

W. Oevel

Mathematik für Physiker I

Veranstaltungsnr: 172020

Skript zur Vorlesung, Universität Paderborn, Wintersemester 2003/2004

Zeit und Ort:	V2	Di	11.15 – 12.45	D1.303	
	V2	Mi	11.15 – 12.45	D1.303	
	V2	Do	9.15 – 10.45	D1.328	
	Ü2	Do	11 – 13	D1.328	(Kai Gehrs)

Zur Homepage der Veranstaltung:
math-www.uni-paderborn.de/~walter (→ Lehrveranstaltungen)

Email: walter@upb.de

Inhalt

1	Sammelsurium	3
1.1	Notation, logische Aussagen	3
1.2	Die natürlichen Zahlen, vollständige Induktion	6
1.3	Die rationalen Zahlen	7
1.4	Die reellen Zahlen	9
1.5	Die komplexen Zahlen	16
1.5.1	Definitionen	16
1.5.2	Polynomwurzeln, Fundamentalsatz der Algebra	21
2	Folgen und Grenzwerte	29
2.1	Definitionen, Beispiele, einige Sätze	29
2.2	Weitere Konvergenzsätze	40
2.2.1	Konvergenz monotoner reeller Folgen	40
2.2.2	Cauchy-Folgen	41
2.2.3	Teilfolgen und Häufungspunkte	44
2.3	Unendliches, uneigentliche Konvergenz	46
3	Reihen	49
3.1	Definitionen, Beispiele, Sätze	49
3.2	Rechenregeln und das Cauchy-Produkt	54
3.3	Spezielle Konvergenzkriterien	56
3.4	Bedingte Konvergenz, Umordnungen	61
3.5	Summation per Partialbruchzerlegung	63
4	Funktionen und Stetigkeit	67
4.1	Funktionen	67
4.2	Stetigkeit	68
4.3	Grenzwerte	73
4.4	Der Zwischenwertsatz, das Min/Max-Prinzip	75
4.5	Umkehrfunktionen	78
4.6	Wachstum von Funktionen, Landau-Symbole	81

5	Spezielle Funktionen	83
5.1	Exponentialfunktion und Logarithmus	83
5.2	Trigonometrische Funktionen	86
5.3	Die komplexe Exponentialfunktion	91
5.4	Die Inversen der trigonometrische Funktionen	94
5.5	Diverse spezielle Funktionen: Gamma, Bessel etc.	95
6	Differentialrechnung	99
6.1	Definitionen und Sätze	99
6.2	Der Mittelwertsatz	107
6.3	Taylor-Reihen	108
6.4	Monotonie, Extremwerte	115
6.5	Die de l'Hospitalsche Regel	118
7	Potenzreihen	121
7.1	Der Konvergenzradius	121
7.2	Eigenschaften von Potenzreihen	124
8	Banach-Iteration	129
8.1	Der Banachsche Fixpunktsatz	129
8.2	Das Newton-Verfahren	132
9	Integration	137
9.1	Stammfunktionen: das unbestimmte Integral	137
9.1.1	Definitionen, Grundintegrale	137
9.1.2	Partielle Integration	139
9.1.3	Substitution	141
9.1.4	Rationale Integranden: Partialbruchzerlegung	143
9.2	Das bestimmte Integral	146
9.3	Der Hauptsatz	155
9.4	Uneigentliche Integrale	158
9.5	Die Diracsche Delta-Funktion	160
9.6	Einige spezielle Anwendungen	162
10	Gewöhnliche Differentialgleichungen	167
10.1	DGLen 1-ter Ordnung	167
10.1.1	Definitionen	167
10.1.2	Graphische Lösung	169
10.1.3	Separation (Trennung der Variablen)	171
10.1.4	Variation der Konstanten	175
10.2	Lineare DGLen höherer Ordnung	179
10.3	Lineare DGLen mit konstanten Koeffizienten	184

11 Fourier-Entwicklungen	191
11.1 Hilbert-Raumstrukturen	191
11.1.1 Definitionen: Skalarprodukt und Norm	191
11.1.2 Bestapproximation und Konvergenz	196
11.2 Trigonometrische Fourier-Reihen	200
11.2.1 Definition der trigonometrischen Fourier-Reihen	201
11.2.2 Konvergenz im quadratischen Mittel	204
11.2.3 Punktweise Konvergenz	204
11.2.4 Abklingverhalten der Fourier-Koeffizienten	214
11.2.5 Das Gibbs'sche Phänomen	217
11.2.6 Beliebige Perioden	219
12 Lineare Algebra	223
12.1 Vektorräume	223
12.2 Etwas Geometrie im \mathbb{R}^3	230
12.2.1 Skalar-, Vektor- und Spatprodukt	230
12.2.2 Geraden und Ebenen	235
12.3 Matrizen	236
12.4 Determinanten	241
12.5 Lineare Gleichungssysteme	254
12.5.1 Quadratische Dreieckssysteme	255
12.5.2 Der Gauß-Algorithmus	257
12.5.3 Die allgemeine Lösung des Dreieckssystems	263
12.6 Inverse von Matrizen	270
12.6.1 Die Gauß-Jordan-Variante	270
12.6.2 LR -Zerlegung	271
12.7 Eigenwerte und -vektoren	276
12.7.1 Definitionen	276
12.7.2 Symmetrische/hermitesche Matrizen	283
12.7.3 Diagonalisierung und Funktionalkalkül	286

Literatur

Die Vorlesung baut nicht streng auf irgendeinem Buch auf, sondern geht ihren eigenen Weg. Die angegebenen Referenzen dienen dazu, sich *unabhängig* vom Skript entsprechende Grundlagen anzueignen oder spezielle Inhalte zu vertiefen. Es handelt sich um eine recht willkürliche Auswahl: Neben den angegebenen Büchern gibt es sicherlich jede Menge weiterer Literatur, die den behandelten Stoff analog abdeckt.

Die Jänich-Bücher sind allgemein recht elementar geschrieben und leicht und angenehm zu lesen. Empfehlenswert. Für den Stoff dieses Semesters sind 1), 2), 3) besonders interessant.

- 1) K. JÄNICH : *Mathematik 1* („geschrieben für Physiker“), Springer, 2001.
Für uns relevanter Inhalt: Kapitel 1 - 5: Analysis in \mathbb{R} ; Kapitel 13: Fourier-Reihen; Kapitel 8 - 9: Matrizen; Kapitel 11: Determinanten; Kapitel 12: Skalar- und Kreuzprodukt; Kapitel 21: Eigenwerte- und -Vektoren.
Für später relevanter Inhalt: Analysis im \mathbb{R}^n : diverse Kapitel
- 2) K. JÄNICH : *Mathematik 2* („geschrieben für Physiker“), Springer, 2002.
Für uns relevanter Inhalt: Kapitel 23: Grundlagen der Analysis; Kapitel 24: Funktionenfolgen und Reihen; Kapitel 25: Taylorentwicklung;
Für später relevanter Inhalt: Analysis im \mathbb{R}^n : diverse Kapitel
- 3) K. JÄNICH : *Lineare Algebra*, Springer, 2002.
Alles über Lineare Algebra und Matrizen, was man als Physiker wissen sollte. Sehr zu empfehlen.
- 4) K. JÄNICH : *Analysis für Physiker und Ingenieure*, Springer, 2001. Funktionentheorie, Differentialgleichungen und spezielle Funktionen.
- 5) K. JÄNICH: *Funktionentheorie – Eine Einführung*, Springer, 1999.

Hinweis für später: zum Erlernen der Theorie komplexer Funktionen ist der Funktionentheorie-Band 5) absolut empfehlenswert.

- [Pap] LOTHAR PAPULA: Mathematik für Ingenieure und Naturwissenschaftler. Band 1 - 3 + Mathematische Formelsammlung. Braunschweig/Wiesbaden: Vieweg 2001. (P41 TBG2788)

Recht elementar und mathematisch nicht sehr tief gehend; dafür leicht und angenehm zu lesen. Dies sind Standardbücher und große Renner bei den Ingenieuren. Hier steht „für Ingenieure und Naturwissenschaftler“ drauf: diese Reihe ist allgemein für eine anwendungsorientierte Kundschaft sehr geeignet, die sich weniger für das Abstrakte in der Mathematik interessiert. Übungen und Anwendungsbeispiele sind speziell auf Ingenieure zugeschnitten.

Hier ist zunächst **Band 1** interessant: er umfasst Folgen, Reihen, Stetigkeit, Differentialrechnung, spezielle Funktionen und Integration.

Band 2 umfasst die Lineare Algebra und komplexe Zahlen. Weiterhin viele weitere Dinge, die im ersten Semester noch nicht so interessant sind.

Band 3 umfasst mehrdimensionale Differential- und Integralrechnung (interessant für das 2-te Semester). Darüberhinaus Stochastik (für uns nicht ganz so interessant).

- [TI] S. TIMMAN: Repetitorium der Analysis Springe: Binomi-Verlag.

Eigentlich kein 'Repetitorium', sondern eine vollständige Einführung mit Definitionen etc. Recht elementar geschrieben, sehr übersichtlich. Gelungener Kompromiss zwischen mathematischem Tiefgang und guter Lesbarkeit auf für Nicht-Mathematiker. Grundlagen, Folgen und Reihen, Stetigkeit, Differential- und Integralrechnung. Zahlreiche Übungsaufgaben mit Musterlösungen.

- [Bla] C. BLATTER: Analysis I, II, III Berlin: Springer 1991.

Abstrakter und anspruchsvoller. Recht kompakt.

Band 1 umfasst Folgen, Reihen, Stetigkeit, Differential- und Integralrechnung.

Band 2 enthält mehr zur Integralrechnung, Kurven, Funktionenfolgen, Potenzreihen und Differentialrechnung im \mathbb{R}^n (relevant für das 2-te Semester).

Band 3 behandelt die Integralsätze im \mathbb{R}^n (relevant für das 2-te Semester).

- [For] O. FORSTER: Analysis 1 - 3. Vieweg. 2001. (P41 TIA 2647)

Abstrakt und anspruchsvoller. Recht kompakt. Standardwerk für Mathematikstudenten. **Band 1** umfasst Folgen, Reihen, Stetigkeit, Differential- und Integralrechnung einer Veränderlichen (Stoff des 1-ten Semesters). Die **Bände 2** und **3** umfassen die mehrdimensionale Analysis und sind ab dem 2-ten Semester relevant.

Kapitel 1

Sammelsurium (Notation, logische Aussagen, Zahlenräume etc.)

Dieses Kapitel dient im Wesentlichen dazu, eine einheitliche Notation und Sprechweise für die späteren Kapitel zu schaffen. Die Inhalte (nicht unbedingt die Formalisierungen) sollten aus der Schule bekannt sein. Eventuell sind die komplexen Zahl neu.

↓14.10.03

1.1 Notation, logische Aussagen

Notation 1.1:

Folgende Standardbezeichnungen und Symbole sollten aus der Schule bekannt sein und werden auch hier verwendet werden:

- **Mengen:** $\{1, 2, 3, x, y, z\}$, $\{x \in \mathbb{R}; x^2 + 25 \cdot x > 29\}$,
- **die leere Meng:** \emptyset oder auch $\{\}$,
- **Elemente von Mengen:** $x \in A$ bedeutet „ x ist aus der Menge A “,
- **Vereinigung:** $A \cup B = \{x; x \in A \text{ oder } x \in B\}$,
- **Durchschnitt:** $A \cap B = \{x; x \in A \text{ und } x \in B\}$,
- **Teilmengen:** $A \subset B$ heißt: alle Elemente von A sind auch in B enthalten,
- **Obermengen:** $B \supset A$ ist äquivalent zu $A \subset B$,
- **Differenzmengen:** $A \setminus B = \{x \in A; x \notin B\}$,
- $\mathbb{N} =$ die Menge der natürlichen Zahlen $\{1, 2, 3, \dots\}$,

- $\mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, 2, 3, \dots\}$,
- $\mathbb{Z} =$ die Menge der ganzen Zahlen $= \{\dots, -2, -1, 0, 1, 2, \dots\}$,
- $\mathbb{Q} =$ die Menge der rationalen Zahlen $= \{\frac{z}{n}; z \in \mathbb{Z}; n \in \mathbb{N}\}$,
- $\mathbb{R} =$ die Menge der reellen Zahlen,

Intervalle:

$$(a, b) = \{x \in \mathbb{R}; a < x < b\}, \text{ „offene“ Intervalle,}$$

$$[a, b] = \{x \in \mathbb{R}; a \leq x \leq b\}, \text{ „geschlossene“ Intervalle,}$$

$$(a, b] = \{x \in \mathbb{R}; a < x \leq b\}, \text{ „halboffene“ Intervalle,}$$

$$[a, b) = \{x \in \mathbb{R}; a \leq x < b\}, \text{ „halboffene“ Intervalle,}$$

- $\mathbb{C} =$ die Menge der komplexen Zahlen (siehe Abschnitt 1.5),
- **kartesisches Produkt:** $A \times B = \{(a, b); a \in A, b \in B\}$,
- „Unendlich“: $\pm\infty$,
- **Wurzel:** $x = \sqrt{y}$ mit $y \in [0, \infty)$ ist die **positive** Lösung von $x^2 = y$.

Elementare Rechenregeln 1.2:

Einige aus der Schule bekannte Rechenregeln für's Potenzieren:

- $x^0 = 1$,
- $x^{-\alpha} = \frac{1}{x^\alpha}$,
- $x^{\alpha+\beta} = x^\alpha \cdot x^\beta$,
- $(x^\alpha)^\beta = x^{\alpha \cdot \beta}$.

Eine logische Aussage ist eine Aussage, der man genau einen der Wahrheitswerte „wahr“ oder „falsch“ zuordnen kann. Beispiele:

(A1) Die Zahl 5 ist eine Primzahl.

(A2) Es gibt eine größte Primzahl.

(A3) Die Zahl $2^{(2^{(2^{1234567})})} + 123$ ist eine Primzahl.

Hierbei ist A1 wahr und A2 falsch, während der Wahrheitsgehalt von A3 unklar ist (aber entweder 'wahr' oder 'falsch' sein muss).

Logische Aussagen werden über die Symbole

$$\begin{aligned}
A_1 \Rightarrow A_2 &\equiv \text{„aus } A_1 \text{ folgt } A_2\text{“} \\
&\equiv \text{„}A_2 \text{ ist wahr, wenn } A_1 \text{ wahr ist“} \\
&\equiv \text{„}A_1 \text{ ist falsch, wenn } A_2 \text{ falsch ist“}, \\
A_1 \Leftrightarrow A_2 &\equiv \text{„}A_2 \text{ ist genau dann wahr, wenn } A_1 \text{ wahr ist“} \\
&\equiv \text{„}A_2 \text{ ist genau dann falsch, wenn } A_1 \text{ falsch ist“} \\
&\equiv \text{„}A_1 \text{ und } A_2 \text{ sind äquivalent“}
\end{aligned}$$

zu neuen logischen Aussagen miteinander verknüpft. Beispiele:

- (a) Ist n eine Primzahl, so ist $n + 1$ keine Primzahl.
- (b) Ist n eine ungerade Primzahl, so ist $n + 1$ keine Primzahl.
- (c) $x \in \mathbb{R}$ ist genau dann positiv, wenn $(x + 1)^2 > (x - 1)^2$ gilt.

Mit

$$\begin{aligned}
a_1: & n \text{ ist eine Primzahl,} \\
a_2: & n + 1 \text{ ist keine Primzahl,} \\
a_3: & n \text{ ist eine ungerade Primzahl,} \\
a_4: & x \in \mathbb{R} \text{ ist positiv,} \\
a_5: & \text{für } x \text{ gilt } (x + 1)^2 > (x - 1)^2
\end{aligned}$$

werden die obigen Aussagen kürzer als

$$(a) : a_1 \Rightarrow a_2, \quad (b) : a_3 \Rightarrow a_2, \quad (c) : a_4 \Leftrightarrow a_5$$

formuliert. Hierbei ist (a) nicht für jedes n wahr (die Ausnahme ist $n = 2$), während b) und c) für jedes n bzw. x wahr sind.

Typische Aussagen werden mit Hilfe „logischer Quantoren“ formuliert:

$$\begin{aligned}
\forall & \text{ „für alle“,} \\
\exists & \text{ „es gibt (mindestens) ein“,} \\
\exists! & \text{ „es gibt genau ein“,} \\
\nexists & \text{ „es gibt kein“.}
\end{aligned}$$

Beispiele:

$$(A) \forall x \in \mathbb{R} \exists y \in \mathbb{R} : x = y^2.$$

Klartext: für jedes $x \in \mathbb{R}$ gibt es (mindestens) ein $y \in \mathbb{R}$ mit $x = y^2$.

$$(B) \forall x \in [0, \infty) \exists y \in \mathbb{R} : x = y^2.$$

Klartext: Für jedes $x \in [0, \infty)$ gibt es (mindestens) ein $y \in \mathbb{R}$ mit $x = y^2$.

$$(C) \forall x \in [0, \infty) \exists! y \in \mathbb{R} : x = y^2.$$

Klartext: Für jedes $x \in [0, \infty)$ gibt es genau ein $y \in \mathbb{R}$ mit $x = y^2$.

(D) $\forall x \in (-\infty, 0) \exists y \in \mathbb{R} : x = y^2$.

Klartext: Für jedes $x \in (-\infty, 0)$ gilt: es gibt kein reelles y mit $x = y^2$. In besserem Deutsch: Für kein $x \in (-\infty, 0)$ gibt es ein reelles y mit $x = y^2$.

Hierbei ist (A) falsch, (B) richtig, (C) falsch und (D) richtig. Man sollte diese Schreibweisen kennen, um Aussagen in Mathe-Büchern lesen zu können. Ich werde hier versuchen, die Quantorenschreibweise weitmöglichst zu vermeiden und die Aussagen im Klartext ausformulieren.

1.2 Die natürlichen Zahlen, vollständige Induktion

15.10.03↓

Die Menge der natürlichen Zahlen $\mathbb{N} = \{1, 2, 3, \dots\}$ bzw. $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ ist aus der Schule wohlvertraut. Man kann \mathbb{N} mathematisch durch die folgenden Axiome (die „Peano-Axiome“) festlegen:

Axiomatisierung von \mathbb{N} 1.3:

(P1) Es gibt ein ausgezeichnetes („kleinstes“) Element $1 \in \mathbb{N}$.

(P2) Es gibt eine „Nachfolgerfunktion“ $\nu : \mathbb{N} \rightarrow \mathbb{N} \setminus \{1\}$ mit den Eigenschaften:

(P2a) Für jedes Paar $n_1, n_2 \in \mathbb{N}$ gilt: $n_1 \neq n_2 \Rightarrow \nu(n_1) \neq \nu(n_2)$.

(P2b) Besitzt eine Teilmenge $A \subset \mathbb{N}$ die Eigenschaften

1) $1 \in A$,

2) $n \in A \Rightarrow \nu(n) \in A$,

so folgt $A = \mathbb{N}$.

Die durch diese Eigenschaften bestimmte Menge $\mathbb{N} = \{1, \nu(1), \nu(\nu(1)), \dots\}$ ist uns als $\mathbb{N} = \{1, 2, 3, \dots\}$ vertraut. Wir benutzen dabei die übliche Bezeichnung

$$n = \underbrace{\nu(\dots(\nu(1))\dots)}_{n-1}.$$

Die Nachfolgerfunktion $\nu(n) = n + 1$ ist eine schrecklich umständliche Notation und wird durch die übliche Addition ausgedrückt. Das Axiom (P2b) ist das „Prinzip der vollständigen Induktion“:

Sei $A(n)$ eine Aussage über eine natürliche Zahl n . Können wir $A(1)$ beweisen und zeigen, dass aus der Aussage $A(n)$ die Aussage $A(n+1)$ folgt, so folgt aus (P2), dass die Menge $A := \{n \in \mathbb{N}; A(n) \text{ ist richtig}\}$ ganz \mathbb{N} ist, also $A(n)$ für alle n in \mathbb{N} gilt.

Beispiel 1.4: Behauptung:

$$A(n) : \quad \sum_{i=1}^n i = \frac{n \cdot (n+1)}{2}.$$

Beweis durch Induktion nach n :

Induktionsstart: $A(1)$ ist die Aussage $\sum_{i=1}^1 i = \frac{1 \cdot (1+1)}{2}$. Diese Aussage $A(1) : 1 = 1$ ist trivialerweise richtig.

Schritt $n \rightarrow n+1$: Für irgendein gegebenes n gelte $A(n)$, also $\sum_{i=1}^n i = n \cdot (n+1)/2$. Wir folgern hieraus $A(n+1)$:

$$\begin{aligned} \sum_{i=1}^{n+1} i &= \left(\sum_{i=1}^n i \right) + (n+1) \stackrel{A(n)}{=} \frac{n \cdot (n+1)}{2} + n+1 \\ &= \frac{n^2 + n + 2 \cdot n + 2}{2} = \frac{(n+1) \cdot (n+2)}{2}. \end{aligned}$$

Unter Benutzung der Gültigkeit von $A(n)$ haben wir die Gültigkeit von $A(n+1)$ bewiesen. Das Prinzip der vollständigen Induktion sagt uns, dass $A(n)$ für alle $n \in \mathbb{N}$ richtig ist.

1.3 Die rationalen Zahlen

Wir haben \mathbb{N} über die Peano-Axiome mathematisch sauber definiert. Angenommen, wir hätten die ganzen Zahlen \mathbb{Z} (aus vielen Schuljahren zumindestens intuitiv bekannt) mathematisch sauber definiert (was wir hier nicht haben), so könnten wir die Menge der rationalen Zahlen als die Menge aller Brüche

$$\mathbb{Q} = \left\{ \frac{z}{n}; z \in \mathbb{Z}; n \in \mathbb{N} \right\}$$

definieren. Eine rationale Zahl wäre somit so etwas wie das Paar

$$\frac{z}{q} \equiv (\text{Zähler } z, \text{ Nenner } n).$$

Hier gibt es ein formales Problem, da z.B. $1/2 \equiv (1, 2)$ und $2/4 \equiv (2, 4)$ übereinstimmen sollen. Man kann das „Kürzen“ von Brüchen mathematisch sauber fassen, indem man „Äquivalenzklassen“ von Paaren einführt:

$$(z, n) \sim (\tilde{z}, \tilde{n}), \text{ wenn } z \cdot \tilde{n} = \tilde{z} \cdot n \text{ gilt.}$$

Hier gehen Physikstudent und Mathematikstudentin nun zunächst eine Weile getrennte Wege (können später aber wieder zueinander finden). Wir akzeptieren

als Physiker den aus der Schule bekannten „intuitiven“ Umgang mit rationalen Zahlen über die vertrauten Rechenregeln wie z.B.

$$\frac{z_1}{n_1} + \frac{z_2}{n_2} = \frac{z_1 \cdot n_2 + n_1 \cdot z_2}{n_1 \cdot n_2}, \quad \frac{z_1}{n_1} \cdot \frac{z_2}{n_2} = \frac{(z_1 \cdot z_2)}{(n_1 \cdot n_2)}.$$

Die Mathematikerin würde formal eine Addition und eine Multiplikation auf den Äquivalenzklassen der Paare (Zähler, Nenner) definieren und dann zeigen, dass diese Definitionen sich als die obige Addition/Multiplikation von Brüchen interpretieren lassen. Sind auf diese Weise alle aus der Schule vertrauten Rechenregeln formal bewiesen, trifft man sich wieder und entwickelt die Theorie weiter. Der Physiker-Weg des „es wird schon alles zusammenpassen“ ist dabei deutlich kürzer.

Satz 1.5: (Abzählbarkeit von \mathbb{Q})

Die Menge \mathbb{Q} der rationalen Zahlen ist abzählbar, d.h., es gibt eine invertierbare Abbildung von \mathbb{N} nach \mathbb{Q} .

Beweis: Wir betrachten zunächst die positiven rationalen Zahlen

$$\mathbb{Q}_+ = \left\{ \frac{z}{n}; z \in \mathbb{N}, n \in \mathbb{N} \right\}.$$

Jedes $q \in \mathbb{Q}_+$ findet sich nach Kürzen von Nenner und Zähler an einem eindeutig bestimmten Punkt des Schemas

$n \setminus z$	1	2	3	...
1	*	*	*	...
2	*	*	*	...
3	*	*	*	...
\vdots	\vdots

Dieses Schema wird links oben startend wie folgt durchlaufen:

$n \setminus z$	1	2	3	...
1	→	↘	→	↘
2	↓	↗	↘	
3	↗	↘		
\vdots	↓	↗		
\vdots	↗			

Überspringt man dabei die Stellen, die nach Kürzen von Zähler und Nenner einer früher schon gefundenen rationalen Zahlen entsprechen, hat man die invertierbare Abbildung $\mathbb{N} \rightarrow \mathbb{Q}_+$ explizit konstruiert (wenngleich ich mich scheue,

die Abbildungsvorschrift explizit hinschreiben). Analog ist die Menge \mathbb{Q}_- der negativen rationalen Zahlen abzählbar (denn \mathbb{Q}_+ und \mathbb{Q}_- lassen sich durch $q \in \mathbb{Q}_+ \rightarrow -q \in \mathbb{Q}_-$ ein-eindeutig aufeinander abbilden). Die Vereinigung $\mathbb{Q}_+ \cup \mathbb{Q}_-$ abzählbarer Mengen ist offensichtlich wieder abzählbar (Aufgabe 5, Blatt 1) und damit auch $\mathbb{Q} = \{0\} \cup \mathbb{Q}_+ \cup \mathbb{Q}_-$.

Q.E.D.

1.4 Die reellen Zahlen

Die reellen Zahlen und die komplexen Zahlen des nächsten Abschnitts sind die Zahlen, mit denen es der Physiker meistens zu tun hat.

Axiomatisierung von \mathbb{R} 1.6:

Die Menge \mathbb{R} der reellen Zahlen mit der Addition

$$+ : (x, y) \in \mathbb{R} \times \mathbb{R} \rightarrow x + y \in \mathbb{R}$$

und der Multiplikation

$$\cdot : (x, y) \in \mathbb{R} \times \mathbb{R} \rightarrow x \cdot y \in \mathbb{R}$$

hat folgende Eigenschaften:

(R0) Addition und Multiplikation sind **kommutativ**, **assoziativ** und **distributiv**, d.h., für alle $x, y, z \in \mathbb{R}$ gilt:

$$\begin{aligned} x + y &= y + x, & (x + y) + z &= x + (y + z), \\ x \cdot y &= y \cdot x, & (x \cdot y) \cdot z &= x \cdot (y \cdot z), \\ x \cdot (y + z) &= x \cdot y + x \cdot z. \end{aligned}$$

(R1) Es gibt zwei ausgezeichnete Elemente 0 und 1 in \mathbb{R} , für die $x + 0 = x$ und $1 \cdot x = x$ für alle $x \in \mathbb{R}$ gilt. Bezeichnung: 0 ist das „**neutrale Element der Addition**“, 1 ist das „**neutrale Element der Multiplikation**“.

(R2) (Inverse bzgl. der Addition) Zu jedem $x \in \mathbb{R}$ gibt es genau ein Element \tilde{x} mit $x + \tilde{x} = 0$. Bezeichnung: $\tilde{x} = -x$.

(R3) (Inverse bzgl. der Multiplikation) Zu jedem $x \in \mathbb{R} \setminus \{0\}$ gibt es genau ein Element $\hat{x} \in \mathbb{R} \setminus \{0\}$ mit $x \cdot \hat{x} = 1$. Bezeichnung: $\hat{x} = 1/x = x^{-1}$.

(R4) (Anordnung) Es gibt eine Teilmenge $\mathbb{R}_+ \subset \mathbb{R}$ (die positiven reellen Zahlen), für die gilt:

$$0 \neq x \in \mathbb{R}_+ \Rightarrow -x \notin \mathbb{R}_+,$$

$$0 \neq x \notin \mathbb{R}_+ \Rightarrow -x \in \mathbb{R}_+, \\ x, y \in \mathbb{R}_+ \Rightarrow x + y \in \mathbb{R}_+, x \cdot y \in \mathbb{R}_+.$$

Bezeichnung: wir schreiben $x < y$, wenn $y - x \in \mathbb{R}_+$ gilt. Wir schreiben $x \leq y$, wenn $x < y$ oder $x = y$ gilt.

Definition: Eine Zahl $\beta \in \mathbb{R}$ heißt **obere Schranke** einer Teilmenge $A \subset \mathbb{R}$, wenn $a \leq \beta$ gilt für alle $a \in A$. Eine Teilmenge $A \subset \mathbb{R}$ heißt **nach oben beschränkt**, wenn es (mindestens) eine obere Schranke von A gibt.

(R5) (Supremumsaxiom) Jede nach oben beschränkte nicht-leere Teilmenge $A \subset \mathbb{R}$ besitzt eine kleinste obere Schranke in \mathbb{R} , die als das „**Supremum von A** “

$$\sup A = \min\{\beta \in \mathbb{R}; a \leq \beta \forall a \in A\} \in \mathbb{R}.$$

bezeichnet wird.

Wir sind aus der Schule bereits an viele weitere Vereinbarungen und Notationen gewöhnt. Beispielsweise schreiben wir einfach

$$x + y + z, x \cdot y \cdot z,$$

denn die Assoziativität garantiert, dass eine Klammerung wie z.B. $(x + y) + z$ bzw. $x + (y + z)$ nicht nötig ist. Wir benutzen die abkürzenden Notationen

$$x - y := x \cdot (-y), \quad \frac{x}{y} := x \cdot y^{-1}$$

oder benutzen auch

$$y \geq x \quad \text{statt} \quad x \leq y$$

usw. Die aus der Schule vertrauten Rechenregeln lassen sich durch mühselige Kleinarbeit alle aus den Axiomen (R0) – (R5) herleiten („beweisen“). Hier nur ein paar kleine Beispiele.

Behauptung: Man darf Ungleichungen addieren.

Beweis: Für $i = 1, 2$ gilt: $x_i < y_i$ bedeutet $y_i - x_i \in \mathbb{R}_+$. Nach den Axiomen folgt $y_1 - x_1 + y_2 - x_2 \in \mathbb{R}_+$, was als $x_1 + x_2 < y_1 + y_2$ geschrieben wird, also:

$$x_1 < y_1, x_2 < y_2 \quad \Rightarrow \quad x_1 + x_2 < y_1 + y_2.$$

Q.E.D.

Behauptung: $0 \cdot x = 0 \quad \forall x \in \mathbb{R}$.

Beweis: Aus den Axiomen folgt:

$$0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x.$$

Addiert man das nach den Axiomen existierende $-(0 \cdot x)$ zu dieser Gleichung, erhält man $0 = 0 \cdot x$.

Q.E.D.

Behauptung: $(-1) \cdot x = -x \quad \forall x \in \mathbb{R}$.

Beweis: Mit dem Distributivgesetz:

$$x + (-1) \cdot x = x \cdot (1 + (-1)) = x \cdot 0 = 0.$$

Also ist $(-1) \cdot x$ das Inverse von x bzgl. der Addition, welches als $-x$ bezeichnet wird.

Q.E.D.

Aus der Schule ist folgende "Rechenregel" bekannt:

„Bei Multiplikation einer Ungleichung mit einer negativen Zahl dreht sich die Ungleichung um“: $x < 0, y < z \Rightarrow x \cdot y > x \cdot z$.

Beweis: $x < 0, y < z$ bedeuten $-x \in \mathbb{R}_+, z - y \in \mathbb{R}_+$. Aufgrund der Axiome für \mathbb{R}_+ folgt $(-x) \cdot (z - y) \in \mathbb{R}_+$, also $x \cdot y - x \cdot z \in \mathbb{R}_+$, was wiederum als $x \cdot z < x \cdot y$ geschrieben wird.

Q.E.D.

↓16.10.03

Körper-Axiome 1.7:

*Eine Menge \mathbb{R} mit einer Addition und einer Multiplikation, die die Eigenschaften (R0) – (R3) erfüllt, heißt **Körper**. Ein Körper, der zusätzlich die Eigenschaft (R4) hat, heißt **angeordneter Körper**.*

Die Menge der rationalen Zahlen \mathbb{Q} erfüllt offensichtlich die Eigenschaften (R0) – (R4), also: auch \mathbb{Q} ist ein angeordneter Körper. Was also ist der Unterschied zwischen \mathbb{Q} und \mathbb{R} ? Es ist das Supremumsaxiom (R5)! Historisch gesehen geht es um die Tatsache, dass es Zahlen wie $\sqrt{2} \in \mathbb{R}$ gibt, die sich (zum Entsetzen der alten Griechen) als irrational herausstellten:

Satz 1.8: ($\sqrt{2}$ ist irrational)

Es gibt keine rationale Zahl $q \in \mathbb{Q}$ mit $q^2 = 2$.

Beweis: Wir brauchen nur die Nichtexistenz von positiven $q \in \mathbb{Q}_+$ mit $q^2 = 2$ zu betrachten, denn mit $q^2 = 2$ gilt auch $(-q)^2 = 2$, d.h., mit einer negativen rationalen Lösung von $q^2 = 2$ hätten wir automatisch auch eine positive rationale Lösung.

Die typische Formulierung ist: wir können o.B.d.A. („ohne Beschränkung der Allgemeinheit“) $q > 0$ voraussetzen.

Wir führen einen **indirekten Beweis** zur Nichtexistenz von $q \in \mathbb{Q}_+$ mit $q^2 = 2$. Angenommen, es existiert $q = z/n$ mit $z \in \mathbb{Z}, n \in \mathbb{N}$ und $q^2 = 2$. Wir setzen voraus, dass z und n teilerfremde natürliche Zahlen sind (wenn nicht, ersetzen wir z und n durch die teilerfremden Werte, die nach Kürzen entstehen). Aus $q^2 = 2$ folgt

$$z^2 = 2 \cdot n^2 \quad \text{mit} \quad z, n \in \mathbb{N}.$$

Hieraus folgt, dass z gerade sein muss (sonst könnte z^2 nicht den Faktor 2 enthalten). Damit enthält z^2 den Faktor 4. Damit muss n gerade sein, denn sonst wäre $2 \cdot n^2$ nicht durch 4 teilbar.

Sowohl z als auch n müssen also durch 2 teilbar sein, wenn $(z/n)^2 = 2$ gilt. Widerspruch zur Teilerfremdheit von z und n ! Das einzig zweifelhaftes Glied der Beweiskette ist die Annahme, es gäbe ein $q = z/n \in \mathbb{Q}_+$ mit $q^2 = 2$. Also muss diese Annahme falsch gewesen sein.

Q.E.D.

