |f(x)|\,dx = 0$. Gemäß (a) konvergiert dann auch $\sum\limits_{k=1}^{\infty} f(x) = f(x) \lim\limits_{n\to\infty} n$ fast überall auf I, was nur möglich ist, wenn f fast überall auf I verschwindet.

(c) Zunächst sei $f \geq 0$ angenommen, und es sei

$$f_n(x) := \begin{cases} f(x) & : \ x \in I_n, \\ 0 & : \ x \in I \setminus I_n \end{cases}$$

gesetzt. Dann ist die Folge $\big(f_n\big) \subset L(I)$ monoton wachsend mit beschränkter Integralfolge

$$\Big(\int_I f_n(x)\,dx \Big) = \Big(\int_{I_n} f(x)\,dx \Big),$$

und es gilt $\lim\limits_{n\to\infty} f_n = f$. Der Konvergenzsatz von B. Levi garantiert nun $f \in L(I)$ sowie die behauptete Relation (6.13). Gilt aber nicht $f \geq 0$, so benutze man die Darstellung $f = f^+ - f^-$ und wende das soeben Bewiesene auf f^+ und f^- an. □

Eine der wesentlichen Voraussetzungen zum Konvergenzsatz von B. Levi ist die Monotonie der Folge $\big(f_n\big)$ und die Beschränktheit der Integralfolge $\big(\int_a^b f_n(x)\,dx \big)$. Im nächsten Satz – der sicher zu den wichtigsten Resultaten der Lebesgueschen Integrationstheorie zählt – lösen wir uns von diesen Voraussetzungen.

Satz 8.37 (Lebesguescher Satz von der dominierten Konvergenz)
Gegeben sei eine Folge von Funktionen $(f_n) \subset L(I)$, und es existiere $\lim\limits_{n\to\infty} f_n = f$ fast überall auf I. Gibt es darüber hinaus eine integrable Majorante $g \in L(I)$ mit $|f_n| \leq g$ für alle n, so gilt $f \in L(I)$ sowie

$$\boxed{\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.} \tag{6.14}$$

Begründung: Es bezeichne N die Nullmenge aller $x \in I$, für die $\big(f_n(x)\big)$ nicht gegen $f(x)$ konvergiert. Wir setzen

$$g_n(x) := \begin{cases} \inf\{f_n(x), f_{n+1}(x), \ldots\} & : \ x \in I \setminus N, \\ 0 & : \ x \in N, \end{cases}$$

$$h_n(x) := \begin{cases} \sup\{f_n(x), f_{n+1}(x), \ldots\} & : \ x \in I \setminus N, \\ 0 & : \ x \in N. \end{cases}$$

Nach Konstruktion wächst die Folge $\big(g_n(x)\big)$ für jedes $x \in I \setminus N$ monoton; und wegen ihrer Beschränktheit nach oben (nämlich durch $g(x)$) muss sie konvergieren; ihr Grenzwert kann keine andere Zahl als $f(x)$ sein. Ebenso konvergiert die monoton fallende Folge $\big(h_n(x)\big)$ für jedes $x \in I \setminus N$ gegen $f(x)$. Ganz analog wie in Satz 8.31 zeigt man, dass die Funktionen $u_{nk} := \min\{f_n, f_{n+1}, \ldots, f_{n+k}\}$ in $L(I)$ liegen. Halten wir n fest, so ist die Folge $\big(u_{nk}\big)$ monoton fallend mit $\lim\limits_{k\to\infty} u_{nk} = g_n$ fast überall auf I. Aus $|f_n| \leq g$ folgern wir überdies $-g \leq u_{nk} \leq g$, und somit durch Integration auch

$$-\int_a^b g(x)\,dx \leq \int_a^b u_{nk}(x)\,dx \leq \int_a^b g(x)\,dx.$$

Das heißt, die Integralfolge $\big(\int_a^b u_{nk}(x)\,dx \big)$ ist bei festem n beschränkt, so dass der Konvergenzsatz von B. Levi auf $g_n \in L(I)$ führt. Ganz entsprechend erhält man $h_n \in L(I)$. Die fast überall konvergenten Folgen $\big(g_n\big), \big(h_n\big)$ sind überdies integralbeschänkt:

$$\int_a^b g_1(x)\,dx \leq \int_a^b g_n(x)\,dx \leq \int_a^b h_n(x)\,dx \leq \int_a^b h_1(x)\,dx.$$

Wiederum aus dem Konvergenzsatz von B. LEVI erhalten wir für die Grenzfunktion $f \in L(I)$ die Ungleichungskette

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b g_n(x)\,dx \le \lim_{n\to\infty} \int_a^b f_n(x)\,dx \le \lim_{n\to\infty} \int_a^b h_n(x)\,dx = \int_a^b f(x)\,dx,$$

also die behauptete Relation (6.14). □

Bemerkung 8.20 Wie in der RIEMANNschen Integrationstheorie erklärt man das L–Integral **komplexwertiger** Funktionen $f : I \to \mathbf{C}$ auch durch Trennung von Real– und Imaginärteil. Ist $f = u + i\,v$ mit reellen Funktionen $u, v \in L(I)$, so gelte definitionsgemäß $f \in L(I)$ sowie

$$\boxed{\int_a^b f(x)\,dx := \int_a^b u(x)\,dx + i \int_a^b v(x)\,dx.}$$

Mit dieser Zuordnung können die meisten Aussagen aus der LEBESGUEschen Integrationstheorie (sofern keine Monotonie–Eigenschaften von f involviert sind) direkt auf komplexwertige Funktionen f übertragen werden. Ganz speziell gilt der LEBESGUEsche Satz von der dominierten Konvergenz ohne Einschränkung für $f \in \mathrm{Abb}\,(I, \mathbf{C})$. □

BSP. (8.6.1) In der Theorie der Funktionen einer komplexen Veränderlichen spielen häufig Parameter–Integrale vom Typ

$$J(R) := \frac{1}{2\pi} \int_0^\pi e^{i\,z(t;R)} f\big(z(t;R)\big)\,dt, \quad z(t;R) := Re^{it} = R\,(\cos t + i \sin t), \;\; R > 0,$$

eine Rolle. Wir nehmen $f\big(z(\cdot;R)\big) \in L(0,\pi)$ für jedes $R > 0$ sowie $|f\big(z(t;R)\big)| \le K$ für alle $t \in (0,\pi)$ und alle $R > 0$ an. Zu berechnen ist der Grenzwert $\lim\limits_{R\to\infty} J(R)$. Dazu verwenden wir Satz 8.37. Auf dem Intervall $I := (0,\pi)$ gilt $\sin t > 0 \;\forall\, t \in I$, und die Konstante K ist wegen

$$\Big|e^{i\,z(t;R)} f\big(z(t;R)\big)\Big| = e^{-R\sin t} |f\big(z(t;R)\big| \le K$$

eine integrable Majorante. Darüber hinaus strebt

$$\lim_{R\to\infty} \Big|e^{i\,z(t;R)} f\big(z(t;R)\big)\Big| = 0, \quad \text{punktweise für alle } t \in (0,\pi).$$

Der LEBESGUEsche Satz von der dominierten Konvergenz liefert somit

$$\lim_{R\to\infty} J(R) = \frac{1}{2\pi} \int_0^\pi \lim_{R\to\infty} e^{i\,z(t;R)} f\big(z(t;R)\big)\,dt = \frac{1}{2\pi} \int_0^\pi 0\,dt = 0.$$

Kapitel 9

Funktionenfolgen und Funktionenreihen

9.1 Potenzreihen

Wie in Abschnitt 7.6 ausführlich erörtert wurde, kann eine Funktion $f \in \mathrm{Abb}\,(\mathbf{R}, \mathbf{R})$ unter bestimmten Voraussetzungen in Punkten $x_0 \in D(f)$ in eine TAYLOR–Reihe entwickelt werden:

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!}\, f^{(k)}(x_0)\, (x - x_0)^k.$$

Diese Reihe ist ein Spezialfall von allgemeineren Reihen in der Form

$$\boxed{P(x) := \sum_{k=0}^{\infty} a_k\, (x - x_0)^k, \quad a_k \in \mathbf{K} \text{ gegeben.}} \tag{1.1}$$

Hier bezeichne $\mathbf{K}$ wieder den Körper der reellen ($\mathbf{K} := \mathbf{R}$) bzw. der komplexen ($\mathbf{K} := \mathbf{C}$) Zahlen.

Definition 9.1 *Eine Reihe in der Form* (1.1) *heiße eine* **Potenzreihe** *mit Entwicklungspunkt oder Mittelpunkt x_0 und Koeffizienten a_k. Insbesondere hat eine Potenzreihe mit Entwicklungspunkt $x_0 = 0$ die Form*

$$\boxed{P(x) := \sum_{k=0}^{\infty} a_k\, x^k = a_0 + a_1 x + a_2 x^2 + \cdots} \tag{1.2}$$

Bemerkung 9.1 Die Substitution $\xi := x - x_0$ führt Potenzreihen der Form (1.1) stets in Potenzreihen der Form (1.2) über. Es genügt daher, sich ausschließlich mit Potenzreihen vom Typ (1.2) zu befassen. □

Setzen wir in (1.2) neue Koeffizienten $A_k := a_k x^k$ an, so entscheidet das **Wurzelkriterium** (Satz 3.15) darüber, ob die Reihe $\sum\limits_{k=0}^{\infty} A_k$ absolut konvergent ist oder divergent, und zwar je nach Größe des Grenzwertes

$$q := \limsup_{k\to\infty} \sqrt[k]{|A_k|} = \limsup_{k\to\infty} \sqrt[k]{|a_k||x|^k} = |x| \limsup_{k\to\infty} \sqrt[k]{|a_k|}.$$

Wir führen die von x unabhängige Größe

$$\boxed{\rho := \frac{1}{\limsup\limits_{k\to\infty} \sqrt[k]{|a_k|}}} \tag{1.3}$$

ein. Dann folgt aus der Konvergenzaussage des Satzes 3.15 und aus der Bedingung (2.8) in Abschnitt 3.2:

$q<1$	$\Leftrightarrow$	$\lvert x\rvert<\rho$	:	Reihe (1.2) ist **absolut konvergent**
$q>1$	$\Leftrightarrow$	$\lvert x\rvert>\rho$	:	Reihe (1.2) ist **divergent**.

Der Fall $q = 1$ bleibt mit dem Wurzelkriterium unentscheidbar. Offensichtlich ist es völlig ohne Belang, ob die Variable x reell oder komplex ist. Die Zahl ρ legt im Reellen wie im Komplexen in gleicher Weise den Konvergenz– und Divergenzbereich der Reihe (1.2) fest.

Definition 9.2 *Die der Potenzreihe* (1.2) *durch die Vorschrift* (1.3) *zugeordnete Größe* $\rho \geq 0$ *heiße der* **Konvergenzradius** *der Reihe* (1.2). *Dabei seien die Fälle* $\frac{1}{0} := +\infty$ *und* $\frac{1}{\infty} := 0$ *mit einbezogen. Im Fall* $\rho = +\infty$ *heiße die Potenzreihe* (1.2) **beständig konvergent**.

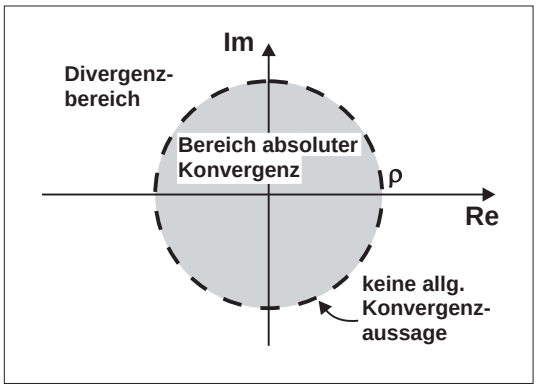
Die folgenden Aussagen ergeben sich unmittelbar aus der Vorbetrachtung.

Satz 9.1 *Es sei* ρ *der Konvergenzradius der Potenzreihe* (1.2). *Dann konvergiert die Potenzreihe* (1.2) *in jedem Punkt* x *innerhalb des* **Konvergenzkreises** $K_\rho(0) := \{x \in \mathbf{C} : |x| < \rho\}$, *und sie divergiert für* $|x| > \rho$, *also außerhalb* $K_\rho(0)$. *Im Fall* $\rho = 0$ *konvergiert die Reihe nur im Punkt* $x = 0$ *zum Summenwert* a_0. *Auf der Kreislinie* $|x| = \rho$ *ist keine allgemeine Konvergenzaussage möglich. Existieren die Grenzwerte*

$$\boxed{\rho_1 := \frac{1}{\lim\limits_{k\to\infty} \sqrt[k]{|a_k|}} \qquad \textit{und/oder} \qquad \rho_2 := \lim_{k\to\infty} \Big|\frac{a_k}{a_{k+1}}\Big|,}$$

so gilt $\rho_1 = \rho = \rho_2$.

Dieser Satz reflektiert lediglich die Konvergenz– und Divergenzaussagen der beiden Sätze 3.15 und 3.17 (Wurzel– bzw. Quotientenkriterium).



Der Konvergenzkreis einer Potenzreihe

BSP. (9.1.1) Wir betrachten die Potenzreihe $P(x) := \sum\limits_{k=1}^{\infty} \frac{x^k}{k^\alpha}$. Es gilt hier

$$\rho := \lim_{k\to\infty} k^{\alpha/k} = \lim_{k\to\infty} e^{(\alpha \ln k)/k} = e^0 = 1 \ \ \forall\, \alpha \in \mathbf{R}.$$

Deshalb resultiert für jedes $\alpha \in \mathbf{R}$:

$$\boxed{P(x) := \sum_{k=1}^{\infty} \frac{x^k}{k^\alpha} \begin{cases} |x| < 1 \ : \ \textbf{absolut konvergent}, \\ |x| > 1 \ : \ \textbf{divergent}. \end{cases}}$$

Für $x \in \mathbf{R}$ lassen sich in den Randpunkten $x = \pm 1$ auch noch Konvergenzaussagen treffen:

$$\begin{aligned} P(1) &:= \sum_{k=1}^{\infty} \frac{1}{k^\alpha} && \text{konvergiert } \forall\ \alpha > 1 \text{ (Integralvergleichskriterium)}, \\ P(-1) &:= \sum_{k=1}^{\infty} \frac{(-1)^k}{k^\alpha} && \text{konvergiert } \forall\ \alpha > 0 \text{ (LEIBNIZ–Kriterium)}. \end{aligned}$$

BSP. (9.1.2) Wir betrachten die Potenzreihe $P(x) := \sum\limits_{k=0}^{\infty} a^{k^2} x^k$ für $a \in \mathbf{R}$. Es gilt hier

$$\rho := \lim_{k\to\infty} |a|^{-k} = \lim_{k\to\infty} e^{-k \ln |a|} = \begin{cases} 0 & : \ |a| > 1, \\ 1 & : \ |a| = 1, \\ +\infty & : \ |a| < 1. \end{cases}$$

Dementsprechend ist der Konvergenzkreis $K_\rho(0)$ entweder leer ($|a| > 1$) oder der Einheitskreis ($|a| = 1$) oder die ganze komplexe Ebene ($|a| < 1$).

BSP. (9.1.3) Wir betrachten die Potenzreihe $P(x) := \sum\limits_{k=0}^{\infty} a_k x^k$ mit $a_k := \cosh k$. Es gilt hier

$$\rho := \lim_{k\to\infty} \left| \frac{a_k}{a_{k+1}} \right| = \lim_{k\to\infty} \frac{e^k + e^{-k}}{e^{k+1} + e^{-(k+1)}} = \lim_{k\to\infty} \frac{1 + e^{-2k}}{e + e^{-(2k+1)}} = \frac{1}{e}.$$

Bemerkung 9.2 (a) Über das Verhalten der Potenzreihe auf dem Rand $|x| = \rho$ des Konvergenzkreises $K_\rho(0)$ wird in Satz 9.1 keine Aussage getroffen. Eine Analyse des Konvergenzverhaltens auf diesem Rand ist Sache der **Funktionentheorie**. Die *reellen* Randpunkte $x = \pm\rho$ sollte man – sofern dies möglich ist – einer gesonderten Betrachtung unter Verwendung der in Abschnitt 3.2 diskutierten Konvergenzkriterien unterziehen, siehe auch Satz 9.10.

(b) Die Funktionen $f_k(x) := a_k x^k,\ k \in \mathbf{N}_0$, sind besonders *gutartig* hinsichtlich der Differenzierbarkeitseigenschaft $f_k \in C^\infty(\mathbf{R})$. Wir werden im nächsten Abschnitt die Frage diskutieren, wie sich diese Eigenschaften von f_k auf die Reihe $\sum\limits_{k=0}^{\infty} f_k(x)$ übertragen. □

9.2 Gleichmäßige Konvergenz

Wir betrachten in diesem Abschnitt **Funktionenfolgen** $\Big(f_k(x)\Big)_{k\in\mathbf{N}_0}$ unter der Voraussetzung, dass die Funktionenfamilie $f_k : D(f_k) \to \mathbf{K}$ einen nichtleeren gemeinsamen Definitionsbereich hat:

$$\emptyset \neq D := \bigcap_{k\in\mathbf{N}_0} D(f_k) \subset \mathbf{R}.$$

Definition 9.3 (a) *Die Menge* $K := \{x \in D : \lim\limits_{k\to\infty} f_k(x)$ *existiert*$\}$ *heiße der* **Konvergenzbereich** *der Funktionenfolge* $\Big(f_k(x)\Big)_{k\in\mathbf{N}_0}$.

(b) *Wir definieren die Folge der Partialsummen*

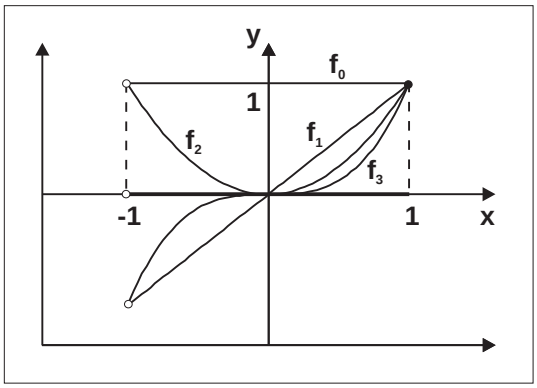
$$\boxed{s_n(x) := \sum_{k=0}^{n} f_k(x), \quad n \in \mathbf{N},\ x \in D.}$$

Die Menge $K := \{x \in D : \lim_{n\to\infty} s_n(x)$ *existiert*$\}$ *heiße der* **Konvergenzbereich** *der Funktionenreihe* $\sum_{k=0}^{\infty} f_k(x)$.

BSP. (9.2.1) Es sei $f_k(x) := x^k \ \forall\ x \in D := \mathbf{R},\ k \in \mathbf{N}_0$. Wir haben offenbar

$$\lim_{k\to\infty} f_k(x) = \begin{cases} 1 & : \ x = 1, \\ 0 & : \ |x| < 1, \\ \text{divergent} & : \ x \notin (-1, +1]. \end{cases}$$

Der Konvergenzbereich der Funktionenfolge $\Big(f_k(x)\Big)_{k\in\mathbf{N}_0}$ ist das Intervall $K := (-1, +1]$.



Der Konvergenzbereich der Folge $(x^k)_{k\geq 0}$

BSP. (9.2.2) Es sei $f_k(x) := (\sin kx)/k^2 \ \forall\ x \in D := \mathbf{R},\ k \in \mathbf{N}$. Wegen

$$\Big|\sum_{k=1}^{\infty} \frac{\sin kx}{k^2}\Big| \leq \sum_{k=1}^{\infty} \frac{1}{k^2} < +\infty$$

konvergiert die Funktionenreihe $\sum_{k=1}^{\infty} f_k(x)$ für alle $x \in \mathbf{R}$ sogar absolut. Der Konvergenzbereich der Funktionenreihe ist $K := \mathbf{R}$.

Auf dem Konvergenzbereich K wird durch die Zuordnungen

$$F(x) := \lim_{k\to\infty} f_k(x) \quad \text{bzw.} \quad F(x) := \sum_{k=0}^{\infty} f_k(x), \quad x \in K,$$

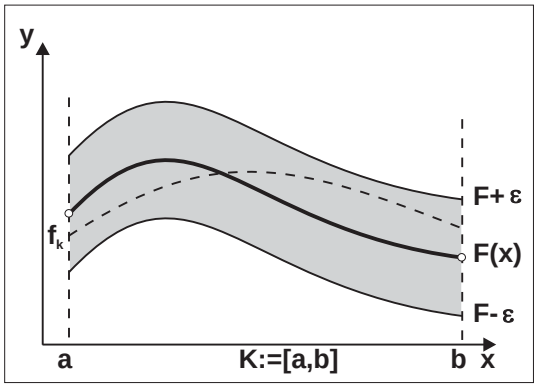
eine Funktion $F : K \to \mathbf{K}$ erklärt. Wir fragen nach den Stetigkeits– und Differenzierbarkeitseigenschaften, die von den Funktionen f_k auf die Grenzfunktion F vererbt werden. Das obige BSP. (9.2.1) zeigt schon die Problematik auf. Obwohl jede Funktion $f_k(x) = x^k$ zur Klasse $C^\infty(\mathbf{R})$ gehört, ist die Grenzfunktion F unstetig. Ähnliche Beispiele lassen sich auch für Funktionenreihen angeben. Wir wollen nun den Konvergenzbegriff so abändern, dass eine konvergente Folge (oder Reihe) stetiger Funktionen auch eine stetige Grenzfunktion besitzt.

Definition 9.4 *Die Funktionenfolge* $\left(f_k(x)\right)_{k\in\mathbf{N}_0}$ *heiße auf der Menge* $K \subset \mathbf{R}$ **gleichmäßig konvergent** *gegen die Grenzfunktion* $F(x)$*, wenn gilt:*

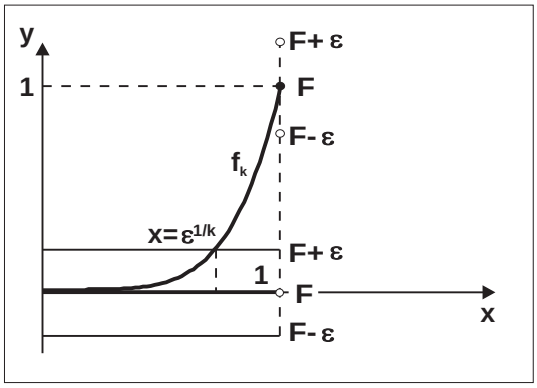
$$\boxed{\forall\, \epsilon > 0\ \exists\, N = N(\epsilon)\ :\ \sup_{x\in K} |F(x) - f_k(x)| \le \epsilon\ \ \forall\, k \ge N.} \tag{2.1}$$

Die Funktionenreihe $\sum\limits_{k=0}^{\infty} f_k(x)$ *heiße auf der Menge* K **gleichmäßig konvergent***gegen die Grenzfunktion* $F(x)$*, wenn die Folge der Partialsummen* $s_n(x) := \sum\limits_{k=0}^{n} f_k(x)$ *dies tut.*

Bemerkung 9.3 Im Unterschied zur gewöhnlichen oder **punktweisen** Konvergenz $f_k(x) \to F(x)$ hängt die Zahl $N(\epsilon)$ in der Bedingung (2.1) *nicht* von der Stelle $x \in K$ ab. $N(\epsilon)$ kann eben **gleichmäßig** bezüglich $x \in K$ gewählt werden. Geometrisch bedeutet diese Bedingung, dass alle Funktionsgraphen $G(f_k)$ ab dem Index $k = N(\epsilon)$ in einem ϵ–Schlauch um den Funktionsgraphen $G(F)$ verlaufen. □



Der ϵ–Schlauch der gleichmäßigen Konvergenz



Nicht gleichmäßige Konvergenz der Folge $f_k(x) := x^k$

BSP. (9.2.3) Wir behaupten, dass die Folge $f_k(x) := x^{1+1/k}$ auf dem Intervall $K := [0,1]$ gleichmäßig gegen die Grenzfunktion $F(x) := x$ konvergiert. Denn mit Hilfe der Differentialrechnung bestimmt man das Maximum der Funktion $g(x) := x - x^{1+1/k}$ an der Stelle $\bar{x} := \left(\frac{k}{k+1}\right)^k \in K$. Es folgt hieraus:

$$\sup_{x\in K} |f_k(x) - F(x)| = g(\bar{x}) = \left(\frac{k}{k+1}\right)^k \left(1 - \frac{k}{k+1}\right) \le \left(1 - \frac{k}{k+1}\right) = \frac{1}{k+1} \le \epsilon\ \ \forall\, k \ge N(\epsilon) := \frac{1}{\epsilon}.$$

BSP. (9.2.4) Wir hatten in BSP. (9.2.1) gezeigt, dass die Funktionenfolge $f_k(x) := x^k$ auf dem Teilintervall $K := [0,1]$ *punktweise* gegen die Grenzfunktion

$$F(x) := \begin{cases} 0 \ :\ x \in [0,1), \\ 1 \ :\ x = 1 \end{cases}$$

konvergiert. Die Konvergenz ist jedoch *nicht gleichmäßig*. Die obige Skizze zeigt, dass jede Funktion $f_k(x)$ den ϵ–Schlauch um die Grenzfunktion $F(x)$ an der Stelle $x := \epsilon^{1/k} \in K$ verlässt, sofern $0 < \epsilon < 1$ gilt.

Für Funktionenreihen existiert ein einfaches Kriterium, mit dessen Hilfe die gleichmäßige Konvergenz nachgeprüft werden kann:

Satz 9.2 (WEIERSTRASS–Kriterium)
Die Funktionenreihe $\sum\limits_{k=0}^{\infty} f_k(x)$ *konvergiert* **gleichmäßig** *auf der Menge* K *gegen die Grenzfunktion* $F(x)$*, wenn es eine Zahlenfolge* $(a_k)_{k\in\mathbf{N}_0}$ *gibt mit den Eigenschaften*

$$\boxed{\text{(i)}\quad |f_k(x)| \le a_k \;\forall\, x\in K \;\forall\, k\in\mathbf{N}_0, \qquad \text{(ii)}\quad \sum_{k=0}^{\infty} a_k \text{ konvergiert.}}$$

Begründung: Wegen (ii) existiert zu jedem $\epsilon > 0$ eine Zahl $N(\epsilon)$ mit $\sum\limits_{k=N+1}^{\infty} a_k \le \epsilon$. Hieraus folgt für alle $n \ge N(\epsilon)$:

$$\sup_{x\in K} |F(x) - s_n(x)| = \sup_{x\in K} \Big| \sum_{k=n+1}^{\infty} f_k(x)\Big| \le \sum_{k=n+1}^{\infty} \sup_{x\in K} |f_k(x)| \le \sum_{k=N+1}^{\infty} a_k \le \epsilon.$$

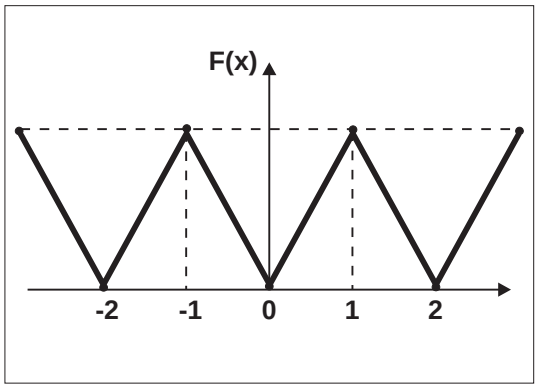
Dies ist aber gerade die Bedingung der gleichmäßigen Konvergenz. □

BSP. (9.2.5) Wir setzen $f_k(x) := \Big(\cos(2k+1)\pi x\Big)\Big/(2k+1)^2,\ x\in\mathbf{R},\ k\ge 0$. Dann gilt $|f_k(x)| \le (2k+1)^{-2} =: a_k\ \forall\, x\in\mathbf{R}$, und die Reihe $\sum\limits_{k=0}^{\infty} a_k$ ist konvergent. Also sind die Bedingungen (i) und (ii) des WEIERSTRASS–Kriteriums erfüllt, welches die gleichmäßige Konvergenz der Funktionenreihe

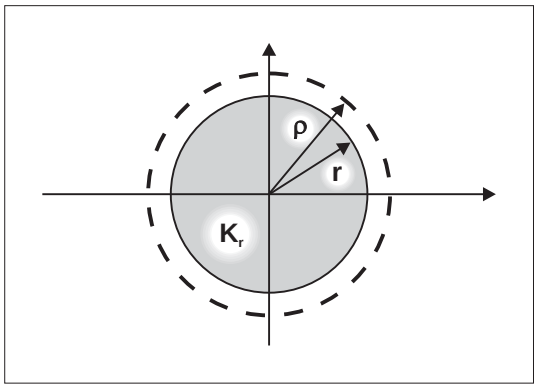
$$F(x) := \frac{1}{2} - \frac{4}{\pi^2}\sum_{k=0}^{\infty} \frac{\cos(2k+1)\pi x}{(2k+1)^2},\quad x\in\mathbf{R},$$

garantiert. Mit Hilfe der Theorie der FOURIER–*Reihen* (Stoff der Ing.–Math.III) wird gezeigt, dass die Grenzfunktion $F(x)$ die stetige, periodische Funktion

$$\boxed{F(x) = |x| \;\forall\, x\in[-1,+1], \quad F(x+2) = F(x)\;\forall\, x\in\mathbf{R},}$$



Die Funktion $F(x) := |x|,\ x\in[-1,+1]$**, mit periodischer Fortsetzung** $F(x+2) = F(x)\ \forall\, x\in\mathbf{R}$**.**



Zur gleichmäßigen Konvergenz von Potenzreihen

ist, siehe obige Skizze. Mit dieser Kenntnis kann $F(x)$ insbesondere an der Stelle $x = 1$ ausgewertet werden. Es gilt $F(1) = 1$, und mit $\cos(2k+1)\pi = -1$ resultiert

$$\boxed{\sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{\pi^2}{8}.}$$

Potenzreihen sind spezielle Funktionenreihen. Mit Hilfe des WEIERSTRASS–Kriteriums zeigt man die folgende Aussage über die gleichmäßige Konvergenz.

Satz 9.3 *Eine Potenzreihe* $P(x) := \sum_{k=0}^{\infty} a_k x^k$ *konvergiert* **gleichmäßig** *auf jeder abgeschlossenen Kreisscheibe* $\overline{K}_r(0) := \{x \in \mathbf{C} : |x| \leq r\}$ *vom Radius* $r < \rho$*, wobei* ρ *den Konvergenzradius der Potenzreihe* $P(x)$ *angibt.*

Begründung: Wir haben in Satz 9.1 gezeigt, dass die Potenzreihe $P(x)$ für $|x| = r < \rho$ absolut konvergiert. Also ist die Reihe $\sum_{k=0}^{\infty} |a_k| r^k$ konvergent, und wegen $|a_k x^k| \leq |a_k| r^k \; \forall \, x \in \overline{K}_r(0), \; k \in \mathbf{N}_0$, sind die Voraussetzungen (i) und (ii) des Satzes 9.2 erfüllt. □

Wir zeigen, dass bei gleichmäßiger Konvergenz Stetigkeitseigenschaften der Funktionen $f_k(x)$ auf die Grenzfunktion $F(x)$ vererbt werden. In diesem Sachverhalt liegt die besondere Bedeutung der gleichmäßigen Konvergenz.

Satz 9.4 (a) *Konvergiert die Folge stetiger Funktionen* $f_k : K \to \mathbf{K}$ *auf dem Intervall* $K := [a, b]$ **gleichmäßig** *gegen eine Grenzfunktion* $F(x)$*, so ist* $F : K \to \mathbf{K}$ *stetig.*

(b) *Für eine Folge stetiger Funktionen* $f_k : K \to \mathbf{K}$ *konvergiere die Funktionenreihe* $\sum_{k=0}^{\infty} f_k(x)$ *auf dem Intervall* $K := [a, b]$ **gleichmäßig** *gegen eine Grenzfunktion* $F(x)$*. Dann ist* $F : K \to \mathbf{K}$ *stetig.*

(c) *Die Grenzfunktion* $P(x) := \sum_{k=0}^{\infty} a_k x^k$ *einer Potenzreihe ist auf dem gesamten Konvergenzkreis* $K_\rho(0) = \{x \in \mathbf{C} : |x| < \rho\}$ *stetig.*

Begründungen: (a) Wir fixieren $x_0 \in K$ und beachten, dass auf Grund der Stetigkeit von f_k die Relation $\lim_{x \to x_0} |f_k(x) - f_k(x_0)| = 0 \; \forall \, k \in \mathbf{N}_0$ gilt. Wir wählen nun zu $\epsilon > 0$ eine Zahl $N(\epsilon)$ gemäß der Vorschrift (2.1). Dann gilt für alle $k \geq N(\epsilon)$:

$$|F(x) - F(x_0)| \leq \underbrace{|F(x) - f_k(x)|}_{\leq \epsilon} + |f_k(x) - f_k(x_0)| + \underbrace{|f_k(x_0) - F(x_0)|}_{\leq \epsilon}.$$

Hieraus erhält man $0 \leq \limsup_{x \to x_0} |F(x) - F(x_0)| \leq 2\epsilon \; \forall \; \epsilon > 0$, und das ist bereits die behauptete Stetigkeit der Funktion F.

(b)Da die n–te Partialsumme $s_n(x) := \sum_{k=0}^{n} f_k(x)$ auf dem Intervall K stetig ist, folgt aus der gleichmäßigen Konvergenz $s_n(x) \to F(x)$ die Stetigkeit der Grenzfunktion F gemäß (a).

(c) Es sei $x_0 \in K_\rho(0)$ fest gewählt. Setzt man $r := (|x_0| + \rho)/2 < \rho$, so konvergiert die Potenzreihe $P(x)$ nach Satz 9.3 gleichmäßig auf der abgeschlossenen Kreisscheibe $\overline{K}_r(0)$. Wegen (b) ist die Grenzfunktion $P(x)$ dort stetig, also insbesondere stetig im Punkt $x_0 \in \overline{K}_r(0)$. □

Bei gleichmäßiger Konvergenz vererben sich auch die Eigenschaften der R–Integrierbarkeit und der Differenzierbarkeit von $f_k(x)$ auf die Grenzfunktion $F(x)$.

Satz 9.5 *Die Funktionen* $f_k : K \to \mathbf{K}$ *seien auf dem Intervall* $K := [a, b]$ *R–integrierbar, und die Funktionenreihe* $\sum_{k=0}^{\infty} f_k(x)$ *konvergiere auf* K **gleichmäßig** *gegen die Grenzfunktion* $F(x)$*. Dann ist auch* F *auf* K *R–integrierbar, und es gilt*

$$\boxed{\int_a^b F(x)\,dx = \int_a^b \Big(\sum_{k=0}^{\infty} f_k(x)\Big)\,dx = \sum_{k=0}^{\infty} \int_a^b f_k(x)\,dx.} \tag{2.2}$$

Das heißt, eine **gleichmäßig konvergente** *Funktionenreihe darf gliedweise (bestimmt) integriert werden.*

Begründung: Da die Folge der Partialsummen $s_n(x) := \sum_{k=0}^{n} f_k(x)$ nach Voraussetzung auf K gleichmäßig gegen die Grenzfunktion $F(x)$ konvergiert, gilt

$$\forall\, \epsilon > 0\ \exists\, N = N(\epsilon) : \sup_{x\in K} |s_n(x) - s_m(x)| \le \sup_{x\in K} |s_n(x) - F(x)| + \sup_{x\in K} |F(x) - s_m(x)| \le 2\epsilon\ \forall\, n, m \ge N. \tag{2.3}$$

Wir setzen

$$S_n := \int_a^b s_n(x)\, dx = \sum_{k=0}^{n} \int_a^b f_k(x)\, dx,\ n \in \mathbf{N}.$$

Dann folgt aus (2.3) für $n, m \ge N(\epsilon)$:

$$|S_n - S_m| \le \int_a^b \sup_{t\in K} |s_n(t) - s_m(t)|\, dx \le 2\epsilon(b-a).$$

Das heißt, die Zahlenfolge $\big(S_n\big)_{n\in\mathbf{N}} \subset \mathbf{K}$ ist eine CAUCHY–Folge, und ihr Grenzwert

$$S := \sum_{k=0}^{\infty} \int_a^b f_k(x)\, dx \in \mathbf{K}$$

existiert. Darüber hinaus gilt wegen (2.3) noch $|S - S_m| \le 2\epsilon(b-a)\ \forall\, m \ge N(\epsilon)$. Wir zeigen hiermit $S = \int_a^b F(x)\, dx$. In der Tat, es gilt für $m \ge N(\epsilon)$:

$$\Big| \int_a^b F(x)\, dx - S \Big| \le \Big| \int_a^b F(x)\, dx - S_m \Big| + |S_m - S| \le \int_a^b \sup_{t\in K} |F(t) - s_m(t)|\, dx + 2\epsilon(b-a) \le 3\epsilon(b-a).$$

Im Limes $\epsilon \to 0+$ folgt hieraus:

$$S = \int_a^b F(x)\, dx = \int_a^b \Big(\sum_{k=0}^{\infty} f_k(x) \Big)\, dx = \sum_{k=0}^{\infty} \int_a^b f_k(x)\, dx.$$

Bemerkung 9.4 Da sich die Funktionenfolge $\big(f_k(x)\big)_{k\in\mathbf{N}_0}$ als Folge der Partialsummen $s_n(x)$ $:= \sum_{k=0}^{n} \big(f_k(x) - f_{k-1}(x)\big)$ mit $f_{-1} = 0$ schreiben lässt, gilt der obige Satz 9.5 auch für Funktionenfolgen: Falls unter den Voraussetzungen von Satz 9.5 die Konvergenz $f_k(x) \to F(x)$ auf dem Intervall K gleichmäßig erfolgt, so ist die Grenzfunktion F R–integrierbar, und es gilt□

$$\boxed{\lim_{k\to\infty} \int_a^b f_k(x)\, dx = \int_a^b F(x)\, dx.}$$

BSP. (9.2.6) Die Funktionenreihe $\sum_{k=1}^{\infty} \frac{1}{k^2} \sin\frac{x}{k^4}$ konvergiert gemäß dem WEIERSTRASS–Kriterium **gleichmäßig** für alle $x \in \mathbf{R}$. Wir können Satz 9.5 anwenden und erhalten

$$\int_0^x \Big(\sum_{k=1}^{\infty} \frac{1}{k^2} \sin\frac{t}{k^4} \Big)\, dt = \sum_{k=1}^{\infty} k^2 \Big(1 - \cos\frac{x}{k^4} \Big).$$

Die Konvergenz dieser Reihe erschließt man aus dem asymptotischen Verhalten

$$k^2 \Big(1 - \cos\frac{x}{k^4} \Big) \sim \frac{x^2}{2k^6},\ k \gg 1.$$

Hätte man hingegen die Funktionenreihe **unbestimmt** integriert unter Verwendung der Stammfunktion $-k^4 \cos\frac{x}{k^4} = \int \sin\frac{x}{k^4}\,dx$, so wäre die resultierende Reihe $-\sum\limits_{k=1}^{\infty} k^2 \cos\frac{x}{k^4}$ für **kein** $x \in \mathbf{R}$ konvergent.

Merke: Satz 9.5 gilt im allgemeinen nicht mehr bei **unbestimmter Integration**.

Satz 9.6 *Die Funktionen $f_k : K \to \mathbf{K}$ seien auf dem Intervall $K := [a,b]$ differenzierbar, und die Ableitungen f_k' seien auf K R–integrierbar. Falls die Funktionenreihe $\sum\limits_{k=0}^{\infty} f_k'(x)$ auf K* **gleichmäßig** *konvergiert und falls die Reihe $\sum\limits_{k=0}^{\infty} f_k(x_0)$ wenigstens für ein $x_0 \in K$ konvergent ist, so ist die Grenzfunktion $F(x) := \sum\limits_{k=0}^{\infty} f_k(x)$ auf K differenzierbar, und es gilt*

$$\boxed{F'(x) = \Big(\sum_{k=0}^{\infty} f_k(x)\Big)' = \sum_{k=0}^{\infty} f_k'(x) \;\; \forall\, x \in K.} \tag{2.4}$$

Das heißt, die Funktionenreihe darf **gliedweise** *differenziert werden. Eine entsprechende Aussage gilt auch für die Grenzfunktion $F(x)$ der Funktionenfolge $\big(f_k(x)\big)_{k\in\mathbf{N}_0}$. Falls die Ableitungen f_k' auf K R–integrierbar sind und eine* **gleichmäßig** *konvergente Folge bilden, falls ferner $\lim\limits_{k\to\infty} f_k(x_0) = F(x_0)$ für wenigstens ein $x_0 \in K$ gilt, so existiert die differenzierbare Grenzfunktion $F(x) = \lim\limits_{k\to\infty} f_k(x)$, und es gilt*

$$\boxed{F'(x) = \lim_{k\to\infty} f_k'(x) \;\; \forall\, x \in K.}$$

Begründung: Die Funktionenreihe $\sum\limits_{k=0}^{\infty} f_k'(x)$ darf wegen Satz 9.5 gliedweise integriert werden:

$$\int_{x_0}^{x} \Big(\sum_{k=0}^{\infty} f_k'(t)\Big)\,dt = \sum_{k=0}^{\infty} \Big(f_k(x) - f_k(x_0)\Big), \quad x \in K.$$

Die Funktionenreihe $\sum\limits_{k=0}^{\infty} f_k(x)$ ist konvergent, da dies nach Voraussetzung auf die obige Reihe und die Reihe $\sum\limits_{k=0}^{\infty} f_k(x_0)$ zutrifft. Darüber hinaus gilt:

$$F(x) := \sum_{k=0}^{\infty} f_k(x) = \sum_{k=0}^{\infty} f_k(x_0) + \int_{x_0}^{x} \Big(\sum_{k=0}^{\infty} f_k'(t)\Big)\,dt, \quad x \in K,$$

und durch Differentiation erhält man daraus die Relation (2.4). □

BSP. (9.2.7) Es seien $f_k(x)$ und $F(x)$ die Funktionen aus BSP. (9.2.3). Durch gliedweise Differentiation erhalten wir

$$-\frac{4}{\pi^2} \sum_{k=0}^{\infty} f_k'(x) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{\sin(2k+1)\pi x}{2k+1}.$$

Für diese Reihe kann die gleichmäßige Konvergenz auf $K := [-1,+1]$ nicht nachgewiesen werden. Deshalb darf die Funktion $F(x) := |x| = \frac{1}{2} - \frac{4}{\pi^2} \sum\limits_{k=0}^{\infty} f_k(x)$ auf K nicht differenziert werden. In der Tat, im Punkt $x = 0$ existiert keine Ableitung $F'(x)$, während $-\frac{4}{\pi^2} \sum\limits_{k=0}^{\infty} f_k'(0) = 0$ liefert. Betrachten wir hingegen die Funktionenreihe

$$F(x) = \sum_{k=1}^{\infty} f_k(x) := \sum_{k=1}^{\infty} \frac{(-1)^k}{\sqrt{k}} e^{-x/k}, \quad x \in K := [0,+\infty),$$

so konvergiert nach dem LEIBNIZ–Kriterium die Reihe

$$\sum_{k=1}^{\infty} f_k(0) = \sum_{k=1}^{\infty} \frac{(-1)^k}{\sqrt{k}}.$$

Wegen $\sum\limits_{k=1}^{\infty} |f_k'(x)| \le \sum\limits_{k=1}^{\infty} \frac{1}{k^{3/2}} < +\infty$ erhalten wir überdies auf K die gleichmäßige Konvergenz der Funktionenreihe $\sum\limits_{k=1}^{\infty} f_k'(x)$. Aus Satz 9.6 folgt deshalb

$$F'(x) = \sum_{k=1}^{\infty} f_k'(x) = -\sum_{k=1}^{\infty} \frac{(-1)^k}{k\sqrt{k}} e^{-x/k}, \quad x \in K.$$

Liegt speziell eine Potenzreihe $P(x) := \sum\limits_{k=0}^{\infty} a_k x^k$ vor, so sind die Funktionen $f_k(x) := a_k x^k$ von jeder Ordnung stetig differenzierbar. Die gliedweise differenzierte Potenzreihe

$$\sum_{k=1}^{\infty} a_k k x^{k-1} = \sum_{k=0}^{\infty} a_{k+1}(k+1) x^k$$

ist wiederum eine Potenzreihe, und deren Konvergenzradius

$$\Big(\limsup_{k\to\infty} \sqrt[k]{|a_{k+1}|(k+1)} \Big)^{-1} = \Big(\limsup_{k\to\infty} \sqrt[k]{|a_{k+1}|} \Big)^{-1} = \rho$$

ist derselbe wie der Konvergenzradius der Ausgangsreihe $P(x)$. Wegen Satz 9.3 konvergiert nun die gliedweise differenzierte Potenzreihe gleichmäßig auf jeder abgeschlossenen Kreisscheibe $\overline{K}_r(0)$ vom Radius $r < \rho$. Somit ist Satz 9.6 anwendbar:

$$P'(x) = \Big(\sum_{k=0}^{\infty} a_k x^k \Big)' = \sum_{k=0}^{\infty} a_k k x^{k-1}.$$

Wenden wir diese Überlegungen nochmals auf $P'(x)$ an, danach auf $P''(x)$, und so fort, so erhalten wir die folgende Aussage:

Satz 9.7 *Die Grenzfunktion $P(x)$ einer Potenzreihe $\sum\limits_{k=0}^{\infty} a_k x^k$ ist innerhalb des Konvergenzkreises $K_\rho(0)$ beliebig oft stetig differenzierbar. Ihre Ableitungen $P^{(n)}(x)$ lassen sich durch gliedweise Differentiation bestimmen:*

$$\boxed{\begin{aligned} P'(x) &= \sum_{k=1}^{\infty} a_k k x^{k-1}, \\ P''(x) &= \sum_{k=2}^{\infty} a_k k(k-1) x^{k-2}, \\ &\vdots \\ P^{(n)}(x) &= \sum_{k=n}^{\infty} a_k \binom{k}{n} n! x^{k-n}, \quad n \in \mathbf{N}. \end{aligned}} \tag{2.5}$$

Jede dieser Reihen hat wiederum denselben Konvergenzkreis $K_\rho(0)$.

Aus der Beziehung (2.5) ergibt sich speziell $P^{(n)}(0) = n! a_n$, und somit $a_n = \frac{1}{n!} P^{(n)}(0)\ \forall\ n \in \mathbf{N}_0$. Es gilt also

$$\boxed{P(x) = \sum_{k=0}^{\infty} \frac{1}{k!} P^{(k)}(0)\, x^k \quad \forall\ x \in K_\rho(0),}$$

und folglich

Satz 9.8 *Jede Potenzreihe ist auf dem Konvergenzkreis die* TAYLOR*–Reihe ihrer Grenzfunktion.*

In Erweiterung des Integrationssatzes 9.5 für allgemeine Funktionenreihen dürfen Potenzreihen auch unbestimmt integriert werden:

Satz 9.9 *Die Grenzfunktion $P(x)$ der Potenzreihe $\sum\limits_{k=0}^{\infty} a_k x^k$ hat auf dem Konvergenzkreis $K_\rho(0)$ eine Stammfunktion $F(x) := \int P(x)\,dx$. Diese kann durch gliedweise Integration aus der Ausgangsreihe gewonnen werden:*

$$\boxed{F(x) := \int P(x)\,dx = \sum_{k=0}^{\infty} a_k \int x^k\,dx = \sum_{k=0}^{\infty} \frac{a_k}{k+1} x^{k+1}.} \tag{2.6}$$

Der Konvergenzkreis der Potenzreihe (2.6) *ist wiederum $K_\rho(0)$.*

Begründung: Die Potenzreihe (2.6) hat den Konvergenzradius

$$\Big(\limsup_{k\to\infty} \sqrt[k]{\frac{|a_k|}{k+1}}\Big)^{-1} = \Big(\underbrace{\lim_{k\to\infty} \sqrt[k]{\frac{1}{k+1}}}_{=1} \limsup_{k\to\infty} \sqrt[k]{|a_k|}\Big)^{-1} = \rho.$$

Wir können Satz 9.7 auf diese Potenzreihe anwenden. Durch gliedweises Differenzieren erhält man $F'(x) = P(x)\ \forall\, x \in K_\rho(0)$. □

Der Satz 9.4 trifft eine Aussage über die Stetigkeit der Grenzfunktion $P(x) := \sum\limits_{k=0}^{\infty} a_k x^k$ nur im Inneren der Kreisscheibe $K_\rho(0)$. Bezüglich der Stetigkeit in den Randpunkten $x = \pm\rho$ gilt der folgende

Satz 9.10 (ABELscher Grenzwertsatz)
Ist die Potenzreihe $P(x) := \sum\limits_{k=0}^{\infty} a_k x^k$ auch noch für $x = +\rho$ oder $x = -\rho$ konvergent, so ist die Grenzfunktion $P(x)$ in dem betreffenden Punkt $x = \pm\rho$ stetig:

$$\boxed{P(\pm\rho) = \lim_{x\to\pm\rho} \sum_{k=0}^{\infty} a_k x^k.}$$

Wir verzichten hier auf eine Begründung und verweisen stattdessen auf die Standardliteratur, zum Beispiel H. HEUSER, Lehrbuch der Analysis, Teil 1. B.G.Teubner Verlag, Stuttgart 1980, S.379.

Mit den hier angegebenen Sätzen können in einfacher Weise die TAYLOR–Reihen zahlreicher Elementarfunktionen berechnet werden. Wir werden dies in einer Reihe von Beispielen aufzeigen.

BSP. (9.2.8) Bestimme die TAYLOR–Reihe der Funktion $F(x) := \ln(1+x)$ im Entwicklungspunkt $x_0 = 0$.
Lösung: Es gilt $F(0) = 0$ sowie

$$\Big(\ln(1+x)\Big)' = \frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k \quad \forall\, |x| < 1.$$

Unter Verwendung von Satz 9.9 erhält man daraus

$$\boxed{\ln(1+x) = F(x) - F(0) = \sum_{k=0}^{\infty} (-1)^k \int_0^x t^k\,dt = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1}\, x^{k+1} \quad \forall\, |x| < 1.}$$

Nach dem LEIBNIZ–Kriterium ist auch die Reihe $\sum\limits_{k=0}^{\infty} \frac{(-1)^k}{k+1}$ konvergent, so dass wir aus dem ABELschen Grenzwertsatz die Stetigkeit der Grenzfunktion $F(x)$ im Punkt $x = 1$ erschließen. Wir folgern

$$\boxed{F(1) = \ln 2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1}.}$$

BSP. (9.2.9) Bestimme die TAYLOR–Reihe der Funktion $F(x) := \text{arc tan}_H\, x$ im Entwicklungspunkt $x_0 = 0$. *Lösung:* Es gilt $F(0) = 0$ sowie

$$\Big(\text{arc tan}_H\, x\Big)' = \frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k} \quad \forall\, |x| < 1.$$

Unter Verwendung von Satz 9.9 erhält man daraus

$$\boxed{\text{arc tan}_H\, x = F(x) - F(0) = \sum_{k=0}^{\infty} (-1)^k \int\limits_0^x t^{2k}\, dt = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}\, x^{2k+1} \quad \forall\, |x| < 1.}$$

In den Punkten $x = \pm 1$ ist wiederum der ABELsche Grenzwertsatz anwendbar:

$$\boxed{F(1) = \text{arc tan}_H\, 1 = \frac{\pi}{4} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} = -\text{arc tan}_H(-1).}$$

BSP. (9.2.10) Bestimme die TAYLOR–Reihe der GAUSS–Fehlerfunktion $F(x) := \text{erf}\, x = \frac{2}{\sqrt{\pi}} \int\limits_0^x e^{-t^2}\, dt$ im Entwicklungspunkt $x_0 = 0$. *Lösung:* Es gilt $F(0) = 0$ sowie

$$\Big(\text{erf}\, x\Big)' = \frac{2}{\sqrt{\pi}} e^{-x^2} = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}\, x^{2k} \quad \forall\, x \in \mathbf{R}.$$

Unter Verwendung von Satz 9.9 erhält man daraus

$$\boxed{\text{erf}\, x = F(x) - F(0) = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int\limits_0^x t^{2k}\, dt = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(2k+1)}\, x^{2k+1} \quad \forall\, x \in \mathbf{R}.}$$

BSP. (9.2.11) Bestimme die TAYLOR–Reihe des Integralsinus $F(x) := Si(x) := \int\limits_0^x \frac{\sin t}{t}\, dt$ im Entwicklungspunkt $x_0 = 0$. *Lösung:* Es gilt $F(0) = 0$ sowie

$$\Big(Si(x)\Big)' = \frac{\sin x}{x} = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\, x^{2k} \quad \forall\, x \in \mathbf{R}.$$

Unter Verwendung von Satz 9.9 erhält man daraus

$$\boxed{Si(x) = F(x) - F(0) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \int\limits_0^x t^{2k}\, dt = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!(2k+1)}\, x^{2k+1} \quad \forall\, x \in \mathbf{R}.}$$

Die Frage, ob verschiedene Potenzreihen auf demselben Konvergenzkreis dieselbe Grenzfunktion haben können, beantworten wir in dem folgenden

Satz 9.11 (Identitätssatz für Potenzreihen)
Die beiden Potenzreihen $P(x) := \sum\limits_{k=0}^{\infty} a_k x^k$ *und* $Q(x) := \sum\limits_{k=0}^{\infty} b_k x^k$ *seien konvergent in* $K_\rho(0)$, $\rho > 0$. *Genau dann haben wir Gleichheit* $P(x) = Q(x)\ \forall\ x \in K_\rho(0)$, *wenn* $a_k = b_k\ \forall\ k \geq 0$ *gilt.*

Begründung: Gilt $a_k = b_k \ \forall\, k \geq 0$, so ist trivialerweise $P = Q$. Gilt umgekehrt $P(x) = Q(x) \ \forall\, x \in K_\rho(0)$, so nehmen wir an, es sei $N \in \mathbf{N}_0$ der kleinste Index, für den $a_N \neq b_N$ erfüllt ist. Dann folgt

$$P(x) - Q(x) = 0 = \sum_{k=N}^{\infty} (a_k - b_k) x^k \ \ \forall\, x \in K_\rho(0).$$

Wird diese Identität durch x^N dividiert, so folgt danach im Limes $x \to 0$ die Bedingung $a_N = b_N$, entgegen der Annahme $a_N \neq b_N$. □

Auf dem Identitätssatz beruht die **Methode des Koeffizientenvergleichs**: Gelten für dieselbe Funktion $P(x)$ zwei Potenzreihenentwicklungen

$$\sum_{k=0}^{\infty} a_k x^k = P(x) = \sum_{k=0}^{\infty} b_k x^k,$$

so folgt stets $a_k = b_k \ \forall\, k \geq 0$.

BSP. (9.2.12) Bestimme die TAYLOR–Reihe der Funktion $F(x) := \tan x$ im Entwicklungspunkt $x_0 = 0$ mit der **Methode der unbestimmten Koeffizienten**. *Lösung:* Da $F(x)$ eine *ungerade* Funktion ist, setzt man eine Potenzreihe mit unbestimmten Koeffizienten in der folgenden Form an:

$$P(x) := \tan x = \sum_{k=0}^{\infty} a_k x^{2k+1} = \frac{\sin x}{\cos x}.$$

Unter Verwendung der bekannten Potenzreihenentwicklungen von $\sin x$ und $\cos x$ erhält man mit Hilfe des CAUCHY–Produktes zweier Reihen:

$$\begin{aligned} \sin x &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\, x^{2k+1} = \cos x \cdot \sum_{k=0}^{\infty} a_k x^{2k+1} = \Big(\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!}\, x^{2k} \Big) \cdot \Big(\sum_{k=0}^{\infty} a_k x^{2k+1} \Big) \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{k} \frac{(-1)^n x^{2n} x^{2k-2n+1}}{(2n)!}\, a_{k-n} = \sum_{k=0}^{\infty} x^{2k+1} \sum_{n=0}^{k} \frac{(-1)^n}{(2n)!}\, a_{k-n}. \end{aligned}$$

Durch Koeffizientenvergleich resultiert nun die folgende **Rekursionsformel**:

$$\boxed{\frac{(-1)^k}{(2k+1)!} = \sum_{n=0}^{k} \frac{(-1)^n}{(2n)!}\, a_{k-n} \ \ \forall\, k \in \mathbf{N}_0.}$$

Aus dieser Formel können die unbestimmten Koeffizienten a_k sukzessive berechnet werden. Man verifiziert mit einigem elementaren Rechenaufwand die folgenden Zahlen:

$$a_0 = 1, \quad a_1 = \frac{1}{3}, \quad a_2 = \frac{2}{3 \cdot 5}, \quad a_3 = \frac{17}{3^2 \cdot 5 \cdot 7},$$

und hieraus folgt

$$\boxed{\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \cdots + \frac{(-1)^n 2^{2n} (2^{2n} - 1) B_{2n}}{(2n)!} x^{2n-1} + \cdots}$$

Die hier verwendeten Zahlen B_{2n} sind die BERNOULLI–Zahlen:

Definition 9.5 *Gegeben seien Zahlen $t \in \mathbf{R}$ und $z \in \mathbf{C}$ mit $|z| < 2\pi$. Die in der Potenzreihenentwicklung*

$$\frac{z e^{tz}}{e^z - 1} = \sum_{j=0}^{+\infty} B_j(t)\, \frac{z^j}{j!} \tag{2.7}$$

auftretenden Polynome $B_j(t)$ mit $\operatorname{Grad} B_j = j$ heißen BERNOULLI–**Polynome.** *Die Zahlen*

$$B_j := B_j(0) \ \ \forall\, j = 0, 1, \ldots, \tag{2.8}$$

heißen BERNOULLI–**Zahlen.**

Bemerkung 9.5 Die ersten BERNOULLI–Zahlen sind:

$$B_0 = 1, \quad B_1 = -\tfrac{1}{2}, \quad B_2 = \tfrac{1}{6}, \quad B_3 = 0, \quad B_4 = -\tfrac{1}{30}, \quad B_5 = 0,$$
$$B_6 = \tfrac{1}{42}, \quad B_7 = 0, \quad B_8 = -\tfrac{1}{30}, \quad B_9 = 0, \quad B_{10} = \tfrac{5}{66}, \quad B_{11} = 0.$$

Stets gilt $B_{2n+1} = 0 \; \forall \, n \in \mathbf{N}$.

Der *Konvergenzradius* der nach der Methode der unbestimmten Koeffizienten berechneten Potenzreihe ist im allgemeinen schwierig zu bestimmen. Sicher wird der Konvergenzradius ρ im Fall der Reihe

$$\sum_{k=0}^{\infty} a_k x^k = \frac{P(x)}{Q(x)}, \;\; Q(0) \neq 0,$$

höchstens bis zur betragskleinsten Nullstelle der Funktion $Q(x)$ reichen. Im Beispiel der Tangens–Reihe gilt also sicher $\rho \leq \pi/2$. □

Aus Satz 9.7 folgt unmittelbar, dass die Grenzfunktion $P(x)$ einer Potenzreihe auf dem Konvergenzkreis $K_\rho(0)$ eine C^∞–Funktion ist. Es wäre umgekehrt falsch zu glauben, dass jede C^∞–Funktion $f(x)$ auch eine Potenzreihenentwicklung zulässt. Formal darf man in jedem C^∞–Punkt x_0 die TAYLOR–Reihe

$$\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0) \, (x - x_0)^k$$

der Funktion $f(x)$ hinschreiben, jedoch braucht diese Reihe für keinen Wert $x \neq x_0$ die Funktion $f(x)$ darzustellen. Wir hatten diese Tatsache bereits in Abschnitt 7.7 diskutiert. Dort wurde die Funktion

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

genannt mit den Eigenschaften $f \in C^\infty(\mathbf{R})$ sowie $f^{(k)}(0) = 0 \; \forall k \in \mathbf{N}_0$. Die formale TAYLOR–Reihe $\sum\limits_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0) \, x^k = 0$ stellt die gegebene Funktion $f(x)$ nur im Punkt $x_0 = 0$ dar. Wir erinnern an Satz 7.20. Dieser besagt, dass die C^∞–Funktion $f(x)$ genau dann an der Stelle x_0 in eine TAYLOR–Reihe entwickelbar ist, wenn für alle x in einer Umgebung des Punktes x_0 gilt:

$$\lim_{n \to \infty} R_n(x; x_0) := \lim_{n \to \infty} \frac{(x - x_0)^{n+1}}{(n+1)!} \, f^{(n+1)}(\xi) = 0. \tag{2.9}$$

Im obigen Beispiel ist diese Bedingung nur im Punkt $x_0 = 0$ erfüllt. Es gilt allgemein der folgende

Satz 9.12 *Zu gegebener Funktion $f \in C^\infty([a,b])$ existiere eine Zahl $M > 0$ mit der Eigenschaft $|f^{(k)}(x)| \leq M < +\infty \; \forall \; x \in [a,b] \; \forall \; k \in \mathbf{N}_0$. Dann gilt an jeder Stelle $x_0 \in (a,b)$ die* TAYLOR*–Entwicklung*

$$\boxed{f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} \, f^{(k)}(x_0) \, (x - x_0)^k \;\; \forall \; x \in [a,b].}$$

Begründung: Wir zeigen, dass das LAGRANGE–Restglied die Bedingung (2.9) erfüllt:

$$0 \leq \lim_{n \to \infty} |R_n(x; x_0)| = \lim_{n \to \infty} \frac{|x - x_0|^{n+1}}{(n+1)!} \, |f^{(n+1)}(\xi)| \leq M \lim_{n \to \infty} \frac{|x - x_0|^{n+1}}{(n+1)!} = 0.$$

Wir treffen in diesem Zusammenhang die folgende

Definition 9.6 *Eine Funktion f heiße in dem Intervall $[a,b]$* **analytisch**, *wenn $f(x)$ in jedem Punkt $x_0 \in (a,b)$ in eine Potenzreihe entwickelbar ist. Die Klasse der über dem Intervall $[a,b]$ analytischen Funktionen bezeichnen wir mit $C^\omega(a,b)$.*

BSP. (9.2.13) Die Funktion $f(x) := \sin x$ gehört zur Klasse $C^\omega(\mathbf{R})$. Denn wegen

$$|f^{(k)}(x)| = \begin{cases} |\cos x| \le 1 \ : \ k \ \text{gerade}, \\ |\sin x| \le 1 \ : \ k \ \text{ungerade}, \end{cases}$$

sind die Voraussetzungen zum Satz 9.12 mit $M = 1$ erfüllt.