Beispiel 1.9: Wir wissen, dass $\sqrt{2} \notin \mathbb{Q}$ gilt. Wir können nun aber die Existenz von $\sqrt{2}$ als die positive Lösung von $x^2 = 2$ mit $x \in \mathbb{R}$ **beweisen**. Wir setzen ganz konstruktiv

$$\beta := \sup \{x \in \mathbb{R}_+; x^2 \leq 2\}$$

und beweisen, dass $\beta \in \mathbb{R}$ existiert und $\beta^2 = 2$ gilt.

Für die Existenz von β ist nur zu zeigen, dass die betrachtete Menge

$$A = \{x \in \mathbb{R}_+; x^2 \leq 2\}$$

nach oben beschränkt ist. Dies ist leicht: aus $x^2 \leq 2$ folgt sicherlich $x < 1.5$, denn für $x \geq 1.5$ würde $x^2 \geq 1.5^2 = 2.25 > 2$ folgen, wir betrachten aber nur x mit $x^2 \leq 2$. Also ist 1.5 eine obere Schranke der Menge A .

Als allgemeines Prinzip gilt, dass das Supremum einer nach oben beschränkten Menge eindeutig ist. (Dies ist ein Quickie. Die einzige wirkliche Herausforderung besteht darin, dies auf einem Zeitungsrand zu beweisen).

Damit stellt obiges β wegen des Supremumsaxioms eine eindeutig definierte Zahl in \mathbb{R} dar. Es bleibt zu beweisen, dass $\beta^2 = 2$ gilt. Das ist etwas mühseliger.

Behauptung: $\beta = \sup \{x \in \mathbb{R}_+; x^2 \leq 2\}$ erfüllt $\beta^2 = 2$.

Beweis: Sicherlich gilt $\beta > 0$, also speziell $2 + \beta > 0$.

Angenommen, es gilt $\beta^2 < 2$. Dann kann β keine obere Schranke von A sein, denn z.B. die Zahl

$$a = \beta + \underbrace{\frac{2 - \beta^2}{2 + \beta}}_{>0} = \frac{2 \cdot \beta + \beta^2 + 2 - \beta^2}{2 + \beta} = \frac{2 + 2 \cdot \beta}{2 + \beta}$$

wäre echt größer als β . Sie liegt aber in A , denn es gilt

$$\begin{aligned} a^2 &= \frac{4 + 8 \cdot \beta + 4 \cdot \beta^2}{(2 + \beta)^2} = \frac{(8 + 8 \cdot \beta + 2 \cdot \beta^2) - 2 \cdot (2 - \beta^2)}{(2 + \beta)^2} \\ &= \frac{2 \cdot (2 + \beta)^2 - 2 \cdot (2 - \beta^2)}{(2 + \beta)^2} = 2 - 2 \cdot \underbrace{\frac{2 - \beta^2}{(2 + \beta)^2}}_{>0} < 2. \end{aligned}$$

Widerspruch zur Annahme, dass β obere Schranke von A ist!

Angenommen, es gilt $\beta^2 > 2$. Dann kann β nicht die kleinste obere Schranke von A sein. Z. B. ist die Zahl

$$\beta' = \beta + \underbrace{\frac{2 - \beta^2}{2 + \beta}}_{<0} = \frac{2 \cdot \beta + \beta^2 + 2 - \beta^2}{2 + \beta} = \frac{2 + 2 \cdot \beta}{2 + \beta}$$

kleiner als β . Sie ist aber eine obere Schranke von A , denn wegen

$$\begin{aligned} \beta'^2 &= \frac{4 + 8 \cdot \beta + 4 \cdot \beta^2}{(2 + \beta)^2} = \frac{(8 + 8 \cdot \beta + 2 \cdot \beta^2) + 2 \cdot (\beta^2 - 2)}{(2 + \beta)^2} \\ &= \frac{2 \cdot (2 + \beta)^2 + 2 \cdot (\beta^2 - 2)}{(2 + \beta)^2} = 2 + 2 \cdot \underbrace{\frac{\beta^2 - 2}{(2 + \beta)^2}}_{>0} > 2 \end{aligned}$$

gilt:

$$a \in A \Rightarrow a^2 \leq 2 \Rightarrow a^2 < \beta'^2 \Rightarrow a < \beta'.$$

(Im letzten Schritt wird ausgenutzt, dass wir bereits wissen, dass $\beta' = (2+2\cdot\beta)/(2+\beta) > 0$ gilt, da sicherlich $\beta > 0$ gilt.) Widerspruch!

Wir haben gezeigt, dass weder $\beta^2 < 2$ noch $\beta^2 > 2$ gelten kann. Es bleibt nur $\beta^2 = 2$.
Q.E.D.

Analog zu oberen Schranken heißt $\alpha \in \mathbb{R}$ **untere Schranke** einer Teilmenge $A \subset \mathbb{R}$, wenn $\alpha \leq a$ für alle $a \in A$ gilt. Wir führen die größte untere Schranke als das **Infimum** von A ein, indem wir es auf das Supremum zurückführen:

$$\inf A = -\sup \{-a; a \in A\}.$$

Man kann leicht nachrechnen, dass dies in der Tat die größte untere Schranke von A ist. Da die Existenz von Suprema/Infima eine so wichtige strukturelle Eigenschaft der reellen Zahlen darstellt, stellen wir das Supremumsaxiom noch einmal heraus:

Das Supremumsaxiom 1.10:

Jede nach oben beschränkte nichtleere Teilmenge A von \mathbb{R} besitzt eine kleinste obere Schranke („das Supremum“ von A):

$$\sup A = \min\{\beta \in \mathbb{R}; a \leq \beta \forall a \in A\}.$$

Jede nach unten beschränkte nichtleere Teilmenge A von \mathbb{R} besitzt eine größte untere Schranke („das Infimum“ von A):

$$\inf A = \max\{\alpha \in \mathbb{R}; \alpha \leq a \forall a \in A\}.$$

Die Existenz des Minimums/Maximums aller oberen/unteren Schranken, welche das Supremum/Infimum definieren, ist die gewünschte Vollständigkeit der reellen Zahlen, die \mathbb{R} z.B. von \mathbb{Q} unterscheidet und uns die Existenz der über \mathbb{Q} hinausgehenden irrationalen Zahlen wie Wurzeln etc. garantiert. In der Tat ist „ \mathbb{R} deutlich größer als \mathbb{Q} “. Beispielsweise können die reellen Zahlen nicht mehr durch \mathbb{N} indiziert werden (d.h., es gibt keine Zahlenfolge (x_n) , die \mathbb{R} vollständig durchläuft):

Satz 1.11: (Überabzählbarkeit von \mathbb{R})

Die Menge \mathbb{R} der reellen Zahlen ist nicht abzählbar, d.h., es gibt keine invertierbare Abbildung von \mathbb{N} nach \mathbb{R} .

Beweis: (Nach Chr. Blatter, Analysis I) Betrachte eine beliebige Abbildung $n \in \mathbb{N} \rightarrow x_n \in \mathbb{R}$. Wir können stets (mindestens) eine Zahl in \mathbb{R} konstruieren, die nicht in der Menge $\{x_n; n \in \mathbb{N}\}$ enthalten ist. Dazu definieren wir die abgeschlossenen Intervalle

$$[a_1, b_1] = [x_1 + 1, x_1 + 2],$$

und

$$[a_{n+1}, b_{n+1}] = \begin{cases} [a_n, \frac{3 \cdot a_n + b_n}{4}], & \text{falls } x_{n+1} \geq \frac{a_n + b_n}{2}, \\ [\frac{a_n + 3 \cdot b_n}{4}, b_n] & \text{falls } x_{n+1} < \frac{a_n + b_n}{2}. \end{cases}$$

Anschaulich: $[a_{n+1}, b_{n+1}]$ ist das linke Viertel von $[a_n, b_n]$, wenn x_{n+1} rechts von der Mitte sitzt und das rechte Viertel, wenn x_{n+1} links von der Mitte sitzt. Damit steigen die a_n monoton, die b_n fallen monoton, und es gilt

$$a_1 \leq a_2 \leq a_3 \leq \dots \leq \dots \leq b_3 \leq b_2 \leq b_1,$$

also

$$[a_1, b_1] \supset [a_2, b_2] \supset [a_3, b_3] \supset \dots \supset [\alpha, \beta],$$

wobei $\alpha = \sup \{\alpha_n; n \in \mathbb{N}\}$, $\beta = \inf \{\beta_n; n \in \mathbb{N}\}$. Ausserdem gilt nach Konstruktion $x_n \notin [a_n, b_n]$, also

$$x_n \notin [a_n, b_n] \supset [a_{n+1}, b_{n+1}] \supset \dots \supset [\alpha, \beta].$$

Damit gibt es die beiden Zahlen $\alpha \in \mathbb{R}$ und $\beta \in \mathbb{R}$ (die evtl. übereinstimmen können), die von allen x_n verschieden sind.

Q.E.D.

Bemerkung 1.12: Existiert in $A \subset \mathbb{R}$ ein maximales Element, so ist dieses Maximum auch das Supremum:

$$\sup A = \max A.$$

Aber nicht jede beschränkte Menge hat ein maximales Element. Z.B. hat $A = [0, 1)$ kein Maximum, denn das Supremum $\sup A = 1$ (der einzige Kandidat für das Maximum) liegt nicht in A .

Merke: Im Gegensatz zu Maxima/Minima existieren Suprema/Infima beschränkter Mengen in \mathbb{R} immer!

Zuletzt noch der für die Analysis unersetzliche Begriff des **Betrags** einer reellen Zahl

$$|x| = \begin{cases} x & \text{falls } x \geq 0, \\ -x & \text{falls } x < 0. \end{cases}$$

Über den Betrag erhält man den Begriff des **Abstands** $|x - y|$ zweier Zahlen. Die fundamentale Eigenschaft des Betrags- bzw. Abstandsbegriffs ist die:

Dreieckungleichung 1.13:

Für alle $x, y \in \mathbb{R}$ gilt die „Dreiecksungleichung“

$$|x + y| \leq |x| + |y|$$

und die „umgekehrte Dreiecksungleichung“:

$$\left| |x| - |y| \right| \leq |x - y|.$$

Beweis: Die Dreiecksungleichung ist durch einige Fallunterscheidungen leicht zu überprüfen (Aufgabe 6, Blatt 1). Die umgekehrte Dreiecksungleichung folgt unmittelbar aus der Dreiecksungleichung. Ersetzt man x durch $x - y$, erhält man

$$|(x - y) + y| \leq |x - y| + |y|, \text{ also } |x| \leq |x - y| + |y|.$$

Ersetzt man y durch $y - x$, erhält man

$$|x + (y - x)| \leq |x| + |y - x|, \text{ also } |y| \leq |x| + |x - y|.$$

Die beiden resultierenden Ungleichungen

$$|x| - |y| \leq |x - y|, \quad |y| - |x| \leq |x - y|$$

lassen sich zur umgekehrten Dreiecksungleichung

$$\left| |x| - |y| \right| \leq |x - y|$$

zusammenfassen.

Q.E.D.

1.5 Die komplexen Zahlen

21.10.03↓

Motivation: die Gleichung $x^2 = -1$ hat offensichtlich keine reellen Lösungen, da $x^2 \geq 0$ für jedes reelle x gilt. Um auch diese Gleichung lösen zu können, muß man neue Zahlen einführen: die **komplexen Zahlen**. Die grundsätzliche Idee ist ganz einfach: man führt ein neues Symbol i ein, das $\sqrt{-1}$ repräsentieren soll. Es wird einzig und allein durch die Rechenregel $i^2 = -1$ festgelegt. Ansonsten behält man alle aus dem Reellen bekannten Rechenregeln einfach bei.

1.5.1 Definitionen

Definition 1.14: (Die komplexen Zahlen)

Die Menge der **komplexen Zahlen** \mathbb{C} ist die Menge aller formalen Summen der Form

$$\mathbb{C} = \{x + i \cdot y; x, y \in \mathbb{R}\}.$$

Für $z = x + i \cdot y \in \mathbb{C}$ nennt man x den **Realteil** und y den **Imaginärteil** von z .

Zahlen $z = x + i \cdot 0$ mit $y = 0$ nennt man **reell**, schreibt auch kurz $z = x$ und identifiziert z mit $x \in \mathbb{R}$.

Zahlen $z = 0 + i \cdot y$ mit $x = 0$ nennt man **imaginär** und schreibt auch kurz $z = i \cdot y$.

Der **Nullpunkt** $z = 0 + i \cdot 0$ wird auch kurz als $z = 0$ geschrieben.

Auf \mathbb{C} definieren wir die Addition

$$z_1 + z_2 = (x_1 + i \cdot y_1) + (x_2 + i \cdot y_2) = \underbrace{(x_1 + x_2)}_{\in \mathbb{R}} + i \cdot \underbrace{(y_1 + y_2)}_{\in \mathbb{R}}$$

sowie die Multiplikation

$$z_1 \cdot z_2 = (x_1 + i \cdot y_1) \cdot (x_2 + i \cdot y_2) = \underbrace{(x_1 \cdot x_2 - y_1 \cdot y_2)}_{\in \mathbb{R}} + i \cdot \underbrace{(x_1 \cdot y_2 + x_2 \cdot y_1)}_{\in \mathbb{R}}.$$

Interpretation 1.15:

Hinter dieser Definition der Multiplikation steckt $i^2 = -1$:

$$i \cdot i = (0 + i \cdot 1) \cdot (0 + i \cdot 1) = (0 \cdot 0 - 1 \cdot 1) + i \cdot (0 \cdot 1 + 1 \cdot 0) = -1.$$

Man braucht sich die formale Definition der Multiplikation nicht zu merken: man benutze einfach die üblichen aus \mathbb{R} bekannten Rechenregeln (Kommutativität $a \cdot b = b \cdot a$, Assoziativität $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, das Distributivgesetz $a \cdot (b + c) = a \cdot b + a \cdot c$ etc.), und setze beim Rechnen

$$i^2 = -1, \quad i^3 = (i^2) \cdot i = -i, \quad i^4 = (i^3) \cdot i = (-i) \cdot i = -(i^2) = 1$$

usw. ein:

$$\begin{aligned} (x_1 + i \cdot y_1) \cdot (x_2 + i \cdot y_2) &= x_1 \cdot x_2 + i \cdot x_1 \cdot y_2 + i \cdot x_2 \cdot y_1 + \underbrace{i^2}_{-1} \cdot y_1 \cdot y_2 \\ &= x_1 \cdot x_2 + i \cdot x_1 \cdot y_2 + i \cdot x_2 \cdot y_1 - y_1 \cdot y_2 = (x_1 \cdot x_2 - y_1 \cdot y_2) + i \cdot (x_1 \cdot y_2 + x_2 \cdot y_1). \end{aligned}$$

Folgerung 1.16:

Wir konstruieren eine Division für $z = x + i \cdot y \neq 0 + i \cdot 0 \equiv 0$:

$$\begin{aligned} \frac{1}{z} &= \frac{1}{x + i \cdot y} = \frac{1}{x + i \cdot y} \cdot \frac{x - i \cdot y}{x - i \cdot y} \\ &= \frac{x - i \cdot y}{(x + i \cdot y) \cdot (x - i \cdot y)} = \frac{x - i \cdot y}{x^2 - (i \cdot y)^2} = \frac{x}{x^2 + y^2} - i \cdot \frac{y}{x^2 + y^2}. \end{aligned}$$

Allgemein:

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{x_1 + i \cdot y_1}{x_2 + i \cdot y_2} = \frac{(x_1 + i \cdot y_1) \cdot (x_2 - i \cdot y_2)}{(x_2 + i \cdot y_2) \cdot (x_2 - i \cdot y_2)} \\ &= \frac{(x_1 \cdot x_2 + y_1 \cdot y_2) + i \cdot (-x_1 \cdot y_2 + x_2 \cdot y_1)}{x_2^2 + y_2^2 + i \cdot (x_2 \cdot y_2 - x_2 \cdot y_2)} \\ &= \frac{x_1 \cdot x_2 + y_1 \cdot y_2}{x_2^2 + y_2^2} + i \cdot \frac{x_2 \cdot y_1 - x_1 \cdot y_2}{x_2^2 + y_2^2}. \end{aligned}$$

Definition 1.17: (komplexe Konjugation etc.)

Es werden folgende speziellen Operationen auf den komplexen Zahlen eingeführt:

$$\begin{aligned} \Re(z) &= \Re(x + i \cdot y) = x && \text{(der Realteil von } z), \\ \Im(z) &= \Im(x + i \cdot y) = y && \text{(der Imaginärteil von } z), \\ |z| &= |x + i \cdot y| = \sqrt{x^2 + y^2} && \text{(der Betrag von } z), \\ \bar{z} &= \overline{x + i \cdot y} = x - i \cdot y && \text{(das komplex Konjugierte von } z). \end{aligned}$$

Merkregel 1.18:

Die Division komplexer Zahlen läuft auf den Standardtrick „Erweitern mit dem komplex konjugierten Nenner hinaus“:

$$\frac{z_1}{z_2} = \frac{x_1 + i \cdot y_1}{x_2 + i \cdot y_2} = \frac{(x_1 + i \cdot y_1) \cdot (x_2 - i \cdot y_2)}{(x_2 + i \cdot y_2) \cdot (x_2 - i \cdot y_2)} = \frac{z_1 \cdot \bar{z}_2}{z_2 \cdot \bar{z}_2}.$$

Satz 1.19: (Rechenregeln)

Für alle $z, z_1, z_2 \in \mathbb{C}$ gilt: „Kommutativität“ und „Assoziativität“ von Multiplikation und Division

$$z_1 \cdot z_2 = z_2 \cdot z_1, \quad (z_1 \cdot z_2) \cdot z_3 = z_1 \cdot (z_2 \cdot z_3),$$

$$\frac{1}{z_1 \cdot z_2} = \frac{1}{z_2} \cdot \frac{1}{z_1}, \quad \left(\frac{1}{z_1} \cdot \frac{1}{z_2}\right) \cdot \frac{1}{z_3} = \frac{1}{z_1} \cdot \left(\frac{1}{z_2} \cdot \frac{1}{z_3}\right),$$

„Linearität“ von \Re , \Im und Konjugation

$$\Re(z_1 + z_2) = \Re(z_1) + \Re(z_2), \quad \Im(z_1 + z_2) = \Im(z_1) + \Im(z_2), \quad \overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2,$$

„Multiplikativität“ des Betrags und der Konjugation

$$|z_1 \cdot z_2| = |z_1| \cdot |z_2|, \quad \left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|}, \quad \overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}, \quad \overline{\left(\frac{z_1}{z_2} \right)} = \frac{\overline{z_1}}{\overline{z_2}}$$

sowie die Beziehungen

$$|z|^2 = |\bar{z}|^2 = z \cdot \bar{z} = \Re(z)^2 + \Im(z)^2, \quad \frac{z_1}{z_2} = \frac{z_1 \cdot \bar{z}_2}{|z_2|^2}$$

und die „Dreiecksungleichung“:

$$|z_1 + z_2| \leq |z_1| + |z_2|.$$

Beweis: Alles ist direkt nachzurechnen, z.B.

$$z \cdot \bar{z} = (x + i \cdot y) \cdot (x - i \cdot y) = x^2 + y^2 = |z|^2 = |\bar{z}|^2$$

oder (wie schon oben durchgeführt):

$$\frac{z_1}{z_2} = \frac{z_1 \cdot \bar{z}_2}{z_2 \cdot \bar{z}_2} = \frac{z_1 \cdot \bar{z}_2}{|z_2|^2}.$$

Q.E.D.

Beispiel 1.20: In MuPAD wird $i = \sqrt{-1}$ durch I dargestellt:

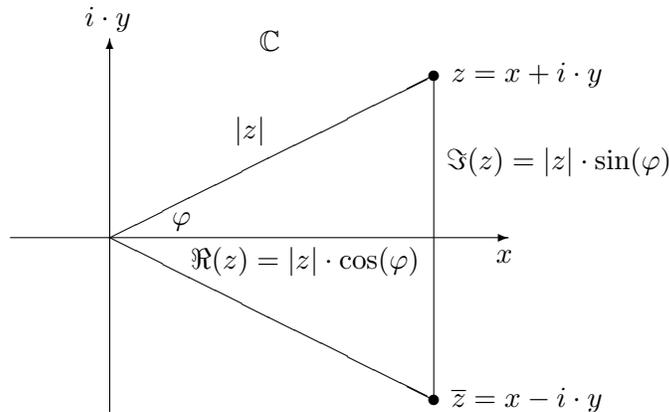
```
>> sqrt(-1)
I
>> I^2
-1
```

Die MuPAD-Funktionen `Re`, `Im`, `conjugate` und `abs` berechnen Real- und Imaginärteil, komplexe Konjugation und den Absolutbetrag:

```
>> z:= 2 + 3*I:
>> Re(z), Im(z), conjugate(z), abs(z)
2, 3, 2 - 3 I, 131/2
```

Geometrische Interpretation 1.21:

Man stellt sich üblicherweise die Menge der komplexen Zahlen als 2-dimensionale Ebene („die komplexe Ebene“) vor:



Der Betrag von z ist der Abstand zum Ursprung, komplexe Konjugation entspricht der Spiegelung an der x -Achse („die reelle Achse“). Die y -Achse wird auch als „imaginäre Achse“ bezeichnet.

Geometrisch ist \mathbb{C} nichts anderes als \mathbb{R}^2 : $x + i \cdot y \in \mathbb{C} \hat{=} (x, y) \in \mathbb{R}^2$. Die komplexe Addition entspricht genau der Addition von Vektoren im \mathbb{R}^2 . Algebraisch besteht der Unterschied zwischen \mathbb{C} und \mathbb{R}^2 darin, dass man auf \mathbb{C} neben der Addition noch eine Multiplikation $\mathbb{C} \times \mathbb{C} \mapsto \mathbb{C}$ hat, wohingegen es auf \mathbb{R}^2 keine interessante Multiplikation zweier Vektoren gibt, die wieder einen Vektor liefert (außer derjenigen, die \mathbb{R}^2 zu \mathbb{C} macht).

Bemerkung 1.22: Führt man den eingezeichneten Winkel φ zwischen dem „Vektor“ z und der reellen Achse ein, so gilt mit den aus der Schule bekannten Winkelfunktionen \sin und \cos :

$$z = x + i \cdot y = |z| \cdot \cos(\varphi) + i \cdot |z| \cdot \sin(\varphi).$$

Die Darstellung $z = x + i \cdot y$ nennt man die **Kartesische Darstellung** der komplexen Zahl z durch Real- und Imaginärteil. Die Darstellung

$$z = |z| \cdot \left(\cos(\varphi) + i \cdot \sin(\varphi) \right)$$

durch den Betrag $|z|$ und den „Polarwinkel“ $\varphi \in [0, 2\pi)$ heißt **Polardarstellung** von z (φ heißt auch „das Argument von z “). Wir werden später in Abschnitt 5.3 auf die Polardarstellung komplexer Zahlen zurückkommen, nachdem wir die komplexe Exponentialfunktion eingeführt haben.

Bemerkung 1.23: Im Gegensatz zu \mathbb{R} ist \mathbb{C} **prinzipiell nicht orientierbar!** Damit ist gemeint, dass es auf \mathbb{C} keine Ordnung $z_1 < z_2$ im Sinne von (R4) der Axiomatisierung 1.6 geben kann, die mit den Körperaxiomen (R0) – (R3) verträglich ist (diese gelten auch für \mathbb{C} : mit der komplexen Addition und Multiplikation ist \mathbb{C} in der Tat ein Körper). Dass prinzipiell keine Anordnung mit (R0) – (R4) auf \mathbb{C} existieren kann, ist leicht gezeigt:

Angenommen, es gäbe \mathbb{C}_+ (die Menge der positiven komplexen Zahlen). Das neutrale Element 1 der Multiplikation (dies ist auch auf \mathbb{C} die reelle Zahl 1) muss in \mathbb{C}_+ liegen, damit liegt -1 nicht in \mathbb{C}_+ . Liegt i in \mathbb{C}_+ ? Nein, denn sonst müsste nach (R4) die Zahl $i \cdot i = -1$ wieder in \mathbb{C}_+ liegen. Mit demselben Argument kann $-i$ wegen $(-i) \cdot (-i) = -1$ auch nicht in \mathbb{C}_+ liegen. Damit haben wir ein Element $z = i \neq 0$ mit der Eigenschaft, dass weder z noch $-z$ in \mathbb{C}_+ liegt. Dies verletzt (R4).

1.5.2 Nullstellen von Polynomen, der Fundamentalsatz der Algebra

↓22.10.03

Die Motivation zur Einführung der komplexen Zahlen war, Polynomgleichungen wie z.B. $x^2 + 1 = 0$ lösen zu können. In der Tat stellt sich nun heraus, dass Polynome vom Grad n immer genau n (evtl. „entartete“) komplexe Nullstellen haben. Wir definieren zunächst „Entartung“ von Nullstellen, wobei wir auf die aus der Schule bekannte Differentiation zurückgreifen:

Definition 1.24: (Vielfachheit von Nullstellen)

Sei $f : \mathbb{R} \mapsto \mathbb{R}$ eine mehrfach differenzierbare Funktion (siehe Kapitel 6). Man nennt x^* eine Nullstelle der „**Vielfachheit**“ k (oder auch „ **k -fache Nullstelle**“), wenn

$$f(x^*) = f'(x^*) = \dots = f^{(k-1)}(x^*) = 0, \quad f^{(k)}(x^*) \neq 0.$$

Beispiel 1.25: (Mehrfache Polynomwurzeln)

Für das Polynom $p(x) = (x - x^*)^n$ mit $n > 0$ ist x^* eine n -fache Nullstelle:

$$p(x) = (x - x^*)^n, \quad p'(x) = n \cdot (x - x^*)^{n-1}, \quad p''(x) = n \cdot (n-1) \cdot (x - x^*)^{n-2}, \\ \dots, \quad p^{(n-1)}(x) = n \cdot (n-1) \cdot \dots \cdot 2 \cdot (x - x^*), \quad p^{(n)}(x) = n!,$$

also

$$p(x^*) = (x^* - x^*)^n = 0, \\ p'(x^*) = n \cdot (x^* - x^*)^{n-1} = 0, \\ p''(x^*) = n \cdot (n-1) \cdot (x^* - x^*)^{n-2} = 0, \\ \dots$$

$$p^{(n-1)}(x^*) = n \cdot (n-1) \cdot \dots \cdot 2 \cdot (x^* - x^*)^1 = 0,$$

$$p^{(n)}(x^*) = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1 \cdot (x^* - x^*)^0 = n! \neq 0.$$

Beispiel 1.26: (Mehrfache Polynomwurzeln)

Seien x_1, \dots, x_k verschieden. Für das Polynom

$$p(x) = (x - x_1)^{n_1} \cdot \dots \cdot (x - x_k)^{n_k}$$

vom Grad $n = n_1 + \dots + n_k$ sind x_1, \dots, x_k Nullstellen der Vielfachheit $n_1, \dots, n_k \in \mathbb{N}$. Der Nachweis geht analog zum letzten Beispiel: Betrachte eine der Nullstellen x_i und schreibe

$$p(x) = (x - x_i)^{n_i} \cdot f(x), \quad f(x) = (x - x_1)^{n_1} \cdot \dots \cdot \cancel{(x - x_i)^{n_i}} \cdot \dots \cdot (x - x_k)^{n_k}.$$

Über die aus der Schule bekannte Produktregel $(g \cdot f)' = g' \cdot f + g \cdot f'$ der Differentiation folgt

$$p(x) = (x - x_i)^{n_i} \cdot f(x),$$

$$p'(x) = n_i \cdot (x - x_i)^{n_i-1} \cdot f(x) + (x - x_i)^{n_i} \cdot f'(x),$$

$$p''(x) = n_i \cdot (n_i - 1) \cdot (x - x_i)^{n_i-2} \cdot f(x) + 2 \cdot n_i \cdot (x - x_i)^{n_i-1} \cdot f'(x) + (x - x_i)^{n_i} \cdot f''(x)$$

usw., wobei f, f', f'' etc. Polynome sind. Die ersten $n_i - 1$ Ableitungen verschwinden an der Stelle $x = x_i$:

$$p(x_i) = 0^{n_i} \cdot f(x_i) = 0,$$

$$p'(x_i) = 0^{n_i-1} \cdot f(x_i) + 0^{n_i} \cdot f'(x_i) = 0$$

$$p''(x_i) = (\dots) \cdot 0^{n_i-2} \cdot f(x_i) + (\dots) \cdot 0^{n_i-1} \cdot f'(x_i) + 0^{n_i} \cdot f''(x_i) = 0$$

usw. Die n_i -te Ableitung verschwindet nicht:

$$p^{(n_i)}(x) = n_i! \cdot f(x) + (\dots) \cdot (x - x_i) \cdot f'(x) + \dots + (x - x_i)^{n_i} \cdot f^{(n_i)}(x),$$

also $p^{(n_i)}(x_i) = n_i! \cdot f(x_i)$, wobei

$$f(x_i) = (x_i - x_1)^{n_1} \cdot \dots \cdot \cancel{(x_i - x_i)^{n_i}} \cdot \dots \cdot (x_i - x_k)^{n_k} \neq 0$$

gilt, da $x_i \neq x_1, \dots, x_i \neq x_k$ vorausgesetzt ist. Damit ist x_i eine Nullstelle der Vielfachheit n_i .

Ein Polynom und seine Ableitungen

$$p(x) = a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + \dots + a_1 \cdot x + a_0,$$

$$p'(x) = a_n \cdot n \cdot x^{n-1} + a_{n-1} \cdot (n-1) \cdot x^{n-2} + \dots + a_1$$

etc. ist natürlich auch für komplexe Zahlen x wohldefiniert und man kann daher nach (mehrfachen) komplexen Nullstellen fragen. Die Definition 1.24 der Vielfachheit wird dabei auch für komplexe Nullstellen beibehalten.

Wir stellen zunächst das Horner-Schema zur Polynomauswertung und Polynomdivision vor:

Satz 1.27: (Polynomauswertung und -division per Horner-Schema)

Für das Polynom $p(x) = a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + \dots + a_0$ mit Koeffizienten $a_k \in \mathbb{C}$ gilt für jedes $x^* \in \mathbb{C}$:

$$\frac{p(x) - p(x^*)}{x - x^*} = b_0 \cdot x^{n-1} + b_1 \cdot x^{n-2} + \dots + b_{n-2} \cdot x + b_{n-1},$$

wobei b_0, b_1 etc. durch die Rekursion („**Horner-Schema**“)

$b_0 := a_n ;$ $\text{for } k := 1 \text{ to } n \text{ do } b_k := b_{k-1} \cdot x^* + a_{n-k} ;$

gegeben sind. Es gilt $b_n = p(x^*)$.

Beweis: Mit $b_0 = a_n$, $b_k - b_{k-1} \cdot x^* = a_{n-k}$ und $-b_{n-1} \cdot x^* = a_0 - b_n$ folgt

$$\begin{aligned} & (x - x^*) \cdot \left(b_0 \cdot x^{n-1} + b_1 \cdot x^{n-2} + \dots + b_{n-2} \cdot x + b_{n-1} \right) \\ &= b_0 \cdot x^n + b_1 \cdot x^{n-1} + b_2 \cdot x^{n-2} + \dots + b_{n-1} \cdot x \\ & \quad - b_0 \cdot x^* \cdot x^{n-1} - b_1 \cdot x^* \cdot x^{n-2} - \dots - b_{n-2} \cdot x^* \cdot x - b_{n-1} \cdot x^* \\ &= a_n \cdot x^n + \underbrace{a_{n-1}} \cdot x^{n-1} + \underbrace{a_{n-2}} \cdot x^{n-2} + \dots + \underbrace{a_1} \cdot x + \underbrace{a_0 - b_n} \\ &= p(x) - b_n. \end{aligned}$$

Für $x = x^*$ folgt $0 = p(x^*) - b_n$ und damit

$$(x - x^*) \cdot \left(b_0 \cdot x^{n-1} + b_1 \cdot x^{n-2} + \dots + b_{n-2} \cdot x + b_{n-1} \right) = p(x) - p(x^*).$$

Q.E.D.

Das Horner-Schema liefert mittels $b_n = p(x^*)$ die Auswertung des Polynoms an einer Stelle x^* mit n Multiplikationen und $n - 1$ Additionen. In der Tat ist es (für „dicht besetzte“ Polynome) das Standardschema, mit dem auf dem Rechner Polynomauswertungen implementiert werden. Bei der Auswertung werden gleichzeitig die Koeffizienten b_0, \dots, b_{n-1} des „**Faktorpolynoms**“ $(p(x) - p(x^*)) / (x - x^*)$ mitgeliefert. Das Horner-Schema läuft auf die folgende Darstellung des Polynoms hinaus:

$$\begin{aligned}
 p(x^*) &= ((\dots((a_n \cdot x^* + a_{n-1}) \cdot x^* + a_{n-2}) \cdot x^* + \dots) \dots) \cdot x^* + a_1) \cdot x^* + a_0. \\
 &\quad \uparrow \\
 &\quad b_0 = a_n \\
 &\quad \underbrace{\hspace{2cm}} \\
 &\quad b_1 = a_n \cdot x^* + a_{n-1} \\
 &\quad \underbrace{\hspace{2cm}} \\
 &\quad b_2 = a_n \cdot x^{*2} + a_{n-1} \cdot x^* + a_{n-2} \\
 &\quad \quad \quad \vdots \\
 &\quad \underbrace{\hspace{2cm}} \\
 &\quad b_n = a_n \cdot x^{*n} + a_{n-1} \cdot x^{*(n-1)} + \dots + a_1 \cdot x^* + a_0 = p(x^*)
 \end{aligned}$$

Ist x^* eine Nullstelle („Wurzel“) des Polynoms, so folgt $p(x)/(x - x^*) = \text{Polynom}(x)$. Es ergibt sich das Grundprinzip, dass man bei einer gegebenen Nullstelle einen „Linearfaktor“ $x - x^*$ vom Polynom abspalten kann:

Folgerung 1.28:

Ist x^ eine Wurzel des Polynoms p vom Grad $n > 0$, so gilt*

$$p(x) = (x - x^*) \cdot q(x)$$

mit einem „Faktorpolynom“ $q(x) = b_0 \cdot x^{n-1} + b_1 \cdot x^{n-2} + \dots + b_{n-1}$ vom Grad $n - 1$, dessen Koeffizienten z.B. durch das Horner-Schema berechenbar sind.

Merke: x^* ist dann und genau dann eine Wurzel, wenn sich der Linearfaktor $x - x^*$ vom Polynom abspalten lässt.

Zwar hat nicht jedes Polynom reelle Nullstellen, aber es gilt das (zu beweisende) wichtige Prinzip:

Jedes Polynom vom Grad > 0 hat (mindestens) eine komplexe Nullstelle.

Setzen wir zur Motivation des kommenden Fundamentalsatzes 1.29 mal dieses Prinzip voraus. Es gilt $p(x) = (x - x^*) \cdot q(x)$ mit einer (garantiert existierenden) Nullstelle x^* von p . Die Nullstellen des Faktorpolynoms q sind offensichtlich wieder Nullstellen des Ausgangspolynoms p . Hat man nun eine Nullstelle x^{**} von q , so kann man nach Folgerung 1.28 angewendet auf q einen weiteren Linearfaktor abspalten:

$$q(x) = (x - x^{**}) \cdot \tilde{q}(x), \quad \text{also} \quad p(x) = (x - x^*) \cdot (x - x^{**}) \cdot \tilde{q}(x)$$

mit einem Restpolynom \tilde{q} vom Grad $n - 2$. Dies setzt man fort, bis man nach n Schritten auf ein konstantes Polynom stößt, das keine Nullstellen mehr besitzt. Es folgt eine Faktordarstellung

$$p(x) = (x - x^*) \cdot (x - x^{**}) \cdot (x - x^{***}) \cdot \dots \cdot (x - x^{*\dots*}) \cdot a_n,$$

wobei a_n das zuletzt verbleibende konstante Restpolynom (vom Grad 0) ist. Vergleicht man die führenden Koeffizienten auf der linken und rechten Seite dieser Gleichung, so sieht man sofort, dass das verbleibende konstante Restpolynom nichts anderes als der führende Koeffizient von $p(x) = a_n \cdot x^n + \dots + a_0$ ist. Es folgt das Grundprinzip:

Jedes Polynom vom Grad $n > 0$ hat genau n komplexe Nullstellen.

Hierbei müssen wir aber etwas vorsichtig zählen, da in der Konstruktion die Nullstellen x^* , x^{**} , x^{***} etc. eventuell übereinstimmen können. Eine saubere Formulierung liefert der folgende fundamentale Satz:

Satz 1.29: (Fundamentalsatz der Algebra, Gauß 1799)

Zu jedem Polynom $p(x) = a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + \dots + a_0$ vom Grad $n > 0$ mit $a_0, \dots, a_n \in \mathbb{C}$ und $a_n \neq 0$ gibt es komplexe Zahlen z_1, \dots, z_k (die Wurzeln) und $n_1, \dots, n_k \in \mathbb{N}$ (die Vielfachheiten) mit $n_1 + \dots + n_k = n$, so dass

$$p(x) = a_n \cdot (x - z_1)^{n_1} \cdot \dots \cdot (x - z_k)^{n_k}.$$

Für die Anzahl k der *unterschiedlichen* Wurzeln gilt hierbei $k \leq n$ wegen $n_1 + \dots + n_k = n$.

zum Beweis: Wie in der Motivation gezeigt, braucht man nur zu beweisen, dass jedes Polynom vom Grad > 0 mindestens eine komplexe Nullstelle besitzt. Dies ist je nach den zur Verfügung stehenden Hilfsmitteln aber gar nicht so einfach und sprengt unseren Rahmen hier. Typischerweise wird der Satz in Lehrbüchern über komplexe Funktionentheorie bewiesen. Ein „elementarer“ Beweis findet sich z.B. unter:

<http://helios.mathematik.uni-kl.de/~luene/kleinodien/laplace.html>

Interpretation 1.30:

Nach Beispiel 1.26 sind z_1, \dots, z_k die Nullstellen von p mit den Vielfachheiten n_1, \dots, n_k . Zählt man z_1 als n_1 Wurzeln, z_2 als n_2 Wurzeln etc., so ergeben sich insgesamt $n_1 + \dots + n_k = n$ komplexe Wurzeln des Polynoms, und in der Tat erhalten wir in dieser Zählweise:

Jedes Polynom vom Grad $n > 0$ hat genau n komplexe Nullstellen!

Über die Anzahl der **reellen** Nullstellen hingegen kann man i.A. wenig aussagen (z.B. hat $p(x) = x^2 + 1$ überhaupt keine reelle Nullstellen).

Merkregel 1.31:

Der Punkt x^* ist dann und genau dann eine Nullstelle eines Polynoms $p(x)$, wenn sich gemäß Folgerung 1.28 der Linearfaktor $x - x^*$ vom Polynom abfaktorieren läßt. Die Vielfachheit von x^* gibt an, wie oft sich dieser Linearfaktor abspalten läßt. Man sollte eine Polynomwurzel besser als einen Linearfaktor ansehen. Der Fundamentalsatz besagt, dass sich ein Polynom vom Grad n immer in genau n Linearfaktoren aufspalten läßt.

Die Existenz der komplexen Wurzeln sagt nichts darüber aus, ob man diese Wurzeln in irgendeiner Weise explizit darstellen kann. In der Tat gibt es z.B. für Polynome vom Grad ≥ 5 keine allgemeingültige Lösungsformel mit Hilfe von (verschachtelten) Wurzeln. Numerisch kann man jedoch stets Gleitpunktapproximationen der Wurzeln finden.

Beispiel 1.32: In MuPAD ist `solve` für exakte Lösungen und `numeric::solve` für numerische Lösungen zuständig. Das folgende Polynom hat 9 Wurzeln, die sich (zufälligerweise) alle explizit darstellen lassen:

```
>> p:= x^9 + 2*x^7 - x^3 - 2*x:
>> solve(p = 0, x)
```

```
{0, -1, 1, - I 21/2, I 21/2, - 1/2 I 31/2 - 1/2,
 1/2 - 1/2 I 31/2, 1/2 I 31/2 - 1/2, 1/2 I 31/2 + 1/2}
```

Der numerische Gleichungslöser liefert Gleitpunktnäherungen der Wurzeln:

```
>> numeric::solve(p = 0, x)
{0.0, - 0.5 - 0.8660254038 I, - 0.5 + 0.8660254038 I,
 0.5 - 0.8660254038 I, 1.0, -1.414213562 I, 1.414213562 I,
 0.5 + 0.8660254038 I, -1.0}
```

Die zurückgegebenen Objekte {...} sind jeweils Mengen, deren Elementanordnung willkürlich vom System nach internen Kriterien bestimmt wird. Diese sehen für exakte Werte anders aus als für Gleitpunktnäherungen, so dass sich die Reihenfolge der Elemente beim exakten und beim numerischen Lösen unterscheiden kann (was im obigen Beispiel auch in der Tat der Fall ist).

Es fällt hierbei auf, dass die komplexen Wurzeln als komplex konjugierte Paare $x_k \pm i \cdot y_k$ auftauchen. Das ist kein Zufall und liegt daran, dass das eben betrachtete Polynom „reell“ ist (damit ist gemeint, dass die Koeffizienten reell sind).

Satz 1.33: (konjugierte Wurzepaare reeller Polynome)

Ist z eine k -fache Nullstelle des Polynoms $p(x) = a_n \cdot x^n + \dots + a_0$ mit reellen Koeffizienten a_0, \dots, a_n , so ist auch \bar{z} eine k -fache Nullstelle des Polynoms. Bei reellen Polynomen tauchen nicht-reelle Wurzeln also immer in komplex-konjugierten Paaren auf.

Beweis: Für ein reelles Polynom gilt wegen $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$ offensichtlich

$$\overline{p(z)} = p(\bar{z}).$$

Also gilt $p(\bar{z}) = 0$ dann und genau dann, wenn $p(z) = 0$ gilt. Da mit p auch alle Ableitungen von p wieder reelle Polynome sind, stimmen auch die Vielfachheiten der Nullstellen z und \bar{z} überein.

Q.E.D.

Kapitel 2

Folgen und Grenzwerte

↓23.10.03

Die Grundlage der Analysis ist der Begriff des Grenzwertes. Er ist aus der Schule bekannt (bzw. sollte bekannt sein) und wird hier rekapituliert. Da es kaum einen Unterschied macht, Folgen und Grenzwerte in \mathbb{R} oder in \mathbb{C} zu betrachten, formulieren wir die folgenden Definitionen und Sätze in \mathbb{C} , was \mathbb{R} als Spezialfall umschließt. In den Beispielen und Übungen werden hauptsächlich reelle Folgen betrachtet.

2.1 Definitionen, Beispiele, einige Sätze

Definition 2.1: (Folgen)

Eine **Folge** $(z_n) = (z_1, z_2, z_3, \dots)$, manchmal auch $(z_n) = (z_0, z_1, z_2, \dots)$, ist eine Zuordnung (Funktion)

Index $n \in \mathbb{N}$ (bzw. \mathbb{N}_0) \longrightarrow Wert $z_n \in \mathbb{C}$.

Beispiel 2.2:

a) $x_n = (-1)^n; n \in \mathbb{N}$. Die Folge (x_n) ist $(-1, 1, -1, 1, \dots)$.

b) $x_n = \frac{1}{n}; n \in \mathbb{N}$. Die Folge (x_n) ist $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$.

c) $x_n = 1 - \frac{1}{n^2}; n \in \mathbb{N}$. Die Folge (x_n) ist $(0, \frac{3}{4}, \frac{8}{9}, \frac{15}{16}, \frac{24}{25}, \dots)$.

d) $x_n = (1 + \frac{1}{n})^n; n \in \mathbb{N}$. Die Folge (x_n) ist

$(2, \frac{9}{4}, \frac{64}{27}, \frac{625}{256}, \frac{7776}{3125}, \dots) \approx (2.0, 2.25, 2.3703\dots, 2.4414\dots, 2.4883\dots, \dots)$.

Beispiel 2.3: Einige simple Berechnungen mit MuPAD. Folgen können z.B. als Funktionen definiert werden:

```
>> x := n -> (1 + 1/n)^n
```

```
      n -> (1 + 1/n)^n
```

Der „Folgenerator“ \$ dient zur Erzeugung von Folgen:

```
>> x(n) $ n = 1..5
```

```
      2, 9/4, 64/27, 625/256, 7776/3125
```

Gleitpunktnäherungen werden durch float erzeugt:

```
>> float(x(n)) $ n = 1..5
```

```
      2.0, 2.25, 2.37037037, 2.44140625, 2.48832
```

Manchmal sind Monotonieeigenschaften von Folgen interessant. Da hierzu Folgliedern verglichen werden müssen, kann Monotonie nur im Reellen betrachtet werden.

Bezeichnung 2.4:

Eine reelle Folge (x_n) heißt „**monoton wachsend**“ bzw. „**monoton fallend**“, wenn $x_n \leq x_m$ bzw. $x_n \geq x_m$ gilt für alle Indexpaare n, m mit $n < m$. Bei $x_n < x_m$ bzw. $x_n > x_m$ spricht man von „**streng monoton wachsend**“ bzw. „**streng monoton fallend**“.

Zunächst die formale Definition von „Konvergenz“ und „Grenzwert“, die etwas abschreckend sein mag, aber (keine Angst!) später nur in (den hier nicht wirklich interessierenden) technischen Beweisen zum Einsatz kommt. Oft reicht es, einfache Vererbungsregeln wie z.B. aus Satz 2.13 zu benutzen, um Grenzwerte mittels Arithmetikregeln zu ermitteln.

Definition 2.5: (Grenzwerte von Folgen)

Eine Folge (z_n) in \mathbb{C} heißt „**konvergent**“, wenn eine Zahl $z^* \in \mathbb{C}$ existiert, so dass sich (intuitiv) „alle Zahlen z_n für großes n dem Wert z^* beliebig genau annähern“.

Formal: zu jedem noch so kleinen $\epsilon > 0$ läßt sich eine reelle Zahl $N(\epsilon)$ angeben, so dass $|z_n - z^*| \leq \epsilon$ gilt für alle Indizes $n \geq N(\epsilon)$.

Anschaulich: alle Werte z_n weichen für $n \geq N(\epsilon)$ maximal um den Wert ϵ vom Grenzwert ab.

Der Wert z^* heißt dann „**Grenzwert**“ („**Limes**“) der Folge (z_n) .

Schreibweisen:

$$z^* = \lim_{n \rightarrow \infty} z_n \quad \text{oder auch} \quad z_n \rightarrow z^* \text{ für } n \rightarrow \infty.$$

Eine nicht konvergierende Folge heißt „**divergent**“. Konvergente Folgen mit dem Grenzwert 0 heißen auch **Nullfolgen**.

Bemerkung 2.6: Die Aussage „für alle $n \geq N(\epsilon)$ “ impliziert, dass nur „hinreichend große Indizes n “ betrachtet zu werden brauchen. **Merke:** für Konvergenz ist das Verhalten der Folge für kleine Indexwerte völlig irrelevant. Genauer: man kann immer endlich viele Folgeelemente abändern, ohne dass sich etwas an der Konvergenz ändert: man kann o.B.d.A. (= ohne Beschränkung der Allgemeinheit) immer $N(\epsilon)$ größer wählen als der größte Index der geänderten Folgenglieder.

Eine intuitive Interpretation der ϵ -Definition der Konvergenz lautet:

Für jedes (noch so kleine) $\epsilon > 0$ haben höchstens endlich viele Folgenglieder einen Abstand zum Grenzwert, der größer ist als ϵ .

Satz 2.7: (Eindeutigkeit von Grenzwerten)

Grenzwerte sind eindeutig, d.h., zu (z_n) gibt es höchstens ein z^* mit der obigen Eigenschaft.

Beweis: Seien z^* und z^{**} zwei Grenzwerte. Zu jedem $\epsilon > 0$ gilt für hinreichend große Indizes n :

$$\begin{aligned} |z_n - z^*| &\leq \epsilon, & |z_n - z^{**}| &\leq \epsilon \\ \Rightarrow |z^* - z^{**}| &= |z^* - z_n + z_n - z^{**}| \leq |z^* - z_n| + |z_n - z^{**}| \leq 2 \cdot \epsilon. \end{aligned}$$

Da $\epsilon > 0$ beliebig klein gewählt werden kann und damit auch $2 \cdot \epsilon$ beliebig klein sein kann, folgt $|z^* - z^{**}| = 0$, also $z^* = z^{**}$.

Q.E.D.

Einige einfache Beispiele mit formalem Beweis:

Beispiel 2.8: Die konstante Folge $(z_n) = (c, c, c, \dots)$ ist konvergent mit dem Grenzwert $z^* = \lim_{n \rightarrow \infty} z_n = c$, denn für alle n gilt

$$|z_n - z^*| = |c - c| = 0 \leq \epsilon,$$

wie auch immer $\epsilon > 0$ vorgegeben wird. Formal: zu $\epsilon > 0$ wähle $N(\epsilon) = 1$.

Beispiel 2.9: Die Folge $x_n = \frac{1}{n}$ ist konvergent mit dem Grenzwert $x^* = \lim_{n \rightarrow \infty} x_n = 0$.
Formaler Beweis: zu beliebigem $\epsilon > 0$ wähle $N(\epsilon) = \frac{1}{\epsilon}$. Dann folgt für alle $n \geq N(\epsilon)$:

$$|x_n - x^*| = |x_n - 0| = |x_n| = \left| \frac{1}{n} \right| = \frac{1}{n} \leq \frac{1}{N(\epsilon)} = \frac{1}{\frac{1}{\epsilon}} = \epsilon.$$

Und noch ein Beispiel mit formalem Beweis:

Beispiel 2.10: Für $c \in \mathbb{C}$ gelte $|c| < 1$. Dann ist die Folge $z_n = c^n$ eine Nullfolge.

Beweis: Für $c = 0$ ist alles klar. Sei nun $c \neq 0$. Definiere $h = \frac{1}{|c|} - 1 > 0$, d.h., $|c| = \frac{1}{1+h}$. Es gilt

$$\frac{1}{|c^n|} = \frac{1}{|c|^n} = (1+h)^n = 1 + n \cdot h + \binom{n}{2} \cdot h^2 + \dots \geq 1 + n \cdot h > n \cdot h.$$

Es folgt $|c^n| < \frac{1}{n \cdot h} \leq \epsilon$ für alle Indizes $n \geq \frac{1}{h \cdot \epsilon} =: N(\epsilon)$.

Q.E.D.

Beispiele:

$$c = 0.5 : (c^n) = (0.5, 0.25, 0.125, 0.0625, 0.03125, \dots).$$

Für $|c| \geq 1$ gilt diese Aussage nicht! Z.B.:

$$c = 1 : (c^n) = (1, 1, 1, 1, \dots) \quad (\text{konvergiert gegen } 1),$$

$$c = i : (c^n) = (i, -1, -i, 1, i, -1, \dots) \quad (\text{konvergiert nicht}),$$

$$c = 2 : (c^n) = (2, 4, 8, 16, \dots) \quad (\text{divergiert, bzw. „konvergiert gegen } \infty\text{“}).$$

Beispiel 2.11: Einige Berechnungen mit MuPAD:

```
>> x := n -> c^n
```

```
      n -> c^n
```

```
>> x(n) $ n = 1..10
```

```
      2  3  4  5  6  7  8  9  10
      c, c , c , c , c , c , c , c , c
```

Grenzwerte werden mit `limit` berechnet. Die Hilfeseite dazu wird mittels `?limit` angefordert:

```
>> ?limit
```

Ohne Weiteres kann der Grenzwert nicht bestimmt werden, da er ja von den Eigenschaften von c abhängt:

```
>> limit(x(n), n = infinity)
```

```
Warning: cannot determine sign of ln(c) [stdlib::limit::limitMRV]
```

```
      n
      limit(c , n = infinity)
```

Nehmen wir an, c sei reell und $0 < c < 1$:

```
>> assume(0 < c < 1):
>> limit(x(n), n = infinity)
```

0

Nehmen wir an, $c > 1$:

```
>> assume(c > 1):
>> limit(x(n), n = infinity)
```

infinity

Ein Beispiel einer nicht konvergierenden Folge:

Beispiel 2.12: Die Folge $x_n = (-1)^n$, also $(x_n) = (-1, 1, -1, 1, \dots)$ ist nicht konvergent (hat keinen Grenzwert). Hier ein formaler Beweis: zu $\epsilon = \frac{1}{2}$ läßt sich kein $N(\epsilon)$ finden. Angenommen, ein Grenzwert x^* existiert. Dann müßte $N(\epsilon)$ existieren mit

$$|x_n - x^*| \leq \epsilon, \quad |x_{n+1} - x^*| \leq \epsilon$$

für alle $n \geq N(\epsilon)$. Es würde folgen:

$$|x_n - x_{n+1}| = |x_n \underbrace{-x^* + x^*}_{=0} - x_{n+1}| \leq |x_n - x^*| + |x^* - x_{n+1}| \leq \epsilon + \epsilon = \frac{1}{2} + \frac{1}{2} = 1.$$

Für die betrachtete Folge gilt aber $|x_n - x_{n+1}| = 2$ für jedes n . Widerspruch! Damit muss die Annahme „es existiert x^* “ falsch gewesen sein.

Die formale Definition mit ϵ und $N(\epsilon)$ ist unangenehm und man möchte diese recht technischen Betrachtungen und Abschätzungen liebend gern vermeiden. Wie geht man beim praktischen Rechnen vor? Es gibt Rechenregeln! Damit läßt sich ϵ und $N(\epsilon)$ oft vollständig verbannen:

Satz 2.13: (Rechenregeln für Grenzwerte)

↓28.10.03

Seien $(x_n), (y_n)$ konvergente Folgen in \mathbb{C} , sei $c \in \mathbb{C}$ eine Konstante. Dann gilt:

- a) $\lim_{n \rightarrow \infty} (c \cdot x_n) = c \cdot \lim_{n \rightarrow \infty} x_n,$
- b) $\lim_{n \rightarrow \infty} (x_n \pm y_n) = \lim_{n \rightarrow \infty} x_n \pm \lim_{n \rightarrow \infty} y_n,$
- c) $\lim_{n \rightarrow \infty} (x_n \cdot y_n) = \lim_{n \rightarrow \infty} x_n \cdot \lim_{n \rightarrow \infty} y_n,$
- d) $\lim_{n \rightarrow \infty} \left(\frac{x_n}{y_n} \right) = \frac{\lim_{n \rightarrow \infty} x_n}{\lim_{n \rightarrow \infty} y_n},$ falls $\lim_{n \rightarrow \infty} y_n \neq 0$ gilt (!),
- e) $\lim_{n \rightarrow \infty} \sqrt{x_n} = \sqrt{\lim_{n \rightarrow \infty} x_n}.$

Eine Beweisandeutung (nur für technisch Interessierte):

Beweisskizze: Seien x^* bzw. y^* die Grenzwerte von (x_n) bzw. (y_n) .

a) Für $c = 0$ ist die Behauptung sicherlich richtig. Sei nun $c \neq 0$. Zu $\epsilon > 0$ gibt es ein N , so dass

$$|x_n - x^*| \leq \frac{\epsilon}{|c|}$$

gilt für alle $n \geq N$. Für diese Indizes folgt

$$|c \cdot x_n - c \cdot x^*| = |c| \cdot |x_n - x^*| \leq |c| \cdot \frac{\epsilon}{|c|} = \epsilon.$$

b) Wähle ein beliebiges $\epsilon > 0$. Da (x_n) und (y_n) als konvergent vorausgesetzt sind, gibt es Werte N_x bzw. N_y mit

$$|x_n - x^*| \leq \frac{\epsilon}{2} \quad \text{bzw.} \quad |y_n - y^*| \leq \frac{\epsilon}{2}$$

für alle $n \geq N_x$ bzw. $n \geq N_y$. Für alle $n \geq N(\epsilon) := \max(N_x, N_y)$ folgt

$$\begin{aligned} |x_n \pm y_n - (x^* \pm y^*)| &= |x_n - x^* \pm (y_n - y^*)| \\ &\leq |x_n - x^*| + |\pm (y_n - y^*)| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Die Aussagen c) – e) lassen sich mit ähnlichen (etwas aufwendigeren) Abschätzungen beweisen.

Q.E.D.

Beispiel 2.14: Wir wissen bereits, dass $x_n = \frac{1}{n}$ eine Nullfolge ist. Durch Einsatz der Rechenregeln folgt unmittelbar:

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \frac{1}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \lim_{n \rightarrow \infty} \frac{1}{n} = 0 \cdot 0 = 0,$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \frac{1}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \lim_{n \rightarrow \infty} \frac{1}{n^2} = 0 \cdot 0 = 0,$$

usw. Durch Induktion nach k ergibt sich:

Alle Folgen der Form $x_n = \frac{1}{n^k}$ mit positiven Potenzen k sind Nullfolgen.

Manchmal muss man etwas manipulieren und umschreiben. Bei rationalen Ausdrücken in n (also $\text{Polynom}(n)/\text{Polynom}(n)$) gilt das allgemeine Rezept: ziehe in Zähler und Nenner die führende Potenz von n raus und kürze. Typischerweise verbleiben dann nur noch Nullfolgen im Ausdruck, die zu 0 werden, wenn man über die obigen Rechenregeln „den Grenzwert in den Ausdruck ’reinzieht“:

Beispiel 2.15:

$$\lim_{n \rightarrow \infty} \frac{2 \cdot n^2 + 1}{n^2 - n} = \lim_{n \rightarrow \infty} \frac{n^2 \cdot (2 + \frac{1}{n^2})}{n^2 \cdot (1 - \frac{1}{n})} = \lim_{n \rightarrow \infty} \frac{2 + \frac{1}{n^2}}{1 - \frac{1}{n}} = \frac{2 + \lim_{n \rightarrow \infty} \frac{1}{n^2}}{1 - \lim_{n \rightarrow \infty} \frac{1}{n}} = \frac{2 + 0}{1 - 0} = 2.$$

Hierbei wurde benutzt, dass wir in den Beispielen 2.9 und 2.14 bereits $1/n$ und $1/n^2$ als Nullfolgen identifiziert haben. Man sieht, mit etwas Geschick eingesetzt, machen die Rechenregeln die Berechnung von Grenzwerten oft sehr einfach. Manchmal muss man allerdings „tricksen“:

Beispiel 2.16:

$$\begin{aligned} \lim_{n \rightarrow \infty} (\sqrt{n+1} - \sqrt{n}) &= \lim_{n \rightarrow \infty} \frac{(\sqrt{n+1} - \sqrt{n}) \cdot (\sqrt{n+1} + \sqrt{n})}{\sqrt{n+1} + \sqrt{n}} = \lim_{n \rightarrow \infty} \frac{\sqrt{n+1}^2 - \sqrt{n}^2}{\sqrt{n+1} + \sqrt{n}} \\ &= \lim_{n \rightarrow \infty} \frac{(n+1) - n}{\sqrt{n+1} + \sqrt{n}} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n+1} + \sqrt{n}} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n \cdot (1 + \frac{1}{n})} + \sqrt{n}} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n} \cdot \sqrt{1 + \frac{1}{n}} + \sqrt{n}} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n} \cdot (\sqrt{1 + \frac{1}{n}} + 1)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \cdot \lim_{n \rightarrow \infty} \frac{1}{\sqrt{1 + \frac{1}{n}} + 1} = \lim_{n \rightarrow \infty} \sqrt{\frac{1}{n}} \cdot \frac{1}{\lim_{n \rightarrow \infty} (\sqrt{1 + \frac{1}{n}} + 1)} \\ &= \sqrt{\lim_{n \rightarrow \infty} \frac{1}{n}} \cdot \frac{1}{\sqrt{1 + \lim_{n \rightarrow \infty} \frac{1}{n}} + 1} = 0 \cdot \frac{1}{\sqrt{1 + 0} + 1} = 0. \end{aligned}$$

Manchmal helfen alle Rechenregeln nichts, und man muss technisch abschätzen. Eine hilfreiche Aussage liefert der folgende Satz, der nur für reelle Folgen gilt. Liegen die Folgenglieder x_n in Intervallen $[a_n, b_n]$ und konvergieren die Intervallenden gegen den selben Wert, so bleibt der Folge nichts anderes übrig, als ebenfalls gegen diesen Wert zu konvergieren (die Intervalllänge $b_n - a_n$ konvergiert gegen 0):

Satz 2.17: (Intervallschachtelung)

Seien (a_n) , (b_n) , (x_n) reelle Folgen. Die Folgen (a_n) und (b_n) mögen gegen den selben Grenzwert konvergieren. Gilt für alle hinreichend großen Indizes $a_n \leq x_n \leq b_n$, so konvergiert auch (x_n) gegen

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n.$$

Beweis: Sei x^* der Grenzwert von (a_n) und (b_n) . Die Folge $(b_n - a_n)$ ist positiv und eine Nullfolge. Zu $\epsilon > 0$ gibt es Werte N_1 bzw. N_2 mit

$$b_n - a_n = |b_n - a_n| \leq \frac{\epsilon}{2}, \quad |a_n - x^*| \leq \frac{\epsilon}{2}$$

für alle $n \geq N_1$ bzw. $n \geq N_2$. Für alle $n \geq N := \max(N_1, N_2)$ folgt

$$\begin{aligned} |x_n - x^*| &= |x_n - a_n + a_n - x^*| \leq |x_n - a_n| + |a_n - x^*| \\ &= x_n - a_n + |a_n - x^*| \leq b_n - a_n + |a_n - x^*| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Q.E.D.

Beispiel 2.18: Sei $x_n = \frac{n}{n^2+1}$. Offensichtlich gilt

$$\underbrace{0}_{a_n} \leq x_n = \frac{n}{n^2+1} \leq \frac{n}{n^2} = \underbrace{\frac{1}{n}}_{b_n}.$$

Die Intervallgrenzen $a_n = 0$ und $b_n = \frac{1}{n}$ sind beides Nullfolgen, also ist auch x_n eine Nullfolge.

Beispiel 2.19: Für positive reelle Zahlen c definieren wir $z_n = c^{1/n}$ als die positive reelle Lösung von $z^n = c$.

Fall 1: Sei $c \geq 1$. Sicherlich gilt $z_n \geq 1$. Setze $z_n = 1 + h_n$ mit $h_n \geq 0$. Es folgt

$$c = (1 + h_n)^n = 1 + n \cdot h_n + \binom{n}{2} \cdot h_n^2 + \dots \geq 1 + n \cdot h_n \quad \Rightarrow \quad 0 \leq h_n \leq \frac{c-1}{n}.$$

Damit ist h_n eine Nullfolge, also $z_n = 1 + h_n \rightarrow 1$ für $n \rightarrow \infty$.

Fall 2: Sei $0 < c \leq 1$. Sicherlich gilt $0 < z_n \leq 1$. Setze $z_n = 1/(1 + h_n)$ mit $h_n \geq 0$. Analog zu Fall 1 folgt

$$\frac{1}{c} = (1 + h_n)^n = 1 + n \cdot h_n + \binom{n}{2} \cdot h_n^2 + \dots \geq 1 + n \cdot h_n \quad \Rightarrow \quad 0 \leq h_n \leq \frac{\frac{1}{c} - 1}{n}.$$

Damit ist h_n eine Nullfolge, also $z_n = 1/(1 + h_n) \rightarrow 1$ für $n \rightarrow \infty$.

Merke:

$$\boxed{\lim_{n \rightarrow \infty} c^{1/n} = 1 \text{ für jedes reelle } c > 0.}$$

Satz und Definition 2.20:

Sei z eine beliebige komplexe Zahl. Die Folge $x_n = (1 + \frac{z}{n})^n$ konvergiert gegen einen von z abhängenden Grenzwert $x^*(z)$, der auch als e^z oder auch als $\exp(z)$ bezeichnet wird. Die Funktion $\exp : z \mapsto e^z$ heißt „**Exponential-Funktion**“. Der spezielle Grenzwert $e = e^1$ für $z = 1$ heißt „**Eulersche Zahl**“:

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n \approx 2.71828\dots$$

Der Beweis ist sehr technisch und bringt keine wirklichen Erkenntnisse. Nur der Vollständigkeit halber wird eine Teilskizze angegeben:

Beweisskizze: Wir betrachten nur den Fall $z = 1$. Man zeigt jeweils per Induktion, dass die reelle Folge $x_n = \left(1 + \frac{1}{n}\right)^n$ streng monoton wachsend in n ist und dass

$$y_n = \left(1 + \frac{1}{n-1}\right)^n = \left(1 + \frac{1}{n-1}\right)^{n-1} \cdot \left(1 + \frac{1}{n-1}\right) = x_{n-1} \cdot \left(1 + \frac{1}{n-1}\right)$$

streng monoton fallend in n ist. Da offensichtlich $x_n < y_n < y_2 = 4$ gilt, ist x_n monoton wachsend und nach oben beschränkt. Dies reicht, um die Konvergenz von (x_n) zu folgern (Satz 2.24). Zusatz: y_n ist monoton fallend und nach unten durch $y_n > x_{n-1} > x_1 = 2$ beschränkt, konvergiert also ebenfalls. Es gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} y_n &= \lim_{n \rightarrow \infty} x_{n-1} \cdot \left(1 + \frac{1}{n-1}\right) = \lim_{n \rightarrow \infty} x_{n-1} \cdot \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n-1}\right) \\ &= \lim_{n \rightarrow \infty} x_{n-1} = \lim_{n \rightarrow \infty} x_n. \end{aligned}$$

Damit liefert jedes x_n eine untere und y_n eine obere Schranke für die Eulersche Zahl, wobei das Intervall $[x_n, y_n]$, in dem sie zu finden ist, auf die Länge 0 zusammenschrumpft.

Q.E.D.

Eine technische Vorüberlegung für den Beweis des kommenden Satzes 2.22:

Technischer Hilfssatz 2.21:

Sei (z_n) eine komplexe Nullfolge mit der Eigenschaft, dass $(n^2 \cdot z_n)$ beschränkt ist (d.h., es gibt eine Konstante $c > 0$, so dass für alle Indizes n $|z_n| \leq \frac{c}{n^2}$ gilt). Dann gilt

$$\lim_{n \rightarrow \infty} (1 + z_n)^n = 1.$$

Der Beweis ist sehr technisch und bringt keine wirklichen Erkenntnisse. Er ist nur der Vollständigkeit halber angegeben:

Beweis: Es gilt (Aufgabe 12) für jedes $z \in \mathbb{C}$:

$$z^n - 1 = (z - 1) \cdot (1 + z + \cdots + z^{n-1}).$$

Für $z = 1 + z_n$ folgt

$$(1 + z_n)^n - 1 = z_n \cdot (1 + (1 + z_n) + \cdots + (1 + z_n)^{n-1}).$$

Die Dreiecksungleichung liefert

$$\begin{aligned} \left| (1 + z_n)^n - 1 \right| &\leq |z_n| \cdot (1 + |1 + z_n| + \cdots + |1 + z_n|^{n-1}) \\ &\leq |z_n| \cdot (1 + (1 + |z_n|) + \cdots + (1 + |z_n|)^{n-1}). \end{aligned}$$

Mit $(1 + |z_n|)^k \leq (1 + |z_n|)^n$ für alle $k = 0, 1, \dots, n - 1$ folgt

$$\begin{aligned} \left| (1 + z_n)^n - 1 \right| &\leq |z_n| \cdot ((1 + |z_n|)^n + (1 + |z_n|)^n + \cdots + (1 + |z_n|)^n) \\ &= n \cdot |z_n| \cdot (1 + |z_n|)^n. \end{aligned}$$

Wegen $|z_n| \leq c/n^2$ gilt auch $|z_n| \leq c/n$:

$$\left| (1 + z_n)^n - 1 \right| \leq n \cdot |z_n| \cdot \left(1 + \frac{c}{n}\right)^n \leq n \cdot |z_n| \cdot e^c.$$

Wegen $|z_n| \leq c/n^2$ gilt $|n \cdot z_n| \leq c/n$:

$$\left| (1 + z_n)^n - 1 \right| \leq \frac{c \cdot e^c}{n}.$$

Damit gilt $\lim_{n \rightarrow \infty} ((1 + z_n)^n - 1) = 0$, und es folgt $\lim_{n \rightarrow \infty} (1 + z_n)^n = 1$.

Q.E.D.