LANDAU–Symbole

In vielen Fällen genügt es, eine gegebene Funktion $f(x)$ durch ihr TAYLOR–Polynom $T_n(x)$ vom Grade n im Entwicklungspunkt x_0 zu ersetzen. *Zum Beispiel* liefert

$$T_3(x) := 1 + x + \frac{x}{2} + \frac{x^3}{6} \approx e^x =: f(x)$$

für $x \to 0$ eine ausreichende Näherung der Exponentialfunktion. Zur Beschreibung dieses Sachverhaltes, das heißt, zur Beschreibung des **asymptotischen Verhaltens** einer Funktion $f(x)$ in der Nähe eines Entwicklungspunktes, verwendet man die LANDAU–Symbole:

Definition 9.7 *Auf der Menge $X \subset \mathbf{R}$ seien Funktionen $f, g : X \to \mathbf{R}$ gegeben, und es sei x_0 ein Häufungspunkt von X.*

(a) *Falls der Grenzwert $\lim\limits_{X \ni x \to x_0} \frac{f(x)}{g(x)} = 0$ existiert, so schreibt man*

$$\boxed{f(x) = o\Big(g(x)\Big) \ \textit{für} \ x \to x_0, \ (\textit{sprich: klein oh von } g(x)).}$$

(b) *Falls $\left|\frac{f(x)}{g(x)}\right| \le M < +\infty$ in einem Intervall $(x_0 - \epsilon, x_0 + \epsilon) \cap X$ gilt, so schreibt man*

$$\boxed{f(x) = \mathcal{O}\Big(g(x)\Big) \ \textit{für} \ x \to x_0, \ (\textit{sprich: groß oh von } g(x)).}$$

Zum Beispiel: Die TAYLOR–Formel mit LAGRANGE–Restglied kann unter geeigneten Voraussetzungen an die Funktion f mit Hilfe der obigen LANDAU–Symbolik formuliert werden:

$$\boxed{f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x_0)\,(x - x_0)^k + \mathcal{O}\Big(|x - x_0|^{n+1}\Big) \ \text{für} \ x \to x_0.}$$

Wir haben speziell für $f(x) := e^x$ und $x_0 := 0$:

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + \mathcal{O}(x^{n+1}) = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + o(x^n) \ \text{für} \ x \to 0.$$

Mit der LANDAU–Symbolik können asymptotische Rechnungen häufig eleganter dargestellt werden.

BSP. (9.2.14) Umgehung der Regeln von L'HOSPITAL:

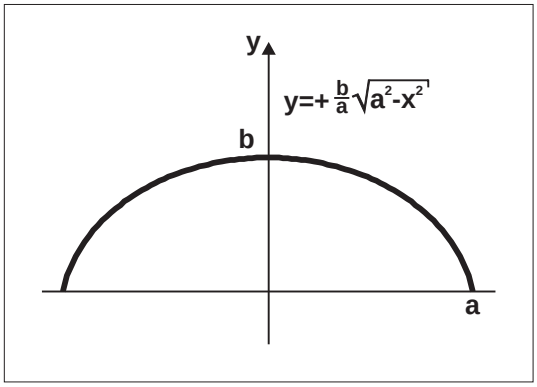
$$\lim_{x \to 0} \frac{2e^{-x} + 2x - x^2 - 2}{\sin x - x} = \lim_{x \to 0} \frac{2 - 2x + x^2 - x^3/3 + \mathcal{O}(x^4) + 2x - x^2 - 2}{x - x^3/6 + \mathcal{O}(x^5) - x} = \lim_{x \to 0} \frac{-1/3 + \mathcal{O}(x)}{-1/6 + \mathcal{O}(x^2)} = 2.$$

Wir haben hier entsprechende TAYLOR–Entwicklungen für e^{-x} und $\sin x$ eingesetzt.

BSP. (9.2.15) Der Umfang L_{Ell} einer Ellipse mit den Halbachsen $a, b > 0$ kann nicht mehr elementar berechnet werden. Wie in Ing.–Math.III zu begründen sein wird, hat man

$$L_{\text{Ell}} = 2 \int_{-a}^{a} \sqrt{1 + y'^2(x)}\, dx$$

mit $y(x) := (b/a)\sqrt{a^2 - x^2}$ und $y'(x) = -bx/(a\sqrt{a^2 - x^2})$.



Zum Umfang einer Ellipse

Führt man die **numerische Exzentrizität** $\epsilon^2 := (a^2 - b^2)/a^2,\ a > b$, ein, so erhält man

$$L_{\text{Ell}} = 2 \int_{-a}^{a} \sqrt{\frac{a^2 - \epsilon^2 x^2}{a^2 - x^2}}\, dx.$$

Dieses Integral transformieren wir mit Hilfe der Substitution $x = g(t) := a \sin t,\ dx = a \cos t\, dt$:

$$\boxed{L_{\text{Ell}} = 2a \int_{-\pi/2}^{\pi/2} \sqrt{1 - \epsilon^2 \sin^2 t}\, dt = 4a \int_{0}^{\pi/2} \sqrt{1 - \epsilon^2 \sin^2 t}\, dt =: 4a\, \mathbf{E}(\epsilon).}$$

Hierin bezeichnet $\mathbf{E}(\epsilon)$ das **vollständige elliptische Integral 2. Gattung**, vgl. Abschnitt 8.3, BSP. (8.3.4). Dieses Integral ist nicht mehr elementar darstellbar. Für *runde Ellipsen* $\epsilon \ll 1$ (das heißt für $a \approx b$) kann der Integrand nach Potenzen $(\epsilon \sin t)^k$ entwickelt werden:

$$\begin{aligned} L_{\text{Ell}} &= 4a \int_{0}^{\pi/2} \Big(1 - \frac{1}{2}\,\epsilon^2 \sin^2 t - \frac{1}{2 \cdot 4}\,\epsilon^4 \sin^4 t - \frac{1 \cdot 3}{2 \cdot 4 \cdot 6}\,\epsilon^6 \sin^6 t - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 8}\,\epsilon^8 \sin^8 t + \mathcal{O}(\epsilon^{10})\Big)\, dt \\ &= 2\pi a \Big(1 - \frac{1}{4}\,\epsilon^2 - \frac{3}{64}\,\epsilon^4 - \frac{5}{256}\,\epsilon^6 - \frac{175}{16\,384}\,\epsilon^8 + \mathcal{O}(\epsilon^{10})\Big). \end{aligned}$$

In der Geodäsie verwendet man häufig die Näherungsformel

$$\boxed{L_{\text{Ell}} = \pi\Big(3\frac{a+b}{2} - \sqrt{ab}\Big) + \mathcal{O}(\epsilon^8).}$$

Warum diese Formel für $\epsilon \to 0$ tauglich ist, folgt aus den asymptotischen Entwicklungen

$$\begin{aligned} \frac{a+b}{2} &= \frac{a}{2}\Big(1 + \sqrt{1 - \epsilon^2}\Big) = a\,\Big(1 - \frac{1}{4}\,\epsilon^2 - \frac{4}{64}\,\epsilon^4 - \frac{8}{256}\,\epsilon^6 - \frac{320}{16\,384}\,\epsilon^8 + \mathcal{O}(\epsilon^{10})\Big), \\ \sqrt{ab} &= a\sqrt[4]{1 - \epsilon^2} = a\,\Big(1 - \frac{1}{4}\,\epsilon^2 - \frac{6}{64}\,\epsilon^4 - \frac{14}{256}\,\epsilon^6 - \frac{616}{16\,384}\,\epsilon^8 + \mathcal{O}(\epsilon^{10})\Big). \end{aligned}$$

Mit diesen beiden Entwicklungen erhalten wir:

$$\pi\Big(3\frac{a+b}{2} - \sqrt{ab}\Big) - L_{\text{Ell}} = \frac{6\pi a}{16\,384}\,\epsilon^8 + \mathcal{O}(\epsilon^{10}) = \mathcal{O}(\epsilon^8).$$

Kapitel 10

Lineare Differentialgleichungen

10.1 Lineare Differentialoperatoren

Bezeichnet $X \subset \mathbf{R}$ ein Intervall oder eine endliche Vereinigung von Intervallen, so hatten wir in Abschnitt 7.3 bereits festgestellt, dass mit jeder der Funktionenklassen

$$\boxed{C^k(X;\mathbf{K}) := C^k(X) := \{f \in \mathrm{Abb}\,(\mathbf{R},\mathbf{K}) \,:\, f \text{ ist } k\text{–mal stetig differenzierbar auf } X\},\ \ k \in \mathbf{N}_0,}$$

ein **Vektorraum** über dem Körper $\mathbf{K}$ ($= \mathbf{R}$ oder $:= \mathbf{C}$) vorliegt. Diese Aussage trifft in gleicher Weise auf die Funktionenklasse

$$C^\infty(X) := \bigcap_{k\in\mathbf{N}_0} C^k(X)$$

zu. Entsprechende Feststellungen können auch für die Klassen $C^k(X;\mathbf{K}^n)$ *vektorwertiger* Funktionen und für die Klassen $C^k(X;\mathbf{K}^{(m,n)})$ *matrixwertiger* Funktionen getroffen werden.

Wir hatten ferner in Abschnitt 5.2 gezeigt, dass in Vektorräumen der Begriff der **linearen Abbildung** sinnvoll erklärt werden kann. Während lineare Abbildungen in *endlichdimensionalen* Vektorräumen genau mit den Matrizen identifiziert werden können, ist die Mannigfaltigkeit der linearen Abbildungen in den *unendlichdimensionalen* Vektorräumen C^k erheblich größer. Wir bezeichnen nachfolgend wie in Abschnitt 7.3 mit D den *Differentialoperator* $D := d/dx$. Ist C^k einer der Vektorräume $C^k(X)$, $C^k(X;\mathbf{K}^n)$ oder $C^k(X;\mathbf{K}^{(m,n)})$, so ist die Abbildung

$$D : \begin{cases} C^{k+1} \to C^k, \\ f \mapsto Df := f' \ \text{ oder punktweise } \ f(x) \mapsto (Df)(x) := f'(x)\ \forall\, x \in X, \end{cases}$$

für jedes feste $k \in \mathbf{N}_0$ linear. Ebenso ist die p–fache Hintereinanderausführung von D linear, und zwar für jedes feste $k \in \mathbf{N}_0$ und jedes $p \in \mathbf{N}$ als Abbildung

$$D^p : \begin{cases} C^{k+p} \to C^k, \\ f \mapsto D^p f := f^{(p)} \ \text{ oder punktweise } \ f(x) \mapsto (D^p f)(x) := f^{(p)}(x)\ \forall\, x \in X. \end{cases}$$

Wir setzen noch vereinbarungsgemäß $D^0 := Id$.

BSP. (10.1.1) Mit Hilfe der LEIBNIZ–Regel aus Satz 7.7 erhält man für $p \in \mathbf{N}$:

$$\begin{aligned} D^p\Big(e^x \cdot \sin x\Big) &= \sum_{j=0}^{p} \binom{p}{j} (D^j \sin x)\,(D^{p-j} e^x) = e^x \sum_{j=0}^{p} \binom{p}{j} D^j \sin x \\ &= e^x \left(\sin x + \binom{p}{1}\cos x - \binom{p}{2}\sin x - \binom{p}{3}\cos x + \cdots + D^p \sin x\right). \end{aligned}$$

Dieses Beispiel legt es nahe, auch lineare Abbildungen von der Form $a(x)D^p$ zu betrachten. Für $a \in C(X;\mathbf{K})$ ist durch

$$a(x)D^p : \begin{cases} C^p \to C^0, \\ f(x) \mapsto a(x)\,(D^p f)(x) := a(x)\,f^{(p)}(x)\ \forall\ x \in X, \end{cases}$$

in der Tat eine lineare Abbildung erklärt.

Beachte jedoch: Für $L := a(x)D^p$ ist die Hintereinanderausführung $L^2 := L \circ L$ im allgemeinen nicht mehr erklärt, da selbst bei Vorgabe von $f \in C^k,\ k \geq p$, stets nur $Lf \in C^0$ gilt. *Verbessert* man aber die Eigenschaften der Koeffizientenfunktion $a(x)$, zum Beispiel durch die Forderung $a \in C^p(X;\mathbf{K})$, so ist auch L^2 erklärt. Insbesondere sind für $a \in C^\infty(X;\mathbf{K})$ *alle Potenzen* L^k erklärt.

BSP. (10.1.2) Wir betrachten den Differentialoperator $L := e^x D^2$. Es gilt hier

$$L^2 = (e^x D^2)\,(e^x D^2) = e^x\,D^2(e^x D^2) = e^x\,D(e^x D^2 + e^x D^3) = e^{2x}\,(D^2 + 2D^3 + D^4).$$

Nun folgt *zum Beispiel* $L^2 \sin x = e^{2x}\,(-\sin x - 2\cos x + \sin x) = -2e^{2x}\cos x$.

Wir können Ausdrücke in der Form $a(x)D^p$ durch Linearkombinationen miteinander verknüpfen. Auf diese Weise lassen sich neue lineare Abbildungen definieren:

Definition 10.1 *Sind Funktionen $a_j \in C(X;\mathbf{K}),\ j = 0, 1, \ldots, n$, gegeben, so heiße*

$$\boxed{L_n := \sum_{j=0}^{n} a_j(x)D^j = a_n(x)D^n + a_{n-1}(x)D^{n-1} + \cdots + a_1(x)D + a_0(x)}$$

ein **linearer gewöhnlicher Differentialoperator n–ter Ordnung** *. Die Abbildung*

$$L_n : \begin{cases} C^n \to C^0, \\ f(x) \mapsto (L_n f)(x) := a_n(x)\,f^{(n)}(x) + \cdots + a_1(x)\,f'(x) + a_0(x)\,f(x)\ \forall\ x \in X, \end{cases}$$

ist linear.

Bemerkung 10.1 (a) In den Nullstellen der Koeffizientenfunktion $a_n(x)$ hat der Differentialoperator L_n nur noch eine Ordnung $\leq n-1$. Um diesen *Ordnungsabfall* auszuschließen, setzt man in der Regel $a_n(x) \neq 0\ \forall\ x \in X$ voraus. Da der Operator L_n in diesem Fall auch durch $a_n(x)$ dividiert werden kann, darf man ohne Beschränkung der Allgemeinheit annehmen, dass L_n in der *kanonischen Form*

$$L_n := \sum_{j=0}^{n-1} a_j(x)D^j + D^n = D^n + a_{n-1}(x)D^{n-1} + \cdots + a_1(x)D + a_0(x) \tag{1.1}$$

vorliegt. Sind die Koeffizienten $a_j(x) \equiv a_j \in \mathbf{K}$ Konstanten, so liegt der Sonderfall eines linearen gewöhnlichen Differentialoperators **mit konstanten Koeffizienten** vor:

$$L_n := \sum_{j=0}^{n-1} a_j D^j + D^n = D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0. \tag{1.2}$$

(b) Sind L_m und L_n lineare gewöhnliche Differentialoperatoren der Ordnungen m bzw. n:

$$L_m := \sum_{j=0}^{m-1} a_j(x)D^j + D^m, \quad L_n := \sum_{j=0}^{n-1} b_j(x)D^j + D^n,$$

so sind die Komposita $L_m \circ L_n$ bzw. $L_n \circ L_m$ wiederum lineare gewöhnliche Differentialoperatoren der Ordnung $n+m$, sofern die Koeffizientenfunktionen die Differenzierbarkeitseigenschaften $a_j \in C^n(X;\mathbf{K})$ und $b_j \in C^m(X;\mathbf{K})$ haben. Man beachte

$$L_m \circ L_n \neq L_n \circ L_m, \qquad (L_n \circ L_m)f \neq (L_m f)\,(L_n f).$$

(b) Eine Ausnahme von der ersten Ungleichung liegt bei linearen gewöhnlichen Differentialoperatoren mit **konstanten Koeffizienten** vor. Solche Differentialoperatoren erfüllen stets $\square$

$$\boxed{L_m \circ L_n = L_n \circ L_m.} \tag{1.3}$$

BSP. (10.1.3) Es seien $L_1 := D-4$ und $L_2 := D^2+3D-2$ gesetzt. Dann erhält man für jede Funktion $f \in C^3(X)$:

$$\begin{aligned} L_1(L_2 f) &= (D-4)\,(f''+3f'-2f) = f'''+3f''-2f'-4f''-12f'+8f = (D^3-D^2-14D+8)f, \\ L_2(L_1 f) &= (D^2+3D-2)\,(f'-4f) = f'''+3f''-2f'-4f''-12f'+8f = (D^3-D^2-14D+8)f. \end{aligned}$$

Im Sinne der gewöhnlichen Multiplikation findet man auch formal

$$L_1 L_2 = (D-4)\,(D^2+3D-2) = D^3-D^2-14D+8.$$

Merke: Für lineare gewöhnliche Differentialoperatoren mit **konstanten Koeffizienten**

$$L_n := \sum_{j=0}^{n} a_j\, D^j, \quad a_n \neq 0,$$

gelten dieselben Rechengesetze wie für Polynome $P_n(\lambda) := \sum\limits_{j=0}^{n} a_j\,\lambda^j$. Insbesondere gelten die Regeln für das Rechnen mit binomischen Ausdrücken. *Zum Beispiel* gilt der binomische Lehrsatz

$$(D-a)^n = \sum_{j=0}^{n} \binom{n}{j} (-a)^{n-j} D^j.$$

10.2 Lineare Differentialgleichungen n–ter Ordnung

Definition 10.2 *Ist* $L_n : C^n \to C^0$ *mit* $L_n := \sum\limits_{j=0}^{n-1} a_j(x)D^j + D^n$ *ein linearer gewöhnlicher Differentialoperator* n*–ter Ordnung, und ist* $f \in C^0$ *gegeben, so heiße eine Gleichung in der Form*

$$\boxed{L_n y = f} \tag{2.1}$$

eine **lineare gewöhnliche Differentialgleichung** (DGl) n*–ter Ordnung für eine gesuchte Funktion* $y(x)$*. Ist* $f = 0$*, so heiße die Gleichung* (2.1) **homogen** *, sonst* **inhomogen**. *Eine Funktion* $y \in C^n(X)$ *heiße auf der Definitionsmenge* $X \subset \mathbf{R}$ *eine* **Lösung** *der DGl* (2.1)*, wenn gilt:*

$$\boxed{(L_n y)(x) = f(x) \;\; \forall\, x \in X.}$$

BSP. (10.2.1) Auf der Menge $X := \mathbf{R} \setminus \{\pm\frac{1}{2}\sqrt{2}\}$ sei der lineare Differentialoperator 2. Ordnung $L_2 : C^2(X) \to C^0(X)$ in der folgenden Weise erklärt:

$$L_2 := D^2 + \frac{2x(1+2x^2)}{1-2x^2}\,D - \frac{2(1+2x^2)}{1-2x^2}, \quad L_2 y = y'' + \frac{2x(1+2x^2)}{1-2x^2}\,y' - \frac{2(1+2x^2)}{1-2x^2}\,y.$$

Man prüft durch Einsetzen sehr einfach nach, dass die beiden Funktionen $y_1(x) := x$ und $y_2(x) := e^{x^2}$ Lösungen der homogenen DGl $L_2 y = 0$ sind. Das allgemeine Problem besteht **immer** in der Bestimmung **aller** Lösungen der inhomogenen DGl $L_2 y = f$.

Da der Operator L_n eine *lineare* Abbildung ist, können Aussagen der linearen Algebra auf die hier vorliegenden linearen DGln übertragen werden.

Satz 10.1 *Sei L_n ein linearer gewöhnlicher Differentialoperator n–ter Ordnung. Dann gilt:*

(a) *Die Lösungen $y_h \in C^n$ der homogenen DGl $L_n y = 0$ bilden einen Unterraum* $\operatorname{Kern} L_n$ *von C^n:*

$$\boxed{L_n y_1 = 0 = L_n y_2 \;\;\Rightarrow\;\; L_n(\lambda\, y_1 + \mu\, y_2) = 0 \;\; \forall\, \lambda, \mu \in \mathbf{K}.}$$

(b) *Ist $y_p \in C^n$ eine partikuläre Lösung der inhomogenen DGl $L_n y = f$, so ist die Lösungsgesamtheit der DGl $L_n y = f$ der affine Unterraum*

$$\boxed{\mathcal{L}(DGl) = y_p + \operatorname{Kern} L_n = \{y \in C^n \,:\, y = y_p + y_h \;\; mit \;\; y_h \in \operatorname{Kern} L_n\}.}$$

Die *Begründung* erfolgt ganz analog zum Satz 5.8.(d) für lineare Gleichungssysteme.

Zum Beispiel: Wir betrachten das BSP. (10.2.1): Gemäß Satz 10.1 ist $y_h(x) := C_1 x + C_2 e^{x^2}$ für beliebige $C_j \in \mathbf{K}$ ebenfalls Lösung der homogenen DGl $L_2 y = 0$. Also gilt $\operatorname{span}\{x,\, e^{x^2}\} \subseteq \operatorname{Kern} L_2$.

Es bleibt die Frage zu beantworten, welche *Dimension* der Lösungsraum $\operatorname{Kern} L_n$ der homogenen DGl $L_n y = 0$ hat. Hierzu ist es erforderlich, die **lineare Abhängigkeit** von Funktionen über einer Definitionsmenge $X \subset \mathbf{R}$ zu untersuchen. Es ist nicht immer einfach, über die lineare Abhängigkeit (**LA**) eines Funktionensystems $f_1, f_2, \ldots, f_n$ bzw. über dessen lineare Unabhängigkeit (**LU**) eine Aussage zu treffen.

BSP. (10.2.2) Die Funktionen $f_1(x) := \sin^2 x$ und $f_2(x) := \cos^2 x$ sind **LU** auf ganz $\mathbf{R}$. Denn aus $C_1 \sin^2 x + C_2 \cos^2 x = 0 \;\forall\, x \in \mathbf{R}$ folgt $C_1 = 0$, wenn $x = \pi/2$ gesetzt wird, und $C_2 = 0$, wenn $x = 0$ gesetzt wird. Hingegen sind die Funktionen $f_1(x), f_2(x)$ und $f_3(x) := \cos 2x$ **LA**, denn es gilt ja $\cos^2 x - \sin^2 x - \cos 2x = 0 \;\forall\, x \in \mathbf{R}$.

Im Sinne einer Analyse nehmen wir an, das Funktionensystem $f_1, f_2, \ldots, f_n \in C^{n-1}(X)$ sei **LA**. Dann existieren Zahlen $C_1, C_2, \ldots, C_n \in \mathbf{K}$ mit $\sum\limits_{j=1}^{n} |C_j| > 0$ derart, dass gilt

$$\sum_{j=1}^{n} C_j f_j(x) = 0 \;\; \forall\, x \in X.$$

Diese Identität differenzieren wir sukzessive $(n-1)$–mal, so dass das lineare Gleichungssystem

$$\sum_{j=1}^{n} C_j f_j^{(k)}(x) = 0, \quad k = 0, 1, \ldots, n-1, \;\; x \in X, \tag{2.2}$$

entsteht. Das System (2.2) hat genau dann für alle $x \in X$ nichttriviale Lösungen $C_1, C_2, \ldots, C_n$, wenn die Koeffizientendeterminante verschwindet:

$$W(x) := \det \begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{bmatrix} = 0 \;\; \forall\, x \in X. \tag{2.3}$$

Definition 10.3 *Die dem Funktionensystem $f_1, f_2, \ldots, f_n \in C^{n-1}(X)$ zugeordnete Determinante*

$$W(x) := \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}, \quad x \in X,$$

heiße die WRONSKI**–Determinante** *von $f_1(x), f_2(x), \ldots, f_n(x)$ auf der Menge X.*

Wir erhalten aus dieser Vorbetrachtung sofort das folgende Resultat:

Satz 10.2 *Ist das Funktionensystem $f_1, f_2, \ldots, f_n \in C^{n-1}(X)$ auf der Menge X* **LA**, *so gilt $W(x) = 0 \ \forall\, x \in X$. Gilt hingegen $W(x) \neq 0 \ \forall\ x \in X$, so ist das Funktionensystem $f_1, f_2, \ldots, f_n$ auf der Menge X* **LU**.

Bemerkung 10.2 Die Bedingung $W(x) \neq 0\ \forall\ x \in X$ ist lediglich *hinreichend* für die lineare Unabhängigkeit eines Funktionensystems $f_1, f_2, \ldots, f_n \in C^{n-1}(X)$. Wie wir in BSP. (10.2.2) festgestellt haben, sind die beiden Funktionen $f_1(x) := \sin^2 x$ und $f_2(x) := \cos^2 x$ auf der Menge $X := \mathbf{R}$ **LU**, während die WRONSKI–Determinante

$$W(x) = \begin{vmatrix} \sin^2 x & \cos^2 x \\ 2\sin x \cos x & -2\sin x \cos x \end{vmatrix} = -2\sin x \cos x = -\sin 2x$$

für jedes $x = k\pi/2,\ k \in \mathbf{Z}$, verschwindet. □

BSP. (10.2.3) Das System der Monome $1, x, x^2, \ldots, x^n,\ n \in \mathbf{N}$, ist **LU** auf jedem Intervall $X \subset \mathbf{R}$. Denn es gilt für die WRONSKI–Determinante:

$$W(x) = \begin{vmatrix} 1 & x & x^2 & x^3 & \cdots & x^n \\ & 1! & 2x & 3x^2 & \cdots & nx^{n-1} \\ & & 2! & 3!x & \cdots & n(n-1)x^{n-2} \\ & & & 3! & \cdots & n(n-1)(n-2)x^{n-3} \\ & & & & \ddots & \vdots \\ O & & & & & n! \end{vmatrix} = 0!\,1!\,2!\cdots n! \neq 0 \ \ \forall\, x \in \mathbf{R}.$$

Da das Funktionensystem der Monome in jedem der Vektorräume $C^k(X)$ enthalten ist, resultiert

$$\boxed{\dim C^k(X) = \infty.}$$

BSP. (10.2.4) Es seien $y_1, y_2 \in C^2([a,b])$ zwei Lösungen der homogenen DGl

$$\boxed{L_2 y := y'' + a_1(x)y' + a_0(x)y = 0.}$$

Dazu existiere ein Punkt $x_0 \in [a,b]$ mit $W(x_0) \neq 0$. Wir zeigen, dass dann die Funktionen y_1, y_2 auf $[a,b]$ bereits **LU** sind. In der Tat, es gilt durch Subtraktion der beiden folgenden Gleichungen:

$$\begin{array}{lcll} L_2 y_1 := y_1'' + a_1(x)y_1' + a_0(x)y_1 & = & 0 & \cdot y_2 \\ L_2 y_2 := y_2'' + a_1(x)y_2' + a_0(x)y_2 & = & 0 & \cdot y_1 \\ \hline \underbrace{y_1''y_2 - y_2''y_1}_{=-W'} + a_1(x)\underbrace{(y_1'y_2 - y_2'y_1)}_{=-W} & = & 0. & \end{array}$$

Die WRONSKI–Determinante erfüllt hier die Differentialgleichung

$$\boxed{\frac{W'}{W} = -a_1(x), \quad x \in [a,b].} \tag{2.4}$$

Wird diese Gleichung integriert, so erhält man

$$\ln\Big|\frac{W(x)}{W(x_0)}\Big| = \int\limits_{x_0}^{x} \frac{W'(t)}{W(t)}\,dt = -\int\limits_{x_0}^{x} a_1(t)\,dt \;\; \forall\, x \in [a,b].$$

Das heißt, die WRONSKI–Determinante lässt sich in Abhängigkeit des bekannten Funktionswertes $W(x_0)$ explizit in der folgenden Form angeben:

$$\boxed{W(x) = W(x_0)\,\exp\Big\{-\int\limits_{x_0}^{x} a_1(t)\,dt\Big\} \;\; \forall\, x \in [a,b].}$$

Da die Exponentialfunktion im Endlichen nicht verschwinden kann, erhält man $W(x) \neq 0\; \forall\, x \in [a,b]$ genau dann, wenn $W(x_0) \neq 0$ in einem einzigen Punkt $x_0 \in [a,b]$ gilt. Da dies nach Voraussetzung hier der Fall ist, sind die beiden Lösungen y_1, y_2 auf $[a,b]$ **LU**.

Das im obigen BSP. (10.2.4) erzielte Resultat gilt auch allgemeiner für jede lineare DGl n–ter Ordnung:

Satz 10.3 *Auf dem Intervall $[a,b]$ seien $y_1, y_2, \ldots, y_n$ Lösungen der homogenen linearen DGl*

$$L_n y := y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = 0.$$

Dann erfüllt die WRONSKI*–Determinante des Funktionensystems $y_1, y_2, \ldots, y_n \in C^n([a,b])$ die Gleichung*

$$\boxed{\frac{W'}{W} = -a_{n-1}(x), \quad x \in [a,b].} \tag{2.5}$$

Genau dann sind die Lösungen $y_1, y_2, \ldots, y_n$ auf dem Intervall $[a,b]$ **LU***, wenn für ein $x_0 \in [a,b]$ die Bedingung $W(x_0) \neq 0$ gilt.*

Begründung: Die Gleichung (2.5) wird mit völlig analoger Rechnung begründet, wie wir sie zur Herleitung von (2.4) durchgeführt haben. Durch Integration von (2.5) erhält man wiederum

$$\boxed{W(x) = W(x_0)\,\exp\Big\{-\int\limits_{x_0}^{x} a_{n-1}(t)\,dt\Big\} \;\; \forall\, x \in [a,b].} \tag{2.6}$$

Hieraus folgt dann schon die behauptete lineare Unabhängigkeit. □

Die Frage nach der Anzahl der linear unabhängigen Lösungen der homogenen DGl $L_n y = 0$ beantworten wir jetzt in dem folgenden zentralen Satz, dessen Beweis allerdings erst mit den Hilfsmitteln der Ing.–Math.III erbracht werden kann. Wir verweisen dort auf den Satz von PICARD–LINDELÖF.

Satz 10.4 *Es sei $X \subset \mathbf{R}$ ein Intervall. Sind Koeffizientenfunktionen $a_j \in C(X)$, $0 \leq j \leq n-1$ gegeben, so hat die homogene lineare DGl*

$$L_n y := y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = 0$$

auf dem Intervall X genau n linear unabhängige Lösungen $y_j \in C^n(X)$. Die **allgemeine Lösung** *der* **homogenen** *DGl $L_y = 0$ hat die Form*

$$\boxed{y_h(x) = C_1 y_1(x) + C_2 y_2(x) + \cdots + C_n y_n(x), \quad C_j \in \mathbf{K}.}$$

Das heißt, es gilt $\operatorname{Kern} L_n = \operatorname{span}\{y_1(x), y_2(x), \ldots, y_n(x)\}$. *Zu gegebener rechter Seite $f \in C(X)$ existiert stets eine partikuläre Lösung $y_p \in C^n(X)$ der inhomogenen DGl $L_n y = f$, und die* **allgemeine Lösung** *der* **inhomogenen** *DGl $L_n y = f$ ist gegeben durch*

$$\boxed{y(x) = y_p(x) + y_h(x) = y_p(x) + C_1 y_1(x) + C_2 y_2(x) + \cdots + C_n y_n(x), \quad x \in X,\ C_j \in \mathbf{K}.}$$

BSP. (10.2.5) In BSP. (10.2.1) gilt $n = 2$. Also bilden die zwei Funktionen $y_1(x) := x$ und $y_2(x) := e^{x^2}$ auf jedem der Teilintervalle $X_1 := \left(-\infty, -\frac{1}{2}\sqrt{2}\right)$, $X_2 := \left(-\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2}\right)$, $X_3 := \left(\frac{1}{2}\sqrt{2}, +\infty\right)$ bereits eine Basis für den Lösungsraum $\operatorname{Kern} L_2$ der homogenen linearen DGl

$$\boxed{L_2 y := y'' + \frac{2x(1+2x^2)}{1-2x^2}\, y' - \frac{2(1+2x^2)}{1-2x^2}\, y = 0.}$$

Die allgemeine Lösung der homogenen DGl hat somit die Form $y_h(x) = C_1 x + C_2 e^{x^2}$. Die spezielle Funktion $y_p(x) := -1/2$ ist eine partikuläre Lösung der inhomogenen DGl

$$L_2 y = \frac{1+2x^2}{1-2x^2},\ x \in X_j,$$

und somit hat diese DGl gemäß Satz 10.4 die allgemeine Lösung

$$\boxed{y(x) = C_1 x + C_2 e^{x^2} - \frac{1}{2},\ x \in X_j,\ C_1, C_2 \in \mathbf{K}.}$$

Problem: Für allgemeine lineare DGln $L_n y = f$ gibt es *keine allgemeinen analytischen Lösungsverfahren*, weder zur Bestimmung der Lösungsgesamtheit der homogenen DGl noch zum Auffinden einer partikulären Lösung der inhomogenen DGl. Im letztgenannten Fall gilt die folgende Ausnahme: Ist die allgemeine Lösung der homogenen DGl bekannt, so kann eine partikuläre Lösung der inhomogenen DGl **stets** mit dem D'Alembertschen Verfahren der **Variation der Konstanten** berechnet werden. Dieses Verfahren wird im Kurs Ing.–Math.III vorgestellt.

10.3 Das Anfangswertproblem

Wie wir in Satz 10.4 behauptet haben, setzt sich die Lösungsgesamtheit der linearen inhomogenen DGl

$$L_n y := y^{(n)} + a_{n-1}(x) y^{(n-1)} + \cdots + a_1(x) y' + a_0(x) y = f(x) \tag{3.1}$$

aus den n linear unabhängigen Lösungen y_j der homogenen DGl $L_n y = 0$ und einer partikulären Lösung y_p der inhomogenen DGl $L_n y = f$ zusammen. Die Lösungsgesamtheit $\mathcal{L}(DGl) = y_p(x) + \operatorname{span}\{y_1(x), y_2(x), \ldots, y_n(x)\}$ hat demnach n Freiheitsgrade. Diese sind durch die n frei wählbaren Konstanten $C_1, C_2, \ldots, C_n$ gegeben, über die in geeigneter Weise verfügt werden kann.

Definition 10.4 *Auf dem Intervall $X \subset \mathbf{R}$ sei $y \in C^n(X)$ eine Lösung der DGl (3.1). Wir sagen, die Funktion $y(x)$ löse an der Stelle $x_0 \in X$ ein* **Anfangswertproblem**, *wenn zu vorgegebenen Zahlen $y_0, y_1, \ldots, y_{n-1} \in \mathbf{K}$ gilt:*

$$\boxed{y(x_0) = y_0,\ y'(x_0) = y_1, \ldots, y^{(n-1)}(x_0) = y_{n-1}.} \tag{3.2}$$

Satz 10.5 *Es sei $X \subset \mathbf{R}$ ein Intervall. Sind Koeffizientenfunktionen $a_j \in C(X)$, $j = 0, 1, \ldots, n$, und eine rechte Seite $f \in C(X)$ vorgegeben, so hat das Anfangswertproblem (3.1), (3.2) für jeden festen Anfangspunkt $x_0 \in X$ und jeden Satz von Anfangsdaten $y_0, y_1, \ldots, y_{n-1} \in \mathbf{K}$ genau eine Lösung $y \in C^n(X)$.*

Begründung: Wegen Satz 10.4 existiert die allgemeine Lösung $y \in C^n(X)$ in der Form

$$y(x) = C_1 y_1(x) + C_2 y_2(x) + \cdots + C_n y_n(x) + y_p(x),\ x \in X. \tag{3.3}$$

Die WRONSKI–Determinante des Funktionensystems $y_1(x), y_2(x), \ldots, y_n(x)$ verschwindet in keinem Punkt $x_0 \in X$. Die Anfangsbedingungen (3.2) führen mit der Funktion (3.3) auf ein lineares Gleichungssystem für die Bestimmung der Konstanten $C_1, C_2, \ldots, C_n$, nämlich:

$$\begin{bmatrix} y_1(x_0) & y_2(x_0) & \cdots & y_n(x_0) \\ y_1'(x_0) & y_2'(x_0) & \cdots & y_n'(x_0) \\ y_1''(x_0) & y_2''(x_0) & \cdots & y_n''(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)}(x_0) & y_2^{(n-1)}(x_0) & \cdots & y_n^{(n-1)}(x_0) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_n \end{bmatrix} = \begin{bmatrix} y_0 - y_p(x_0) \\ y_1 - y_p'(x_0) \\ y_2 - y_p''(x_0) \\ \vdots \\ y_{n-1} - y_p^{(n-1)}(x_0) \end{bmatrix}. \tag{3.4}$$

Die Determinante der Koeffizientenmatrix ist gerade die WRONSKI–Determinante $W(x_0) \neq 0$. Deshalb besitzt das lineare Gleichungssystem (3.4) genau eine Lösung $C_1, C_2, \ldots, C_n$. □

BSP. (10.3.1) Die lineare inhomogene DGl

$$\boxed{L_3 y := x^3 y''' - 3x^2 y'' + 7xy' - 8y = x}$$

erfüllt auf jedem der Intervalle $X_- := (-\infty, 0)$ und $X_+ := (0, +\infty)$ die Voraussetzungen des Satzes 10.5. (Man dividiere dazu die gesamte Gleichung durch x^3.) Wir werden in Abschnitt 10.6 ein Verfahren zur Berechnung der Lösungsgesamtheit dieser sogenannten EULERschen Differentialgleichung kennenlernen. Nach diesem Verfahren bestimmt man die allgemeine Lösung, *zum Beispiel* auf dem Intervall X_+, in der folgenden Form:

$$y(x) = x^2\Big(C_1 + C_2 \ln x + C_3 (\ln x)^2\Big) - x,\ \ x > 0. \tag{3.5}$$

Das **Anfangswertproblem** bestehe hier in einer *geometrischen* Fragestellung. Man bestimme diejenige Lösung $y(x)$, deren Graph $G(y)$ im Punkt $(1, 0)$ eine waagerechte Wendetangente hat. *Lösung:* Die geometrische Problemstellungläuft auf die Erfüllung der Anfangsbedingungen

$$y(1) = 0,\ y'(1) = 0,\ y''(1) = 0 \tag{3.6}$$

hinaus. Zum Abgleich dieser Bedingungen verwenden wir die Lösungsdarstellung (3.5):

$$\begin{aligned} y(1) &= \Big(x^2(C_1 + C_2 \ln x + C_3 (\ln x)^2) - x\Big)\Big|_{x=1} &&= C_1 - 1 &&\stackrel{!}{=} 0, \\ y'(1) &= \Big(2x(C_1 + C_2 \ln x + C_3 (\ln x)^2) + C_2 x + 2C_3 x \ln x - 1\Big)\Big|_{x=1} &&= 2C_1 + C_2 - 1 &&\stackrel{!}{=} 0, \\ y''(1) &= \Big(2(C_1 + C_2 \ln x + C_3 (\ln x)^2) + 3C_2 + 6C_3 \ln x + 2C_3\Big)\Big|_{x=1} &&= 2C_1 + 3C_2 + 2C_3 &&\stackrel{!}{=} 0. \end{aligned}$$

Dieses lineare Gleichungssystem für die Konstanten C_j hat die eindeutig bestimmte Lösung $C_1 = 1$, $C_2 = -1$, $C_3 = 1/2$. Dementsprechend erhält man die folgende Lösung des Anfangswertproblems:

$$y(x) = x^2\Big(1 - \ln x + \frac{1}{2}(\ln x)^2\Big) - x,\ \ x > 0.$$

10.4 Die lineare homogene DGl mit konstanten Koeffizienten

Für lineare homogene Differentialgleichungen mit **konstanten Koeffizienten** kann die Lösungsgesamtheit stets mit algebraischen Mitteln bestimmt werden. Man verwendet hier mit großem Erfolg die Methode des $e^{\lambda x}$–Ansatzes. Wir stellen zunächst ein Ergebnis über die lineare Unabhängigkeit eines speziellen Funktionensystems vor.

Satz 10.6 *Gegeben seien die paarweise verschiedenen Zahlen $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbf{K}$. Dann ist das Funktionensystem $e^{\lambda_1 x}, e^{\lambda_2 x}, \ldots, e^{\lambda_n x}$ auf jedem Intervall $X \subset \mathbf{R}$ linear unabhängig. Dasselbe trifft auf das Funktionensystem $e^{\lambda x}, xe^{\lambda x}, \ldots, x^{n-1}e^{\lambda x}$ bei festem $\lambda \in \mathbf{K}$ zu.*

Begründungen: (a) Wir bilden die WRONSKI–Determinante des ersten Funktionensystems:

$$W(x) = \begin{vmatrix} e^{\lambda_1 x} & e^{\lambda_2 x} & \cdots & e^{\lambda_n x} \\ \lambda_1 e^{\lambda_1 x} & \lambda_2 e^{\lambda_2 x} & \cdots & \lambda_n e^{\lambda_n x} \\ \lambda_1^2 e^{\lambda_1 x} & \lambda_2^2 e^{\lambda_2 x} & \cdots & \lambda_n^2 e^{\lambda_n x} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{(n-1)} e^{\lambda_1 x} & \lambda_2^{(n-1)} e^{\lambda_2 x} & \cdots & \lambda_n^{(n-1)} e^{\lambda_n x} \end{vmatrix} = \exp\Big(x \sum_{j=1}^{n} \lambda_j\Big) \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{(n-1)} & \lambda_2^{(n-1)} & \cdots & \lambda_n^{(n-1)} \end{vmatrix}.$$

Die letzte Determinante heißt VANDERMONDEsche Determinante. Die folgende Identität zeigt man mit den Regeln der Determinantenrechnung:

$$V(\lambda_1, \lambda_2, \ldots, \lambda_n) := \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{(n-1)} & \lambda_2^{(n-1)} & \cdots & \lambda_n^{(n-1)} \end{vmatrix} = \prod_{1 \le j < k \le n} (\lambda_k - \lambda_j).$$

Wegen $\lambda_j \neq \lambda_k$ für $j \neq k$ gilt somit $V(\lambda_1, \lambda_2, \ldots, \lambda_n) \neq 0$, und wir erhalten die lineare Unabhängigkeit auf Grund von

$$W(x) = \exp\Big(x \sum_{j=1}^{n} \lambda_j\Big)\, V(\lambda_1, \lambda_2, \ldots, \lambda_n) \neq 0 \;\; \forall x \in \mathbf{R}.$$

(b) Wäre das Funktionensystem $e^{\lambda x}, xe^{\lambda x}, \ldots, x^{n-1}e^{\lambda x}$ **LA**, so gäbe es Zahlen $C_1, C_2, \ldots, C_n$ mit $\sum\limits_{j=0}^{n} |C_j| > 0$ und $e^{\lambda x}\Big(C_1 + C_2 x + \cdots + C_n x^{n+1}\Big) = 0$. Da die Exponentialfunktion nicht verschwinden kann, widerspräche dies der linearen Unabhängigkeit des Monomensystems $1, x, x^2, \ldots, x^{n-1}$. □

Wir betrachten jetzt für gegebene Koeffizienten $a_0, a_1, \ldots, a_n \in \mathbf{K}$ die homogene lineare gewöhnliche DGl n–ter Ordnung

$$\boxed{L_n y := \sum_{k=0}^{n} a_k y^{(k)} = a_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0, \;\; a_n \neq 0.} \tag{4.1}$$

Wir suchen Lösungen $y \in C^n(\mathbf{R})$ in der Form

$$\boxed{y(x) = e^{\lambda x}, \;\; x \in \mathbf{R}, \;\; (e^{\lambda x}\text{–Ansatz}),} \tag{4.2}$$

mit unbekanntem Exponenten $\lambda \in \mathbf{C}$. Setzen wir (4.2) in die Gleichung (4.1) ein, so resultiert:

$$L_n y = e^{\lambda x}\Big(\sum_{k=0}^{n} a_k \lambda^k\Big) =: e^{\lambda x}\, P_n(\lambda) \stackrel{!}{=} 0. \tag{4.3}$$

Definition 10.5 *Das Polynom*

$$\boxed{P_n(\lambda) := \sum_{k=0}^{n} a_k \lambda^k = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0} \tag{4.4}$$

heißt das der DGl (4.1) *zugeordnete* **charakteristische Polynom**.

Offenbar bestimmen wegen (4.3) die **Nullstellen** des charakteristischen Polynoms $P_n(\lambda)$ den im Ansatz (4.2) gesuchten Exponenten λ. Der Fundamentalsatz der Algebra garantiert nun die Existenz von genau n Nullstellen, wenn jede Nullstelle entsprechend ihrer Vielfachheit oft gezählt wird. Wir treffen hier zwei Fallunterscheidungen gemäß dem Auftreten *einfacher* oder *mehrfacher* Nullstellen.

Fall (I): Das charakteristische Polynom $P_n(\lambda)$ hat n **einfache** Nullstellen $\lambda_1, \lambda_2, \ldots, \lambda_n$. Aus dem Ansatz (4.2) resultiert in diesem Fall wegen Satz 10.6 ein Funktionensystem von n linear unabhängigen Lösungen

$$y_1(x) := e^{\lambda_1 x},\ y_2(x) := e^{\lambda_2 x}, \ldots, y_n(x) := e^{\lambda_n x},\quad x \in \mathbf{R}.$$

Diese spannen den Lösungsraum der homogene DGl $L_n y = 0$ auf. Ein solches System heißt auch ein **Fundamentalsystem** für die DGl (4.1):

$$\boxed{y_h(x) = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x} + \cdots + C_n e^{\lambda_n x},\quad x \in \mathbf{R}.}$$

BSP. (10.4.1) Man beachte im folgenden Beispiel den formalen Übergang von der Differentialgleichung zum charakteristischen Polynom:

$$\begin{array}{rrccccccccccc}
\text{DGl:} & L_4 y := & y^{(4)} & + & y''' & - & 7y'' & - & y' & + & 6y & = & 0, \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
\text{charakt. Polynom:} & P_4(\lambda) := & \lambda^4 & + & \lambda^3 & - & 7\lambda^2 & - & \lambda & + & 6 & = & 0.
\end{array}$$

Die beiden Nullstellen $\lambda_1 = 1$ und $\lambda_2 = -1$ des charakteristischen Polynoms sind leicht zu erraten. Wir spalten die Linearfaktoren $(\lambda - 1)$ und $(\lambda + 1)$ mit Hilfe des HORNER–Schemas ab:

	1	1	−7	−1	6
$\lambda = 1$	*	1	2	−5	−6
	1	2	−5	−6	0
$\lambda = -1$	*	−1	−1	6	
	1	1	−6	0	

Wir erhalten nun die Linearfaktorzerlegung

$$P_4(\lambda) = (\lambda - 1)(\lambda + 1)(\lambda^2 + \lambda - 6) = (\lambda - 1)(\lambda + 1)(\lambda - 2)(\lambda + 3),$$

und aus ihr resultiert die allgemeine Lösung

$$\boxed{y_h(x) = C_1 e^{x} + C_2 e^{-x} + C_3 e^{2x} + C_4 e^{-3x},\ x \in \mathbf{R}.}$$

BSP. (10.4.2) Wir verfahren nach dem gleichen Schema:

$$\begin{array}{rrccccccc}
\text{DGl:} & L_4 y := & 4y^{(4)} & + & 3y'' & - & y & = & 0, \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\text{charakt. Polynom:} & P_4(\lambda) := & 4\lambda^4 & + & 3\lambda^2 & - & 1 & = & 0.
\end{array}$$

Die beiden Nullstellen $\lambda_1 = i$ und $\lambda_2 = -i$ des charakteristischen Polynoms sind leicht zu erraten. Wir spalten die Linearfaktoren $(\lambda - i)$ und $(\lambda + i)$ mit Hilfe des HORNER–Schemas ab:

	4	0	3	0	-1
$\lambda = i$	$*$	$4i$	-4	$-i$	1
	4	$4i$	-1	$-i$	$\boxed{0}$
$\lambda = -i$	$*$	$-4i$	0	i	
	4	0	-1	$\boxed{0}$	

Wir erhalten nun die Linearfaktorzerlegung

$$P_4(\lambda) = (\lambda - i)(\lambda + i)(4\lambda^2 - 1) = 4(\lambda - i)(\lambda + i)(\lambda - \frac{1}{2})(\lambda + \frac{1}{2}),$$

und aus ihr resultiert die allgemeine Lösung

$$\boxed{y_h(x) = C_1 e^{ix} + C_2 e^{-ix} + C_3 e^{x/2} + C_4 e^{-x/2}, \;\; x \in \mathbf{R}.}$$

Wegen $e^{\pm ix} = \cos x \pm i \sin x$ können wir die allgemeine Lösung auch in **reeller Form** darstellen. Dazu führen wir neue Konstanten $A := C_1 + C_2,\; B := i(C_1 - C_2)$ ein:

$$\boxed{y_h(x) = A \cos x + B \sin x + C_3 e^{x/2} + C_4 e^{-x/2}, \;\; x \in \mathbf{R}.}$$

Fall (II): Das charakteristische Polynom $P_n(\lambda)$ hat **mehrfache** Nullstellen. Zunächst stellt man durch Vergleich der beiden Darstellungen (4.1) und (4.3) fest, dass die Differentialgleichung (4.1) formal in der Form

$$\boxed{L_n y \equiv P_n(D)y = 0} \tag{4.5}$$

geschrieben werden kann. Es seien jetzt $\lambda_1, \lambda_2, \ldots, \lambda_m$ die paarweise verschiedenen Nullstellen von $P_n(\lambda)$ mit Vielfachheiten $k_1, k_2, \ldots, k_m$. Dann gilt $k_1 + k_2 + \cdots + k_m = n$ sowie

$$P_n(\lambda) = a_n(\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_m)^{k_m}. \tag{4.6}$$

In genau derselben Weise kann der Differentialoperator $L_n = P_n(D)$ faktorisiert werden, so dass folgt:

$$L_n y = P_n(D)y = a_n(D - \lambda_1)^{k_1}(D - \lambda_2)^{k_2} \cdots (D - \lambda_m)^{k_m} y = 0. \tag{4.7}$$

Hierbei kommt es wegen der Vertauschungsregel $(D - \lambda_p)^{k_p}(D - \lambda_q)^{k_q} = (D - \lambda_q)^{k_q}(D - \lambda_p)^{k_p}$ auf die Reihenfolge der Faktoren **nicht** an. Offensichtlich sind alle Lösungen $y \in C^n(\mathbf{R})$ der DGl

$$\boxed{(D - \lambda_q)^{k_q} y = 0} \tag{4.8}$$

auch Lösungen der Differentialgleichung (4.7). Zur Lösung der DGl (4.8) setzen wir an:

$$\boxed{y(x) = e^{\lambda_q x} \cdot \varphi(x), \;\; x \in \mathbf{R}.}$$

Es folgt unter Verwendung der Identität $(D - \lambda_q)\big(e^{\lambda_q x} \cdot \varphi(x)\big) = e^{\lambda_q x} \cdot \varphi'(x)$ durch Einsetzen in die Gleichung (4.8):

$$(D - \lambda_q)^{k_q} y(x) = e^{\lambda_q x} \cdot \varphi^{(k_q)}(x) \stackrel{!}{=} 0.$$

Da die Exponentialfunktion im Endlichen nicht verschwindet, muss folglich $\varphi^{(k_q)}(x) = 0$ gelten, und dies führt auf die polynomiale Lösung

$$\varphi(x) = Q_{k_q - 1}(x) = C_1 + C_2 x + \cdots + C_{k_q} x^{k_q - 1}, \;\; x \in \mathbf{R}.$$

Wie wir in Satz 10.6 gezeigt haben, ist das Funktionensystem $e^{\lambda_q x}, xe^{\lambda_q x}, \ldots, x^{k_q-1}e^{\lambda_q x}$ **LU**, so dass dieses System den k_q–dimensionalen Lösungsraum der DGl (4.8) aufspannt. In ganz analoger Weise verfährt man mit den weiteren Wurzeln des charakteristischen Polynoms. Wir fassen zusammen zu folgender Aussage:

Satz 10.7 *Es seien $\lambda_1, \lambda_2, \ldots, \lambda_m$, $m \le n$, die paarweise verschiedenen Nullstellen des charakteristischen Polynoms $P_n(\lambda)$ mit Vielfachheiten $k_1, k_2, \ldots, k_m$. Ist λ_q, $1 \le q \le m$, eine* **einfache** *Wurzel, so ist durch sie eine Lösung*

$$\boxed{y_q(x) = A_q e^{\lambda_q x}, \quad x \in \mathbf{R},}$$

*der homogenen DGl $L_n y = 0$ bestimmt. Ist λ_q, $1 \le q \le m$, eine k_q–***fache** *Wurzel, so ist durch sie eine Lösung*

$$\boxed{y_q(x) = \left(C_{q1} + C_{q2}x + \cdots + C_{qk_q}x^{k_q-1}\right) e^{\lambda_q x}, \quad x \in \mathbf{R},}$$

der homogenen DGl $L_n y = 0$ bestimmt. Dabei sind A_q bzw. C_{qj} frei wählbare (komplexe) Konstanten. Die allgemeine Lösung der homogenen DGl $L_n y = 0$ ist dann in der Form

$$\boxed{y_h(x) = y_1(x) + y_2(x) + \cdots + y_m(x)}$$

gegeben.

BSP. (10.4.3) Wir betrachten die folgende Differentialgleichung:

$$\begin{array}{rrccccccccccc}
\text{DGl:} & L_5 y := & y^{(5)} & - & y^{(4)} & + & 2y''' & - & 2y'' & + & y' & - & y & = & 0, \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
\text{charakt. Polynom:} & P_5(\lambda) := & \lambda^5 & - & \lambda^4 & + & 2\lambda^3 & - & 2\lambda^2 & + & \lambda & - & 1 & = & 0.
\end{array}$$

Man errät hier die Nullstelle $\lambda_1 = 1$, und nach Abspalten des Linearfaktors $(\lambda - 1)$ verbleibt ein Restpolynom $\lambda^4 + 2\lambda^2 + 1 = (\lambda^2+1)^2 = (\lambda+i)^2(\lambda-i)^2$. Es liegen somit die Nullstellen $\lambda_1 = 1$ (einfach), $\lambda_2 = i$ (doppelt) und $\lambda_3 = -i$ (doppelt) vor. Gemäß Satz 10.7 hat die homogene DGl $L_5 y = 0$ die allgemeine Lösung

$$\boxed{y_h(x) = C_1 e^x + (C_2 + C_3 x)e^{ix} + (C_4 + C_5 x)e^{-ix}, \quad x \in \mathbf{R}.}$$

Man erkennt an den bisherigen Erörterungen, dass die gesamte Problematik bei linearen gewöhnlichen DGln im Auffinden der Nullstellen des charakteristischen Polynoms $P_n(\lambda)$ besteht, also ein *rein algebraisches Problem* ist.