Der folgende Satz ist fundamental und sehr wichtig:

29.10.03↓

Satz 2.22: (Funktionalgleichungen der Exponentialfunktion)

Für alle $z, z_1, z_2 \in \mathbb{C}$, $n \in \mathbb{Z}$ gilt:

$$e^0 = 1, \quad e^{-z} = \frac{1}{e^z}, \quad e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}, \quad (e^z)^n = e^{n \cdot z}.$$

Trotz aller Wichtigkeit des Satzes: der Beweis bringt keine wirklichen Erkenntnisse und ist nur der Vollständigkeit halber für technisch Interessierte angegeben:

Beweis: Es gilt

$$\begin{aligned} & \left(1 + \frac{z_1 + z_2}{n}\right)^n \cdot \left(1 - \frac{z_1}{n}\right)^n \cdot \left(1 - \frac{z_2}{n}\right)^n \\ &= \left(\left(1 + \frac{z_1 + z_2}{n}\right) \cdot \left(1 - \frac{z_1}{n}\right) \cdot \left(1 - \frac{z_2}{n}\right)\right)^n \\ &= \left(1 - \frac{z_1^2 + z_1 \cdot z_2 + z_2^2}{n^2} + \frac{z_1 \cdot z_2 \cdot (z_1 + z_2)}{n^3}\right)^n. \end{aligned}$$

Die Folge

$$\begin{aligned} x_n &= -\frac{z_1^2 + z_1 \cdot z_2 + z_2^2}{n^2} + \frac{z_1 \cdot z_2 \cdot (z_1 + z_2)}{n^3} \\ &= \frac{1}{n^2} \cdot \left(z_1^2 + z_1 \cdot z_2 + z_2^2 + \frac{z_1 \cdot z_2 \cdot (z_1 + z_2)}{n}\right) \end{aligned}$$

erfüllt die im Hilfssatz 2.21 geforderte Bedingung $|x_n| \leq c/n^2$ mit

$$\begin{aligned} \left|z_1^2 + z_1 \cdot z_2 + z_2^2 + \frac{z_1 \cdot z_2 \cdot (z_1 + z_2)}{n}\right| &\leq |z_1^2 + z_1 \cdot z_2 + z_2^2| + \frac{|z_1 \cdot z_2 \cdot (z_1 + z_2)|}{n} \\ &\leq |z_1^2 + z_1 \cdot z_2 + z_2^2| + |z_1 \cdot z_2 \cdot (z_1 + z_2)| =: c. \end{aligned}$$

Hilfssatz 2.21 liefert damit

$$\lim_{n \rightarrow \infty} \left(1 + \frac{z_1 + z_2}{n}\right)^n \cdot \left(1 - \frac{z_1}{n}\right)^n \cdot \left(1 - \frac{z_2}{n}\right)^n = \lim_{n \rightarrow \infty} (1 + x_n)^n = 1$$

und folglich gilt:

$$\begin{aligned} 1 &= \lim_{n \rightarrow \infty} \left(1 + \frac{z_1 + z_2}{n}\right)^n \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{z_1}{n}\right)^n \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{z_2}{n}\right)^n \\ &= e^{z_1 + z_2} \cdot e^{-z_1} \cdot e^{-z_2}. \end{aligned}$$

Mit $e^0 = \lim_{n \rightarrow \infty} \left(1 + \frac{0}{n}\right)^n = \lim_{n \rightarrow \infty} 1^n = 1$ folgt für $z_1 = z, z_2 = -z$:

$$1 = e^0 \cdot e^{-z} \cdot e^z = e^{-z} \cdot e^z \quad \Rightarrow \quad e^{-z} = \frac{1}{e^z}.$$

Für allgemeine z_1, z_2 folgt dann:

$$1 = e^{z_1 + z_2} \cdot e^{-z_1} \cdot e^{-z_2} \quad \Rightarrow \quad e^{z_1 + z_2} = e^{z_1} \cdot e^{z_2}.$$

Hiermit folgt für $n \in \mathbb{N}$:

$$(e^z)^n = e^z \cdot e^z \cdot \dots \cdot e^z = e^{z+z+\dots+z} = e^{n \cdot z}.$$

Mit $e^{-z} = 1/e^z$ folgt die selbe Eigenschaft auch für negative ganzzahlige Potenzen n .

Q.E.D.

Beispiel 2.23: Einige Rechnungen mit MuPAD. Die Exponentialfunktion heißt `exp`:

```
>> limit((1 + 1/n)^n, n = infinity)
```

```
exp(1)
```

Mit `%` wird auf den letzten Wert zugegriffen:

```
>> float(%)
```

```
2.718281829
```

```
>> exp(20) = float(exp(20))
```

```
exp(20) = 485165195.4
```

```
>> exp(3 + I/2) = float(exp(3 + I/2))
```

```
exp(3 + 1/2 I) = 17.62671695 + 9.629519358 I
```

Für reelle Argumente kann die Exponentialfunktion mittels `plotfunc2d` gezeichnet werden. Falls `x` vorher einen Wert zugewiesen bekommen hatte, muss dieser zunächst mittels `delete` gelöscht werden:

```
>> delete x:
```

```
>> plotfunc2d(exp(x), x = -2..3)
```

2.2 Weitere Konvergenzsätze

In diesem Abschnitt werden einige allgemeine Sätze formuliert, die hilfreich sind, die Konvergenz von Folgen zu prüfen.

2.2.1 Konvergenz monotoner reeller Folgen

Die folgende Aussage ist äußerst hilfreich, denn sie garantiert Konvergenz, ohne dass der konkrete Grenzwert bekannt zu sein braucht. Die Aussage beruht auf Monotonie und ist daher nur auf reelle Folgen anwendbar. Die Konvergenz basiert auf dem Supremumsaxiom 1.10 für \mathbb{R} .

Satz 2.24: (Konvergenz monotoner Folgen)

Sei (x_n) eine monoton steigende bzw. fallende reelle Folge. Ist die Folge nach oben bzw. unten beschränkt, also $x_n \leq M$ bzw. $m \leq x_n$ für alle Indizes n , so ist (x_n) konvergent. Es gilt

$$\lim_{n \rightarrow \infty} x_n = \sup \{x_n; n \in \mathbb{N}\} \leq M \quad \text{bzw.} \quad m \leq \lim_{n \rightarrow \infty} x_n = \inf \{x_n; n \in \mathbb{N}\}.$$

Beweis: Betrachte eine monoton steigende und durch M nach oben beschränkte Folge (x_n) . Setze $A = \{x_n; n \in \mathbb{N}\}$. Der gesuchte Grenzwert ist $x^* = \sup A$. Zum Beweis der Konvergenz gegen x^* sei $\epsilon > 0$ beliebig vorgegeben. Da x^* die *kleinste* obere Schranke von A ist, ist $x^* - \epsilon$ keine obere Schranke, d.h., es existiert ein Index $N(\epsilon)$ mit $x_{N(\epsilon)} > x^* - \epsilon$. Wegen der Monotonie gilt für alle Indizes $n \geq N(\epsilon)$:

$$x^* \geq x_n \geq x_{N(\epsilon)} \geq x^* - \epsilon \quad \Rightarrow \quad 0 \leq x^* - x_n \leq \epsilon \quad \Rightarrow \quad |x^* - x_n| \leq \epsilon.$$

Da x^* die kleinste obere Schranke von A ist, gilt für die obere Schranke M die Ungleichung $x^* \leq M$.

Die Konvergenz monoton fallender, nach unten beschränkter Folgen ist analog zu beweisen.

Q.E.D.

Beispiel 2.25: Betrachte

$$x_n = \sum_{k=0}^n \frac{1}{k!} = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}.$$

Offensichtlich ist (x_n) monoton steigend und nach oben beschränkt:

$$x_n = 1 + \underbrace{\frac{1}{1!}}_{=\frac{1}{2^0}} + \underbrace{\frac{1}{2!}}_{=\frac{1}{2^1}} + \underbrace{\frac{1}{3!}}_{<\frac{1}{2^2}} + \cdots + \underbrace{\frac{1}{n!}}_{<\frac{1}{2^{n-1}}} \leq 1 + \sum_{k=0}^{n-1} \frac{1}{2^k} = 1 + \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} \leq 3.$$

Die Folge konvergiert gegen einen Grenzwert ≤ 3 (es ist die Eulersche Zahl 2.71828...).

2.2.2 Cauchy-Folgen

Wir betrachten einige Aussagen, die sowohl in \mathbb{R} als auch allgemeiner in \mathbb{C} gelten. Zunächst wird der Zusammenhang zwischen reeller und komplexer Konvergenz von Folgen durch den folgenden Satz aufgedeckt:

Satz 2.26: (Komplexe und reelle Konvergenz)

Sei $z_n = x_n + i \cdot y_n \in \mathbb{C}$ mit $x_n, y_n \in \mathbb{R}$. Die Folge (z_n) konvergiert dann und genau dann gegen $z^* = x^* + i \cdot y^*$ ($x^*, y^* \in \mathbb{R}$), wenn Real- und Imaginärteil einzeln konvergieren: $\lim_{n \rightarrow \infty} x_n = x^*$, $\lim_{n \rightarrow \infty} y_n = y^*$.

Beweis: Es gilt $|z_n - z^*|^2 = (x_n - x^*)^2 + (y_n - y^*)^2$.

Gilt $(x_n) \rightarrow x^*$ und gleichzeitig $(y_n) \rightarrow y^*$, so ist $|z_n - z^*|^2$ eine Nullfolge, also auch $|z_n - z^*|$. Dies ist per Definition die Konvergenz $(z_n) \rightarrow z^*$.

Umgekehrt, es gelte $(z_n) \rightarrow z^*$. Mit

$$0 \leq |x_n - x^*| \leq |z_n - z^*|, \quad 0 \leq |y_n - y^*| \leq |z_n - z^*|$$

folgt mit Satz 2.17 unmittelbar, dass $|x_n - x^*|$ und $|y_n - y^*|$ Nullfolgen sein müssen. Dies ist per Definition die Konvergenz $(x_n) \rightarrow x^*$, $(y_n) \rightarrow y^*$.

Q.E.D.

Die Definition der Konvergenz 2.5 benötigt die Kenntnis des Grenzwerts. Der folgende Satz 2.28 ist eine Existenzaussage, mit der auch ohne Kenntnis des konkreten Grenzwerts die Konvergenz abgelesen werden kann. Zunächst die entscheidende Begriffsbildung:

Definition 2.27: (Cauchy-Folgen)

Eine Folge (z_n) in \mathbb{C} heißt „**Cauchy-Folge**“, wenn zu jedem $\epsilon > 0$ eine reelle Zahl $N(\epsilon)$ existiert, so dass für alle $n, m \geq N(\epsilon)$ gilt: $|z_n - z_m| \leq \epsilon$.

Satz 2.28: (Die konvergenten Folgen sind die Cauchy-Folgen)

Eine Folge in \mathbb{C} konvergiert dann und genau dann, wenn sie eine Cauchy-Folge ist.

Der Beweis ist technisch und bringt keine wirklichen Erkenntnisse. Er ist nur der Vollständigkeit halber angegeben:

Beweis: Wir betrachten zunächst Folgen (x_n) in \mathbb{R} .

Konvergenz \Rightarrow Cauchy-Folge: Ist (x_n) konvergent mit Grenzwert x^* , so gibt es zu $\epsilon > 0$ ein $N(\epsilon)$, so dass

$$|x_n - x^*| \leq \epsilon, \quad |x_m - x^*| \leq \epsilon$$

gilt für alle $n, m \geq N(\epsilon)$. Für alle $n, m \geq N(\epsilon/2)$ folgt

$$|x_n - x_m| = |x_n - x^* + x^* - x_m| \leq |x_n - x^*| + |x^* - x_m| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

d.h., (x_n) ist eine Cauchy-Folge.

Cauchy-Folge \Rightarrow Konvergenz: Sei (x_n) eine Cauchy-Folge mit $|x_n - x_m| \leq \epsilon$ für alle $n, m \geq N_x(\epsilon)$. Hieraus folgt, dass die Menge $A_n = \{x_m; m \geq n\}$ für jedes n beschränkt ist:

$$|x_m| = |x_m - x_n + x_n| \begin{cases} \leq |x_n| + |x_m - x_n|, \\ \geq |x_n| - |x_m - x_n|, \end{cases}$$

wobei z.B. $|x_m - x_n| \leq 1$ für $m, n \geq N_x(1)$ gilt. Man kann also definieren:

$$a_n := \inf \{x_m; m \geq n\}, \quad b_n := \sup \{x_m; m \geq n\}.$$

Offensichtlich gilt $a_n \leq x_n \leq b_n$. Die Folge b_n ist monoton fallend, da die Suprema immer kleinerer Mengen betrachtet werden. Analog ist die Folge a_n monoton wachsend. Nach Satz 2.24 konvergieren damit (a_n) und (b_n) gegen gewisse Grenzwerte a^* und b^* mit $a^* \leq b^*$. Wir zeigen, dass $a^* = b^*$ gilt. Angenommen, es gilt $a^* < b^*$. Betrachte $\epsilon = (b^* - a^*)/4 > 0$. Wähle ein $n \geq N_x(\epsilon)$. Da a_n als Infimum die *größte* untere Schranke von A_n ist, ist $a_n + \epsilon$ keine untere Schranke von A_n mehr: es gibt ein $m_1 \geq n$, so dass $x_{m_1} < a_n + \epsilon$. Analog gibt es ein $m_2 \geq n$, so dass $x_{m_2} > b_n - \epsilon$. Wegen der Monotonie von (a_n) und (b_n) gilt $a_n \leq a^*$ und $b_n \geq b^*$, also:

$$b_n - a_n \geq b^* - a^* = 4 \cdot \epsilon.$$

Damit folgt

$$x_{m_2} - x_{m_1} = \underbrace{x_{m_2} - b_n}_{>-\epsilon} + \underbrace{b_n - a_n}_{\geq 4\epsilon} + \underbrace{a_n - x_{m_1}}_{>-\epsilon} \geq -\epsilon + 4 \cdot \epsilon - \epsilon > \epsilon.$$

Hierbei gilt $m_1, m_2 \geq n \geq N_x(\epsilon)$. Mit der Cauchy-Eigenschaft von (x_n) müßte für solche Indizes aber $|x_{m_2} - x_{m_1}| \leq \epsilon$ gelten. Widerspruch!

Damit ist gezeigt, dass *reelle* Folgen (x_n) genau dann konvergieren, wenn sie Cauchy-Folgen sind. Analog zu Satz 2.26 ist leicht gezeigt, dass eine komplexe Folge genau dann eine Cauchy-Folge ist, wenn die Folgen der Real- und Imaginärteile beide Cauchy-Folgen sind. Zusammen mit Satz 2.26 ergibt sich damit, dass auch komplexe Folgen genau dann konvergieren, wenn sie Cauchy-Folgen sind.

Q.E.D.

Einige weitere Begriffe:

Definition 2.29: (Häufungspunkte von Mengen)

Ein Punkt $z^* \in \mathbb{C}$ heißt „**Häufungspunkt**“ einer Menge $A \subset \mathbb{C}$, wenn für jedes $\epsilon > 0$ die sogenannte „ **ϵ -Umgebung**“ von z^*

$$\bar{U}_\epsilon(z^*) = \{z \in \mathbb{C}; |z - z^*| \leq \epsilon\}$$

mindestens einen Punkt in A enthält: $\bar{U}_\epsilon(z^*) \cap A \neq \emptyset$.

Geometrisch ist die ϵ -Umgebung eines Punktes z^* ein Kreisgebiet mit Mittelpunkt z^* und Radius ϵ , wobei in der obigen Definition der Kreisrand

$$\{z \in \mathbb{C}; |z - z^*| = \epsilon\}$$

mit zur Umgebung gerechnet wird.

Beispiel 2.30: Offensichtlich ist jeder Punkt in A ein Häufungspunkt von A (denn dieser Punkt liegt in jeder ϵ -Umgebung von sich selbst). Es kann aber auch Punkte außerhalb von A geben, die Häufungspunkte von A sind. In \mathbb{R} sind z.B. die Endpunkte von Intervallen stets Häufungspunkte, selbst wenn das Intervall $A = (a, b) \subset \mathbb{R}$ offen ist. Z.B.: offensichtlich liegt für $\epsilon > 0$ der Punkt $x = \min((a + b)/2, a + \epsilon)$ sowohl in $A = (a, b)$ als auch in der ϵ Umgebung von a . Also ist a ein Häufungspunkt von A .

Definition 2.31: (abgeschlossene Mengen)

Eine Menge $A \in \mathbb{C}$ heißt „**abgeschlossen**“, wenn jeder ihrer Häufungspunkte in A liegt. Die Menge heißt „**offen**“, wenn ihr Komplement $\mathbb{C} \setminus A = \{z \in \mathbb{C}; z \notin A\}$ abgeschlossen ist.

Beispiel 2.32: Abgeschlossene Intervalle $[a, b]$ in \mathbb{R} sind abgeschlossene Mengen. Die hier definierten ϵ -Umgebungen

$$\bar{U}_\epsilon(z^*) = \{z \in \mathbb{C}; |z - z^*| \leq \epsilon\}$$

sind abgeschlossene Mengen in \mathbb{C} .

Achtung: in der Literatur werden als ϵ -Umgebungen oft die Mengen

$$U_\epsilon(z^*) = \{z \in \mathbb{C}; |z - z^*| < \epsilon\}$$

betrachtet, die den Kreisrand $\{z \in \mathbb{C}; |z - z^*| = \epsilon\}$ nicht enthalten. Diese Mengen sind **nicht abgeschlossen**, denn die Punkte des Kreisrands sind Häufungspunkte. Die U_ϵ sind offene Mengen.

30.10.03↓

2.2.3 Teilfolgen und Häufungspunkte

Neben der Konvergenz gibt es den (schwächeren) Begriff von „Teilkonvergenz“, der sich in sogenannten „Häufungspunkten von Folgen“ manifestiert.

Definition 2.33: (Häufungspunkte von Folgen)

Ein Punkt $z^* \in \mathbb{C}$ heißt „**Häufungspunkt**“ der Folge (z_n) , wenn in jeder ϵ -Umgebung von z^* unendlich viele Folgenglieder liegen, also: zu jedem $\epsilon > 0$ existieren unendlich viele Folgenglieder z_n mit $|z_n - z^*| \leq \epsilon$.

Bemerkung 2.34: Die Definitionen 2.29 (Häufungspunkte von Mengen) und 2.33 (Häufungspunkte von Folgen) sind miteinander verwandt: Ist z^* ist Häufungspunkt der Folge (z_n) , so ist z^* Häufungspunkt der Menge $\{z_n; n \in \mathbb{N}\} \subset \mathbb{C}$.

Beispiel 2.35: Die Folge $x_n = (-1)^n$, also $(x_n) = (-1, 1, -1, 1, \dots)$ hat die beiden Häufungspunkte $x_1^* = 1$ und $x_2^* = -1$.

Der Punkt $x_1^* = 1$ ist Häufungspunkt, denn für alle Folgenglieder mit geradem Index n (dies sind unendlich viele) gilt $|x_n - x_1^*| = 0 \leq \epsilon$ für jedes $\epsilon > 0$.

Der Punkt $x_2^* = -1$ ist Häufungspunkt, denn für alle Folgenglieder mit ungeradem Index n (dies sind unendlich viele) gilt $|x_n - x_2^*| = 0 \leq \epsilon$ für jedes $\epsilon > 0$.

Bemerkung 2.36: Sei $n_1 < n_2 < \dots$ eine streng monoton steigende Folge von Indizes in \mathbb{N} . Die Folge $(z_{n_i}) = (z_{n_1}, z_{n_2}, \dots)$ heißt „**Teilfolge**“ der Folge (z_n) .

z^* ist genau dann Häufungspunkt der Folge (z_n) , wenn es eine gegen z^* konvergierende Teilfolge von (z_n) gibt.

Zu einem gegebenen Häufungspunkt z^* folgt eine explizite Konstruktion einer konvergenten Teilfolge. Zu $\epsilon = 1/k$ gibt es unendlich viele Folgenglieder z_n mit $|z_n - z^*| \leq 1/k$. Wähle n_i als den ersten Folgenindex, für den der Abstand zwischen z_n und dem Häufungspunkt den Wert $1/k$ unterschreitet:

$$n_k := \min \left\{ n; n > n_{k-1} \text{ und } |z_n - z^*| \leq \frac{1}{k} \right\} \quad (n_0 := 0).$$

Nach Konstruktion gilt $|z_{n_k} - z^*| \leq \frac{1}{k}$ für alle $k = 1, 2, \dots$ und damit auch $|z_{n_j} - z^*| \leq \frac{1}{j} \leq \frac{1}{k}$ für alle $n_j \geq n_k$. Damit konvergiert (z_{n_k}) gegen z^* .

Umgekehrt, gibt es eine gegen z^* konvergente Teilfolge von (z_n) , so liegen in jeder ϵ -Umgebung von z^* alle bis auf endliche viele Glieder der Teilfolge. Also ist z^* ein Häufungspunkt von (z_n) .

Beispiel 2.37: Für die Folge $(x_n) = (-1, 1, -1, 1, \dots)$ aus Beispiel 2.35 konvergiert die Teilfolge $(x_2, x_4, \dots) = (1, 1, 1, \dots)$ gegeben den Häufungspunkt 1 und die Teilfolge $(x_1, x_3, \dots) = (-1, -1, -1, \dots)$ gegeben den Häufungspunkt -1 .

Satz 2.38:

Eine konvergente Folge besitzt genau einen Häufungspunkt: den Grenzwert.

Beweis: Für den Grenzwert z^* einer konvergierenden Folge (z_n) gibt es für jedes $\epsilon > 0$ ein $N(\epsilon)$, so dass alle Folgenglieder mit Indizes $\geq N(\epsilon)$ in der ϵ -Umgebung von z^* liegen. Damit ist der Grenzwert ein Häufungspunkt. Gäbe es einen weiteren davon verschiedenen Häufungspunkt z^{**} , gäbe es für $\epsilon = |z^{**} - z^*|/3 > 0$ mindestens einen Folgenpunkt z_n mit Index $n \geq N(\epsilon)$ in dieser

ϵ -Umgebung von z^{**} , und es würde folgen:

$$3 \cdot \epsilon = |z^* - z^{**}| = |z^* - z_n + z_n - z^{**}| \leq |z^* - z_n| + |z_n - z^{**}| \leq 2 \cdot \epsilon.$$

Widerspruch!

Q.E.D.

Satz 2.39: (Bolzano (1781–1848) und Weierstrass (1815–1897))

*Eine Folge heißt **beschränkt**, wenn die Menge aller Folgenpunkte beschränkt ist (im Komplexen: die reelle Menge $\{|z_n|; n \in \mathbb{N}\}$ ist beschränkt). Es gilt: Jede beschränkte Folge in \mathbb{R} oder \mathbb{C} besitzt mindestens einen Häufungspunkt.*

Wir verzichten auf die strenge technische Durchführung des Beweises und geben nur die (sehr einfache) Idee an:

Beweisidee: Die Folge liege innerhalb eines Quadrates in der komplexen Ebene. Zerlege dieses Quadrat in 4 gleichgrosse Teilquadrate der halben Seitenlänge. Mindestens eines der Teilquadrate enthält unendliche viele der Folgenglieder. Wähle eines dieser Teilquadrate aus und zerlege es wiederum in 4 Teilquadrate usw. Die so konstruierte Folge von Quadraten schrumpft auf einen Punkt zusammen (dies ist der Häufungspunkt), in dessen Nähe nach Konstruktion unendlich viele der Folgenpunkte existieren.

Q.E.D.

Bemerkung 2.40: Nach Bemerkung 2.36 heißt dies:

Jede beschränkte Folge besitzt eine konvergente Teilfolge.

2.3 Unendliches, uneigentliche Konvergenz

In diesem Abschnitt geht es um eine oft nützliche Schreibweise, die in der hier vorgestellten Form nur bei reellen Folgen sinnvoll ist. Die „unendlichen Werte“ $\pm\infty$ sind keine reellen Zahlen, sondern dienen nur als nützliche Abkürzungen, um gewisse Situationen zu beschreiben. Wir lassen $\pm\infty$ als („uneigentliche“) Grenzwerte reeller Folgen zu:

Definition 2.41: ($\pm\infty$ als Grenzwert)

- Eine reelle Folge (x_n) „**konvergiert (uneigentlich) gegen ∞** “, wenn die Folgenglieder jede beliebig vorgegebene Schranke $c > 0$ überschreiten: zu jedem reellen c existiert eine reelle Zahl $N(c)$, so dass $x_n \geq c$ gilt für alle Indizes $n \geq N(c)$. Schreibweise:

$$\lim_{n \rightarrow \infty} x_n = \infty .$$

- Eine reelle Folge (x_n) „**konvergiert (uneigentlich) gegen $-\infty$** “, wenn die Folgenglieder jede beliebig vorgegebene Schranke $c < 0$ unterschreiten: zu jedem reellen c existiert eine reelle Zahl $N(c)$, so dass $x_n \leq c$ gilt für alle Indizes $n \geq N(c)$. Schreibweise:

$$\lim_{n \rightarrow \infty} x_n = -\infty .$$

Beispiel 2.42: Die Folgen $x_n = n$, $x_n = n^2$, $x_n = \sqrt{n}$, $x_n = 2^n$ konvergieren gegen ∞ . Die Folgen $x_n = -n$, $x_n = -n^2$, $x_n = -\sqrt{n}$, $x_n = -(2^n)$ konvergieren gegen $-\infty$.

Beispiel 2.43: Achtung: die Folgen $x_n = (-1)^n \cdot n$ (also $(-1, 2, -3, 4, -5, \dots)$) oder auch $x_n = (-2)^n$ (also $(-2, 4, -8, 16, -32, \dots)$) konvergieren **nicht** gegen ∞ oder $-\infty$, sie divergieren!

Man darf getrost mit ∞ und $-\infty$ rechnen, wobei folgende Rechenregeln gelten:

Rechenregeln für $\pm\infty$ 2.44:

Sei c eine reelle Zahl.

- $c \pm \infty = \pm\infty$,
- $c \cdot (\pm\infty) = \pm \text{sign}(c) \cdot \infty$ für $c \neq 0$. Hierbei ist $\text{sign}(c)$ das Vorzeichen von c .
- $\frac{1}{\pm\infty} = 0$,
- $\infty + \infty = \infty$, $-\infty - \infty = -\infty$,
- $\infty \cdot \infty = (-\infty) \cdot (-\infty) = \infty$, $\infty \cdot (-\infty) = (-\infty) \cdot \infty = -\infty$,
- $\infty^\infty = \infty$, $\infty^{-\infty} = 0$,
- $c^\infty = \infty$ für $c > 1$, $c^\infty = 0$ für $0 < c < 1$,
- $c^{-\infty} = 0$ für $c > 1$, $c^{-\infty} = \infty$ für $0 < c < 1$.

Beispiel 2.45: Die Folge $x_n = n^3 + n$ konvergiert gegen ∞ :

$$\lim_{n \rightarrow \infty} (n^3 + n) = \lim_{n \rightarrow \infty} n^3 + \lim_{n \rightarrow \infty} n = \infty + \infty = \infty .$$

Aus dem obigen Ergebnis folgt sofort das nächste Ergebnis:

Beispiel 2.46: Die Folge $x_n = \frac{1}{n^3 + n}$ konvergiert gegen 0:

$$\lim_{n \rightarrow \infty} \frac{1}{n^3 + n} = \frac{1}{\lim_{n \rightarrow \infty} (n^3 + n)} = \frac{1}{\infty} = 0 .$$

Beim Rechnen mit $\pm\infty$ muss man aber etwas Vorsicht walten lassen. Wenn man auf eine der folgenden Situationen stößt, darf man nicht weiterrechnen, sondern muss die betrachteten Grenzwerte anders ermitteln:

Undefinierte Ergebnisse beim Rechnen mit $\pm\infty$ 2.47:

- $0 \cdot (\pm\infty) = \text{„undefiniert“}$,
- $\infty - \infty = \text{„undefiniert“}$, $-\infty + \infty = \text{„undefiniert“}$,
- $\frac{\pm\infty}{\pm\infty} = \text{„undefiniert“}$,
- $c^\infty = \text{„undefiniert“}$ für $c \leq 0$ und $c = 1$,
- $c^{-\infty} = \text{„undefiniert“}$ für $c \leq 0$ und $c = 1$,
- $\frac{1}{0} = \text{„undefiniert“}$.

Beispiel 2.48: Betrachte die Folge $x_n = n^3 - n$:

$$\lim_{n \rightarrow \infty} (n^3 - n) \stackrel{(\text{??})}{=} \lim_{n \rightarrow \infty} n^3 - \lim_{n \rightarrow \infty} n \stackrel{(\text{??})}{=} \infty - \infty \stackrel{(\text{??})}{=} \text{„undefiniert“}.$$

Dies heißt nicht, dass kein Grenzwert existiert, sondern nur, dass wir den Grenzwert über die Rechenregeln mit $\pm\infty$ nicht berechnen können. Man muss in einem solchen Fall genauer untersuchen. Z.B funktioniert folgendes Argument:

$$\begin{aligned} \lim_{n \rightarrow \infty} (n^3 - n) &= \lim_{n \rightarrow \infty} n^3 \cdot \left(1 - \frac{1}{n^2}\right) = \lim_{n \rightarrow \infty} n^3 \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n^2}\right) \\ &= \infty \cdot \left(1 - \frac{1}{\lim_{n \rightarrow \infty} n^2}\right) = \infty \cdot \left(1 - \frac{1}{\infty}\right) = \infty \cdot (1 - 0) = \infty. \end{aligned}$$

Ein weiteres solches Beispiel:

Beispiel 2.49: Betrachte die Folge $x_n = \frac{2 \cdot n^3 + n}{n^4 + 1}$:

$$\lim_{n \rightarrow \infty} \frac{2 \cdot n^3 + n}{n^4 + 1} \stackrel{(\text{??})}{=} \frac{\lim_{n \rightarrow \infty} (2 \cdot n^3 + n)}{\lim_{n \rightarrow \infty} (n^4 + 1)} \stackrel{(\text{??})}{=} \frac{\infty}{\infty} \stackrel{(\text{??})}{=} \text{„undefiniert“}.$$

Dies sagt wiederum gar nichts darüber aus, ob ein Grenzwert existiert oder nicht. In diesem Fall führt wieder ein wenig Manipulation zum Erfolg:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{2 \cdot n^3 + n}{n^4 + 1} &= \lim_{n \rightarrow \infty} \frac{n^3 \cdot \left(2 + \frac{1}{n^2}\right)}{n^4 \cdot \left(1 + \frac{1}{n^4}\right)} = \lim_{n \rightarrow \infty} \frac{2 + \frac{1}{n^2}}{n \cdot \left(1 + \frac{1}{n^4}\right)} \\ &= \frac{2 + \frac{1}{\infty}}{\infty \cdot \left(1 + \frac{1}{\infty}\right)} = \frac{2}{\infty} = 0. \end{aligned}$$

Kapitel 3

Reihen

↓4.11.03

Es geht nun um spezielle Folgen, deren Glieder durch Summation entstehen. Für diese „Reihen“ gibt es spezielle Konvergenzkriterien.

3.1 Definitionen, Beispiele, Sätze

Definition 3.1: (Reihen)

Die einer komplexen Folge (z_n) zugeordnete „**Reihe**“ ist die Folge (S_n) der „**Partialsommen**“

$$S_n = \sum_{k=1}^n z_k.$$

Existiert ein Grenzwert der Partialsommen, so nennt man ihn den Wert der „**unendlichen Reihe**“ und schreibt auch $\lim_{n \rightarrow \infty} \sum_{k=1}^n z_k = \sum_{k=1}^{\infty} z_k$. Eine

Reihe heißt „**absolut konvergent**“, wenn der Grenzwert $\sum_{k=1}^{\infty} |z_k|$ existiert.

Merke: Reihenkonvergenz ist Folgenkonvergenz, nämlich Konvergenz der Folge der Partialsommen S_n .

Beispiel 3.2: Die sogenannte „arithmetische Reihe“ besitzt eine explizite Darstellung:

$$\begin{aligned}
 S_n &= \sum_{k=1}^n k = 1 + 2 + \cdots + (n-1) + n \\
 &= \frac{1}{2} \cdot \left(\begin{array}{ccccccc} 1 & + & 2 & + & \cdots & + & (n-1) & + & n \\ & & & & & & n & + & (n-1) & + & \cdots & + & 2 & + & 1 \end{array} \right) \\
 &= \frac{1}{2} \cdot \underbrace{\left((n+1) + (n+1) + \cdots + (n+1) + (n+1) \right)}_{n \text{ Summanden}} \\
 &= \frac{1}{2} \cdot n \cdot (n+1).
 \end{aligned}$$

Halten wir fest:

$$\boxed{\sum_{k=1}^n k = \frac{n \cdot (n+1)}{2}.}$$

Die arithmetische Reihe konvergiert damit uneigentlich gegen ∞ .

Beispiel 3.3: Sei $z \in \mathbb{C}$. Eine „geometrische Reihe“ ist von der Form

$$S_n = 1 + z + z^2 + \cdots + z^n = \sum_{k=0}^n z^k.$$

Auch in diesem Fall kann man eine explizite Formel für S_n angeben:

$$\boxed{S_n = \sum_{k=0}^n z^k = \frac{1 - z^{n+1}}{1 - z} = \frac{z^{n+1} - 1}{z - 1} .}$$

Dies ist leicht nachzuvollziehen:

$$\begin{aligned}
 (1-z) \cdot S_n &= (1-z) \cdot (1 + z + z^2 + \cdots + z^n) \\
 &= \begin{array}{l} 1 \cdot (1 + z + z^2 + \cdots + z^n) \\ -z \cdot (1 + z + z^2 + \cdots + z^n) \end{array} \\
 &= \begin{array}{l} 1 + z + z^2 + \cdots + z^n \\ -z - z^2 - \cdots - z^n - z^{n+1} \end{array} \\
 &= \begin{array}{l} 1 \\ \cdots \phantom{z^{n+1}} \end{array} - z^{n+1}.
 \end{aligned}$$

Mit der expliziten Summenformel ist die Konvergenz geometrischer Reihen leicht zu überprüfen. Für $|z| < 1$ konvergiert z^{n+1} gegen 0 (Beispiel 2.10):

$$\boxed{\sum_{k=0}^{\infty} z^k = \frac{1}{1-z} \quad \text{für } |z| < 1 .}$$

Diese Reihe konvergiert absolut, denn mit denselben Argumenten konvergiert

$$\sum_{k=0}^{\infty} |z|^k = \frac{1}{1-|z|} \quad \text{für } |z| < 1 .$$

Beispiel 3.4: Einige Berechnungen mit MuPAD. Für die symbolische Berechnung von Summen ist die Funktion `sum` zuständig:

```
>> sum(k, k = 1..n)
```

$$\frac{n(n+1)}{2}$$

Durch Faktorisierung mittels `factor` ergibt sich oft eine einfachere Form:

```
>> factor(%)
```

$$1/2 n (n + 1)$$

Die geometrische Reihe:

```
>> sum(z^k, k = 0..n)
```

$$\frac{z^{n+1} - 1}{z - 1}$$

```
>> assume(0 < z < 1):
```

```
>> sum(z^k, k = 0..infinity)
```

$$\frac{1}{z - 1}$$

Beispiel 3.5: Die periodische Dezimaldarstellung $0.\overline{d_1d_2d_3}$ mit Dezimalziffern $d_k \in \{0, 1, \dots, 9\}$ steht für $0.d_1d_2d_3d_1d_2d_3d_1d_2d_3\dots$. Solche zyklischen Dezimalentwicklungen sind rationale Zahlen. Sei $n = „d_1d_2d_3“ = d_1 \cdot 100 + d_2 \cdot 10 + d_3 \in \{0, \dots, 999\}$:

$$\begin{aligned} 0.\underbrace{d_1d_2d_3}_n \underbrace{d_1d_2d_3}_n \underbrace{d_1d_2d_3}_n \dots &= \frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \frac{d_1}{10^4} + \frac{d_2}{10^5} + \frac{d_3}{10^6} + \dots \\ &= \frac{d_1 \cdot 100 + d_2 \cdot 10 + d_3}{10^3} + \frac{d_1 \cdot 100 + d_2 \cdot 10 + d_3}{10^6} + \dots \\ &= \frac{n}{1000} + \frac{n}{1000^2} + \frac{n}{1000^3} + \dots = n \cdot \sum_{k=1}^{\infty} \frac{1}{1000^k} \end{aligned}$$

$$= n \cdot \left(\sum_{k=0}^{\infty} \frac{1}{1000^k} - 1 \right) = n \cdot \left(\frac{1}{1 - \frac{1}{1000}} - 1 \right) = n \cdot \left(\frac{1000}{999} - 1 \right) = \frac{n}{999}.$$

Bemerkung: statt der formalen Rechnung kann man die Beweisidee für die Summenformel der geometrischen Reihe explizit nachvollziehen, was in diesem Fall als "Rechenrick" sogar ganz einfach zu merken ist:

$$\begin{array}{r} 1000 \cdot x = d_1 d_2 d_3 \cdot d_1 d_2 d_3 d_1 d_2 d_3 \dots \\ - \quad x = \quad \quad \quad 0 \cdot d_1 d_2 d_3 d_1 d_2 d_3 \dots \\ \hline 999 \cdot x = \underbrace{d_1 d_2 d_3}_{n} \cdot 000\,000 \dots \end{array} \Rightarrow x = \frac{n}{999}.$$

Das Cauchy-Kriterium (Definition 2.27 und Satz 2.28) liefert folgendes Konvergenzkriterium für Reihen:

Satz 3.6: (Das Cauchy-Kriterium für Reihen)

Die Reihe $\sum_k z_k$ konvergiert genau dann, wenn es zu jedem $\epsilon > 0$ ein $N(\epsilon)$ gibt, so dass $\left| \sum_{k=n}^m z_k \right| \leq \epsilon$ gilt für alle $m \geq n \geq N(\epsilon)$.

Beweis: Nach dem Cauchy-Kriterium konvergiert die Folge $S_n = \sum_{k=1}^n z_k$, wenn es zu jedem $\epsilon > 0$ ein $N(\epsilon)$ gibt, so dass

$$|S_m - S_n| \leq \epsilon \text{ für alle } m, n \geq N(\epsilon).$$

Hierbei ist $S_m - S_n = \sum_{k=n+1}^m z_k$ für $m > n$. Ersetzt man n durch $n-1$ ergibt sich das angegebene Kriterium.

Q.E.D.

Als Folgerung ergibt sich, dass absolut konvergente Reihen konvergieren:

Satz 3.7: (Absolute Konvergenz impliziert Konvergenz)

Wenn $\sum_{k=1}^{\infty} |z_k|$ konvergiert, dann auch $\sum_{k=1}^{\infty} z_k$.

Beweis: Die Dreiecksungleichung liefert $\left| \sum_{k=n}^m z_k \right| \leq \sum_{k=n}^m |z_k|$. Erfüllt $\sum_k |z_k|$ das Cauchy-Kriterium 3.6, so ist dieses Kriterium automatisch auch für $\sum_k z_k$ erfüllt.

Q.E.D.

Nur Reihen über Nullfolgen können konvergieren:

Satz 3.8:

Wenn die Reihe $\sum_{k=1}^{\infty} z_k$ konvergiert, dann ist (z_k) eine Nullfolge.

Beweis: Das Cauchy-Kriterium 3.6 mit $m = n$ besagt, dass es zu jedem $\epsilon > 0$ ein $N(\epsilon)$ gibt, so dass

$$\left| \sum_{k=n}^n z_k \right| = |z_n| \leq \epsilon$$

für alle $n \geq N(\epsilon)$ gilt. Dies ist die Konvergenz $(z_n) \rightarrow 0$.

Q.E.D.

Dies ist kein Konvergenz- sondern ein Divergenzkriterium: bilden die Summanden keine Nullfolge, muss die Reihe divergieren!

Beispiel 3.9: Die Reihe $\sum_k \frac{k}{k+1}$ konvergiert nicht, da die Summanden $\frac{k}{k+1}$ nicht gegen 0 konvergieren (sie konvergieren gegen 1).

Die Umkehrung gilt nicht: bilden die Summanden eine Nullfolge, so kann man nicht darauf schließen, dass die Reihe konvergiert. Ein Gegenbeispiel:

Beispiel 3.10: Die sogenannte „**harmonische Reihe**“ $S_n = \sum_{k=1}^n \frac{1}{k}$ ist unbeschränkt, d.h., sie divergiert.

Beweis: Betrachte die Teilfolge (S_{2^m}) :

$$\begin{aligned} S_{2^m} &= 1 + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{> \frac{2}{2^2} = \frac{1}{2}} + \underbrace{\frac{1}{5} + \cdots + \frac{1}{8}}_{> \frac{2^2}{2^3} = \frac{1}{2}} + \cdots + \underbrace{\frac{1}{2^{m-1}+1} + \cdots + \frac{1}{2^m}}_{> \frac{2^{m-1}}{2^m} = \frac{1}{2}} \\ &\geq 1 + \underbrace{\frac{1}{2} + \cdots + \frac{1}{2}}_{m \text{ Terme}} = 1 + \frac{m}{2}. \end{aligned}$$

Da (S_n) monoton wächst, konvergiert (S_n) uneigentlich gegen ∞ .

Q.E.D.

Auf dieses Argument kommen wir später beim „Kondensationskriterium“ 3.20 zurück.

Der Satz 2.24 liefert für reelle Reihen folgendes Konvergenzkriterium:

Satz 3.11: (Reihenkonvergenz bei positiven Summanden)

Sei (x_n) eine reelle Nullfolge mit $x_n \geq 0$. Die Reihe $\sum_k x_k$ konvergiert dann und genau dann, wenn die Partialsummen $\sum_{k=1}^n x_k$ nach oben beschränkt sind.

Beweis: Wegen $x_k \geq 0$ ist die Partialsummenfolge $S_n = \sum_{k=1}^n x_k$ monoton steigend. Ist sie beschränkt, konvergiert sie nach Satz 2.24. Konvergiert sie, ist sie selbstverständlich beschränkt.

Q.E.D.

Wir erinnern an Beispiel 2.25:

Beispiel 3.12: Mit $k! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot k \geq 2^{k-1}$ sind die folgenden Partialsummen nach oben beschränkt:

$$\begin{aligned} S_n &= \sum_{k=0}^n \frac{1}{k!} = 1 + \underbrace{\frac{1}{1!}}_{=\frac{1}{2^0}} + \underbrace{\frac{1}{2!}}_{=\frac{1}{2^1}} + \underbrace{\frac{1}{3!}}_{<\frac{1}{2^2}} + \dots + \underbrace{\frac{1}{n!}}_{<\frac{1}{2^{n-1}}} \\ &\leq 1 + \sum_{k=0}^{n-1} \frac{1}{2^k} = 1 + \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} \leq 3. \end{aligned}$$

Damit konvergiert $\sum_{k=0}^{\infty} \frac{1}{k!}$ gegen einen Wert ≤ 3 (es ist die Eulersche Zahl 2.718...).

3.2 Rechenregeln und das Cauchy-Produkt

4.11.03↓

Die Rechenregeln 2.13 liefern sofort:

Satz 3.13: (Rechenregeln)

a) Wenn $\sum_k z_k$ (absolut) konvergiert, dann auch $\sum_k c \cdot z_k$ für jedes $c \in \mathbb{C}$:

$$\sum_{k=1}^{\infty} c \cdot z_k = c \cdot \sum_{k=1}^{\infty} z_k.$$

b) Wenn $\sum_k z_k$ und $\sum_k \tilde{z}_k$ (absolut) konvergieren, dann auch $\sum_k (z_k \pm \tilde{z}_k)$:

$$\sum_{k=1}^{\infty} (z_k \pm \tilde{z}_k) = \sum_{k=1}^{\infty} z_k \pm \sum_{k=1}^{\infty} \tilde{z}_k.$$

Beweis: Satz 2.13 a) + b) angewendet auf die Partialsummen.

Q.E.D.

Es macht Sinn, Reihen miteinander zu multiplizieren. Da das Distributivgesetz aus dem Produkt zweier Partialsummen eine endliche Doppelsumme

$$(a_1 + \dots + a_n) \cdot (b_1 + \dots + b_n) = \sum_{i=1}^n \sum_{j=1}^n a_i \cdot b_j$$

ergibt, muss man die Summanden zunächst gezielt zusammenfassen, um zu einer Folge von Partialsummen für das Produkt zu kommen. Wir führen eine „Anordnung“ ein, indem wir einen formalen Ordnungsparameter ζ einführen:

$$\sum_{k=1}^n a_k = \sum_{k=1}^n a_k \cdot \zeta^k \mid \zeta = 1 .$$

Das Produkt ordnen wir dann nach Potenzen von ζ :

$$\begin{aligned} & (a_1 \cdot \zeta + a_2 \cdot \zeta^2 + \dots) \cdot (b_1 \cdot \zeta + b_2 \cdot \zeta^2 + \dots) \\ &= a_1 \cdot b_1 \cdot \zeta^2 + (a_1 \cdot b_2 + a_2 \cdot b_1) \cdot \zeta^3 + (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1) \cdot \zeta^4 + \dots \end{aligned}$$

Die einzelnen ζ -Potenzen liefern die Summanden der Partialsummen. Dies motiviert die folgende Definition:

Definition 3.14: (Das Cauchy-Produkt von Reihen)

Das „**Cauchy-Produkt**“ (die „**Faltung**“) zweier Reihen $\sum_k a_k$, $\sum_k b_k$ ist die Reihe $\sum_k c_k$ mit

$$c_k = \sum_{\substack{i,j \\ i+j=k}} a_i \cdot b_j.$$

Satz 3.15: (Konvergenz des Cauchy-Produkts)

Sind die Reihen $\sum_{k=1}^{\infty} a_k$ und $\sum_{k=1}^{\infty} b_k$ absolut konvergent, so auch das Cauchy-

Produkt $\sum_{k=2}^{\infty} c_k$. Es gilt $\sum_{k=2}^{\infty} c_k = \left(\sum_{k=1}^{\infty} a_k \right) \cdot \left(\sum_{k=1}^{\infty} b_k \right)$.

Beweis: (Für technisch Interessierte) Für die Partialsummen des Cauchy-Produkts gilt

$$\begin{aligned} \sum_{k=2}^n |c_k| &= \sum_{k=2}^n \left| \sum_{\substack{i,j \\ i+j=k}} a_i \cdot b_j \right| \leq \sum_{k=2}^n \sum_{\substack{i,j \\ i+j=k}} |a_i| \cdot |b_j| \\ &\leq \sum_{\substack{i,j \\ i \leq n, j \leq n}} |a_i| \cdot |b_j| = \left(\sum_{i=1}^n |a_i| \right) \cdot \left(\sum_{j=1}^n |b_j| \right). \end{aligned}$$

Da die rechte Seite für $n \rightarrow \infty$ konvergiert, ist die rechte Seite beschränkt. Nach Satz 3.11 liefert dies die Konvergenz von $\sum_k |c_k|$, also die absolute Konvergenz

des Cauchy-Produkts.

Das Cauchy-Produkt konvergiert gegen das Produkt der einzelnen Summen:

$$\begin{aligned} \left| \underbrace{\left(\sum_{i=1}^{2n} a_i \right)}_{S_{2n}^{(a)}} \cdot \underbrace{\left(\sum_{j=1}^{2n} b_j \right)}_{S_{2n}^{(b)}} - \underbrace{\sum_{k=2}^{2n} c_k}_{S_{2n}^{(c)}} \right| &= \left| \sum_{i=1}^{2n} \sum_{j=1}^{2n} a_i \cdot b_j - \sum_{k=2}^{2n} \sum_{\substack{i,j \\ i+j=k}} a_i \cdot b_j \right| \\ &= \left| \sum_{\substack{i,j \\ i \leq 2n, j \leq 2n \\ i+j > 2n}} a_i \cdot b_j \right| \leq \sum_{\substack{i,j \\ i \leq 2n, j \leq 2n \\ i+j > 2n}} |a_i| \cdot |b_j| \\ &\leq \max(|a_n|, \dots, |a_{2n}|) \cdot \sum_{j=1}^{2n} |b_j| + \max(|b_n|, \dots, |b_{2n}|) \cdot \sum_{i=1}^{2n} |a_i|. \end{aligned}$$

Da (a_n) , (b_n) Nullfolgen sind, bilden die Maxima im letzten Ausdruck eine Nullfolge in n , während die Summen beschränkt sind. Für die Partialsummen folgt

$$S_{2n}^{(c)} = S_{2n}^{(a)} \cdot S_{2n}^{(b)} + \text{Nullfolge}_n,$$

was zur Behauptung $S_n^{(c)} \rightarrow (\sum_i a_i) \cdot (\sum_j b_j)$ führt.

Q.E.D.

3.3 Spezielle Konvergenzkriterien

Es gibt eine Anzahl spezieller Kriterien für die Konvergenz/Divergenz von Reihen:

Satz 3.16: (Majorantenkriterium)

Sei $\sum_{k=1}^{\infty} x_k$ eine konvergente Reihe mit reellen Summanden $x_k \geq 0$. Sei (z_n) eine komplexe Folge. Gilt für alle bis auf endlich viele Indizes $|z_k| \leq x_k$, so konvergiert die Reihe $\sum_{k=1}^{\infty} z_k$ absolut.

Bezeichnung: $\sum_k x_k$ heißt „konvergente Majorante“ für $\sum_k z_k$.

Beweis: Die Partialsummen von $\sum_k |z_k|$ sind beschränkt (o.B.d.A. nehmen wir an, $|z_k| \leq x_k$ gilt für alle Indizes):

$$\sum_{k=1}^n |z_k| \leq \sum_{k=1}^n x_k \leq \sum_{k=1}^{\infty} x_k.$$

Nach Satz 3.11 konvergiert $\sum_k |z_k|$, d.h., die Reihe $\sum_k z_k$ konvergiert absolut.

Q.E.D.

Satz 3.17: (Minorantenkriterium)

Sei $\sum_{k=1}^{\infty} x_k$ eine divergente Reihe mit reellen Summanden $x_k \geq 0$. Sei (y_n) eine reelle Folge. Gilt für alle bis auf endlich viele Indizes $x_k \leq y_k$, so divergiert die Reihe $\sum_{k=1}^{\infty} y_k$ (genauer: sie konvergiert uneigentlich gegen ∞).

Bezeichnung: $\sum_k x_k$ heißt „divergente Minorante“ für $\sum_k y_k$.

Beweis: Würde die Reihe $\sum_k y_k$ konvergieren, wäre sie eine konvergente Majorante für $\sum_k x_k$, und nach Satz 3.16 müsste $\sum_k x_k$ konvergieren.

Q.E.D.

Satz 3.18: (Quotientenkriterium)

Sei (z_n) eine komplexe Folge. Gilt für alle bis auf endlich viele Indizes $\frac{|z_{k+1}|}{|z_k|} \leq c$ mit einem Wert $c \in (0, 1)$, so konvergiert die Reihe $\sum_{k=1}^{\infty} z_k$ absolut.

Beweis: Aus der Abschätzung folgt

$$|z_k| \leq c \cdot |z_{k-1}| \leq c^2 \cdot |z_{k-2}| \leq \dots \leq c^{k-1} \cdot |z_1|.$$

Die geometrische Reihe $\sum_k |z_1| \cdot c^{k-1}$ konvergiert nach Beispiel 3.3 für $c \in (0, 1)$ und ist damit eine konvergente Majorante für $\sum_k z_k$.

Q.E.D.

Satz 3.19: (Wurzelkriterium)

Sei (z_n) eine komplexe Folge. Gilt für alle bis auf endlich viele Indizes $\sqrt[k]{|z_k|} \leq c$ mit einem Wert $c \in (0, 1)$, so konvergiert die komplexe Reihe $\sum_{k=1}^{\infty} z_k$ absolut.

Beweis: Aus der Abschätzung folgt $|z_k| \leq c^k$. Damit ist die geometrische Reihe $\sum_k c^k$ für $c \in (0, 1)$ eine konvergente Majorante für $\sum_k z_k$.

Q.E.D.

Satz 3.20: (Kondensationskriterium)

Ist (x_n) eine monoton fallende Folge positiver reeller Zahlen, so konvergiert $\sum_k x_k$ dann und genau dann, wenn die Reihe

$$\sum_{m=0}^{\infty} 2^m \cdot x_{2^m}$$

konvergiert.

Beweis: Nach Satz 3.11 ist zu entscheiden, ob die Partialsummen $S_n = \sum_{k=1}^n x_k$ beschränkt sind. Betrachte

$$\begin{aligned} S_{2^{n+1}-1} &= \underbrace{x_1}_{=1 \cdot x_1} + \underbrace{x_2 + x_3}_{\leq 2 \cdot x_2} + \underbrace{x_4 + \cdots + x_7}_{\leq 4 \cdot x_4} + \cdots + \underbrace{x_{2^n} + \cdots + x_{2^{n+1}-1}}_{\leq 2^n \cdot x_{2^n}} \\ &\leq \sum_{m=0}^n 2^m \cdot x_{2^m}. \end{aligned}$$

Konvergiert $\sum_m 2^m \cdot x_{2^m}$, so sind alle Partialsummen S_n beschränkt und $\sum_k x_k$ konvergiert. Umgekehrt gilt:

$$\begin{aligned} S_{2^n} &= x_1 + \underbrace{x_2}_{=1 \cdot x_2} + \underbrace{x_3 + x_4}_{\geq 2 \cdot x_4} + \underbrace{x_5 + \cdots + x_8}_{\geq 4 \cdot x_8} + \cdots + \underbrace{x_{2^{n-1}+1} + \cdots + x_{2^n}}_{\geq 2^{n-1} \cdot x_{2^n}} \\ &\geq x_1 + \sum_{m=1}^n 2^{m-1} \cdot x_{2^m} \geq \frac{1}{2} \cdot \sum_{m=1}^n 2^m \cdot x_{2^m}. \end{aligned}$$

Divergiert $\sum_m 2^m \cdot x_{2^m}$, so sind alle Partialsummen unbeschränkt und $\sum_k x_k$ divergiert.

Q.E.D.

Hier eine Reihe von Beispielen zu den diversen Kriterien:

Beispiel 3.21: Wir setzen das Majorantenkriterium 3.16 ein, um zu zeigen, dass die Reihe $\sum_{k=1}^{\infty} \frac{1}{k^2}$ konvergiert. Dazu machen wir eine Anleihe beim kommenden Beispiel 3.31, wo die Konvergenz

$$\sum_{k=1}^{\infty} \frac{1}{k \cdot (k+1)} = 1$$

gezeigt wird. Dazu schätzen wir $x_k = \frac{1}{k^2}$ gegen $\tilde{y}_k = \frac{1}{k \cdot (k+1)}$ ab ($\sum_k \tilde{y}_k$ soll als konvergente Majorante für $\sum_k x_k$ dienen). Es gilt zwar nicht unmittelbar $|x_k| = x_k \leq \tilde{y}_k$, aber mit

$$k^2 \geq \frac{k \cdot (k+1)}{2}$$

($\Leftrightarrow 2 \cdot k^2 \geq k^2 + k \Leftrightarrow k^2 \geq k$; dies ist für alle $k \geq 1$ erfüllt) folgt

$$x_k = \frac{1}{k^2} \leq \frac{2}{k \cdot (k+1)} = 2 \cdot \tilde{y}_k =: y_k.$$

Mit Beispiel 3.31 folgt

$$\sum_{k=1}^{\infty} x_k \leq \sum_{k=1}^{\infty} y_k = \sum_{k=1}^{\infty} \frac{2}{k \cdot (k+1)} = 2.$$

Das Majorantenkriterium garantiert hiermit die Konvergenz von $\sum_k \frac{1}{k^2}$. Welchen Wert diese Reihe hat, haben wir damit allerdings nicht herausbekommen.

Beispiel 3.22: Die in Beispiel 3.21 betrachtete Summe wird mit MuPAD berechnet:

```
>> sum(1/k^2, k = 1..infinity)
      2
      PI
      ---
      6
```

Hierbei ist $PI = \pi = 3.1415\dots$. Zur Kontrolle vergleichen wir diesen Wert mit einer langen, aber endlichen Summe:

```
>> float(%)
      1.644934067
```

```
>> sum(1.0/k^2, k = 1..1000)
      1.643934567
```

(Das passt einigermaßen.) Einige weitere Summen, z.B. $\sum_{k=1}^{\infty} \frac{k+1}{k \cdot (k+2) \cdot (k+5)}$:

```
>> sum((k + 1)/k/(k + 2)/(k + 5), k = 1..infinity)
      323/900
```

Oder auch $\sum_{k=1}^{\infty} \frac{1}{k^3}$:

```
>> sum(1/k^3, k = 1..infinity)
      zeta(3)
```

Dieser Reihenwert hat keine elementare Darstellung. Stattdessen stellt MuPAD ihn mittels der (unter Mathematikern) berühmten speziellen Funktion **zeta** (die sogenannte Riemannsche Zeta-Funktion) dar. Das nützt uns hier relativ wenig, da wir mit dieser Funktion nicht näher vertraut sind. Zumindestens kann man hiermit aber bequem Gleitpunktnäherungen berechnen:

```
>> float(%)
      1.202056903
```

Die Web-Seite mathworld.wolfram.com/RiemannZetaFunction.html liefert weitere Informationen zur Zeta-Funktion.

Beispiel 3.23: Die Reihe $\sum_{k=1}^n \frac{1}{\sqrt{k}}$ konvergiert nicht für $n \rightarrow \infty$: zwar konvergieren die Summanden $x_k = \frac{1}{\sqrt{k}}$ gegen 0, aber „nicht schnell genug“:

$$\sum_{k=1}^{100} \frac{1}{\sqrt{k}} = 18.5896\dots, \quad \sum_{k=1}^{1000} \frac{1}{\sqrt{k}} = 61.8010\dots, \quad \sum_{k=1}^{10000} \frac{1}{\sqrt{k}} = 198.5446\dots$$

Genauer gesagt: die Reihe „konvergiert uneigentlich gegen ∞ “. Beweis: die nach Beispiel 3.10 divergierende harmonische Reihe ist eine divergente Minorante:

$$\frac{1}{k} \leq \frac{1}{\sqrt{k}}.$$

Beispiel 3.24: (!! Wichtiges Beispiel !!) Betrachte die Reihe

$$\sum_{k=0}^n \frac{z^k}{k!},$$

wo z eine beliebige feste komplexe Zahl ist (beachte: $0! = 1$). Diese Reihe konvergiert nach dem Quotientenkriterium: Mit $z_k = \frac{z^k}{k!}$ ist der Quotient zweier aufeinander folgender Summanden

$$\frac{|z_{k+1}|}{|z_k|} = \frac{\frac{|z|^{k+1}}{(k+1)!}}{\frac{|z|^k}{k!}} = \frac{|z|^{k+1} \cdot k!}{|z|^k \cdot (k+1)!} = \frac{|z|^k \cdot |z| \cdot k!}{|z|^k \cdot (k+1) \cdot k!} = \frac{|z|}{k+1}.$$

Für hinreichend große k (nämlich $k \geq 2 \cdot |z|$) gilt

$$\frac{|z_{k+1}|}{|z_k|} \leq \frac{|z|}{2 \cdot |z| + 1} < \frac{|z|}{2 \cdot |z|} = \frac{1}{2} =: c < 1,$$

womit das Quotientenkriterium erfüllt ist.

Der Grenzwert heißt „**Exponentialfunktion**“ e^z bzw. $\exp(z)$. In der Tat stimmt die Reihe mit der in Satz 2.20 benutzten Definition überein (was noch zu zeigen wäre):

$$e^z = \exp(z) = \lim_{n \rightarrow \infty} \left(1 + \frac{z}{n}\right)^n = \sum_{k=0}^{\infty} \frac{z^k}{k!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \dots$$

Die Reihendarstellung der exp-Funktion bietet einige Vorteile. Für kleine Argumente z gilt z.B. die Näherung

$$\exp(z) = 1 + \underbrace{z}_{\text{klein}} + \underbrace{\frac{z^2}{2!}}_{\text{noch kleiner}} + \underbrace{\frac{z^3}{3!}}_{\text{noch viel kleiner}} + \dots \approx 1 + z + \frac{z^2}{2}.$$

Wir ersparen uns hier den Beweis für

$$\lim_{n \rightarrow \infty} \left(1 + \frac{z}{n}\right)^n = \sum_{k=0}^{\infty} \frac{z^k}{k!},$$

der einigen technischen Abschätzungsaufwand erfordert.

Beispiel 3.25: Die Reihe $\sum_{k=1}^{\infty} \frac{1}{k^p}$ konvergiert genau dann, wenn $p > 1$ gilt.

↓6.11.03

Beweis: für festes $p > 0$ ist die Folge $x_k = 1/k^p$ monoton fallend. Das Kondensationskriterium liefert die Konvergenz, wenn

$$\sum_{m=1}^{\infty} 2^m \cdot x_{2^m} = \sum_{m=1}^{\infty} 2^m \cdot \frac{1}{(2^m)^p} = \sum_{m=1}^{\infty} \frac{1}{(2^m)^{p-1}} = \sum_{m=1}^{\infty} \frac{1}{(2^{p-1})^m}$$

konvergiert. Diese geometrische Reihe konvergiert, wenn $\frac{1}{2^{p-1}} < 1$ gilt, d.h., für $p > 1$. Für $p \leq 1$ ist die harmonische Reihe $\sum_k \frac{1}{k}$ eine divergierende Minorante.

3.4 Bedingte Konvergenz, Umordnungen

Es gibt einen Spezialfall, wo die Tatsache, dass die Summanden eine Nullfolge bilden, für die Konvergenz der Reihe ausreicht: alternierende Reihen:

Satz 3.26: (Das Leibniz-Kriterium für alternierende reelle Reihen)

Eine reelle Folge (x_n) heißt „alternierend“, wenn für jeden Index x_n und x_{n+1} unterschiedliche Vorzeichen haben. Ist zusätzlich $|x_n|$ monoton fallend und (x_n) eine Nullfolge, so konvergiert die zugeordnete „alternierende Reihe“ $\sum_k x_k$.

Beweis: Fixiere ein beliebiges n . Durch Induktion nach m ist leicht zu zeigen, dass

$$\left| \sum_{k=n}^m x_k \right| \leq |x_n|$$

für alle $m \geq n$ gilt. Da $|x_n|$ eine Nullfolge bildet, ist das Cauchy-Kriterium 3.6 erfüllt.

Q.E.D.

Beispiel 3.27: Die „alternierende harmonische Reihe“

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

erfüllt das Leibniz-Kriterium und konvergiert (der Grenzwert ist $\ln(2)$).

Die alternierende harmonische Reihe konvergiert, aber sie konvergiert nicht absolut (die harmonische Reihe ist bekanntlich divergent). Für solche „bedingt“ (= nicht absolut) konvergente Reihen ist höchste Vorsicht geboten: die Reihenfolge der Summation ist wichtig!

Beobachtung 3.28:

Für die alternierende harmonische Reihe des letzten Beispiels gilt:

$$\begin{array}{r} S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \dots \\ \frac{S}{2} = \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \dots \\ \hline \frac{3 \cdot S}{2} = 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \dots \end{array}$$

Man kann sich leicht überlegen, dass in der letzten Reihe der Kehrwert jeder natürlichen Zahl genau einmal auftaucht. In der Tat erhält sie alle Summanden der alternierenden Reihe S , nur dass die Summanden anders angeordnet sind:

„Nehme 2 positive Summanden von S , dann einen negativen, dann die beiden nächsten positiven, dann den nächsten negativen usw.“

Diese Umordnung hat den Grenzwert verändert!

Erstaunlicherweise kann man durch eine geeignete Umsummation *jeden beliebigen* Grenzwert erreichen:

Satz 3.29: (Riemannscher Umordnungssatz für bedingt konvergente Reihen)

Sei $\sum_k x_k$ eine konvergierende reelle Reihe, die nicht absolut konvergiert: $\sum_k |x_k| = \infty$. Dann gibt es zu jedem $S \in \mathbb{R}$ eine bijektive Abbildung $P: \mathbb{N} \rightarrow \mathbb{N}$ (eine **“Permutation“** von \mathbb{N}), so dass

$$\sum_{k=1}^{\infty} x_{P(k)} = \lim_{n \rightarrow \infty} \sum_{k=1}^n x_{P(k)} = S$$

gilt.

Die **Beweisidee** ist sehr einfach, der Beweis ist konstruktiv. Sei

$$N^+ = \{n \in \mathbb{N}, x_n \geq 0\}, \quad N^- = \{n \in \mathbb{N}, x_n < 0\}.$$

Man überlegt sich, dass die Divergenz von $\sum_k |x_k|$ zusammen mit der Konvergenz von $\sum_k x_k$ bedeutet, dass $\sum_{k \in N^+} x_k$ gegen ∞ und $\sum_{k \in N^-} x_k$ gegen $-\infty$ konvergiert. Zu gegebenem S wähle solange Indizes in N^+ , bis die Summe über die entsprechenden positiven x_k zum ersten Mal S überschreitet (wegen $\sum_{k \in N^+} x_k = \infty$ wird dies sicherlich irgendwann geschehen). Dann wähle solange Indizes in N^- , bis durch Hinzuaddieren der entsprechenden negativen x_k zum ersten Mal S unterschritten wird (wegen $\sum_{k \in N^-} x_k = -\infty$ wird dies sicherlich irgendwann geschehen). Dann wähle wieder Indizes auf N^+ , bis S überschritten wird usw. Die Differenz zwischen S und der überschreitenden bzw. unterschreitenden Zwischensumme ist jeweils kleiner als das letzte Folgeelement, das addiert bzw. subtrahiert wurde. Die Folgenglieder sind aber eine Nullfolge, da $\sum_k x_k$ konvergiert. Damit konvergieren die konstruierten Zwischensummen gegen S . Details: siehe z.B. Chr. Blatter, Analysis 1, Springer.

Q.E.D.

Glücklicherweise ergibt sich dieses Umordnungsproblem bei absolut konvergierenden Reihen nicht. Jede Umordnung der Reihenglieder liefert die selbe Summe:

Satz 3.30: (Umordnungssatz für absolut konvergente Reihen)

Sei $\sum_k z_k$ eine (komplexe) absolut konvergierende Reihe. Dann konvergiert $\sum_k z_{P(k)}$ für jede Permutation P absolut gegen den selben Grenzwert.

Beweis: siehe z.B. Chr. Blatter, Analysis 1, Springer.

3.5 Summation per Partialbruchzerlegung

Es gibt einige Situationen, wo man (endliche) Reihen explizit berechnen kann. Der Reihenwert ergibt sich als Grenzwert des expliziten Ausdrucks:

Beispiel 3.31: Betrachte $\sum_{k=1}^n \frac{1}{k \cdot (k+1)}$. Die entscheidende Beobachtung ist:

$$\frac{1}{k \cdot (k+1)} = \frac{1}{k} - \frac{1}{k+1}$$

(man bringe $\frac{1}{k} - \frac{1}{k+1}$ auf den Hauptnenner). Hiermit ergibt sich

$$\sum_{k=1}^n \frac{1}{k \cdot (k+1)} = \sum_{k=1}^n \left(\frac{1}{k} - \frac{1}{k+1} \right) = \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \frac{1}{k+1}$$

$$\begin{aligned}
&= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \\
&\quad - \frac{1}{2} - \frac{1}{3} - \cdots - \frac{1}{n} - \frac{1}{n+1} \\
&= 1 - \frac{1}{n+1}.
\end{aligned}$$

Man nennt so eine Summe auch „Teleskopsumme“: sie läßt sich zu einigen wenigen Termen „zusammenschieben“, da sich fast alle Summanden aufheben. Es folgt:

$$\sum_{k=1}^{\infty} \frac{1}{k \cdot (k+1)} = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k \cdot (k+1)} = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n+1}\right) = 1.$$

Der im obigen Beispiel angewendete Trick läßt sich systematisch anwenden:

Rezept (Summation durch „Partialbruchzerlegung“) 3.32:

Betrachte die Summe $\sum_k a_k$ über einen „rationalen“ Ausdruck in k :

$$a_k = \frac{c_0 + c_1 \cdot k + \cdots + c_p \cdot k^p}{d_0 + d_1 \cdot k + \cdots + d_q \cdot k^q}$$

mit $p+2 \leq q$ (für $p+2 > q$ divergiert die Reihe $\sum_k a_k$).

- *Schritt 1:* Bestimme die Nullstellen k_1, k_2, \dots, k_q des Nennerpolynoms $P(k) = d_0 + d_1 \cdot k + \cdots + d_q \cdot k^q = d_q \cdot (k - k_1) \cdots (k - k_q)$.
- *Schritt 2:* Sind alle Nullstellen einfach¹, so kann man den Ausdruck stets folgendermaßen additiv zerlegen: es gibt Werte e_1, e_2, \dots, e_q , so dass

$$a_k = \frac{c_0 + c_1 \cdot k + \cdots + c_p \cdot k^p}{d_0 + d_1 \cdot k + \cdots + d_q \cdot k^q} = \frac{e_1}{k - k_1} + \frac{e_2}{k - k_2} + \cdots + \frac{e_q}{k - k_q}.$$

Finde diese Werte e_1, \dots, e_q ! Man bringt dazu die rechte Seite dieses Ansatzes auf den Hauptnenner (das ergibt nach Konstruktion das Nennerpolynom $P(k)$). Das Zählerpolynom muss mit dem Zähler der linken Seite übereinstimmen. Vergleiche in den Zählern die Koeffizienten der k -Potenzen, die einzeln übereinstimmen müssen. Dies führt zu einem (stets lösbaren) **linearen Gleichungssystem** für e_1, \dots, e_q .