Fall (III): **Reelle Lösungsgesamtheit.** Hat der lineare Differentialoperator $L_n := P_n(D) = \sum\limits_{k=0}^{n} a_k D^k$, $a_n \neq 0$, ausschließlich **reelle Koeffizienten** $a_k \in \mathbf{R}$, so treten komplexe Nullstellen des charakteristischen Polynoms $P_n(\lambda)$ stets nur als *konjugiert komplexe* Paare auf: Mit $\lambda_q := \alpha_q + i\,\beta_q$, $\beta_q \neq 0$, ist auch $\overline{\lambda}_q = \alpha_q - i\,\beta_q$ Nullstelle. Hat λ_q die Vielfachheit k_q, so sind die komplexwertigen Funktionen

$$x^j\, e^{\lambda_q\, x}, \quad x^j\, e^{\overline{\lambda}_q\, x}, \quad j = 0, 1, \ldots, k_q - 1,$$

Lösungen der homogenen DGl $L_n y = 0$. Durch Zerlegung in Real– und Imaginärteil erhält man dazu äquivalente Paare **reeller** Funktionen, nämlich

$$x^j\, e^{\alpha_q\, x} \cdot \cos \beta_q\, x, \quad x^j\, e^{\alpha_q\, x} \cdot \sin \beta_q\, x, \quad j = 0, 1, \ldots, k_q - 1.$$

Wir folgern hieraus:

Satz 10.8 *Die Koeffizienten a_k der linearen homogenen DGl*

$$\boxed{L_n y := P_n(D)y = \sum_{k=0}^{n} a_k D^k y = 0,\ a_n \neq 0,}$$

seien ausschließlich **reell**. *Dann wird die Lösungsgesamtheit dieser DGl ausschließlich von reellen Funktionen aufgespannt, und zwar von den Funktionen*

$$\boxed{e^{rx},\ x\,e^{rx}, \ldots, x^{k-1}\,e^{rx},\ \ x \in \mathbf{R},}$$

falls $\lambda = r \in \mathbf{R}$ eine k**–fache reelle** *Nullstelle des charakteristischen Polynoms $P_n(\lambda)$ ist, beziehungsweise von den Funktionen*

$$\boxed{\left.\begin{array}{l} e^{\alpha x}\cdot\cos\beta x,\ x\,e^{\alpha x}\cdot\cos\beta x, \ldots, x^{k-1}\,e^{\alpha x}\cdot\cos\beta x, \\ e^{\alpha x}\cdot\sin\beta x,\ x\,e^{\alpha x}\cdot\sin\beta x, \ldots, x^{k-1}\,e^{\alpha x}\cdot\sin\beta x, \end{array}\right\}\ x \in \mathbf{R},}$$

falls $\lambda = \alpha \pm i\,\beta$ ein Paar **konjugiert komplexer Nullstellen der Vielfachheit** k *ist.*

BSP. (10.4.4) Wir betrachten die folgende lineare homogene Differentialgleichung:

$$\begin{array}{rrccccccccccc} \text{DGl:} & P_4(D)y := & \big(D^4 & - & 4D^3 & + & 15D^2 & - & 22D & + & 10\big)y & = & 0, \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\ \text{charakt. Polynom:} & P_4(\lambda) := & \lambda^4 & - & 4\lambda^3 & + & 15\lambda^2 & - & 22\lambda & + & 10 & = & 0. \end{array}$$

Man kann die doppelte Nullstelle $\lambda_1 = 1$ des charakteristischen Polynoms leicht erraten. Wir spalten den Linearfaktor $(\lambda-1)^2$ mit Hilfe des HORNER–Schemas ab:

	1	−4	15	−22	10
$\lambda = 1$	*	1	−3	12	−10
	1	−3	12	−10	0
$\lambda = 1$	*	1	−2	10	
	1	−2	10	0	

Wir erhalten nun die Linearfaktorzerlegung

$$P_4(\lambda) = (\lambda-1)^2(\lambda^2 - 2\lambda + 10) = (\lambda-1)^2(\lambda - 1 - 3i)(\lambda - 1 + 3i),$$

und aus ihr resultiert die allgemeine Lösung in der *komplexen* Form

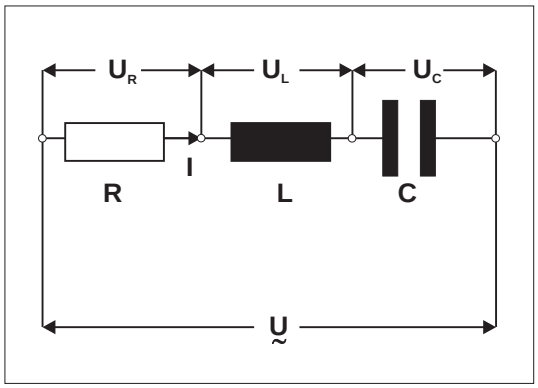
$$y_h(x) = (C_1 + C_2 x)e^x + C_3 e^{(1+3i)x} + C_4 e^{(1-3i)x}.$$

Hierzu äquivalent ist die *reelle* Form

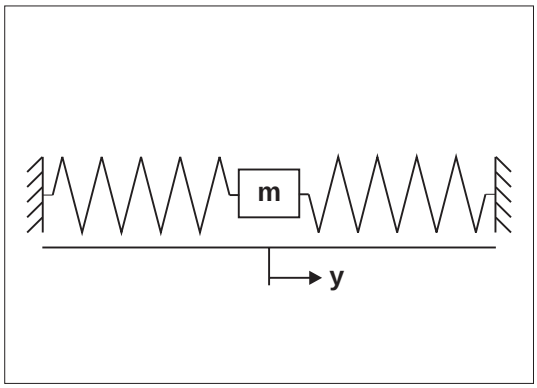
$$\boxed{y_h(x) = e^x\left(C_1 + C_2 x + \tilde{C}_3\cos 3x + \tilde{C}_4 \sin 3x\right).}$$

BSP. (10.4.5) **Die Schwingungsdifferentialgleichung.** Ein RCL–Kreis ist ein elektrischer Schaltkreis, bestehend aus der Hintereinanderschaltung eines OHMschen Widerstandes R, einer Induktivität L und eines Kondensators C. Dieser Schaltkreis sei an eine Wechselspannung $U = U(t)$ angeschlossen. Gemäß den KIRCHHOFFschen Gesetzen der Elektrodynamik gelten die folgenden Gesetze für die zeitliche Veränderung der Teilspannungen $U_R(t), U_L(t), U_C(t)$ und des Stromes $I(t)$, wenn auf den Kondensator die Ladung $Q(t)$ aufgebracht wird:

$$\text{(i)}\ \ U = U_R + U_L + U_C, \quad \text{(ii)}\ \ U_R = R\cdot I, \quad U_L = L\frac{dI}{dt}, \quad U_C = \frac{Q}{C}.$$



RCL–Kreis an einer Wechselspannung



Der Feder–Masse–Schwinger als mechanisches Analogon

Unter Berücksichtigung der Relation $I(t) = \frac{dQ}{dt}$ gelangt man durch Differenzieren der Gleichung (i) und Einsetzen von (ii) zu einer linearen gewöhnlichen Differentialgleichung mit konstanten Koeffizienten, nämlich

$$L\,\frac{d^2I}{dt^2} + R\,\frac{dI}{dt} + \frac{1}{C}\,I = \frac{dU}{dt}. \tag{4.9}$$

Bei gegebener Spannung $U = U(t)$ ist dies die Bestimmungsgleichung für die Stromstärke $I = I(t)$. Ist $U = const$ eine *Gleichspannung*, so resultiert die lineare homogene DGl

$$\boxed{\Big(L\,D^2 + R\,D + \frac{1}{C}\Big)\,I := L\,\frac{d^2I}{dt^2} + R\,\frac{dI}{dt} + \frac{1}{C}\,I = 0.} \tag{4.10}$$

Nach Division dieser DGl durch $L \neq 0$ und Verwendung neuer Bezeichnungen $y(t) := I(t),\ \rho := R/2L,\ \omega_0^2 := 1/LC$, gelangt man zu einer linearen homogenen DGl vom Typ

$$\boxed{\ddot{y} + 2\rho\,\dot{y} + \omega_0^2\,y = 0.} \tag{4.11}$$

Definition 10.6 *Die lineare homogene DGl* (4.11) *heiße die* **Differentialgleichung der freien Schwingung**. *Die Größe* $\rho > 0$ *heiße* **Dämpfungskonstante**, *und die Größe* ω_0 *heiße die* **Kenn–Frequenz** *oder* **Eigenfrequenz** *der Schwingung.*

Bemerkung 10.3 Die DGl (4.11) beschreibt in gleicher Weise die Bewegung einer Masse m zwischen zwei Federn, deren Rückstellkraft $K = -k^2\,y$ proportional zur zeitlichen Auslenkung $y(t)$ ist. Dabei wirke eine geschwindigkeitsproportionale Reibung mit der Reibungskraft $R = -r\,\dot{y}$. Dieses Gebilde nennt man einen mechanischen Feder–Masse–Schwinger. Aus dem NEWTONschen Kraftgesetz folgt:

$$m\ddot{y} = K + R = -(r\,\dot{y} + k^2\,y) \quad \text{oder} \quad \ddot{y} + \frac{r}{m}\,\dot{y} + \frac{k^2}{m}\,y = 0.$$

Man erhält hieraus mit den Größen $\rho := r/2m,\ \omega_0^2 := k^2/m$ wieder die DGl (4.11). □

Zur Lösung der DGl (4.11) stellen wir ihr charakteristisches Polynom $P_2(\lambda) := \lambda^2 + 2\rho\,\lambda + \omega_0^2$ auf, dessen Nullstellen durch

$$\lambda_\pm := -\rho \pm \sqrt{\rho^2 - \omega_0^2}$$

gegeben sind. Die allgemeine Lösung hat somit die Form

$$\boxed{y_h(t) = \begin{cases} e^{-\rho t}\Big(A\,e^{t\sqrt{\rho^2-\omega_0^2}} + B\,e^{-t\sqrt{\rho^2-\omega_0^2}}\Big) & :\ \rho > \omega_0, \\ e^{-\rho t}\Big(A + B\,t\Big) & :\ \rho = \omega_0, \\ e^{-\rho t}\Big(A\,\cos t\sqrt{\omega_0^2-\rho^2} + B\,\sin t\sqrt{\omega_0^2-\rho^2}\Big) & :\ \rho < \omega_0. \end{cases}} \tag{4.12}$$

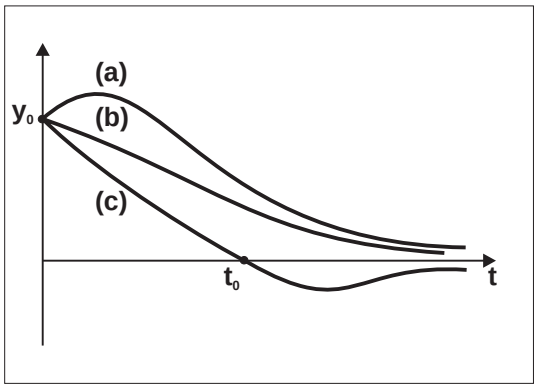
Durch die Vorgaben von **Anfangsbedingungen** $y(0) = y_0$, $\dot{y}(0) = y_1$ lassen sich die Konstanten A, B in eindeutiger Weise bestimmen. Im Fall des RCL–Kreises müssten also $I(0)$ und $\dot{I}(0)$ vorgegeben werden. Die Lösungen (4.12) können in diesem Fall interpretiert werden als das *Nachschwingen* des Schaltkreises, wenn man zum Zeitpunkt $t = 0$ die Spannung U abschaltet. Entsprechend den drei verschiedenenen Lösungsformen (4.12) klassifizieren wir das zeitliche Verhalten der freien Schwingung $y(t)$ in der folgenden Weise:

1.Fall: **Aperiodischer Kriechfall.** Dieser Fall liegt bei der Parameterkonstellation $\rho > \omega_0$ vor. Er entspricht dem Auftreten einer **starken Dämpfung**, also im RCL–Kreis der Vorgabe eines großen OHMschen Widerstandes R. Die Lösung $y_h(t) = Ae^{t\,\lambda_+} + Be^{t\,\lambda_-}$ hat die beiden *negativen* Exponenten

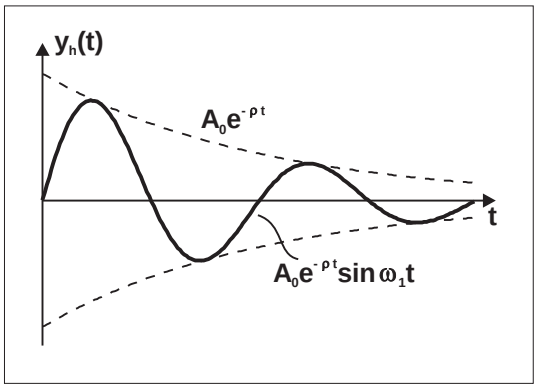
$$\boxed{\lambda_\pm := -\rho \pm \omega_2 < 0, \quad \omega_2 := \sqrt{\rho^2 - \omega_0^2}.}$$

Es gilt $\lim\limits_{t\to+\infty} y_h(t) = 0$ für jede Wahl der Konstanten A, B. Gibt man die Anfangsbedingungen $y(0) = y_0$ und $\dot{y}(0) = y_1$ vor, so hat die Lösung die beiden folgenden äquivalenten Darstellungen:

$$\boxed{y_h(t) = \begin{cases} e^{-\rho t}\Big(A\,e^{t\omega_2} + B\,e^{-t\omega_2}\Big) & \text{mit } A := \dfrac{y_0(\rho+\omega_2)+y_1}{2\omega_2},\ B := -\dfrac{y_0(\rho-\omega_2)+y_1}{2\omega_2}, \\ e^{-\rho t}\Big(D\,\cosh t\omega_2 + E\,\sinh t\omega_2\Big) & \text{mit } D := y_0,\ E := \dfrac{\rho\,y_0+y_1}{\omega_2}. \end{cases}}$$



Der aperiodische Kriechfall



Die gedämpfte Schwingung

Ist wenigstens eine der Konstanten A, B von Null verschieden, so tritt eine Nullstelle $y_h(t_0) = 0$ nur dann auf, wenn $A\,e^{2t_0\omega_2} = -B$ gilt. Im Endlichen kann dies für höchstens ein t_0 eintreten. Durch Diskussion der Parameter A und B erhält man den oben skizzierten zeitlichen Lösungsverlauf mit den folgenden Spezifikationen:

$$\begin{aligned} &\text{Kurve (a)} \ : \ y_1 > 0, \\ &\text{Kurve (b)} \ : \ y_1 \le 0 \ \text{ und } \ -y_1 < y_0(\rho+\omega_2), \\ &\text{Kurve (c)} \ : \ y_1 \le 0 \ \text{ und } \ -y_1 > y_0(\rho+\omega_2). \end{aligned}$$

2.Fall: **Aperiodischer Grenzfall.** Dieser liegt im Fall $\rho = \omega_0$ vor. Im RCL–Kreis muss $R^2 = 4L/C$ gelten. Die Lösung hat die Form $y_h(t) = (A + Bt)\,e^{-\rho t}$, und es gilt wiederum ein zeitliches Abklingen $\lim\limits_{t\to+\infty} y_h(t) = 0$. Mit den Anfangsbedingungen $y(0) = y_0$, $\dot{y}(0) = y_1$ erhält die Lösung die Gestalt

$$\boxed{y_h(t) = \Big(y_0 + (\rho\,y_0 + y_1)t\Big)\,e^{-\rho t}.}$$

Es liegt ein ähnlicher zeitlicher Lösungsverlauf wie im aperiodischen Kriechfall vor, jedoch mit den folgenden Spezifikationen:

$$\begin{aligned} &\text{Kurve (a)} \ : \ y_1 > 0, \\ &\text{Kurve (b)} \ : \ y_1 \le 0 \ \text{ und } \ -y_1 < \rho\,y_0, \\ &\text{Kurve (c)} \ : \ y_1 \le 0 \ \text{ und } \ -y_1 > \rho\,y_0. \end{aligned}$$

3.Fall: **Gedämpfte Schwingung.** Dieser Fall liegt bei der Parameterkonstellation $0 < \rho < \omega_0$ vor. Er entspricht dem Auftreten einer **kleinen Dämpfung**, also im RCL–Kreis der Vorgabe eines kleinen OHMschen

Widerstandes R. Die Lösung $y_h(t) = Ae^{t\,\lambda_+} + Be^{t\,\lambda_-}$ hat die beiden *konjugiert komplexen* Exponenten

$$\boxed{\lambda_\pm := -\rho \pm i\omega_1, \quad \omega_1 := \sqrt{\omega_0^2 - \rho^2} > 0.}$$

Schreibt man die Anfangsbedingungen $y(0) = y_0$ und $\dot{y}(0) = y_1$ vor, so gibt es für die Lösung die beiden folgenden äquivalenten Darstellungen:

$$\boxed{y_h(t) = \begin{cases} e^{-\rho t}\Big(A\,\cos\omega_1 t + B\,\sin\omega_1 t\Big) & \text{mit } A := y_0,\ B := \dfrac{\rho\, y_0 + y_1}{\omega_1}, \\ A_0\, e^{-\rho t}\sin(\omega_1 t + \varphi) & \text{mit } A_0 := \sqrt{A^2 + B^2},\ \varphi := \operatorname{arc\,tan}_H \dfrac{A}{B}. \end{cases}} \tag{4.13}$$

Die Konstante A_0 heißt die **Amplitude** und der Winkel $\varphi \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$ die **Nullphase** der gedämpften Schwingung. Für $\rho = 0$ (in diesem Fall liegt ein reiner LC–Kreis ohne OHMschen Widerstand vor), schwingt der Schaltkreis ungedämpft in seinem angeregten Zustand mit der Eigenfrequenz $\omega_0 = 1/\sqrt{LC}$:

$$\boxed{y_h(t) = y_0\,\cos\omega_0 t + \frac{y_1}{\omega_0}\,\sin\omega_0 t.}$$

Der Vorgang heißt **ungedämpfte freie Schwingung**. Gilt jedoch $\rho > 0$, so haben wir wiederum ein zeitliches Abklingverhalten $\lim\limits_{t\to+\infty} y_h(t) = 0$. Die Lösung $y_h(t)$ schwingt mit zeitlich abnehmender Amplitude in der Frequenz $\omega_1 = \sqrt{\omega_0^2 - \rho^2}$. Der Vorgang heißt **gedämpfte Schwingung**.

Bemerkung 10.4 Der aperiodische Grenzfall (Fall 2) kann auch als Grenzwert $\omega_1 \to 0$ aus der gedämpften Schwingung (4.13) abgeleitet werden. Mit der Regel von L'HOSPITAL erhält man nämlich: □

$$\begin{aligned} \lim_{\omega_1\to 0} &= e^{-\rho t}\lim_{\omega_1\to 0}\Big(y_0\cos\omega_1 t + (\rho\, y_0 + y_1)\frac{\sin\omega_1 t}{\omega_1}\Big) = e^{-\rho t}\Big(y_0 + (\rho\, y_0 + y_1)\lim_{\omega_1\to 0}\frac{t\cos\omega_1 t}{1}\Big) \\ &= e^{-\rho t}\Big(y_0 + (\rho\, y_0 + y_1)\,t\Big). \end{aligned}$$

Wir diskutieren jetzt die Frage, welchen Einfluss eine am RCL–Kreis anliegende Wechselspannung $U = U(t)$ auf das Lösungsverhalten der Differentialgleichung (4.9) hat. Wir nehmen zum Beispiel eine cosinusförmige Wechselspannung $U(t) := U_0\,\cos\omega t$ an. Dann tritt an die Stelle der Gleichung (4.9) die inhomogene DGl

$$L\,\frac{d^2 I}{dt^2} + R\,\frac{dI}{dt} + \frac{1}{C}\,I = -U_0\omega\,\sin\omega t,$$

beziehungsweise an die Stelle der homogenen Schwingungsdifferentialgleichung (4.11) die inhomogene DGl

$$\boxed{\ddot{y} + 2\rho\,\dot{y} + \omega_0^2\,y = p_0\,\sin\omega t} \tag{4.14}$$

(mit $p_0 := -U_0\omega/L$ im Fall des RCL–Kreises). Nach unseren bisherigen Erkenntnissen haben wir nur noch eine *partikuläre Lösung* der inhomogenen DGl zu bestimmen, da wir die Lösungsgesamtheit der homogenen DGl bereits ausführlich diskutiert haben. Es ist physikalisch einleuchtend, dass die periodische Wechselspannung $U(t) = U_0\,\cos\omega t$ einen periodischen Stromverlauf $I_p(t)$ mit derselben Periode ω erzwingen wird. Diese Überlegung rechtfertigt einen **Ansatz von der Form der rechten Seite** der DGl (4.14), wobei eine Linearkombination der beiden ω–periodischen Funktionen $\sin\omega t$ und $\cos\omega t$ als Ansatzfunktion sicherlich nicht falsch ist:

$$\begin{array}{lcrcrl} y_p(t) & = & A\cos\omega t & + & B\sin\omega t & \cdot\omega_0^2 \\ \dot{y}_p(t) & = & -A\omega\sin\omega t & + & B\omega\cos\omega t & \cdot 2\rho \\ \ddot{y}_p(t) & = & -A\omega^2\cos\omega t & - & B\omega^2\sin\omega t & \cdot 1 \\ \hline p_0\,\sin\omega t & \stackrel{!}{=} & \Big(A(\omega_0^2-\omega^2) + 2\rho\omega B\Big)\cos\omega t & + & \Big(B(\omega_0^2-\omega^2) - 2\rho\omega A\Big)\sin\omega t. & \end{array} \tag{4.15}$$

Durch Koeffizientenvergleich erhält man das folgende lineare Gleichungssystem für die Bestimmung der Konstanten A und B:

$$\begin{array}{rcrcl} (\omega_0^2-\omega^2)A & + & 2\rho\omega B & = & 0, \\ -2\rho\omega A & + & (\omega_0^2-\omega^2)B & = & p_0. \end{array} \tag{4.16}$$

Die Determinante $D := 4\rho^2\omega^2 + (\omega_0^2 - \omega^2)^2$ der zugeordneten Koeffizientenmatrix kann im Fall $\rho > 0$ für keine Erregerfrequenz $\omega > 0$ verschwinden. Deshalb existieren in diesem Fall stets eindeutige Lösungen

$$A = -\frac{1}{D}\,2\rho\omega p_0, \quad B = \frac{1}{D}\,(\omega_0^2 - \omega^2)p_0.$$

Der Ansatz (4.15) liefert somit eine wohlbestimmte partikuläre Lösung $y_p(t)$ der inhomogenen DGl (4.14):

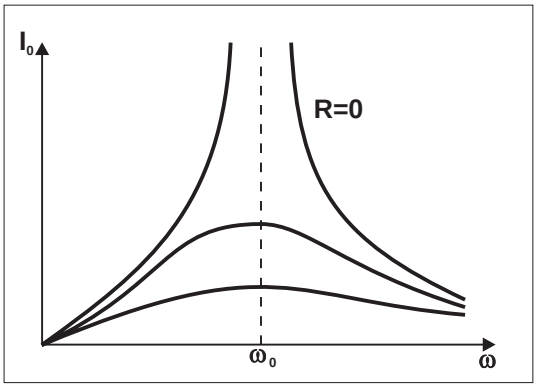
$$\boxed{y_p(t) = \frac{p_0}{D}\Big((\omega_0^2 - \omega^2)\sin\omega t - 2\rho\omega\cos\omega t\Big).}$$

Mit den Parametern des RCL–Kreises hat also die periodische Wechselspannung $U(t) = U_0\cos\omega t$ einen periodischen Stromfluss $I_p(t)$ erzwungen, wobei gilt

$$\boxed{I_p(t) = \frac{p_0}{D}\Big((\omega_0^2 - \omega^2)\sin\omega t - 2\rho\omega\cos\omega t\Big) = -\frac{p_0}{\sqrt{D}}\cos(\omega t + \varphi), \;\; \varphi := \text{arc tan}_H\,\frac{\omega_0^2 - \omega^2}{2\rho\omega}.}$$

Der Strom $I_p(t)$ schwingt mit derselben Frequenz ω wie die Erregerspannung, jedoch mit einer **Phasenverschiebung** φ. Die Amplitude des Stroms ist

$$\boxed{I_0 = -\frac{p_0}{\sqrt{D}} = \frac{U_0\omega}{L\sqrt{4\rho^2\omega^2 + (\omega_0^2 - \omega^2)^2}} = \frac{U_0}{\sqrt{R^2 + \Big(\omega L - \frac{1}{\omega C}\Big)^2}}.}$$



Das Amplitudenportrait in Abhängigkeit vom OHMschen Widerstand R

Man erkennt sofort, dass die Amplitude I_0 ihren Maximalwert bei $\omega = \omega_0 = 1/\sqrt{LC}$ annimmt. Dort gilt $(I_0)_{\max} = U_0/R$. Wir haben ferner die Grenzwerte $\lim\limits_{\omega\to 0+} I_0 = 0 = \lim\limits_{\omega\to+\infty} I_0$. Für kleine Dämpfung $0 < R \ll 1$ wächst die Amplitude I_0 in der Nähe der Kenn–Frequenz ω_0 sehr stark an. Diesen Sachverhalt bezeichnet man als **Resonanzphänomen** oder kurz als **Resonanz**. Im dämpfungsfreien Fall $R = 0$ wächst I_0 unbeschränkt, wenn ω die Kenn–Frequenz ω_0 erreicht. Es kommt zur sogenannten **Resonanzkatastrophe.**
Für $\rho = 0$ und $\omega = \omega_0$ führt der Ansatz (4.15) auf keine bestimmbaren Koeffizienten A und B. Denn wegen $D = 0$ ist das lineare Gleichungssystem (4.16) nicht mehr beständig lösbar. Der tiefere Grund ist in der Tatsache zu sehen, dass in diesem Fall auf der rechten Seite der inhomogenen DGl

$$L_2 y := \ddot{y} + \omega_0^2\, y = p_0 \sin\omega_0 t$$

eine *Lösung* $y_h(t) := p_0\sin\omega_0 t$ *der homogenen DGl* $L_2 y = 0$ steht. Es ist klar, dass Ansätze in der Form $A\cos\omega_0 t$ und $B\sin\omega_0 t$ lediglich die homogene DGl $L_2 y = 0$ erfüllen. Man wird jedoch durch einen sogenannten **Resonanzansatz** zum Erfolg geführt:

$y_p(t)$	$=$	$t(A\cos\omega_0 t + B\sin\omega_0 t)$			$\cdot\omega_0^2$
$\dot{y}_p(t)$	$=$	$t\omega_0(-A\sin\omega_0 t + B\cos\omega_0 t)$	$+$	$A\cos\omega_0 t + B\sin\omega_0 t$	$\cdot 0$
$\ddot{y}_p(t)$	$=$	$-t\omega_0^2(A\cos\omega_0 t + B\sin\omega_0 t)$	$-$	$2A\omega_0\sin\omega_0 t + 2B\omega_0\cos\omega_0 t$	$\cdot 1$
$p_0\,\sin\omega_0 t$	$\stackrel{!}{=}$		$-$	$2A\omega_0\sin\omega_0 t + 2B\omega_0\cos\omega_0 t.$	

Durch Koeffizientenvergleich erhalten wir

$$A = -\frac{p_0}{2\omega_0} = \frac{U_0}{2L}, \quad B = 0,$$

und hieraus resultiert die sogenannte **Resonanzlösung**

$$\boxed{y_p(t) = -\frac{p_0 t}{2\omega_0}\cos\omega_0 t \quad \text{bzw.} \quad I_p(t) = \frac{U_0 t}{2L}\cos\Big(\frac{t}{\sqrt{LC}}\Big).}$$

Im Resonanzfall erhält man eine mit der Zeit linear anwachsende Amplitude $I_0 = U_0 t/2L$.

10.5 Die lineare DGl mit konstanten Koeffizienten und speziellen Inhomogenitäten

Wir betrachten in diesem Abschnitt die **inhomogene** lineare gewöhnliche DGl mit konstanten Koeffizienten

$$\boxed{L_n y := P_n(D)y = \sum_{k=0}^{n} a_k D^k y = R(x), \quad a_n \neq 0,\ x \in \mathbf{R},} \tag{5.1}$$

wo wir auf der rechten Seite nur spezielle Funktionen $R(x)$ zulassen wollen. Und zwar soll $R(x)$ zu einem der folgenden drei Typen gehören:

Typ (I)	$R_I(x) := Q_m(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0,$
Typ (II)	$R_{II}(x) := e^{\alpha x}\cdot Q_m(x),\ \alpha \in \mathbf{R},$
Typ (III)	$R_{III}(x) := Q_m(x)\cdot e^{\alpha x}\cos\beta x$ oder $Q_m(x)\cdot e^{\alpha x}\sin\beta x,\ \alpha,\beta \in \mathbf{R}.$

Wir hatten in Satz 10.8 festgestellt, dass Funktionen vom Typ R_I, R_{II}, R_{III} den Lösungsraum der *homogenen* DGl $L_n y = 0$ aufspannen. Andererseits beobachtet man, dass sich die Funktionen R_I, R_{II}, R_{III} bei Anwendung des Differentialoperators $P_n(D)$ *reproduzieren*. Dies gibt Anlass zur Hoffnung, dass partikuläre Lösungen der *inhomogenen* DGl $P_n(D)y = R(x)$ bei Vorgabe einer der Inhomogenitäten R_I, R_{II}, R_{III} durch einen Ansatz *vom selben Typ* bestimmt werden können. Daher wird für den jeweiligen Typ ein **Direktansatz** von derselben Form, jedoch mit unbestimmten Koeffizienten, nahegelegt. Durch Einsetzen in die DGl sollten sich genügend viele Bedingungen für die Bestimmung der freien Koeffizienten ergeben. Problematisch ist lediglich der **Resonanzfall**, wie das BSP. (10.4.5) der Schwingungsdifferentialgleichung in Abschnitt 10.4 gezeigt hat. Wir präzisieren hier:

Definition 10.7 *Eine Inhomogenität vom Typ*

$$\boxed{R(x) := Q_m(x)\cdot e^{\alpha x}\cos\beta x \quad \textit{oder} \quad R(x) := Q_m(x)\cdot e^{\alpha x}\sin\beta x,\ \alpha,\beta \in \mathbf{R},}$$

erzeuge eine k–fache Resonanz *in der DGl $P_n(D)y = R(x)$, wenn $\lambda_0 := \alpha + i\,\beta \in \mathbf{C}$ eine* **k–fache Nullstelle** *des charakteristischen Polynoms $P_n(\lambda)$ ist, das heißt, wenn gilt:*

$$P_n(\lambda) = (\lambda - \lambda_0)^k\,\tilde{P}(\lambda) \ \ \textit{mit}\ \ \tilde{P}(\lambda_0) \neq 0.$$

Beachte: Mit dieser Definition sind auch die Inhomogenitäten vom Typ (I) ($\alpha = 0 = \beta$) und vom Typ (II) ($\beta = 0$) erfasst.

Satz 10.9 *Sind Inhomogenitäten vom Typ*

$$\boxed{R(x) := Q_m(x) \cdot e^{\alpha x} \cos \beta x \quad \text{bzw.} \quad R(x) := Q_m(x) \cdot e^{\alpha x} \sin \beta x,\ \alpha, \beta \in \mathbf{R},}$$

gegeben, so führen die folgenden **Direktansätze** *immer zu einer partikulären Lösung $y_p(x)$ der inhomogenen DGl $P_n(D)y = R(x)$:*

(a) *Falls $\lambda_0 := \alpha + i\,\beta \in \mathbf{C}$* **keine Nullstelle** *des charakteristischen Polynoms $P_n(\lambda)$ ist (d.h. $P_n(\lambda_0) \neq 0$), so setze man*

$$\boxed{\begin{array}{rcl} S_m(x) & := & A_0 + A_1 x + A_2 x^2 + \cdots + A_m x^m, \\ T_m(x) & := & B_0 + B_1 x + B_2 x^2 + \cdots + B_m x^m, \\ y_p(x) & := & e^{\alpha x} \Big(S_m(x) \cdot \cos \beta x + T_m(x) \cdot \sin \beta x \Big). \end{array}} \tag{5.2}$$

Die Koeffizienten A_j, B_j bestimmt man durch Koeffizientenvergleich nach Einsetzen des Ansatzes (5.2) in die DGl (5.1).

(b) *Liegt der* **Resonanzfall** *vor, das heißt, ist $\lambda_0 := \alpha + i\,\beta \in \mathbf{C}$ eine k–***fache Nullstelle** *des charakteristischen Polynoms $P_n(\lambda)$, so setze man nun:*

$$\boxed{y_p(x) := x^k\, e^{\alpha x} \Big(S_m(x) \cdot \cos \beta x + T_m(x) \cdot \sin \beta x \Big).} \tag{5.3}$$

Man verifiziert mit elementarer Rechnung, dass die Ansätze (5.2) und (5.3) tatsächlich eine partikuläre Lösung der inhomogenen DGl $P_n(D)y = R(x)$ liefern.

Bemerkung 10.5 (a) Es ist zu beachten, dass in den Ansätzen (5.2) und (5.3) **stets sowohl die Sinus– als auch die Cosinus–Terme** mitzuführen sind, selbst wenn in der Inhomogenität $R(x)$ nur einer dieser beiden Terme auftritt.

(b) Die lineare inhomogene DGl (5.1) unterliegt dem **Superpositionsprinzip**. Setzt sich die Inhomogenität $R(x)$ aus einer endlichen Linearkombination von Funktionen des Typs R_I, R_{II}, R_{III} zusammen, zum Beispiel in der Form

$$R(x) = C_1 R_I(x) + C_2 R_{II}(x) + C_3 R_{III}(x),$$

so gewinnt man eine partikuläre Lösung $y_p(x)$ durch **Superposition** der drei partikulären Lösungen von

$$L_n y_{p_1} = R_I, \quad L_n y_{p_2} = R_{II}, \quad L_n y_{p_3} = R_{III},$$

das heißt durch die Linearkombination □

$$\boxed{y_p(x) = C_1 y_{p_1}(x) + C_2 y_{p_2}(x) + C_3 y_{p_3}(x).}$$

BSP. (10.5.1) Wir berechnen die allgemeine Lösung der linearen inhomogenen DGl

$$\boxed{L_3 y := y''' - y'' + y' - y = 2\,\cosh x + x \sin x + \cos x, \quad x \in \mathbf{R}.}$$

In einem 1.Schritt lösen wir die homogene Differentialgleichung:

$$\begin{array}{rrcccccccc} \text{DGl:} & P_3(D)y & := & \Big(D^3 & - & D^2 & + & D & - & 1\Big)y & = & 0, \\ & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\ \text{charakt. Polynom:} & P_3(\lambda) & := & \lambda^3 & - & \lambda^2 & + & \lambda & - & 1 & = & 0. \end{array}$$

Man kann die Nullstelle $\lambda_1 = 1$ des charakteristischen Polynoms leicht erraten. Wir spalten den Linearfaktor $(\lambda - 1)$ mit Hilfe des HORNER–Schemas ab:

	1	-1	1	-1
$\lambda = 1$	$*$	1	0	1
	1	0	1	$\boxed{0}$

Wir erhalten nun die Linearfaktorzerlegung

$$P_3(\lambda) = (\lambda - 1)(\lambda^2 + 1) = (\lambda - 1)(\lambda - i)(\lambda + i),$$

und aus ihr resultiert die allgemeine Lösung der homogenen DGl in der *reellen* Form

$$\boxed{y_h(x) = C_1\, e^x + C_2\, \cos x + C_3\, \sin x, \quad x \in \mathbf{R}.}$$

Im 2.Schritt berechnen wir nach dem Superpositionsprinzip eine partikuläre Lösung der inhomogenen Differentialgleichung

$$P_3(D)y = e^x + e^{-x} + x\, \sin x + \cos x =: R_1(x) + R_2(x) + R_3(x) + R_4(x), \quad x \in \mathbf{R}.$$

(a) Wir betrachten $P_3(D)y = R_1(x) := e^x$. Da $\lambda_1 = 1$ eine einfache Nullstelle des charakteristischen Polynoms ist, erzeugt die Inhomogenität $R_1(x)$ **einfache Resonanz**. Der folgende **Resonanzansatz** ist erforderlich:

$$\begin{array}{lcrcrl} y_{p_1}(x) & = & A_0 x e^x & & & \cdot(-1) \\ y'_{p_1}(x) & = & A_0 x e^x & + & A_0 e^x & \cdot 1 \\ y''_{p_1}(x) & = & A_0 x e^x & + & 2A_0 e^x & \cdot(-1) \\ y'''_{p_1}(x) & = & A_0 x e^x & + & 3A_0 e^x & \cdot 1 \\ \hline e^x & \stackrel{!}{=} & & & 2A_0 e^x & \Rightarrow A_0 = 1/2. \end{array}$$

Wir erhalten eine Teillösung

$$y_{p_1}(x) = \frac{1}{2}\, x e^x, \quad x \in \mathbf{R}.$$

(b) Wir betrachten $P_3(D)y = R_2(x) := e^{-x}$. Die Inhomogenität $R_2(x)$ erzeugt keine Resonanz. Der folgende Direktansatz ist erforderlich:

$$\begin{array}{lcrl} y_{p_2}(x) & = & A_0 e^{-x} & \cdot(-1) \\ y'_{p_2}(x) & = & -A_0 e^{-x} & \cdot 1 \\ y''_{p_2}(x) & = & A_0 e^{-x} & \cdot(-1) \\ y'''_{p_2}(x) & = & -A_0 e^{-x} & \cdot 1 \\ \hline e^{-x} & \stackrel{!}{=} & -4A_0 e^{-x} & \Rightarrow A_0 = -1/4. \end{array}$$

Wir erhalten eine Teillösung

$$y_{p_2}(x) = -\frac{1}{4}\, e^{-x}, \quad x \in \mathbf{R}.$$

(c) Wir betrachten $P_3(D)y = R_3(x) + R_4(x) := x\, \sin x + \cos x$. Da $\lambda_2 = i$ eine einfache Nullstelle des charakteristischen Polynoms ist, erzeugen beide Inhomogenitäten $R_3(x), R_4(x)$ **einfache Resonanz**. Die folgenden **Resonanzansätze** sind erforderlich:

$$y_{p_3}(x) = x\Big((A_0 + A_1 x)\cos x + (B_0 + B_1 x)\sin x\Big), \quad y_{p_4}(x) = x\Big(A_0 \cos x + B_0 \sin x\Big).$$

Man fasst diese beiden Ansätze zusammen zu einem einzigen Ansatz, nämlich

$$\begin{array}{rcl|l}
y_{p_3}(x) = & (A_0x + A_1x^2)\cos x + (B_0x + B_1x^2)\sin x & & \cdot(-1)\\
y'_{p_3}(x) = & -(A_0x + A_1x^2)\sin x + (B_0x + B_1x^2)\cos x + (A_0 + 2A_1x)\cos x + (B_0 + 2B_1x)\sin x & & \cdot 1\\
y''_{p_3}(x) = & -(A_0x + A_1x^2)\cos x - (B_0x + B_1x^2)\sin x - 2(A_0 + 2A_1x)\sin x + 2(B_0 + 2B_1x)\cos x & & \\
& + 2A_1\cos x + 2B_1\sin x & & \cdot(-1)\\
y'''_{p_3}(x) = & (A_0x + A_1x^2)\sin x - (B_0x + B_1x^2)\cos x - 3(A_0 + 2A_1x)\cos x - 3(B_0 + 2B_1x)\sin x & & \\
& - 6A_1\sin x + 6B_1\cos x & & \cdot 1\\
\hline
x\sin x + \cos x & \stackrel{!}{=} & -4(A_1 + B_1)x\cos x + 4(A_1 - B_1)x\sin x & \\
& & -(2A_0 + 2B_0 + 2A_1 - 6B_1)\cos x + (2A_0 - 2B_0 - 6A_1 - 2B_1)\sin x. &
\end{array}$$

Durch Koeffizientenvergleich resultiert das lineare Gleichungssystem

A_0	A_1	B_0	B_1	1
0	1	0	1	0
0	4	0	−4	1
−2	−2	−2	6	1
2	−6	−2	−2	0

mit den Lösungen

$$A_0 = -\frac{3}{8}, \quad A_1 = \frac{1}{8}, \quad , B_0 = -\frac{5}{8}, \quad B_1 = -\frac{1}{8}.$$

Hiermit erhalten wir eine Teillösung

$$y_{p_3}(x) = \frac{1}{8}\left((x^2 - 3x)\,\cos x - (x^2 + 5x)\,\sin x\right), \quad x \in \mathbf{R}.$$

Durch Superposition $y_p(x) := y_{p_1}(x) + y_{p_2}(x) + y_{p_3}(x)$ haben wir eine partikuläre Lösung der inhomogenen DGl gewonnen, und die allgemeine Lösung hat nun die folgende Form

$$\boxed{y(x) = y_h(x) + y_p(x) = \left(C_1 + \frac{x}{2}\right)e^x - \frac{1}{4}\,e^{-x} + \left(C_2 + \frac{x^2}{8} - \frac{3x}{8}\right)\cos x + \left(C_3 - \frac{x^2}{8} - \frac{5x}{8}\right)\sin x, \quad x \in \mathbf{R}.}$$

BSP. (10.5.2) Wir berechnen hier die allgemeine Lösung der linearen inhomogenen DGl

$$\boxed{L_4y := y^{(4)} - 4y''' + 4y'' = 2 + 6x + xe^{2x}, \quad x \in \mathbf{R}.}$$

In einem 1.Schritt lösen wir wiederum die homogene Differentialgleichung:

$$\begin{array}{rrcccccccc}
\text{DGl:} & P_4(D)y & := & \left(D^4 & - & 4D^3 & + & 4D^2\right)y & = & 0,\\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\text{charakt. Polynom:} & P_4(\lambda) & := & \lambda^4 & - & 4\lambda^3 & + & 4\lambda^2 & = & 0.
\end{array}$$

Wir erhalten die Linearfaktorzerlegung

$$P_4(\lambda) = \lambda^2(\lambda^2 - 4\lambda + 4) = \lambda^2(\lambda - 2)^2,$$

und aus ihr resultiert die allgemeine Lösung der homogenen DGl

$$\boxed{y_h(x) = C_1 + C_2x + (C_3 + C_4x)\,e^{2x}, \quad x \in \mathbf{R}.}$$

Im 2.Schritt berechnen wir nach dem Superpositionsprinzip eine partikuläre Lösung der inhomogenen Differentialgleichung

$$P_4(D)y = (2 + 6x) + xe^{2x} =: R_1(x) + R_2(x), \quad x \in \mathbf{R}.$$

(a) Wir betrachten $P_4(D)y = R_1(x) := 2 + 6x$. Da $\lambda_1 = 0$ eine doppelte Nullstelle des charakteristischen Polynoms ist, erzeugt die Inhomogenität $R_1(x)$ **Doppelresonanz**. Der folgende **Resonanzansatz** ist erforderlich:

$$\begin{array}{lcl|l}
y_{p_1}(x) & = & A_0x^2 + A_1x^3 & \cdot 0 \\
y'_{p_1}(x) & = & 2A_0x + 3A_1x^2 & \cdot 0 \\
y''_{p_1}(x) & = & 2A_0 + 6A_1x & \cdot 4 \\
y'''_{p_1}(x) & = & 6A_1 & \cdot(-4) \\
y^{(4)}_{p_1}(x) & = & 0 & \cdot 1 \\
\hline
2 + 6x & \stackrel{!}{=} & 8A_0 - 24A_1 + 24A_1x & \Rightarrow A_0 = 1,\ A_1 = 1/4.
\end{array}$$

Wir erhalten eine Teillösung

$$y_{p_1}(x) = x^2 + \frac{1}{4}x^3, \quad x \in \mathbf{R}.$$

(b) Wir betrachten $P_4(D)y = R_2(x) := xe^{2x}$. Da $\lambda_2 = 2$ ebenfalls eine doppelte Nullstelle des charakteristischen Polynoms ist, erzeugt die Inhomogenität $R_2(x)$ wiederum **Doppelresonanz**. Der folgende **Resonanzansatz** ist erforderlich:

$$\begin{array}{lcl|l}
y_{p_2}(x) & = & (A_0x^2 + A_1x^3)e^{2x} & \cdot 0 \\
y'_{p_2}(x) & = & 2(A_0x^2 + A_1x^3)e^{2x} + (2A_0x + 3A_1x^2)e^{2x} & \cdot 0 \\
y''_{p_2}(x) & = & 4(A_0x^2 + A_1x^3)e^{2x} + 4(2A_0x + 3A_1x^2)e^{2x} + (2A_0 + 6A_1x)e^{2x} & \cdot 4 \\
y'''_{p_2}(x) & = & 8(A_0x^2 + A_1x^3)e^{2x} + 12(2A_0x + 3A_1x^2)e^{2x} + 6(2A_0 + 6A_1x)e^{2x} + 6A_1e^{2x} & \cdot(-4) \\
y^{(4)}_{p_2}(x) & = & 16(A_0x^2 + A_1x^3)e^{2x} + 32(2A_0x + 3A_1x^2)e^{2x} + 24(2A_0 + 6A_1x)e^{2x} + 48A_1e^{2x} & \cdot 1 \\
\hline
xe^{2x} & \stackrel{!}{=} & 4(2A_0 + 6A_1x)e^{2x} + 24A_1e^{2x}. &
\end{array}$$

Durch Koeffizientenvergleich erhält man die Lösungen

$$A_0 = -\frac{1}{8}, \quad A_1 = \frac{1}{24}.$$

Hieraus resultiert eine Teillösung

$$y_{p_2}(x) = \frac{1}{24}\left(x^3 - 3x^2\right)e^{2x}, \quad x \in \mathbf{R}.$$

Durch Superposition $y_p(x) := y_{p_1}(x) + y_{p_2}(x)$ haben wir eine partikuläre Lösung der inhomogenen DGl gewonnen, und die allgemeine Lösung hat nun die folgende Form

$$\boxed{y(x) = y_h(x) + y_p(x) = C_1 + C_2x + x^2 + \frac{x^3}{4} + \left(C_3 + C_4x - \frac{x^2}{8} + \frac{x^3}{24}\right)e^{2x}, \quad x \in \mathbf{R}.}$$

10.6 Die Eulersche Differentialgleichung

In einigen Spezialfällen gelingt es, die lineare gewöhnliche Differentialgleichung mit **nichtkonstanten Koeffizienten**

$$L_ny := \sum_{k=0}^{n} a_k(x)D_x^k y = f(x), \quad x \in X,\ a_n(x) \neq 0\ \forall\ x \in X, \tag{6.1}$$

durch eine bijektive Variablentransformation

$$x = \varphi(t),\ t \in I \quad \Leftrightarrow \quad t = \varphi^{-1}(x) =: \psi(x),\ x \in X, \tag{6.2}$$

in eine Differentialgleichung mit **konstanten Koeffizienten** in der neuen Variablen t zu überführen. Dabei müssen die Ableitungen $D_x^k y(x)$ umgerechnet werden auf die Ableitungen nach

der neuen Variablen $t = \psi(x)$. Fasst man in der folgenden Analyse die Funktion $y(x) = y\big(\varphi(t)\big)$ als Funktion der Variablen t auf (ohne das Funktionssymbol y zu ändern), so kann diese Umrechnung mit Hilfe der Kettenregel bewerkstelligt werden, zum Beispiel:

$$\frac{dy}{dx} = \frac{dy}{dt}\cdot\frac{dt}{dx} = \frac{dy}{dt}\cdot\frac{d\psi}{dx}.$$

Wir setzen voraus, dass $\varphi \in C^n(I)$ gelte sowie $(d/dt)\varphi(t) \neq 0 \;\forall\; t \in I$. Dann ist $\varphi : I \to X$ streng monoton. Die Umkehrfunktion $t = \psi(x)$ existiert, und es gilt $\psi \in C^n(X)$ sowie $dt/dx = 1/\dot\varphi(t)$. Hier bezeichne "$\cdot$" $:= d/dt$ die Ableitung nach der Variablen t. Man erhält nun unter diesen Voraussetzungen durch wiederholte Anwendung der Kettenregel:

$$\left.\begin{aligned}
\frac{dy}{dx} &= \frac{dy}{dt}\cdot\frac{dt}{dx} &&= \dot y\,\frac{dt}{dx},\\
\frac{d^2y}{dx^2} &= \frac{d}{dx}\Big(\dot y\,\frac{dt}{dx}\Big) &&= \ddot y\Big(\frac{dt}{dx}\Big)^2 + \dot y\,\frac{d^2t}{dx^2},\\
\frac{d^3y}{dx^3} &= \frac{d}{dx}\Big(\ddot y\Big(\frac{dt}{dx}\Big)^2 + \dot y\,\frac{d^2t}{dx^2}\Big) &&= y^{(3)}\Big(\frac{dt}{dx}\Big)^3 + 3\ddot y\,\frac{d^2t}{dx^2}\cdot\frac{dt}{dx} + \dot y\,\frac{d^3t}{dx^3},
\end{aligned}\right\} \tag{6.3}$$

usf. Wir betrachten jetzt die spezielle Variablentransformation

$$\boxed{\varphi : \begin{cases} \mathbf{R} \to (0,+\infty), \\ t \mapsto x = \varphi(t) := e^t, \end{cases} \qquad \varphi^{-1} = \psi : \begin{cases} (0,+\infty) \to \mathbf{R}, \\ x \mapsto t = \psi(x) := \ln x. \end{cases}} \tag{6.4}$$

Es ist nicht schwierig, die folgenden Ableitungen zu ermitteln:

$$\frac{dt}{dx} = \frac{1}{x}, \quad \frac{d^2t}{dx^2} = -\frac{1}{x^2}, \quad \frac{d^3t}{dx^3} = \frac{2}{x^3}, \quad \cdots \quad \frac{d^nt}{dx^n} = (-1)^{n-1}\,\frac{(n-1)!}{x^n}.$$

Wir verwenden diese Identitäten in den Formeln (6.3) und bezeichnen dazu

$$\boxed{D := \frac{d}{dt}.}$$

Dann folgt:

$$\begin{aligned}
x\,\frac{dy}{dx} &= \frac{x}{x}\,\dot y &&= Dy,\\
x^2\frac{d^2y}{dx^2} &= x^2\Big(\frac{1}{x^2}\,\ddot y - \frac{1}{x^2}\,\dot y\Big) &&= \ddot y - \dot y = D(D-1)y,\\
x^3\frac{d^3y}{dx^3} &= x^3\Big(\frac{1}{x^3}D^3y - \frac{3}{x^3}\ddot y + \frac{2}{x^3}\dot y\Big) &&= D^3y - 3\ddot y + 2\dot y = D(D-1)(D-2)y,
\end{aligned}$$

usf. Allgemein zeigt man durch vollständige Induktion nach n:

$$\boxed{x^n\,\frac{d^ny}{dx^n} = D(D-1)(D-2)\cdots(D-n+1)y, \;\; n \in \mathbf{N},\; D := \frac{d}{dt}.} \tag{6.5}$$

Wegen dieser Relation gelingt es, die Differentialgleichung mit nichtkonstanten Koeffizienten

$$\boxed{\begin{aligned}
L_ny &:= \sum_{k=0}^{n} a_kx^ky^{(k)}\\
&= a_nx^ny^{(n)} + a_{n-1}x^{n-1}y^{(n-1)} + \cdots + a_1xy' + a_0y \overset{!}{=} f(x),\; x > 0,\\
&\quad\; a_n \neq 0,\; a_k \in \mathbf{K},
\end{aligned}} \tag{6.6}$$

mit Hilfe der Transformation (6.4) in eine DGl mit konstanten Koeffizienten zu überführen.

Definition 10.8 *Die lineare gewöhnliche DGl* (6.6) *heißt* EULER**sche Differentialgleichung** *n–ter Ordnung. Mit Hilfe der Variablentransformation $t := \ln x$, $x > 0$, transformiert man die* EULER*sche Differentialgleichung* (6.6) *in eine lineare gewöhnliche DGl mit konstanten Koeffizienten, nämlich*

$$\boxed{P_n(D)y := \sum_{k=1}^{n} a_k D(D-1)\cdots(D-k+1)y(t) + a_0 y(t) = f(e^t),\ t \in \mathbf{R},}$$

wobei $D := d/dt$ zu setzen ist.

Nach erfolgter Transformation behandelt man die EULERsche DGl mit den Methoden aus den Abschnitten 10.4 und 10.5.

BSP. (10.6.1) Es ist zweckmäßig, die Transformation einer gegebenen EULERschen DGl nach folgendem Formalismus durchzuführen, den wir am Beispiel der DGl

$$\boxed{L_3 y := x^3 y''' - x^2 y'' + 6xy' - 10y = x^2 + \ln x^2,\ \ x > 0,}$$

vorführen:

$$\begin{array}{rlcccccccccl}
\text{DGl:} & L_3 y & := & x^3y''' & - & x^2y'' & + 6xy' & - & 10y & = & x^2 + \ln x^2, \\
 & & & \downarrow & & \downarrow & \downarrow & & \downarrow & & \\
\text{Transf. } x = e^t\text{:} & P_3(D)y & = & \underbrace{\Big(D(D-1)(D-2)} & - & \underbrace{D(D-1)} & + 6D & - & 10\Big)y & = & e^{2t} + 2t, \\
 & & = & \Big(D^3 & - & 4D^2 & + 9D & - & 10\Big)y & & \\
 & & & \downarrow & & \downarrow & \downarrow & & \downarrow & & \\
\text{charakt. Polynom:} & P_3(\lambda) & = & \lambda^3 & - & 4\lambda^2 & + 9\lambda & - & 10 & = & 0.
\end{array}$$

Man kann die Nullstelle $\lambda_1 = 2$ des charakteristischen Polynoms leicht erraten. Wir spalten den Linearfaktor $(\lambda - 2)$ mit Hilfe des HORNER–Schemas ab:

	1	−4	9	−10
$\lambda = 2$	*	2	−4	10
	1	−2	5	0

Wir erhalten nun die Linearfaktorzerlegung

$$P_3(\lambda) = (\lambda - 2)(\lambda^2 - 2\lambda + 5) = (\lambda - 2)(\lambda - 1 - 2i)(\lambda - 1 + 2i),$$

und aus ihr resultiert die allgemeine Lösung der homogenen DGl in der *reellen* Form

$$\boxed{y_h(t) = C_1\, e^{2t} + e^t\Big(C_2\, \cos 2t + C_3\, \sin 2t\Big),\ \ t \in \mathbf{R}.}$$

Im nächsten Schritt berechnen wir eine partikuläre Lösung der inhomogenen Differentialgleichung $P_3(D)y = e^{2t} + 2t$. Da $\lambda_1 = 2$ eine einfache Nullstelle des charakteristischen Polynoms ist, erzeugt die Inhomogenität e^{2t} **einfache Resonanz**, während die Inhomogenität $2t$ resonanzfrei ist. Der folgende kombinierte **Resonanzansatz** ist erforderlich:

$$\begin{array}{lcrcrcrl}
y_p(t) & = & A_0 t e^{2t} & + & B_0 & + & B_1 t & \cdot(-10) \\
Dy_p(t) & = & A_0(2t+1)e^{2t} & & & + & B_1 & \cdot 9 \\
D^2y_p(t) & = & A_0(4t+4)e^{2t} & & & & & \cdot(-4) \\
D^3y_p(t) & = & A_0(8t+12)e^{2t} & & & & & \cdot 1 \\
\hline
e^{2t} + 2t & \stackrel{!}{=} & 5A_0e^{2t} & - & 10B_0 & + & B_1(9 - 10t). &
\end{array}$$

Durch Koeffizientenvergleich erhält man

$$A_0 = \frac{1}{5}, \quad B_0 = -\frac{9}{50}, \quad B_1 = -\frac{1}{5},$$

und hieraus resultiert die allgemeine Lösung

$$y(t) = y_h(t) + y_p(t) = e^{2t}\Big(C_1 + \frac{1}{5}\,t\Big) + e^t\Big(C_2\,\cos 2t + C_3\,\sin 2t\Big) - \frac{1}{5}\Big(t + \frac{9}{10}\Big), \quad t \in \mathbf{R}.$$

Durch Rücktransformation $t = \ln x$ erhält man diese Lösung in Abhängigkeit von der Variablen $x > 0$:

$$\boxed{y(x) = x^2\Big(C_1 + \frac{1}{5}\,\ln x\Big) + x\Big(C_2\,\cos(\ln x^2) + C_3\,\sin(\ln x^2)\Big) - \frac{1}{5}\Big(\ln x + \frac{9}{10}\Big), \quad x > 0.}$$

BSP. (10.6.2) Wir berechnen hier die Lösung der EULERschen DGl aus BSP. (10.3.1), Abschnitt 10.3, nämlich

$$\boxed{L_3 y := x^3 y''' - 3x^2 y'' + 7xy' - 8y = x, \quad x > 0.}$$

Wir verfahren analog zum vorangegangenen Beispiel:

$$\begin{array}{rrrcccccccc}
\text{DGl:} & L_3 y & := & x^3y''' & - & 3x^2y'' & + 7xy' & - 8y & = & x, \\
 & & & \downarrow & & \downarrow & \downarrow & \downarrow & & \\
\text{Transf. } x = e^t\text{:} & P_3(D)y & = & \underbrace{\big(D(D-1)(D-2) - 3D(D-1) + 7D - 8\big)}\,y & & & & & = & e^t, \\
 & & = & \big(D^3 & - & 6D^2 & + 12D - 8\big)y & & & \\
 & & & \downarrow & & \downarrow & \downarrow & \downarrow & & \\
\text{charakt. Polynom:} & P_3(\lambda) & = & \lambda^3 & - & 6\lambda^2 & + 12\lambda - & 8 & = & 0.
\end{array}$$

Man kann die Nullstelle $\lambda_1 = 2$ des charakteristischen Polynoms leicht erraten. Wir spalten den Linearfaktor $(\lambda - 2)$ mit Hilfe des HORNER–Schemas ab:

	1	−6	12	−8
$\lambda = 2$	*	2	−8	8
	1	−4	4	0

Wir erhalten nun die Linearfaktorzerlegung

$$P_3(\lambda) = (\lambda - 2)(\lambda^2 - 4\lambda + 4) = (\lambda - 2)^3,$$

und aus ihr resultiert die allgemeine Lösung der homogenen DGl

$$\boxed{y_h(t) = \Big(C_1 + C_2 t + C_3 t^2\Big)\,e^{2t}, \quad t \in \mathbf{R}.}$$

Im nächsten Schritt berechnen wir eine partikuläre Lösung der inhomogenen Differentialgleichung $P_3(D)y = e^t$. Da $\lambda = 1$ keine Nullstelle des charakteristischen Polynoms ist, ist die Inhomogenität e^t resonanzfrei. Der folgende **Direktansatz** ist erforderlich:

$$\begin{array}{lcl|l}
y_p(t) & = & A_0 e^t & \cdot(-8) \\
Dy_p(t) & = & A_0 e^t & \cdot 12 \\
D^2 y_p(t) & = & A_0 e^t & \cdot(-6) \\
D^3 y_p(t) & = & A_0 e^t & \cdot 1 \\
\hline
e^t & \stackrel{!}{=} & -A_0 e^t & \Rightarrow A_0 = -1.
\end{array}$$

Hieraus resultiert die allgemeine Lösung

$$y(t) = y_h(t) + y_p(t) = e^{2t}\Big(C_1 + C_2 t + C_3 t^2\Big) - e^t, \quad t \in \mathbf{R}.$$

Durch Rücktransformation $t = \ln x$ erhält man diese Lösung in Abhängigkeit von der Variablen $x > 0$:

$$\boxed{y(x) = x^2\Big(C_1 + C_2 \ln x + C_3 (\ln x)^2\Big) - x, \quad x > 0.}$$

Kapitel 11

Eigenwerte und Eigenvektoren von Matrizen

11.1 Das Eigenwertproblem

Die homogene lineare gewöhnliche Differentialgleichung n–ter Ordnung mit konstanten Koeffizienten

$$\boxed{L_n y := y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0} \tag{1.1}$$

kann in ein äquivalentes **System** von n gewöhnlichen Differentialgleichungen 1.Ordnung überführt werden. Dazu führe man neue abhängige Veränderliche

$$y_1(x) := y(x),\ y_2(x) := y'(x), \ldots, y_n(x) := y^{(n-1)}(x) \tag{1.2}$$

ein. Aus dieser Zuordnung ergibt sich ganz offensichtlich das zu (1.1) äquivalente System

$$\left.\begin{array}{lcl} y_1'(x) & = & y_2(x), \\ y_2'(x) & = & y_3(x), \\ y_3'(x) & = & y_4(x) \\ \vdots & & \vdots \\ y_{n-1}'(x) & = & y_n(x), \\ y_n'(x) & = & RS(x), \end{array}\right\} \quad \Leftrightarrow \quad \vec{y}\,'(x) = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & & & 0 \\ 0 & 0 & 0 & \ddots & & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & \cdots & -a_{n-1} \end{bmatrix}}_{A} \vec{y}(x),$$

wobei wir die Bezeichnungen

$$RS(x) := -\sum_{k=1}^{n} a_{k-1} y_k(x), \quad \vec{y}(x) := \Big(y_1(x), y_2(x), \ldots, y_n(x)\Big)^T$$

verwendet haben.

BSP. (11.1.1) Es sei die Differentialgleichung

$$\boxed{L_3 y := y''' - 4y'' + 5y' - y = 0}$$

vorgelegt. Wir setzen $y_1(x) := y(x)$, $y_2(x) := y'(x)$, $y_3(x) := y''(x)$. Nun kann die gegebene DGl 3.Ordnung in der folgenden Form als System 1.Ordnung geschrieben werden:

$$\boxed{\vec{y}\,'(x) = \begin{bmatrix} y_1'(x) \\ y_2'(x) \\ y_3'(x) \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -5 & 4 \end{bmatrix}}_{=:A} \vec{y}(x) =: A\vec{y}(x).}$$

In Verallgemeinerung dieses Zusammenhangs können wir uns mit der folgenden Aufgabenstellung auseinandersetzen. Gegeben sei eine Matrix $A \in \mathbf{K}^{(n,n)}$. Man bestimme eine differenzierbare Vektorfunktion $\vec{y}(x) := \Big(y_1(x), y_2(x), \ldots, y_n(x)\Big)^T$ so, dass gilt:

$$\boxed{\vec{y}\,'(x) = A\vec{y}(x), \quad x \in \mathbf{R}.} \tag{1.3}$$

Der bei linearen gewöhnlichen Differentialgleichungen mit konstanten Koeffizienten erprobte "$e^{\lambda x}$–Ansatz" führt auch hier zu einer *Algebraisierung* der Gleichung (1.3). Wird nämlich

$$\vec{y}(x) = e^{\lambda x}\vec{v}, \quad \vec{v} \in \mathbf{C}^n, \tag{1.4}$$

in (1.3) eingesetzt, so resultiert

$$\vec{0} = e^{\lambda x}(A - \lambda\, Id)\vec{v}.$$

Wegen $e^{\lambda x} \neq 0 \ \forall\, \lambda \in \mathbf{C}$ hat man also folgendes **Eigenwertproblem** zu lösen:

$$\boxed{\text{Finde } \lambda \in \mathbf{C} \text{ und } \vec{0} \neq \vec{v} \in \mathbf{C}^n \text{ so, dass } A\vec{v} = \lambda\vec{v}.}$$

Definition 11.1 *Zu gegebener quadratischer Matrix $A \in \mathbf{K}^{(n,n)}$ heiße eine Zahl $\lambda \in \mathbf{C}$* **Eigenwert** *(Ew) von A, wenn es ein $\vec{v} \in \mathbf{C}^n$, $\vec{v} \neq \vec{0}$, derart gibt, dass*

$$\boxed{A\vec{v} = \lambda\,\vec{v}} \tag{1.5}$$

gilt. Der so definierte Vektor $\vec{v}$ heiße **Eigenvektor** *(Ev) zum* **Eigenwert** *λ.*

BSP. (11.1.2) Die **Diagonalmatrix** $\Lambda := \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$ erfüllt offenbar $\Lambda\vec{e}_j = \lambda_j\,\vec{e}_j \ \forall\ j = 1, \ldots, n$, worin $\vec{e}_j := (0, \ldots, 0, \underbrace{1}_{j\text{–te Stelle}}, 0, \ldots, 0)^T$ den j–ten **Standardbasisvektor** des $\mathbf{K}^n$ bezeichnet. Das heißt, $\vec{e}_j$ ist Eigenvektor zum Eigenwert λ_j. Ist $\lambda_j = \lambda \ \forall\ j = 1, \ldots, n$, so zeigt dieses Beispiel insbesondere, dass möglicherweise nur ein einziger Eigenwert λ existiert und dazu mehrere Eigenvektoren.

Bemerkung 11.1 Äquivalent mit (1.5) ist die Lösung des linearen homogenen Gleichungssystems

$$\boxed{(A - \lambda\, Id)\vec{v} = \vec{0}.} \tag{1.6}$$

Nach den aus der linearen Algebra bekannten Lösbarkeitskriterien (vgl. Satz 5.34) ist das Gleichungssystem (1.6) genau dann nichttrivial lösbar, wenn die Determinante von $A - \lambda\, Id$ verschwindet: $\det(A - \lambda\, Id) = 0$. Aus diesem Sachverhalt resultiert ein Kriterium zur Bestimmung der Eigenwerte. □

Satz 11.1 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$.*

(a) *Genau dann ist $\lambda \in \mathbf{C}$ ein Eigenwert von A, wenn gilt:*

$$\boxed{P_n(\lambda) := \det(A - \lambda\, Id) = 0.} \tag{1.7}$$

(b) *Wir haben*

$$\boxed{P_n(\lambda) = \det(A - \lambda\, Id) = (-1)^n\lambda^n + p_{n-1}\lambda^{n-1} + \cdots + p_1\lambda + \det A,} \tag{1.8}$$

das heißt, $\det(A - \lambda\, Id)$ *ist ein Polynom in λ vom Grade genau n.*

Begründung: Für die Spaltenvektoren $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$ der Matrix A gilt ja $\vec{a}_j = A\vec{e}_j \; \forall\, j = 1, 2, \ldots, n$, so dass aus den Rechenregeln über Determinanten Folgendes resultiert:

$$\left.\begin{aligned} \det(A - \lambda\, Id) &= \det\Big[(A - \lambda\, Id)\vec{e}_1, \ldots, (A - \lambda\, Id)\vec{e}_n\Big] \\ &= \det(\vec{a}_1 - \lambda\vec{e}_1, \vec{a}_2 - \lambda\vec{e}_2, \ldots, \vec{a}_n - \lambda\vec{e}_n) = \det(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n) \\ &+\lambda \sum_{j=1}^{n} \det(\vec{a}_1, \ldots, \vec{a}_{j-1}, -\vec{e}_j, \vec{a}_{j+1}, \ldots, \vec{a}_n) + \cdots \\ &+(-1)^{n-1}\lambda^{n-1} \sum_{j=1}^{n} \underbrace{\det(\vec{e}_1, \ldots, \vec{e}_{j-1}, \vec{a}_j, \vec{e}_{j+1}, \ldots, \vec{e}_n)}_{=a_{jj}} \\ &+(-1)^n \lambda^n \underbrace{\det\, Id}_{=1}. \end{aligned}\right\} \tag{1.9}$$

Hieraus folgen schon die behaupteten Relationen. □

Bemerkung 11.2 (a) Die Formeln (1.9) liefern **explizite Darstellungen** für die Koeffizienten p_k des Polynoms $P_n(\lambda)$. Insbesondere gilt

$$\left.\begin{aligned} p_0 &= \det A, \quad p_1 = -\sum_{j=1}^{n} \det(\vec{a}_1, \ldots, \vec{a}_{j-1}, \vec{e}_j, \vec{a}_{j+1}, \ldots, \vec{a}_n), \\ p_{n-1} &= (-1)^{n-1} \sum_{j=1}^{n} a_{jj}. \end{aligned}\right\} \tag{1.10}$$

(b) Der Fundamentalsatz der Algebra weist dem Polynom $P_n(\lambda)$ genau n Nullstellen zu, und diese sind gemäß Satz 11.1 genau die Eigenwerte der Matrix A. □

Definition 11.2 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$.*

(a) *Das Polynom von Grade genau n :*

$$\boxed{P_n(\lambda) := \det(A - \lambda\, Id)}$$

heiße **charakteristisches Polynom** *der Matrix A.*

(b) *Die Zahl*

$$\boxed{\mathrm{Sp}(A) := \sum_{j=1}^{n} a_{jj}}$$

heiße **Spur** *von A (manchmal auch "$\mathrm{tr}(A)$" von* **trace**).