- *Schritt 3:* Es gilt

$$\sum_{k=1}^n a_k = \sum_{k=1}^n \left(\frac{e_1}{k - k_1} + \frac{e_2}{k - k_2} + \cdots + \frac{e_q}{k - k_q} \right)$$

¹Bei entarteten Nullstellen muss man den Ansatz für die Partialsummenzerlegung etwas modifizieren, damit die folgenden Schritte funktionieren (man muss auf ein lösbares lineares Gleichungssystem für die unbekannt Parameter der Zerlegung stoßen). Vergleiche mit Aufgabe 12a auf Blatt 4a.

$$= \sum_{k=1}^n \frac{e_1}{k - k_1} + \sum_{k=1}^n \frac{e_2}{k - k_2} + \cdots + \sum_{k=1}^n \frac{e_n}{k - k_n}.$$

Unterscheiden sich die Nullstellen um ganze Zahlen, so lässt sich diese Summe von Summen als „Teleskopsumme“ zu einem expliziten Ausdruck in n vereinfachen.

Beispiel 3.33: Betrachte $\sum_{k=3}^n \frac{k}{k^3 - 2 \cdot k^2 - k + 2}$.

Schritt 1: Die Nullstellen des Nennerpolynoms sind $k_1 = 1$, $k_2 = -1$, $k_3 = 2$:

$$k^3 - 2 \cdot k^2 - k + 2 = (k - 1) \cdot (k + 1) \cdot (k - 2).$$

Schritt 2: Mache den Ansatz:

$$\frac{k}{k^3 - 2 \cdot k^2 - k + 2} = \frac{e_1}{k - 1} + \frac{e_2}{k + 1} + \frac{e_3}{k - 2}.$$

Bringe die rechte Seite auf den Hauptnenner und ordne den Zähler nach k -Potenzen:

$$\begin{aligned} & \frac{e_1}{k - 1} + \frac{e_2}{k + 1} + \frac{e_3}{k - 2} \\ &= \frac{e_1 \cdot (k + 1) \cdot (k - 2) + e_2 \cdot (k - 1) \cdot (k - 2) + e_3 \cdot (k - 1) \cdot (k + 1)}{(k - 1) \cdot (k + 1) \cdot (k - 2)} \\ &= \frac{e_1 \cdot k^2 - e_1 \cdot k - 2 \cdot e_1 + e_2 \cdot k^2 - 3 \cdot e_2 \cdot k + 2 \cdot e_2 + e_3 \cdot k^2 - e_3}{(k - 1) \cdot (k + 1) \cdot (k - 2)} \\ &= \frac{(e_1 + e_2 + e_3) \cdot k^2 + (-e_1 - 3 \cdot e_2) \cdot k + (-2 \cdot e_1 + 2 \cdot e_2 - e_3)}{(k - 1) \cdot (k + 1) \cdot (k - 2)}. \end{aligned}$$

Dies muss als Polynom in k mit dem Zähler der Summanden der Reihe übereinstimmen, also

$$k = 0 \cdot k^2 + 1 \cdot k + 0 = \underbrace{(e_1 + e_2 + e_3)}_{=0} \cdot k^2 + \underbrace{(-e_1 - 3 \cdot e_2)}_{=1} \cdot k + \underbrace{(-2 \cdot e_1 + 2 \cdot e_2 - e_3)}_{=0}.$$

Durch Vergleich der k -Potenzen ergibt sich:

$$e_1 + e_2 + e_3 = 0, \quad -e_1 - 3 \cdot e_2 = 1, \quad -2 \cdot e_1 + 2 \cdot e_2 - e_3 = 0.$$

Die Lösung dieses linearen Gleichungssystems ist

$$e_1 = -\frac{1}{2}, \quad e_2 = -\frac{1}{6}, \quad e_3 = \frac{2}{3},$$

also

$$\frac{k}{k^3 - 2 \cdot k^2 - k + 2} = -\frac{1}{2} \cdot \frac{1}{k - 1} - \frac{1}{6} \cdot \frac{1}{k + 1} + \frac{2}{3} \cdot \frac{1}{k - 2}.$$

Schritt 3: Reduktion der Teleskopsumme. Beachte, dass eine der Gleichungen $e_1 + e_2 + e_3 = 0$ war. Deshalb ist es kein Zufall, dass sich in der Tat eine Teleskopsumme ergibt:

$$\sum_{k=3}^n \frac{k}{k^3 - 2 \cdot k^2 - k + 2} = \sum_{k=3}^n \frac{(-1/2)}{k - 1} + \sum_{k=3}^n \frac{(-1/6)}{k + 1} + \sum_{k=3}^n \frac{(2/3)}{k - 2}$$

$$\begin{aligned}
&= \frac{(-1/2)}{2} + \frac{(-1/2)}{3} + \frac{(-1/2)}{4} + \dots + \frac{(-1/2)}{n-2} + \frac{(-1/2)}{n-1} \\
&+ \frac{(-1/6)}{4} + \dots + \frac{(-1/6)}{n-2} + \frac{(-1/6)}{n-1} + \frac{(-1/6)}{n} + \frac{(-1/6)}{n+1} \\
&+ \frac{(2/3)}{1} + \frac{(2/3)}{2} + \frac{(2/3)}{3} + \underbrace{\frac{(2/3)}{4}}_0 + \underbrace{\dots}_0 + \underbrace{\frac{(2/3)}{n-2}}_0 \\
&= \frac{(-1/2)}{2} + \frac{(-1/2)}{3} + \frac{(-1/2)}{n-1} \\
&+ \frac{(-1/6)}{n-1} + \frac{(-1/6)}{n} + \frac{(-1/6)}{n+1} \\
&+ \underbrace{\frac{(2/3)}{1} + \frac{(2/3)}{2} + \frac{(2/3)}{3}}_{\frac{29}{36}} + \underbrace{\frac{(-2/3)}{n-1} + \frac{(-1/6)}{n} + \frac{(-1/6)}{n+1}}_{\frac{(-2/3)}{n-1} + \frac{(-1/6)}{n} + \frac{(-1/6)}{n+1}}.
\end{aligned}$$

Der Grenzwert für $n \rightarrow \infty$ liefert

$$\sum_{k=3}^{\infty} \frac{k}{k^3 - 2 \cdot k^2 - k + 2} = \frac{29}{36}.$$

Beispiel 3.34: In MuPAD ist die Funktion `partfrac` („partial fraction“) für die Partialbruchzerlegung zuständig:

`>> partfrac(k/(k^3 - 2*k^2 - k + 2))`

$$\frac{2}{3(k-2)} - \frac{1}{6(k+1)} - \frac{1}{2(k-1)}$$

Kapitel 4

Funktionen und Stetigkeit

4.1 Funktionen

↓11.11.03

Definition 4.1:

Eine Funktion $f : D \mapsto \mathbb{C}$ ist eine Zuordnung $f : z \mapsto f(z)$ einer Zahl $z \in D \subset \mathbb{C}$ zu einem „Bildwert“ $f(z) \in \mathbb{C}$. Der Punkt z heißt auch „Urbild“ von $f(z)$. Die Menge $D \subset \mathbb{C}$ heißt „Definitionsbereich“, die Menge

$$f(D) := \{f(z); z \in D\}$$

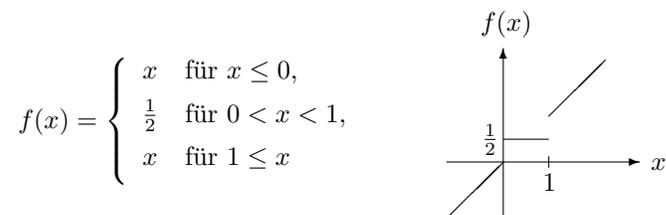
heißt „Bildbereich“ oder auch „Wertebereich“ der Funktion.

Eine reelle Funktion $f : D \subset \mathbb{R} \mapsto \mathbb{R}$ heißt

- **monoton steigend**, wenn $f(x) \leq f(y)$ gilt
- **streng monoton steigend**, wenn $f(x) < f(y)$ gilt
- **monoton fallend**, wenn $f(x) \geq f(y)$ gilt
- **streng monoton fallend**, wenn $f(x) > f(y)$ gilt

für alle $x, y \in D$ mit $x < y$.

Beispiel 4.2: a) Die (stückweise definierte) Funktion $f : \mathbb{R} \mapsto \mathbb{R}$

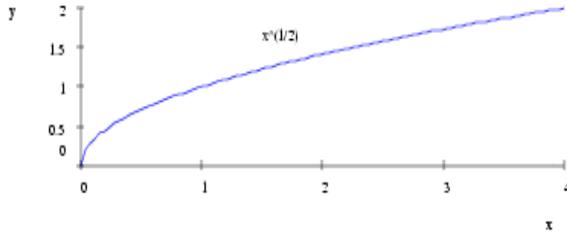


ist monoton steigend (aber nicht streng monoton steigend). Der Definitionsbereich ist \mathbb{R} , der Bildbereich ist $f(\mathbb{R}) = (-\infty, 0] \cup \left\{\frac{1}{2}\right\} \cup [1, \infty)$.

Die MuPAD-Graphik dazu (`piecewise` erzeugt stückweise definierte Funktionen):

```
>> f:= piecewise([x <= 0, x],
                 [0 < x and x < 1, 1/2],
                 [1 <= x, x])
>> plotfunc2d(f(x), x = -2..2)
```

b) Die Funktion $f : [0, \infty) \mapsto [0, \infty)$, $f(x) = \sqrt{x}$ ist streng monoton steigend. Die MuPAD-Graphik dazu (`sqrt` ist die Wurzelfunktion):



4.2 Stetigkeit

Definition 4.3: (Stetigkeit)

Eine Funktion $f : D \subset \mathbb{C} \mapsto \mathbb{C}$ heißt **stetig am Punkt** $z^* \in D$, wenn für jede gegen z^* konvergierende Folge (z_n) mit $z_n \in D$ gilt:

$$\lim_{n \rightarrow \infty} f(z_n) = f(z^*).$$

Die Funktion f heißt **stetig auf dem Bereich** D , wenn sie an allen Punkten $x^* \in D$ stetig ist.

Definition 4.3 ist ein Möglichkeit, die Stetigkeit an einem Punkt zu definieren. Hier eine andere Variante:

Definition 4.4: (Stetigkeit, Version 2)

Eine Funktion $f : D \subset \mathbb{C} \mapsto \mathbb{C}$ heißt **stetig am Punkt** $z^* \in D$, wenn für jedes $\epsilon > 0$ ein $\delta > 0$ existiert, so dass

$$|f(z) - f(z^*)| < \epsilon$$

gilt für alle z in der Menge $\{z \in D; |z - z^*| < \delta\}$.

Anschaulich: gibt man sich eine Abweichung ϵ der Funktionswerte $f(z)$ vom Wert $f(z^*)$ vor, so gibt es stets eine Umgebung von z^* , auf der die Funktionswerte nicht mehr als ϵ von $f(z^*)$ abweichen.

Satz 4.5: (Äquivalenz der Stetigkeitsdefinitionen)

Die Definitionen 4.3 und 4.4 sind äquivalent.

Beweis: (für technisch Interessierte)

4.3 \Rightarrow 4.4 Sei f im Sinne von 4.3 an der Stelle z^* stetig. Sei $\epsilon > 0$ vorgegeben. Angenommen, es gäbe kein $\delta > 0$ mit der Eigenschaft, dass $|f(z) - f(z^*)| < \epsilon$ gilt für alle z in D mit $|z - z^*| < \delta$. D.h., zu jedem $\delta > 0$ gibt es (mindestens) ein z in D mit $|z - z^*| < \delta$, für das $|f(z) - f(z^*)| \geq \epsilon$ gilt. Zu jedem $\delta = 1/n$, $n \in \mathbb{N}$, sei z_n ein solches Element. Damit haben wir eine gegen z^* konvergierende Folge (z_n) , für die nicht $f(z_n)$ gegen $f(z^*)$ konvergiert. Widerspruch zur Stetigkeit am Punkt z^* nach Definition 4.3.

4.4 \Rightarrow 4.3 Sei f im Sinne von 4.4 an der Stelle z^* stetig. Sei z_n eine beliebige gegen z^* konvergierende Folge. Es ist zu zeigen, dass $f(z_n)$ gegen $f(z^*)$ konvergiert. Zu $\epsilon > 0$ existiert nach Definition 4.4 ein δ , so dass $|f(z_n) - f(z^*)| < \epsilon$ gilt für alle z_n in D mit $|z_n - z^*| < \delta$. Da z_n gegen z^* konvergiert, gibt es zu δ ein N , so dass $|z_n - z^*| < \delta$ für alle $n \geq N$. Damit haben wir zu $\epsilon > 0$ ein $N = N(\delta(\epsilon))$ gefunden, so dass $|f(z_n) - f(z^*)| < \epsilon$ gilt für alle $n \geq N$. Also konvergiert $f(z_n)$ gegen $f(z^*)$.

Q.E.D.

Die formale Definition 4.3 der Stetigkeit sollte man sich so merken:

Merkregel 4.6:

Für beliebige konvergente Folgen z_n gilt

$$\lim_{n \rightarrow \infty} f(z_n) = f\left(\lim_{n \rightarrow \infty} z_n\right),$$

wenn die Funktion f an der Stelle $\lim_{n \rightarrow \infty} z_n$ stetig ist.

Definition 4.7: (einseitige Stetigkeit)

Für reelle Funktionen $f : D \subset \mathbb{R} \mapsto \mathbb{R}$ wird zusätzlich definiert:

Die Funktion f heißt **rechtsseitig stetig am Punkt** $x^* \in D$, wenn $\lim_{n \rightarrow \infty} f(x_n) = f(x^*)$ gilt für alle gegen x^* konvergierenden Folgen (x_n) mit $x_n \geq x^*$.

Die Funktion f heißt **linksseitig stetig am Punkt** $x^* \in D$, wenn $\lim_{n \rightarrow \infty} f(x_n) = f(x^*)$ gilt für alle gegen x^* konvergierenden Folgen (x_n) mit $x_n \leq x^*$.

Ähnlich wie die ϵ - $N(\epsilon)$ -Definition eines Grenzwertes für Folgen ist diese Definition von Stetigkeit (entweder 4.3 oder 4.4) technisch und nur in sehr einfachen Fällen praktisch handhabbar (zumindestens für Nicht-Mathematiker). Man verläßt sich in der Praxis wiederum auf Rechenregeln, mit denen Stetigkeit vererbt wird, siehe Satz 4.10. Zunächst einige einfache Beispiele mit der formalen Definition:

Beispiel 4.8: a) Betrachte die konstante Funktion $f : z \in \mathbb{C} \mapsto c$ (mit einer konstanten Zahl $c \in \mathbb{C}$). Sei (z_n) eine beliebige gegen z^* konvergierende Folge. Es gilt

$$\lim_{n \rightarrow \infty} f(z_n) = \lim_{n \rightarrow \infty} c = c = f(z^*).$$

Damit ist f an jedem Punkt $z^* \in \mathbb{C}$ stetig.

b) Betrachte die Funktion $f(z) = z$. Sei (z_n) eine beliebige gegen z^* konvergierende Folge. Es gilt

$$\lim_{n \rightarrow \infty} f(z_n) = \lim_{n \rightarrow \infty} z_n = z^* = f(z^*).$$

Damit ist f an jedem Punkt $z^* \in \mathbb{C}$ stetig.

c) Betrachte die Funktion $f(z) = z^2 + 1$. Sei (z_n) eine beliebige gegen z^* konvergierende Folge. Mit den Rechenregeln für Grenzwerte gilt

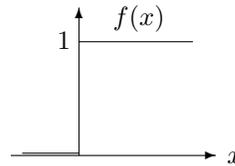
$$\lim_{n \rightarrow \infty} f(z_n) = \lim_{n \rightarrow \infty} (z_n^2 + 1) = \left(\lim_{n \rightarrow \infty} z_n \right)^2 + 1 = (z^*)^2 + 1 = f(z^*).$$

Damit ist f an jedem Punkt $z^* \in \mathbb{C}$ stetig.

Man sieht an diesen Beispielen bereits, dass die Rechenregeln für Grenzwerte sofort zu analogen Rechenregeln für die Vererbung von Stetigkeit führen. Vorher aber noch ein Beispiel zur Unstetigkeit und „einseitigen Stetigkeit“:

Beispiel 4.9: Betrachte die reelle Funktion

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



Diese Funktion ist überall stetig, außer am Punkt $x = 0$. Dort ist sie aber immer noch rechtsseitig stetig: nähert man sich dem Punkt $x = 0$ von rechts, so sind die Funktionswerte konstant 1. Der Grenzwert der Funktionswerte ist wiederum 1 und stimmt mit dem Funktionswert $f(0) = 1$ überein.

Die Funktion ist aber nicht linksseitig stetig: nähert man sich dem Punkt $x = 0$ von links, so sind die Funktionswerte konstant 0. Der Grenzwert der Funktionswerte ist wiederum 0 und stimmt **nicht** mit dem Funktionswert $f(0) = 1$ überein.

Eine stetige Funktion muss aber offensichtlich sowohl links- als auch rechtsseitig stetig sein, damit ist f am Punkt $x = 0$ unstetig.

Nun die Rechenregeln:

Satz 4.10: (Rechenregeln zur Stetigkeit)

Seien f und g Funktionen. Sei z^* ein Punkt aus dem Schnitt der Definitionsbereiche von f und g (d.h., sowohl $f(z^*)$ als auch $g(z^*)$ ist definiert). Seien f und g am Punkt z^* stetig. Sei c eine Konstante. Dann gilt:

- Die Funktion $z \mapsto c \cdot f(z)$ ist am Punkt z^* stetig.
- Die Funktion $z \mapsto f(z) + g(z)$ ist am Punkt z^* stetig.
- Die Funktion $z \mapsto f(z) - g(z)$ ist am Punkt z^* stetig.
- Die Funktion $z \mapsto f(z) \cdot g(z)$ ist am Punkt z^* stetig.
- Die Funktion $z \mapsto \frac{f(z)}{g(z)}$ ist am Punkt z^* stetig, falls $g(z^*) \neq 0$.
- Die Funktion $z \mapsto \sqrt{f(z)}$ ist am Punkt z^* stetig.

Weiterhin gilt: ist g am Punkt z^* stetig und f am Punkt $g(z^*)$, so ist $z \mapsto f(g(z))$ am Punkt z^* stetig.

Beweis: Betrachte eine beliebige Folge $(z_n) \rightarrow z^*$ und wende die Rechenregeln 2.13 an.

Q.E.D.

Beispiel 4.11: Die Funktion $f(x) = \frac{x+1}{x^2+1}$ ist überall auf \mathbb{R} stetig: Da konstante Funktionen sowie $g(x) = x$ stetig sind, ist auch $h(x) = x + 1$ stetig. Analog ist $k(x) = x^2$ und damit auch $j(x) = x^2 + 1$ stetig. Außerdem gilt $j(x) > 0$ für alle $x \in \mathbb{R}$, womit der Quotient $f(x) = \frac{h(x)}{j(x)}$ ebenfalls überall stetig ist.

Betrachtet man die Funktion in der komplexen Ebene, so ist sie überall stetig bis auf die beiden Punkte $\pm i$, wo der Nenner verschwindet.

An diesem Beispiel merkt man, dass folgende „Pi mal Daumen-Regel“ gilt:

Merkregel 4.12:

Aus stetigen Funktionen „zusammengesetzte“ Funktionen sind wieder stetig. Lediglich an den Stellen, wo man durch 0 teilt, kann die Funktion unstetig sein.

Manchmal helfen die Rechenregeln nicht, und man muss technisch abschätzen:

Beispiel 4.13: Die in Definition 2.20/Beispiel 3.24 eingeführte Exponentialfunktion $z \mapsto \exp(z)$ ist stetig am Nullpunkt. Betrachte dazu eine beliebige Nullfolge h_n , für die o.B.d.A. $|h_n| \leq 1$ gelte. Wegen

$$|e^{h_n} - 1| = \left| 1 + h_n + \frac{h_n^2}{2!} + \frac{h_n^3}{3!} + \dots - 1 \right| = |h_n| \cdot \left| 1 + \frac{h_n}{2!} + \frac{h_n^2}{3!} + \dots \right|$$

$\leq |h_n| \cdot \left(1 + \frac{|h_n|}{2!} + \frac{|h_n|^2}{3!} + \dots\right) \leq |h_n| \cdot \left(1 + \frac{1}{2!} + \frac{1}{3!} + \dots\right) = |h_n| \cdot (e^1 - 1) \leq 2 \cdot |h_n|$
 ist $e^{h_n} - 1$ eine Nullfolge, also $\lim_{n \rightarrow \infty} e^{h_n} = 1 = e^0$. Dies ist die Stetigkeit am Nullpunkt.

Satz 4.14: (Stetigkeit der Exponentialfunktion)

Die in Definition 2.20/Beispiel 3.24 eingeführte Exponentialfunktion $z \mapsto \exp(z)$ mit dem Definitionsbereich \mathbb{C} ist an allen Punkten $z \in \mathbb{C}$ stetig.

Beweis: Sei (h_n) eine beliebige Nullfolge. Wegen der Funktionalgleichung 2.22 $e^{a+b} = e^a \cdot e^b$ und der gerade gezeigten Stetigkeit im Nullpunkt folgt

$$\lim_{n \rightarrow \infty} e^{z+h_n} = \lim_{n \rightarrow \infty} e^z \cdot e^{h_n} = e^z \cdot \lim_{n \rightarrow \infty} e^{h_n} = e^z \cdot e^{\lim_{n \rightarrow \infty} h_n} = e^z \cdot e^0 = e^z.$$

Q.E.D.

12.11.03↓

Die Merkregel 4.12 besagt, dass es potentielle Unstetigkeiten gibt, wenn man durch 0 teilt. **Aber:** es kann auch passieren, dass an diesen Stellen Stetigkeit vorliegt (nämlich bei speziellen $\frac{0}{0}$ -Situation):

Beispiel 4.15: Die Funktion

$$f(z) = \begin{cases} \frac{z^2 - 1}{z - 1} & \text{für } z \neq 1, \\ 2 & \text{für } z = 1 \end{cases}$$

ist überall (auch an der Stelle $z = 1$) stetig. Dies ist leicht gezeigt: Wegen $z^2 - 1 = (z + 1) \cdot (z - 1)$ ist f nichts anderes als eine komplizierte Schreibweise für $f(z) = z + 1$.

Etwas komplizierter ist

$$f(z) = \begin{cases} \frac{e^z - 1}{z} & \text{für } z \neq 0, \\ 1 & \text{für } z = 0. \end{cases}$$

Auch diese Funktion ist überall (auch an der Stelle $z = 0$) stetig, was durch folgende Betrachtung plausibel wird:

$$\frac{e^z - 1}{z} = \frac{1}{z} \cdot \left(1 + z + \frac{z^2}{2!} + \dots - 1\right) = \frac{1}{z} \cdot \left(z + \frac{z^2}{2!} + \dots\right) = 1 + \frac{z}{2!} + \dots$$

Eine $\frac{0}{0}$ -Situation läßt sich mit Hilfe der „l'Hospitalschen Regel“ systematisch untersuchen, siehe Beispiel 6.36.

4.3 Grenzwerte

Betrachtet man $f(z) = (e^z - 1)/z$, so ist diese Funktion zunächst mal für $z = 0$ nicht definiert, sie hat dort eine „**Definitionslücke**“. In Beispiel 4.15 haben wir einen geeigneten Wert definiert, der die Funktion insgesamt stetig macht. Dieser Wert ergibt sich als „Grenzwert“ der Funktion, wenn das Argument gegen den kritischen Wert strebt.

Definition 4.16: (Grenzwerte bei Funktionen)

Betrachte eine Funktion f auf dem Definitionsbereich $D = \mathbb{C} \setminus \{z^*\}$. Der Wert f^* heißt „**Grenzwert (Limes)**“ von f für $z \rightarrow z^*$, wenn für jede gegen z^* konvergierende Folge (z_n) mit $z_n \in D$ gilt:

$$\lim_{n \rightarrow \infty} f(z_n) = f^*.$$

Die Schreibweise ist dann: $f^* = \lim_{z \rightarrow z^*} f(z)$.

Die Funktion

$$z \in D \cup \{z^*\} \mapsto \begin{cases} f(z) & \text{für } z \in D, \\ f^* & \text{für } z = z^* \end{cases}$$

nennt man die „**stetige Fortsetzung**“ von f auf den erweiterten Definitionsbereich $D \cup \{z^*\}$. Nach Konstruktion ist die Fortsetzung stetig am Punkt z^* .

Definition 4.17: (Einseitige Grenzwerte bei reellen Funktionen)

Für reelle Funktionen $f : D \setminus \{x^*\} \subset \mathbb{R} \mapsto \mathbb{R}$ wird weiterhin definiert::

Der Wert f^* heißt „**rechtsseitiger Grenzwert**“ von f für $x \rightarrow x^*$, wenn $\lim_{n \rightarrow \infty} f(x_n) = f^*$ gilt für alle gegen x^* konvergierende Folgen (x_n) mit $x_n > x^*$. Schreibweise:

$$f^* = \lim_{x \rightarrow x^*+0} f(x).$$

Der Wert f^* heißt „**linksseitiger Grenzwert**“ von f für $x \rightarrow x^*$, wenn $\lim_{n \rightarrow \infty} f(x_n) = f^*$ gilt für alle gegen x^* konvergierende Folgen (x_n) mit $x_n < x^*$. Schreibweise:

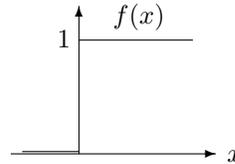
$$f^* = \lim_{x \rightarrow x^*-0} f(x).$$

Beispiel 4.18: Für eine am Punkt x^* definierte und dort stetige reelle Funktion gilt immer

$$\lim_{x \rightarrow x^*-0} f(x) = \lim_{x \rightarrow x^*+0} f(x) = \lim_{x \rightarrow x^*} f(x) = f(x^*).$$

Beispiel 4.19: Betrachte

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



Hier gilt für die Sprungstelle $x^* = 0$:

$$\lim_{x \rightarrow 0-0} f(x) = 0, \quad \lim_{x \rightarrow 0+0} f(x) = 1, \quad \lim_{x \rightarrow 0} f(x) \text{ existiert nicht.}$$

Beispiel 4.20: Für die reelle Funktion $f(x) = \frac{1}{x}$ gilt

$$\lim_{x \rightarrow \infty} f(x) = 0.$$

Formale Begründung: Sei (x_n) eine beliebige gegen ∞ konvergierende Folge:

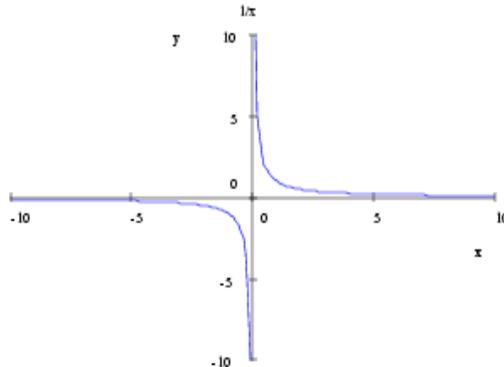
$$\lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} \frac{1}{x_n} = \frac{1}{\infty} = 0.$$

Am Punkt $x = 0$ ist f unstetig („singulär“): die Funktion hat eine sogenannte Polstelle. Wir lassen die Werte $\pm\infty$ wieder als Grenzwerte zu. Dann existieren einseitige Grenzwerte:

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

Das Argument `ViewingBox = [-10..10, -10..10]` im folgenden Befehl weist MuPAD an, alles außerhalb der angegebenen Bereiche zu ignorieren, wodurch sich eine gut skalierte Graphik ergibt:

```
>> plotfunc2d(1/x, x = -10..10,
ViewingBox = [-10..10, -10..10])
```



Mit dem Grenzwertbegriff für Funktionen können wir die Stetigkeit an einem Punkt auch folgendermaßen charakterisieren:

Satz 4.21: (Stetigkeit von Funktionen über \mathbb{R})

Eine auf einem Intervall $D = (a, b) \subset \mathbb{R}$ definierte Funktion f ist am Punkt $x^* \in D$ genau dann linksseitig stetig, wenn

$$\lim_{x \rightarrow x^* - 0} f(x) = f(x^*)$$

gilt. Sie ist genau dann rechtsseitig stetig, wenn

$$\lim_{x \rightarrow x^* + 0} f(x) = f(x^*)$$

gilt. Sie ist genau dann stetig, wenn der links- und rechtsseitige Grenzwert existiert und beide Grenzwerte mit dem Funktionswert übereinstimmen:

$$\lim_{x \rightarrow x^* - 0} f(x) = \lim_{x \rightarrow x^* + 0} f(x) = f(x^*).$$

Beweis: Das folgt unmittelbar aus den Definitionen. Für die letzte Aussage beachte, dass eine beliebige gegen x^* konvergierende Folge aufgespalten werden kann in die Teilfolge aller Elemente, die kleiner sind als x^* und die Teilfolge aller Elemente, die größer als x^* sind. Die Konvergenz der Teilfolgen gegen $f(x^*)$ ist die links- bzw. rechtsseitige Stetigkeit, die Konvergenz der Gesamtfolge gegen $f(x^*)$ ist die Stetigkeit.

Q.E.D.

4.4 Der Zwischenwertsatz, das Min/Max-Prinzip

Es folgen zwei sehr wichtige und fundamentale Sätze für reelle stetige Funktionen.

Satz 4.22: (Der Zwischenwertsatz für stetige Funktionen)

Sei $f : [a, b] \mapsto \mathbb{R}$ auf dem Intervall $[a, b] \subset \mathbb{R}$ stetig. Dann nimmt f auf dem Intervall alle Werte zwischen $f(a)$ und $f(b)$ an: zu jedem y zwischen den Werten $f(a)$ und $f(b)$ existiert mindestens ein $x \in [a, b]$ mit $f(x) = y$.

Beweis: (nicht nur für technisch Interessierte)

Wir benutzen einen expliziten Algorithmus („**Intervallhalbierung**“), um die Lösung von $f(x) = y$ zu finden.

Sei $f(a) \neq f(b)$ (sonst gibt es nichts zu zeigen). O.B.d.A. gelte $f(a) < f(b)$ (sonst betrachte statt $-f$ statt f). Gegeben sei y mit $f(a) \leq y \leq f(b)$. Für $y = f(a)$ bzw. $y = f(b)$ ist die Behauptung sicher mit $x = a$ bzw. $x = b$ erfüllt. Es gelte also nun $f(a) < y < f(b)$.

Betrachte den Mittelpunkt $m = (a+b)/2$ des Intervalls. Gilt $f(m) = y$, sind wir fertig. Für $f(m) > y$ betrachten wir die linke Intervallhälfte $[a_1, b_1] := [a, m]$, für $f(m) < y$ betrachten wir die rechte Intervallhälfte $[a_1, b_1] := [m, b]$. Nach Konstruktion ist die Ausgangssituation

$$f(a_1) \leq y \leq f(b_1)$$

für das neue Intervall $[a_1, b_1]$ wieder hergestellt. Das betrachtete Intervall $[a_1, b_1]$ wird nun erneut zu einem Intervall $[a_2, b_2]$ halbiert usw.

Es ergibt sich eine Folge von immer kleineren Intervallen $[a_n, b_n]$, deren linke Enden a_n eine monoton steigende und deren rechten Enden b_n eine monoton fallende Folge bildet. Nach Konstruktion gilt für alle Intervallenden

$$f(a_n) \leq y \leq f(b_n).$$

Nach Satz 2.24 konvergieren die monotonen beschränkten Folgen (a_n) und (b_n) gegen Grenzwerte a^* bzw b^* , die übereinstimmen müssen, da die Intervalllängen $b_n - a_n = (b - a)/2^n$ gegen Null konvergieren. Da f am Punkt $x := a^* = b^*$ stetig ist, folgt

$$f(x) = \lim_{n \rightarrow \infty} f(a_n) = f(\lim_{n \rightarrow \infty} a_n) \leq y \leq f(\lim_{n \rightarrow \infty} b_n) = \lim_{n \rightarrow \infty} f(b_n) = f(x),$$

also $f(x) = y$.

Q.E.D.

Bemerkung 4.23: Der im Beweis des Zwischenwertsatzes verwendete Algorithmus („Intervallhalbierung“, „Bi-Sektion“) ist ein auch in der Praxis anwendbarer Suchalgorithmus zum approximativen Lösen einer Gleichung $f(x) = y$. Er liefert eine Folge von Intervallschachtelungen $[a_n, b_n]$ für die Lösung. Die Genauigkeit ist die Länge des Intervalls, auf das die Lösung eingeschränkt werden konnte. Mit $2^{10} = 1024 \approx 10^3$ gilt die

Faustregel: Durch je 10 Halbierungsschritte gewinnt man jeweils etwa 3 Dezimalstellen Genauigkeit hinzu.

Bemerkung 4.24: Der Beweis verwendet über Satz 2.24 das Supremumsaxiom für \mathbb{R} . In der Tat hat beispielsweise die stetige Funktion $f(x) = x^2 - 2$ auf dem Intervall $[0, 2] \cap \mathbb{Q}$ keine Nullstelle, obwohl $f(0) = -2 < 0 < f(2) = 2$ gilt, da die Nullstelle $x = \sqrt{2}$ nicht rational ist.

Ein weiteres wichtiges Ergebnis für stetige reelle Funktionen ist, dass der Bildbereich eines beschränkten abgeschlossenen Intervalls wieder ein beschränktes abgeschlossenes Intervall ist. Mit anderen Worten: die Funktion nimmt auf einem abgeschlossenen Intervall immer (mindestens) ein globales Minimum und ein globales Maximum an:

Satz 4.25: (Das Min/Max-Prinzip für stetige Funktionen)

Sei $f : [a, b] \mapsto \mathbb{R}$ stetig auf dem Intervall $[a, b] \subset \mathbb{R}$. Dann existiert ein $x_{min} \in [a, b]$ und ein $x_{max} \in [a, b]$ mit

$$f(x_{min}) \leq f(x) \leq f(x_{max})$$

für alle $x \in [a, b]$.

Beweis: (für technisch Interessierte) Wir konstruieren x_{max} . Die Bildmenge $f([a, b]) := \{f(x); x \in [a, b]\}$ ist nach oben beschränkt. Sonst gäbe es nämlich eine uneigentlich nach ∞ konvergierende Folge (y_n) in $f([a, b])$ mit (nicht unbedingt eindeutig bestimmten) Urbildern $x_n \in [a, b]$. Nach Bolzano-Weierstrass 2.39/Bemerkung 2.40 gibt es eine gegen einen Grenzwert $x^* \in [a, b]$ konvergierende Teilfolge (x_{n_k}) in $[a, b]$, für die

$$\infty = \lim_{k \rightarrow \infty} y_{n_k} = \lim_{k \rightarrow \infty} f(x_{n_k}) = f\left(\lim_{k \rightarrow \infty} x_{n_k}\right) = f(x^*)$$

gelten müsste. Widerspruch!

Da $f([a, b])$ nach oben beschränkt ist, existiert gemäß des Supremumsaxioms 1.10 das Supremum $Y = \sup f([a, b])$ aller Bildpunkte. Es gilt zu zeigen, dass dieses Supremum in der Menge $f([a, b])$ liegt, also ein Maximum ist:

Da $Y - \frac{1}{n}$ keine obere Schranke von $f([a, b])$ sein kann (Y ist als Supremum die kleinste obere Schranke), gibt es zu jedem $n \in \mathbb{N}$ ein $x_n \in [a, b]$ mit

$$Y - \frac{1}{n} < f(x_n) \leq Y.$$

Wiederum existiert nach Bolzano-Weierstrass 2.39/Bemerkung 2.40 eine gegen einen Grenzwert $x^* \in [a, b]$ konvergente Teilfolge (x_{n_k}) , für die

$$Y = \lim_{k \rightarrow \infty} \left(Y - \frac{1}{n_k}\right) \leq \lim_{k \rightarrow \infty} f(x_{n_k}) \leq Y$$

gilt. Mit der Stetigkeit von f folgt

$$Y = \lim_{k \rightarrow \infty} f(x_{n_k}) = f\left(\lim_{k \rightarrow \infty} x_{n_k}\right) = f(x^*).$$

Also ist $f(x^*) = \max f([a, b])$, d.h., $x_{max} = x^*$ ist die gesuchte Maximumsstelle.

Die Minimumsstelle x_{min} mit $f(x_{min}) = \min f([a, b])$ ergibt sich sofort als die Maximumsstelle von $-f$.

Q.E.D.

Bemerkung 4.26: Die Abgeschlossenheit des Intervalls $[a, b]$ ist wesentlich für die Existenz von Minimum und Maximum. Beispielsweise hat für das offene Intervall $(0, 1)$ die auf $(0, 1)$ stetige Funktion $f(x) = 1/x$ offensichtlich weder ein Maximum noch ein Minimum!

4.5 Umkehrfunktionen

Definition 4.27: (Invertierbarkeit von Funktionen)

Eine Funktion $f : D \mapsto W$ von einem Definitionsbereich D in den Wertebereich $W = f(D) = \{f(x); x \in D\}$ heißt **invertierbar**, wenn zu jedem Wert $y \in W$ genau ein Urbild $x \in D$ mit $f(x) = y$ existiert.

Beispiel 4.28: Die Funktion $f(x) = x^2$ auf dem Definitionsbereich $D = [0, \infty)$ mit dem Wertebereich $f(D) = [0, \infty)$ ist invertierbar: zu $y = f(x) = x^2$ gehört genau ein Urbild $x = \sqrt{y}$ im Definitionsbereich D .

Die Funktion $f(x) = x^2$ auf dem Definitionsbereich $D = (-\infty, 0]$ mit dem Wertebereich $f(D) = [0, \infty)$ ist invertierbar: zu $y = f(x) = x^2$ gehört genau ein Urbild $x = -\sqrt{y}$ im Definitionsbereich D .

Die Funktion $f(x) = x^2$ ist nicht invertierbar, wenn man sie auf dem Definitionsbereich $D = \mathbb{R}$ betrachtet: Jetzt gibt es zu jedem $y = f(x) = x^2$ aus dem Wertebereich $f(D) = [0, \infty)$ zwei Urbilder $x = \sqrt{y}$ und $x = -\sqrt{y}$.

Definition 4.29: (Inverse einer Funktion)

Die Funktion $f : D \mapsto W$ von einem Definitionsbereich D in den Wertebereich $W = f(D) = \{f(x); x \in D\}$ sei invertierbar. Die „**Umkehrabbildung**“ („**Inverse**“) von f ist die Funktion $f^{-1} : W \mapsto D$, die dem Punkt $y = f(x) \in W$ den (eindeutig bestimmten) Wert x zuordnet.

Beispiel 4.30: Die Funktion $f(x) = x^2$ auf dem Definitionsbereich $D = [0, \infty)$ mit dem Wertebereich $W = f(D) = [0, \infty)$ hat die durch $f^{-1}(y) = \sqrt{y}$ gegebene Inverse $f^{-1} : W \mapsto D$.

Die Funktion $f(x) = x^2$ auf dem Definitionsbereich $D = (-\infty, 0]$ mit dem Wertebereich $f(D) = [0, \infty)$ hat die durch $f^{-1}(y) = -\sqrt{y}$ gegebene Inverse $f^{-1} : W \mapsto D$.

Die Funktion $f(x) = x^2$ auf dem Definitionsbereich $D = \mathbb{R}$ hat keine Inverse.

Die Funktion $f(x) = 2 - 3 \cdot x$ auf dem Wertebereich $D = \mathbb{R}$ hat die Inverse $f^{-1}(y) = \frac{2-y}{3}$.

Um die Inverse zu bestimmen, muss man $y = f(x)$ nach x auflösen:

$$y = 2 - 3 \cdot x \quad \Longrightarrow \quad 3 \cdot x = 2 - y \quad \Longrightarrow \quad x = \frac{2 - y}{3}.$$

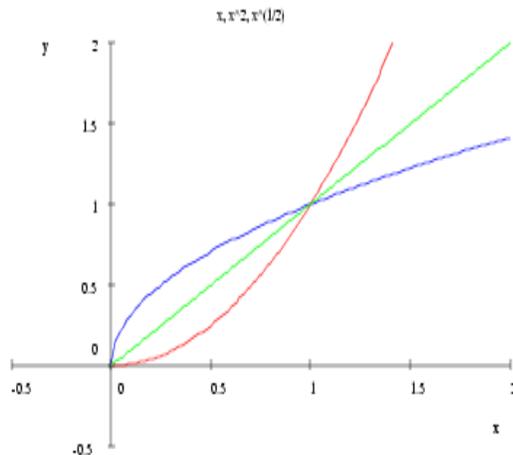
Graphische Darstellung der Inversen 4.31:

Hat man eine invertierbare Funktion f graphisch dargestellt, so hat man auch sofort den Graphen von f^{-1} . Der Graph von f ist eine Punktmenge (x, y) mit $y = f(x)$ in der x - y -Ebene. Der Graph von f^{-1} ist die Punktmenge (y, x) mit $y = f(x)$. Diese ergibt sich einfach durch Spiegelung an der „ersten Winkelhalbierenden“ (dies ist die durch $y = x$ gegebene Gerade).

Der Graph der Umkehrfunktion f^{-1} ist die Spiegelung des Graphen der Funktion f an der ersten Winkelhalbierenden.

Beispiel 4.32: Zur Demonstration hierzu einige MuPAD Graphiken. Betrachte $f(x) = x^2$ auf $D = [0, \infty)$, $f^{-1}(y) = \sqrt{y}$. Statt $f^{-1}(y) = \sqrt{y}$ wird $f^{-1}(x) = \sqrt{x}$ eingegeben (Goethe sagt dazu treffend: „Name ist Schall und Rauch“). Die Winkelhalbierende $y = x$ wird zusätzlich eingezeichnet:

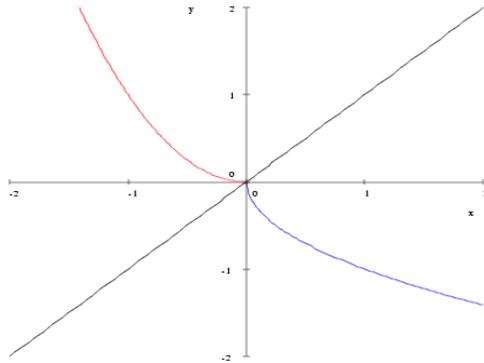
```
>> plotfunc2d(x, x^2, sqrt(x), x = 0..2,
             ViewingBox = [-0.5..2, -0.5..2])
```



Das selbe noch einmal, diesmal wird $f(x) = x^2$ aber auf dem Definitionsbereich $D = (-\infty, 0]$ betrachtet. Da die Inverse $f^{-1}(y) = -\sqrt{y}$ auf einem anderen Definitionsbereich

lebt ($y \geq 0, x \leq 0$), `plotfunc2d` aber alle Funktionen über einem gemeinsamen Bereich zeichnet, wird nun das folgende flexiblere `plot`-Konstrukt benutzt:

```
>> plot(// die Winkelhalbierende:
        plot::Function2d(x, x = -2..2, Color = RGB::Black),
        // f(x):
        plot::Function2d(x^2, x = -2..0, Color = RGB::Red),
        // die Inverse von f:
        plot::Function2d(-sqrt(y), y = 0..2, Color = RGB::Blue),
        ViewingBox = [-2..2, -2..2])
```



Bei streng monotonen Funktionen ist die Invertierbarkeit garantiert:

Satz 4.33: (Invertierbarkeit bei Monotonie)

Streng monotone reelle Funktionen $f : [a, b] \mapsto f([a, b])$ sind immer invertierbar. Ist f streng monoton steigend, dann auch f^{-1} . Ist f streng monoton fallend, dann auch f^{-1} . Ist f monoton und stetig, dann auch f^{-1} .

Beweis: (für technisch Interessierte) Die Eindeutigkeit der Urbilder folgt unmittelbar aus der Monotonie, denn aus $x_1 \neq x_2$ (also entweder $x_1 < x_2$ oder $x_1 > x_2$) folgt per strenger Monotonie $f(x_1) \neq f(x_2)$. Die Monotonie von f^{-1} ist offensichtlich.

Zur Stetigkeit von f^{-1} . Sei o.B.d.A. f streng monoton wachsend (sonst betrachte $-f$). Wähle einen beliebigen Punkt y aus dem Wertebereich $f([a, b]) = [f(a), f(b)]$. Sei (y_n) eine beliebige gegen y konvergierende Folge, sei $x = f^{-1}(y)$. Wegen $(y_n) \rightarrow y$ gibt es zu jedem $\delta > 0$ ein $N(\delta)$, so dass

$$y_n \in [y - \delta, y + \delta]$$

gilt für alle $n \geq N(\delta)$. Zu $\epsilon > 0$ setze

$$\delta(\epsilon) = \min(y - f(x - \epsilon), f(x + \epsilon) - y) > 0.$$

Für alle $n \geq N(\delta(\epsilon))$ folgt dann

$$y_n \in [y - \delta(\epsilon), y + \delta(\epsilon)] \subset [f(x - \epsilon), f(x + \epsilon)],$$

also

$$\begin{aligned} f^{-1}(y_n) &\in f^{-1}([f(x - \epsilon), f(x + \epsilon)]) = [f^{-1}(f(x - \epsilon)), f^{-1}(f(x + \epsilon))] \\ &= [x - \epsilon, x + \epsilon]. \end{aligned}$$

Also: zu $\epsilon > 0$ haben wir ein $N(\delta(\epsilon))$ konstruiert, so dass

$$|f^{-1}(y_n) - x| = |f^{-1}(y_n) - f^{-1}(y)| \leq \epsilon$$

gilt für alle $n \geq N(\delta(\epsilon))$. Also ist f^{-1} stetig: $\lim_{n \rightarrow \infty} f^{-1}(y_n) = f^{-1}(\lim_{n \rightarrow \infty} y_n)$.
Q.E.D.

4.6 Wachstum von Funktionen, Landau-Symbole

Es gibt eine nützliche Schreibweise, mit der das Wachstum von Funktionen an interessanten Stellen elegant und schreibtechnisch äußerst effizient beschrieben wird. „Interessante Stellen“ sind dabei meist Nullstellen, Extrema oder Singularitäten.

Notation 4.34:

Seien f und g Funktionen, die in der Umgebung eines Punktes z_0 definiert seien. Bei reellen Funktionen betrachtet man oft auch die Punkte $z_0 = \pm\infty$.

- $f(z) = O(g(z))$ im Limes $z \rightarrow z_0$ bedeutet, dass die Funktion $|f(z)|/|g(z)|$ auf einer Umgebung von z_0 nach oben beschränkt ist.
- $f(z) = o(g(z))$ im Limes $z \rightarrow z_0$ bedeutet $\lim_{z \rightarrow z_0} f(z)/g(z) = 0$.

Beispiel 4.35:

$$\begin{aligned} e^z &= O(1) \text{ im Limes } z \rightarrow 0, & e^z &= 1 + O(z) \text{ im Limes } z \rightarrow 0, \\ e^z &= 1 + z + O(z^2) \text{ im Limes } z \rightarrow 0, & e^z &= 1 + z + o(z) \text{ im Limes } z \rightarrow 0, \\ \frac{x}{x+1} &= O(x) \text{ im Limes } x \rightarrow 0, & \frac{x}{x+1} &= O(1) \text{ im Limes } x \rightarrow \infty, \end{aligned}$$

Beispiel 4.36: Für alle positiven k gilt

$$x^k = o(e^x) \text{ im Limes } x \rightarrow \infty$$

sowie

$$e^{-x} = o\left(\frac{1}{x^k}\right) \text{ im Limes } x \rightarrow \infty,$$

denn

$$\left| \frac{e^{-x}}{1/x^k} \right| = \left| \frac{x^k}{e^x} \right| = \left| \frac{x^k}{1 + x + \dots + \frac{x^{k+1}}{(k+1)!} + \dots} \right| \leq \left| \frac{(k+1)!}{x} \right| \rightarrow 0 \quad \text{für } x \rightarrow \infty.$$

Anschaulich: die Funktion e^x wächst für gegen ∞ wachsendes x schneller als jede positive Potenz von x . Die Funktion e^{-x} fällt für gegen ∞ wachsendes x schneller ab als jede negative Potenz von x .

Kapitel 5

Einige spezielle Funktionen: exp, ln, sin, cos etc.

↓18.11.03

5.1 Exponentialfunktion und Logarithmus

Die überaus wichtige Exponentialfunktion soll nun etwas genauer diskutiert werden. Die ursprüngliche Definition 2.20 ist für die Diskussion zu unhandlich. Die in Beispiel 3.24 eingeführte Reihendarstellung ist wesentlich nützlicher. Wir haben sie bereits benutzt, um in Satz 4.14 die Stetigkeit über ganz \mathbb{C} zu beweisen. Wir betrachten die Exponentialfunktion nun zunächst im Reellen genauer:

Satz 5.1: (Eigenschaften der reellen Exponentialfunktion)

Die Exponentialfunktion $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \dots$ ist für $x \in \mathbb{R}$ streng monoton steigend. Es gilt

$$\lim_{x \rightarrow -\infty} e^x = 0, \quad \lim_{x \rightarrow \infty} e^x = \infty.$$

Der Wertebereich ist $(0, \infty)$.

Beweis: Für $0 \leq x < y$ ist $e^x < e^y$ offensichtlich, denn die Summanden der Partialsummen sind streng monoton wachsend:

$$e^y - e^x = \left(1 + y + \frac{y^2}{2} + \dots\right) - \left(1 + x + \frac{x^2}{2} + \dots\right) = \underbrace{y - x}_{>0} + \underbrace{\frac{y^2 - x^2}{2}}_{>0} + \dots > 0.$$

Für $x < y < 0$ folgt die Monotonie aus der Funktionalgleichung 2.22: $e^x = 1/e^{-x} < 1/e^{-y} = e^y$. Nach Beispiel 4.36 wächst e^x (stärker als jede positive x -Potenz) gegen ∞ für $x \rightarrow \infty$. Wegen $e^x = 1/e^{-x}$ fällt e^x gegen 0 für $x \rightarrow -\infty$. Damit ist der Wertebereich $(0, \infty)$.

Q.E.D.

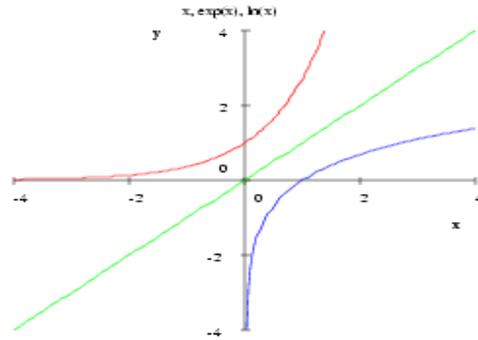
Definition 5.2: (Der natürliche Logarithmus)

Wegen der strengen Monotonie der reellen Exponentialfunktion $\exp : \mathbb{R} \mapsto (0, \infty)$ gibt es eine Umkehrfunktion, die man den „natürlichen Logarithmus“ $\ln : (0, \infty) \mapsto \mathbb{R}$ nennt:

$$\ln(\exp(x)) = x \text{ für alle } x \in \mathbb{R}, \quad \exp(\ln(y)) = y \text{ für alle } y \in (0, \infty).$$

Beispiel 5.3: Durch Spiegelung an der Winkelhalbierenden ergibt sich sofort der Graph von \ln aus dem Graphen von \exp :

```
>> plotfunc2d(x, exp(x), ln(x), x = -4..4,
              ViewingBox = [-4..4, -4..4])
```



Da die Exponentialfunktion nach Satz 4.14 monoton und stetig ist, ist mit Satz 4.33 auch der Logarithmus monoton und stetig:

Merke 5.4:

- \exp und \ln sind stetig und streng monoton wachsend.
- Es gilt $e^x > 1$ für alle $x > 0$, es gilt $\ln(y) > 0$ für alle $y > 1$.
- Es gilt $e^0 = 1$ und $\ln(1) = 0$.
- Es gilt $e^x < 1$ für alle $x < 0$ und $\ln(y) < 0$ für alle y mit $0 < y < 1$.

Bemerkung 5.5: Es ist klar, was mit x^y gemeint ist, wenn $x \in \mathbb{R}$ positiv und y eine ganze oder eine rationale Zahl ist (z.B. $x^{\frac{3}{4}} = \sqrt[4]{x^3}$). Was aber ist $x^{\sqrt{2}}$? Betrachte eine rationale Potenz $y = p/q$ mit $p, q \in \mathbb{N}$, dann ist $a = x^y = x^{p/q} > 0$ als die (eindeutige) positive Lösung von $a^q = x^p$ definiert. Setzen wir $x = e^{\ln(x)}$, so folgt mit den Funktionalgleichungen 2.22:

$$a^q = x^p = (e^{\ln(x)})^p = e^{p \cdot \ln(x)} = \left(e^{\frac{p}{q} \cdot \ln(x)} \right)^q.$$

Die einzige reelle positive Lösung a dieser Gleichung ist offensichtlich

$$x^{\frac{p}{q}} = a = e^{\frac{p}{q} \cdot \ln(x)}.$$

Also: für jedes rationale $y = p/q$ gilt:

$$x^y = e^{y \cdot \ln(x)} \text{ für jedes reelle } x > 0 \text{ und jedes rationale } y.$$

Man benutzt die obige Formel, um Potenzen von $x > 0$ auch für nicht-rationale reelle Werte y zu **definieren**, was nach obiger Überlegung mit der intuitiven „Wurzeldefinition“ für rationales y verträglich ist. Z. B.:

>> float(2^PI) = float(exp(PI*ln(2)))

$$8.824977827 = 8.824977827$$

Satz 5.6: (Rechenregeln für exp und ln)

Für beliebiges $x, y \in \mathbb{R}$ gilt:

$$e^{x+y} = e^x \cdot e^y, \quad (e^x)^y = e^{x \cdot y}, \quad e^{-x} = \frac{1}{e^x}.$$

Für beliebiges $x > 0, y > 0$ gilt:

$$\ln(x \cdot y) = \ln(x) + \ln(y), \quad \ln(x^y) = y \cdot \ln(x), \quad \ln\left(\frac{1}{x}\right) = -\ln(x).$$

Beweis: Die Funktionalgleichungen $e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}$ und $e^{-z} = 1/e^z$ waren schon in Satz 2.22 über \mathbb{C} gezeigt worden. Sind $z_1, z_2 \in \mathbb{R}$, folgt durch Logarithmieren

$$z_1 + z_2 = \ln\left(e^{z_1} \cdot e^{z_2}\right).$$

Mit $x = e^{z_1}$, $y = e^{z_2}$, also $z_1 = \ln(x)$, $z_2 = \ln(y)$, folgt $\ln(x) + \ln(y) = \ln(x \cdot y)$. Für $y = 1/x$ ergibt sich $\ln(x) + \ln(1/x) = \ln(1) = 0$. Nach Definition beliebiger reeller Potenzen gemäß Bemerkung 5.5 ergibt sich

$$(e^x)^y = e^{y \cdot \ln(e^x)} = e^{y \cdot x}.$$

Durch Logarithmieren folgt für beliebiges reelles $z = e^x > 0$:

$$\ln(z^y) = y \cdot x = y \cdot \ln(z).$$

Q.E.D.

Beispiel 5.7: Die Regel $\ln(x^y) = y \cdot \ln(x)$ ist nützlich, um Gleichungen aufzulösen, wo die gesuchte Größe in einem Exponenten auftaucht. Z.B.:

$$\begin{aligned} 2^x = 8 &\Rightarrow \ln(2^x) = \ln(8) \Rightarrow x \cdot \ln(2) = \ln(8) \\ \Rightarrow x &= \frac{\ln(8)}{\ln(2)} = \frac{\ln(2^3)}{\ln(2)} = \frac{3 \cdot \ln(2)}{\ln(2)} = 3. \end{aligned}$$

Bemerkung 5.8: Aus der Schulzeit mag man gewöhnt sein, statt mit dem natürlichen Logarithmus mit dem Zehner-Logarithmus \log_{10} umzugehen. Bei Informatikern ist (aus naheliegenden Gründen) der Logarithmus \log_2 zur Basis 2 populär. Hier ist der Zusammenhang zwischen dem natürlichen Logarithmus und dem Logarithmus zu einer beliebigen (positiven) Basis $b \neq 1$:

$$x = \log_b(y) \Leftrightarrow y = b^x \Leftrightarrow \ln(y) = \ln(b^x) = x \cdot \ln(b) \Leftrightarrow x = \frac{\ln(y)}{\ln(b)},$$

also

$$\log_b(y) = \frac{\ln(y)}{\ln(b)} \quad \text{für alle } y > 0, b > 0, b \neq 1.$$

Beispiel 5.9: Neben dem natürlichen Logarithmus `ln` hat MuPAD Logarithmen `log(b, y)` zu beliebigen positiven Basen $b \neq 1$:

```
>> log(10, 25.0) = ln(25.0)/ln(10.0)
```

```
1.397940009 = 1.397940009
```

```
>> log(2, 25.0) = ln(25.0)/ln(2.0)
```

```
4.64385619 = 4.64385619
```

5.2 Trigonometrische Funktionen

In der Schule waren im Kontext „Geometrie“ die Winkelfunktionen `sin` und `cos` eingeführt worden. Hier unsere Versionen:

Satz und Definition 5.10:

Die folgenden Reihen konvergieren für jeden Wert $z \in \mathbb{C}$. Die Reihenwerte heißen $\sin(z)$ bzw. $\cos(z)$ (die „trigonometrischen Funktionen“ Sinus und Cosinus):

$$\begin{aligned}\sin(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k \cdot z^{2k+1}}{(2 \cdot k + 1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} \pm \dots, \\ \cos(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k \cdot z^{2k}}{(2 \cdot k)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} \pm \dots.\end{aligned}$$

Zusätzlich führt man die nützliche Funktion $\tan(z) = \frac{\sin(z)}{\cos(z)}$ ein.

Beweis: Es ist zu zeigen, dass die definierenden Reihen konvergieren. In der Tat konvergieren sie absolut, was analog zu Beispiel 3.24 aus dem Quotientenkriterium folgt. Für die sin-Reihe:

$$\begin{aligned}\left| \frac{(-1)^{k+1} \cdot z^{2k+3} / (2 \cdot k + 3)!}{(-1)^k \cdot z^{2k+1} / (2 \cdot k + 1)!} \right| &= \frac{|z|^2 \cdot (2 \cdot k + 1)!}{(2 \cdot k + 3)!} \\ &= \frac{|z|^2}{(2 \cdot k + 2) \cdot (2 \cdot k + 3)} \leq \frac{|z|^2}{4 \cdot k^2} \leq \frac{1}{4}\end{aligned}$$

für $k \geq |z|$. Die Konvergenz der cos-Reihe folgt analog.

Q.E.D.

Der folgende Zusammenhang ist eine der wichtigsten Formeln überhaupt für exp, sin und cos:

Satz 5.11: (Die Euler-Formel)

Für jedes $z \in \mathbb{C}$ gilt folgende Beziehung zwischen der Exponentialfunktion und den trigonometrischen Funktionen:

$$\boxed{e^{i \cdot z} = \cos(z) + i \cdot \sin(z)}.$$

Für $x \in \mathbb{R}$ folgt $\cos(x) = \Re(e^{i \cdot x})$, $\sin(x) = \Im(e^{i \cdot x})$.

Beweis:

$$\begin{aligned}\cos(z) + i \cdot \sin(z) &= 1 - \frac{z^2}{2!} + \frac{z^4}{4!} \pm \dots \\ &\quad + i \cdot z - \frac{i \cdot z^3}{3!} + \frac{i \cdot z^5}{5!} \mp \dots \\ &= 1 + i \cdot z + \frac{(i \cdot z)^2}{2!} + \frac{(i \cdot z)^3}{3!} + \frac{(i \cdot z)^4}{4!} + \frac{(i \cdot z)^5}{5!} + \dots = e^{i \cdot z}.\end{aligned}$$

Q.E.D.

Satz 5.12:

Für jedes $z \in \mathbb{C}$ gilt $\sin(z) = \frac{e^{iz} - e^{-iz}}{2 \cdot i}$, $\cos(z) = \frac{e^{iz} + e^{-iz}}{2}$.

Beweis:

$$\begin{aligned} e^{iz} \pm e^{-iz} &= \cos(z) + i \cdot \sin(z) \pm \cos(-z) \pm i \cdot \sin(-z) \\ &= \cos(z) + i \cdot \sin(z) \pm \cos(z) \mp i \cdot \sin(z) = \begin{cases} 2 \cdot \cos(z) & \text{für } +, \\ 2 \cdot i \cdot \sin(z) & \text{für } -. \end{cases} \end{aligned}$$

Q.E.D.

Satz 5.13: (Stetigkeit der trigonometrischen Funktion)

Die trigonometrischen Funktionen \sin und \cos sind auf \mathbb{C} stetig.

Beweis: Da die Exponentialfunktion auf \mathbb{C} stetig ist, folgt dies über die Rechenregeln 4.10 für Stetigkeit aus den Darstellungen in Satz 5.12.

Q.E.D.

Satz 5.14: (Die Additionstheoreme der trigonometrischen Funktionen)

Für beliebiges $z_1, z_2 \in \mathbb{C}$ gilt:

$$\begin{aligned} \sin(z_1 + z_2) &= \sin(z_1) \cdot \cos(z_2) + \cos(z_1) \cdot \sin(z_2), \\ \cos(z_1 + z_2) &= \cos(z_1) \cdot \cos(z_2) - \sin(z_1) \cdot \sin(z_2). \end{aligned}$$

Beweis: Für $z_1, z_2 \in \mathbb{R}$ sind wegen $\cos(x) = \Re(e^{ix})$, $\sin(x) = \Im(e^{ix})$ die Additionstheoreme nichts Anderes als die Funktionalgleichung für \exp :

$$\begin{aligned} \cos(z_1 + z_2) &= \Re(e^{i(z_1+z_2)}) = \Re(e^{iz_1} \cdot e^{iz_2}) \\ &= \Re\left((\cos(z_1) + i \cdot \sin(z_1)) \cdot (\cos(z_2) + i \cdot \sin(z_2))\right) \\ &= \cos(z_1) \cdot \cos(z_2) - \sin(z_1) \cdot \sin(z_2). \end{aligned}$$

Das Additionstheorem für den reellen Sinus folgt analog über $\sin(z_1 + z_2) = \Im(e^{i(z_1+z_2)})$.

Für beliebiges $z_1, z_2 \in \mathbb{C}$ nehme man die Darstellung aus Satz 5.12, um die Additionstheoreme auf $e^{i(z_1+z_2)} = e^{iz_1} \cdot e^{iz_2}$ zurückzuführen.

Q.E.D.

Satz 5.15: (Symmetrien der trigonometrischen Funktionen)

Für beliebiges $z \in \mathbb{C}$ gilt: $\sin(-z) = -\sin(z)$, $\cos(-z) = \cos(z)$.

Beweis: Die Sinus-Reihe enthält nur ungerade Potenzen: $(-z)^{2\cdot k+1} = -z^{2\cdot k+1}$.
Die Cosinus-Reihe enthält nur gerade Potenzen: $(-z)^{2\cdot k} = z^{2\cdot k}$.

Q.E.D.

Satz 5.16: (Der Satz des Pythagoras)

Für jedes $z \in \mathbb{C}$ gilt: $\boxed{\sin^2(z) + \cos^2(z) = 1}$.

Beweis: Dies ist das Additionstheorem des Cosinus für $z_1 = z$, $z_2 = -z$ zusammen mit $\cos(0) = 1$:

$$1 = \cos(z - z) = \cos(z) \cdot \cos(-z) - \sin(z) \cdot \sin(-z) = \cos^2(z) + \sin^2(z).$$

Q.E.D.

 $\boxed{\downarrow 20.11.03}$

Wir brauchen die Kreiszahl π . Da wir hier keine Geometrie treiben und π über das Verhältnis von Kreisumfang zu Kreisdurchmesser einführen können, müssen wir π anders definieren:

Satz und Definition 5.17:

Auf der positiven reellen Achse besitzt der Cosinus mindestens eine Nullstelle. Sei $x_1 = \inf \{x \in \mathbb{R}; \cos(x) = 0; x > 0\}$ die kleinste positive Nullstelle des Cosinus. Definiere $\pi = 2 \cdot x_1 \approx 3.1415\dots$.

Beweis: Die Summanden der Cosinus-Reihe $\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} \mp \dots$ haben wechselnde Vorzeichen. Für kleines $|x|$ sind die Summanden monoton fallend.

Damit gilt $\cos(x) = 1 - \frac{x^2}{2} + f(x)$, wobei speziell für $|x| \leq 2$ gilt:

$$0 \leq f(x) = \frac{x^4}{4!} - \frac{x^6}{6!} \pm \dots \leq \frac{x^4}{4!}.$$

Es folgt

$$\begin{aligned} \cos(1) &= 1 - \frac{1}{2!} + f(1), & 0 \leq f(1) &\leq \frac{1}{24}, \\ \cos(2) &= 1 - \frac{4}{2!} + f(2), & 0 \leq f(2) &\leq \frac{16}{24}, \end{aligned}$$

also

$$\cos(1) \geq 1 - \frac{1}{2} = \frac{1}{2} > 0, \quad \cos(2) \leq 1 - \frac{4}{2} + \frac{16}{24} = -\frac{1}{3} < 0.$$

Der Zwischenwertsatz 4.22 für stetige Funktionen garantiert (mindestens) eine Nullstelle im Intervall $(1, 2)$. Damit ist die Menge $\{x \in \mathbb{R}; \cos(x) = 0; x > 0\}$ nicht leer und besitzt ein Infimum.

Q.E.D.