(c) *Ist λ_j eine k_j–fache Nullstelle des charakteristischen Polynoms $P_n(\lambda)$, das heißt, gilt*

$$\boxed{P_n(\lambda) = (\lambda - \lambda_j)^{k_j} Q(\lambda) \quad mit \quad Q(\lambda_j) \neq 0,}$$

so heiße k_j die **algebraische Dimension** *des Eigenwertes λ_j. Stets gilt $\sum\limits_j k_j = n$.*

Bemerkung 11.3 Es seien $\lambda_1, \lambda_2, \ldots, \lambda_m$, $m \leq n$, die *paarweise verschiedenen* Nullstellen des charakteristischen Polynoms $P_n(\lambda)$ mit ihren *Vielfachheiten* $k_1, \ldots, k_m$. Dann gilt offenbar die Linearfaktorzerlegung

$$P_n(\lambda) = (-1)^n(\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_m)^{k_m} = (-1)^n \prod_{j=1}^{m} (\lambda - \lambda_j)^{k_j}. \tag{1.11}$$

Wegen (1.10) folgt deshalb aus den VIËTAschen Wurzelsätzen: □

$$\boxed{\begin{aligned} P_n(0) &= p_0 = \det A = \prod_{j=1}^{m} \lambda_j^{k_j}; \\ \mathrm{Sp}(A) &= (-1)^{n-1} p_{n-1} = \sum_{j=1}^{m} k_j\, \lambda_j. \end{aligned}}$$

Satz 11.2 *Gegeben sei die Matrix* $A \in \mathbf{K}^{(n,n)}$.

(a) *Genau dann existiert die inverse Matrix* $A^{-1} \in \mathbf{K}^{(n,n)}$, *wenn alle Eigenwerte von* A *von Null verschieden sind.*

(b) *Es gelten die Beziehungen*

$$\boxed{\det A = \prod_{j=1}^{n} \lambda_j \quad \textit{und} \quad \mathrm{Sp}(A) \equiv \sum_{j=1}^{n} a_{jj} = \sum_{j=1}^{n} \lambda_j,}$$

wobei jeder Eigenwert so oft zu zählen ist, wie seine (algebraische) Vielfachheit angibt.

BSP. (11.1.3) Die Matrix $A := \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ hat das charakteristische Polynom $P_2(\lambda) = \det(A - \lambda\, Id) = \begin{vmatrix} 1-\lambda & 3 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2$, und demzufolge einen doppelten Eigenwert $\lambda_1 = 1$. Wir bestätigen $\mathrm{Sp}(A) = 2 = 2 \cdot \lambda_1$ sowie $\det A = 1 = \lambda_1^2$.

Hat man die Eigenwerte λ_j von A gefunden, so folgt aus (1.6), dass die jedem λ_j zugeordneten Eigenvektoren $\vec{v}$ den **Kern** der Abbildung $A - \lambda_j\, Id : \mathbf{C}^n \to \mathbf{C}^n$ aufspannen. Wegen der Unterraumeigenschaft $\mathrm{Kern}\,(A - \lambda_j\, Id) \subseteq \mathbf{C}^n$ gilt natürlich

$$1 \le \rho(\lambda_j) := \dim \mathrm{Kern}\,(A - \lambda_j\, Id) \le n.$$

Definition 11.3 *Gegeben sei die Matrix* $A \in \mathbf{K}^{(n,n)}$. *Sei* λ_j *Eigenwert von* A. *Dann heiße der von den zu* λ_j *gehörigen Eigenvektoren* $\vec{v} \in \mathbf{C}^n$ *aufgespannte Unterraum*

$$\boxed{\mathrm{Kern}\,(A - \lambda_j\, Id) \subseteq \mathbf{C}^n}$$

der **Eigenraum** *von* λ_j. *Seine Dimension* $\rho(\lambda_j) := \dim \mathrm{Kern}\,(A - \lambda_j\, Id)$ *heiße die* **geometrische Dimension** *des Eigenwertes* λ_j.

Die Berechnung des Eigenraumes zum Eigenwert λ läuft also auf die Lösung des homogenen linearen Gleichungssystems $(A - \lambda\, Id)\vec{v} = \vec{0}$ hinaus. Die folgenden Beispiele zeigen, dass algebraische und geometrische Dimension eines Eigenwertes λ im allgemeinen verschieden sind.

BSP. (11.1.4) Gegeben sei die Matrix $A := \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 4 \\ -1 & -1 & -2 \end{bmatrix}$ mit dem charakteristischen Polynom

$$P_3(\lambda) = \det(A - \lambda\, Id) = \begin{vmatrix} 2-\lambda & 1 & 1 \\ 2 & 3-\lambda & 4 \\ -1 & -1 & -2-\lambda \end{vmatrix} = -(\lambda-3)(\lambda-1)(\lambda+1).$$

Es gibt drei verschiedene Eigenwerte $\lambda_1 = 3,\ \lambda_2 = 1,\ \lambda_3 = -1$ mit algebraischer $\dim \lambda_j = 1$. Die zugeordneten Eigenvektoren $\vec{v}_j$ sind die Lösungen der homogenen Gleichungssysteme $(A - \lambda_j\, Id)\vec{v}_j = \vec{0}$. Man verifiziert ohne Schwierigkeiten:

$$\mathrm{Kern}\,(A - \lambda_1\, Id) = \mathrm{span}\Big\{(2,3,-1)^T\Big\}; \quad \mathrm{Kern}\,(A - \lambda_2\, Id) = \mathrm{span}\Big\{(1,-1,0)^T\Big\};$$

$$\text{Kern}\,(A-\lambda_3\,Id) = \text{span}\Big\{(0,1,-1)^T\Big\}.$$

Demzufolge gilt hier $\rho(\lambda_j) = \dim \text{Kern}\,(A-\lambda_j\,Id) = 1 = \text{algebr. dim}\,\lambda_j$.

Beachte: Die Eigenvektoren $\vec{v}_j,\ j=1,2,3$, sind *linear unabhängig.*

BSP. (11.1.5) Wir betrachten

$$A := \begin{bmatrix} 1 & 4 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{mit} \quad P_3(\lambda) = \det(A-\lambda\,Id) = \begin{vmatrix} 1-\lambda & 4 & 3 \\ 0 & 1-\lambda & 2 \\ 0 & 0 & 1-\lambda \end{vmatrix} = -(\lambda-1)^3.$$

Es gibt nur einen einzigen Eigenwert $\lambda_1 = 1$ mit algebraischer $\dim\lambda_1 = 3$. Die Bestimmung des Eigenraumes $\text{Kern}\,(A-\lambda_1\,Id)$ läuft auf die Lösung des homogenen Gleichungssystems (1.6) hinaus, also

$$(A-\lambda_1\,Id)\vec{v} = \vec{0} = (A-\,Id)\vec{v} = \begin{bmatrix} 0 & 4 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \Leftrightarrow v_2 = v_3 = 0,\ \ \vec{v} = v_1 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ \ v_1 \neq 0.$$

Wir erhalten $\text{Kern}\,(A-\lambda_1\,Id) = \text{span}\Big\{(1,0,0)^T\Big\}$, und somit $\rho(\lambda_1) = 1 < 3 = \text{algebr. dim}\,\lambda_1$. Die Matrix A hat nur einen einzigen Eigenvektor.

Wir haben allgemein:

Satz 11.3 *Seien λ_j Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$.*

(a) *Es gilt $\lambda_1 \neq \lambda_2$ genau, wenn* $\text{Kern}\,(A-\lambda_1\,Id) \cap \text{Kern}\,(A-\lambda_2\,Id) = \{\vec{0}\}$.

(b) *Eigenvektoren zu* **verschiedenen** *Eigenwerten sind linear unabhängig* **(LU)**.

(c) $\rho(\lambda_j) = \text{geom. dim}\,\lambda_j \leq \text{algebr. dim}\,\lambda_j = k_j$.

(d) *Genau dann gilt $\rho(\lambda_j) = k_j\ \forall\, j$, wenn $n = \sum\limits_j \rho(\lambda_j)$ ist. Dies ist genau dann richtig, wenn es in $\mathbf{C}^n$ eine Basis aus Eigenvektoren der Matrix A gibt.*

(e) *Die Diagonalmatrix $\Lambda := \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$ hat genau die Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_n$, und die Vektoren der Standardbasis $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ sind die zugeordneten Eigenvektoren.*

Begründungen: (a) Wäre $\vec{0} \neq \vec{v} \in \text{Kern}\,(A-\lambda_1\,Id) \cap \text{Kern}\,(A-\lambda_2\,Id)$, so wäre $A\vec{v}-\lambda_1\vec{v} = A\vec{v}-\lambda_2\vec{v} = \vec{0}$, also $(\lambda_2-\lambda_1)\vec{v} = \vec{0}$, und somit $\lambda_1 = \lambda_2$.

(b) Diese Aussage ist lediglich eine Interpretation der Aussage (a).

(c) Wir fixieren einen Eigenwert λ von A mit $\rho = \text{geom. dim}\,\lambda$ und $k = \text{algebr. dim}\,\lambda$. Gelte $\text{Kern}\,(A-\lambda\,Id) = \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_\rho\}$. Nach dem Basisergänzungssatz (Satz 4.12) gibt es $n-\rho$ weitere LU Vektoren $\vec{v}_j \in \mathbf{C}^n,\ j = \rho+1, \ldots, n$, so dass $\mathbf{C}^n = \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ gilt. Für die Matrix $T := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) \in \mathbf{C}^{(n,n)}$ gilt $\text{Rang}\,T = n$, und somit $T \in \text{Inv}\,(\mathbf{C}^n)$ sowie $T\vec{e}_j = \vec{v}_j$ bzw. $\vec{e}_j = T^{-1}\vec{v}_j\ \forall\, j = 1, 2, \ldots, n$. Wir folgern

$$T^{-1}AT\vec{e}_j = T^{-1}A\vec{v}_j = \lambda T^{-1}\vec{v}_j = \lambda\vec{e}_j\ \ \forall\, j = 1, 2, \ldots, \rho.$$

Also muss $T^{-1}AT$ die Gestalt

$$T^{-1}AT = \left[\begin{array}{c|c} \lambda\,Id_\rho & B \\ \hline 0 & C \end{array}\right] \quad \text{mit} \quad Id_\rho := \text{diag}\,(1, 1, \ldots, 1) \in \mathbf{K}^{(\rho,\rho)}$$

haben. Wir erschließen

$$\left.\begin{array}{rcl} P_n(\mu) := & \det(T^{-1}AT - \mu\,Id) = \det[T^{-1}(A-\mu\,Id)T] \\ = & (\det T^{-1})[\det(A-\mu\,Id)](\det T) = \det(A-\mu\,Id) \\ = & (\lambda-\mu)^\rho \cdot \det(C-\mu\,Id). \end{array}\right\} \tag{1.12}$$

Das heißt, λ ist mindestens eine ρ–fache NS von $\det(A - \mu\, Id)$. Also muss $k \geq \rho$ gelten.

(d) Wegen $\rho(\lambda_j) \leq k_j$ und $n = \sum\limits_j k_j$ kann $n = \sum\limits_j \rho(\lambda_j)$ nur dann gelten, wenn $\rho(\lambda_j) = k_j \; \forall\, j$ erfüllt ist. Das heißt, die Eigenräume $\mathrm{Kern}\,(A - \lambda_j\, Id)$ von A spannen den Vektorraum $\mathbf{C}^n$ auf.

(e) folgt aus BSP. (11.1.2). Da $\mathrm{span}\,\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\} = \mathbf{K}^n$ gilt, liegen alle Eigenvektoren vor. □

Bemerkung 11.4 (a) Aus der Rechnung (1.12) resultiert folgende Tatsache: Gegeben seien $A \in \mathbf{K}^{(n,n)}$ und $T \in \mathrm{Inv}\,(\mathbf{K}^n)$. Dann haben A und $B := T^{-1}AT$ dasselbe charakteristische Polynom $P_n(\lambda)$, also auch dieselben Eigenwerte:

$$\boxed{P_n(\lambda) = \det(A - \lambda\, Id) = \det(T^{-1}AT - \lambda\, Id).} \tag{1.13}$$

(b) Gilt geom. $\dim \lambda_j =$ algebr. $\dim \lambda_j \; \forall\, j$, so bilden die zugeordneten Eigenvektoren $\vec{v}_1, \ldots, \vec{v}_n$ gemäß Satz 11.3 eine Basis des Vektorraumes $\mathbf{C}^n$. Die Matrix

$$\boxed{T := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) \in \mathbf{C}^{(n,n)}} \tag{1.14}$$

hat $\mathrm{Rang}\, T = n$, so dass $T \in \mathrm{Inv}\,(\mathbf{C}^n)$ resultiert. T heißt **Modalmatrix** von A. Mit der Diagonalmatrix $\Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$ gilt dann □

$$\boxed{AT = (A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_n) = (\lambda_1\vec{v}_1, \lambda_2\vec{v}_2, \ldots, \lambda_n\vec{v}_n) = T\Lambda.} \tag{1.15}$$

BSP. (11.1.6) Die Matrix $A := \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix}$ hat das charakteristische Polynom

$$P_3(\lambda) = \det(A - \lambda\, Id) = \begin{vmatrix} 3-\lambda & 0 & -1 \\ 0 & 3-\lambda & 4 \\ 0 & 0 & 2-\lambda \end{vmatrix} = (3-\lambda)^2(2-\lambda).$$

Es liegen die Eigenwerte $\lambda_1 = 2$ (einfach) und $\lambda_2 = 3$ (doppelt) vor. Man erhält mit leichter Rechnung $\mathrm{Kern}\,(A - \lambda_1 Id) = \mathrm{span}\left\{(1, -4, 1)^T\right\}$. Zur Bestimmung des Eigenraumes $\mathrm{Kern}\,(A - \lambda_2\, Id)$ lösen wir das homogene Gleichungssystem

$$\vec{0} = (A - \lambda_2\, Id)\vec{v} = (A - 3\, Id)\vec{v} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 4 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \Leftrightarrow v_3 = 0; \quad v_1, v_2 \quad \text{beliebig.}$$

Es existieren zwei linear unabhängige Eigenvektoren, nämlich $\vec{v}_2 := v_1\,(1,0,0)^T$ und $\vec{v}_3 := v_2\,(0,1,0)^T$. Demzufolge haben wir

$$\mathrm{Kern}\,(A - \lambda_2\, Id) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

d.h. algebr. $\dim \lambda_2 =$ geom. $\dim \lambda_2$. Es trifft Satz 11.3.(d) zu. Die Eigenvektoren $\vec{v}_1 := (1, -4, 1)^T$, $\vec{v}_2 := (1, 0, 0)^T$ und $\vec{v}_3 := (0, 1, 0)^T$ bilden eine Basis des Vektorraumes $\mathbf{C}^3$. Die **Modalmatrix** von A lautet

$$T = \begin{bmatrix} 1 & 1 & 0 \\ -4 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}; \quad T^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 4 \end{bmatrix}.$$

11.2 Der Satz von Cayley–Hamilton

Zur Motivation greifen wir zunächst nochmals die Situation auf, in welcher der Raum $\mathbf{C}^n$ eine Basis aus Eigenvektoren einer Matrix $A \in \mathbf{K}^{(n,n)}$ besitzt. Es seien $\lambda_j,\ j = 1, 2, \ldots, m$, die *paarweise verschiedenen* Eigenwerte von A mit Vielfachheiten k_j, und seien $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{k_j}$ die dem Eigenwert λ_j zugeordneten Eigenvektoren. Dann folgt mit der Linearfaktorzerlegung (1.11) des charakteristischen Polynoms für jeden Index $l = 1, 2, \ldots, k_j$:

$$\left.\begin{aligned} P_n(A)\vec{v}_l &:= (-1)^n (A - \lambda_1\, Id)^{k_1} (A - \lambda_2\, Id)^{k_2} \cdots (A - \lambda_m\, Id)^{k_m} \vec{v}_l \\ &= (-1)^n (A - \lambda_1\, Id)^{k_1} \cdots \underbrace{Id}_{j\text{–ter Faktor}} \cdots (A - \lambda_m\, Id)^{k_m} \underbrace{(A - \lambda_j\, Id)^{k_j}}_{j\text{–ter Faktor}} \vec{v}_l \\ &= \vec{0}. \end{aligned}\right\} \qquad (2.1)$$

Hier haben wir die *Kommutator–Eigenschaft*

$$(A - \lambda_j\, Id)(A - \lambda_k\, Id) = A^2 - (\lambda_j + \lambda_k)A + \lambda_j\lambda_k\, Id = (A - \lambda_k\, Id)(A - \lambda_j\, Id) \quad (j \neq k)$$

eingearbeitet. Da die Gesamtheit der Eigenvektoren den Vektorraum $\mathbf{C}^n$ aufspannt, resultiert nun aus (2.1) $P_n(A) = O$. Das heißt, die Matrix A **erfüllt ihre eigene charakteristische Gleichung**. Die Eigenschaft gilt auch im allgemeinen Fall:

Satz 11.4 (Cayley–Hamilton)
Gegeben seien $A \in \mathbf{K}^{(n,n)}$ und dazu die paarweise verschiedenen Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbf{C}$ mit Vielfachheiten $k_1, k_2, \ldots, k_m$. Dann gilt:

$$\boxed{P_n(A) = (-1)^n (A - \lambda_1\, Id)^{k_1} (A - \lambda_2\, Id)^{k_2} \cdots (A - \lambda_m\, Id)^{k_m} = O,} \qquad (2.2)$$

das heißt, die Matrix A erfüllt ihre eigene charakteristische Gleichung.

Diesen (nicht vollständig) bewiesenen Satz belegen wir durch das folgende Beispiel.

BSP. (11.2.1) Es sei $A := \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ die Matrix aus BSP. (11.1.3), Abschnitt 11.1. Diese hat den doppelten Eigenwert $\lambda_1 = 1$. Man errechnet sofort, dass $\mathrm{Kern}\,(A - \lambda_1\, Id) = \mathrm{span}\left\{(1,0)^T\right\}$ gilt. Das heißt, wir haben $\rho(\lambda_1) = 1 < 2 = k_1$. Es ist aber $P_2(\lambda) = (-1)^2(\lambda - 1)^2 = (\lambda - 1)^2$, und auch

$$P_2(A) = (A - Id)^2 = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = O.$$

Die Beziehung (2.2) besagt insbesondere, dass die Eigenwerte λ_j der Matrix A vermöge P_n auf die Eigenwerte $P_n(\lambda_j) = 0$ der Matrix $P_n(A) = O$ abgebildet werden. Diese Transformationseigenschaft gilt allgemeiner:

Satz 11.5 *Gegeben seien ein Eigenwert λ der Matrix $A \in \mathbf{K}^{(n,n)}$ und der zugeordnete Eigenvektor $\vec{v}$.*
(a) *Für ein beliebiges Polynom $Q_m(\mu) := \sum_{j=0}^{m} q_j \mu^j$ hat die Matrix $Q_m(A) \in \mathbf{C}^{(n,n)}$ den Eigenvektor $\vec{v}$ zum Eigenwert $Q_m(\lambda)$. Insbesondere besitzen $q \cdot A$ den Eigenwert $q \cdot \lambda$ und $A + \mu\, Id$ den Eigenwert $\lambda + \mu$.*
(b) *Die* **transponierte** *Matrix A^T hat den Eigenwert λ, und die* **adjungierte** *Matrix $A^* = \overline{A}^T$ den Eigenwert $\overline{\lambda}$.*

Begründungen: (a) Klar, aus $A\vec{v} = \lambda\vec{v}$ folgt sofort $A^j\vec{v} = A^{j-1}\lambda\vec{v} = \cdots = \lambda^j\vec{v} \;\; \forall\, j \in \mathbf{N}_0$. Also gilt $Q_m(A)\vec{v} = \sum\limits_{j=0}^{m} q_j A^j\vec{v} = \sum\limits_{j=0}^{m} q_j\lambda^j\vec{v} = Q_m(\lambda)\vec{v}$.

(b) Die Determinantenregeln liefern:

$$\begin{aligned} \det(A - \lambda\, Id) &= \det[(A - \lambda\, Id)^T] = \det(A^T - \lambda\, Id), \\ \det(A^* - \overline{\lambda}\, Id) &= \det[(A - \lambda\, Id)^*] = \det[\,\overline{(A - \lambda\, Id)^T}\,] = \overline{\det(A - \lambda\, Id)}. \end{aligned}$$

Hieraus folgen die Behauptungen. □

Die Relation (2.2) ermöglicht es, die Potenz A^n der Matrix A durch die niederen Potenzen $A^{n-1}, A^{n-2}, \ldots, A$, Id auszudrücken. Allgemein kann somit jede Potenz von A als Linearkombination des Systems $Id, A, A^2, \ldots, A^{n-1}$ dargestellt werden. Auf diese Weise lassen sich zum Beispiel **matrixwertige Funktionen** über Potenzreihen erklären:

$$f_A(t) := \sum_{k=0}^{\infty} a_k (tA)^k, \;\; t \in D(f_A).$$

BSP. (11.2.2) Es sei $A := \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ die Matrix aus BSP. (11.2.1). Wir hatten gezeigt, dass $O = (A - Id)^2 = A^2 - 2A + Id$ gilt, also $A^2 = 2A - Id$. Hieraus resultiert:

$$A^3 = 2A^2 - A = 3A - 2\, Id, \quad A^4 = 3A^2 - 2A = 4A - 3\, Id, \ldots, \text{allgemein: } A^k = kA - (k-1)\, Id, \;\; k \in \mathbf{N}.$$

Nun ist die folgende Funktion sinnvoll definiert:

$$\begin{aligned} e^{tA} &:= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = Id + A \underbrace{\sum_{k=1}^{\infty} \frac{kt^k}{k!}}_{=te^t} - Id \underbrace{\sum_{k=2}^{\infty} \frac{(k-1)t^k}{k!}}_{=(t-1)e^t+1} \\ &= te^t A + (1-t)e^t\, Id = \begin{bmatrix} e^t & 3te^t \\ 0 & e^t \end{bmatrix} = e^t \begin{bmatrix} 1 & 3t \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

BSP. (11.2.3) Die Matrix $A := \begin{bmatrix} 2 & 2 \\ -1 & -2 \end{bmatrix}$ hat das charakteristische Polynom

$$P_2(\lambda) := \det(A - \lambda\, Id) = \begin{vmatrix} 2-\lambda & 2 \\ -1 & -2-\lambda \end{vmatrix} = \lambda^2 - 2.$$

Somit gilt hier $P_2(A) = O = A^2 - 2\, Id$. Daraus errechnet man sukzessive

$$A^2 = 2\, Id, \quad A^3 = 2A, \quad A^4 = 4\, Id, \ldots, \text{allgemein: } A^{2k} = 2^k\, Id, \quad A^{2k+1} = 2^k A, \; k \in \mathbf{N}_0,$$

und weiterhin

$$\begin{aligned} e^{tA} &:= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} A^{2k} + \sum_{k=0}^{\infty} \frac{t^{2k+1}}{(2k+1)!} A^{2k+1} = \sum_{k=0}^{\infty} \frac{(\sqrt{2}t)^{2k}}{(2k)!}\, Id + \frac{A}{\sqrt{2}} \sum_{k=0}^{\infty} \frac{(\sqrt{2}t)^{2k+1}}{(2k+1)!} \\ &= Id \cdot \cosh(\sqrt{2}t) + \frac{1}{\sqrt{2}} A \cdot \sinh(\sqrt{2}t) = \begin{bmatrix} \cosh(\sqrt{2}t) + \sqrt{2}\,\sinh(\sqrt{2}t) & \sqrt{2}\,\sinh(\sqrt{2}t) \\ -\frac{1}{2}\sqrt{2}\,\sinh(\sqrt{2}t) & \cosh(\sqrt{2}t) - \sqrt{2}\,\sinh(\sqrt{2}t) \end{bmatrix}. \end{aligned}$$

Definition 11.4 *Zu gegebener Matrix $A \in \mathbf{K}^{(n,n)}$ heiße die matrixwertige Funktion*

$$\boxed{e^{tA} := \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k, \;\; t \in \mathbf{R},\; A^0 := Id,} \tag{2.3}$$

die **Matrix–Exponentialfunktion** *von A.*

Bemerkung 11.5 Ist $\|\cdot\| : \mathbf{K}^{(n,n)} \to \mathbf{R}$ eine *submultiplikative Matrixnorm*, so gilt ja $\|A^2\| \leq \|A\|^2$, und weiter durch Induktion $\|A^k\| \leq \|A\|^k$. Hieraus erschließen wir unter Verwendung der Dreiecksungleichung:

$$\|e^{tA}\| \leq \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \|A\|^k = e^{|t|\,\|A\|} < +\infty \;\forall\, t \in \mathbf{R}.$$

Dies begründet die Existenz der Matrix–Exponentialfunktion für jedes $t \in \mathbf{R}$. □

Es sei $\vec{v} \in \mathbf{K}^n$ ein fester Vektor. Dann wird durch

$$\vec{y}(t) := e^{tA}\vec{v} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \vec{v}, \quad \vec{y}(0) = \vec{v}, \tag{2.4}$$

eine vektorwertige Funktion $\vec{y} : \mathbf{R} \to \mathbf{K}^n$ erklärt. Die Reihe (2.4) darf gliedweise differenziert werden, und es folgt

$$\frac{d}{dt}\,\vec{y}(t) = \sum_{k=1}^{\infty} \frac{k t^{k-1}}{k!} A^k \vec{v} = A \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \vec{v} = A\vec{y}(t).$$

Satz 11.6 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Dann hat das Anfangswertproblem*

$$\boxed{\vec{y}\,'(t) = A\vec{y}(t), \quad \vec{y}(t_0) = \vec{y}_0 \in \mathbf{K}^n,} \tag{2.5}$$

die eindeutig bestimmte Lösung $\vec{y}(t) = e^{(t-t_0)A}\vec{y}_0$.

Begründung: Offensichtlich ist die Funktion $\vec{y}(t)$ eine Lösung, und wir brauchen nur ihre Eindeutigkeit zu zeigen. Wäre auch $\vec{y}_1(t)$ eine Lösung, so könnten wir für festes $t > t_0$ setzen:

$$\vec{w}(s) := e^{(t-s)A}\vec{y}_1(s), \quad t_0 \leq s \leq t.$$

Durch Differentiation folgt nun

$$\frac{d}{ds}\,\vec{w}(s) = e^{(t-s)A}\Big(- A\vec{y}_1(s) + \vec{y}\,'_1(s)\Big) = \vec{0},$$

und somit $\vec{w}(s) = const \;\forall\, t_0 \leq s \leq t$. Das heißt,

$$\vec{w}(t_0) = e^{(t-t_0)A}\vec{y}_1(t_0) = e^{(t-t_0)A}\vec{y}_0 = \vec{y}(t) = \vec{w}(t) = e^{0\,A}\vec{y}_1(t) = \vec{y}_1(t).$$

Wir haben also $\vec{y}(t) = \vec{y}_1(t)$, und somit Eindeutigkeit vorliegen. □

Die Berechnung der Matrix–Exponentialfunktion e^{tA} kann im Fall $n \geq 3$ doch mit erheblichem Rechenaufwand verbunden sein, da es am konkreten Beispiel $A \in \mathbf{K}^{(n,n)}$ nicht einfach ist, die Potenzen A^k allgemein zu berechnen. Die Situation vereinfacht sich ganz beträchtlich, wenn die Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ der Matrix A eine **Basis des $\mathbf{C}^n$** bilden. Sind nämlich $\lambda_1, \lambda_2, \ldots, \lambda_n$ die (nicht notwendig voneinander verschiedenen) Eigenwerte von A, so gilt:

$$\boxed{e^{tA}\vec{v}_j = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \vec{v}_j = \sum_{k=0}^{\infty} \frac{t^k \lambda_j^k}{k!}\,\vec{v}_j = e^{t\,\lambda_j}\,\vec{v}_j, \quad j = 1, 2, \ldots, n.}$$

Da nun jeder Vektor $\vec{y}_0 \in \mathbf{C}^n$ in der Basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ aufgespannt werden kann, resultiert aus Satz 11.6 sofort:

Satz 11.7 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$, deren Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ eine* **Basis** *des Vektorraumes $\mathbf{C}^n$ bilden mögen. Es bezeichne λ_j den Eigenwert, der dem Eigenvektor $\vec{v}_j$ zugeordnet ist. Dann hat das Anfangswertproblem (2.5) die eindeutig bestimmte Lösung*

$$\boxed{\vec{y}(t) = C_1 e^{\lambda_1 (t-t_0)} \vec{v}_1 + C_2 e^{\lambda_2 (t-t_0)} \vec{v}_2 + \cdots + C_n e^{\lambda_n (t-t_0)} \vec{v}_n.} \tag{2.6}$$

Die Konstanten C_j sind die Koeffizienten der Linearkombination $\vec{y}_0 = C_1\vec{v}_1 + C_2\vec{v}_2 + \cdots + C_n\vec{v}_n$.

BSP. (11.2.4) Es sei $A := \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix}$ die Matrix aus BSP. (11.1.6), Abschnitt 11.1. Dort wurden bereit die Eigenwerte $\lambda_1 = \lambda_2 = 3$, $\lambda_3 = 2$ und die zugeordneten Eigenvektoren $\vec{v}_1 = (1,0,0)^T$, $\vec{v}_2 = (0,1,0)^T$, $\vec{v}_3 = (1,-4,1)^T$ berechnet. Wir lösen das Anfangswertproblem

$$\vec{y}\,'(t) = A\vec{y}(t), \quad \vec{y}(0) = \vec{y}_0 = (a,b,c)^T \in \mathbf{R}^3.$$

Gemäß Satz 11.7 hat das DGl–System die allgemeine Lösung

$$\boxed{\vec{y}(t) = \Big(C_1\vec{v}_1 + C_2\vec{v}_2\Big)\, e^{3t} + C_3 e^{2t}\vec{v}_3.}$$

Nun gilt offensichtlich

$$\vec{y}_0 = \underbrace{(a-c)}_{=:C_1}\vec{v}_1 + \underbrace{(4c+b)}_{=:C_2}\vec{v}_2 + \underbrace{c}_{=:C_3}\vec{v}_3,$$

und aus dieser Zerlegung erhält man die folgende Lösung des Anfangswertproblems:

$$\vec{y}(t) = (a-c)e^{3t}\vec{v}_1 + (4c+b)e^{3t}\vec{v}_2 + ce^{2t}\vec{v}_3 = e^{3t}\begin{bmatrix} a-c \\ 4c+b \\ 0 \end{bmatrix} + e^{2t}\begin{bmatrix} c \\ -4c \\ c \end{bmatrix}.$$

Sind **nicht genügend viele Eigenvektoren** vorhanden, so gestaltet sich die Lösung des Anfangswertproblems (2.5) wesentlich schwieriger. Dieser Fall ist in Ing.–Math.III zu behandeln.

11.3 Ähnliche Matrizen

Offenbar können bei Diagonalmatrizen $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ Eigenwerte und Eigenvektoren direkt abgelesen werden. Wir untersuchen deshalb die Aufgabe, eine Matrix $A \in \mathbf{K}^{(n,n)}$ mit Hilfe geeigneter Transformationen auf Diagonalgestalt zu bringen, und zwar unter Beibehaltung der Eigenwerte. Anlass dazu bietet die Relation (1.13), nach der die Matrizen A und $T^{-1}AT$ für beliebiges $T \in \operatorname{Inv}(\mathbf{C}^n)$ dasselbe charakteristische Polynom haben.

Definition 11.5 *Zwei Matrizen $A, B \in \mathbf{K}^{(n,n)}$ heißen* **ähnlich**, *wenn es eine Transformationsmatrix $T \in \operatorname{Inv}(\mathbf{C}^n)$ gibt mit*

$$\boxed{B = T^{-1}AT.} \tag{3.1}$$

Die Beziehung (3.1) heißt **Ähnlichkeitstransformation.** *Eine Matrix $A \in \mathbf{K}^{(n,n)}$ heiße* **diagonalähnlich** (*kurz:* **diagonalisierbar**), *falls A ähnlich mit einer Diagonalmatrix Λ ist.*

Bemerkung 11.6 Die Matrix T in (3.1) ist nicht eindeutig bestimmt. Gibt es ein solches T, so leistet auch $\tilde{T} := \lambda \cdot T$ für jedes $\lambda \neq 0$ das in (3.1) Verlangte. □

Der folgende Satz gibt ein Kriterium für die Diagonalisierbarkeit einer Matrix an.

Satz 11.8 (a) *Sind $A, B \in \mathbf{K}^{(n,n)}$ ähnlich, so gilt*

$$\boxed{\det(A - \lambda\, Id) = \det(B - \lambda\, Id)\ \forall\ \lambda \in \mathbf{C};}$$

das heißt, die Matrizen A und B haben dieselbe Determinante und dasselbe charakteristische Polynom und somit dieselben Eigenwerte.

(b) *Ist $B = T^{-1}AT$, und ist $\vec{v}$ ein Eigenvektor von B zum Eigenwert λ, so ist $T\vec{v}$ ein Eigenvektor von A zum selben Eigenwert λ.*

(c) *Genau dann ist A diagonalisierbar, wenn die Eigenvektoren von A eine Basis des $\mathbf{C}^n$ bilden. Bezeichnet $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ diese Basis, so gilt mit der* **Modalmatrix** *$T := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ die Beziehung*

$$\boxed{T^{-1}AT = \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),} \tag{3.2}$$

sofern die Zuordnung $A\vec{v}_j = \lambda_j \vec{v}_j\ \forall\ j = 1, \ldots, n$, angenommen wird. Die Matrix Λ heißt die **Spektralmatrix** *von A.*

Begründungen: (a) ergibt sich unmittelbar aus (1.13).

(b) Aus $\vec{0} \neq \vec{v}$ und $B\vec{v} = \lambda\vec{v}$ resultiert $AT\vec{v} = TB\vec{v} = T\lambda\vec{v} = \lambda T\vec{v}$.

(c) Ist A diagonalisierbar, so existieren $T := (\vec{t}_1, \vec{t}_2, \ldots, \vec{t}_n) \in \mathbf{C}^{(n,n)}$ und $\Lambda := \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbf{C}^{(n,n)}$ mit der Beziehung (3.2). Gemäß (a) hat A die Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_n$, und gemäß (b) sind $T\vec{e}_1 = \vec{t}_1, \ldots, T\vec{e}_n = \vec{t}_n$ die zugeordneten Eigenvektoren. Wegen $\operatorname{Rang} T = n$ spannen diese den $\mathbf{C}^n$ auf.
Es seien umgekehrt $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ Eigenvektoren von A, die den $\mathbf{C}^n$ aufspannen. Ist dann $T := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ die Modalmatrix und Λ die Spektralmatrix von A, so gilt (1.15), oder äquivalent $T^{-1}AT = \Lambda$. □

Bemerkung 11.7 Sind alle n Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$ *voneinander verschieden*, so folgt aus Satz 11.3.(b), dass die Eigenvektoren eine Basis des $\mathbf{C}^n$ bilden. Also ist A *diagonalisierbar.* □

BSP. (11.3.1) Zur Matrix $A := \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix}$ hatten wir bereits in Abschnitt 11.1, BSP. (11.1.6), die Modalmatrix T und ihre Inverse berechnet. Es folgt

$$T^{-1}AT = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ -4 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \Lambda.$$

Das Problem der Diagonalisierbarkeit kann weiter spezialisiert werden, indem von der Modalmatrix T weitere Eigenschaften gefordert werden.

Definition 11.6 (a) *Die Matrix $Q \in \mathbf{K}^{(n,n)}$ heiße* **orthogonal** *($\mathbf{K} := \mathbf{R}$) bzw.* **unitär** *($\mathbf{K} := \mathbf{C}$), wenn gilt*

$$\boxed{Q^* = Q^{-1}.}$$

(b) *Die Matrizen $A, B \in \mathbf{K}^{(n,n)}$ heißen* **unitär ähnlich**, *wenn es eine unitäre Transformationsmatrix $Q \in \mathbf{C}^{(n,n)}$ gibt mit*

$$\boxed{B = Q^*AQ.} \tag{3.3}$$

(c) *Die Matrix A heiße* **unitär diagonalisierbar** *oder* **normal**, *wenn eine Diagonalmatrix Λ und eine unitäre Matrix Q existieren mit*

$$\boxed{\Lambda = Q^* A Q.} \tag{3.4}$$

Eine Matrix $Q = (\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n) \in \mathbf{C}^{(n,n)}$ ist gemäß Satz 5.22.(d) genau dann unitär, wenn die Spaltenvektoren $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n$ eine **Orthonormalbasis** (ON–Basis) des $\mathbf{C}^n$ bilden: Bezeichnet $\langle \cdot, \cdot \rangle$ das *Standardskalarprodukt* des $\mathbf{C}^n$, so gilt $\langle \vec{q}_j, \vec{q}_k \rangle = \delta_{jk} \ \forall \ j,k = 1,2,\ldots,n$. Ist also A normal, so sind die Vektoren $\vec{q}_j = Q\vec{e}_j \ \forall \ j = 1,2,\ldots,n$, gemäß Satz 11.8.(b) die Eigenvektoren von A. Bilden umgekehrt die Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots \vec{v}_n$ von A eine ON–Basis des $\mathbf{C}^n$, so ist die Modalmatrix $Q := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ unitär, und es gilt (3.4). Zusammenfassend erhält man den ersten Teil des folgenden Satzes:

Satz 11.9 (a) *Genau dann ist $A \in \mathbf{K}^{(n,n)}$ normal, wenn die Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ von A eine ON–Basis des $\mathbf{C}^n$ bilden.*

(b) *Ist $A \in \mathbf{K}^{(n,n)}$ normal, so gestattet die Matrix A die* **Spektralzerlegung**

$$\boxed{A = \sum_{k=0}^{n} \lambda_j \left(\vec{v}_j \otimes \vec{v}_j \right).}$$

Hierin bezeichnet $\lambda_1, \lambda_2, \ldots, \lambda_n$ die Eigenwerte der Matrix A und $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ das zugeordnete ON–System der Eigenvektoren.

Begründung: (b) Da jeder Vektor $\vec{x} \in \mathbf{C}^n$ in der ON–Basis die eindeutige Zerlegung $\vec{x} = C_1\vec{v}_1 + C_2\vec{v}_2 + \cdots + C_n\vec{v}_n$ hat, gilt einerseits $A\vec{x} = C_1\lambda_1\vec{v}_1 + C_2\lambda_2\vec{v}_2 + \cdots + C_n\lambda_n\vec{v}_n$. Andererseits gilt auch

$$\sum_{j=1}^{n} \lambda_j \Big(\vec{v}_j \otimes \vec{v}_j \Big) \vec{v}_k = \sum_{j=1}^{n} \lambda_j \, \langle \vec{v}_k, \vec{v}_j \rangle \, \vec{v}_j = \lambda_k \, \vec{v}_k, \quad k = 1,2,\ldots,n,$$

und somit $\sum\limits_{j=1}^{n} \lambda_j \Big(\vec{v}_j \otimes \vec{v}_j \Big) \vec{x} = C_1\lambda_1\vec{v}_1 + C_2\lambda_2\vec{v}_2 + \cdots + C_n\lambda_n\vec{v}_n.$ □

BSP. (11.3.2) Die Matrix $A := \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ hat das charakteristische Polynom

$$P_3(\lambda) := \det(A - \lambda\, Id) = \begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 2-\lambda & 1 \\ 0 & 1 & 1-\lambda \end{vmatrix} = (1-\lambda)\lambda(\lambda - 3).$$

Wir haben drei verschiedene Eigenwerte $\lambda_1 = 0, \ \lambda_2 = 1, \ \lambda_3 = 3$, und dazu gehören drei normierte Eigenvektoren

$$\vec{v}_1 := \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}; \quad \vec{v}_2 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}; \quad \vec{v}_3 := \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},$$

die offenbar eine ON–Basis des $\mathbf{C}^3$ bilden. Man erhält weiterhin mit $Q := (\vec{v}_1, \vec{v}_2, \vec{v}_3)$ die behauptete Relation

$$Q^*AQ = \frac{1}{6} \begin{bmatrix} \sqrt{2} & -\sqrt{2} & \sqrt{2} \\ \sqrt{3} & 0 & -\sqrt{3} \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ -\sqrt{2} & 0 & 2 \\ \sqrt{2} & -\sqrt{3} & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \Lambda.$$

Wie das Beispiel zeigt, ist es recht mühsam, Normalität einer Matrix A über die Orthonormalität der Eigenvektoren nachzuprüfen. Ein einfaches Kriterium liefert der

Satz 11.10 *Genau dann ist $A \in \mathbf{K}^{(n,n)}$ normal, wenn $A^*A = AA^*$ gilt.*

Begründung: Wir zeigen, dass die Bedingung $A^*A = AA^*$ **notwendig** für die Normalität von A ist. Auf den Nachweis der Hinlänglichkeit soll hier verzichtet werden. Wir verweisen auf die Literatur (z.B.J. Stoer/R. Bulirsch, Einführung in die Numerische Mathematik II, dort Satz (6.4.3)). Sei also A normal. Dann gilt (3.4), und somit auch

$$Q\Lambda Q^* = A;\ Q\Lambda^* Q^* = A^*;\ AA^* = Q\Lambda Q^* Q\Lambda^* Q^* = Q\Lambda\Lambda^* Q^*;\ A^*A = Q\Lambda^*\Lambda Q^*.$$

Wegen $\Lambda\Lambda^* = \Lambda^*\Lambda$ folgt $AA^* = A^*A$. □

BSP. (11.3.3) Für die Matrizen

$$A := \begin{bmatrix} 2\,i & 0 & i \\ 0 & 1 & 0 \\ i & 0 & 2\,i \end{bmatrix}, \qquad A^* = \begin{bmatrix} -2\,i & 0 & -i \\ 0 & 1 & 0 \\ -i & 0 & -2\,i \end{bmatrix}$$

hat man ganz offensichtlich $AA^* = A^*A$, wie folgende Rechnung zeigt:

$$AA^* = \begin{bmatrix} 2\,i & 0 & i \\ 0 & 1 & 0 \\ i & 0 & 2\,i \end{bmatrix} \begin{bmatrix} -2\,i & 0 & -i \\ 0 & 1 & 0 \\ -i & 0 & -2\,i \end{bmatrix} = \begin{bmatrix} 5 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 5 \end{bmatrix} = \begin{bmatrix} -2\,i & 0 & -i \\ 0 & 1 & 0 \\ -i & 0 & -2\,i \end{bmatrix} \begin{bmatrix} 2\,i & 0 & i \\ 0 & 1 & 0 \\ i & 0 & 2\,i \end{bmatrix} = A^*A.$$

Gemäß Satz 11.10 ist A normal. Tatsächlich hat A die Eigenwerte $\lambda_1 = 1;\ \lambda_2 = i;\ \lambda_3 = 3\,i$, und dazu gehören die drei normierten Eigenvektoren

$$\vec{v}_1 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix};\quad \vec{v}_2 := \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix};\quad \vec{v}_3 := \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Diese bilden eine ON–Basis des $\mathbf{C}^3$.

Sonderfall I. Gelte $A = A^*$, das heißt, $A \in \mathbf{K}^{(n,n)}$ sei **symmetrisch** ($\mathbf{K} := \mathbf{R}$) bzw. **hermitesch** ($\mathbf{K} := \mathbf{C}$).

Satz 11.11 *Es sei $A = A^* \in \mathbf{K}^{(n,n)}$. Dann gelten die folgenden Aussagen:*

(a) *A ist normal.*

(b) *Es gibt eine ON–Basis des $\mathbf{C}^n$ aus Eigenvektoren von A.*

(c) *Alle Eigenwerte von A sind reell.*

Begründungen: (a) Wegen $A^* = A$ gilt trivialerweise $AA^* = A^*A$. Man verwende nun Satz 11.10. Aussage (b) ist eine Folgerung aus Satz 11.9. Zu (c): Für einen Eigenwert $\lambda = 0$ ist nichts zu zeigen. Sei also $\lambda \neq 0$ Eigenwert von A, und sei $\vec{0} \neq \vec{v}$ zugeordneter Eigenvektor. Dann folgt

$$\lambda\langle\vec{v},\vec{v}\rangle = \langle A\vec{v},\vec{v}\rangle = \langle\vec{v},A^*\vec{v}\rangle = \langle\vec{v},A\vec{v}\rangle = \overline{\lambda}\langle\vec{v},\vec{v}\rangle,$$

also $\lambda = \overline{\lambda}$. □

BSP. (11.3.4) Gegeben sei die symmetrische Matrix $A := \begin{bmatrix} 3 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & 3 \end{bmatrix} = A^*$, deren charakteristisches Polynom wie folgt lautet:

$$P_3(\lambda) = \det(A - \lambda\,Id) = \begin{vmatrix} 3-\lambda & 0 & -1 \\ 0 & 4-\lambda & 0 \\ -1 & 0 & 3-\lambda \end{vmatrix} = -(\lambda-4)^2(\lambda-2).$$

Es liegen die Eigenwerte $\lambda_1 = 2$ (einfach) und $\lambda_2 = 4$ (doppelt) vor. Man erhält mit einfacher Rechnung

$$\mathrm{Kern}\,(A - \lambda_1\,Id) = \mathrm{span}\left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Zur Bestimmung von Kern $(A - \lambda_2\, Id)$ ist das homogene Gleichungssystem

$$\vec{0} = (A - \lambda_2\, Id)\vec{v} = (A - 4\, Id)\vec{v} = \begin{bmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \vec{0}$$

zu lösen. Offensichtlich ist v_2 beliebig wählbar, während $v_1 = -v_3$ gelten muss. Mit $v_2 = 0$ und $v_1 = 1$ erhält man den Eigenvektor

$$\vec{v}_2 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \perp \vec{v}_1 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Da wir aus Satz 11.11 wissen, dass ein dritter Eigenvektor senkrecht auf $\vec{v}_1$ und $\vec{v}_2$ stehen muss, setzen wir

$$\vec{v}_3 = \vec{v}_1 \times \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \in \text{Kern}\,(A - \lambda_2\, Id).$$

Die orthogonale Modalmatrix $Q := (\vec{v}_1, \vec{v}_2, \vec{v}_3)$ liefert jetzt die gewünschte Diagonalisierung

$$Q^* A Q = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \Lambda.$$

Das obige Beispiel zeigt, dass die Eigenräume symmetrischer Matrizen paarweise senkrecht zueinander sind. Dies gilt allgemeiner für normale Matrizen.

Satz 11.12 *Gegeben sei die normale Matrix $A \in \mathbf{K}^{(n,n)}$. Bezeichne $\|\vec{v}\|_2 := \sqrt{\langle \vec{v}, \vec{v} \rangle}$, $\vec{v} \in \mathbf{C}^n$, die Euklidische Norm in $\mathbf{C}^n$. Dann gilt:*

(a) $\|A\vec{v}\|_2 = \|A^*\vec{v}\|_2\ \forall\ \vec{v} \in \mathbf{C}^n$ *und* $\text{Kern}\,(A - \lambda\, Id) = \text{Kern}\,(A^* - \overline{\lambda}\, Id)\ \forall\ \lambda \in \mathbf{C}$.

(b) *Für verschiedene Eigenwerte $\lambda_1 \neq \lambda_2$ von A gilt* $\text{Kern}\,(A - \lambda_1\, Id) \perp \text{Kern}\,(A - \lambda_2\, Id)$.

Begründungen: (a) Für beliebiges $\vec{v} \in \mathbf{C}^n$ haben wir wegen $A^*A = AA^*$:

$$\|A\vec{v}\|_2^2 = \langle A\vec{v}, A\vec{v} \rangle = \langle \vec{v}, A^*A\vec{v} \rangle = \langle \vec{v}, AA^*\vec{v} \rangle = \langle A^*\vec{v}, A^*\vec{v} \rangle = \|A^*\vec{v}\|_2^2.$$

Sei nun $\lambda \in \mathbf{C}$, und sei $A_\lambda := A - \lambda\, Id$ gesetzt. Dann gilt

$$A_\lambda (A_\lambda)^* = (A - \lambda\, Id)(A^* - \overline{\lambda}\, Id) = AA^* - (\lambda A^* + \overline{\lambda} A) + \lambda\overline{\lambda}\, Id = (A_\lambda)^*\, A_\lambda.$$

Da also A_λ normal ist, folgt aus dem ersten Teil der Behauptung

$$\|(A - \lambda\, Id)\vec{v}\|_2 = \|(A^* - \overline{\lambda}\, Id)\vec{v}\|_2\ \ \forall\ \vec{v} \in \mathbf{C}^n,$$

mithin $\text{Kern}\, A_\lambda = \text{Kern}\,(A_\lambda)^*$.

(b) Wir wählen $\vec{v}_1 \in \text{Kern}\,(A - \lambda_1\, Id) = \text{Kern}\,(A^* - \overline{\lambda}_1\, Id)$ und $\vec{v}_2 \in \text{Kern}\,(A - \lambda_2\, Id)$. Wir erhalten $\langle \vec{v}_2, \vec{v}_1 \rangle = 0$ aus folgender Rechnung:

$$\lambda_2 \langle \vec{v}_2, \vec{v}_1 \rangle = \langle A\vec{v}_2, \vec{v}_1 \rangle = \langle \vec{v}_2, A^*\vec{v}_1 \rangle = \lambda_1 \langle \vec{v}_2, \vec{v}_1 \rangle.$$

Sonderfall II. $A \in \mathbf{K}^{(n,n)}$ sei **positiv definit**, das heißt, es gelte $A = A^*$ und $\langle A\vec{x}, \vec{x} \rangle > 0\ \forall\ \vec{0} \neq \vec{x} \in \mathbf{K}^n$.

Satz 11.13 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Dann gelten die folgenden Aussagen:*

(a) *A positiv definit* $\quad \Rightarrow \quad$ *A normal.*

(b) *A positiv definit genau, wenn $A = A^*$ gilt und alle Eigenwerte von A positiv sind.*

(c) $A \in \text{Inv}\,(\mathbf{K}^n) \quad \Rightarrow \quad$ *A^*A positiv definit.*

Begründungen: (a) Wegen $A = A^*$ ist jede positiv definite Matrix trivialerweise normal.

(b) Gemäß Satz 11.11.(b) gibt es eine ON–Basis des $\mathbf{C}^n$ aus Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ von A. Seien $\lambda_1, \lambda_2, \ldots, \lambda_n$ die zugeordneten Eigenwerte. Für jeden Vektor $\vec{x} \in \mathbf{C}^n,\ \vec{x} \neq \vec{0}$, gilt eine Zerlegung $\vec{x} = \sum_{j=1}^{n} \alpha_j \vec{v}_j$ mit $\sum_{j=1}^{n} |\alpha_j|^2 > 0$. Wir folgern

$$\langle A\vec{x}, \vec{x}\rangle = \Big\langle \sum_{j=1}^{n} \alpha_j \lambda_j \vec{v}_j, \sum_{k=1}^{n} \alpha_k \vec{v}_k \Big\rangle = \sum_{j=1}^{n} \lambda_j |\alpha_j|^2.$$

Man erkennt, $\langle A\vec{x}, \vec{x}\rangle > 0\ \forall\, \vec{0} \neq \vec{x} \in \mathbf{C}^n$ gilt genau dann, wenn $\lambda_j > 0\ \forall\, j = 1, \ldots, n$, erfüllt ist.

(c) Wegen $A \in \mathrm{Inv}\,(\mathbf{K}^n)$ gilt $A\vec{x} \neq \vec{0}\ \forall\, \vec{0} \neq \vec{x} \in \mathbf{K}^n$, und somit $\langle A^*A\vec{x}, \vec{x}\rangle = \|A\vec{x}\|_2^2 > 0$. □

Bemerkung 11.8 Wenn A positiv definit ist, kann $\lambda = 0$ **kein** Eigenwert sein. Das heißt, es gilt $\mathrm{Kern}\, A = \{\vec{0}\}$, und somit $A \in \mathrm{Inv}\,(\mathbf{K}^n)$. □

BSP. (11.3.5) Die Matrix $A := \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ mit dem charakteristischen Polynom

$$P_3(\lambda) = \det(A - \lambda\, Id) = \begin{vmatrix} 2-\lambda & 0 & 1 \\ 0 & 2-\lambda & 0 \\ 1 & 0 & 2-\lambda \end{vmatrix} = (2-\lambda)(\lambda-1)(\lambda-3)$$

hat die drei positiven Eigenwerte $\lambda_1 = 1,\ \lambda_2 = 2,\ \lambda_3 = 3$. Da A symmetrisch ist, muss A auch positiv definit sein. Tatsächlich gilt

$$\begin{aligned} \langle A\,(x_1, x_2, x_3)^T,\ (x_1, x_2, x_3)^T\rangle &= \langle (2x_1 + x_3,\ x_2,\ x_1 + 2x_3)^T,\ (x_1, x_2, x_3)^T\rangle \\ &= |x_1 + x_3|^2 + |x_1|^2 + |x_2|^2 + |x_3|^2 > 0\ \forall\, \vec{0} \neq \vec{x} \in \mathbf{K}^3. \end{aligned}$$

Sonderfall III. Es gelte $A^* = A^{-1}$, das heißt, $A \in \mathbf{K}^{(n,n)}$ sei **orthogonal** ($\mathbf{K} := \mathbf{R}$) bzw. **unitär** ($\mathbf{K} := \mathbf{C}$).

Satz 11.14 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Dann gelten die folgenden Aussagen:*
(a) *$A^* = A^{-1}$ impliziert A normal.*
(b) *Genau dann gilt $A^* = A^{-1}$, wenn $AA^* = Id$ erfüllt ist.*
(c) *$A^* = A^{-1}$ impliziert, dass alle Eigenwerte λ von A auf der* **Einheitskreislinie** *liegen: $|\lambda| = 1$.*

Begründungen: (a) Klar, es gilt $AA^* = AA^{-1} = Id = A^{-1}A = A^*A$. Also ist A normal.
(b) Ist $A^* = A^{-1}$, so gilt trivialerweise $AA^* = Id$, siehe (a). Ist andererseits $AA^* = Id$, so erhalten wir $1 = \det(AA^*) = |\det A|^2$. Also gilt $\det A \neq 0$, mithin $A \in \mathrm{Inv}\,(\mathbf{K}^n)$. Wir folgern $A^{-1} = A^{-1}\, Id = A^{-1}(AA^*) = A^*$.
(c) Sei λ Eigenwert von A und $\vec{0} \neq \vec{v}$ zugeordneter Eigenvektor. Dann gilt:

$$\|\vec{v}\|_2^2 = \langle A^*A\vec{v}, \vec{v}\rangle = \langle A\vec{v}, A\vec{v}\rangle = \langle \lambda\vec{v}, \lambda\vec{v}\rangle = |\lambda|^2\, \|\vec{v}\|_2^2,$$

also $|\lambda| = 1$. □

BSP. (11.3.6) Die Matrix $A \in \mathbf{R}^{(3,3)}$ mit

$$A := \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad A^* = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad AA^* = Id,$$

ist für $\alpha \in \mathbf{R}$ offensichtlich *orthogonal*. Wegen

$$P_3(\lambda) = \det(A - \lambda\, Id) = \begin{vmatrix} \cos\alpha - \lambda & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha - \lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)(\lambda - e^{i\alpha})(\lambda - e^{-i\alpha}),$$

liegen die drei Eigenwerte $\lambda_1 = 1,\ \lambda_2 = e^{i\alpha},\ \lambda_3 = e^{-i\alpha}$ vor, und es gilt $|\lambda_j| = 1$. Für $\alpha = 0$ haben wir $\lambda_1 = \lambda_2 = \lambda_3 = 1$, für $\alpha = \pi$ resultiert $\lambda_1 = 1,\ \lambda_2 = \lambda_3 = -1$.

Bemerkung 11.9 (a) Das charakteristische Polynom $P_n(\lambda) = \det(A - \lambda\, Id)$ einer *reellen* Matrix $A \in \mathbf{R}^{(n,n)}$ hat offenbar nur reelle Koeffizienten. Also können *komplexe* Eigenwerte nur als *konjugierte Paare* auftreten:

$$\boxed{\begin{array}{l} \lambda \text{ ist Eigenwert von } A \in \mathbf{R}^{(n,n)} \text{ und } \vec{v} \text{ ist zugeordneter Eigenvektor} \\ \Leftrightarrow \quad \overline{\lambda} \text{ ist Eigenwert von } A \text{ und } \overline{(\vec{v})} \text{ ist zugeordneter Eigenvektor.} \end{array}}$$

Klar, wegen $A = \overline{A}$ gilt $A\,\overline{(\vec{v})} = \overline{(A\vec{v})} = \overline{(\lambda\vec{v})} = \overline{\lambda}\ \overline{(\vec{v})}$. Speziell hat eine orthogonale Matrix $A \in \mathbf{R}^{(3,3)}$ entweder **einen** oder **drei** Eigenwerte mit den Werten ± 1.

(b) Ist $A \in \mathbf{K}^{(n,n)}$ diagonalisierbar: $\Lambda = T^{-1}AT$, so ist jede Potenz $A^k,\ k \in \mathbf{N}$, mit derselben Transformationsmatrix T diagonalisierbar. Es gilt nämlich

$$(T^{-1}AT)^k = (T^{-1}A\underbrace{T)(T^{-1}}_{=Id}A\underbrace{T)(}_{=Id}\cdots)\underbrace{(T^{-1}}_{=Id}AT) = T^{-1}A^kT.$$

Folglich gilt:

$$\boxed{\Lambda^k = \operatorname{diag}(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k) = T^{-1}A^kT \ \ \forall\ k \in \mathbf{N}.} \tag{3.5}$$

Ist $A \in \operatorname{Inv}(\mathbf{K}^n)$, so bleibt (3.5) sogar für alle $k \in \mathbf{Z}$ richtig, also insbesondere auch für negative Potenzen. Allgemein gilt auf Grund von Satz 11.5 sogar für jedes beliebige Polynom $Q_m(\mu) := \sum_{j=0}^{m} q_j\mu^j$ die Beziehung □

$$\boxed{Q_m(\Lambda) = \operatorname{diag}[Q_m(\lambda_1), Q_m(\lambda_2), \ldots, Q_m(\lambda_n)] = T^{-1}Q_m(A)T.} \tag{3.6}$$

11.4 Die Schursche Normalform einer Matrix

Für **nichtnormale** Matrizen $A \in \mathbf{K}^{(n,n)}$ gelten einige Eigenschaften nicht mehr, die auf normale Matrizen zutreffen, zum Beispiel:

- Zu verschiedenen Eigenwerten der Matrix A gehörende Eigenvektoren sind nicht notwendig orthogonal, wohl aber linear unabhängig.
- Es existieren nicht notwendig n linear unabhängige Eigenvektoren der Matrix A.
- Im allgemeinen hat man für die Eigenwerte λ der Matrix A Ungleichheit von algebraischer und geometrischer Dimension.
- Die Matrix A braucht nicht diagonalisierbar zu sein.

Insbesondere der letzte Punkt macht es im allgemeinen Fall unmöglich, den Vektorraum $\mathbf{C}^n$ durch eine Basis von Eigenvektoren der Matrix A aufzuspannen. An die Stelle der Diagonalisierbarkeit einer Matrix A tritt aber die folgende Transformationseigenschaft auf obere Dreiecksgestalt.

Satz 11.15 *Zu jeder gegebenen Matrix $A \in \mathbf{K}^{(n,n)}$ existiert eine Matrix $T \in \mathrm{Inv}\,(\mathbf{C}^n)$ mit*

$$T^{-1}AT = \begin{bmatrix} \lambda_1 & * & * & \cdots & * \\ & \lambda_2 & * & \cdots & * \\ & & \lambda_3 & \cdots & * \\ & & & \ddots & \vdots \\ O & & & & \lambda_n \end{bmatrix} =: D_0 \in U_R(\mathbf{C}^n). \tag{4.1}$$

Dabei bezeichnen wir mit λ_j, $j = 1, 2, \ldots, n$, die nicht notwendig verschiedenen Eigenwerte der Matrix A.

Bemerkung 11.10 (a) Die obere Dreiecksmatrix D_0 heißt **Schursche Normalform** der Matrix A. Die Matrizen $T \in \mathrm{Inv}\,(\mathbf{C}^n)$ und $D_0 \in U_R(\mathbf{C}^n)$ sind *nicht eindeutig* festgelegt.

(b) Es kann sogar speziell eine *unitäre* Matrix $Q \in \mathbf{C}^{(n,n)}$ mit $Q^*AQ = D_0 \in U_R(\mathbf{C}^n)$ bestimmt werden, vgl. J. Stoer/R. Bulirsch, Einführung in die Numerische Mathematik II, Satz (6.4.1). □

Begründung für Satz 11.15: Wir führen *vollständige Induktion* nach der Raumdimension n durch. Hierbei ist $n = 1$ der triviale Fall. Der Satz gelte jetzt bereits für Matrizen $A \in \mathbf{K}^{(n-1,n-1)}$.

Vererbung: Es sei λ_1 ein Eigenwert von A, und es sei $\vec{0} \neq \vec{v}_1 \in \mathbf{C}^n$ ein zugeordneter Eigenvektor. Wir ergänzen $\vec{v}_1$ durch Vektoren $\vec{p}_2, \vec{p}_3, \ldots, \vec{p}_n$ zu einer Basis des $\mathbf{C}^n$ und setzen $P := (\vec{v}_1, \vec{p}_2, \ldots, \vec{p}_n)$. Wegen $\mathrm{Rang}\, P = n$ haben wir $P \in \mathrm{Inv}\,(\mathbf{C}^n)$ sowie

$$P\vec{e}_1 = \vec{v}_1, \quad P^{-1}\vec{v}_1 = \vec{e}_1, \quad P^{-1}AP\vec{e}_1 = P^{-1}A\vec{v}_1 = \lambda_1 P^{-1}\vec{v}_1 = \lambda_1 \vec{e}_1.$$

Hieraus ergibt sich die folgende Form der Matrix $P^{-1}AP$, nämlich

$$P^{-1}AP = \left[\begin{array}{c|c} \lambda_1 & \vec{a}^T \\ \hline \vec{0} & A_1 \end{array}\right],$$

mit $A_1 \in \mathbf{C}^{(n-1,n-1)}$ und $\vec{a} \in \mathbf{C}^{n-1}$. Nach Induktionsannahme existieren zu A_1 Matrizen $T_1 \in \mathrm{Inv}\,(\mathbf{C}^{n-1})$ und $D_1 \in U_R(\mathbf{C}^{n-1})$ mit

$$T_1^{-1}A_1T_1 = \begin{bmatrix} \lambda_2 & * & * & \cdots & * \\ & \lambda_3 & * & \cdots & * \\ & & \lambda_4 & \cdots & * \\ & & & \ddots & \vdots \\ O & & & & \lambda_n \end{bmatrix} = D_1.$$

Wir setzen jetzt

$$T := P\left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1 \end{array}\right], \quad T^{-1} = \left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1^{-1} \end{array}\right] P^{-1}.$$

Es gilt also offenbar $T \in \mathrm{Inv}\,(\mathbf{C}^n)$ sowie

$$T^{-1}AT = \left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1^{-1} \end{array}\right] P^{-1}AP \left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1 \end{array}\right] = \left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1^{-1} \end{array}\right] \left[\begin{array}{c|c} \lambda_1 & \vec{a}^T \\ \hline \vec{0} & A_1 \end{array}\right] \left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1 \end{array}\right]$$

$$= \left[\begin{array}{c|c} \lambda_1 & \vec{a}^T \\ \hline \vec{0} & T_1^{-1}A_1 \end{array}\right] \left[\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{0} & T_1 \end{array}\right] = \left[\begin{array}{c|c} \lambda_1 & * \cdots * \\ \hline \vec{0} & T_1^{-1}A_1T_1 \end{array}\right] =: D_0.$$

Dies ist die behauptete obere Dreiecksgestalt. □

Bemerkung 11.11 (a) Der Beweis zum obigen Satz 11.15 enthält ein **konstruktives Verfahren** zur Berechnung der SCHURschen Normalform. Dazu modifizieren wir den Beweis gemäß (b):

(b) Wir bestimmen zunächst **alle** linear unabhängigen Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m,\ m \leq n$, der Matrix A. Es seien $\lambda_1, \lambda_2, \ldots, \lambda_m$ ihre zugeordneten (nicht notwendig voneinander verschiedenen) Eigenwerte. Wir wählen weitere $n - m$ Vektoren $\vec{p}_{m+1}, \ldots, \vec{p}_n$ derart, dass das Vektorsystem $\vec{v}_1, \ldots, \vec{v}_m, \vec{p}_{m+1}, \ldots, \vec{p}_n$ eine Basis des Vektorraumes $\mathbf{C}^n$ bildet. Setzen wir $P_1 := (\vec{v}_1, \ldots, \vec{v}_m, \vec{p}_{m+1}, \ldots, \vec{p}_n)$, so folgt nun ganz analog wie im Beweis zu Satz 11.3:

$$P_1^{-1}AP_1 = \left[\begin{array}{c|c} \Lambda_m & B_1 \\ \hline O & A_1 \end{array}\right], \quad \Lambda_m := \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_m).$$

Mit der Restmatrix $A_1 \in \mathbf{C}^{(n-m,n-m)}$ verfahren wir in derselben Weise. Wir bestimmen ihre linear unabhängigen Eigenvektoren $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_r \in \mathbf{C}^{n-m}$ und die zugeordneten Eigenwerte $\mu_1, \mu_2, \ldots, \mu_r$ sowie eine Basisergänzung $\vec{q}_{r+1}, \ldots, \vec{q}_{n-m-r} \in \mathbf{C}^{n-m}$. Mit der Matrix $P_2 := (\vec{w}_1, \ldots, \vec{w}_r, \vec{q}_{r+1}, \ldots, \vec{q}_{n-m-r})$ erhalten wir sodann:

$$P_2^{-1}A_1P_2 = \left[\begin{array}{c|c} \Lambda_r & B_2 \\ \hline O & A_2 \end{array}\right], \quad \Lambda_r := \mathrm{diag}\,(\mu_1, \mu_2, \ldots, \mu_r),$$

usf. Nach $k \leq n$ Schritten haben wir mit P_k eine Reduzierung auf die 1×1–Matrix $A_k := \lambda_n$ bewirkt, und an dieser Stelle kann das Verfahren abgebrochen werden. Wir setzen schließlich

$$T_1 := P_1, \quad T_2 := \left[\begin{array}{c|c} Id_m & O \\ \hline O & P_2 \end{array}\right], \quad T_3 := \left[\begin{array}{c|c} Id_{m+r} & O \\ \hline O & P_3 \end{array}\right], \ldots,$$

$$T_2^{-1} = \left[\begin{array}{c|c} Id_m & O \\ \hline O & P_2^{-1} \end{array}\right], \quad T_3^{-1} = \left[\begin{array}{c|c} Id_{m+r} & O \\ \hline O & P_3^{-1} \end{array}\right], \ldots,$$

und erhalten mit $T := T_1T_2 \cdots T_k$ die gesuchte Transformationsmatrix, die tatsächlich das Verlangte leistet:

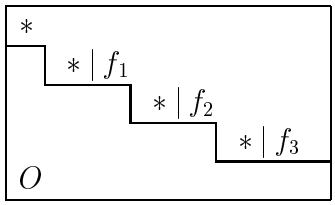
$$T^{-1}AT = T_k^{-1} \cdots \left(T_2^{-1}(T_1^{-1}AT_1)T_2\right) \cdots T_k = D_0.$$

(c) Die Basisergänzung zu den Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ bestimmen wir mit dem in Abschnitt 4.5 beschriebenen Verfahren. Wir setzen

$$V^T := \left[\begin{array}{c} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_m^T \end{array}\right] \in \mathbf{C}^{(m,n)}$$

und führen V^T durch *elementaren Zeilenumformungen* in eine Staffelform über:

$$V^T = \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_m^T \end{bmatrix} \overset{\text{Gauss–Schritte}}{\Rightarrow}$$



S–System, Typ (I) oder Typ (II).

Die Basisergänzungsvektoren $\vec{e}_{j_1}, \vec{e}_{j_2}, \ldots, \vec{e}_{j_{n-m}}$ sind jetzt genau diejenigen Einheitsvektoren $\vec{e}_{j_l}^T = (0, \ldots, 0, 1, 0, \ldots, 0)$ der Standardbasis, deren 1 an der Position f_l des obigen Staffelsystems steht, $l = 1, 2, \ldots, n-m$. □

BSP. (11.4.1) Zu bestimmen ist die SCHURsche Normalform der folgenden Matrix

$$A := \begin{bmatrix} 0 & 0 & 0 & -16 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Lösung: Wir bestimmen zunächst die **Eigenwerte** der Matrix A:

$$\det(A - \lambda\, Id) = \begin{vmatrix} -\lambda & 0 & 0 & -16 \\ 1 & -\lambda & 0 & 0 \\ 0 & 1 & -\lambda & 8 \\ 0 & 0 & 1 & -\lambda \end{vmatrix} = -\lambda \begin{vmatrix} -\lambda & 0 & 0 \\ 1 & -\lambda & 8 \\ 0 & 1 & -\lambda \end{vmatrix} - \begin{vmatrix} 0 & 0 & -16 \\ 1 & -\lambda & 8 \\ 0 & 1 & -\lambda \end{vmatrix}$$

$$= -\lambda\Big((-\lambda)^3 + 8\lambda\Big) + 16 = \lambda^4 - 8\lambda^2 + 16 = (\lambda^2 - 4)^2 = (\lambda - 2)^2(\lambda + 2)^2.$$

Im nächsten Schritt bestimmen wir die Eigenvektoren und eine Basisergänzung:

$$(A - 2\, Id)\vec{v}_1 = \vec{0} \Leftrightarrow \left[\begin{array}{cccc|c} -2 & 0 & 0 & -16 & 0 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & -2 & 8 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ \hline 1 & 0 & 0 & 8 & 0 \\ \hline 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \vec{v}_1 = \begin{bmatrix} -8 \\ -4 \\ 2 \\ 1 \end{bmatrix},$$

$$(A + 2\, Id)\vec{v}_2 = \vec{0} \Leftrightarrow \left[\begin{array}{cccc|c} 2 & 0 & 0 & -16 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 8 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ \hline 1 & 0 & 0 & -8 & 0 \\ \hline 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \vec{v}_2 = \begin{bmatrix} 8 \\ -4 \\ -2 \\ 1 \end{bmatrix}.$$

Es liegen hier lediglich zwei linear unabhängige Eigenvektoren vor, zu denen wir nun eine Basisergänzung bestimmen:

$$V^T := \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \end{bmatrix} \Leftrightarrow \left[\begin{array}{cccc|c} -8 & -4 & 2 & 1 & \\ 8 & -4 & -2 & 1 & \\ \hline -8 & -4 & 2 & 1 & \\ \hline 0 & -8 & 0 & 1 & \\ \hline \end{array}\right] \Rightarrow \begin{array}{l} \vec{p}_3^T = (0,0,1,0)^T, \\ \vec{p}_4^T = (0,0,0,1)^T. \end{array}$$

Im folgenden Schritt führen wir bereits die erste Reduktion der Matrix A durch:

$$P_1 := T_1 := (\vec{v}_1, \vec{v}_2, \vec{p}_3, \vec{p}_4) = \begin{bmatrix} -8 & 8 & 0 & 0 \\ -4 & -4 & 0 & 0 \\ 2 & -2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}, \quad T_1^{-1} = \frac{1}{16} \begin{bmatrix} -1 & -2 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 4 & 0 & 16 & 0 \\ 0 & 4 & 0 & 16 \end{bmatrix}.$$

Wir erhalten hieraus:

$$\begin{aligned} T_1^{-1} A T_1 &= \frac{1}{16} \begin{bmatrix} -1 & -2 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 4 & 0 & 16 & 0 \\ 0 & 4 & 0 & 16 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & -16 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -8 & 8 & 0 & 0 \\ -4 & -4 & 0 & 0 \\ 2 & -2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \\ &= \left[\begin{array}{cc|cc} 2 & 0 & 0 & 1 \\ 0 & -2 & 0 & -1 \\ \hline 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{array}\right]. \end{aligned}$$

Die Restmatrix $A_1 := \left(\begin{smallmatrix} 0 & 4 \\ 1 & 0 \end{smallmatrix}\right)$ hat die folgenden Eigenwerte:

$$\det(A_1 - \lambda\, Id) = \begin{vmatrix} -\lambda & 4 \\ 1 & -\lambda \end{vmatrix} = (\lambda^2 - 4) = (\lambda - 2)(\lambda + 2),$$

also $\lambda_1 = 2$ und $\lambda_2 = -2$. Man bestimmt mit einfacher Rechnung die zugeordneten Eigenvektoren $\vec{w}_1 = (2,1)^T$ und $\vec{w}_2 = (-2,1)^T$. Wir haben nun bereits genügend viele linear unabhängige Eigenvektoren bestimmt, so dass eine Basisergänzung nicht erforderlich ist. Wir können den zweiten Reduktionsschritt durchführen. Mit

$$P_2 := (\vec{w}_1, \vec{w}_2) = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}, \quad P_2^{-1} = \frac{1}{4} \begin{bmatrix} 1 & 2 \\ -1 & 2 \end{bmatrix}$$

sowie

$$T_2 := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad T_2^{-1} = \frac{1}{4} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

erhalten wir nun

$$\begin{aligned} T_2^{-1}(T_1^{-1} A T_1) T_2 &= \frac{1}{4} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & -2 & 0 & -1 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & -2 & -1 & -1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} = R_0, \end{aligned}$$

und dies ist die gesuchte SCHURsche Normalform. Die Transformation T ist gegeben durch

$$T = T_1 T_2 = \begin{bmatrix} -8 & 8 & 0 & 0 \\ -4 & -4 & 0 & 0 \\ 2 & -2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -8 & 8 & 0 & 0 \\ -4 & -4 & 0 & 0 \\ 2 & -2 & 2 & -2 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Bemerkung 11.12 Neben der SCHURschen Normalform einer Matrix $A \in \mathbf{K}^{(n,n)}$ gibt es die weitaus komplizierter strukturierte JORDAN–Normalform von A, die wir in Abschnitt 11.8.1 genauer studieren werden. Wir teilen vorab den folgenden Satz ohne Beweis mit und verweisen bezüglich weiterer Details zum Beispiel auf R. WALTER, Einführung in die lineare Algebra. Friedr.Vieweg, Braunschweig 1982. □

Satz 11.16 (JORDAN–Normalform einer Matrix)
Die Matrix $A \in \mathbf{K}^{(n,n)}$ habe die paarweise verschiedenen Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_m$, $m \leq n$, mit den Vielfachheiten $k_1, k_2, \ldots, k_m$, $k_1 + k_2 + \cdots + k_m = n$. Dann existiert eine Matrix $T \in \mathrm{Inv}\,(\mathbf{C}^n)$ mit $T^{-1}AT = J$, wobei J die JORDAN–**Normalform** *der Matrix A heiße und in der Form*

$$J := \begin{bmatrix} J_1 & 0 & \cdots & 0 & 0 \\ 0 & J_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & J_{s-1} & 0 \\ 0 & 0 & \cdots & 0 & J_s \end{bmatrix}, \quad J_j := \begin{bmatrix} \lambda_j & 1 & & 0 & 0 \\ & \lambda_j & 1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \lambda_j & 1 \\ 0 & 0 & & & \lambda_j \end{bmatrix} \quad \textit{oder} \quad J_j := (\lambda_j), \tag{4.2}$$

$(j = 1, 2, \ldots, s)$ mit $s \geq m$ eindeutig (bis auf die Reihenfolge der JORDAN*–Blöcke J_j) gegeben ist. Dabei brauchen die Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_s$ nicht paarweise verschieden zu sein. Die Anzahl s der* JORDAN*–Blöcke entspricht genau der Anzahl der linear unabhängigen Eigenvektoren der Matrix A. Das heißt, im Sonderfall* $\mathrm{geom.dim}\,\lambda_j = 1 \ \forall\, j = 1, 2, \ldots, m$ *gilt $s = m$, die $\lambda_1, \lambda_2, \ldots, \lambda_m$ sind die oben angegebenen paarweise verschiedenen Eigenwerte, und es gilt $J_j \in \mathbf{C}^{(k_j, k_j)}$.*

11.5 Jacobi–Verfahren für reelle symmetrische Matrizen

Ist $A \in \mathbf{R}^{(n,n)}$ eine **reelle symmetrische** Matrix, so folgt aus Satz 11.11, dass alle Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_n$ von A reell sind und dass es eine ON–Basis des $\mathbf{R}^n$ aus Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ der Matrix A gibt. In diesem Falle ist A mit Hilfe der *orthogonalen* Matrix $Q := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ diagonalisierbar:

$$Q^T A Q = \Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n); \tag{5.1}$$

vgl. Satz 11.9. Ziel ist es, die Transformation (5.1) *iterativ* zu realisieren gemäß folgender Konstruktion:

$$A^{(k)} := Q_k^T A^{(k-1)} Q_k, \quad k \in \mathbf{N}; \quad A^{(0)} := A. \tag{5.2}$$

Dabei soll die orthogonale Matrix Q_k so bestimmt werden, dass im k–ten Schritt durch die Transformation (5.2) das betragsgrößte Nichtdiagonalelement $a_{pq}^{(k-1)}$ der Matrix $A^{(k-1)}$, also

$$|a_{pq}^{(k-1)}| := \max_{i>j} |a_{ij}^{(k-1)}| \tag{5.3}$$

annulliert wird. Um dies zu bewerkstelligen, führen wir die **JACOBI–Rotationen** ein; das sind spezielle orthogonale Matrizen mit nahezu Diagonalstruktur. Wir beginnen mit einem motivierenden Beispiel.

BSP. (11.5.1) Gegeben sei die Matrix $Q := \begin{bmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{bmatrix}$. Es gilt

$$Q^T Q = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix} \begin{bmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

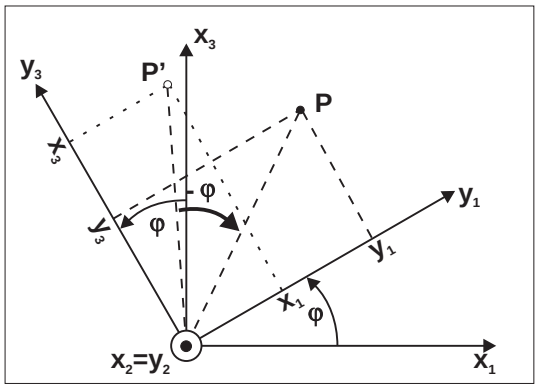
das heißt, $Q^T = Q^{-1}$, so dass die Matrix $Q \in \mathbf{R}^{(3,3)}$ *orthogonal* ist. Die Matrix Q vermittelt eine bijektive lineare Abbildung des $\mathbf{R}^3$ auf sich:

$$\vec{y} := \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = Q\vec{x} = \begin{bmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1\cos\varphi + x_3\sin\varphi \\ x_2 \\ -x_1\sin\varphi + x_3\cos\varphi \end{bmatrix}.$$

Die neuen Koordinatenachsen $y_1 = 0$ bzw. $y_3 = 0$ sind durch die Gleichungen

$$x_1 = -x_3 \tan\varphi \quad \text{bzw.} \quad x_3 = x_1 \tan\varphi$$

bestimmt. Das heißt, die Abbildung Q bewirkt eine *Drehung* der zweidimensionalen (x_1, x_3)–Ebene um den Winkel $-\varphi$.



Drehung der (x_1, x_3)–Ebene

Die Matrizen $A = (a_{jk})$, $A' = (a'_{jk})$, $A'' = (a''_{jk})$ mit

$$A' = Q^T A \quad \text{und} \quad A'' = A'Q$$

erfüllen offenbar die Relationen

$$a'_{jk} = \begin{cases} a_{1k}\cos\varphi - a_{3k}\sin\varphi & , \ j = 1, \\ a_{2k} & , \ j = 2; \ \ k = 1,2,3; \\ a_{1k}\sin\varphi + a_{3k}\cos\varphi & , \ j = 3, \end{cases}$$

$$a''_{jk} = \begin{cases} a'_{j1}\cos\varphi - a'_{j3}\sin\varphi & , \ k = 1, \\ a'_{j2} & , \ k = 2; \ \ j = 1,2,3. \\ a'_{j1}\sin\varphi + a'_{j3}\cos\varphi & , \ k = 3, \end{cases}$$

Allgemein definieren wir:

Definition 11.7 *Eine Matrix $Q := Q(p,q;\varphi) \in \mathbf{R}^{(n,n)}$ in der speziellen Form*

$$Q(p,q;\varphi) := \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & u_{pp} & \cdots & \cdots & u_{pq} & & \\ & & & \vdots & 1 & & \vdots & & \\ & & & \vdots & & \ddots & \vdots & & \\ & & & \vdots & & & 1 \ \vdots & & \\ & & & u_{qp} & \cdots & \cdots & u_{qq} & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{bmatrix}$$

$\leftarrow$ p–te Zeile

$\leftarrow$ q–te Zeile

$\uparrow$ p–te Spalte, $\uparrow$ q–te Spalte

mit

$$\boxed{u_{pp} := u_{qq} := \cos\varphi, \quad u_{pq} := \sin\varphi, \quad u_{qp} := -\sin\varphi; \ \ \varphi \in \mathbf{R},}$$

*heiße eine (p,q)–***Rotationsmatrix** *oder kurz eine* JACOBI**–Rotation**.

In diesem allgemeinen Fall haben wir folgende Eigenschaften:

Satz 11.17 *Für $1 \le p < q \le n$ sei $Q := Q(p,q;\varphi) \in \mathbf{R}^{(n,n)}$ eine (p,q)–Rotationsmatrix. Dann gilt:*

(a) *Q ist orthogonal. Das heißt, es gilt $Q^T = Q^{-1}$.*

(b) *Gegeben sei die Matrix $A = (a_{jk}) \in \mathbf{K}^{(n,n)}$, und es seien $A' = (a'_{jk})$, $A'' = (a''_{jk})$ gemäß*

$$A' = Q^T A, \quad A'' = A'Q \tag{5.4}$$

definiert. Dann gelten die folgenden Relationen:

$$\boxed{\begin{aligned} a'_{pk} &= a_{pk}\cos\varphi - a_{qk}\sin\varphi, \\ a'_{qk} &= a_{pk}\sin\varphi + a_{qk}\cos\varphi, \qquad k = 1,2,\ldots,n; \\ a'_{jk} &= a_{jk} \quad \textit{für} \quad j \neq p,q, \end{aligned}} \tag{5.5}$$

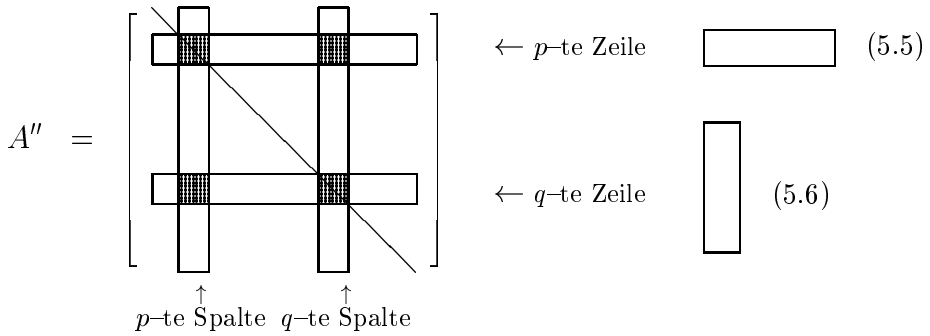
$$\boxed{\begin{aligned} a''_{jp} &= a'_{jp}\cos\varphi - a'_{jq}\sin\varphi, \\ a''_{jq} &= a'_{jp}\sin\varphi + a'_{jq}\cos\varphi, \qquad j = 1,2,\ldots,n. \\ a''_{jk} &= a'_{jk} \quad \textit{für} \quad k \neq p,q, \end{aligned}} \tag{5.6}$$

(c) *Ist $A \in \mathbf{R}^{(n,n)}$ symmetrisch: $A = A^T$, so ist auch $A'' = Q^T AQ$ symmetrisch.*

Begründungen: Die Behauptungen (a) und (b) lassen sich leicht durch direkte Rechnung verifizieren, während (c) sich aus folgender Beziehung ergibt:

$$(A'')^T = (Q^T AQ)^T = (AQ)^T Q = Q^T A^T Q = Q^T AQ = A''.$$

Bemerkung 11.13 Durch die Transformation $A'' = Q^T AQ$ werden in der Matrix A nur die Elemente in den p–ten und q–ten Zeilen und Spalten verändert:



Die Zeilentransformationen verlaufen nach den Formeln (5.5), die Spaltentransformationen gemäß (5.6). Ist A symmetrisch, so erhält man aus (5.5) und (5.6) die Elemente in den vier Kreuzungspunkten zu

$$\boxed{\begin{aligned} a''_{pp} &= a_{pp}\cos^2\varphi - 2a_{pq}\cos\varphi\sin\varphi + a_{qq}\sin^2\varphi, \\ a''_{qq} &= a_{pp}\sin^2\varphi + 2a_{pq}\cos\varphi\sin\varphi + a_{qq}\cos^2\varphi, \\ a''_{pq} &= a''_{qp} = (a_{pp} - a_{qq})\cos\varphi\sin\varphi + a_{pq}(\cos^2\varphi - \sin^2\varphi). \end{aligned}} \tag{5.7}$$

Aufwandsanalyse. Wegen der Symmetrie von A'' braucht man A'' nur in und unterhalb der Diagonalen zu berechnen. Nach obiger Figur sind dazu $2(n-2)$ Elemente gemäß den

Formeln (5.5) und (5.6) zu bestimmen sowie drei Elemente in den Kreuzungspunkten gemäß den Formeln (5.7). Wegen $a''_{qq} = a_{pp} + a_{qq} - a''_{pp}$ ist hierfür der folgende Aufwand erforderlich: □

$$\boxed{Z_{A''} = 4n + 5 \quad \text{Multiplikationen.}} \tag{5.8}$$

Im nächsten Schritt werden wir das **betragsgrößte Nichtdiagonalelement** $a_{pq} = a_{qp}$ der symmetrischen Matrix A unter Verwendung einer JACOBI–Rotation annullieren.

Satz 11.18 *Gegeben sei die symmetrische Matrix $A = A^T \in \mathbf{R}^{(n,n)}$, und es seien Indizes $1 \le p < q \le n$ so gewählt, dass gilt:*

$$|a_{pq}| = \max_{i>j} |a_{ij}|.$$

Wir setzen

$$\boxed{\theta := \frac{a_{qq} - a_{pp}}{2\, a_{qp}}; \quad t := \tan\varphi := \begin{cases} \dfrac{1}{\theta + \operatorname{sign}(\theta) \cdot \sqrt{1+\theta^2}} & \text{für} \quad \theta \neq 0, \\ 1 & \text{für} \quad \theta = 0. \end{cases}} \tag{5.9}$$

$$\boxed{\cos\varphi := \frac{1}{\sqrt{1+t^2}}, \quad \sin\varphi := t \cdot \cos\varphi.} \tag{5.10}$$

Wird mit diesen Daten die JACOBI*–Rotation $Q = Q(p, q; \varphi)$ definiert, so bewirkt die Transformation $A'' = Q^T A Q$, dass gilt:*

$$\boxed{a''_{pq} = a''_{qp} = 0, \quad a''_{pp} = a_{pp} - a_{qp}\tan\varphi, \quad a''_{qq} = a_{qq} + a_{qp}\tan\varphi.} \tag{5.11}$$

Ferner lassen sich mit dem Parameter

$$\boxed{r := \frac{\sin\varphi}{1 + \cos\varphi} \quad \left[= \tan\left(\tfrac{\varphi}{2}\right)\right]} \tag{5.12}$$

die Formeln (5.5) *und* (5.6) *in der folgenden Form darstellen:*

$$\boxed{\left.\begin{aligned} a'_{pk} &= a_{pk} - [a_{qk} + r\, a_{pk}] \cdot \sin\varphi, \\ a'_{qk} &= a_{qk} + [a_{pk} - r\, a_{qk}] \cdot \sin\varphi, \end{aligned}\right\} k = 1, 2, \ldots, n.} \tag{5.13}$$

$$\boxed{\left.\begin{aligned} a''_{jp} &= a'_{jp} - [a'_{jq} + r\, a'_{jp}] \cdot \sin\varphi, \\ a''_{jq} &= a'_{jq} + [a'_{jp} - r\, a'_{jq}] \cdot \sin\varphi, \end{aligned}\right\} j = 1, 2, \ldots, n.} \tag{5.14}$$

Begründungen: Gemäß (5.7) und unter Verwendung von $\sin(2\varphi) = 2\sin\varphi \cdot \cos\varphi$ sowie $\cos(2\varphi) = \cos^2\varphi - \sin^2\varphi$ ist die Bedingung $a''_{pq} = 0$ äquivalent mit der folgenden Bedingungsgleichung für den Drehwinkel φ :

$$\cot(2\varphi) = \frac{a_{qq} - a_{pp}}{2\, a_{pq}} =: \theta. \tag{5.15}$$

Setzt man $t := \tan\varphi$, so folgt hieraus die algebraische Gleichung $(1 - t^2)/2\, t = \theta$ oder äquivalent $t^2 + 2\theta t - 1 = 0$ mit den beiden Wurzeln

$$t_{1/2} = -\theta \pm \sqrt{1+\theta^2} = \frac{1}{\theta \pm \sqrt{1+\theta^2}}.$$

Die Wahl der betragskleineren Lösung führt auf den in (5.9) angegebenen Wert für $t = \tan\varphi$, und es wird $-1 < \tan\varphi \leq 1$ erreicht, so dass der Drehwinkel φ auf das Intervall $-\frac{\pi}{4} < \varphi \leq \frac{\pi}{4}$ beschränkt bleibt. In diesem Intervall gilt $0 < \cos\varphi = \cos\varphi \Big/ \sqrt{\cos^2\varphi + \sin^2\varphi} = 1 \Big/ \sqrt{1 + \tan^2\varphi}$ und $\sin\varphi = \tan\varphi \cdot \cos\varphi$, also (5.10). Ferner resultiert aus (5.7):

$$\begin{aligned} a''_{pp} &= a_{pp} - a_{qp} \left\{ 2\cos\varphi \cdot \sin\varphi - \frac{a_{qq} - a_{pp}}{a_{qp}} \sin^2\varphi \right\} \overset{(5.15)}{=} a_{pp} - a_{qp} \cdot \tan\varphi, \\ a''_{qq} &= a_{pp} + a_{qq} - a''_{pp} = a_{qq} + a_{qp} \cdot \tan\varphi, \end{aligned}$$

also (5.11). Schließlich verwenden wir die Identität

$$\cos\varphi = \frac{\cos\varphi\ (1 + \cos\varphi)}{1 + \cos\varphi} = \frac{1 + \cos\varphi - \sin^2\varphi}{1 + \cos\varphi} = 1 - \frac{\sin\varphi}{1 + \cos\varphi}\ \sin\varphi = 1 - r\ \sin\varphi,$$

um aus den Relationen (5.5) und (5.6) die Formeln (5.13) und (5.14) herzuleiten. □

Bemerkung 11.14 Man beachte, dass sich der Rechenaufwand in den Formeln (5.11) gegenüber den Formeln (5.7) verringert. Darüber hinaus sind die Darstellungen (5.11) unempfindlicher gegen Rundungsfehler. Die Anzahl der Multiplikationen in (5.13) und (5.14) hingegen bleibt die gleiche wie in den Formeln (5.5) und (5.6). □

Wir kommen jetzt zurück zur Ausgangsproblemstellung, nämlich zum Iterationsprozess (5.2). Im k–ten Schritt sei das betragsgrößte Nichtdiagonalelement $a_{pq}^{(k-1)}$ der Matrix $A^{(k-1)}$ so bestimmt, dass (5.3) gilt. Dieses Element annullieren wir durch die JACOBI–Rotation $Q_k := Q(p, q; \varphi)$ nach der Vorschrift von Satz 11.18. Im $(k+1)$–ten Schritt wiederholen wir dieses Verfahren für die Matrix $A^{(k)}$ aus (5.2). Obwohl das vorher annullierte Element im allgemeinen wieder ungleich Null wird, hat man doch den folgenden Konvergenzsatz:

Satz 11.19 *Für die symmetrische Matrix $A = A^T \in \mathbf{R}^{(n,n)}$ konvergiert die Folge* (5.2) *von ähnlichen Matrizen $A^{(k)}$ gegen eine Diagonalmatrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.*

Begründung: Die Quadratsumme der Nichtdiagonalelemente von $A^{(k)}$ bezeichnen wir mit

$$S\left(A^{(k)}\right) := \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[a_{ij}^{(k)}\right]^2, \quad k = 1, 2, \ldots$$

Zu zeigen ist nun, dass $S(A^{(k)}),\ \ k \in \mathbf{N}$, eine monoton fallende Nullfolge bildet. Dazu betrachten wir die Differenz $S(A'') - S(A)$ unter einer JACOBI–Rotation $A'' = Q^T A Q$ zum Indexpaar (p, q). Es gilt

$$\begin{aligned} S(A'') &= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[a''_{ij}\right]^2 = \sum_{\substack{i=1 \\ i \neq p,q}}^{n} \sum_{\substack{j=1 \\ j \neq i,p,q}}^{n} \left[a''_{ij}\right]^2 + \sum_{\substack{i=1 \\ i \neq p,q}}^{n} \left\{ \left[a''_{ip}\right]^2 + \left[a''_{iq}\right]^2 \right\} \\ &\quad + \sum_{\substack{j=1 \\ j \neq p,q}}^{n} \left\{ \left[a''_{pj}\right]^2 + \left[a''_{qj}\right]^2 \right\} + 2\left[a''_{qp}\right]^2. \end{aligned}$$

Wir haben wegen (5.5) und (5.6) die Beziehungen $a''_{ij} = a_{ij}\ \ \forall\ i \neq p, q \wedge j \neq i, p, q$. Ferner überlegt man sich, dass wegen der Orthogonalität von Q die Beziehungen

$$\begin{aligned} \left[a''_{ip}\right]^2 + \left[a''_{iq}\right]^2 &= a_{ip}^2 + a_{iq}^2, \quad i \neq p, q, \\ \left[a''_{pj}\right]^2 + \left[a''_{qj}\right]^2 &= a_{pj}^2 + a_{qj}^2, \quad j \neq p, q \end{aligned}$$

gelten. Deshalb ergibt sich aus obiger Rechnung:

$$S(A'') = S(Q^T A Q) = [S(A) - 2a_{qp}^2] + 2\left[a''_{qp}\right]^2.$$

Da nun die Transformation Q_k im k–ten Schritt gerade die Annullierung von $a_{qp}^{(k-1)}$ bewirkt, erhält man

$$S\Big(A^{(k)}\Big) = S\Big(A^{(k-1)}\Big) - 2\left[a_{qp}^{(k-1)}\right]^2, \quad k = 1, 2, \ldots \tag{5.16}$$

Das heißt, die Folge $0 \leq S(A^{(k)})$ ist streng monoton fallend, und die Abnahme von $S(A^{(k-1)})$ ist wegen (5.3) sogar optimal. Nach dem Hauptsatz über die monotone Konvergenz muss $0 \leq S := \lim\limits_{k\to\infty} S(A^{(k)})$ existieren. Wir müssen noch $S = 0$ zeigen. Wegen (5.3) gilt

$$S\Big(A^{(k-1)}\Big) \leq (n^2 - n)\left[a_{qp}^{(k-1)}\right]^2,$$

und somit gemäß (5.16):

$$S\Big(A^{(k)}\Big) \leq \left[1 - \frac{2}{n^2 - n}\right]\ S\Big(A^{(k-1)}\Big) \leq \left[1 - \frac{2}{n^2 - n}\right]^k\ S\Big(A^{(0)}\Big). \tag{5.17}$$

Für $n = 2$ gilt $\kappa := 1 - 2/(n^2 - n) = 0$, im Einklang mit der Tatsache, dass eine einzige JACOBI–Rotation zur Diagonalisierung der symmetrischen (2×2)–Matrix A ausreicht. In diesem Fall gilt bereits $S(A^{(1)}) = 0$. Für $n > 2$ folgt $\kappa < 1$, und dies führt wegen (5.17) zur Behauptung $S = 0$. □

Bemerkung 11.15 (a) Das Produkt der JACOBI–Rotationen

$$V_k := Q_1 Q_2 \cdots Q_k, \quad k = 1, 2, \ldots,$$

definiert eine orthogonale Matrix $V_k \in \mathbf{R}^{(n,n)}$ mit $A^{(k)} = V_k^T A V_k$. Gemäß Satz 11.19 approximiert $A^{(k)}$ für hinreichend große Zahlen k die Spektralmatrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ von A mit beliebiger Genauigkeit. Das heißt, die Diagonalelemente $a_{jj}^{(k)}$ liefern Approximationen der Eigenwerte λ_j von A, während die Spaltenvektoren von V_k Näherungen der zugeordneten orthonormierten Eigenvektoren sind.

(b) Gemäß (5.17) konvergiert die Folge $S(A^{(k)})$ mindestens wie eine geometrische Folge mit dem Quotienten $\kappa = 1 - 2/(n^2 - n)$. Um etwa die Bedingung

$$\frac{S\Big(A^{(k)}\Big)}{S\Big(A^{(0)}\Big)} \leq \epsilon^2 \tag{5.18}$$

zu erfüllen, muss also $(1 - \frac{1}{N})^k \leq \epsilon^2$ gelten, wobei $N := \frac{1}{2}(n^2 - n)$ gerade die Anzahl der Nichtdiagonalelemente in der unteren Hälfte von A angibt. Für größeres N gilt $\ln(1 - \frac{1}{N}) \approx -\frac{1}{N}$, so dass

$$k \geq \frac{2\ln(\epsilon)}{\ln(1 - \frac{1}{N})} \approx 2N\ \ln(\frac{1}{\epsilon}) = (n^2 - n)\ \ln(\frac{1}{\epsilon})$$

die erforderliche Anzahl der JACOBI–Rotationen zur Erfüllung von (5.18) angibt. Bei Vorgabe von $\epsilon := 10^{-d}$ und bei einem Aufwand von ca. $4n$ Multiplikationen pro JACOBI–Rotation benötigt man für das JACOBI–Verfahren etwa

$$\boxed{Z_{\text{JACOBI}} \approx d \cdot (n^3 - n^2) \cdot 4\ln 10 \doteq 9.21 \cdot d\,(n^3 - n^2)} \tag{5.19}$$

Multiplikationen.

(c) Die lineare Konvergenz (5.17) der Folge $S(A^{(n)})$ ist zu pessimistisch, denn man kann sogar quadratische Konvergenz zeigen, sobald $S(A^{(n)})$ genügend klein geworden ist. Wir verweisen auf die Literatur.

(d) Die Güte der Approximation der Eigenwerte λ_j durch die Diagonalelemente $a_{jj}^{(k)}$ wird durch folgenden Satz charakterisiert:

Satz 11.20 *Gegeben sei die symmetrische Matrix $A^T = A \equiv A^{(0)} \in \mathbf{R}^{(n,n)}$, deren Eigenwerte λ_j in aufsteigender Folge $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ angeordnet seien. Bezeichne $l_j^{(k)}$ die der Größe nach geordneten Diagonalelemente der Matrix $A^{(k)}$: $l_1^{(k)} \leq l_2^{(k)} \leq \cdots \leq l_n^{(k)}$. Dann gilt*

$$|l_j^{(k)} - \lambda_j| \leq \sqrt{S\left(A^{(k)}\right)} \;\; \forall\; j = 1, 2, \ldots, n; \;\; k = 0, 1, \ldots \tag{5.20}$$

Zur *Begründung* verweisen wir wiederum auf die Literatur. Wegen $S(A^{(k)}) \leq (n^2 - n)\,[a_{qp}^{(k)}]^2$ ist es einfacher, anstelle von (5.20) die schlechtere Abschätzung

$$|l_j^{(k)} - \lambda_j| < n \cdot |a_{qp}^{(k)}| \;\; \forall j = 1, 2, \ldots, n, \;\; k = 0, 1, \ldots,$$

zu verwenden, die auch als *Abbruchkriterium* dienen kann.

(e) Die Matrixfolge V_k berechnet sich rekursiv gemäß

$$V_0 := Id, \quad V_k := V_{k-1} Q_k \;\; \forall\; k = 1, 2, \ldots$$

Die Matrizenmultiplikation $V_{k-1}Q_k$ kann gemäß den Formeln (5.6) oder (5.14) bewerkstelligt werden; sie erfordert $4n$ Multiplikationen. Damit verdoppelt sich der Gesamtaufwand des JACOBI–Verfahrens, wenn die Eigenvektoren mitberechnet werden. □

Zyklisches JACOBI–Verfahren: Die Suche nach dem betragsgrößten Nichtdiagonalelement erfordert gemäß (5.3) in jedem Iterationsschritt $N = \frac{1}{2}(n^2 - n)$ Vergleichsoperationen, was in Relation zu den $4n$ Multiplikationen der JACOBI–Rotation einen unverhältnismäßig großen Aufwand bedeutet. Man umgeht die Abfrage (5.3), wenn man jedes der N Nichtdiagonalelemente in der Linksdreiecksmatrix A_L in einem Zyklus von N Rotationen je einmal annulliert. In diesem Falle erhält man ein *zyklisches* JACOBI*–Verfahren.* Ein spezieller Zyklus besteht zum Beispiel in dem spaltenweisen Abarbeiten der Linksdreiecksmatrix A_L; das heißt, die JACOBI–Rotationen$Q(p, q; \varphi)$ durchlaufen sequentiell die Indexfolge

$$(p, q) = (1,2), (1,3), \ldots, (1,n), (2,3), (2,4), \ldots, (2,n), (3,4), \ldots, (n-1, n). \tag{5.21}$$

Dabei soll $Q_k(p, q; \varphi_k)$ *nicht* ausgeführt werden, wenn das Element $a_{qp}^{(k-1)}$ verschwindet. Die Werte von $\cos\varphi$ und $\sin\varphi$ werden wieder nach der Vorschrift (5.9) und (5.10) festgelegt.

Satz 11.21 *Für eine symmetrische Matrix $A = A^T \in \mathbf{R}^{(n,n)}$ konvergiert die Folge $A^{(k)} = Q_k^T A^{(k-1)} Q_k$, $k \in \mathbf{N}$, $A^{(0)} := A$, gegen eine Diagonalmatrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, wobei die* JACOBI*–Rotationen $Q_k = Q_k(p, q; \varphi_k)$ nach dem speziellen zyklischen* JACOBI*–Verfahren* (5.21) *gebildet werden.* □

Bezüglich eines *Beweises* verweisen wir auf die in H.R. SCHWARZ, Numerische Mathematik, S.243/244, angegebene Literatur. Wir haben dem Buch von H.R. SCHWARZ auch den folgenden Algorithmus für das spezielle zyklische JACOBI–Verfahren entnommen.

Algorithmus für das spezielle zyklische JACOBI–Verfahren.

Vorgabe: Die Elemente a_{jk}, $1 \le k \le j \le n$, (linkes unteres Dreieck) der symmetrischen Matrix $A = A^T \in \mathbf{R}^{(n,n)}$.

```
 1: INIT:  für i := 1,2,...,n :
 2:           für j := 1,2,...,n :
 3:             v_ij := 0; (Ende j)
 4:           v_ii := 1; (Ende i)
 5: CYCL:  wiederhole:
 6:           sum := 0
 7:           für i := 2,3,...,n :
 8:             für j := 1,2,...,i-1 :
 9:               sum := sum + a_ij^2; (Ende j, Ende i)
10:           für p := 1,2,...,n-1 :
11:             für q := p+1,p+2,...,n :
12:               falls |a_qp| ≥ ε^2 dann:
13:                 θ := (a_qq - a_pp)/(2 * a_qp); t := 1;
14:                 falls |θ| > δ dann:
15:                   t := 1/(θ + sign(θ)√(1+θ^2)); (Ende falls)
16:                 c := 1/√(1+t^2); s := c * t; r := s/(1+c);
17:                 a_pp := a_pp - t * a_qp; a_qq := a_qq + t * a_qp; a_qp := 0;
18:                 für j := 1,2,...,p-1 :
19:                   g := a_qj + r * a_pj; h := a_pj - r * a_qj;
20:                   a_pj := a_pj - s * g; a_qj := a_qj + s * h; (Ende j)
21:                 für i := p+1,p+2,...,q-1 :
22:                   g := a_qi + r * a_ip; h := a_ip - r * a_qi;
23:                   a_ip := a_ip - s * g; a_qi := a_qi + s * h; (Ende i)
24:                 für i := q+1,q+2,...,n :
25:                   g := a_iq + r * a_ip; h := a_ip - r * a_iq;
26:                   a_ip := a_ip - s * g; a_iq := a_iq + s * h; (Ende i)
27:                 für i := 1,2,...,n :
28:                   g := v_iq + r * v_ip; h := v_ip - r * v_iq;
29:                   v_ip := v_ip - s * g; v_iq := v_iq + s * h;
               (Ende i, Ende falls, Ende q, Ende p)
30: TEST:  bis 2 * sum < ε^2; STOP
```

(5.22)

Hierin bezeichnet ϵ die absolute Toleranz, mit welcher die Eigenwerte berechnet werden sollen. Diese stehen nach Ablauf des Verfahrens in der Diagonalen der Matrix A, während die Matrix V spaltenweise die zugeordneten orthonormierten Näherungen der Eigenvektoren enthält. Mit δ wird die Maschinengenauigkeit bezeichnet, also die kleinste positive Zahl, für die der Rechner $1 + \delta \neq 1$ liefert. Eine JACOBI–Rotation wird übersprungen, falls $|a_{qp}| < \epsilon^2$ ausfällt. Obwohl der numerische Aufwand beim zyklischen JACOBI–Verfahren relativ hoch ist, wird das Verfahren gekennzeichnet durch Einfachheit, hohe numerische Stabilität und simple Computer–Implementierung. Dies macht das JACOBI–Verfahren in den Anwendungen sehr beliebt.

BSP. (11.5.2) Wir berechnen die Eigenwerte und Eigenvektoren der folgenden Matrix mit dem zyklischen JACOBI–Verfahren für zwei verschiedene Genauigkeitsvorgaben $\epsilon := 10^{-2}$ und $\epsilon := 10^{-6}$. Im ersten Fall werden vier volle Zyklen bis zur Erreichung der geforderten Genauigkeit benötigt, im zweiten Fall fünf Zyklen.

$$A := \begin{bmatrix} -1 & 1 & 2 & 3 & 5 \\ 1 & -4 & 5 & 9 & 14 \\ 2 & 5 & -7 & 12 & 19 \\ 3 & 9 & 12 & -21 & 33 \\ 5 & 14 & 19 & 33 & -52 \end{bmatrix}$$

Vorgegebene Toleranz: $\epsilon := 10^{-02}$. Anzahl der durchlaufenen Zyklen: $it := 4$.

Spektralmatrix Λ von A

$$\begin{bmatrix}
2.661\,846\mathrm{E}^{+01} & 0.000\,000\mathrm{E}^{+00} & -6.997\,978\mathrm{E}^{-07} & -3.316\,493\mathrm{E}^{-08} & -2.709\,262\mathrm{E}^{-10} \\
0.000\,000\mathrm{E}^{+00} & -1.752\,892\mathrm{E}^{+00} & -1.542\,687\mathrm{E}^{-07} & -1.433\,845\mathrm{E}^{-11} & 1.391\,400\mathrm{E}^{-15} \\
-6.997\,978\mathrm{E}^{-07} & -1.542\,687\mathrm{E}^{-07} & -9.817\,706\mathrm{E}^{+00} & 0.000\,000\mathrm{E}^{+00} & 1.380\,238\mathrm{E}^{-05} \\
-3.316\,493\mathrm{E}^{-08} & -1.433\,845\mathrm{E}^{-11} & 0.000\,000\mathrm{E}^{+00} & -2.484\,249\mathrm{E}^{+01} & -4.075\,662\mathrm{E}^{-06} \\
-2.709\,262\mathrm{E}^{-10} & 1.391\,400\mathrm{E}^{-15} & 1.380\,238\mathrm{E}^{-05} & -4.075\,662\mathrm{E}^{-06} & -7.520\,536\mathrm{E}^{+01}
\end{bmatrix}$$

Eigenwerte von A

λ_1	λ_2	λ_3	λ_4	λ_5
$2.661\,846\mathrm{E}^{+01}$	$-1.752\,892\mathrm{E}^{+00}$	$-9.817\,706\mathrm{E}^{+00}$	$-2.484\,249\mathrm{E}^{+01}$	$-7.520\,536\mathrm{E}^{+01}$

Modalmatrix V von A

$$\begin{bmatrix}
1.929\,423\mathrm{E}^{-01} & -9.714\,106\mathrm{E}^{-01} & -1.140\,296\mathrm{E}^{-01} & -7.073\,394\mathrm{E}^{-02} & -3.359\,338\mathrm{E}^{-02} \\
4.525\,275\mathrm{E}^{-01} & 2.099\,135\mathrm{E}^{-01} & -8.066\,920\mathrm{E}^{-01} & -3.011\,019\mathrm{E}^{-01} & -9.869\,504\mathrm{E}^{-02} \\
5.216\,230\mathrm{E}^{-01} & 8.747\,452\mathrm{E}^{-02} & 5.637\,527\mathrm{E}^{-01} & -6.167\,442\mathrm{E}^{-01} & -1.485\,495\mathrm{E}^{-01} \\
5.370\,761\mathrm{E}^{-01} & 5.920\,082\mathrm{E}^{-02} & 1.188\,265\mathrm{E}^{-01} & 6.853\,187\mathrm{E}^{-01} & -4.735\,640\mathrm{E}^{-01} \\
4.443\,542\mathrm{E}^{-01} & 3.378\,063\mathrm{E}^{-02} & 6.563\,728\mathrm{E}^{-02} & 2.330\,216\mathrm{E}^{-01} & 8.618\,589\mathrm{E}^{-01}
\end{bmatrix}$$

Diese Resultate werden durch Heraufsetzung der Genauigkeitsvorgabe nur unwesentlich verbessert.

Vorgegebene Toleranz: $\epsilon := 10^{-06}$. Anzahl der durchlaufenen Zyklen: $it := 5$.

Spektralmatrix Λ von A

$$\begin{bmatrix}
2.661\,846\mathrm{E}^{+01} & 2.962\,915\mathrm{E}^{-15} & 6.115\,309\mathrm{E}^{-22} & 1.840\,409\mathrm{E}^{-27} & 0.000\,000\mathrm{E}^{+00} \\
2.962\,915\mathrm{E}^{-15} & -1.752\,892\mathrm{E}^{+00} & 5.346\,429\mathrm{E}^{-25} & 0.000\,000\mathrm{E}^{+00} & 3.395\,412\mathrm{E}^{-14} \\
6.115\,309\mathrm{E}^{-22} & 5.346\,429\mathrm{E}^{-25} & -9.817\,706\mathrm{E}^{+00} & -8.609\,500\mathrm{E}^{-13} & -5.269\,630\mathrm{E}^{-18} \\
1.840\,409\mathrm{E}^{-27} & 0.000\,000\mathrm{E}^{+00} & -8.609\,500\mathrm{E}^{-13} & -2.484\,249\mathrm{E}^{+01} & -1.745\,097\mathrm{E}^{-19} \\
0.000\,000\mathrm{E}^{+00} & 3.395\,412\mathrm{E}^{-14} & -5.269\,630\mathrm{E}^{-18} & -1.745\,097\mathrm{E}^{-19} & -7.520\,536\mathrm{E}^{+01}
\end{bmatrix}$$

Eigenwerte von A

λ_1	λ_2	λ_3	λ_4	λ_5
$2.661\,846\mathrm{E}^{+01}$	$-1.752\,892\mathrm{E}^{+00}$	$-9.817\,706\mathrm{E}^{+00}$	$-2.484\,249\mathrm{E}^{+01}$	$-7.520\,536\mathrm{E}^{+01}$

Modalmatrix V von A

$$\begin{bmatrix}
1.929\,423\mathrm{E}^{-01} & -9.714\,106\mathrm{E}^{-01} & -1.140\,296\mathrm{E}^{-01} & -7.073\,394\mathrm{E}^{-02} & -3.359\,336\mathrm{E}^{-02} \\
4.525\,276\mathrm{E}^{-01} & 2.099\,135\mathrm{E}^{-01} & -8.066\,920\mathrm{E}^{-01} & -3.011\,019\mathrm{E}^{-01} & -9.869\,490\mathrm{E}^{-02} \\
5.216\,230\mathrm{E}^{-01} & 8.747\,451\mathrm{E}^{-02} & 5.637\,527\mathrm{E}^{-01} & -6.167\,442\mathrm{E}^{-01} & -1.485\,497\mathrm{E}^{-01} \\
5.370\,761\mathrm{E}^{-01} & 5.920\,082\mathrm{E}^{-02} & 1.188\,264\mathrm{E}^{-01} & 6.853\,188\mathrm{E}^{-01} & -4.735\,640\mathrm{E}^{-01} \\
4.443\,542\mathrm{E}^{-01} & 3.378\,062\mathrm{E}^{-02} & 6.563\,748\mathrm{E}^{-02} & 2.330\,215\mathrm{E}^{-01} & 8.618\,589\mathrm{E}^{-01}
\end{bmatrix}$$

11.6 Anwendung: Die Flächen 2.Ordnung

Wir setzen in diesem Abschnitt voraus, dass **Kegelschnitte** in $\mathbf{R}^2$ aus der Schulgeometrie bekannt sind. Kegelschnitte werden in der Regel in *kartesischen Koordinaten* in ihrer **Normalform** dargestellt:

- Normalform der **Ellipsengleichung:**

$$\boxed{\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1, \qquad a, b > 0.}$$

  Im Sonderfall $a = b$ erhält man die Normalform der **Kreisgleichung.**

- Normalform der **Hyperbelgleichung:**

$$\boxed{\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1, \qquad a, b > 0.}$$

- Normalform der **Parabelgleichung:**

$$\boxed{2px_1 - x_2^2 = 0, \qquad p > 0.}$$

Eine *geometrische Definition* der Kegelschnitte kann auch in der folgenden Weise vorgenommen werden:

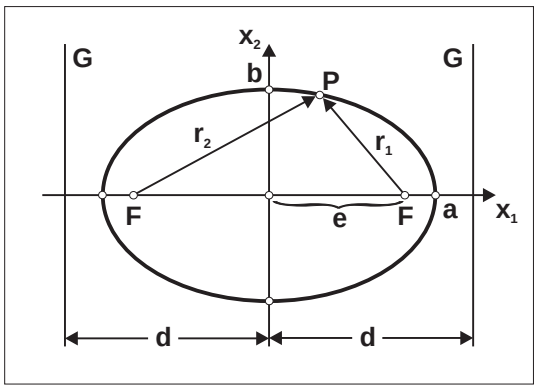
Definition 11.8 *Die Menge aller Punkte P, für die das Verhältnis $\overline{PF}/\overline{PG}$ der Abstände zu einem gegebenen Punkt F (dem* **Brennpunkt**) *und zu einer gegebenen Geraden G (der* **Leitlinie**) *eine Konstante ϵ ist, ist ein* **Kegelschnitt** *mit der* **numerischen Exzentrizität** *ϵ. Spezialfälle sind:*

(a) *Der* **Kreis** *mit $\epsilon := 0$: F ist der Mittelpunkt, und G liegt im Unendlichen.*

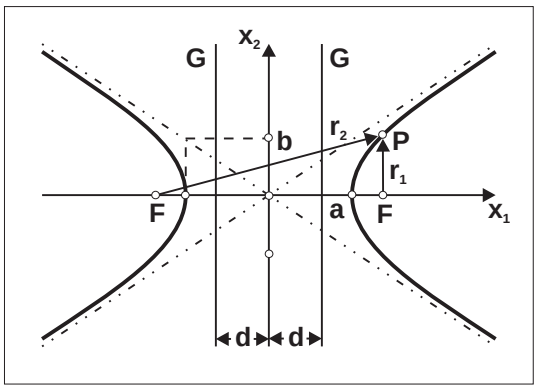
(b) *Die* **Ellipse:** *Sind $a > b > 0$ und $\overline{OF} = \sqrt{a^2 - b^2} =: e$, so gilt $\epsilon := e/a$, $0 < \epsilon < 1$. Die Gerade G liegt im Abstand $d := a/\epsilon$ parallel zur kleinen Halbachse.*

(c) *Die* **Hyperbel:** *Sind nun $a > b > 0$ und $\overline{OF} = \sqrt{a^2 + b^2} =: e$, so gilt $\epsilon := e/a$, $\epsilon > 1$. Die Gerade G liegt im Abstand $d := a/\epsilon$ parallel zur x_2–Achse.*

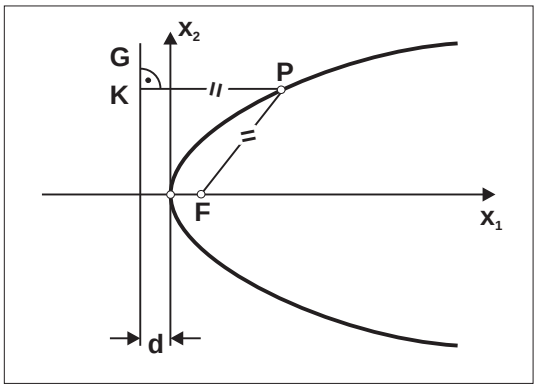
(d) *Die* **Parabel:** *Es gelten $\epsilon := 1$, $\overline{OF} =: p/2$, und G ist die Gerade parallel zur x_2–Achse im Abstand $d := -p/2$.*



Die Ellipse: $r_1 + r_2 = 2a = const$



Die Hyperbel: $r_2 - r_1 = 2a = const$



Die Parabel: $\overline{FP} = \overline{KP}$

Bemerkung 11.16 Definiert man

$$\boxed{r(\varphi) := \frac{p}{1 - \epsilon \cos \varphi}, \qquad \epsilon \geq 0,}$$

für solche $\varphi \in [0, 2\pi)$, für die $r(\varphi) \geq 0$ gilt, so liefert die Relation $r = r(\varphi)$ eine **Polardarstellung** der Kegelschnitte, bei der ein Brennpunkt stets im Ursprung liegt. Man erhält im Einzelnen

- einen **Kreis**, falls $\epsilon = 0$ und $p > 0$ gelten
- eine **Ellipse**, falls $0 < \epsilon < 1$ und $p = b^2/a = (1 - \epsilon^2)a$ gelten; der Ursprung O liegt im linken Brennpunkt
- eine **Hyperbel**, falls $\epsilon > 1$ und $p = b^2/a = (\epsilon^2 - 1)a$ gelten; der Ursprung O liegt im rechten Brennpunkt
- eine **Parabel**, falls $\epsilon = 1$ und $p > 0$ gelten.

Eine Gleichung in der Form

$$\boxed{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_1x_1 + 2a_2x_2 + b = 0} \tag{6.1}$$

heißt **allgemeine Kegelschnittgleichung**. In ihr sind die oben eingeführten Normalformen als Spezialfälle enthalten. Im allgemeinen Fall kann die Gleichung (6.1) durch eine **Hauptachsentransformation** auf eine der Normalformen reduziert werden (man vergleiche z.B. I.N. Bronstein/K.A. Semendjajew, Taschenbuch der Mathematik, 22. Auflage, 1985; dort Abschnitt 2.6.6.1). Wir werden nachfolgend anstelle der zweidimensionalen Kegelschnittgleichung (6.1) die allgemeinere *dreidimensionale* Gleichung studieren:

$$\boxed{\begin{array}{rcl} 0 & = & a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + a_{33}x_3^2 \\ & + & 2a_1x_1 + 2a_2x_2 + 2a_3x_3 + b. \end{array}} \tag{6.2}$$

Definition 11.9 *Es seien* $O \neq A \in \mathbf{R}^{(3,3)}$ *eine* **symmetrische** *Matrix,* $\vec{a} \in \mathbf{R}^3$ *ein fester Vektor und* $b \in \mathbf{R}$ *eine feste Zahl,*

$$A := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \qquad \vec{a} := \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \qquad \vec{x} := \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbf{R}^3.$$

Dann heiße die Menge aller Punkte $P \in \mathbf{R}^3$*, deren Ortsvektor* $OP =: \vec{x}$ *der Gleichung*

$$\langle A\vec{x}, \vec{x} \rangle + 2 \langle \vec{a}, \vec{x} \rangle + b = 0 \tag{6.3}$$

genügt, eine **Fläche 2.Ordnung** *oder eine* **Quadrik** *in* $\mathbf{R}^3$*. Die Gleichungen* (6.2) *und* (6.3) *sind äquivalent – da* A *symmetrisch ist, gilt* $a_{jk} = a_{kj}$*.*

Unser Ziel ist es, die Quadriken (6.3) auf Normalformen zu transformieren, so dass eine Klassifizierung ermöglicht wird. Dazu lösen wir in einem *1.Schritt* das Eigenwertproblem für die symmetrische Matrix A. Gemäß Satz 11.11 hat A drei reelle Eigenwerte $\lambda_1, \lambda_2, \lambda_3$, und es existiert eine ON–Basis des $\mathbf{R}^3$ aus den zugeordneten normierten Eigenvektoren $\vec{y}_1, \vec{y}_2, \vec{y}_3$. Die orthogonale Matrix $Y := (\vec{y}_1, \vec{y}_2, \vec{y}_3)$ induziert eine **Basistransformation** $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\} \mapsto \{\vec{y}_1, \vec{y}_2, \vec{y}_3\}$ der natürlichen Basis des $\mathbf{R}^3$ auf eine ON–Basis. Gemäß Satz 5.15 (mit $A := Id = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$ und $B := Y = (\vec{y}_1, \vec{y}_2, \vec{y}_3)$) ist Y die Matrix der **Koordinaten**transformation zum Basiswechsel $B \to A$; d.h. die Vektoren $\vec{x} = \sum_{j=1}^{3} x_j \vec{e}_j$ und $\vec{y} = \sum_{j=1}^{3} y_j \vec{y}_j$ stehen in der folgenden Beziehung zueinander:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = Y \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = Y^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{6.4}$$

Man beachte, dass wegen der Orthogonalität von Y die Gleichung $Y^{-1} = Y^T$ gilt.

Satz 11.22 *In der neuen Basis* $\{\vec{y}_1, \vec{y}_2, \vec{y}_3\}$ *hat die Gleichung* (6.3) *der Quadrik die Form*

$$\boxed{\langle \Lambda \vec{y}, \vec{y} \rangle + 2 \langle \vec{b}, \vec{y} \rangle + b = 0,} \tag{6.5}$$

worin $\Lambda := \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$, $\vec{y} := (y_1, y_2, y_3)^T$ *und* $\vec{b} := Y^T \vec{a} = (b_1, b_2, b_3)^T$ *zu setzen sind.*

Begründung: Wegen Satz 11.8(c) haben wir nämlich $\Lambda = Y^T A Y$, und somit unter Beachtung von (6.4) die Äquivalenz

$$\begin{aligned} (6.3) \quad &\Leftrightarrow \quad \langle AY\vec{y}, Y\vec{y} \rangle + 2 \langle \vec{a}, Y\vec{y} \rangle + b = 0 \quad \Leftrightarrow \quad \langle Y^T A Y \vec{y}, \vec{y} \rangle + 2 \langle Y^T \vec{a}, \vec{y} \rangle + b = 0 \\ &\Leftrightarrow \quad \langle \Lambda \vec{y}, \vec{y} \rangle + 2 \langle \vec{b}, \vec{y} \rangle + b = 0. \end{aligned}$$

Dies ist die behauptete Gleichung (6.5); sie lautet in expliziter Form

$$\boxed{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + 2b_1 y_1 + 2b_2 y_2 + 2b_3 y_3 + b = 0.} \tag{6.6}$$

Wegen $A \neq O$ können nicht alle Eigenwerte gleichzeitig verschwinden. Wir nehmen o.B.d.A. $\lambda_1 \neq 0$ an. Dies ist stets durch geeignete Numerierung der neuen Basisvektoren $\vec{y}_1, \vec{y}_2, \vec{y}_3$ erreichbar. Wir unterscheiden nunmehr diverse Hauptfälle gemäß dem Nichtverschwinden der weiteren Eigenwerte.

Hauptfall I: Es gilt $\lambda_j \neq 0$ für $j = 1, 2, 3$. Der Basiswechsel $(O; \vec{y}_1, \vec{y}_2, \vec{y}_3) \mapsto (O_1; \vec{z}_1, \vec{z}_2, \vec{z}_3)$ mit zugeordneter Koordinatentransformation

$$z_i := y_i + \frac{b_i}{\lambda_i}, \qquad i = 1, 2, 3, \tag{6.7}$$

bewirkt eine *Parallelverschiebung* des Ursprungs ($O_1 = (\frac{b_1}{\lambda_1}, \frac{b_2}{\lambda_2}, \frac{b_3}{\lambda_3})^T$). Man erhält nun anstelle der Gleichung (6.6):

$$\boxed{\lambda_1 z_1^2 + \lambda_2 z_2^2 + \lambda_3 z_3^2 + \alpha = 0, \quad \alpha := b - \Big(\frac{b_1^2}{\lambda_1} + \frac{b_2^2}{\lambda_2} + \frac{b_3^2}{\lambda_3}\Big).}$$

Fall I.1: $\alpha \neq 0$. Nach Division der obigen Gleichung durch $-\alpha$ und Einführung neuer Parameter gemäß

$$\frac{\lambda_1}{\alpha} =: \pm\frac{1}{a^2}, \quad \frac{\lambda_2}{\alpha} =: \pm\frac{1}{b^2}, \quad \frac{\lambda_3}{\alpha} =: \pm\frac{1}{c^2}$$

(mit $a^2 > 0$, $b^2 > 0$, $c^2 > 0$) erhält man je nach Lage der Vorzeichen die folgenden Gleichungsformen:

I.1.1: Ellipsoid

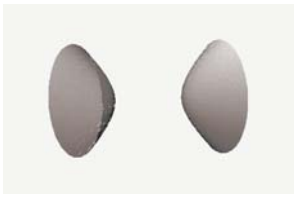
$$\boxed{\frac{z_1^2}{a^2} + \frac{z_2^2}{b^2} + \frac{z_3^2}{c^2} = 1}$$

I.1.2: einschaliges Hyperboloid

$$\boxed{\frac{z_1^2}{a^2} + \frac{z_2^2}{b^2} - \frac{z_3^2}{c^2} = 1}$$

I.1.3: zweischaliges Hyperboloid

$$\boxed{\frac{z_1^2}{a^2} - \frac{z_2^2}{b^2} - \frac{z_3^2}{c^2} = 1}$$



I.1.4: nullteilige Fläche (**nicht** reell)

$$\boxed{-\frac{z_1^2}{a^2} - \frac{z_2^2}{b^2} - \frac{z_3^2}{c^2} = 1}$$

Bemerkung 11.17 (a) Weitere mögliche Vorzeichenkombinationen führen – nach Umnumerierung der Koordinaten – auf einen der Fälle **I.1.2** oder **I.1.3**.

(b) Das **Ellipsoid** I.1.1 ist eine geschlossene, im Endlichen liegende Fläche. Jede Ebene $z_1 = h$ mit $h^2 < a^2$ schneidet diese Fläche in einer Ellipse. Dasselbe gilt für Schnitte senkrecht zu den anderen Koordinatenachsen.

(c) Das **einschalige Hyperboloid** I.1.2 wird von den Ebenen $z_3 = h$, $h \in \mathbf{R}$, in Ellipsen geschnitten, die für $h = 0$ am kleinsten sind. Ebenen $z_1 = h$ bzw. $z_2 = h$ schneiden diese Fläche in Hyperbeln. Eine *Ausnahme* bilden die Schnittlinien mit den Ebenen $z_1 = \pm a$ bzw. $z_2 = \pm b$. Diese bilden ein sich schneidendes Geradenpaar.

(d) Das **zweischalige Hyperboloid** I.1.3 zerfällt in zwei getrennte Teilflächen. Ebenen $z_1 = h$ mit $h^2 < a^2$ haben *keinen* gemeinsamen Schnittpunkt mit der Fläche; Ebenen $z_1 = h$ mit $h^2 \geq a^2$ schneiden die Fläche in Ellipsen. Die Schnittlinien der Fläche mit den Ebenen $z_2 = h$ bzw. $z_3 = h$ sind stets Hyperbeln. □

Fall I.2: $\alpha = 0$. Mit entsprechend modifizierten Bezeichnungen aus dem vorangegangenen **Fall I.1** erhält man hier:

I.2.1: komplexer Kegel mit reeller Spitze in $(0,0,0)^T$

$$\boxed{\frac{z_1^2}{a^2}+\frac{z_2^2}{b^2}+\frac{z_3^2}{c^2}=0}$$

I.2.2: reeller Kegel

$$\boxed{\frac{z_1^2}{a^2}+\frac{z_2^2}{b^2}-\frac{z_3^2}{c^2}=0}$$

Bemerkung 11.18 (a) Alle weiteren Vorzeichenkombinationen sind durch den Fall **I.2.2** erfasst.

(b) Mit dem Punkt $(z_1,z_2,z_3)^T$ liegt auch jeder Punkt $(\lambda z_1,\lambda z_2,\lambda z_3)^T$, $\lambda\in\mathbf{R}$, auf dem **reellen Kegel**; d.h. Geraden durch die Spitze $(0,0,0)^T$ und einen beliebigen Kegelpunkt $(z_1,z_2,z_3)^T$ liegen ganz auf dieser Fläche. Ebenen $z_3=h$ schneiden den Kegel in Ellipsen; Ebenen $z_1=h$ oder $z_2=h$ schneiden den Kegel in Hyperbeln. □

Hauptfall II: Es gilt $\lambda_1\neq 0\neq\lambda_2$, aber $\lambda_3=0$. Wir wenden die Transformation (6.7) nur noch auf die beiden ersten Koordinatenachsen an, was einer Translation der (y_1,y_2)–Ebene entspricht. Man erhält nun anstelle der Gleichung (6.6):

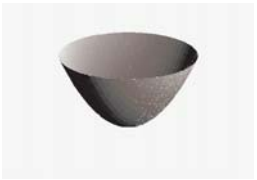
$$\boxed{\lambda_1 z_1^2+\lambda_2 z_2^2+2b_3y_3+\alpha=0,\quad \alpha:=b-\big(\frac{b_1^2}{\lambda_1}+\frac{b_2^2}{\lambda_2}\big).}$$

Wir unterscheiden Unterfälle gemäß $b_3\neq 0$ und $b_3=0$, $\alpha\neq 0$ sowie $b_3=0=\alpha$.

Fall II.1: $b_3\neq 0$. Die Translation $z_3:=y_3+\frac{\alpha}{2b_3}$ führt mit $p:=\pm b_3\neq 0$ auf folgende Gleichungsformen:

II.1.1: elliptisches Paraboloid

$$\boxed{\frac{z_1^2}{a^2}+\frac{z_2^2}{b^2}=2pz_3}$$



II.1.2: hyperbolisches Paraboloid

$$\boxed{\frac{z_1^2}{a^2}-\frac{z_2^2}{b^2}=2pz_3}$$

Bemerkung 11.19 (a) Falls $p>0$ gilt, so besitzt das **elliptische Paraboloid** nur für $z_3\geq 0$ reelle Punkte. Ebenen $z_3=h>0$ schneiden die Fläche in Ellipsen. Die Ebenen $z_1=h$ bzw. $z_2=h$ schneiden die Fläche in Parabeln. Entsprechendes gilt für $p<0$.

(b) Das **hyperbolische Paraboloid** wird von Ebenen $z_3=h$ in Hyperbeln, und von Ebenen $z_1=h$ bzw. $z_2=h$ in Parabeln geschnitten. □

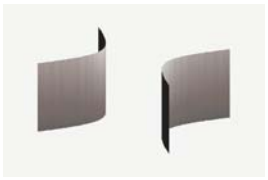
Fall II.2: $b_3 = 0$ **und** $\alpha \neq 0$. Nach Division durch $-\alpha$ erhält man in Analogie zum **Fall I.1:**

II.2.1: elliptischer Zylinder

$$\frac{z_1^2}{a^2} + \frac{z_2^2}{b^2} = 1$$

II.2.2: hyperbolischer Zylinder

$$\frac{z_1^2}{a^2} - \frac{z_2^2}{b^2} = 1$$



II.2.3: nullteiliger Zylinder (**nicht** reell)

$$-\frac{z_1^2}{a^2} - \frac{z_2^2}{b^2} = 1$$

Bemerkung 11.20 (a) Der **elliptische Zylinder** II.2.1 hat prismatische Form; die Schnittlinien mit Ebenen $z_3 = h$ sind Ellipsen. Ebenen $z_1 = h$ bzw. $z_2 = h$ schneiden die Fläche in zwei parallelen Geraden.

(b) Beim **hyperbolischen Zylinder** II.2.2 sind die Schnittlinien mit den Ebenen $z_3 = h$ Hyperbeln. Ansonsten gilt das unter (a) Gesagte. □

Fall II.3: $b_3 = 0$ **und** $\alpha = 0$. Hier treten zwei mögliche Fälle auf:

II.3.1: zwei komplexe Ebenen, die sich in der reellen z_3–Achse schneiden

$$\frac{z_1^2}{a^2} + \frac{z_2^2}{b^2} = 0$$

II.3.2: zwei reelle, sich schneidende Ebenen

$$\frac{z_1^2}{a^2} - \frac{z_2^2}{b^2} = 0$$

Bemerkung 11.21 Wir schreiben **II.3.2** in der Form $(\frac{z_1}{a} + \frac{z_2}{b})(\frac{z_1}{a} - \frac{z_2}{b}) = 0$ und ersehen, dass in der (z_1, z_2)–Ebene die sich schneidenden Geraden $\frac{z_1}{a} + \frac{z_2}{b} = 0$, $\frac{z_1}{a} - \frac{z_2}{b} = 0$ durch den Ursprung $(0, 0)^T$ definiert werden. Durch Verschiebung dieser Geraden längs der z_3–Achse entstehen dann zwei sich **schneidende Ebenen.** □

Hauptfall III: Es gilt $\lambda_1 \neq 0$, aber $\lambda_2 = 0 = \lambda_3$. Wir erhalten hier die folgende Form der Gleichung (6.6):

$$\lambda_1 y_1^2 + 2b_1 y_1 + 2b_2 y_2 + 2b_3 y_3 + b = 0.$$

Durch Drehung des (y_2, y_3)–Koordinatensystems in ein (z_2, z_3)–System kann erreicht werden, dass der Koeffizient von z_3 verschwindet. Hierzu setzen wir

$$\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} z_2 \\ z_3 \end{bmatrix}.$$

Dann folgt $b_2 y_2 + b_3 y_3 = (b_2 \cos\varphi + b_3 \sin\varphi) z_2 + (-b_2 \sin\varphi + b_3 \cos\varphi) z_3$. Mit

$$\varphi := \begin{cases} \frac{\pi}{2} & : \ b_2 = 0, \\ \arctan_H \frac{b_3}{b_2} & : \ b_2 \neq 0 \end{cases}$$

wird das Verlangte erreicht: man erhält

$$\lambda_1 y_1^2 + 2b_1 y_1 + 2\beta z_2 + b = 0, \qquad \beta := b_2 \cos\varphi + b_3 \sin\varphi.$$

Nach einer weiteren Parallelverschiebung $y_1 =: z_1 - \frac{b_1}{\lambda_1}$ resultiert schließlich die Normalform

$$\boxed{\lambda_1 z_1^2 + 2\beta z_2 + \alpha = 0, \quad \alpha := b - \frac{b_1^2}{\lambda_1}.}$$

Fall III.1: $\beta \neq 0$. Durch nochmalige Parallelverschiebung $\hat{z}_2 := z_2 + \frac{\alpha}{2\beta}$ gelangt man zu:

III.1.1: parabolischer Zylinder

$$\boxed{z_1^2 = 2p\hat{z}_2, \qquad p := -\frac{\beta}{\lambda_1} \neq 0}$$

Bemerkung 11.22 Der **parabolische Zylinder** entsteht durch Verschiebung der Parabel $z_1^2 = 2p\hat{z}_2$ längs der z_3–Achse. □

Fall III.2: $\beta = 0$ **und** $\alpha \neq 0$. Nach Division durch $-\alpha$ erhält man die folgenden zwei Gleichungsformen:

III.2.1: reelle parallele Ebenen $z_1 = a, z_1 = -a$

$$\boxed{\frac{z_1^2}{a^2} = 1}$$

III.2.2: Paar konjugiert komplexer Ebenen

$$\boxed{-\frac{z_1^2}{a^2} = 1}$$

Fall III.3: $\beta = 0$ **und** $\alpha = 0$.

III.3.1: Doppelebene

$$\boxed{z_1^2 = 0}$$

Es tritt die doppelt zu zählende (z_2, z_3)–Ebene mit der Gleichung $z_1 = 0$ als (entartete) Fläche 2.Ordnung auf.

Hiermit haben wir eine vollständige Aufzählung aller Flächen 2.Ordnung in $\mathbf{R}^3$ erbracht, welche durch die Gleichung (6.2) definiert werden. Die Hintereinanderausführung der beiden Transformationen (6.4) und (6.7) heißt **Hauptachsentransformation**; die in den einzelnen Fällen angegebenen Koordinatenachsen $\vec{z}_1, \vec{z}_2, \vec{z}_3$ heißen **Hauptachsen** der Quadrik.

Eine analoge Betrachtung kann auch in $\mathbf{R}^n$ durchgeführt werden. Allgemein heißt die Menge aller Punkte $P \in \mathbf{R}^n$ eine **Fläche 2.Ordnung** oder eine **Quadrik** in $\mathbf{R}^n$, wenn der Ortsvektor $OP =: \vec{x} = (x_1, x_2, \ldots, x_n)^T$ einer Gleichung

$$\sum_{i,k=1}^{n} a_{ik} x_i x_k + 2 \sum_{i=1}^{n} a_i x_i + b = 0$$

genügt. Dabei wird vorausgesetzt, dass die Matrix $A := (a_{ik}) \in \mathbf{R}^{(n,n)}$ symmetrisch ist. Durch Bestimmung der Eigenvektoren von A erhält man wieder die Hauptachsenrichtungen der Quadrik, und durch anschließende Parallelverschiebung gewinnt man die Normalformen

$$\sum_{i=1}^{n} \lambda_i z_i^2 + \alpha = 0 \quad \text{oder} \quad \sum_{i=1}^{n-1} \lambda_i z_i^2 = 2 p z_n,$$

in der nicht alle λ_i verschwinden. Den Sonderfall $n = 2$ der Kegelschnitte erhält man sofort aus dem behandelten Fall, wenn die dritte Koordinate (x_3 bzw. y_3 bzw. z_3) Null gesetzt wird.

BSP. (11.6.1) Gegeben sei in $\mathbf{R}^3$ eine Quadrik mit der Gleichung

$$-x_1^2 + 2\sqrt{3} x_1 x_3 + 2x_2^2 + x_3^2 - 4x_2 - 8\sqrt{3} x_3 + 12 = 0.$$

Hier gilt

$$A = \begin{bmatrix} -1 & 0 & \sqrt{3} \\ 0 & 2 & 0 \\ \sqrt{3} & 0 & 1 \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} 0 \\ -2 \\ -4\sqrt{3} \end{bmatrix}, \quad b = 12.$$

Wir berechnen die Eigenwerte von A und die zugeordneten Eigenvektoren. Aus der charakteristischen Gleichung

$$\det(A - \lambda\, Id) = \begin{vmatrix} -1-\lambda & 0 & \sqrt{3} \\ 0 & 2-\lambda & 0 \\ \sqrt{3} & 0 & 1-\lambda \end{vmatrix} = (2-\lambda)(\lambda-2)(\lambda+2) \stackrel{!}{=} 0$$

resultieren die Eigenwerte $\lambda_1 = \lambda_2 = 2$ sowie $\lambda_3 = -2$. Die Eigenvektoren zu $\lambda_{1/2} = 2$ berechnet man aus dem homogenen linearen Gleichungssystem

$$(A - 2\, Id)\vec{y} = \begin{bmatrix} -3 & 0 & \sqrt{3} \\ 0 & 0 & 0 \\ \sqrt{3} & 0 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \stackrel{!}{=} \vec{0}.$$

Es ergibt sich $\sqrt{3} y_1 = y_3$, während y_2 frei wählbar ist. Hieraus rekrutieren sich zum Beispiel die zwei orthonormierten Eigenvektoren

$$\vec{y}_1 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{y}_2 := \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ \sqrt{3} \end{bmatrix}.$$

Einen Eigenvektor zu $\lambda_3 = -2$ erhält man aus dem linearen Gleichungssystem

$$(A + 2\,Id)\vec{y} = \begin{bmatrix} 1 & 0 & \sqrt{3} \\ 0 & 4 & 0 \\ \sqrt{3} & 0 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \stackrel{!}{=} \vec{0},$$

nämlich

$$\vec{y}_3 := \frac{1}{2}\,(-\sqrt{3}, 0, 1)^T.$$

Es resultiert die Koordinatentransformation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\frac{1}{2}\begin{bmatrix} 0 & 1-\sqrt{3} \\ 2 & 0 & 0 \\ 0 & \sqrt{3} & 1 \end{bmatrix}}_{=:Y} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \vec{b} = Y^T\vec{a} = \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & \sqrt{3} \\ -\sqrt{3} & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \\ -4\sqrt{3} \end{bmatrix} = \begin{bmatrix} -2 \\ -6 \\ -2\sqrt{3} \end{bmatrix}.$$

Die transformierte Gleichung der Quadrik lautet jetzt

$$2y_1^2 + 2y_2^2 - 2y_3^2 - 4y_1 - 12y_2 - 4\sqrt{3}y_3 + 12 = 0.$$

Nach Durchführung der Parallelverschiebung $z_1 := y_1 - 1,\ z_2 := y_2 - 3,\ z_3 := y_3 + \sqrt{3}$ ergibt sich die Normalform $2z_1^2 + 2z_2^2 - 2z_3^2 - 2 = 0$ oder äquivalent

$$\boxed{z_1^2 + z_2^2 - z_3^2 = 1.}$$

Es liegt der Fall I.1.2 eines einschaligen (Rotations–)Hyperboloids vor mit der z_3–Achse als Rotationsachse. Um die Normalform der Quadrik zu erhalten, wurde hier insgesamt die Transformation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & 1-\sqrt{3} \\ 2 & 0 & 0 \\ 0 & \sqrt{3} & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ \sqrt{3} \end{bmatrix}$$

durchgeführt.

11.7 Hauptvektoren

Wir bezeichnen hier wieder mit $\mathbf{K}$ entweder den Körper $\mathbf{R}$ der reellen Zahlen oder den Körper $\mathbf{C}$ der komplexen Zahlen. Eine Matrix $A \in \mathbf{K}^{(n,n)}$ hat stets genau n Eigenwerte $\lambda_j \in \mathbf{C}$, wenn jeder Eigenwert gemäß seiner Vielfachheit gezählt wird. Sind also $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbf{C}$ die paarweise verschiedenen Eigenwerte von A mit Vielfachheiten $k_1, k_2, \ldots, k_m$, $1 \le m \le n$, so gilt stets $k_1 + k_2 + \cdots + k_m = n$.

Wir hatten k_j die **algebraische Dimension** des Eigenwertes λ_j genannt:

$$\boxed{k_j = \text{algebr. dim}\,\lambda_j.}$$

Jedem Eigenwert λ_j der Matrix A ist (mindestens) ein Eigenvektor $\vec{v}_j \in \mathbf{C}^n$ zugeordnet; dieser ist Lösung des homogenen linearen Gleichungssystems $(A - \lambda_j\,Id)\vec{v}_j = \vec{0}$, also ein Element des Unterraumes $\text{Kern}\,(A - \lambda_j\,Id) \subset \mathbf{C}^n$. Ist $\mathbf{K} = \mathbf{R}$, so muss dem Eigenwert λ_j einer Matrix $A \in \mathbf{R}^{(n,n)}$ im allgemeinen **kein** Eigenvektor $\vec{v}_j \in \mathbf{R}^n$ zugeordnet sein:

BSP. (11.7.1) Die reelle Matrix $A := \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix} \in \mathbf{R}^{(2,2)}$ hat das charakteristische Polynom $P_2(\lambda) = \det(A - \lambda\, Id) = \lambda^2 + 1$ und somit das Paar konjugiert komplexer Eigenwerte $\lambda_\pm := \pm i$. Diesem Paar ist ein Paar konjugiert komplexer Eigenvektoren zugeordnet, nämlich

$$\vec{v}_+ := \begin{bmatrix} 2 \\ 1-i \end{bmatrix}, \quad \vec{v}_- := \begin{bmatrix} 2 \\ 1+i \end{bmatrix}.$$

Reelle Eigenvektoren existieren nicht!

Wir betrachten im folgenden Kern $(A - \lambda_j\, Id)$ stets als Unterraum des Vektorraumes $\mathbf{C}^n$ und nennen seine Dimension die **geometrische Dimension** des Eigenwertes λ_j:

$$\boxed{\rho(\lambda_j) := \text{geom. dim}\, \lambda_j = \dim \text{Kern}\, (A - \lambda_j\, Id).}$$

Wir hatten in Satz 11.3(d) festgestellt, dass genau dann $\rho(\lambda_j) = k_j$ für alle $j = 1, 2, \ldots, m$ gilt, wenn es in $\mathbf{C}^n$ eine Basis von Eigenvektoren der Matrix $A \in \mathbf{K}^{(n,n)}$ gibt. Das heißt, im "schlimmen" Fall $\rho(\lambda_j) < k_j$ (für mindestens ein j) kann der Vektorraum $\mathbf{C}^n$ nicht mehr vollständig durch Eigenvektoren aufgespannt werden. In diesem Fall gelten die meisten Sätze aus Abschnitt 11.3 nicht mehr, und viele Eigenschaften aus Abschnitt 11.2 gelten ebenfalls nicht. Der Begriff des Eigenvektors kann jedoch in der folgenden Weise verallgemeinert werden:

Definition 11.10 *Ein Vektor $\vec{0} \neq \vec{w} \in \mathbf{C}^n$ heiße* **Hauptvektor der Stufe** *$k \in \mathbf{N}$ zum Eigenwert λ der Matrix $A \in \mathbf{K}^{(n,n)}$, wenn gilt*

$$\boxed{(A - \lambda\, Id)^k \vec{w} = \vec{0} \quad \textit{und} \quad (A - \lambda\, Id)^{k-1} \vec{w} \neq \vec{0}.}$$

Aus dieser Definition resultieren einige einfache Folgerungen.

Folgerung 11.1 *Die Hauptvektoren erster Stufe zum Eigenwert λ sind genau die Eigenvektoren zum Eigenwert λ.*

Folgerung 11.2 *Ist $\vec{w} \neq \vec{0}$ ein Hauptvektor der Stufe $k \geq 2$ zum Eigenwert λ, so ist $(A - \lambda\, Id)\vec{w}$ ein Hauptvektor der Stufe $k-1$ zum selben Eigenwert λ.*

Begründung: Setzt man $\vec{v} := (A - \lambda\, Id)\vec{w}$, so ist $(A - \lambda\, Id)^{k-1}\vec{v} = \vec{0}$, aber $(A - \lambda\, Id)^{k-2}\vec{v} \neq \vec{0}$, und somit auch $\vec{v} \neq \vec{0}$. □

Folgerung 11.3 *Die Hauptvektoren der Stufe $k \geq 2$ zum Eigenwert λ sind von den Hauptvektoren der Stufen $\leq k-1$ zum selben Eigenwert λ linear unabhängig.*

Begründung: Sind $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$, $r \leq n$, die Hauptvektoren der Stufen $\leq k-1$, so folgte aus dem Ansatz der linearen Abhängigkeit

$$\vec{w} = \sum_{j=1}^{r} \mu_j \vec{v}_j \quad \text{mit} \quad |\mu_1| + |\mu_2| + \cdots + |\mu_r| \neq 0$$

die widersprüchliche Bedingung

$$(A - \lambda\, Id)^{k-1}\vec{w} = \sum_{j=1}^{r} \mu_j \underbrace{(A - \lambda\, Id)^{k-1}\vec{v}_j}_{=\vec{0}} = \vec{0}.$$

Folgerung 11.4 *Die Gesamtheit aller Hauptvektoren der Stufe $\leq k$ zum Eigenwert λ spannt den Unterraum* $\operatorname{Kern}(A-\lambda\,Id)^k \subset \mathbf{C}^n$ *auf.*

Folgerung 11.5 *Für jeden Eigenwert λ der Matrix $A \in \mathbf{K}^{(n,n)}$ und jede Zahl $k \in \mathbf{N}$ gilt:*

$$\boxed{\{\vec{0}\} \subset \operatorname{Kern}(A-\lambda\,Id) \subseteq \operatorname{Kern}(A-\lambda\,Id)^2 \subseteq \cdots \subseteq \operatorname{Kern}(A-\lambda\,Id)^k.} \tag{7.1}$$

Ist $\operatorname{Kern}(A-\lambda\,Id)^k = \operatorname{Kern}(A-\lambda\,Id)^{k+1}$ *für ein $k \in \mathbf{N}$ erfüllt, so folgt*

$$\boxed{\operatorname{Kern}(A-\lambda\,Id)^r = \operatorname{Kern}(A-\lambda\,Id)^{r+1} \;\; \forall\, r \geq k.} \tag{7.2}$$

Begründung: Sicher gilt (7.1), und deswegen auch $\operatorname{Kern}(A-\lambda\,Id)^r \subseteq \operatorname{Kern}(A-\lambda\,Id)^{r+1}$. Wir zeigen nun durch vollständige Induktion nach r die umgekehrte Inklusion $\operatorname{Kern}(A-\lambda\,Id)^{r+1} \subseteq \operatorname{Kern}(A-\lambda\,Id)^r,\ r \geq k$. Für $r=k$ folgt diese aus der Voraussetzung.
Vererbung: Gelte schon $\operatorname{Kern}(A-\lambda\,Id)^{r+1} \subseteq \operatorname{Kern}(A-\lambda\,Id)^r$ für ein $r > k$. Sei $\vec{v} \in \operatorname{Kern}(A-\lambda\,Id)^{r+2}$, also $\vec{0} = (A-\lambda\,Id)^{r+2}\vec{v} = (A-\lambda\,Id)^{r+1}(A-\lambda\,Id)\vec{v}$. Somit ist $(A-\lambda\,Id)\vec{v} \in \operatorname{Kern}(A-\lambda\,Id)^{r+1}$, und nach Induktionsannahme auch $(A-\lambda\,Id)\vec{v} \in \operatorname{Kern}(A-\lambda\,Id)^r$. Dies impliziert $(A-\lambda\,Id)^{r+1}\vec{v} = \vec{0}$, also $\vec{v} \in \operatorname{Kern}(A-\lambda\,Id)^{r+1}$. □

Sicher gilt $\dim \operatorname{Kern}(A-\lambda\,Id)^k \leq n$ für jedes $k \in \mathbf{N}$. Also können wir aus den Folgerungen 11.3, 11.4 und 11.5 die Existenz eines **kleinsten** Index $f = f(\lambda)$ erschließen, für den gilt:

$$\boxed{\{\vec{0}\} \subset \operatorname{Kern} A_\lambda \subset \operatorname{Kern} A_\lambda^2 \subset \cdots \subset \operatorname{Kern} A_\lambda^f = \operatorname{Kern} A_\lambda^{f+1},} \tag{7.3}$$

worin die ersten f Inklusionen echt sind. Hier und im folgenden gelte

$$\boxed{A_\lambda := A - \lambda\,Id.}$$

Definition 11.11 *Die durch die Eigenschaft (7.3) charakterisierte Zahl $f = f(\lambda) \in \mathbf{N}$ heiße der* **Fittingindex** *oder kurz der* **Index** *des Eigenwertes λ einer Matrix $A \in \mathbf{K}^{(n,n)}$. Für diesen Index gilt stets $f(\lambda) \leq n$. Der Unterraum* $\operatorname{Kern}(A-\lambda\,Id)^f \subseteq \mathbf{C}^n$ *heiße der* **verallgemeinerte Eigenraum** *oder der* **Hauptraum** *von A zum Eigenwert λ.*

Gemäß Folgerung 11.4 wird der verallgemeinerte Eigenraum $\operatorname{Kern}(A-\lambda\,Id)^f$ von der Gesamtheit der Hauptvektoren der Stufe $\leq f$ zum Eigenwert λ aufgespannt. Dabei ist eine Antwort auf die folgenden Fragen zu finden:

(I) Wie groß ist $\dim \operatorname{Kern}(A-\lambda\,Id)^f$?

(II) Wie berechnet man die Hauptvektoren der Stufe $\leq f$ zum Eigenwert λ?

Die Antwort auf die Frage (I) wird in dem folgenden Satz gegeben, dessen Beweis wir hier nicht führen können (vgl. zum Beispiel R. Walter, Einführung in die Lineare Algebra. Vieweg–Verlag, Braunschweig 1982).

Satz 11.23 *Gegeben seien die Matrix $A \in \mathbf{K}^{(n,n)}$ und ein Eigenwert λ von A mit* geom. dim λ $<$ algebr. dim $\lambda = k$. *Dann gibt es zum Eigenwert λ genau k linear unabhängige Hauptvektoren von A und somit eine Zahl $f \leq n$ mit*

$$\boxed{\dim \operatorname{Kern}(A-\lambda\,Id) < \cdots < \dim \operatorname{Kern}(A-\lambda\,Id)^f = \dim \operatorname{Kern}(A-\lambda\,Id)^{f+1} = k.}$$

Die Antwort auf Frage (II) finden wir in der obigen Folgerung 11.2. Die Hauptvektoren $\vec{w}_{j+1}$ der Stufe $j+1$ zum Eigenwert λ lösen das inhomogene lineare Gleichungssystem

$$\boxed{A_\lambda \vec{w}_{j+1} = \vec{b}_j,} \tag{7.4}$$

worin $\vec{b}_j$ ein Hauptvektor (Hv) der Stufe j zum selben Eigenwert λ ist. Man beachte, dass wegen $\det A_\lambda = 0$ notwendigerweise **Lösbarkeitsbedingungen** an die rechte Seite $\vec{b}_j$ der Gleichung (7.4) zu stellen sind. Und zwar folgt aus Satz 5.19, dass das lineare Gleichungssystem (7.4) genau dann lösbar ist, wenn gilt

$$\boxed{\vec{b}_j \perp \operatorname{Kern} A_\lambda^* = \operatorname{Kern}(A^* - \bar{\lambda}\, Id).} \tag{7.5}$$

Die Erfüllbarkeit dieser Lösbarkeitsbedingung setzt voraus, dass $\vec{b}_j$ ein **allgemeiner Hauptvektor** der Stufe j ist. Das heißt, wird der Eigenraum $\operatorname{Kern}(A - \lambda\, Id)$ zum Eigenwert λ von den linear unabhängigen Eigenvektoren $\vec{w}_1^{\,1}, \vec{w}_1^{\,2}, \ldots, \vec{w}_1^{\,s}$ aufgespannt, so ist $\vec{b}_j$ eindeutig nur bis auf eine Linearkombination $\sum\limits_{i=1}^{s} \rho_i \vec{w}_1^{\,i}$.

Satz 11.24 *Gegeben seien die Matrix $A \in \mathbf{K}^{(n,n)}$ sowie ein Eigenwert $\lambda \in \mathbf{C}$ von A mit* algebr. $\dim \lambda = k$. *Gilt* $\dim \operatorname{Kern} A_\lambda^j < k$ *und sind $\vec{w}_1^{\,1}, \vec{w}_1^{\,2}, \ldots, \vec{w}_1^{\,s}$ die linear unabhängigen Eigenvektoren zum Eigenwert λ von A sowie $\vec{w}_j^{\,1}, \vec{w}_j^{\,2}, \ldots, \vec{w}_j^{\,r}$ die linear unabhängigen Hauptvektoren der Stufe j zum selben λ, so erhält man alle Hauptvektoren $\vec{w}_{j+1}$ der Stufe $j+1$ zum Eigenwert λ durch Lösen des linearen Gleichungssystems* (7.4). *Dabei muss*

$$\boxed{\vec{b}_j := \begin{cases} \displaystyle\sum_{i=1}^{r} \mu_i \vec{w}_j^{\,i} & : j = 1 \ \textit{und} \ r = s, \\ \displaystyle\sum_{i=1}^{s} \rho_i \vec{w}_1^{\,i} + \sum_{i=1}^{r} \mu_i \vec{w}_j^{\,i} & : j > 1 \end{cases} \quad \textit{mit} \quad |\mu_1| + |\mu_2| + \cdots + |\mu_r| \neq 0}$$

der Lösbarkeitsbedingung (7.5) *unterworfen werden.*

Begründung: (a) Zunächst ist klar, dass $\vec{b}_j$ für jede Wahl von μ_i gemäß $|\mu_1| + |\mu_2| + \cdots + |\mu_r| \neq 0$ ein Hauptvektor der Stufe j zum Eigenwert λ ist. Sei nun $\vec{w}_{j+1}$ ein Hv der Stufe $j+1$ zum selben Eigenwert λ. Dann gilt wegen Folgerung 11.2 die Gleichung (7.4). Wir erhalten $\vec{b}_j \in \operatorname{Bild} A_\lambda = (\operatorname{Kern} A_\lambda^*)^\perp$, und somit die Lösbarkeitsbedingung (7.5). Gilt umgekehrt die Lösbarkeitsbedingung (7.5), so ist das lineare Gleichungssystem (7.4) lösbar. Für jede Lösung $\vec{w}_{j+1}$ gilt $A_\lambda^{j+1} \vec{w}_{j+1} = A_\lambda^j \vec{b}_j = \vec{0}$ sowie $A_\lambda^j \vec{w}_{j+1} = \sum\limits_{i=1}^{r} \mu_i A_\lambda^{j-1} \vec{w}_j^{\,i} \neq \vec{0}$. Dies ergibt sich aus der Folgerung 11.3 und der linearen Unabhängigkeit der Vektoren $\vec{w}_j^{\,i}$. Also ist $\vec{w}_{j+1}$ ein Hv der Stufe $j+1$ zum Eigenwert λ.

(b) Wegen Satz 11.23 gilt $(\operatorname{Kern} A_\lambda^{j+1}) \setminus (\operatorname{Kern} A_\lambda^j) \neq \emptyset$. Also muss es Hauptvektoren der Stufe $j+1$ zum Eigenwert λ geben, d.h. das lineare Gleichungssystem (7.4) muss Lösungen zulassen. □

Beginnend mit den Eigenvektoren $\vec{w}_1$ (den Hv der Stufe 1, die das lineare Gleichungssystem $A_\lambda \vec{w}_1 = \vec{0}$ lösen), können also alle Hauptvektoren zum Eigenwert λ durch die Vorschrift (7.4) sukzessive berechnet werden. Hat man insgesamt $k =$ algebr. $\dim \lambda$ linear unabhängige Hauptvektoren berechnet, so bricht das Verfahren von selbst ab: Man erhält aus (7.4) keine weiteren linear unabhängigen Hauptvektoren.

Wir zeigen nun, dass die verallgemeinerten Eigenräume einer Matrix $A \in \mathbf{K}^{(n,n)}$ den Vektorraum $\mathbf{C}^n$ aufspannen.

Satz 11.25 *Es sei $A \in \mathbf{K}^{(n,n)}$ gegeben.*

(a) *Ist $\vec{v}$ ein Hauptvektor der Stufe $j+1$ zum Eigenwert λ, so ist das Vektorsystem $\vec{v}, A_\lambda\vec{v}, \ldots, A_\lambda^j\vec{v}$ linear unabhängig.*

(b) *Für je zwei Eigenwerte $\lambda \neq \mu$ der Matrix A mit Indizes f bzw. g gilt* $\operatorname{Kern}(A-\lambda\,Id)^f \cap \operatorname{Kern}(A-\mu\,Id)^g = \{\vec{0}\}$.

Begründungen: (a) Wir haben $A_\lambda^{j+1}\vec{v} = \vec{0}$. Test auf lineare Abhängigkeit: Angenommen, es wäre

$$\vec{b}_j := \sum_{i=0}^{j} \xi_i A_\lambda^i \vec{v} = \vec{0}.$$

Dann erhielten wir sukzessive $\vec{0} = A_\lambda^j\vec{b}_j = \xi_0 A_\lambda^j\vec{v}$, also $\xi_0 = 0$; weiter $\vec{0} = A_\lambda^{j-1}\vec{b}_j = \xi_1 A_\lambda^j\vec{v}$, also $\xi_1 = 0$; ...; und schließlich $\vec{0} = \vec{b}_j = \xi_j A_\lambda^j\vec{v}$, also $\xi_j = 0$. Somit muss das angegebene Vektorsystem linear unbhängig sein.

(b) Sei zum Beispiel $f \geq g$. Wäre $\vec{0} \neq \vec{v} \in (\operatorname{Kern} A_\lambda^f) \cap (\operatorname{Kern} A_\mu^g)$, so gäbe es eine Zahl $0 \leq j \leq f-1$ mit $A_\lambda^{j+1}\vec{v} = \vec{0}$ und $A_\lambda^j\vec{v} \neq \vec{0}$. Dann wäre

$$\vec{0} = A_\mu^g\vec{v} = \Big((\lambda-\mu)\,Id + (A-\lambda\,Id)\Big)^g\vec{v} = \sum_{i=0}^{g}\binom{g}{i}(\lambda-\mu)^{g-i}A_\lambda^i\vec{v} = \sum_{i=0}^{\min\{g,j\}}\xi_i A_\lambda^i\vec{v},$$

wobei $\xi_i := \binom{g}{i}(\lambda-\mu)^{g-i} \neq 0$ gilt. Dies stände im Widerspruch zum Ergebnis (a). □

Da also die Hauptvektoren zu verschiedenen Eigenwerten von A linear unabhängig sind, erhalten wir schließlich:

Satz 11.26 *Sind $\lambda_1, \lambda_2, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$ mit Vielfachheiten $k_1, k_2, \ldots, k_m$ und bezeichnen $E(\lambda_j) := \operatorname{Kern}(A-\lambda_j\,Id)^{f(\lambda_j)}$ die verallgemeinerten Eigenräume von A, so gilt die direkte Zerlegung*

$$\boxed{\mathbf{C}^n = E(\lambda_1) \oplus E(\lambda_2) \oplus \cdots \oplus E(\lambda_m),}$$

d.h. der Vektorraum $\mathbf{C}^n$ besitzt eine Basis aus Hauptvektoren von A.

Begründung: Satz 11.25(b) stellt zunächst sicher, dass die Summe $E(\lambda_1) \oplus E(\lambda_2) =: U$ direkt ist. Wir zeigen, dass auch $U \cap E(\lambda_3) = \{\vec{0}\}$ gilt. Daraus ergibt sich die Direktheit der Summe $E(\lambda_1) \oplus E(\lambda_2) \oplus E(\lambda_3)$ und somit durch Induktion die Behauptung, weil ja $\dim E(\lambda_j) = k_j$ und $k_1 + k_2 + \cdots + k_m = n$ gelten. Wäre also $U \cap E(\lambda_3) \neq \{\vec{0}\}$, so gäbe es ein $\vec{0} \neq \vec{v} \in E(\lambda_3)$ sowie Vektoren $\vec{w}_1 \in E(\lambda_1)$ und $\vec{w}_2 \in E(\lambda_2)$ mit $\vec{v} = \vec{w}_1 + \vec{w}_2 \in U$. Sei $g := \max\{f(\lambda_1), f(\lambda_2), f(\lambda_3)\}$. Dann wäre

$$\vec{0} = A_{\lambda_3}^g\vec{v} = \Big((\lambda_1-\lambda_3)\,Id + A_{\lambda_1}\Big)^g\vec{w}_1 + \Big((\lambda_2-\lambda_3)\,Id + A_{\lambda_2}\Big)^g\vec{w}_2 = \sum_{i=0}^{g}\xi_i A_{\lambda_1}^i\vec{w}_1 + \sum_{i=0}^{g}\eta_i A_{\lambda_2}^i\vec{w}_2 =: \vec{z}_1 + \vec{z}_2$$

mit $\xi_i := \binom{g}{i}(\lambda_1-\lambda_3)^{g-i} \neq 0 \neq \binom{g}{i}(\lambda_2-\lambda_3)^{g-i} =: \eta_i$ und $\vec{z}_1 \in E(\lambda_1)$ sowie $\vec{z}_2 \in E(\lambda_2)$. Wegen $E(\lambda_1) \cap E(\lambda_2) = \{\vec{0}\}$ folgte $\vec{z}_1 = \vec{0} = \vec{z}_2$, im Widerspruch zu Satz 11.25(a). □

BSP. (11.7.2) Wir betrachten die folgende Matrix $A \in \mathbf{R}^{(3,3)}$ und bestimmen ihr charakteristisches Polynom $P_3(\lambda) = \det(A-\lambda\,Id)$:

$$A := \begin{bmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & -2 \end{bmatrix}, \quad P_3(\lambda) = \begin{vmatrix} -(2+\lambda) & 1 & 0 \\ 0 & -(2+\lambda) & 1 \\ 0 & 0 & -(2+\lambda) \end{vmatrix} = -(2+\lambda)^3.$$

Es resultiert der einzige Eigenwert $\lambda = -2$ mit $k = \text{algebr. dim}\,\lambda = 3$. Die zugeordneten Eigenvektoren $\vec{w}_1$ (= Hv der Stufe 1) sind die Lösungen des homogenen linearen Gleichungssystems

$$\vec{0} = (A+2\,Id)\vec{w}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \quad \text{also} \quad \vec{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Es folgt geom. dim $\lambda = 1$, so dass wir den verallgemeinerten Eigenraum von λ berechnen müssen. Hierzu stellen wir zunächst fest, dass für $\lambda = -2$ gilt:

$$A_\lambda = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_\lambda^T = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \operatorname{Kern} A_\lambda^T = \operatorname{span}\{\vec{v}\}, \quad \vec{v} := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Ganz offensichtlich gilt $\vec{w}_1 \perp \operatorname{Kern} A_\lambda^T$, so dass das inhomogene lineare Gleichungssystem $A_\lambda \vec{w}_2 = \vec{w}_1$ lösbar ist, also Hauptvektoren der Stufe 2 liefert:

$$A_\lambda \vec{w}_2 = \vec{w}_1 \Leftrightarrow \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_2 = \begin{bmatrix} c \\ 1 \\ 0 \end{bmatrix}, \ c \in \mathbf{C}.$$

Es gilt hier $\vec{w}_2 = c\vec{w}_1 + (0,1,0)^T$, wobei der Vektor $\vec{w}_1$ – wie schon festgestellt – den Unterraum $\operatorname{Kern} A_\lambda$ aufspannt und die Bedingung $\vec{w}_1 \perp \operatorname{Kern} A_\lambda^T$ erfüllt. Weil $\vec{w}_2 \perp \operatorname{Kern} A_\lambda^T$ für jedes $c \in \mathbf{C}$ gilt, können wir $c = 0$ wählen. Die Vektoren $\vec{w}_1$ und $\vec{w}_2 := (0,1,0)^T$ sind nun linear unabhängig. Die Lösbarkeitsbedingung $\vec{w}_2 \perp \operatorname{Kern} A_\lambda^T$ stellt sicher, dass das inhomogene lineare Gleichungssystem $A_\lambda \vec{w}_3 = \vec{w}_2$ lösbar ist, also Hauptvektoren der Stufe 3 liefert:

$$A_\lambda \vec{w}_3 = \vec{w}_2 \Leftrightarrow \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_3 = \begin{bmatrix} c_1 \\ 0 \\ 1 \end{bmatrix}, \ c_1 \in \mathbf{C}.$$

Nun kann die Lösbarkeitsbedingung $\vec{w}_3 \perp \operatorname{Kern} A_\lambda^T$ für kein $c_1 \in \mathbf{C}$ mehr erfüllt werden. Das Verfahren der Hauptvektorberechnung bricht ab; allerdings haben wir bereits genügend viele Hauptvektoren bestimmt. Wegen $\vec{w}_3 = c_1\vec{w}_1 + (0,0,1)^T$ können wir $c_1 = 0$ wählen, und wir erhalten so die folgende Basis von Hauptvektoren, die den Vektorraum $\mathbf{C}^3$ aufspannen:

$$\vec{w}_1 = (1,0,0)^T, \quad \vec{w}_2 = (0,1,0)^T, \quad \vec{w}_3 = (0,0,1)^T.$$

Wir haben in dem vorliegenden Beispiel bereits einen wichtigen Sonderfall der Hauptvektorberechnung behandelt, nämlich den

$$\boxed{\textbf{Sonderfall: } 1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda =: k.}$$

In diesem Fall gilt $\dim \operatorname{Kern} A_\lambda = 1$, d.h. es existiert nur ein einziger linear unabhängiger Eigenvektor $\vec{w}_1$ zum Eigenwert λ. Dieser muss wegen Satz 11.24 **automatisch** die Lösbarkeitsbedingung $\vec{w}_1 \perp \operatorname{Kern} A_\lambda^*$ erfüllen, so dass das lineare Gleichungssystem $A_\lambda \vec{w} = \vec{w}_1$ stets eine eindimensionale Lösungsmenge $\vec{w} = \vec{w}_2 + \operatorname{span}\{\vec{w}_1\}$ besitzt, wobei $\vec{w}_2$ eine beliebige partikuläre Lösung des inhomogenen Gleichungssystems ist. Es existiert also genau ein von $\vec{w}_1$ linear unabhängiger Hauptvektor $\vec{w}_2$ der Stufe 2. Im Falle algebr. dim > 2 muss $\vec{w}_2$ die Lösbarkeitsbedingung $\vec{w}_2 \perp \operatorname{Kern} A_\lambda^*$ wieder **automatisch** erfüllen, so dass das lineare Gleichungssystem $A_\lambda \vec{w} = \vec{w}_2$ wiederum eine eindimensionale Lösungsmenge $\vec{w} = \vec{w}_3 + \operatorname{span}\{\vec{w}_1\}$ besitzt, usf. Die Analysis der Hauptvektorberechnung vereinfacht sich in diesem Sonderfall also ganz erheblich:

- das Überprüfen der Lösbarkeitsbedingungen entfällt
- in jeder Stufe j, $2 \le j \le k$, gibt es genau einen von den bereits bestimmten Hauptvektoren linear unabhängigen Hauptvektor $\vec{w}_j$ zum Eigenwert λ. Dieser ist (eine beliebige) partikuläre Lösung des inhomogenen linearen Gleichungssystems

$$\boxed{(A - \lambda\, Id)\vec{w}_j = \vec{w}_{j-1}.} \tag{7.6}$$

BSP. (11.7.3) Die folgende Matrix $A \in \mathbf{R}^{(4,4)}$ hat das charakteristische Polynom $P_4(\lambda) = \det(A - \lambda\, Id) = (3-\lambda)^4$, und somit den vierfachen Eigenwert $\lambda = 3$:

$$A := \begin{bmatrix} 3 & 1 & -2 & 3 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}, \quad P_4(\lambda) = \begin{vmatrix} 3-\lambda & 1 & -2 & 3 \\ 0 & 3-\lambda & -1 & 1 \\ 0 & 0 & 3-\lambda & 0 \\ 0 & 0 & 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^4.$$

Wir bestimmen zunächst die Eigenvektoren $\vec{w}_1$ zum Eigenwert $\lambda = 3$:

$$A_\lambda \vec{w}_1 = \vec{0} \Leftrightarrow \begin{bmatrix} 0 & 1 & -2 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{w}_1 = \begin{bmatrix} c \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad c \in \mathbf{C}.$$

Hier liegt offensichtlich der Sonderfall $1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda = 4$ vor. Wir wählen nun $c = 1$ und gehen wie oben beschrieben vor, d.h. wir bestimmen den Hauptvektor $\vec{w}_2$ der Stufe 2 gemäß

$$A_\lambda \vec{w}_2 = \vec{w}_1 \Leftrightarrow \begin{bmatrix} 0 & 1 & -2 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Hauptvektor $\vec{w}_3$ der Stufe 3:

$$A_\lambda \vec{w}_3 = \vec{w}_2 \Leftrightarrow \begin{bmatrix} 0 & 1 & -2 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_3 = \begin{bmatrix} 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}.$$

Hauptvektor $\vec{w}_4$ der Stufe 4:

$$A_\lambda \vec{w}_4 = \vec{w}_3 \Leftrightarrow \begin{bmatrix} 0 & 1 & -2 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}, \quad \text{also} \quad \vec{w}_4 = \begin{bmatrix} 0 \\ 8 \\ 1 \\ -2 \end{bmatrix}.$$

Hiermit haben wir in $\mathbf{C}^4$ eine Basis $\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\}$ aus Hauptvektoren der Matrix A bestimmt.

11.7.1 Das Verfahren des Kernaustauschs

Im allgemeinen Fall $1 < \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda =: k$ wird es unabdingbar sein, in jeder Stufe die Lösbarkeitsbedingung (7.5) nachzuprüfen. Da wir jetzt auch den Kern der adjungierten Matrix A_λ^* bestimmen müssen, wird das Verfahren zur Berechnung der Hauptvektoren höherer Stufe sehr aufwendig. Das folgende Beispiel soll einen Überblick über den zu betreibenden Aufwand vermitteln. Die dargestellte Vorgehensweise hat zwar allgemeingültigen Charakter, wir weisen aber schon jetzt vorsorglich daraufhin, dass wir mit dem sich anschließenden **Verfahren des Kernaustauschs** zu einer wesentlichen Vereinfachung gelangen werden: Die Berechnung von Kern A_λ^* wird sich vollständig erübrigen, und die Lösbarkeitsbedingungen (7.5) werden sich jeweils zwangsläufig aus dem GAUSS–Algorithmus ergeben.

BSP. (11.7.4) Wir betrachten die folgende Matrix $A \in \mathbf{R}^{(5,5)}$, deren charakteristisches Polynom $P_5(\lambda) = \det(A - \lambda\, Id) = \lambda^5$ ist und die somit genau einen Eigenwert $\lambda = 0$ mit $\text{algebr. dim}\,\lambda = 5 =: k$

besitzt:

$$A := \begin{bmatrix} -2 & 0 & 2 & 1 & 4 \\ 4 & 0 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ -4 & 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad P_5(\lambda) = \begin{vmatrix} -2-\lambda & 0 & 2 & 1 & 4 \\ 4 & -\lambda & -4 & -2 & 0 \\ 0 & 0 & -\lambda & 0 & -4 \\ -4 & 0 & 4 & 2-\lambda & 0 \\ 0 & 0 & 0 & 0 & -\lambda \end{vmatrix}.$$

Mit Hilfe der Matrix $A_\lambda := (A - \lambda\, Id) = A$ berechnen wir zunächst alle zugeordneten Eigenvektoren $\vec{w}_1$. Es gilt

$$A_\lambda \vec{w}_1 = \vec{0} \;\rightleftharpoons\; \left[\begin{array}{ccccc|c} -2 & 0 & 2 & 1 & 4 & 0 \\ 4 & 0 & -4 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 \\ -4 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \overset{\text{Gauss}}{\rightleftharpoons} \left[\begin{array}{ccccc|c} -2 & \boxed{0} & \boxed{2} & \boxed{1} & 4 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \rightleftharpoons \vec{w}_1 = \begin{bmatrix} c_2 + c_3 \\ c_1 \\ c_2 \\ 2c_3 \\ 0 \end{bmatrix},$$

($c_j \in \mathbf{C}$). Es resultieren drei linear unabhängige Eigenvektoren $\vec{w}_1^{\,1}$, $\vec{w}_1^{\,2}$, $\vec{w}_1^{\,3}$, die wir zum Beispiel durch die Parameterwahl $(c_1, c_2, c_3) := (1,0,0),\ := (0,1,0),\ := (0,0,1)$ spezifizieren können:

$$\vec{w}_1^{\,1} = (0,1,0,0,0)^T, \quad \vec{w}_1^{\,2} = (1,0,1,0,0)^T, \quad \vec{w}_1^{\,3} = (1,0,0,2,0)^T.$$

Offenbar gilt $3 = \text{geom. dim}\,\lambda < k = 5$, so dass noch zwei weitere Hauptvektoren zum Eigenwert $\lambda = 0$ berechnet werden müssen. Wir haben $\text{Kern}\, A_\lambda^T = \text{span}\,\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ mit

$$A_\lambda^T = A^T = \begin{bmatrix} -2 & 4 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -4 & 0 & 4 & 0 \\ 1 & -2 & 0 & 2 & 0 \\ 4 & 0 & -4 & 0 & 0 \end{bmatrix}, \quad \vec{u}_1 := \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \ \vec{u}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \vec{u}_3 := \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}.$$

Gemäß Satz 11.24 lautet die Lösbarkeitsbedingung (7.5) hier

$$\vec{b}_1 := \mu_1 \vec{w}_1^{\,1} + \mu_2 \vec{w}_1^{\,2} + \mu_3 \vec{w}_1^{\,3} = (\mu_2 + \mu_3,\ \mu_1,\ \mu_2,\ 2\mu_3,\ 0)^T \perp \text{span}\,\{\vec{u}_1, \vec{u}_2, \vec{u}_3\},$$

und sie führt wegen $\vec{b}_1 \perp \vec{u}_1$ auf die beiden Gleichungen

$$0 = \langle \vec{b}_1, \vec{u}_2 \rangle = \mu_1 + 2\mu_3, \quad 0 = \langle \vec{b}_1, \vec{u}_3 \rangle = \mu_1 + 4\mu_2 + 2\mu_3$$

mit den Lösungen $\mu_2 = 0$ und $\mu_1 = -2\mu_3$. Wählen wir z.B. $\mu_3 := 1$, so haben wir jetzt $\vec{b}_1 = (1,-2,0,2,0)^T \perp \text{Kern}\, A_\lambda^T$, und somit die Lösbarkeit von

$$A_\lambda \vec{w}_2 = \vec{b}_1 \;\rightleftharpoons\; \left[\begin{array}{ccccc|c} -2 & 0 & 2 & 1 & 4 & 1 \\ 4 & 0 & -4 & -2 & 0 & -2 \\ 0 & 0 & 0 & 0 & -4 & 0 \\ -4 & 0 & 4 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \overset{\text{Gauss}}{\rightleftharpoons} \left[\begin{array}{ccccc|c} -2 & \boxed{0} & \boxed{2} & \boxed{1} & 4 & 1 \\ 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \rightleftharpoons \vec{w}_2 = \begin{bmatrix} -\frac{1}{2} + c_2 + c_3 \\ c_1 \\ c_2 \\ 2c_3 \\ 0 \end{bmatrix},$$

($c_j \in \mathbf{C}$). Wegen $\vec{w}_2 = c_1 \vec{w}_1^{\,1} + c_2 \vec{w}_1^{\,2} + c_3 \vec{w}_1^{\,3} + (-\frac{1}{2}, 0, 0, 0, 0)^T$ bekommen wir nur einen einzigen, von den bereits berechneten Eigenvektoren linear unabhängigen Hauptvektor der Stufe 2, nämlich (wenn wir z.B. $c_1 = c_2 = c_3 = 0$ wählen)

$$\vec{w}_2^{\,1} = \Big(-\tfrac{1}{2}, 0, 0, 0, 0\Big)^T.$$

Es ist erforderlich, noch einen weiteren Hauptvektor der Stufe 3 zu berechnen. Gemäß Satz 11.24 lautet die Lösbarkeitsbedingung (7.5) nun:

$$\vec{b}_1 := \rho_1 \vec{w}_1^{\,1} + \rho_2 \vec{w}_1^{\,2} + \rho_3 \vec{w}_1^{\,3} + \mu \vec{w}_2^{\,1} = \Big(-\tfrac{\mu}{2} + \rho_2 + \rho_3,\ \rho_1,\ \rho_2,\ 2\rho_3,\ 0\Big)^T \perp \text{span}\,\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}.$$

Da bereits $\vec{b}_1 \perp \vec{u}_1$ erfüllt ist, haben wir noch

$$0 = \langle \vec{b}_1, \vec{u}_2 \rangle = \rho_1 + 2\rho_3, \quad 0 = \langle \vec{b}_1, \vec{u}_3 \rangle = -\mu + \rho_1 + 4\rho_2 + 2\rho_3$$

mit den Lösungen $\rho_2 = \frac{\mu}{4}$ und $\rho_1 = -2\rho_3$. Der Satz 11.24 verbietet es, $\mu = 0$ zu setzen. Hingegen ist es erlaubt, $\rho_3 = 0$ zu wählen. Für $\rho_3 \neq 0$ enthält nämlich die Lösungsmenge des inhomogenen Systems $A_\lambda \vec{w}_3 = \vec{b}_1$ den schon berechneten Lösungsanteil $\rho_3 \vec{w}_2^1$, den wir hier weglassen dürfen. Setzen wir also z.B. $\mu = 4$ und $\rho_3 = 0$, so haben wir jetzt $\vec{b}_1 = (-1, 0, 1, 0, 0)^T \perp \operatorname{Kern} A_\lambda^T$, und somit wiederum die Lösbarkeit von

$$A_\lambda \vec{w}_3 = \vec{b}_1 \rightleftharpoons \left(\begin{array}{rrrrr|r} -2 & 0 & 2 & 1 & 4 & -1 \\ 4 & 0 & -4 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 1 \\ -4 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \overset{\text{Gauss}}{\rightleftharpoons} \left(\begin{array}{rrrrr|r} -2 & \boxed{0} & \boxed{2} & \boxed{1} & 4 & -1 \\ 0 & 0 & 0 & 0 & 8 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \rightleftharpoons \vec{w}_3 = \begin{bmatrix} c_2 + c_3 \\ c_1 \\ c_2 \\ 2c_3 \\ -\frac{1}{4} \end{bmatrix},$$

($c_j \in \mathbf{C}$). Wegen $\vec{w}_3 = c_1 \vec{w}_1^1 + c_2 \vec{w}_1^2 + c_3 \vec{w}_1^3 + (0, 0, 0, 0, -\frac{1}{4})^T$ können wir $c_1 = c_2 = c_3 = 0$ setzen. Die Vektoren $\vec{w}_1^1$, $\vec{w}_1^2$, $\vec{w}_1^3$, $\vec{w}_2^1$ und

$$\vec{w}_3^1 := \left(0, 0, 0, 0, -\tfrac{1}{4}\right)^T$$

spannen nun den verallgemeinerten Eigenraum $E(\lambda) = \operatorname{Kern} (A - \lambda\, Id)^{f(\lambda)}$ auf, und es gilt $f(\lambda) = 3$.

Wir fassen die am BSP. (11.7.4) exemplarisch durchgeführten Rechenschritte zu folgender **Vorschrift** für den allgemeinen Fall zusammen. Es bezeichne dabei $\{\mathrm{Hv}_j\}$ die Menge der Hauptvektoren j–ter Stufe zum festen Eigenwert $\lambda \in \mathbf{C}$ einer gegebenen Matrix A.

- **1. Schritt:** Bestimme $\{\mathrm{Hv}_1\} := \{\vec{w}_1^1, \vec{w}_1^2, \ldots, \vec{w}_1^{r_1}\}$ für $r_1 := \text{geom. dim}\, \lambda$, also die Gesamtheit der linear unabhängigen Ev zum Ew λ. Bestimme eine Basis für $\operatorname{Kern} A_\lambda^*$.

- **2. Schritt:** Bestimme in
$$\vec{b}_1 := \sum_{i=1}^{r_1} \mu_i \vec{w}_1^i$$
alle $\mu_i \in \mathbf{C}$ so, dass die Lösbarkeitsbedingung $\vec{b}_1 \perp \operatorname{Kern} A_\lambda^*$ erfüllt ist. Löse zu diesen μ_i mit dem Gauss–Algorithmus das lineare Gleichungssystem $A_\lambda \vec{w}_2 = \vec{b}_1$. Man erhält die Lösungsmenge $\{\mathrm{Hv}_2\} = \{\vec{w}_2^1, \vec{w}_2^2, \ldots, \vec{w}_2^{r_2}\}$.

- **3. Schritt:** Bestimme in
$$\vec{b}_2 := \sum_{i=1}^{r_1} \rho_i \vec{w}_1^i + \sum_{i=1}^{r_2} \mu_i \vec{w}_2^i$$
alle ρ_i, $\mu_i \in \mathbf{C}$ so, dass die Lösbarkeitsbedingung $\vec{b}_2 \perp \operatorname{Kern} A_\lambda^*$ erfüllt ist. Löse zu diesen ρ_i, μ_i mit dem Gauss–Algorithmus das lineare Gleichungssystem $A_\lambda \vec{w}_3 = \vec{b}_2$ und bestimme somit die Lösungsmenge $\{\mathrm{Hv}_3\} = \{\vec{w}_3^1, \vec{w}_3^2, \ldots, \vec{w}_3^{r_3}\}$.

Im $(j+1)$–ten Schritt muss also ein

$$\vec{b}_j := \sum_{i=1}^{r_1} \rho_i \vec{w}_1^i + \sum_{i=1}^{r_j} \mu_i \vec{w}_j^i$$

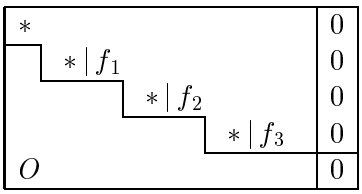
bestimmt werden, welches die Lösbarkeitsbedingung $\vec{b}_j \perp \operatorname{Kern} A_\lambda^*$ erfüllt. Danach ist die Lösungsmenge $\{\mathrm{Hv}_{j+1}\} = \{\vec{w}_{j+1}^{\,1}, \vec{w}_{j+1}^{\,2}, \ldots, \vec{w}_{j+1}^{\,r_{j+1}}\}$ des inhomogenen linearen Gleichungssystems $A_\lambda \vec{w}_{j+1} = \vec{b}_j$ zu ermitteln. Das Verfahren bricht automatisch ab, wenn $k =$ algebr. dim λ linear unabhängige Hauptvektoren bestimmt sind.

Verfahren des Kernaustauschs: Charakteristisch für die oben erklärte Verfahrensvorschrift ist das Auftreten der **Eigenvektoren** $\vec{w}_1^{\,1}, \vec{w}_1^{\,2}, \ldots, \vec{w}_1^{\,r_1}$ in **jedem** Schritt des Verfahrens. Deshalb ist es unabdingbar, auch das Verfahren des Kernaustauschs mit dem folgenden Schritt zu initialisieren:

- **1. Schritt:** Bestimme $\operatorname{Kern} A_\lambda = \operatorname{span}\{\vec{w}_1^{\,1}, \vec{w}_1^{\,2}, \ldots, \vec{w}_1^{\,r_1}\}$.

Dazu führen wir die erweiterte Matrix $(A_\lambda, \vec{0})$ durch *elementare Zeilenumformungen* in eine Staffelform über:

$$(A_\lambda, \vec{0}) \overset{\text{Gauss–Schritte}}{\Rightarrow}$$



Die Spalten S_i in den Positionen f_l dieses Staffelsystems (und somit auch die Spalten S_i der Matrix A_λ) bestimmen immer genau die gesuchten Eigenvektoren $\vec{w}_1^{\,l}$, $1 \le l \le r_1$, also den Unterraum $\operatorname{Kern} A_\lambda$. Ist dieser ein für allemal berechnet, so werden die Spalten S_i zur Lösung der linearen Gleichungssysteme $A_\lambda \vec{w}_{j+1} = \vec{b}_j$ fortan nicht mehr benötigt: Der Lösungsraum $\operatorname{Kern} A_\lambda$ des homogenen Systems liefert zu den bereits bekannten Hauptvektoren keine neuen linear unabhängigen hinzu. Das heißt, die Spalten S_i dürfen **nach dem 1. Schritt** ausgetauscht werden. Und zwar tauschen wir in A_λ die Spalte S_i in der Position $i = f_l$ gegen den Eigenvektor $\vec{w}_1^{\,l}$, $1 \le l \le r_1$ aus. (Kernaustausch!)

- **2. Schritt:** Bezeichnet $\tilde{A}_\lambda$ die durch Kernaustausch aus A_λ entstandene Matrix, so lösen wir nun mit dem GAUSS–Algorithmus das **homogene System** $\tilde{A}_\lambda \vec{v}_2 = \vec{0}$. Die Hauptvektoren $\vec{w}_2$ der Stufe 2 erhält man aus den Lösungsvektoren $\vec{v}_2$, indem man die Komponenten des Zeilenvektors $\vec{v}_2^{\,T}$ in den Positionen f_l durch 0 ersetzt.

Durch den Kernaustausch wird bewirkt, dass das zu lösende inhomogene lineare Gleichungssystem

$$A_\lambda \vec{w}_2 - \sum_{i=1}^{r_1} \mu_i \vec{w}_1^{\,i} = \vec{0}$$

die äquivalente Form $\tilde{A}_\lambda \vec{v}_2 = \vec{0}$ erhält, sofern von der bereits bekannten Lösungsmannigfaltigkeit des homogenen Systems (nämlich von $\operatorname{Kern} A_\lambda$) abgesehen wird. Die Lösungsvektoren $\vec{v}_2 = (\ldots, -\mu_1, \ldots, -\mu_{r_1}, \ldots)^T$ enthalten in den Komponentenpositionen f_i genau die Parameter μ_i derjenigen Linearkombination $\vec{b}_1 = \sum_{i=1}^{r_1} \mu_i \vec{w}_1^{\,i}$, welche die Lösbarkeitsbedingung $\vec{b}_1 \perp \operatorname{Kern} A_\lambda^*$ befriedigt. Die Bedingungen (7.5) werden hier also **automatisch** mitkontrolliert, und man braucht $\operatorname{Kern} A_\lambda^*$ überhaupt nicht zu berechnen. Da die μ_i nur von theoretischem Interesse sind, müssen ihre Zahlenwerte gar nicht mitbestimmt werden. Die spezielle Wahl von $\mu_i = 0$, $1 \le i \le r_1$ in $\vec{v}_2$ liefert die Hauptvektoren $\vec{w}_2$ der Stufe 2 mit i.a. $r_2 \ge 1$ freien Parametern.

- **3. Schritt:** Bezeichnet $\tilde{A}_\lambda$ wie vorher die durch Kernaustausch aus A_λ entstandene Matrix, so lösen wir nun mit dem GAUSS–Algorithmus das **inhomogene System** $\tilde{A}_\lambda \vec{v}_3 = \vec{w}_2 := \sum_{i=1}^{r_2} \mu_i \vec{w}_2^{\,i}$. Die Hauptvektoren $\vec{w}_3$ der Stufe 3 erhält man aus den Lösungsvektoren $\vec{v}_3$, indem man die Komponenten des Zeilenvektors $\vec{v}_3^{\,T}$ in den Positionen f_l durch 0 ersetzt.

Hier bewirkt der Kernaustausch, dass das zu lösende inhomogene lineare Gleichungssystem

$$A_\lambda \vec{w}_3 - \sum_{i=1}^{r_1} \rho_i \vec{w}_1^{\,i} = \sum_{i=1}^{r_2} \mu_i \vec{w}_2^{\,i}$$

die äquivalente Form $\tilde{A}_\lambda \vec{v}_3 = \vec{w}_2$ erhält. Der GAUSS–Algorithmus führt auf Lösbarkeitsbedingungen für die in $\vec{w}_2$ auftretenden Parameter μ_i. Für diese Parameter erhalten wir Lösungsvektoren $\vec{v}_3 = (\ldots, -\rho_1, \ldots, -\rho_{r_1}, \ldots)^T$. Werden die Parameter ρ_i in den Positionen f_i wiederum durch $\rho_i = 0$ ersetzt, so resultieren aus $\vec{v}_3$ die Hauptvektoren $\vec{w}_3$ dritter Stufe, usw.

Nach endlich vielen Schritten gelangt man zu einem unlösbaren System $\tilde{A}_\lambda \vec{v}_p = \vec{w}_{p-1}$: Das Verfahren terminiert genau mit der erforderlichen Anzahl $k = \text{algebr. dim}\,\lambda$ von Hauptvektoren. Man beachte, dass die inhomogenen Systeme $\tilde{A}_\lambda \vec{v}_p = \vec{w}_{p-1}$ ab dem 2. Schritt stets dieselbe Systemmatrix $\tilde{A}_\lambda$ besitzen. Man merke sich deshalb die erforderlichen Zeilenumformungen zur Erstellung der Stufenform. Diese müssen dann nur noch auf die wechselnden rechten Seiten $\vec{w}_{p-1}$ übertragen werden.

BSP. (11.7.5) Wir demonstrieren das Verfahren des Kernaustauschs an der Matrix $A \in \mathbf{R}^{(5,5)}$ des vorangegangenen Beispiels BSP. (11.7.4). Das heißt, wir betrachten

$$A := \begin{bmatrix} -2 & 0 & 2 & 1 & 4 \\ 4 & 0 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ -4 & 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

mit dem einzigen Eigenwert $\lambda = 0$ der algebr. $\dim \lambda = 5 =: k$. Im **1. Schritt** bestimmen wir wie vorher die zugeordneten Eigenvektoren:

$A_\lambda \vec{w}_1 = \vec{0} \;\rightleftharpoons$

−2	0	2	1	4	0
4	0	−4	−2	0	0
0	0	0	0	−4	0
−4	0	4	2	0	0
0	0	0	0	0	0

Gauss $\rightleftharpoons$

c_1	c_2	c_3			
−2	[0]	[2]	[1]	4	0
0	0	0	0	8	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

Aus der obigen Stufenform resultieren, etwa durch die spezielle Parameterwahl $(c_1, c_2, c_3) := (1,0,0)$, $=: (0,1,0)$, $:= (0,0,1)$, die drei linear unabhängigen Eigenvektoren

$$\vec{w}_1^{\,1} = (0,1,0,0,0)^T, \quad \vec{w}_1^{\,2} = (1,0,1,0,0)^T, \quad \vec{w}_1^{\,3} = (1,0,0,2,0)^T.$$

2. Schritt: Kernaustausch. In A_λ werden die Spalten S_2, S_3, S_4 (entsprechend den Spalten unter c_1, c_2, c_3) gegen die Eigenvektoren $\vec{w}_1^{\,1}$, $\vec{w}_1^{\,2}$, $\vec{w}_1^{\,3}$ ausgetauscht. Danach lösen wir unter Verwendung der elementaren Zeilenumformungen

$$\boxed{Z_2 + Z_4 \Rightarrow Z_2,} \qquad \boxed{-2Z_1 + 2Z_3 + Z_4 \Rightarrow Z_4}$$

das lineare Gleichungssystem

$\tilde{A}_\lambda \vec{v}_2 = \vec{0} \;\rightleftharpoons$

c_1	c_2	c_3			
−2	0	1	1	4	0
4	1	0	0	0	0
0	0	1	0	−4	0
−4	0	0	2	0	0
0	0	0	0	0	0

Gauss $\rightleftharpoons$

c_1	c_2	c_3				$\vec{w}_2^{\,1}$	$\vec{w}_3^{\,1}$
−2	0	1	1	4	0	1	0
0	1	0	2	0	0	0	0
0	0	1	[0]	−4	0	0	0
0	0	0	0	−16	0	−2	0
0	0	0	0	0	0	0	[$\frac{1}{8}$]

W!

Aus Gründen der Bezeichnungskonsistenz haben wir hier $c_i := -\mu_i$ gesetzt. Wir ignorieren zunächst die beiden Spalten nach dem Doppelstrich.

Offenbar ist der Parameter c_3 frei wählbar, und es resultiert

$$\vec{v}_2 = \Big(\frac{c_3}{2},\, -2c_3,\, 0,\, c_3,\, 0\Big)^T.$$

Das heißt, die Linearkombination $\vec{b}_1 = \sum_{i=1}^{3} \mu_i \vec{w}_1^{\,i}$ mit den Parametern $\mu_1 := 2c_3,\ \mu_2 := 0,\ \mu_3 := -c_3$ erfüllt die Lösbarkeitbedingung $\vec{b}_1 \perp \operatorname{Kern} A_\lambda^*$. Vorschriftsgemäß ersetzen wir die Komponenten von $\vec{v}_2^{\,T}$ in den Spalten $S_2,\ S_3,\ S_4$ durch Null, und wir erhalten den einzigen Hauptvektor 2. Stufe, etwa durch Wahl von $c_3 := 2$

$$\vec{w}_2^{\,1} = (1,0,0,0,0)^T.$$

Die Wahl von $c_3 := -1$ führt wieder auf das Ergebnis aus BSP. (11.7.4).

3. Schritt: Wir führen in $\vec{w}_2^{\,1}$ die oben angegebenen Zeilenumformungen durch und tragen das Ergebnis in das obige Staffelsystem unter der Spalte "$\vec{w}_2^{\,1}$" ein. Das System $\tilde{A}_\lambda \vec{v}_3 = \vec{w}_2^{\,1}$ liegt nun schon in Stufenform vor und besitzt bei freier Wahl von c_3 die Lösung

$$\vec{v}_3 = \Big(\frac{c_3}{2},\, -2c_3,\, \frac{1}{2},\, c_3,\, \frac{1}{8}\Big)^T.$$

Vorschriftsgemäß ersetzen wir die Komponenten von $\vec{v}_3^{\,T}$ in den Spalten $S_2,\ S_3,\ S_4$ wieder durch Null, und wir erhalten den einzigen Hauptvektor 3. Stufe, etwa durch Wahl von $c_3 := 0$

$$\vec{w}_3^{\,1} = \Big(0,0,0,0,\frac{1}{8}\Big)^T.$$

Man beachte, dass der Hauptvektor $\vec{w}_2^{\,1} := (-2,0,0,0,0)^T$ mit der gleichen Rechnung auf das in BSP. (11.7.4) gefundene Resultat $\vec{w}_3^{\,1} = (0,0,0,0,-\frac{1}{4})^T$ führt.

4. Schritt: Wir führen in $\vec{w}_3^{\,1}$ die oben angegebenen Zeilenumformungen durch und tragen das Ergebnis in das Staffelsystem unter der Spalte "$\vec{w}_3^{\,1}$" ein. Man erkennt sofort die Unlösbarkeit des resultierenden Systems: Das Verfahren terminiert mit dem Ergebnis

$$\{\mathrm{Hv}_2\} = \{\vec{w}_2^{\,1}\}, \quad \{\mathrm{Hv}_3\} = \{\vec{w}_3^{\,1}\}, \quad E(\lambda) = \operatorname{Kern}(A - \lambda\, Id)^3 = \operatorname{span}\{\vec{w}_1^{\,1}, \vec{w}_1^{\,2}, \vec{w}_1^{\,3}, \vec{w}_2^{\,1}, \vec{w}_3^{\,1}\}.$$

11.7.2 Anfangswertaufgaben für Systeme mit konstanten Koeffizienten

Wir greifen hier nochmals das Anfangswertproblem (2.5) aus Abschnitt 11.2 auf. Dort hatten wir Systeme von Differentialgleichungen 1. Ordnung mit konstanten Koeffizienten diskutiert. In Satz 11.6 konnten wir zeigen, dass die Anfangswertaufgabe

$$\boxed{\vec{y}\,'(t) = A\vec{y}(t), \qquad \vec{y}(t_0) = \vec{y}_0 \in \mathbf{K}^n} \tag{7.7}$$

für eine gegebene Matrix $A \in \mathbf{K}^{(n,n)}$ und für jeden Anfangsvektor $\vec{y}_0 \in \mathbf{K}^n$ die eindeutig bestimmte Lösung $\vec{y}(t) = e^{(t-t_0)A}\vec{y}_0$ besitzt. Falls der Vektorraum $\mathbf{C}^n$ von einer Basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ aus **Eigenvektoren** der Matrix A aufgespannt wird, so kann die Berechnung der Exponentialmatrix $e^{(t-t_0)A}$ in sehr einfacher Weise durch die Eigenvektoren $\vec{v}_j$ und die zugehörigen Eigenwerte λ_j erfolgen. Wir hatten in Satz 11.7 die folgende Darstellung der Lösung $\vec{y}(t)$ hergeleitet:

$$\boxed{\vec{y}(t) = C_1 e^{\lambda_1(t-t_0)}\vec{v}_1 + C_2 e^{\lambda_2(t-t_0)}\vec{v}_2 + \cdots + C_n e^{\lambda_n(t-t_0)}\vec{v}_n.} \tag{7.8}$$

Darin sind die Konstanten C_j die Koeffizienten der Linearkombination des Anfangsvektors:

$$\vec{y}_0 = C_1\vec{v}_1 + C_2\vec{v}_2 + \cdots + C_n\vec{v}_n.$$

Wir interpretieren das hier Gesagte stets so, dass auch im reellen Fall $\mathbf{K} := \mathbf{R}$ eine reelle Lösung der Anfangswertaufgabe (7.7) in der Form (7.8) gefunden werden kann, selbst wenn die Matrix $A \in \mathbf{R}^{(n,n)}$ komplexe Eigenwerte und somit auch komplexe Eigenvektoren besitzt. Wie wir bereits in Bemerkung 11.9 festgestellt haben, ist mit $\lambda_j := \alpha_j + i\beta_j$, $\alpha_j, \beta_j \in \mathbf{R}$, auch die konjugiert komplexe Zahl $\overline{\lambda_j} = \alpha_j - i\beta_j$ ein Eigenwert der Matrix A. Ist dem Eigenwert λ_j der Eigenvektor $\vec{v}_j := \vec{u}_j + i\vec{w}_j$, $\vec{u}_j, \vec{w}_j \in \mathbf{R}^n$, zugeordnet, so gehört zum Eigenwert $\overline{\lambda_j}$ der konjugierte Eigenvektor $\overline{(\vec{v}_j)} = \vec{u}_j - i\vec{w}_j$. Demgemäß gilt für jedes Zahlenpaar C, D:

$$\begin{aligned} Ce^{\lambda_j(t-t_0)}\vec{v}_j \;&+\; De^{\overline{\lambda_j}(t-t_0)}\overline{(\vec{v}_j)} \\ &= e^{\alpha_j(t-t_0)}\cos\beta_j(t-t_0)\,(\tilde{C}\vec{u}_j + \tilde{D}\vec{w}_j) + e^{\alpha_j(t-t_0)}\sin\beta_j(t-t_0)\,(\tilde{D}\vec{u}_j - \tilde{C}\vec{w}_j), \end{aligned}$$

worin $\tilde{C} := C + D$ und $\tilde{D} := i(C - D)$ zu setzen sind. Nun sind die *reellen* Zahlen $\tilde{C}, \tilde{D}$ die Koeffizienten der Linearkombination des (reellen) Anfangsvektors $\vec{y}_0$ in Richtung der beiden reellen Basisvektoren $\vec{u}_j, \vec{w}_j$.

Besitzt die Matrix $A \in \mathbf{K}^{(n,n)}$ **keine** n linear unabhängigen Eigenvektoren, so kann die Lösung $\vec{y}(t)$ der Anfangswertaufgabe (7.7) nicht mehr in der Form (7.8) dargestellt werden. Wir zeigen im folgenden eine modifizierte Darstellung unter Heranziehung der Hauptvektoren von A. Es sei also $\lambda \in \mathbf{C}$ ein Eigenwert der Vielfachheit $k \geq 2$, und es sei $\vec{w}_j \in \mathbf{C}^n$ ein Hauptvektor der Stufe $j \geq 1$ zum Eigenwert λ. Sicher gilt $j \leq k$ und somit

$$\boxed{\begin{aligned} e^{tA}\vec{w}_j &= \sum_{i=0}^{\infty}\frac{t^i}{i!}(A - \lambda\,Id + \lambda\,Id)^i\vec{w}_j = \sum_{i=0}^{\infty}\frac{t^i}{i!}\sum_{r=0}^{i}\binom{i}{r}\lambda^{i-r}(A - \lambda\,Id)^r\vec{w}_j \\ &= \sum_{r=0}^{\infty}\frac{t^r}{r!}\,(A - \lambda\,Id)^r\vec{w}_j\sum_{i=r}^{\infty}\frac{t^{i-r}\lambda^{i-r}}{(i-r)!} \\ &= e^{\lambda t}\sum_{r=0}^{j-1}\frac{t^r}{r!}\,(A - \lambda\,Id)^r\vec{w}_j. \end{aligned}} \tag{7.9}$$

Gemäß (7.6) vereinfacht sich die Darstellung (7.9) ganz erheblich, wenn wir wieder den

$$\boxed{\textbf{Sonderfall: } 1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda =: k}$$

betrachten. In diesem Fall gilt nämlich wegen (7.6) die Relation $(A - \lambda\,Id)^r\vec{w}_j = \vec{w}_{j-r}$, und aus (7.9) folgt:

Satz 11.27 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Sei $\lambda \in \mathbf{C}$ ein Eigenwert von A mit $1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda =: k$. Sei $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ die gemäß (7.6) bestimmte Kette von Hauptvektoren $\vec{w}_j \in \mathbf{C}^n$ der Stufen $j = 1, 2, \ldots, k$ zum Eigenwert λ. Sind die Zahlen $C_j \in \mathbf{C}$ die Koeffizienten des Anfangsvektors $\vec{y}_0 \in \mathbf{K}^n$ in der Teilbasis $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ des Vektorraumes $\mathbf{C}^n$, so ist dem Eigenwert λ der folgende Lösungsanteil der Anfangswertaufgabe (7.7) eindeutig zugeordnet:*

$$\boxed{\vec{y}_\lambda(t) = e^{\lambda(t-t_0)}\sum_{j=1}^{k}C_j\Big(\sum_{r=0}^{j-1}\frac{(t-t_0)^r}{r!}\,\vec{w}_{j-r}\Big).} \tag{7.10}$$

Begründung: Es genügt, den Fall $t_0 = 0$ zu betrachten. Klar, die Hauptvektoren $\vec{w}_j$ spannen gemäß Satz 11.26 den k–dimensionalen Unterraum $E(\lambda) \subseteq \mathbf{C}^n$ auf. Die in $E(\lambda)$ liegende Komponente $\vec{y}_{0\lambda}$ des Anfangsvektors $\vec{y}_0$ besitzt eine eindeutige Linearkombination $\vec{y}_{0\lambda} = C_1\vec{w}_1 + C_2\vec{w}_2 + \cdots + C_k\vec{w}_k$. Dieser Komponenten ist der Lösungsanteil $\vec{y}_\lambda(t) = e^{tA}\vec{y}_{0\lambda}$ zugeordnet. Somit folgt aus (7.9)

$$\vec{y}_\lambda(t) = \sum_{j=1}^{k} C_j e^{tA}\vec{w}_j = e^{\lambda t}\sum_{j=1}^{k} C_j\Big(\sum_{r=0}^{j-1}\frac{t^r}{r!}\,\vec{w}_{j-r}\Big),$$

und dies ist die behauptete Darstellung (7.10). □

Sind nun $\lambda_1, \lambda_2, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$ mit Vielfachheiten $k_1, k_2, \ldots, k_m$, und ist jeweils die Bedingung

$$\boxed{\text{geom. dim}\,\lambda_j = 1, \quad j = 1, 2, \ldots, m,}$$

erfüllt, so erhält man durch *Superposition* der gemäß Satz 11.27 bestimmten Teillösungen $\vec{y}_{\lambda_j}$ die eindeutig bestimmte Lösung der Anfangswertaufgabe (7.7) in der Form

$$\boxed{\vec{y}(t) = \vec{y}_{\lambda_1}(t) + \vec{y}_{\lambda_2}(t) + \cdots + \vec{y}_{\lambda_m}(t).} \tag{7.11}$$

Bemerkung 11.23 Lässt man die Konstanten $C_j \in \mathbf{C}$ in (7.10) unbestimmt und setzt man $t_0 = 0$, so hat man mit (7.11) die **allgemeine Lösung** des Systems $\vec{y}\,'(t) = A\vec{y}(t)$ linearer Differentialgleichungen 1.Ordnung vorliegen. □

BSP. (11.7.6) Es sei $A \in \mathbf{R}^{(4,4)}$ die Matrix aus BSP. (11.7.3) mit dem vierfachen Eigenwert $\lambda = 3$ der geom. dim $\lambda = 1$ und den zugeordneten Hauptvektoren $\vec{w}_j$ der Stufen $1-4$:

$$\vec{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_3 = \begin{bmatrix} 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{w}_4 = \begin{bmatrix} 0 \\ 8 \\ 1 \\ -2 \end{bmatrix}.$$

Dann hat das lineare DGl–System

$$\vec{y}\,'(t) = A\vec{y}(t)$$

gemäß Satz 11.27 die allgemeine Lösung

$$\boxed{\vec{y}(t) = e^{3t}\Big(C_1\vec{w}_1 + C_2(\vec{w}_2 + t\vec{w}_1) + C_3(\vec{w}_3 + t\vec{w}_2 + \tfrac{t^2}{2!}\vec{w}_1) + C_4(\vec{w}_4 + t\vec{w}_3 + \tfrac{t^2}{2!}\vec{w}_2 + \tfrac{t^3}{3!}\vec{w}_1)\Big).} \tag{7.12}$$

Die Anfangswertaufgabe (7.7) zum Anfangsvektor $\vec{y}_0 = (1, 6, 1, -1)^T$ wird durch (7.12) gelöst, wenn wir überall t durch $t - t_0$ ersetzen und wenn die Konstanten C_j das folgende inhomogene lineare Gleichungssystem lösen:

$$C_1\vec{w}_1 + C_2\vec{w}_2 + C_3\vec{w}_3 + C_4\vec{w}_4 = \vec{y}_0.$$

Dieses hat die explizite Form

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -3 & 8 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 1 \\ -1 \end{bmatrix}$$

mit der eindeutig bestimmten Lösung $C_1 = C_2 = C_3 = C_4 = 1$. Setzen wir zur Abkürzung $p_j(t) := e^{3(t-t_0)}(1 + \frac{t-t_0}{1!} + \frac{(t-t_0)^2}{2!} + \cdots + \frac{(t-t_0)^j}{j!})$, so erhalten wir die gesuchte Lösung der Anfangswertaufgabe (7.7) zum obigen Anfangsvektor $\vec{y}_0$ in der Form

$$\boxed{\vec{y}(t) = p_3(t)\vec{w}_1 + p_2(t)\vec{w}_2 + p_1(t)\vec{w}_3 + p_0(t)\vec{w}_4.}$$

BSP. (11.7.7) Es ist die Anfangswertaufgabe (7.7) für den Anfangsvektor $\vec{y}_0 := (4, 1, 2, 1)^T$ und die folgende Matrix $A \in \mathbf{R}^{(4,4)}$ zu lösen, deren charakteristisches Polynom $P_4(\lambda) = \det(A - \lambda\, Id) = (2-\lambda)^2(1-\lambda)^2$ ist:

$$A := \begin{bmatrix} 2 & 0.5 & -1 & -2.5 \\ 0 & 2 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad P_4(\lambda) = \begin{vmatrix} 2-\lambda & 0.5 & -1 & -2.5 \\ 0 & 2-\lambda & -2 & -1 \\ 0 & 0 & 1-\lambda & 0 \\ 0 & 0 & 1 & 1-\lambda \end{vmatrix}.$$

Es liegen je ein doppelter Eigenwert $\lambda_1 = 2$ und $\lambda_2 = 1$ vor. Wir bestimmen zunächst die Eigenvektoren $\vec{w}_1$ zum Eigenwert $\lambda_1 = 2$:

$$A_{\lambda_1}\vec{w}_1 = \vec{0} \Leftrightarrow \begin{bmatrix} 0 & 0.5 & -1 & -2.5 \\ 0 & 0 & -2 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{w}_1 = \begin{bmatrix} c \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad c \in \mathbf{C}.$$

Hier liegt offensichtlich der Sonderfall $1 = \text{geom. dim}\,\lambda_1 < \text{algebr. dim}\,\lambda_1 = 2$ vor. Wir wählen nun $c = 1$ und bestimmen den Hauptvektor $\vec{w}_2$ der Stufe 2 gemäß

$$A_{\lambda_1}\vec{w}_2 = \vec{w}_1 \Leftrightarrow \begin{bmatrix} 0 & 0.5 & -1 & -2.5 \\ 0 & 0 & -2 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}.$$

Wir bestimmen des weiteren die Eigenvektoren $\vec{v}_1$ zum Eigenwert $\lambda_2 = 1$:

$$A_{\lambda_2}\vec{v}_1 = \vec{0} \Leftrightarrow \begin{bmatrix} 1 & 0.5 & -1 & -2.5 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{v}_1 = c \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad c \in \mathbf{C}.$$

Auch hier liegt der Sonderfall $1 = \text{geom. dim}\,\lambda_2 < \text{algebr. dim}\,\lambda_2 = 2$ vor. Wir wählen wiederum $c = 1$ und bestimmen den Hauptvektor $\vec{v}_2$ der Stufe 2 gemäß

$$A_{\lambda_2}\vec{v}_2 = \vec{v}_1 \Leftrightarrow \begin{bmatrix} 1 & 0.5 & -1 & -2.5 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \text{also} \quad \vec{v}_2 = \begin{bmatrix} 1.5 \\ 3 \\ 1 \\ 0 \end{bmatrix}.$$

Hiermit haben wir in $\mathbf{C}^4$ eine Basis $\{\vec{w}_1, \vec{w}_2, \vec{v}_1, \vec{v}_2\}$ aus Hauptvektoren der Matrix A bestimmt. Das lineare DGl–System

$$\vec{y}\,'(t) = A\vec{y}(t)$$

hat folglich gemäß Satz 11.27 die allgemeine Lösung

$$\boxed{\vec{y}(t) = e^{2t}\Big(C_1\vec{w}_1 + C_2(\vec{w}_2 + t\vec{w}_1)\Big) + e^{t}\Big(C_3\vec{v}_1 + C_4(\vec{v}_2 + t\vec{v}_1)\Big).} \tag{7.13}$$

Die Anfangswertaufgabe (7.7) zum Anfangsvektor $\vec{y}_0 = (4,1,2,1)^T$ wird durch (7.13) gelöst, wenn wir überall t durch $t - t_0$ ersetzen und wenn die Konstanten C_j das folgende inhomogene lineare Gleichungssystem lösen:

$$C_1\vec{w}_1 + C_2\vec{w}_2 + C_3\vec{v}_1 + C_4\vec{v}_2 = \vec{y}_0.$$

Dieses hat die explizite Form

$$\begin{bmatrix} 1 & 0 & 2 & 1.5 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

mit der eindeutig bestimmten Lösung $C_1 = -1$, $C_2 = -3$, $C_3 = 1$, $C_4 = 2$. Hiermit erhalten wir die gesuchte Lösung der Anfangswertaufgabe (7.7) zum obigen Anfangsvektor $\vec{y}_0$ in der Form

$$\boxed{\vec{y}(t) = e^{t-t_0}\Big(5 + 4(t-t_0),\ 7 + 2(t-t_0),\ 2,\ 1 + 2(t-t_0)\Big)^T - e^{2(t-t_0)}\Big(1 + 3(t-t_0),\ 6,\ 0,\ 0\Big)^T.}$$

Der **allgemeine Fall** $2 \leq \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda$ lässt sich nicht mehr so einfach behandeln wie der oben diskutierte Sonderfall. Wir werden jedoch in Abschnitt 11.8.2 zeigen, dass die Kenntnis speziell strukturierter Ketten von Hauptvektoren zum Eigenwert $\lambda \in \mathbf{C}$ einer Matrix $A \in \mathbf{K}^{(n,n)}$ zu einer erheblichen Vereinfachung der Darstellung von Lösungen zur Anfangswertaufgabe (7.7) führen kann. Man erhält aber auch in denjenigen Fällen aus der Beziehung (7.9) eine – wenn auch kompliziertere – Darstellung, in denen nur irgendeine (unstrukturierte) Basis des verallgemeinerten Eigenraumes $E(\lambda)$ bekannt ist:

Satz 11.28 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Sei $\lambda \in \mathbf{C}$ ein Eigenwert von A mit der Vielfachheit $k = \text{algebr. dim}\,\lambda$. Sei $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ das System von Hauptvektoren zum Eigenwert λ, das den verallgemeinerten Eigenraum $E(\lambda) := \text{Kern}\,A_\lambda^{f(\lambda)}$ aufspannt. Dann ist dem Eigenwert λ der folgende Lösungsanteil der Anfangswertaufgabe (7.7) eindeutig zugeordnet:*

$$\boxed{\vec{y}_\lambda(t) = e^{\lambda(t-t_0)} \sum_{j=1}^{k} C_j \Big(\sum_{r=0}^{k-1} \frac{(t-t_0)^r}{r!} A_\lambda^r \vec{w}_j \Big) = e^{\lambda(t-t_0)} \sum_{r=0}^{k-1} \frac{(t-t_0)^r}{r!} A_\lambda^r \Big(\sum_{j=1}^{k} C_j \vec{w}_j \Big).} \qquad (7.14)$$

Darin sind die Zahlen $C_j \in \mathbf{C}$ die Koeffizienten des Anfangsvektors $\vec{y}_0 \in \mathbf{K}^n$ in der Teilbasis $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ des Vektorraumes $\mathbf{C}^n$.

Bemerkung 11.24 Wie vorher gilt das **Superpositionsprinzip**: Sind $\lambda_1, \lambda_2, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$ mit den Vielfachheiten $k_1, k_2, \ldots, k_m$, so erhält man durch Superposition der gemäß Satz 11.28 bestimmten Teillösungen $\vec{y}_{\lambda_j}$ die eindeutig bestimmte Lösung der Anfangswertaufgabe (7.7) in der Form

$$\boxed{\vec{y}(t) = \vec{y}_{\lambda_1}(t) + \vec{y}_{\lambda_2}(t) + \cdots + \vec{y}_{\lambda_m}(t).} \qquad (7.15)$$

Bleiben die Konstanten $C_j \in \mathbf{C}$ in (7.14) unbestimmt und setzt man $t_0 = 0$, so liegt mit (7.15) die **allgemeine Lösung** des DGl–Systems $\vec{y}\,'(t) = A\vec{y}(t)$ vor. □

BSP. (11.7.8) Wir betrachten die folgende Matrix $A \in \mathbf{R}^{(4,4)}$, deren charakteristisches Polynom $P_4(\lambda) = \det(A - \lambda\,Id) = (1-\lambda)^3(4-\lambda)$ ist:

$$A := \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad P_4(\lambda) = \begin{vmatrix} 1-\lambda & 0 & 3 & 0 \\ 0 & 1-\lambda & 2 & 0 \\ 0 & 0 & 1-\lambda & 0 \\ 0 & 0 & 0 & 4-\lambda \end{vmatrix}.$$

Es liegen ein dreifacher Eigenwert $\lambda_1 = 1$ und ein einfacher Eigenwert $\lambda_2 = 4$ vor. Wir bestimmen zunächst die Eigenvektoren zum Eigenwert $\lambda_1 = 1$:

$$A_{\lambda_1}\vec{w}^{\,1} = \vec{0} \quad \rightleftharpoons \quad \left[\begin{array}{cccc|c} 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \end{array}\right] \overset{\text{Gauss}}{\rightleftharpoons} \left[\begin{array}{cccc|c} c_1 & c_2 & & & \\ \hline \boxed{0} & \boxed{0} & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Es resultieren die zwei linear unabhängigen Eigenvektoren

$$\vec{w}_1^{\,1} := (1,0,0,0)^T, \quad \vec{w}_2^{\,1} := (0,1,0,0)^T,$$

so dass gilt: $2 = \text{geom. dim}\,\lambda_1 < \text{algebr. dim}\,\lambda_1 = 3$. Demzufolge muss ein weiterer Hauptvektor $\vec{w}_3^{\,1}$ der Stufe 2 zum Eigenwert λ_1 be stimmt werden, den wir mit dem Verfahren des Kernaustauschs berechnen. Dazu tauschen wir in A_{λ_1} die Spalten S_1 und S_2 (entsprechend den Spalten unter c_1, c_2) gegen die Eigenvektoren $\vec{w}_1^{\,1}$ und $\vec{w}_2^{\,1}$ aus und lösen danach das folgende lineare Gleichungssystem:

$$\tilde{A}_{\lambda_1}\vec{v} = \vec{0} \quad \rightleftharpoons \quad \left[\begin{array}{cccc|c} c_1 & c_2 & & & \\ \hline 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \end{array}\right] \overset{\text{Gauss}}{\rightleftharpoons} \left[\begin{array}{cccc|c} c_1 & c_2 & & & \\ \hline 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & \boxed{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Dieses Stufensystem wird gelöst durch den Vektor $\vec{v} = (-3,\,-2,\,1,\,0)^T$. Vorschriftsgemäß haben wir die Komponenten des Zeilenvektors $\vec{v}^{\,T}$ in den Spalten S_1 und S_2 durch Null zu ersetzen. Es resultiert der einzige Hauptvektor 2.Stufe zum Eigenwert $\lambda_1 = 1$:

$$\vec{w}_3^{\,1} = (0,0,1,0)^T.$$

Die Vektoren $\vec{w}_1^{\,1}, \vec{w}_2^{\,1}, \vec{w}_3^{\,1}$ spannen nun den verallgemeinerten Eigenraum zum Eigenwert $\lambda_1 = 1$ auf, und es gilt $f(\lambda_1) = 2$. Schließlich bestimmen wir noch den Eigenvektor $\vec{w}^{\,2}$ zum Eigenwert $\lambda_2 = 4$:

$$A_{\lambda_2}\vec{w}^{\,2} = \vec{0} \quad \rightleftharpoons \quad \left[\begin{array}{cccc|c} -3 & 0 & 3 & 0 & 0 \\ \hline 0 & -3 & 2 & 0 & 0 \\ 0 & 0 & -3 & \boxed{0} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \text{also} \quad \vec{w}^{\,2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ c \end{bmatrix}, \quad c \in \mathbf{C}.$$

Mit der Wahl $c = 1$ erhalten wir $\vec{w}_1^{\,2} = (0,0,0,1)^T$, so dass letztendlich die Vektoren $\vec{w}_1^{\,1}, \vec{w}_2^{\,1}, \vec{w}_3^{\,1}, \vec{w}_1^{\,2}$ mit der Standardbasis des $\mathbf{C}^4$ zusammenfallen. Gemäß (7.14) berechnen wir

$$A_{\lambda_1}\vec{w}_1^{\,1} = \vec{0} = A_{\lambda_1}\vec{w}_2^{\,1} = A_{\lambda_2}\vec{w}_1^{\,2}, \quad A_{\lambda_1}\vec{w}_3^{\,1} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \end{bmatrix} := \vec{b}_1,$$

und erhalten daraus die allgemeine Lösung des DGl–Systems $\vec{y}\,'(t) = A\vec{y}(t)$ in der Form

$$\boxed{\vec{y}(t) = e^t\Big(C_1\vec{w}_1^{\,1} + C_2\vec{w}_2^{\,1} + C_3(\vec{w}_3^{\,1} + t\vec{b}_1)\Big) + e^{4t}C_4\vec{w}_1^{\,2}.}$$

Zu lösen sei nun die Anfangswertaufgabe (7.7) für den Anfangsvektor $\vec{y}_0 := (1,2,3,4)$. Das lineare Gleichungssystem $\vec{y}_0 = C_1\vec{w}_1^{\,1} + C_2\vec{w}_2^{\,1} + C_3\vec{w}_3^{\,1} + C_4\vec{w}_1^{\,2}$ hat offensichtlich die eindeutige Lösung

$C_1 = 1$, $C_2 = 2$, $C_3 = 3$ sowie $C_4 = 4$. Also resultiert aus (7.16) die Lösung der Anfangswertaufgabe (7.7) zum obigen Anfangsvektor in der Form

$$\boxed{\begin{aligned} \vec{y}(t) &= e^{t-t_0}\Big(1+9(t-t_0), 2+6(t-t_0), 3, 0\Big)^T + e^{4(t-t_0)}(0,0,0,4)^T \\ &= \begin{bmatrix} e^{t-t_0} & 0 & 3(t-t_0)e^{t-t_0} & 0 \\ 0 & e^{t-t_0} & 2(t-t_0)e^{t-t_0} & 0 \\ 0 & 0 & e^{t-t_0} & 0 \\ 0 & 0 & 0 & e^{4(t-t_0)} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} =: Y(t-t_0)\vec{y}_0. \end{aligned}} \tag{7.16}$$

Bemerkung 11.25 Mit der in (7.16) angegebenen Matrixfunktion $Y(t)$ verifiziert man durch direkte Rechnung für das obige BSP. (11.7.8) die Beziehung

$$\boxed{Y'(t) = AY(t), \quad t \in \mathbf{R}, \qquad Y(0) = Id.} \tag{7.17}$$

Das heißt, die Vektorfunktion $\vec{y}(t) := Y(t-t_0)\vec{y}_0$ erfüllt $\vec{y}\,'(t) = A\vec{y}(t)$ und $\vec{y}(t_0) = \vec{y}_0$; sie löst somit die Anfangswertaufgabe (7.7) □

Definition 11.12 *Die so bestimmte Matrixfunktion $Y(t)$ heißt die* **Fundamentalmatrix** *des DGl-Systems $\vec{y}\,'(t) = A\vec{y}(t)$.*

BSP. (11.7.9) Gesucht ist die Lösung der Anfangswertaufgabe

$$\vec{y}\,'(t) = A\vec{y}(t) := \begin{bmatrix} -2 & 0 & 2 & 1 & 4 \\ 4 & 0 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ -4 & 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \vec{y}(t), \quad \vec{y}(0) = \vec{y}_0 := \begin{bmatrix} 9 \\ 1 \\ 2 \\ 6 \\ 5 \end{bmatrix}.$$

Die Systemmatrix $A \in \mathbf{R}^{(5,5)}$, die wir bereits in BSP. (11.7.5) behandelt haben, besitzt den Eigenwert $\lambda = 0$ der Vielfachheit $k = 5 =$ algebr. dim $\lambda >$ geom. dim $\lambda = 3$. Wie in BSP. (11.7.5) gezeigt wurde, spannen die folgenden fünf Hauptvektoren den verallgemeinerten Eigenraum $E(\lambda) = \operatorname{Kern} A_\lambda^3 = \mathbf{C}^5$ auf:

$$\vec{w}_1^{\,1} := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_1^{\,2} := \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_1^{\,3} := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \quad \vec{w}_2^{\,1} := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_3^{\,1} := \frac{1}{8}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Aus der Tatsache, dass die Vektoren $\vec{w}_j^{\,i}$ Hauptvektoren der Stufen $j = 1, 2, 3$ sind, erhält man sofort

$$\vec{0} = A_\lambda \vec{w}_1^{\,1} = A_\lambda \vec{w}_1^{\,2} = A_\lambda \vec{w}_1^{\,3} = A_\lambda^2 \vec{w}_2^{\,1} = A_\lambda^3 \vec{w}_3^{\,1}.$$

Darüber hinaus findet man mit einfacher Rechnung

$$A_\lambda \vec{w}_2^{\,1} = \begin{bmatrix} -2 \\ 4 \\ 0 \\ -4 \\ 0 \end{bmatrix} =: \vec{b}_1, \quad A_\lambda \vec{w}_3^{\,1} = \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} =: \vec{b}_2, \quad A_\lambda^2 \vec{w}_3^{\,1} = \begin{bmatrix} -2 \\ 4 \\ 0 \\ -4 \\ 0 \end{bmatrix} =: \vec{b}_3.$$

Die Lösungsdarstellung (7.14) führt somit auf die **allgemeine Lösung**

$$\boxed{\vec{y}(t) = e^{0t}\Big\{C_1\vec{w}_1^{\,1} + C_2\vec{w}_1^{\,2} + C_3\vec{w}_1^{\,3} + C_4(\vec{w}_2^{\,1} + t\vec{b}_1) + C_5\Big(\vec{w}_3^{\,1} + t\vec{b}_2 + \frac{t^2}{2!}\,\vec{b}_3\Big)\Big\}.}$$

Zur Lösung der Anfangswertaufgabe $\vec{y}(0) = \vec{y}_0 = (9,1,2,6,5)^T$ berechnen wir die Koeffizienten C_j aus dem linearen Gleichungssystem

$$\vec{y}(0) = C_1\vec{w}_1^{\,1}+C_2\vec{w}_1^{\,2}+C_3\vec{w}_1^{\,3}+C_4\vec{w}_2^{\,1}+C_5\vec{w}_3^{\,1} \stackrel{!}{=} \vec{y}_0 \quad \rightleftharpoons \quad \left(\begin{array}{ccccc|c} 0 & 1 & 1 & 1 & 0 & 9 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & \frac{1}{8} & 5 \end{array}\right) \quad \rightleftharpoons \quad \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 & \frac{1}{8} & 5 \end{array}\right).$$

Man ersieht hier unschwer das Lösungstupel $(C_1, C_2, C_3, C_4, C_5)^T = (1,2,3,4,40)^T$, und demgemäß resultiert als eindeutig bestimmte Lösung der Anfangswertaufgabe:

$$\boxed{\vec{y}(t) = \Big((9+12t-40t^2),\ (1+16t+80t^2),\ (2-20t),\ (6-16t-80t^2),\ 5\Big)^T.}$$

Beachte: Die allgemeine Lösung kann auch in einer Matrixform $\vec{y}(t) = \tilde{Y}(t)\vec{C}$ geschrieben werden mit

$$\tilde{Y}(t) = \Big(\vec{w}_1^{\,1},\ \vec{w}_1^{\,2},\ \vec{w}_1^{\,3},\ (\vec{w}_2^{\,1}+t\vec{b}_1),\ (\vec{w}_3^{\,1}+t\vec{b}_2+\tfrac{t^2}{2!}\vec{b}_3)\Big) \quad \text{und} \quad \vec{C} := (C_1,C_2,C_3,C_4,C_5)^T.$$

Da ein linearer Zusammenhang $\vec{C} = B\vec{y}_0$ zwischen dem Anfangsvektor $\vec{y}_0$ und dem Koeffizientenvektor $\vec{C}$ besteht, besitzt die Anfangswertaufgabe die Lösungsdarstellung $\vec{y}(t) = \tilde{Y}(t)B\vec{y}_0$. Wir erhalten hieraus $\tilde{Y}(0)B = Id$, also $B = \tilde{Y}^{-1}(0)$. Die Matrixfunktion

$$Y(t) := \tilde{Y}(t)\tilde{Y}^{-1}(0)$$

ist genau die **Fundamentalmatrix** des DGl–Systems $\vec{y}\,'(t) = A\vec{y}(t)$.

Die Lösung der Anfangswertaufgabe (7.7) gestaltet sich nach der Vorschrift des Satzes 11.28 relativ kompliziert. Um einen anderen Lösungszugang aufzuzeigen, greifen wir nochmals das obige Beispiel (11.7.8) auf, in welchem wir den speziellen Anfangspunkt $t_0 = 0$ betrachten.

BSP. (11.7.10) Zu lösen ist die Anfangswertaufgabe

$$\vec{y}\,'(t) = A\vec{y}(t) = \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}\vec{y}(t), \qquad \vec{y}(0) = \vec{y}_0 := \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}. \tag{7.18}$$

Da die Matrix A eine **Rechtsdreiecksmatrix** ist, kann das DGl–System (7.18), beginnend mit der letzten Zeile, **sukzessive von unten nach oben** gelöst werden. In expliziter Schreibweise lautet (7.18):

$$\begin{bmatrix} y_1'(t) \\ y_2'(t) \\ y_3'(t) \\ y_4'(t) \end{bmatrix} = \begin{bmatrix} y_1(t)+3y_3(t) \\ y_2(t)+2y_3(t) \\ y_3(t) \\ 4y_4(t) \end{bmatrix}, \qquad \begin{bmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \\ y_4(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Wir lösen gemäß der Strategie *4.Zeile* $\Rightarrow$ *3.Zeile* $\Rightarrow$ *2.Zeile* $\Rightarrow$ *1.Zeile.*

4.Zeile: Die DGl $y_4' - 4y_4 = 0$ hat die allgemeine Lösung $y_4(t) = C_4e^{4t}$ mit dem Anfangswert $y_4(0) = C_4 \stackrel{!}{=} 4$. Somit erhalten wir als Lösungskomponente der Anfangswertaufgabe (7.18) die Funktion

$$\boxed{y_4(t) = 4e^{4t}.}$$

3.Zeile: Die DGl $y_3' - y_3 = 0$ hat die allgemeine Lösung $y_3(t) = C_3 e^t$ mit dem Anfangswert $y_3(0) = C_3 \stackrel{!}{=} 3$. Hieraus resultiert als Lösungskomponente der Anfangswertaufgabe (7.18) die Funktion

$$\boxed{y_3(t) = 3e^t.}$$

2.Zeile: Die DGl $y_2' - y_2 = 2y_3 = 2C_3 e^t$ hat den homogenen Lösungsanteil $y_{2h}(t) = C_2 e^t$. Da der Resonanzfall vorliegt, gewinnt man eine partikuläre Lösung der inhomogenen DGl durch einen Ansatz $y_{2p}(t) = Ate^t$, der zu $A = 2C_3$ führt. Man erhält die allgemeine Lösung $y_2(t) = y_{2h}(t) + y_{2p}(t) = (C_2 + 2C_3 t)e^t$ mit dem Anfangswert $y_2(0) = C_2 \stackrel{!}{=} 2$. Hieraus resultiert als Lösungskomponente der Anfangswertaufgabe (7.18) die Funktion

$$\boxed{y_2(t) = (2 + 6t)e^t.}$$

1.Zeile: Die DGl $y_1' - y_1 = 3y_3 = 3C - 3e^t$ wird ganz analog wie in der 2.Zeile behandelt. Man erhält die allgemeine Lösung $y_1(t) = (C_1 + 3C_3 t)e^t$ mit dem Anfangswert $y_1(0) = C_1 \stackrel{!}{=} 1$, und es ergibt sich als Lösungskomponente der Anfangswertaufgabe (7.18) die Funktion

$$\boxed{y_1(t) = (1 + 9t)e^t.}$$

Die **allgemeine Lösung** des DGl–Systems (7.18) resultiert nun in der Form

$$\begin{aligned}
\vec{y}(t) &= C_1 e^t \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + C_2 e^t \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + C_3 e^t \begin{bmatrix} 3t \\ 2t \\ 1 \\ 0 \end{bmatrix} + C_4 e^{4t} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} e^t & 0 & 3te^t & 0 \\ 0 & e^t & 2te^t & 0 \\ 0 & 0 & e^t & 0 \\ 0 & 0 & 0 & e^{4t} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = Y(t) \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix}.
\end{aligned}$$

Hierin ist $Y(t)$ die bereits bekannte Fundamentalmatrix aus (7.16). Die spezielle Lösung der Anfangswertaufgabe (7.18) hingegen lautet

$$\boxed{\vec{y}(t) = Y(t)\vec{y}_0 = \left((1 + 9t)e^t, (2 + 6t)e^t, 3e^t, 4e^{4t}\right)^T,}$$

in Übereinstimmung mit dem Resultat in BSP. (11.7.8), wenn wir dort $t_0 = 0$ setzen.

Im Hinblick auf das obige Beispiel erinnern wir hier nochmals an die durch Satz 11.15 begründete Transformation einer Matrix $A \in \mathbf{K}^{(n,n)}$ auf die SCHURsche **Normalform**. Wir haben in Satz 11.15 die Existenz einer Matrix $T \in \mathbf{K}^{(n,n)}$ begründet mit

$$T^{-1}AT = \begin{bmatrix} \lambda_1 & * & * & \cdots & * \\ & \lambda_2 & * & \cdots & * \\ & & \lambda_3 & \cdots & * \\ & & & \ddots & \vdots \\ O & & & & \lambda_n \end{bmatrix} =: D_0. \tag{7.19}$$

Durch eine Variablentransformation

$$\boxed{\vec{y}(t) =: T\vec{z}(t)} \tag{7.20}$$

resultiert die mit (7.7) äquivalente Anfangswertaufgabe

$$\boxed{\vec{z}\,'(t) = T^{-1}AT\vec{z}(t) = D_0\vec{z}(t), \qquad \vec{z}(t_0) = T^{-1}\vec{y}_0 =: \vec{z}_0.} \tag{7.21}$$

Da die Matrix D_0 wiederum eine Rechtdreiecksmatrix ist, können wir das DGl–System wie im vorangegangenen BSP. (11.7.10) sukzessive von unten nach oben auflösen. Die Berechnung der SCHURschen Normalform einer Matrix A ist jedoch keineswegs trivial, so dass man diesen besonderen Lösungsweg nur beschreiten wird, wenn aus anderen Problemstellungen die Transformationsmatrix T bereits bekannt ist.

BSP. (11.7.11) Wir betrachten die folgende Matrix $A \in \mathbf{R}^{(3,3)}$, für die man die SCHURsche Normalform und die Transformationsmatrix T ganz analog wie in BSP. (11.4.1) berechnen kann:

$$A := \begin{bmatrix} 1 & -3 & 1 \\ 2 & -4 & 1 \\ 2 & -3 & 0 \end{bmatrix}, \quad T := \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix}.$$

Man ermittelt hieraus

$$D_0 = T^{-1}AT = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

Die Anfangswertaufgabe (7.7) mit dem speziellen Anfangsvektor $\vec{y}_0 := (-1, 1, 6)^T$ und dem Anfangszeitpunkt $t_0 := 0$ wird nun durch die Transformation (7.20) in die folgende Form übergeführt:

$$\begin{bmatrix} z_1'(t) \\ z_2'(t) \\ z_3'(t) \end{bmatrix} = \begin{bmatrix} -z_1(t) \\ -z_2(t) + z_3(t) \\ -z_3(t) \end{bmatrix}, \qquad T^{-1}\vec{y}_0 = \vec{z}(0) = \begin{bmatrix} z_1(0) \\ z_2(0) \\ z_3(0) \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}.$$

Wir lösen dieses System gemäß der Strategie *3.Zeile* $\Rightarrow$ *2.Zeile* $\Rightarrow$ *1.Zeile*.

3.Zeile: Die DGl $z_3' + z_3 = 0$ hat die allgemeine Lösung $z_3(t) = C_3e^{-t}$ mit dem Anfangswert $z_3(0) = C_3 \stackrel{!}{=} 1$. Somit erhalten wir als Lösungskomponente der transformierten Anfangswertaufgabe die Funktion

$$\boxed{z_3(t) = e^{-t}.}$$

2.Zeile: Die DGl $z_2' + z_2 = z_3 = C_3e^{-t}$ hat den homogenen Lösungsanteil $z_{2h}(t) = C_2e^{-t}$. Da der Resonanzfall vorliegt, gewinnt man eine partikuläre Lösung der inhomogenen DGl durch einen Ansatz $z_{2p}(t) = Ate^{-t}$, der zu $A = C_3$ führt. Man erhält die allgemeine Lösung $z_2(t) = z_{2h}(t) + z_{2p}(t) = (C_2 + C_3t)e^{-t}$ mit dem Anfangswert $z_2(0) = C_2 \stackrel{!}{=} 1$. Hieraus resultiert als Lösungskomponente der transformierten Anfangswertaufgabe die Funktion

$$\boxed{z_2(t) = (1 + t)e^{-t}.}$$

1.Zeile: Die DGl $z_1' + z_1 = 0$ wird ganz analog wie in der 3.Zeile behandelt. Man erhält die allgemeine Lösung $z_1(t) = C_1e^{-t}$ mit dem Anfangswert $z_1(0) = C_1 \stackrel{!}{=} -2$, und es ergibt sich als Lösungskomponente der transformierten Anfangswertaufgabe die Funktion

$$\boxed{z_1(t) = -2e^{-t}.}$$

Die **allgemeine Lösung** des transformierten DGl–Systems resultiert nun in der Form

$$\vec{z}(t) = C_1e^{-t}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + C_2e^{-t}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + C_3e^{-t}\begin{bmatrix} 0 \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t} & te^{-t} \\ 0 & 0 & e^{-t} \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} =: Z(t)\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}.$$

Die spezielle Lösung der transformierten Anfangswertaufgabe lautet

$$\vec{z}(t) = Z(t)\vec{z}_0 = \Big(-2e^{-t}, (1+t)e^{-t}, e^{-t}\Big)^T.$$

Nach Rücktransformation $\vec{y}(t) = T\vec{z}(t)$ hat man $\vec{y}(t) = Y(t)\vec{y}_0$ mit der Fundamentalmatrix

$$Y(t) = TZ(t)T^{-1} = \begin{bmatrix} (1+2t)e^{-t} & -3te^{-t} & te^{-t} \\ 2te^{-t} & (1-3t)e^{-t} & te^{-t} \\ 2te^{-t} & -3te^{-t} & (1+t)e^{-t} \end{bmatrix}.$$

Die spezielle Lösung der Anfangswertaufgabe ist somit

$$\boxed{\vec{y}(t) = Y(t)\vec{y}_0 = e^{-t}(t-1, t+1, t+6)^T.}$$

11.8 Ketten von Hauptvektoren

11.8.1 Die Berechnung der Jordan–Normalform einer Matrix

Wir haben in Satz 11.16 behauptet, dass jeder Matrix $A \in \mathbf{K}^{(n,n)}$ eindeutig (bis auf die Reihenfolge der JORDAN–Blöcke) eine Matrix $J \in \mathbf{C}^{(n,n)}$ in der Form (4.2) zugeordnet werden kann, welche vermöge einer (nicht eindeutig bestimmten) Transformation $T \in \mathrm{Inv}\,(\mathbf{C}^n)$ durch

$$\boxed{J = T^{-1}AT} \tag{8.1}$$

aus A hervorgeht. J heißt eine JORDAN**–Normalform** der Matrix A. Jede der in der Relation (4.2) auftretenden Matrizen in der Form

$$J_j := \begin{bmatrix} \lambda_j & 1 & & 0 & 0 \\ & \lambda_j & 1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \lambda_j & 1 \\ 0 & 0 & & & \lambda_j \end{bmatrix} \quad \text{bzw. } J_j := (\lambda_j), \qquad j = 1, 2, \ldots, s, \tag{8.2}$$

heiße ein JORDAN**–Block** oder ein JORDAN**–Kasten**.

BSP. (11.8.1) Mit

$$A := \begin{bmatrix} 2 & 0.5 & -1 & -2.5 \\ 0 & 2 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad T := \begin{bmatrix} 1 & 0 & 2 & 1.5 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad T^{-1} := \begin{bmatrix} 1 & 0 & -1.5 & -2 \\ 0 & 0.5 & -1.5 & -0.5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

folgt durch einfache Rechnung

$$J = T^{-1}AT = \begin{bmatrix} \boxed{\begin{matrix} 2 & 1 \\ 0 & 2 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} 1 & 1 \\ 0 & 1 \end{matrix}} \end{bmatrix} = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}.$$

Bei der **Konstruktion** dieses Demonstrationsbeispiels sind wir von der Vorgabe der Matrix J ausgegangen. Danach haben wir mit einer invertierbaren Matrix T das Matrizenprodukt $A := TJT^{-1}$ berechnet. In der Praxis ist die Aufgabe aber in umgekehrter Richtung zu lösen: Zu vorgegebenem A sollen J und T berechnet werden! Satz 11.16 sichert zunächst die Existenz und die Eindeutigkeit (bis auf die oben erwähnte Reihenfolge der JORDAN–Kästen) einer JORDAN–Normalform J der Matrix A sowie die Existenz einer zugehörigen Transformationsmatrix T. Die Gleichung (8.1) liefert einen Ansatz, wie J und T berechnet werden können: Die zu (8.1) äquivalente Gleichung $AT = TJ$ ist dabei nützlich. Wir behandeln vorab zum besseren Verständnis zwei Sonderfälle.

1. Sonderfall: Im ersten Sonderfall nehmen wir an, die Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ der Matrix A bilden eine **Basis** des $\mathbf{C}^n$.

Satz 11.29 *Bilden die Eigenvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ der Matrix $A \in \mathbf{K}^{(n,n)}$ eine Basis des $\mathbf{C}^n$ mit der Zuordnung $A\vec{v}_j = \lambda_j \vec{v}_j \;\; \forall \; 1 \leq j \leq n$, so leistet die* **Modalmatrix** *$T := (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) \in \mathrm{Inv}\,(\mathbf{C}^n)$ die in* Satz 11.16 *behauptete Transformation auf eine* JORDAN*–Normalform, nämlich*

$$\boxed{T^{-1}AT = J = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n).} \tag{8.3}$$

Begründung: Diese ergibt sich unmittelbar aus Satz 11.8(c). □

BSP. (11.8.2) Die folgende Matrix $A \in \mathbf{R}^{(4,4)}$ hat das charakteristische Polynom $P_4(\lambda) = \det(A - \lambda\, Id) = (2-\lambda)^2(-2-\lambda)^2$:

$$A := \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ -4 & 0 & -2 & 0 \\ 0 & 4 & 0 & -2 \end{bmatrix}, \quad P_4(\lambda) = \begin{vmatrix} 2-\lambda & 0 & 0 & 0 \\ 0 & 2-\lambda & 0 & 0 \\ -4 & 0 & -2-\lambda & 0 \\ 0 & 4 & 0 & -2-\lambda \end{vmatrix}.$$

Es liegt somit je ein doppelter Eigenwert $\lambda_1 = 2$ und $\lambda_2 = -2$ vor. Wir bestimmen zunächst die Eigenvektoren $\vec{v}$ zum Eigenwert $\lambda_1 = 2$:

$$A_{\lambda_1}\vec{v} = \vec{0} \Leftrightarrow \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -4 & 0 & -4 & 0 \\ 0 & 4 & 0 & -4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \;\; \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Wir bestimmen des weiteren die Eigenvektoren $\vec{v}$ zum Eigenwert $\lambda_2 = -2$:

$$A_{\lambda_2}\vec{v} = \vec{0} \Leftrightarrow \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \;\; \vec{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Hiermit haben wir in $\mathbf{C}^4$ eine Basis $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ aus Eigenvektoren der Matrix A bestimmt, so dass die in Satz 11.29 geschilderte Situation vorliegt. Mit Hilfe der Modalmatrix

$$T := (\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

zeigt man nun mit elementarer Rechnung die behauptete Relation (8.3), nämlich

$$T^{-1}AT = J = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} = \mathrm{diag}\,(2, 2, -2, -2).$$

2.Sonderfall: Die Konstruktion einer JORDAN–Normalform wird erheblich komplizierter, wenn die Matrix A Eigenwerte $\lambda \in \mathbf{C}$ besitzt mit geom. dim $\lambda <$ algebr. dim λ. Relativ einfach überschaubar in dieser Klasse von Eigenwerten ist wiederum der folgende

$$\boxed{\textbf{Sonderfall: } 1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda =: k.}$$

In diesem Fall gilt nämlich die Relation (7.6). Das heißt, neben dem Eigenvektor $\vec{w}_1$ zum Eigenwert λ gibt es in jeder Stufe j, $2 \le j \le k$, genau einen von den bereits bestimmten Hauptvektoren **linear unabhängigen Hauptvektor** $\vec{w}_j$ zum Eigenwert λ. Diesen erhält man als (beliebige) partikuläre Lösung des inhomogenen linearen Gleichungssystems

$$\boxed{A_\lambda \vec{w}_j := (A - \lambda\, Id)\vec{w}_j = \vec{w}_{j-1}, \;\; \text{äquivalent} \;\; A\vec{w}_j = \vec{w}_{j-1} + \lambda\vec{w}_j, \quad j = 2, 3, \ldots, k.} \tag{8.4}$$

Die Rekursionsformel (8.4), nämlich $A_\lambda \vec{w}_j = \vec{w}_{j-1}$, führt durch wiederholte Anwendung zur expliziten Darstellung $\vec{w}_j = A_\lambda^{k-j}\vec{w}_k$, $j = 1, 2, \ldots, k-1$. Das heißt, das hier vorliegende System von Hauptvektoren $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ können wir auch in der Form $A_\lambda^{k-1}\vec{w}_k, A_\lambda^{k-2}\vec{w}_k, \ldots, \vec{w}_k$ schreiben. Wir sind für die folgende Definition motiviert:

Definition 11.13 *Eine endliche Folge von Hauptvektoren $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ der Stufen 1 bis k (zum selben Eigenwert $\lambda \in \mathbf{C}$ der Matrix $A \in \mathbf{K}^{(n,n)}$) heiße eine* **Kette von Hauptvektoren** *zum Eigenwert λ, wenn gilt:*

$$\boxed{\vec{0} = A_\lambda^k \vec{w}_k, \;\; \vec{w}_1 = A_\lambda^{k-1}\vec{w}_k, \;\; \vec{w}_2 = A_\lambda^{k-2}\vec{w}_k, \;\; \ldots, \;\; \vec{w}_{k-1} = A_\lambda \vec{w}_k.} \tag{8.5}$$

Dabei fordern wir zusätzlich, dass die endliche Folge nicht verlängerbar ist, d.h. dass $A_\lambda \vec{w} = \vec{w}_k$ keine Lösung $\vec{w}$ besitzt. In diesem Fall heiße $\vec{w}_k$ ein **Hauptvektor höchster Stufe**.

Wie wir in Satz 11.25(a) gezeigt haben, ist eine Kette von Hauptvektoren zum Eigenwert λ stets linear unabhängig. Der verallgemeinerte Eigenraum $E(\lambda) := \text{Kern}\, A_\lambda^f$ besitzt dann gemäß Satz 11.26 eine Basis aus solchen Ketten von Hauptvektoren zum selben Eigenwert.

Im vorliegenden Sonderfall $1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda = k$ wird $E(\lambda)$ durch **genau** eine Kette von Hauptvektoren aufgespannt, nämlich von der durch (8.4) bestimmten Kette $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$. Bilden wir mit dieser Kette die Matrix $\tilde{T} := (\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k) \in \mathbf{C}^{(n,k)}$, so ergibt sich aus (8.4) mit einigen elementaren Rechnungen die Beziehung

$$\begin{aligned} A\tilde{T} &= (A\vec{w}_1, A\vec{w}_2, \ldots, A\vec{w}_k) = (\lambda\vec{w}_1, \vec{w}_1 + \lambda\vec{w}_2, \ldots, \vec{w}_{k-1} + \lambda\vec{w}_k) \\ &= (\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k) \begin{bmatrix} \lambda & 1 & & 0 & 0 \\ & \lambda & 1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \lambda & 1 \\ 0 & 0 & & & \lambda \end{bmatrix} = \tilde{T} J_\lambda. \end{aligned} \tag{8.6}$$

Aus dieser Relation leiten wir die folgende Aussage ab.

Satz 11.30 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Sei $\lambda \in \mathbf{C}$ ein Eigenwert von A, welcher der Bedingung $1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda = k$ genügt. Sei $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ die gemäß* (8.4)

bestimmte Kette von Hauptvektoren $\vec{w}_j \in \mathbf{C}^n$ der Stufen $j = 1, 2, \ldots, k$ zum Eigenwert λ. Dann ist dem Eigenwert λ eindeutig der JORDAN*–Kasten*

$$J_\lambda := \begin{bmatrix} \lambda & 1 & & 0 & 0 \\ & \lambda & 1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \lambda & 1 \\ 0 & 0 & & & \lambda \end{bmatrix} \in \mathbf{C}^{(k,k)}$$

zugeordnet, und die aus den Hauptvektoren gebildete Matrix $\tilde{T} := (\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k) \in \mathbf{C}^{(n,k)}$ leistet die Transformation $A\tilde{T} = \tilde{T}J_\lambda$. Ist $T := \tilde{T}$ invertierbar, so resultiert aus $T^{-1}AT = J_\lambda =: J$ bereits die JORDAN*–Normalform der Matrix A.*

BSP. (11.8.3) Wir betrachten die Matrix $A \in \mathbf{R}^{(4,4)}$ aus BSP. (11.7.3), nämlich

$$A := \begin{bmatrix} 3 & 1 & -2 & 3 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}.$$

Wie dort bereits gezeigt wurde, besitzt A den vierfachen Eigenwert $\lambda = 3$, für den der Sonderfall $1 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda = 4$ vorliegt. Mit der zugeordneten Kette von Hauptvektoren $\vec{w}_1$, $\vec{w}_2$, $\vec{w}_3$, $\vec{w}_4$ (vgl. BSP. (11.7.3)) bilden wir die Matrizen

$$T := (\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -3 & 8 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \in \text{Inv}\,(\mathbf{R}^4), \quad T^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Es ergibt sich nun mit einfacher Rechnung die JORDAN–Normalform

$$J = T^{-1}AT = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Sind $\lambda_1, \lambda_2 \in \mathbf{C}$ **zwei verschiedene** Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$ mit $1 = \text{geom. dim}\,\lambda_j < \text{algebr. dim}\,\lambda_j = k_j$, so bilden wir in Analogie zu (8.2) die beiden JORDAN–Kästen $J_j \in \mathbf{C}^{(k_j,k_j)}$, $j = 1, 2$ sowie die Matrix $\tilde{T} := (\vec{w}_1^1, \ldots, \vec{w}_{k_1}^1, \vec{w}_1^2, \ldots, \vec{w}_{k_2}^2) \in \mathbf{C}^{(n,k_1+k_2)}$, wobei $\vec{w}_1^j, \vec{w}_2^j, \ldots, \vec{w}_{k_j}^j$ die dem Eigenwert λ_j gemäß (8.4) zugeordnete Kette von Hauptvektoren ist. Nun ist es leicht, die folgende Identität zu verifizieren:

$$A\tilde{T} = \tilde{T} \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}.$$

Das heißt, ist $T := \tilde{T}$ invertierbar, so resultiert aus $T^{-1}AT = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} =: J$ eine JORDAN–Normalform der Matrix A. Eine Verallgemeinerung dieser Situation auf endlich viele solcher Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_m$ liegt nun auf der Hand.

BSP. (11.8.4) Die Matrix $A \in \mathbf{R}^{(4,4)}$ aus BSP. (11.8.1) ist identisch mit der in BSP. (11.7.7) untersuchten Matrix A. Wir hatten dort gezeigt, dass A die beiden Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = 1$ mit $1 = \text{geom. dim}\,\lambda_j < \text{algebr. dim}\,\lambda_j = 2$ besitzt. Die zugeordneten Hauptvektor–Ketten lauten (vgl. BSP. (11.7.6)):

$$\vec{w}_1^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_2^1 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}; \quad \vec{w}_1^2 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{w}_2^2 = \begin{bmatrix} 1.5 \\ 3 \\ 1 \\ 0 \end{bmatrix}.$$

Die Matrix $T := (\vec{w}_1^{\,1}, \vec{w}_2^{\,1}, \vec{w}_1^{\,2}, \vec{w}_2^{\,2})$ ist genau die in BSP. (11.8.1) angegebene Matrix T. Dort hatten wir bereits die Transformation auf JORDAN–Normalform vorgenommen:

$$J = T^{-1}AT = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} = \left[\begin{array}{cc|cc} \boxed{\begin{matrix} 2 & 1 \\ 0 & 2 \end{matrix}} & & 0 & 0 \\ & & 0 & 0 \\ \hline 0 & 0 & \boxed{\begin{matrix} 1 & 1 \\ 0 & 1 \end{matrix}} & \\ 0 & 0 & & \end{array}\right].$$

Es bedarf keiner großen Einsicht, dass auch eine Kombination des Satzes 11.29 mit dem hier diskutierten Sonderfall $1 = \text{geom. dim}\,\lambda_1 < \text{algebr. dim}\,\lambda_1 = k_1$ möglich ist. Haben wir einen weiteren Eigenwert $\lambda_2 \in \mathbf{C}$ mit $\text{geom. dim}\,\lambda_2 = \text{algebr. dim}\,\lambda_2 = k_2 \geq 1$ vorliegen und bezeichnen $\vec{v}_1^{\,2}, \vec{v}_2^{\,2}, \ldots, \vec{v}_{k_2}^{\,2}$ die dem Eigenwert λ_2 zugeordneten linear unabhängigen Eigenvektoren, so bilden wir jetzt die Matrix $\tilde{T} := (\vec{w}_1^{\,1}, \ldots, \vec{w}_{k_1}^{\,1}, \vec{v}_1^{\,2}, \ldots, \vec{v}_{k_2}^{\,2}) \in \mathbf{C}^{(n,k_1+k_2)}$ sowie die Diagonalmatrix $D_2 := \text{diag}\,(\lambda_2, \lambda_2, \ldots, \lambda_2) \in \mathbf{C}^{(k_2,k_2)}$. Dann überzeugt man sich leicht von der folgenden Identität

$$A\tilde{T} = \tilde{T} \begin{bmatrix} J_1 & 0 \\ 0 & D_2 \end{bmatrix}.$$

Das heißt wiederum, falls $T := \tilde{T}$ invertierbar ist, so resultiert aus $T^{-1}AT = \begin{bmatrix} J_1 & 0 \\ 0 & D_2 \end{bmatrix} =: J$ eine JORDAN–Normalform der Matrix A. Eine Verallgemeinerung dieser Situation auf endlich viele solcher Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_m$ liegt abermals auf der Hand.

BSP. (11.8.5) Die folgende Matrix $A \in \mathbf{R}^{(4,4)}$ hat das charakteristische Polynom $P_4(\lambda) = \det(A - \lambda\, Id) = (1-\lambda)^2(\lambda-2)^2$:

$$A := \begin{bmatrix} 2.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \\ -0.5 & 0 & 1.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad P_4(\lambda) = \begin{vmatrix} 2.5-\lambda & 0 & 0.5 & 0 \\ 0 & 1-\lambda & 0 & 0 \\ -0.5 & 0 & 1.5-\lambda & 0 \\ 0 & 0 & 0 & 1-\lambda \end{vmatrix}.$$

Es liegen je ein doppelter Eigenwert $\lambda_1 = 2$ und $\lambda_2 = 1$ vor. Wir bestimmen zunächst die Eigenvektoren $\vec{w}_1$ zum Eigenwert $\lambda_1 = 2$:

$$A_{\lambda_1}\vec{w}_1 = \vec{0} \Leftrightarrow \begin{bmatrix} 0.5 & 0 & 0.5 & 0 \\ 0 & -1 & 0 & 0 \\ -0.5 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{w}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}.$$

Hier liegt offensichtlich der **Sonderfall** $1 = \text{geom. dim}\,\lambda_1 < \text{algebr. dim}\,\lambda_1 = 2$ vor. Wir bestimmen nun den Hauptvektor $\vec{w}_2$ der Stufe 2 gemäß

$$A_{\lambda_1}\vec{w}_2 = \vec{w}_1 \Leftrightarrow \begin{bmatrix} 0.5 & 0 & 0.5 & 0 \\ 0 & -1 & 0 & 0 \\ -0.5 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_2 = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Wir bestimmen des weiteren die Eigenvektoren $\vec{v}$ zum Eigenwert $\lambda_2 = 1$:

$$A_{\lambda_2}\vec{v} = \vec{0} \Leftrightarrow \begin{bmatrix} 1.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \\ -0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Wir haben hier $\text{geom. dim}\,\lambda_2 = \text{algebr. dim}\,\lambda_2 = 2$. Nun bilden wir mit den Vektoren $\vec{w}_1, \vec{w}_2, \vec{v}_1, \vec{v}_2$ die Matrizen

$$T := (\vec{w}_1, \vec{w}_2, \vec{v}_1, \vec{v}_2) = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Die behauptete JORDAN–Normalform ergibt sich gemäß

$$J = T^{-1}AT = \left[\begin{array}{|cc|cc} \hline 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline \end{array}\right] = \begin{bmatrix} J_1 & 0 \\ 0 & D_2 \end{bmatrix}.$$

In der Zusammenfassung haben wir gezeigt:

Satz 11.31 *Die Matrix $A \in \mathbf{K}^{(n,n)}$ habe die paarweise verschiedenen komplexen Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_m$ mit algebraischen Vielfachheiten $k_1, k_2, \ldots, k_m$; $k_1 + k_2 + \cdots + k_m = n$. Jeder Eigenwert λ_j erfülle* **entweder** $1 = \text{geom. dim}\,\lambda_j < k_j$ **oder** $1 \leq \text{geom. dim}\,\lambda_j = k_j$. *Dann besitzt die* JORDAN*–Normalform der Matrix A die Gestalt*

$$J = \begin{bmatrix} J_1 & 0 & \cdots & 0 & 0 \\ 0 & J_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & J_{m-1} & 0 \\ 0 & 0 & \cdots & 0 & J_m \end{bmatrix} \tag{8.7}$$

mit den JORDAN*–Kästen*

$$J_j := \begin{cases} \begin{bmatrix} \lambda_j & 1 & & 0 & 0 \\ & \lambda_j & 1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \lambda_j & 1 \\ 0 & 0 & & & \lambda_j \end{bmatrix} \in \mathbf{C}^{(k_j,k_j)} & : \text{ falls } 1 = \text{geom. dim}\,\lambda_j < k_j, \\ \text{diag}\,(\lambda_j, \lambda_j, \cdots, \lambda_j) \in \mathbf{C}^{(k_j,k_j)} & : \text{ falls } 1 \leq \text{geom. dim}\,\lambda_j = k_j. \end{cases}$$

Eine Matrix $T \in \text{Inv}\,(\mathbf{C}^n)$, welche die Transformation $J = T^{-1}AT$ bewerkstelligt, ist wie folgt bestimmt: Die Sequenz der k_j Spaltenvektoren $T = (\ldots, \vec{t}_{\ell+1}, \vec{t}_{\ell+2}, \ldots, \vec{t}_{\ell+k_j}, \ldots)$, $\ell := k_1 + k_2 + \cdots + k_{j-1}$, besteht entweder aus der Kette der Hauptvektoren $\vec{w}_1^{\,j}, \vec{w}_2^{\,j}, \ldots, \vec{w}_{k_j}^{\,j}$, die im Falle $1 = \text{geom. dim}\,\lambda_j < k_j$ dem Eigenwert λ_j zugeordnet ist, oder aus den k_j linear unabhängigen Eigenvektoren $\vec{v}_1^{\,j}, \vec{v}_2^{\,j}, \ldots, \vec{v}_{k_j}^{\,j}$, die im Falle geom. dim $\lambda_j = k_j$ *zum Eigenwert λ_j gehören.*

BSP. (11.8.6) Die folgende Matrix $A \in \mathbf{R}^{(5,5)}$ besitzt das charakteristisches Polynom $P_5(\lambda) = \det(A - \lambda\, Id) = (3-\lambda)^2(2-\lambda)^2(-\lambda)$:

$$A := \frac{1}{2}\begin{bmatrix} 5 & 1 & 1 & 1 & -1 \\ 0 & 5 & 1 & 0 & -1 \\ 0 & 1 & 5 & 0 & -5 \\ 1 & 1 & 1 & 5 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad P_5(\lambda) = \begin{vmatrix} 2.5-\lambda & 0.5 & 0.5 & 0.5 & -0.5 \\ 0 & 2.5-\lambda & 0.5 & 0.5 & -0.5 \\ 0 & 0.5 & 2.5-\lambda & 0 & -2.5 \\ 0.5 & 0.5 & 0.5 & 2.5-\lambda & -0.5 \\ 0 & 0 & 0 & 0 & -\lambda \end{vmatrix}.$$

Es liegen somit je ein doppelter Eigenwert $\lambda_1 = 3$ und $\lambda_2 = 2$ sowie ein einfacher Eigenwert $\lambda_3 = 0$ vor. Wir bestimmen zunächst die Eigenvektoren $\vec{w}_1$ zum Eigenwert $\lambda_1 = 3$:

$$A_{\lambda_1}\vec{w}_1 = \vec{0} \Leftrightarrow \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & -5 \\ 1 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{w}_1^{\,1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Hier liegt offensichtlich der **Sonderfall** $1 = \text{geom. dim}\,\lambda_1 < \text{algebr. dim}\,\lambda_1 = 2$ vor. Wir müssen deshalb den Hauptvektor $\vec{w}_2^1$ der Stufe 2 bestimmen:

$$A_{\lambda_1}\vec{w}_2 = \vec{w}_1^1 \Leftrightarrow \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & -5 \\ 1 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w5 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{also} \quad \vec{w}_2^1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Wir bestimmen des weiteren die Eigenvektoren $\vec{v}$ zum Eigenwert $\lambda_2 = 2$:

$$A_{\lambda_2}\vec{v} = \vec{0} \Leftrightarrow \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & -5 \\ 1 & 1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{v}_1^2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad \vec{v}_2^2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}.$$

Wir haben hier $2 = \text{geom. dim}\,\lambda_2 = \text{algebr. dim}\,\lambda_2$. Nun berechnen wir schließlich den Eigenvektor $\vec{v}$ zum Eigenwert $\lambda_3 = 0$:

$$A_{\lambda_3}\vec{v} = \vec{0} \Leftrightarrow \frac{1}{2}\begin{bmatrix} 5 & 1 & 1 & 1 & -1 \\ 0 & 5 & 1 & 0 & -1 \\ 0 & 1 & 5 & 0 & -5 \\ 1 & 1 & 1 & 5 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} \stackrel{!}{=} \vec{0}, \quad \text{also} \quad \vec{v}_1^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Wir bilden mit den Vektoren $\vec{w}_1^1, \vec{w}_2^1, \vec{v}_1^2, \vec{v}_2^2, \vec{v}_1^3$ die Transformationsmatrizen

$$T := (\vec{w}_1^1, \vec{w}_2^1, \vec{v}_1^2, \vec{v}_2^2, \vec{v}_1^3) = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad T^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Die der Matrix A zugeordnete JORDAN–Normalform ergibt sich im Einklang mit Satz 11.31 zu

$$J = T^{-1}AT = \begin{bmatrix} \boxed{\begin{matrix} 3 & 1 \\ 0 & 3 \end{matrix}} & 0 & 0 & 0 \\ & \boxed{\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}} & \boxed{\begin{matrix} 2 & 0 \\ 0 & 2 \end{matrix}} & 0 & 0 \\ & & & \boxed{0} & \end{bmatrix}.$$

Allgemeiner Fall: Im allgemeinen Fall $1 < \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda =: k$ bestimmen wir zunächst eine Basis $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k$ des verallgemeinerten Eigenraumes $E(\lambda)$ zum Eigenwert $\lambda \in \mathbf{C}$, zum Beispiel mit dem **Verfahren des Kernaustauschs**. Diese Basis muss **nicht mehr notwendig** aus einer einzigen Kette von Hauptvektoren bestehen; sie enthält jedoch mindestens einen Hauptvektor $\vec{b}_{\kappa_1}$ höchster Stufe $\kappa_1 := f \leq k$, wobei $f = f(\lambda)$ den Fittingindex von λ bezeichnet. Wir gehen **vorläufig** nach folgender Vorschrift vor:

- **Step 1:** Wähle aus der Basis $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k$ des verallgemeinerten Eigenraumes $E(\lambda)$ einen Hauptvektor $\vec{b}_{\kappa_1}$ höchster Stufe und bilde die Kette von Hauptvektoren

$$\vec{w}_{\kappa_1}^1 := \vec{b}_{\kappa_1}, \ \vec{w}_{\kappa_1-1}^1 := A_\lambda \vec{w}_{\kappa_1}^1, \ \vec{w}_{\kappa_1-2}^1 := A_\lambda \vec{w}_{\kappa_1-1}^1, \ \ldots, \ \vec{w}_1^1 := A_\lambda \vec{w}_2^1. \tag{8.8}$$

  Setze $E_{\kappa_1}(\lambda) := \text{span}\,\{\vec{w}_1^1, \vec{w}_2^1, \ldots, \vec{w}_{\kappa_1}^1\}$. Wegen des Austauschsatzes 4.9 können genau κ_1 der Basisvektoren $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k$ gestrichen werden, und zwar so, dass das Restsystem den Ergänzungsraum $E_{k-\kappa_1}(\lambda)$ mit $E(\lambda) = E_{\kappa_1}(\lambda) \oplus E_{k-\kappa_1}(\lambda)$ aufspannt.

- **Step 2:** Wähle aus dem Restsystem der Basisvektoren einen zweiten Hauptvektor $\vec{b}_{\kappa_2}$ ($\kappa_2 \leq \kappa_1$) höchster Stufe und bilde ebenfalls die Kette von Hauptvektoren

$$\vec{w}^2_{\kappa_2} := \vec{b}_{\kappa_2},\ \vec{w}^2_{\kappa_2-1} := A_\lambda \vec{w}^2_{\kappa_2},\ \vec{w}^2_{\kappa_2-2} := A_\lambda \vec{w}^2_{\kappa_2-1},\ \ldots,\ \vec{w}^2_1 := A_\lambda \vec{w}^2_2. \tag{8.9}$$

  Setze $E_{\kappa_2}(\lambda) := \text{span}\{\vec{w}^2_1, \vec{w}^2_2, \ldots, \vec{w}^2_{\kappa_2}\}$ und streiche diejenigen κ_2 Basisvektoren des Restsystems, die nicht den Ergänzungsraum von $E_{\kappa_1}(\lambda) \oplus E_{\kappa_2}(\lambda)$ aufspannen, usf.

Das Verfahren endet nach $s < k$ solchen Schritten, nämlich wenn $\kappa_1 + \kappa_2 + \cdots + \kappa_s = k$ gilt, oder äquivalent, wenn wir

$$E_{\kappa_1}(\lambda) \oplus E_{\kappa_2}(\lambda) \oplus \cdots \oplus E_{\kappa_s}(\lambda) = E(\lambda)$$

erreicht haben. Das ist genau dann der Fall, wenn das Restsystem keine Basisvektoren mehr enthält.

- **Step 3:** Bilde mit diesen s Ketten von Hauptvektoren die Matrix

$$\tilde{T} := (\underbrace{\vec{w}^1_1, \ldots, \vec{w}^1_{\kappa_1}}_{\text{1. Kette}}, \underbrace{\vec{w}^2_1, \ldots, \vec{w}^2_{\kappa_2}}_{\text{2. Kette}}, \ldots, \underbrace{\vec{w}^s_1, \ldots, \vec{w}^s_{\kappa_s}}_{s\text{-te Kette}}) \in \mathbf{C}^{(n,k)}.$$

  Dann ergibt sich in Analogie zu (8.6) die Beziehung

$$A\tilde{T} = \tilde{T} \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_s \end{bmatrix} =: \tilde{T}\tilde{J} \ \text{ mit } \ J_j := \begin{bmatrix} \lambda & 1 & & & 0 & 0 \\ & \lambda & 1 & & & 0 \\ & & \ddots & \ddots & & \\ & & & & \lambda & 1 \\ 0 & & & & & \lambda \end{bmatrix} \in \mathbf{C}^{(\kappa_j,\kappa_j)}, \tag{8.10}$$

  für $j = 1, 2, \ldots, s$.

Ist die Matrix $T := \tilde{T}$ schon invertierbar, so haben wir mit der Matrix $J := \tilde{J}$ bereits eine JORDAN–Normalform von A vorliegen.

BSP. (11.8.7) Wir betrachten hier nochmals die Matrix $A \in \mathbf{R}^{(5,5)}$ aus BSP. (11.7.5) mit dem fünffachen Eigenwert $\lambda = 0$, für den die Relation $3 = \text{geom. dim}\,\lambda < \text{algebr. dim}\,\lambda = 5 =: k$ erfüllt ist (vgl. auch BSP.(11.7.4)). Gemäß BSP. (11.7.5) wird der verallgemeinerte Eigenraum $E(\lambda) = \text{Kern}\,(A - \lambda\, Id)^{f(\lambda)}$ von den folgenden Vektoren aufgespannt:

$$\vec{b}^1_1 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ \vec{b}^2_1 := \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ \vec{b}^3_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 0 \end{bmatrix},\ \vec{b}^1_2 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ \vec{b}^1_3 := \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{8} \end{bmatrix}.$$

Es gilt ferner $f(\lambda) = 3 =: \kappa_1$, und ein Hauptvektor der höchsten Stufe κ_1 ist durch $\vec{b}^1_3$ gegeben. Also setzen wir vorschriftsgemäß $\vec{w}^1_3 := \vec{b}^1_3$. Die Matrix

$$A_\lambda := A = \begin{bmatrix} -2 & 0 & 2 & 1 & 4 \\ 4 & 0 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ -4 & 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

gestattet die unmittelbare Überprüfung der Relationen

$$A_\lambda \vec{w}^1_3 = \vec{w}^1_2 := \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad A_\lambda \vec{w}^1_2 = \vec{w}^1_1 := \begin{bmatrix} -2 \\ 4 \\ 0 \\ -4 \\ 0 \end{bmatrix}.$$

Somit haben wir eine erste Kette von Hauptvektoren, die den Teilraum $E_{\kappa_1}(\lambda) := \{\vec{w}_1^1, \vec{w}_2^1, \vec{w}_3^1\}$ aufspannen. Wegen $\vec{b}_2^1 = \frac{1}{2}\vec{b}_1^2 + \vec{w}_2^1$ und $\vec{b}_1^3 = 2\vec{b}_1^1 - \frac{1}{2}\vec{w}_1^1$ spannen die beiden Vektoren $\vec{w}_1^2 := \vec{b}_1^1$ und $\vec{w}_1^3 := \vec{b}_1^2$ den Ergänzungsraum von $E_{\kappa_1}(\lambda)$ auf. Hierin sind die beiden Eigenvektoren $\vec{w}_1^2$ und $\vec{w}_1^3$ jeweils Hauptvektoren höchster Stufe, so dass nun mit $\{\vec{w}_1^1, \vec{w}_2^1, \vec{w}_3^1, \vec{w}_1^2, \vec{w}_1^3\}$ eine Basis des verallgemeinerten Eigenraumes $E(\lambda)$ aus Ketten von Hauptvektoren vorliegt. Wir bilden mit dieser Basis die Transformationsmatrizen

$$T := (\vec{w}_1^1, \vec{w}_2^1, \vec{w}_3^1, \vec{w}_1^2, \vec{w}_1^3) = \begin{bmatrix} -2 & \frac{1}{2} & 0 & 0 & 1 \\ 4 & 0 & 0 & 1 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 1 \\ -4 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & 0 \end{bmatrix}, \quad T^{-1} = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & -1 & 0 \\ 4 & 0 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & 32 \\ 0 & 4 & 0 & 4 & 0 \\ 2 & 0 & 2 & -1 & 0 \end{bmatrix}.$$

Die der Matrix A zugeordnete JORDAN–Normalform ergibt sich im Einklang mit (8.10) zu

$$J = T^{-1}AT = \begin{bmatrix} \boxed{\begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{matrix} \boxed{0} & 0 \\ 0 & \boxed{0} \end{matrix} \end{bmatrix} = \begin{bmatrix} J_1 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_3 \end{bmatrix}.$$

Das obige BSP. (11.8.7) zeigt, dass die Überprüfung der linearen Abhängigkeiten zwischen der Basis $\{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k\}$ des verallgemeinerten Eigenraumes $E(\lambda)$ und den Teilräumen $E_{\kappa_j}(\lambda)$ sehr mühevoll sein kann. Deshalb ist es sinnvoll, bereits bei der Berechnung der Hauptvektorbasis $\{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k\}$ die Bildung von Hauptvektor–**Ketten** vorzunehmen. Entscheidend dabei ist, dass nur diejenigen Eigenvektoren nicht schon Hauptvektoren **höchster Stufe** sind, die in der Schnittmenge $\operatorname{Kern} A_\lambda \cap (\operatorname{Kern} A_\lambda^*)^\perp$ liegen. Durch das Verfahren des Kernaustauschs werden aber genau diese Eigenvektoren festgelegt. Im 2. Schritt des Kernaustauschverfahrens werden nämlich die Parameter $\mu_i \in \mathbf{C}$ derjenigen Linearkombination $\vec{b}_1 = \sum_{i=1}^{r_1} \mu_i \vec{w}_1^i$ von Eigenvektoren $\vec{w}_1^1, \vec{w}_1^2, \ldots, \vec{w}_1^{r_1}$ berechnet, welche die Lösbarkeitsbedingung $\vec{b}_1 \perp \operatorname{Kern} A_\lambda^*$ befriedigen. Für die Dimension s_1 des Unterraums $\left\{\vec{b}_1 = \sum_{i=1}^{r_1} \mu_i \vec{w}_1^i : \vec{b}_1 \perp \operatorname{Kern} A_\lambda^*\right\}$ gilt stets $1 \le s_1 \le r_1$. Das heißt, es gibt s_1 Eigenvektoren $\vec{w}_1^j \in \operatorname{Bild} A_\lambda = (\operatorname{Kern} A_\lambda^*)^\perp$, $1 \le j \le s_1$ sowie einen Rest Eigenvektoren $\vec{w}_1^j \notin \operatorname{Bild} A_\lambda$, $s_1 + 1 \le j \le r_1$. Wegen dieses Sachverhalts geben wir das folgende **revidierte Verfahren** zur Berechnung der Hauptvektorketten zum Eigenwert $\lambda \in \mathbf{C}$ an:

- **Step 1:** Bestimme zum Eigenwert $\lambda \in \mathbf{C}$ eine Eigenvektor–Basis $\vec{w}_1^1, \vec{w}_1^2, \ldots, \vec{w}_1^{r_1}$ für den Eigenraum $\operatorname{Kern} A_\lambda = \operatorname{span}\{\vec{w}_1^1, \vec{w}_1^2, \ldots, \vec{w}_1^{r_1}\}$.

- **Step 2:** Es bezeichne $\tilde{A}_\lambda$ die durch Kernaustausch aus A_λ entstandene Matrix. Bestimme mit dem Verfahren des Kernaustauschs wie vorher partikuläre Lösungen $\vec{v}_p$ der (in)homogenen Systeme $\tilde{A}_\lambda \vec{v}_p = \vec{w}_{p-1}$ und die zugeordneten Hauptvektoren $\vec{w}_p$. Bestimme dazu insbesondere diejenigen Parameter $\mu_1^{p-1}, \mu_2^{p-1}, \ldots, \mu_{r_1}^{p-1}$ der Linearkombination $\vec{b}_{p-1} := \sum_{i=1}^{r_1} \mu_i^{p-1} \vec{w}_1^i$, die die Lösbarkeitsbedingung $\vec{b}_{p-1} \in \operatorname{Bild} A_\lambda = (\operatorname{Kern} A_\lambda^*)^\perp$ liegen. Das Verfahren des Kernaustauschs liefert diese Parameter ohne Mehraufwand.

- **Step 3:** Beginnend mit den Hauptvektoren höchster Stufe $\vec{u}_p := \vec{w}_p$ bilde man die zugeordnete Hauptvektorkette $\vec{u}_{p-j} := \vec{w}_{p-j} + \sum_{i=1}^{r_1} \mu_i^{p-j} \vec{w}_1^i$, $j = 1, 2, \ldots, p-2$ sowie $\vec{u}_1 := \sum_{i=1}^{r_1} \mu_i^1 \vec{w}_1^i$, worin die Koeffizienten μ_i^{p-j} genau die in Step 2 bestimmten Parameter sind. Die von $\vec{u}_1$ linear unabhängigen Eigenvektoren $\vec{w}_1^i$ sind bereits Hauptvektoren höchster Stufe.

BSP. (11.8.8) Wir berechnen erneut die Hauptvektorketten der Matrix $A \in \mathbf{R}^{(5,5)}$ aus BSP. (11.7.5), und zwar nach dem revidierten Verfahren. Die Matrix A hat den fünffachen Eigenwert $\lambda = 0$. Wie wir bereits gezeigt haben, gelten $A_\lambda = A$, $\operatorname{Kern} A_\lambda = \operatorname{span}\{\vec{w}_1^1, \vec{w}_1^2, \vec{w}_1^3\}$ mit

$$\vec{w}_1^1 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_1^2 := \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_1^3 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}.$$

Step 2: Kernaustausch. In BSP. (11.7.5) wurden die partikulären Lösungen

$$\vec{w}_2^1 := (1,0,0,0,0)^T \qquad \text{zum Parametersatz} \qquad (\mu_1^1, \mu_2^1, \mu_3^1) := (4,0,-2),$$
$$\vec{w}_3^1 := (0,0,0,0,\tfrac{1}{8})^T \qquad \text{zum Parametersatz} \qquad (\mu_1^2, \mu_2^2, \mu_3^2) := (0,-\tfrac{1}{2},0)$$

berechnet.

Step 3: Es ist $\vec{u}_3 := \vec{w}_3^1 = (0,0,0,0,\frac{1}{8})^T$ ein Hauptvektor höchster Stufe. Dazu konstruieren wir die Kette

$$\vec{u}_2 := \vec{w}_2^1 + \sum_{i=1}^{3} \mu_i^2 \vec{w}_1^i = \vec{w}_2^1 - \tfrac{1}{2}\,\vec{w}_1^2 = (\tfrac{1}{2},\, 0,\, -\tfrac{1}{2},\, 0,\, 0)^T,$$
$$\vec{u}_1 := \sum_{i=1}^{3} \mu_i^1 \vec{w}_1^i = 4\vec{w}_1^1 - 2\,\vec{w}_1^3 = (-2,\, 4,\, 0,\, -4,\, 0)^T.$$

Da die drei Vektoren $\vec{u}_1, \vec{w}_1^1, \vec{w}_1^2$ linear unabhängige Eigenvektoren sind, haben wir dasselbe Resultat wie in BSP. (11.8.8) vorliegen: Die drei Hauptvektorketten

$$\vec{u}_1 := \begin{bmatrix} -2 \\ 4 \\ 0 \\ -4 \\ 0 \end{bmatrix}, \quad \vec{u}_2 := \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{u}_3 := \frac{1}{8}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{w}_1^1 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_1^2 := \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

spannen den verallgemeinerten Eigenraum $E(\lambda) := \operatorname{Kern} A_\lambda^3$ auf.

BSP. (11.8.9) Die folgende Matrix $A \in \mathbf{R}^{(5,5)}$ besitzt das charakteristisches Polynom $P_5(\lambda) = \det(A - \lambda\, Id) = (2-\lambda)^5$ und somit genau einen Eigenwert $\lambda = 2$ mit algebr. $\dim \lambda = 5 =: k$:

$$A := \frac{1}{2}\begin{bmatrix} 4 & 1 & 1 & 0 & -1 \\ 1 & 4 & 0 & -1 & 2 \\ 1 & 0 & 4 & -1 & -2 \\ 0 & 1 & 1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}, \quad P_5(\lambda) = \begin{vmatrix} 2-\lambda & 0.5 & 0.5 & 0 & -0.5 \\ 0.5 & 2-\lambda & 0 & -0.5 & 1 \\ 0.5 & 0 & 2-\lambda & -0.5 & -1 \\ 0 & 0.5 & 0.5 & 2-\lambda & -0.5 \\ 0 & 0 & 0 & 0 & 2-\lambda \end{vmatrix}.$$

Step 1: Mit Hilfe der Matrix $A_\lambda := (A - 2\ Id)$ berechnen wir zunächst den Eigenraum $\operatorname{Kern} A_\lambda$. Es gilt

$$A_\lambda = \frac{1}{2}\begin{bmatrix} 0 & 1 & 1 & 0 & -1 \\ 1 & 0 & 0 & -1 & 2 \\ 1 & 0 & 0 & -1 & -2 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

und somit

$$2A_\lambda \vec{w}_1 = \vec{0} \quad \rightleftharpoons \quad \left[\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 2 & 0 \\ 1 & 0 & 0 & -1 & -2 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \text{Gauss} \rightleftharpoons \quad \left[\begin{array}{ccccc|c} & & c_1 & c_2 & & \\ \hline 1 & 0 & 0 & -1 & 2 & 0 \\ 0 & 1 & \boxed{1} & \boxed{0} & -1 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Wir erhalten 2 = geom. $\dim\lambda <$ algebr. $\dim\lambda = 5 =: k$. Die Parameterwahl $(c_1, c_2) := (1,0)$, $:= (0,1)$ führt auf die beiden Eigenvektoren

$$\vec{w}_1^{\,1} := (0,\,-1,\,1,\,0,\,0)^T, \quad \vec{w}_1^{\,2} := (1,\,0,\,0,\,1,\,0)^T.$$

Step 2: Kernaustausch. Wir ersetzen in A_λ die Spalten S_3 und S_4 (entsprechend den Spalten unter c_1, c_2) durch die Eigenvektoren $\vec{w}_1^{\,1}, \vec{w}_1^{\,2}$ und erhalten die Matrix $\tilde{A}_\lambda$. Danach reduzieren wir das homogene System $\tilde{A}_\lambda$ durch die elementaren Zeilenumformungen

$$\boxed{Z_1 \Leftrightarrow Z_2,} \qquad \boxed{Z_3 - Z_1 \Rightarrow Z_3,} \qquad \boxed{Z_4 - Z_2 \Rightarrow Z_4,}$$

auf eine Stufenform, wobei wir zur Vereinfachung wieder $c_i := -\mu_i$ setzen:

$\tilde{A}_\lambda \vec{v}_2 = \vec{0} \quad \rightleftharpoons$

		c_1	c_2		
0	$\frac{1}{2}$	0	1	$-\frac{1}{2}$	0
$\frac{1}{2}$	0	-1	0	1	0
$\frac{1}{2}$	0	1	0	-1	0
0	$\frac{1}{2}$	0	1	$-\frac{1}{2}$	0
0	0	0	0	0	0

Gauss $\rightleftharpoons$

		c_1	c_2	c_3		$\rho_1\vec{w}_2^{\,1} + \rho_2\vec{w}_2^{\,2}$
$\frac{1}{2}$	0	-1	0	1	0	$-2\rho_1 + \rho_2$
0	$\frac{1}{2}$	0	1	$-\frac{1}{2}$	0	0
0	0	2	$\boxed{0}$	$\boxed{-2}$	0	$2\rho_1 - \rho_2$
0	0	0	0	0	0	0
0	0	0	0	0	0	$\boxed{\rho_2}$

Die Lösungsmenge des Stufensystems vor dem Doppelstrich ist zweidimensional: Die Parameterwahl $(c_2, c_3) := (1,0)$, $:= (0,1)$ führt auf die beiden Vektoren

$$\vec{v}_2^{\,1} := (0,\,-2,\,0,\,1,\,0)^T, \quad \vec{v}_2^{\,2} := (0,\,1,\,1,\,0,\,1)^T.$$

Wir ersetzen in den Zeilenvektoren $(\vec{v}_2^{\,i})^T$ die beiden Spalten S_3, S_4 vorschriftsgemäß durch 0 und erhalten so zwei Hauptvektoren der Stufe 2, nämlich

$$\vec{w}_2^{\,1} := (0,\,-2,\,0,\,0,\,0)^T \qquad \text{zum Parametersatz} \qquad (\mu_{11}^1, \mu_{21}^1) := (0,-1),$$
$$\vec{w}_2^{\,2} := (0,\,1,\,0,\,0,\,1)^T \qquad \text{zum Parametersatz} \qquad (\mu_{12}^1, \mu_{22}^1) := (-1,0).$$

Wir führen in $\rho_1\vec{w}_2^{\,1} + \rho_2\vec{w}_2^{\,2}$ die obigen Zeilenvertauschungen durch und tragen das Resultat in das Stufensystem unter der Spalte "$\rho_1\vec{w}_2^{\,1} + \rho_2\vec{w}_2^{\,2}$" ein. Wir erkennen sofort, dass das inhomogene System $\tilde{A}_\lambda\vec{v}_3 = \rho_1\vec{w}_2^{\,1} + \rho_2\vec{w}_2^{\,2}$ für $\rho_2 \neq 0$ unlösbar ist. Das heißt, $\vec{w}_2^{\,2}$ ist bereits ein Hauptvektor höchster Stufe. Für $\rho_2 = 0$ und bei Wahl von $\rho_1 := 1$ hat das System $\tilde{A}_\lambda\vec{v}_3 = \vec{w}_2^{\,1}$ eine Lösung

$$\vec{v}_3 = (-2,\,0,\,1,\,0,\,0)^T,$$

die aus der Spezifikation $c_2 = c_3 = 0$ resultiert. Wir ersetzen in dem Zeilenvektor $\vec{v}_3^{\,T}$ die Spalten S_3, S_4 vorschriftsgemäß durch 0 und erhalten den Hauptvektor 3. Stufe

$$\vec{w}_3^{\,1} := (-2,\,0,\,0,\,0,\,0)^T \quad \text{zum Parametersatz} \quad (\mu_1^2, \mu_2^2) = (-1,0).$$

Step 3: Hauptvektorkette der Länge 3 ist

$$\vec{u}_1^{\,1} := -\vec{w}_1^{\,2} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad \vec{u}_2^{\,1} := \vec{w}_2^{\,1} - \vec{w}_1^{\,1} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{u}_3^{\,1} := \vec{w}_3^{\,1} = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Eine zweite Hauptvektorkette der Länge 2 erhalten wir gemäß

$$\vec{u}_1^{\,2} := -\vec{w}_1^{\,1} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{u}_2^{\,2} := \vec{w}_2^{\,2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Somit liegt mit den Vektoren $\vec{u}_1^1, \vec{u}_2^1, \vec{u}_3^1, \vec{u}_1^2, \vec{u}_2^2$ bereits eine Basis von Hauptvektorketten des verallgemeinerten Eigenraumes $E(\lambda)$ vor. Wir bilden mit diesen beiden Ketten die Transformationsmatrizen

$$T := (\vec{u}_1^1, \vec{u}_2^1, \vec{u}_3^1, \vec{u}_1^2, \vec{u}_2^2) = \begin{bmatrix} -1 & 0 & -2 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 \\ 0 & -1 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad T^{-1} = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & -2 & 0 \\ 0 & -1 & -1 & 0 & 1 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Die der Matrix A zugeordnete JORDAN–Normalform ergibt sich im Einklang mit (8.10) zu

$$J = T^{-1}AT = \left[\begin{array}{ccc|cc} \boxed{\begin{matrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{matrix}} & & & 0 & 0 \\ & & & 0 & 0 \\ & & & 0 & 0 \\ \hline 0 & 0 & 0 & \boxed{\begin{matrix} 2 & 1 \\ 0 & 2 \end{matrix}} & \\ 0 & 0 & 0 & & \end{array}\right] = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}.$$

Es bedarf sicher keiner großen Einsicht, dass beliebige Kombinationen der beiden vorab diskutierten Sonderfälle mit dem hier vorliegenden allgemeinen Fall jederzeit ohne Einschränkung möglich sind. Generell lässt sich die folgende *Kasten–Regel* über die Struktur der JORDAN–Normalform J einer Matrix $A \in \mathbf{K}^{(n,n)}$ aussprechen:

- *Die Anzahl der* JORDAN*–Kästen in der Matrix J entspricht genau der Gesamtzahl der linear unabhängigen* **Eigenvektoren**, *die die Matrix A besitzt*

- *Gehört zu einem Eigenvektor kein weiterer Hauptvektor, d.h. steht der Eigenvektor am Anfang einer Kette von Hauptvektoren der* **Länge** 1, *so bildet der zugeordnete Eigenwert $\lambda \in \mathbf{C}$ den* JORDAN*–Kasten* $\boxed{\lambda}$

- *Steht der Eigenvektor am Anfang einer Kette von Hauptvektoren der* **Länge** 2, *so bildet der zugeordnete Eigenwert $\lambda \in \mathbf{C}$ den* JORDAN*–Kasten* $\boxed{\begin{matrix} \lambda & 1 \\ 0 & \lambda \end{matrix}}$

- *Steht der Eigenvektor am Anfang einer Kette von Hauptvektoren der* **Länge** k, *so bildet der zugeordnete Eigenwert $\lambda \in \mathbf{C}$ den* JORDAN*–Kasten*

$$\boxed{\begin{matrix} \lambda & 1 & & & 0 & 0 \\ & \lambda & 1 & & & 0 \\ & & \ddots & \ddots & & \\ & & & & \lambda & 1 \\ 0 & & & & & \lambda \end{matrix}} \in \mathbf{C}^{(k,k)}.$$

BSP. (11.8.10) Eine Matrix $A \in \mathbf{K}^{(10,10)}$ habe vier linear unabhängige Eigenvektoren, und zwar zwei Eigenvektoren zum Eigenwert $\lambda_1 := 3$, wobei einer der Eigenvektoren am Anfang einer Kette von Hauptvektoren der Länge 4 stehe und der andere Eigenvektor am Anfang einer Kette von Hauptvektoren der Länge 2. Ein weiterer Eigenwert mit Vielfachheit 1 sei $\lambda_2 := 5$ und ein dritter Eigenwert $\lambda_3 := 2$, dem nun zwangsläufig eine Kette von Hauptvektoren der Länge 3 zugeordnet sein

muss. Eine JORDAN–Normalform der Matrix A ist dann wie folgt gegeben:

$$J = \begin{bmatrix} \boxed{\begin{matrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{matrix}} & & & \\ & \boxed{\begin{matrix} 3 & 1 \\ 0 & 3 \end{matrix}} & & \\ & & \boxed{5} & \\ & & & \boxed{\begin{matrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{matrix}} \end{bmatrix}.$$

O

11.8.2 Anfangswertaufgaben: Der allgemeine Fall revidiert

Wir kommen in diesem Unterabschnitt noch einmal auf die Lösung der Anfangswertaufgabe (7.7) zurück, nämlich auf

$$\boxed{\vec{y}\,'(t) = A\vec{y}(t), \qquad \vec{y}(t_0) = \vec{y}_0 \in \mathbf{K}^n} \tag{7.7}$$

für eine gegebene Matrix $A \in \mathbf{K}^{(n,n)}$ und für einen gegebenen Anfangsvektor $\vec{y}_0 \in \mathbf{K}^n$. Und zwar wollen wir hier den **allgemeinen Fall** eines Eigenwertes $\lambda \in \mathbf{C}$ mit $1 <$ geom. dim $\lambda <$ algebr. dim $\lambda =: k \leq n$ unter Einbezug der Hauptvektorketten studieren. Wir haben in Abschnitt 11.8.1 gesehen, dass dem Eigenwert λ in diesem Fall $s :=$ geom. dim λ Ketten von Hauptvektoren der Matrix A zugeordnet sind:

$$\vec{w}_1^{\,j}, \vec{w}_2^{\,j}, \ldots, \vec{w}_{\kappa_j}^{\,j}, \quad j = 1, 2, \ldots, s; \qquad \kappa_1 + \kappa_2 + \cdots + \kappa_s = k. \tag{8.11}$$

Aus der Definition 11.10 sowie aus (8.8) und (8.9) folgern wir die Relationen

$$A_\lambda^r \vec{w}_i^{\,j} \equiv (A - \lambda\, Id)^r \vec{w}_i^{\,j} = \vec{w}_{i-r}^{\,j}, \quad r = 0, 1, \ldots, i-1; \qquad A_\lambda^i \vec{w}_i^{\,j} = \vec{0}, \quad i = 1, 2, \ldots, \kappa_j.$$

Bringen wir diesen Zusammenhang in die Beziehung (7.9) ein, so erkennen wir die folgende Darstellung der Matrix–Exponentialfunktion:

$$e^{tA}\vec{w}_i^{\,j} = e^{\lambda t} \sum_{r=0}^{i-1} \frac{t^r}{r!}\, \vec{w}_{i-r}^{\,j}, \quad i = 1, 2, \ldots, \kappa_j; \;\; j = 1, 2, \ldots, s. \tag{8.12}$$

Deshalb gilt in Analogie zum Satz 11.27 die folgende Aussage:

Satz 11.32 *Gegeben sei die Matrix $A \in \mathbf{K}^{(n,n)}$. Sei $\lambda \in \mathbf{C}$ ein Eigenwert von A mit $1 < s :=$* geom. dim $\lambda <$ algebr. dim $\lambda =: k$. *Seien*

$$\vec{w}_1^{\,j}, \vec{w}_2^{\,j}, \ldots, \vec{w}_{\kappa_j}^{\,j}, \quad j = 1, 2, \ldots, s; \qquad \kappa_1 + \kappa_2 + \cdots + \kappa_s = k$$

die gemäß Abschnitt 11.8.1 *bestimmten Ketten von Hauptvektoren zum Eigenwert λ. Sind die Zahlen $C_i^j \in \mathbf{C}$ die Koeffizienten des Anfangsvektors $\vec{y}_0 \in \mathbf{K}^n$ in der Teilbasis*

$$\vec{w}_1^{\,1}, \ldots, \vec{w}_{\kappa_1}^{\,1}, \vec{w}_1^{\,2}, \ldots, \vec{w}_{\kappa_2}^{\,2}, \ldots, \vec{w}_1^{\,s}, \ldots, \vec{w}_{\kappa_s}^{\,s}$$

des Vektorraumes $\mathbf{C}^n$*, so ist dem Eigenwert* λ *der folgende Lösungsanteil der Anfangswertaufgabe* (7.7) *eindeutig zugeordnet:*

$$\boxed{\vec{y}_\lambda(t) = e^{\lambda(t-t_0)} \sum_{j=1}^{s} \sum_{i=1}^{\kappa_j} C_i^j \Big(\sum_{r=0}^{i-1} \frac{(t-t_0)^r}{r!} \vec{w}_{i-r}^{\,j} \Big).} \tag{8.13}$$

Begründung: Diese ergibt sich völlig analog zur Begründung des Satzes 11.27.

Bemerkung 11.26 (a) Lässt man die Koeffizienten $C_i^j \in \mathbf{C}$ in (8.13) unbestimmt, und setzt man $t_0 = 0$, so liegt mit $\vec{y}_\lambda(t)$ aus (8.13) der dem Eigenwert λ entsprechende Lösungsanteil in der **allgemeinen Lösung** des Systems $\vec{y}\,'(t) = A\vec{y}(t)$ vor.

(b) Sind $\lambda_1, \lambda_2, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte der Matrix $A \in \mathbf{K}^{(n,n)}$ mit Vielfachheiten $k_1, k_2, \ldots, k_m$, denen gemäß den Sätzen 11.7, 11.27 oder 11.32 Teillösungen $\vec{y}_{\lambda_j}$ der Anfangswertaufgabe (7.7) zugeordnet sind, so erhält man wieder durch **Superposition** dieser Teillösungen die eindeutig bestimmte Gesamtlösung □

$$\boxed{\vec{y}(t) = \vec{y}_{\lambda_1}(t) + \vec{y}_{\lambda_2}(t) + \cdots + \vec{y}_{\lambda_m}(t).}$$

BSP. (11.8.11) Es sei $A \in \mathbf{R}^{(5,5)}$ die Matrix aus BSP. (11.8.9), welche den fünffachen Eigenwert $\lambda = 2$ der geom. dim $\lambda = 2$ besitzt. Diesem Eigenwert sind die beiden folgenden Ketten von Hauptvektoren zugeordnet (vgl. BSP. (11.8.9)):

$$\vec{w}_1^{\,1} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad \vec{w}_2^{\,1} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_3^{\,1} = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \vec{w}_1^{\,2} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_2^{\,2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Das lineare DGl–System $\vec{y}\,'(t) = A\vec{y}(t)$ hat somit gemäß Satz 11.31 die allgemeine Lösung

$$\boxed{\vec{y}(t) = e^{2t}\Big(C_1^1\vec{w}_1^{\,1} + C_2^1(\vec{w}_2^{\,1} + t\vec{w}_1^{\,1}) + C_3^1(\vec{w}_3^{\,1} + t\vec{w}_2^{\,1} + \tfrac{t^2}{2!}\vec{w}_1^{\,1}) + C_1^2\vec{w}_1^{\,2} + C_2^2(\vec{w}_2^{\,2} + t\vec{w}_1^{\,2})\Big).} \tag{8.14}$$

Die Anfangswertaufgabe (7.7) zum Anfangsvektor $\vec{y}_0 = (-3,\, 1,\, -2,\, -1,\, 1)^T$ wird durch (8.14) gelöst, wenn wir t überall durch $t - t_0$ ersetzen und wenn die Konstanten C_i^j das folgende inhomogene lineare Gleichungssystem lösen:

$$C_1^1\vec{w}_1^{\,1} + C_2^1\vec{w}_2^{\,1} + C_3^1\vec{w}_3^{\,1} + C_1^2\vec{w}_1^{\,2} + C_2^2\vec{w}_2^{\,2} = \vec{y}_0.$$

Dieses hat die explizite Form

$$\begin{bmatrix} -1 & 0 & -2 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 \\ 0 & -1 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_1^1 \\ C_2^1 \\ C_3^1 \\ C_1^2 \\ C_2^2 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ -2 \\ -1 \\ 1 \end{bmatrix}$$

mit der eindeutig bestimmten Lösung $C_1^1 = C_2^1 = C_3^1 = C_1^2 = C_2^2 = 1$. Setzen wir zur Abkürzung $p_j(t) := e^{2(t-t_0)}(1 + \frac{t-t_0}{1!} + \frac{(t-t_0)^2}{2!} + \cdots + \frac{(t-t_0)^j}{j!})$, so erhalten wir die gesuchte Lösung der Anfangswertaufgabe (7.7) zum obigen Anfangsvektor $\vec{y}_0$ in der Form

$$\boxed{\vec{y}(t) = p_2(t)\vec{w}_1^{\,1} + p_1(t)(\vec{w}_2^{\,1} + \vec{w}_1^{\,2}) + p_0(t)(\vec{w}_3^{\,1} + \vec{w}_2^{\,2}).}$$

BSP. (11.8.12) Die folgende Matrix $A \in \mathbf{R}^{(5,5)}$ besitzt das charakteristische Polynom $P_5(\lambda) = \det(A - \lambda\, Id) = (2-\lambda)^3(4-\lambda)^2$ und somit die beiden Eigenwerte $\lambda_1 = 2$, $\lambda_2 = 4$ mit algebr. dim $\lambda_1 = 3 =: k_1$ und

algebr. $\dim \lambda_2 = 2 =: k_2$:

$$A := \frac{1}{2}\begin{bmatrix} 4 & 1 & 1 & 0 & -1 \\ 0 & 6 & -2 & 0 & 4 \\ 0 & -2 & 6 & 0 & 0 \\ 0 & 1 & 1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}, \quad P_5(\lambda) = \begin{vmatrix} 2-\lambda & 0.5 & 0.5 & 0 & -0.5 \\ 0 & 3-\lambda & -1 & 0 & 2 \\ 0 & -1 & 3-\lambda & 0 & 0 \\ 0 & 0.5 & 0.5 & 2-\lambda & -0.5 \\ 0 & 0 & 0 & 0 & 4-\lambda \end{vmatrix}.$$

Mit Hilfe der Matrizen $A_{\lambda_j} := (A - \lambda_j\, Id),\ j = 1, 2$ berechnen wir die Eigenräume Kern A_{λ_1} und Kern A_{λ_2} sowie gegebenenfalls zugeordnete Hauptvektorketten. Es gelten zunächst für $\lambda_1 = 2$:

$$A_{\lambda_1} = \frac{1}{2}\begin{bmatrix} 0 & 1 & 1 & 0 & -1 \\ 0 & 2 & -2 & 0 & 4 \\ 0 & -2 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

sowie

$$2A_{\lambda_1}\vec{w}_{11} = \vec{0} \quad \rightleftharpoons \quad \left[\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 2 & -2 & 0 & 4 & 0 \\ 0 & -2 & 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \end{array}\right] \quad \overset{\text{Gauss}}{\rightleftharpoons} \quad \left[\begin{array}{ccccc|c} \multicolumn{1}{c}{c_1} & & & \multicolumn{1}{c}{c_2} & & \\ \boxed{0} & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 4 & \boxed{0} & -2 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Wir erhalten $s_1 := 2 = \text{geom. dim}\,\lambda_1 < 3 = k_1$, d.h. der Eigenraum Kern A_{λ_1} ist zweidimensional. Die Parameterwahl $(c_1, c_2) := (1, 0),\ := (0, 1)$ führt auf die beiden Eigenvektoren

$$\vec{w}_{11}^{\,1} := (1,\, 0,\, 0,\, 0,\, 0)^T, \quad \vec{w}_{11}^{\,2} := (0,\, 0,\, 0,\, 1,\, 0)^T.$$

Kernaustausch: Einen Hauptvektor der Stufe 2 ermitteln wir über die Lösung des homogenen Systems $\tilde{A}_{\lambda_1}\vec{v}_{21} = \vec{0}$, wobei die Matrix $\tilde{A}_{\lambda_1}$ aus A_{λ_1} entsteht, indem wir die Spalten S_1 und S_4 (entsprechend den Spalten unter c_1, c_2) durch die Eigenvektoren $\vec{w}_{11}^{\,1}, \vec{w}_{11}^{\,2}$ ersetzen:

$$\tilde{A}_{\lambda_1}\vec{v}_{21} = \vec{0} \quad \rightleftharpoons \quad \left[\begin{array}{ccccc|c} \multicolumn{1}{c}{c_1} & & & \multicolumn{1}{c}{c_2} & & \\ 1 & \frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & -1 & 0 & 2 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \end{array}\right] \quad \overset{\text{Gauss}}{\rightleftharpoons} \quad \left[\begin{array}{ccccc|c} \multicolumn{1}{c}{c_1} & & & \multicolumn{1}{c}{c_2} & & \\ 1 & \frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Für $c_2 := -1$ resultiert die Lösung

$$\vec{v}_{21} := (-1,\, 1,\, 1,\, -1,\, 0)^T.$$

Im Zeilenvektor $\vec{v}_{21}^{\,T}$ ersetzen wir die Spalten $S_1,\ S_4$ durch 0, und wir erhalten einen Hauptvektor der (höchsten) Stufe 2, nämlich

$$\vec{w}_{21}^{\,1} := (0,\, 1,\, 1,\, 0,\, 0)^T \quad \text{zum Parametersatz} \quad (\mu_1^1, \mu_2^1) := (1, 1).$$

Somit sind dem Eigenwert $\lambda_1 = 2$ die beiden folgenden Hauptvektorketten der Länge 2 bzw. der Länge 1 zugeordnet:

$$\vec{u}_{11}^{\,1} := \vec{w}_{11}^{\,1} + \vec{w}_{11}^{\,2} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{u}_{21}^{\,1} := \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{u}_{11}^{\,2} := \vec{w}_{11}^{\,1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

und der verallgemeinerte Eigenraum $E(\lambda_1)$ wird von $\vec{u}_{11}^{\,1}, \vec{u}_{21}^{\,1}, \vec{u}_{11}^{\,2}$ aufgespannt. Nun verfahren wir mit dem Eigenwert $\lambda_2 = 4$ ganz analog. Es gelten

$$A_{\lambda_2} = \frac{1}{2}\begin{bmatrix} -4 & 1 & 1 & 0 & -1 \\ 0 & -2 & -2 & 0 & 4 \\ 0 & -2 & -2 & 0 & 0 \\ 0 & 1 & 1 & -4 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

sowie

$$2A_{\lambda_2}\vec{w}_{12}=\vec{0}\ \rightleftharpoons\ \left|\begin{array}{ccccc|c} -4 & 1 & 1 & 0 & -1 & 0\\ 0 & -2 & -2 & 0 & 4 & 0\\ 0 & -2 & -2 & 0 & 0 & 0\\ 0 & 1 & 1 & -4 & -1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right| \overset{\text{Gauss}}{\rightleftharpoons} \left|\begin{array}{ccccc|c} \underline{-4} & 1 & 1 & 0 & -1 & 0\\ 0 & \underline{-2} & \boxed{-2} & 0 & 0 & 0\\ 0 & 0 & 0 & \underline{-4} & -1 & 0\\ 0 & 0 & 0 & 0 & \underline{4} & 0\\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right| \rightleftharpoons\ \vec{w}_{12}^{\,1}=\begin{bmatrix} 0\\1\\-1\\0\\0 \end{bmatrix}.$$

Hier liegt der Sonderfall $s_2 := 1 = \text{geom. dim}\,\lambda_2 < \text{algebr. dim}\,\lambda_2 = 2$ vor. Die Berechnung eines Hauptvektors der Stufe 2 erfordert keine Lösbarkeitsbedingungen:

$$2A_{\lambda_2}\vec{w}_{22}^{\,1}=2\vec{w}_{12}^{\,1}\ \rightleftharpoons\ \left|\begin{array}{ccccc|c} -4 & 1 & 1 & 0 & -1 & 0\\ 0 & -2 & -2 & 0 & 4 & 2\\ 0 & -2 & -2 & 0 & 0 & -2\\ 0 & 1 & 1 & -4 & -1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right| \overset{\text{Gauss}}{\rightleftharpoons} \left|\begin{array}{ccccc|c} \underline{-4} & 1 & 1 & 0 & -1 & 0\\ 0 & \underline{-2} & \boxed{-2} & 0 & 0 & -2\\ 0 & 0 & 0 & \underline{-4} & -1 & -1\\ 0 & 0 & 0 & 0 & \underline{4} & 4\\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right| \rightleftharpoons\ \vec{w}_{22}^{\,1}=\begin{bmatrix} 0\\1\\0\\0\\1 \end{bmatrix}.$$

Dem Eigenwert $\lambda_2 = 4$ ist also genau die folgende Kette von Hauptvektoren zugeordnet:

$$\vec{w}_{12}^{\,1}=(0,\,1,\,-1,\,0,\,0)^T,\quad \vec{w}_{22}^{\,1}=(0,\,1,\,0,\,0,\,1)^T.$$

Das lineare DGl–System $\vec{y}\,'(t)=A\vec{y}(t)$ hat nun gemäß Satz 11.32 die folgende allgemeine Lösung

$$\boxed{\vec{y}(t)=e^{2t}\Big(C_1^1\vec{u}_{11}^{\,1}+C_2^1(\vec{u}_{21}^{\,1}+t\vec{u}_{11}^{\,1})+C_3^2\vec{w}_{11}^{\,2}\Big)+e^{4t}\Big(D_1^1\vec{w}_{12}^{\,1}+D_2^1(\vec{w}_{22}^{\,1}+t\vec{w}_{12}^{\,1})\Big).} \tag{8.15}$$

Index