Über die Additionstheoreme und Pythagoras folgt nun eine Vielzahl von speziellen Resultaten, z.B.:

$$\sin(2 \cdot x) = 2 \cdot \sin(x) \cdot \cos(x) \quad \Rightarrow \quad \sin(\pi) = 2 \cdot \sin\left(\frac{\pi}{2}\right) \cdot \cos\left(\frac{\pi}{2}\right) = 0,$$

$$\cos(2 \cdot x) = \cos^2(x) - \sin^2(x) = 2 \cdot \cos^2(x) - 1$$

$$\Rightarrow \quad \cos(\pi) = 2 \cdot \cos^2\left(\frac{\pi}{2}\right) - 1 = -1,$$

$$\sin(x + \pi) = \sin(x) \cdot \underbrace{\cos(\pi)}_{-1} + \cos(x) \cdot \underbrace{\sin(\pi)}_0 = -\sin(x),$$

$$\cos(x + \pi) = \cos(x) \cdot \underbrace{\cos(\pi)}_{-1} - \sin(x) \cdot \underbrace{\sin(\pi)}_0 = -\cos(x)$$

etc. Hieraus folgt dann weiterhin die Periodizität

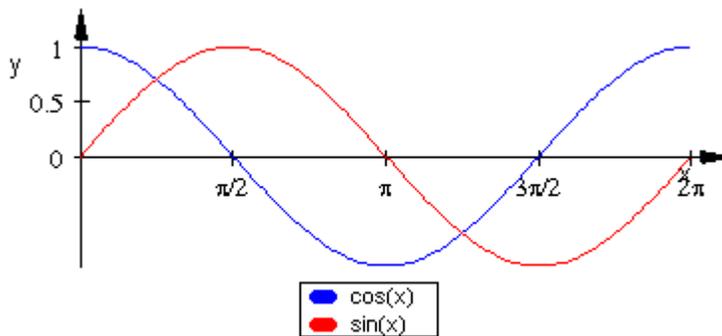
$$\sin(x + 2 \cdot \pi) = \sin(x), \quad \cos(y + 2 \cdot \pi) = \cos(x).$$

Die Einzelergebnisse aus Satz 5.11 bis Satz 5.16 werden zusammengefasst:

Merke 5.18:

Graphisch:

```
>> plotfunc2d(cos(x), sin(x), x=0..2*PI, TicksNumber = None,
             XTicksAt = [0 = "0", PI/2 = "p/2", PI = "p",
                        3*PI/2 = "3p/2", 2*PI = "2p"],
             YTicksAt = [-1, -1/2, 0, 1/2, 1],
             TicksLabelFont = ["Symbol"],
             Height = 5*unit::cm, Width = 10*unit::cm)
```



Einige spezielle Werte:

$$\sin(0) = 0, \quad \sin\left(\frac{\pi}{2}\right) = 1, \quad \sin(\pi) = 0, \quad \sin\left(\frac{3 \cdot \pi}{2}\right) = -1,$$

$$\cos(0) = 1, \quad \cos\left(\frac{\pi}{2}\right) = 0, \quad \cos(\pi) = -1, \quad \cos\left(\frac{3 \cdot \pi}{2}\right) = 0.$$

Periodizität (man braucht die Funktionen nur auf $[0, 2 \cdot \pi)$ zu kennen):

$$\sin(x + 2 \cdot \pi) = \sin(x), \quad \cos(y + 2 \cdot \pi) = \cos(x).$$

Additionstheoreme:

$$\sin(x + y) = \sin(x) \cdot \cos(y) + \cos(x) \cdot \sin(y),$$

$$\cos(x + y) = \cos(x) \cdot \cos(y) - \sin(x) \cdot \sin(y).$$

Symmetrieeigenschaften:

$$\sin(-x) = -\sin(x), \quad \cos(-x) = \cos(x).$$

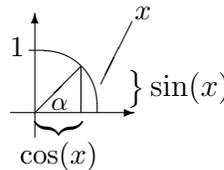
Pythagoras: $\sin^2(x) + \cos^2(x) = 1.$

Euler-Formel: $e^{i \cdot x} = \cos(x) + i \cdot \sin(x).$

Bemerkung 5.19: Vielleicht ist man aus der Schule noch gewohnt, die Argumente der trigonometrischen Funktion in Winkelgraden $\alpha = 0^\circ, \dots, 360^\circ$ anzugeben. Mathematiker nehmen statt des Winkels α die zugehörige Bogenlänge x auf dem Einheitskreis (Einheit: „Radian“, „im Bogenmaß“), der Zusammenhang ist

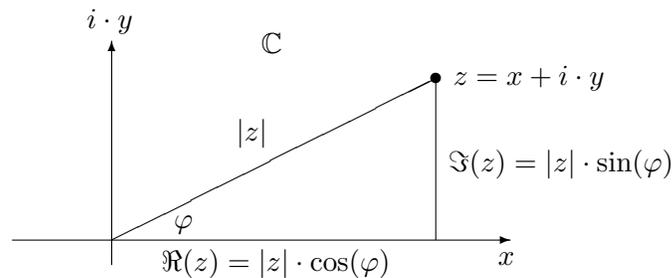
$$x = \frac{\pi}{180} \cdot \alpha,$$

d.h., $90^\circ \hat{=} \frac{\pi}{2}$, $180^\circ \hat{=} \pi$, $360^\circ \hat{=} 2 \cdot \pi$:



5.3 Die komplexe Exponentialfunktion, Polardarstellungen, komplexe Wurzeln

In der Geometrischen Interpretation 1.21 der komplexen Zahlen



war die **Polardarstellung**

$$z = |z| \cdot (\cos(\varphi) + i \cdot \sin(\varphi)), \quad \varphi \in [0, 2 \cdot \pi)$$

komplexer Zahlen eingeführt worden. Mit der Euler-Formel 5.11 ergibt sich die kompakte Polardarstellung:

$$z = |z| \cdot e^{i \cdot \varphi}, \quad \varphi \in [0, 2 \cdot \pi).$$

Man beachte, dass Polarwinkel nur bis auf ganzzahlige Vielfache von $2 \cdot \pi$ bestimmt sind (Periodizität von Sinus und Cosinus):

$$e^{i \cdot (\varphi + k \cdot 2 \cdot \pi)} = e^{i \cdot \varphi} \cdot e^{i \cdot k \cdot 2 \cdot \pi} = e^{i \cdot \varphi} \cdot \underbrace{(e^{i \cdot 2 \cdot \pi})^k}_1 = e^{i \cdot \varphi} \text{ für alle } k \in \mathbb{Z}.$$

Wir vereinbaren, dass unsere Polarwinkel im Intervall $[0, 2 \cdot \pi)$ liegen.

Geometrische Interpretation der komplexen Multiplikation 5.20:

Mit $z_1 = |z_1| \cdot e^{i \cdot \varphi_1}$, $z_2 = |z_2| \cdot e^{i \cdot \varphi_2}$, $\frac{1}{z_2} = \frac{1}{|z_2|} \cdot e^{-i \cdot \varphi_2}$ gilt

$$z_1 \cdot z_2 = |z_1| \cdot |z_2| \cdot e^{i \cdot (\varphi_1 + \varphi_2)}, \quad \frac{z_1}{z_2} = \frac{|z_1|}{|z_2|} \cdot e^{i \cdot (\varphi_1 - \varphi_2)}.$$

Also: die Multiplikation mit einer Zahl mit dem Polarwinkel φ dreht einen komplexen Vektor um den Winkel φ gegen den Uhrzeigersinn, die Division durch diese Zahl dreht den Vektor um den Winkel φ im Uhrzeigersinn. Multiplikation mit i bzw. Division durch i dreht speziell um 90° . Das ist leicht zu merken:

Ein Mathematiker ruft an und hört: „Die gewählte Nummer ist imaginär. Bitte drehen Sie ihren Apparat um 90° !“

Bemerkung 5.21: Für Potenzen von $z = |z| \cdot e^{i \cdot \varphi}$ folgt

$$z^n = |z|^n \cdot e^{i \cdot n \cdot \varphi}.$$

Damit sind wir nun in der Lage, komplexe Wurzeln zu berechnen. Die Aufgabe sei: Finde alle Lösungen von $z^n = a$.

Schritt 1: Stelle a in Polarkoordinaten dar: $a = |a| \cdot e^{i \cdot \alpha}$ mit $\alpha \in [0, 2 \cdot \pi)$.

Schritt 2: Ansatz für die Wurzeln: $z = r \cdot e^{i \cdot \varphi}$ mit $\varphi \in [0, 2 \cdot \pi)$. Vergleiche

$$z^n = r^n \cdot e^{i \cdot n \cdot \varphi} = |a| \cdot e^{i \cdot \alpha}.$$

Beispiel 5.22: Die n -ten „Einheitswurzeln“ der Gleichung $z^n = 1 = 1 \cdot e^{i \cdot 0}$ sind

$$z_k = \cos\left(\frac{k \cdot 2 \cdot \pi}{n}\right) + i \cdot \sin\left(\frac{k \cdot 2 \cdot \pi}{n}\right), \quad k = 0, 1, \dots, n-1.$$

Z.B. für $n = 4$;

$$z_k = \cos\left(\frac{k \cdot \pi}{2}\right) + i \cdot \sin\left(\frac{k \cdot \pi}{2}\right) = \begin{cases} 1 & \text{für } k = 0, \\ i & \text{für } k = 1, \\ -1 & \text{für } k = 2, \\ -i & \text{für } k = 3. \end{cases}$$

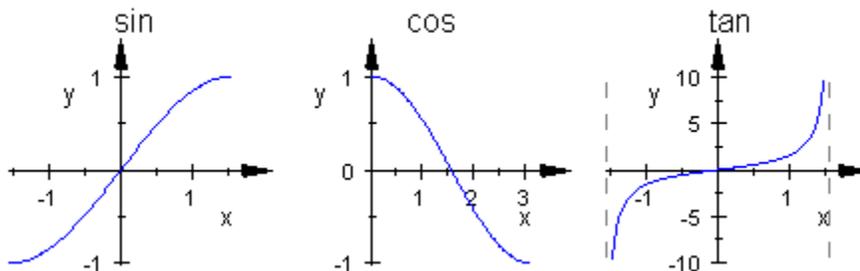
Für $n = 6$:

$$z_k = \cos\left(\frac{k \cdot \pi}{3}\right) + i \cdot \sin\left(\frac{k \cdot \pi}{3}\right) = \begin{cases} 1 & \text{für } k = 0, \\ \frac{1+i\sqrt{3}}{2} & \text{für } k = 1, \\ \frac{-1+i\sqrt{3}}{2} & \text{für } k = 2, \\ -1 & \text{für } k = 3, \\ \frac{-1-i\sqrt{3}}{2} & \text{für } k = 4, \\ \frac{1-i\sqrt{3}}{2} & \text{für } k = 5. \end{cases}$$

5.4 Die Inversen der trigonometrische Funktionen

Die Funktionen \sin und \tan sind auf dem Intervall $[-\pi/2, \pi/2]$ bzw. $(-\pi/2, \pi/2)$ monoton steigend, \cos ist auf auf dem Intervall $[0, \pi]$ monoton fallend:

```
>> plot(plot::Scene2d(plot::Function2d(sin(x), x = -PI/2..PI/2),
    Header = "sin"),
    plot::Scene2d(plot::Function2d(cos(x), x = 0..PI),
    Header = "cos"),
    plot::Scene2d(plot::Function2d(tan(x), x = -PI/2..PI/2),
    Header = "tan"),
    Layout = Horizontal,
    Height = 4*unit::cm, Width = 12*unit::cm)
```



Nach Satz 4.33 existieren damit die Umkehrfunktionen:

Definition 5.23: (Die inversen trigonometrischen Funktionen)

Wir definieren

$$\arcsin : [-1, 1] \rightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right],$$

$$\arccos : [-1, 1] \rightarrow [0, \pi],$$

$$\arctan : \mathbb{R} \rightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

durch die Forderungen

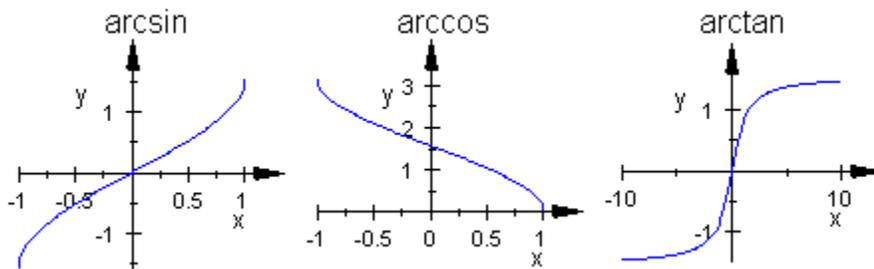
$$\sin(\arcsin(x)) = x \quad \forall x \in [-1, 1], \quad \arcsin(\sin(x)) = x \quad \forall x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right],$$

$$\cos(\arccos(x)) = x \quad \forall x \in [-1, 1], \quad \arccos(\cos(x)) = x \quad \forall x \in [0, \pi],$$

$$\tan(\arctan(x)) = x \quad \forall x \in \mathbb{R}, \quad \arctan(\tan(x)) = x \quad \forall x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Hier MuPAD-Graphiken dieser Funktionen:

```
>> plot(plot::Scene2d(plot::Function2d(arcsin(x), x = -1..1),
    Header = "arcsin"),
    plot::Scene2d(plot::Function2d(arccos(x), x = -1..1),
    Header = "arccos"),
    plot::Scene2d(plot::Function2d(arctan(x), x = -10..10),
    Header = "arctan"),
    Layout = Horizontal,
    Height = 4*unit::cm, Width = 12*unit::cm)
```



5.5 Diverse spezielle Funktionen: Gamma, Bessel etc.

Neben \exp , \ln , \sin , \cos etc. gibt es eine Unzahl weiterer spezieller Funktionen. Das Standardwerk, in dem die meisten der für die Physik relevanten Funktionen sehr detailliert zusammen mit ihren Rechenregeln („Additionstheoremen“ etc.) aufgeführt sind, ist:

[AS] M. ABRAMOWITZ AND I.A. STEGUN, *Handbook of Mathematical Functions*, New York: Dover Publications. 1964 (1st edition).

Es gibt mittlerweile die $(n + 1)$ -te Auflage (mit $n \gg 1$).

Hier nur zwei spezielle Funktionen, die besonders häufig in für die Physik relevanten Formeln auftauchen.

Definition 5.24: (Die Gamma-Funktion)

Für $x \in (0, \infty)$ definiere: $\Gamma(x) = \int_0^{\infty} t^{x-1} \cdot e^{-t} dt$.

Hierbei machen wir schon mal ein Anleihe bei der Integration, die im Kapitel 9 näher besprochen wird. Durch partielle Integration des definierenden Integrals bekommt man sofort die folgende Funktionalgleichung (vergl. Aufgabe ??, Blatt ??):

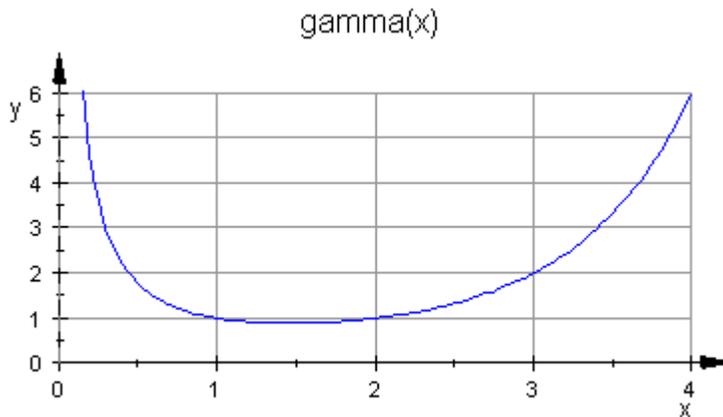
$$\Gamma(x) = (x - 1) \cdot \Gamma(x - 1).$$

Zusammen mit $\Gamma(1) = \int_0^{\infty} e^{-t} dt = 1$ folgt

$$\Gamma(n) = (n - 1)! \quad \text{für } n = 1, 2, 3, \dots$$

Die Gamma-Funktion ist damit eine Verallgemeinerung der Fakultätsfunktion auf nicht-ganzzahlige Argumente x . In MuPAD:

```
>> plotfunc2d(gamma(x), x = 0..4, YRange = 0..6, GridVisible,
              Height = 6*unit::cm, Width = 10*unit::cm,
              Header = "gamma(x)")
```



Definition 5.25: (Die Bessel-Funktionen)

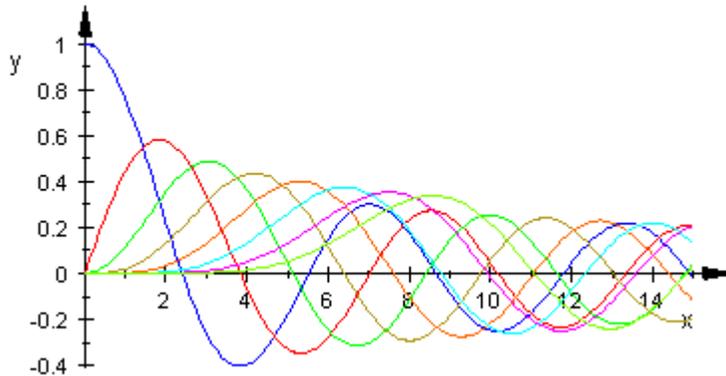
Definiere die **Bessel-Funktionen** (der ersten Art) mit „Index“ $\nu \in \{0, 1, 2, \dots\}$ als

$$J_\nu(z) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j! \cdot (j + \nu)!} \cdot \left(\frac{z}{2}\right)^{\nu+2j}, \quad z \in \mathbb{C}.$$

Die definierenden Reihen konvergieren für alle $z \in \mathbb{C}$, wie man (z.B. über das Quotientenkriterium) leicht nachweist. Hier ist eine MuPAD-Graphik der Bessel-Funktionen J_0, J_1, \dots, J_7 :

```
>> plotfunc2d(besselJ(nu, x) $ nu = 0..7, x = 0..15,
              Height = 6*unit::cm, Width = 10*unit::cm,
              Header = "J0(x), J1(x), .. , J7(x)",
              LegendVisible = FALSE)
```

J0(x), J1(x), .. , J7(x)



Kapitel 6

Differentialrechnung

6.1 Definitionen und Sätze

↓25.11.03

Im Prinzip könnten die meisten der folgenden Überlegungen und Definitionen ohne große Änderungen für komplexe Funktionen $f : \mathbb{C} \rightarrow \mathbb{C}$ durchgeführt werden. Wir beschränken uns hier jedoch auf reelle Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$. Zunächst die Definition einer Ableitung als Grenzwert von „Sekantensteigungen“:

Definition 6.1: (Die Ableitung einer Funktion)

Eine Funktion $f : D \mapsto \mathbb{R}$ heißt „**differenzierbar am Punkt x** “, wenn der Grenzwert

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

existiert. Der Grenzwert $f'(x)$ heißt „**Ableitung von f am Punkt x** “. Alternative Schreibweisen (mit $y = f(x)$):

$$\frac{dy}{dx} = y'(x) = \frac{d}{dx} f(x) = f'(x).$$

Ist f an jedem Punkt x des Definitionsbereichs D differenzierbar, so heißt die Abbildung $f' : x \mapsto f'(x)$ „**Ableitungsfunktion**“ (kurz: „**Ableitung von f** “).

Bemerkung 6.2: Ist eine Funktion an einem Punkt differenzierbar, so ist sie dort auch stetig:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \text{ existiert} &\Rightarrow f(x+h) - f(x) = O(h) \\ &\Rightarrow f(x+h) = f(x) + O(h) \Rightarrow \lim_{h \rightarrow 0} f(x+h) = f(x). \end{aligned}$$

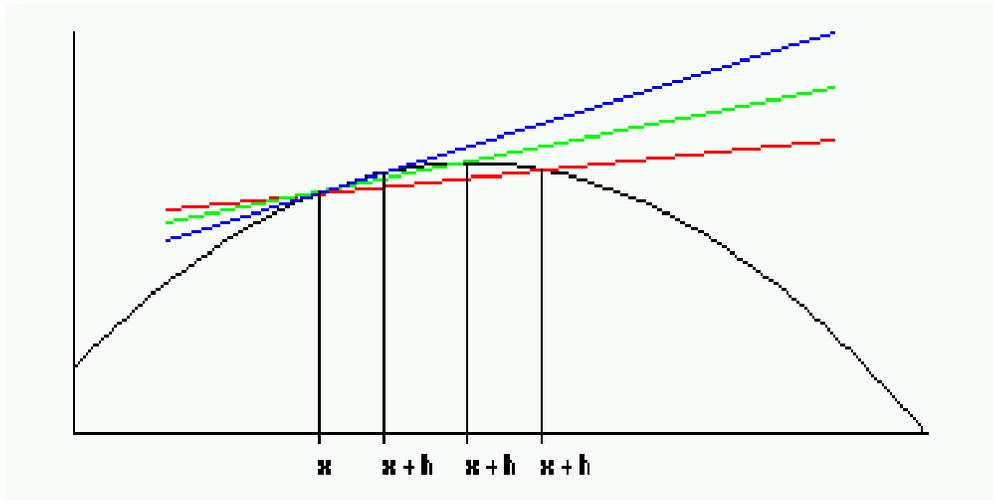
Damit kann eine Funktion nur an Stetigkeitspunkten differenzierbar sein.

Geometrische Interpretation der Ableitung 6.3:

Für kleines $\Delta x = h \neq 0$ ist der „Differenzenquotient“

$$\frac{\Delta f}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{(x + \Delta x) - x} = \frac{f(x + h) - f(x)}{h} \approx f'(x)$$

die Sekantensteigung vom Punkt $(x, f(x))$ zum Punkt $(x + h, f(x + h))$ auf dem Graphen von x :



Die Ableitung $f'(x)$ selbst, d.h., der Grenzwert der Sekantensteigung für $\Delta x = h \rightarrow 0$, ist die Steigung der Tangente an den Graphen von f am Punkt x .

Zur Erinnerung an die Schule: die Tangente T durch den Punkt $(x_0, f(x_0))$ mit der Steigung $f'(x_0)$ ist der Graph der linearen Funktion

$$T(x) = f(x_0) + f'(x_0) \cdot (x - x_0).$$

Interpretation der Ableitung 6.4:

Die Ableitung gibt an, wie stark sich $f(x)$ ändert, wenn sich x um einen kleinen Wert Δx ändert:

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \approx f'(x),$$

d.h.

$$f(x + \Delta x) \approx f(x) + f'(x) \cdot \Delta x .$$

Die Definition der Ableitung über den Grenzwert von Sekantensteigungen ist praktisch unnützlich, da nur in den allereinfachsten Fällen handhabbar, z.B., bei:

Beispiel 6.5: Betrachte $f(x) = x^2$:

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{x^2 + 2 \cdot x \cdot h + h^2 - x^2}{h} = \lim_{h \rightarrow 0} \frac{2 \cdot x \cdot h + h^2}{h} = \lim_{h \rightarrow 0} (2 \cdot x + h) = 2 \cdot x. \end{aligned}$$

Für das praktische Rechnen verläßt man sich wiederum auf Rechenregeln:

Satz 6.6: (Rechenregeln für's Ableiten)

Ableitungen einiger spezieller Funktionen (sei hierbei c eine konstante Zahl):

$$\begin{aligned} \frac{d}{dx} c &= 0, & \frac{d}{dx} x^n &= n \cdot x^{n-1}, & \frac{d}{dx} e^x &= e^x, & \frac{d}{dx} \ln(x) &= \frac{1}{x}, \\ \frac{d}{dx} \sin(x) &= \cos(x), & \frac{d}{dx} \cos(x) &= -\sin(x). \end{aligned}$$

Die Ableitung einer aus einfachen Funktionen zusammengesetzten Funktion ist über folgende Regeln zu berechnen. Seien f und g differenzierbare Funktionen. Die Ableitung der zusammengesetzten Funktion ($f + g$, $f \cdot g$ etc.) existiert jeweils, wenn f und g ableitbar sind:

- $\frac{d}{dx} c \cdot f(x) = c \cdot f'(x)$,
- $\frac{d}{dx} (f(x) + g(x)) = f'(x) + g'(x)$ („**Summenregel**“),
- $\frac{d}{dx} f(x) \cdot g(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$ („**Produktregel**“)
- $\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g(x)^2}$ („**Quotientenregel**“).

Bei der Quotientenregel wird $g(x) \neq 0$ vorausgesetzt (sonst teilt man durch 0).

Beweis: Die Ableitung von x^n , e^x , $\sin(x)$, $\cos(x)$ wird in Übungsaufgaben behandelt. Die Ableitung von $\ln(x)$ wird später in Beispiel 6.18 hergeleitet. Die Linearität $(c \cdot f)' = c \cdot f'$ und $(f + g)' = f' + g'$ folgt unmittelbar aus den Rechenregeln für Grenzwerte von Funktionen. Die Produktregel ergibt sich durch den Grenzwert von

$$\frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x)}{h}$$

$$= \frac{f(x+h) - f(x)}{h} \cdot g(x+h) + f(x) \cdot \frac{g(x+h) - g(x)}{h}$$

für $h \rightarrow 0$. Die Ableitung von $1/g(x)$ ergibt sich aus

$$\frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} = -\frac{g(x+h) - g(x)}{h} \cdot \frac{1}{g(x+h) \cdot g(x)}$$

zu

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

Zusammen mit der Produktregel liefert dies die Quotientenregel

$$\left(\frac{f(x)}{g(x)}\right)' = \left(f(x) \cdot \frac{1}{g(x)}\right)' = \frac{f'(x)}{g(x)} - f(x) \cdot \frac{g'(x)}{g(x)^2}.$$

Q.E.D.

Beispiel 6.7:

$$\frac{d}{dx} \sqrt[3]{x} = \frac{d}{dx} x^{\frac{1}{3}} = \frac{1}{3} \cdot x^{\frac{1}{3}-1} = \frac{1}{3} \cdot x^{-\frac{2}{3}} = \frac{1}{3} \cdot \frac{1}{x^{\frac{2}{3}}} = \frac{1}{3} \cdot \frac{1}{\sqrt[3]{x^2}}.$$

Beispiel 6.8: Summen- und Produktregel:

$$\begin{aligned} \frac{d}{dx} (x + x^2 \cdot e^x) &= \left(\frac{d}{dx} x\right) + \frac{d}{dx} (x^2 \cdot e^x) \\ &= \left(\frac{d}{dx} x\right) + \left(\frac{d}{dx} x^2\right) \cdot e^x + x^2 \cdot \left(\frac{d}{dx} e^x\right) = 1 + 2 \cdot x \cdot e^x + x^2 \cdot e^x. \end{aligned}$$

Beispiel 6.9: Quotientenregel:

$$\frac{d}{dx} \frac{e^x}{x} = \frac{\left(\frac{d}{dx} e^x\right) \cdot x - e^x \cdot \left(\frac{d}{dx} x\right)}{x^2} = \frac{e^x \cdot x - e^x \cdot 1}{x^2} = \frac{e^x}{x} - \frac{e^x}{x^2}.$$

Beispiel 6.10:

$$\begin{aligned}
 \frac{d}{dx} \frac{\cos(x) \cdot e^x}{x} &= \frac{\left(\frac{d}{dx} (\cos(x) \cdot e^x)\right) \cdot x - \cos(x) \cdot e^x \cdot \left(\frac{d}{dx} x\right)}{x^2} \\
 &= \frac{\left(\left(\frac{d}{dx} \cos(x)\right) \cdot e^x + \cos(x) \cdot \left(\frac{d}{dx} e^x\right)\right) \cdot x - \cos(x) \cdot e^x \cdot \left(\frac{d}{dx} x\right)}{x^2} \\
 &= \frac{\left(-\sin(x) \cdot e^x + \cos(x) \cdot e^x\right) \cdot x - \cos(x) \cdot e^x \cdot 1}{x^2} \\
 &= \frac{-\sin(x) \cdot e^x \cdot x + \cos(x) \cdot e^x \cdot x - \cos(x) \cdot e^x}{x^2} \\
 &= -\frac{\sin(x) \cdot e^x}{x} + \frac{\cos(x) \cdot e^x}{x} - \frac{\cos(x) \cdot e^x}{x^2}.
 \end{aligned}$$

Beispiel 6.11: Bequemer geht's mit MuPAD. Die Funktion `diff` ist für's Differenzieren von Ausdrücken zuständig:

```
>> diff(cos(x)*exp(x)/x, x)
```

$$\frac{\cos(x) \exp(x)}{x} - \frac{\cos(x) \exp(x)}{x^2} - \frac{\sin(x) \exp(x)}{x}$$

(Vergleiche mit Beispiel 6.10.) Alternativ können Funktionen (aber keine Ausdrücke) mittels `'` differenziert werden:

```
>> f:= x -> cos(x)*exp(x)/x:
```

```
>> f'(x)
```

$$\frac{\cos(x) \exp(x)}{x} - \frac{\cos(x) \exp(x)}{x^2} - \frac{\sin(x) \exp(x)}{x}$$

So setzt man konkrete Werte in die Ableitung ein:

```
>> f'(1), f'(2)
```

$$-\sin(1) \exp(1), \frac{\cos(2) \exp(2)}{4} - \frac{\sin(2) \exp(2)}{2}$$

>> f'(PI) = float(f'(PI))

$$\frac{\exp(\text{PI})}{2} - \frac{\exp(\text{PI})}{\text{PI}} = -5.02126887$$

26.11.03↓

Wie steht's mit der Ableitung von „Hintereinanderschaltungen“ („**Komposition**“) von Funktionen wie z.B. $\sin(\sqrt{x})$?

Satz 6.12: (Die Kettenregel)

Sei $g : D_g \mapsto D_f \subset \mathbb{R}$ differenzierbar am Punkt $x \in D_g$. Sei $f : D_f \mapsto \mathbb{R}$ differenzierbar am Punkt $g(x) \in D_f$. Dann ist die Funktion $h(x) = f(g(x))$ differenzierbar am Punkt x , und es gilt:

$$\frac{d}{dx} h(x) = \frac{d}{dx} f(g(x)) = \underbrace{f'(g(x))}_{\text{„äußere Ableitung“}} \cdot \underbrace{g'(x)}_{\text{„innere Ableitung“}}.$$

Als Merkgel für $y = g(x)$, $z = f(y) = f(g(x))$:

$$\frac{d}{dx} f(g(x)) = \boxed{\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}} = f'(y) \cdot g'(x).$$

Beweis: Es gilt

$$\begin{aligned} & \frac{f(g(x+h)) - f(g(x))}{h} \\ &= \frac{f\left(g(x) + h \cdot \frac{g(x+h) - g(x)}{h}\right) - f(g(x))}{h \cdot \frac{g(x+h) - g(x)}{h}} \cdot \frac{g(x+h) - g(x)}{h}. \end{aligned}$$

Für $h \rightarrow 0$ konvergiert $\frac{g(x+h) - g(x)}{h}$ gegen $g'(x)$ und $k := h \cdot \frac{g(x+h) - g(x)}{h}$ gegen 0:

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{f\left(g(x) + h \cdot \frac{g(x+h) - g(x)}{h}\right) - f(g(x))}{h \cdot \frac{g(x+h) - g(x)}{h}} \\ &= \lim_{k \rightarrow 0} \frac{f(g(x) + k) - f(g(x))}{k} = f'(g(x)). \end{aligned}$$

Q.E.D.

Beispiel 6.13: Für $g(x) = \sqrt{x}$ gilt

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

Zusammen mit $f(y) = \sin(y)$, $f'(y) = \cos(y)$ folgt:

$$\frac{d}{dx} \sin(\underbrace{\sqrt{x}}_y) = \left(\frac{d}{dy} \sin(y) \right) \cdot \left(\frac{d}{dx} \sqrt{x} \right) = \cos(y) \cdot \frac{1}{2} \cdot \frac{1}{\sqrt{x}} = \frac{\cos(\sqrt{x})}{2 \cdot \sqrt{x}}.$$

Definition 6.14: (Höhere Ableitungen)

Die Funktion f sei differenzierbar, sei f' die Ableitungsfunktion. Ist diese wiederum differenzierbar, so heißt $f'' = (f')'$ die „zweite Ableitung von f “. Ist diese wiederum differenzierbar, so heißt $f''' = (f'')'$ die „dritte Ableitung von f “. Usw. Schreibweisen für die n -te Ableitung einer Funktion f :

$$\frac{d^n}{dx^n} f(x) = f^{(n)}(x) = \overbrace{f^{\prime \dots \prime}}^n(x).$$

Die „nullte“ Ableitung $f^{(0)}$ ist die Funktion f selbst. Ist die n -te Ableitung $f^{(n)}$ eine stetige Funktion in x , so heißt f „ n -fach stetig differenzierbar“.

Beispiel 6.15: Offensichtlich gilt $\exp = \exp' = \exp'' = \exp''' = \dots$ etc. Die 4-te Ableitung der trigonometrischen Funktionen ist jeweils wieder die Ausgangsfunktion:

$$\begin{aligned} \frac{d}{dx} \sin(x) &= \cos(x), & \frac{d^2}{dx^2} \sin(x) &= -\sin(x), \\ \frac{d^3}{dx^3} \sin(x) &= -\cos(x), & \frac{d^4}{dx^4} \sin(x) &= \sin(x), \\ \frac{d}{dx} \cos(x) &= -\sin(x), & \frac{d^2}{dx^2} \cos(x) &= -\cos(x), \\ \frac{d^3}{dx^3} \cos(x) &= \sin(x), & \frac{d^4}{dx^4} \cos(x) &= \cos(x). \end{aligned}$$

Beispiel 6.16: Höhere Ableitungen in MuPAD:

```
>> diff(exp(x^2), x, x) // zweite Ableitung
```

$$2 \exp(x^2) + 4 x \exp(x^2)$$

```
>> n := 6:
```

```
>> diff(exp(x^2), x $ n) // n-te Ableitung
```

$$120 \exp(x^2) + 720 x^2 \exp(x^2) + 480 x^4 \exp(x^2) + 64 x^6 \exp(x^2)$$

Mit der Funktion `subs` (engl.: substitute = ersetze; gemeint ist: ersetze x durch einen Wert) kann man konkrete Werte in Ausdrücke einsetzen. Berechne den Wert der 50-ten Ableitung von $\sin(x^2)e^x$ an der Stelle $x = 0$:

```
>> diff(sin(x^2)*exp(x), x $ 50):
```

```
>> subs(%, x = 0)
```

```
- 32812427642492524028780884258717885804750 cos(0) exp(0) -
```

```
9681156701774438433479738001098392167599 sin(0) exp(0)
```

Hier kommt eine Besonderheit von `subs` zutage: der ersetzte Ausdruck wird nicht sofort „ausgewertet“. D.h. in diesem Fall, dass die Vereinfachungen $\cos(0) = 1, \exp(0) = 1, \sin(0) = 0$ nicht automatisch geschehen. Die Funktion `eval` (engl.: evaluate = werte aus) erzwingt die Evaluation:

```
>> eval(%)
```

```
-32812427642492524028780884258717885804750
```

Kennt man die Ableitung einer invertierbaren Funktion f , so kennt man auch die Ableitung der Umkehrabbildung f^{-1} . Es gilt

$$f^{-1}(f(y)) = y.$$

Leitet man beide Seiten der Gleichung nach y ab, so liefert die Kettenregel

$$f^{-1'}(f(y)) \cdot f'(y) = \frac{d}{dy} y = 1 \quad \implies \quad f^{-1'}(f(y)) = \frac{1}{f'(y)}.$$

Satz 6.17: (Ableitung der Inversen)

Sei f differenzierbar und invertierbar, sei f^{-1} die Umkehrfunktion. Ist $f'(y) \neq 0$, so ist f^{-1} an der Stelle $x = f(y)$ differenzierbar, und es gilt

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

Merkregel: mit $y = f^{-1}(x)$, $x = f(y)$:
$$(f^{-1})'(x) = \frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{f'(y)}.$$

Beispiel 6.18: Für $f^{-1} = \ln$ als Umkehrfunktion der Funktion $f = \exp$ mit $f' = \exp$ folgt mit $x = \exp(y)$, $y = \ln(x)$:

$$\frac{d}{dx} \ln(x) = \frac{1}{f'(y)} = \frac{1}{\exp(y)} = \frac{1}{\exp(\ln(x))} = \frac{1}{x}.$$

Hierbei ist $x > 0$ vorausgesetzt (damit $\ln(x)$ definiert ist). Für $x < 0$ gilt

$$\frac{d}{dx} \ln(-x) = \ln'(-x) \cdot \frac{d}{dx} (-x) = \frac{1}{-x} \cdot (-1) = \frac{1}{x}.$$

Für $x > 0$ ist $|x| = x$, für $x < 0$ ist $|x| = -x$. Zusammengefasst gilt damit:

$$\frac{d}{dx} \ln(|x|) = \frac{1}{x} \quad \text{für alle } x \neq 0.$$

An der Stelle $x = 0$ ist $\ln(|x|)$ unstetig und damit erst recht nicht differenzierbar.

6.2 Der Mittelwertsatz

Satz 6.19: (Der Satz von Rolle)

Sei $f : [a, b] \mapsto \mathbb{R}$ differenzierbar auf dem Intervall $[a, b]$. Es gelte $f(a) = f(b)$. Dann gibt es ein $\xi \in (a, b)$ mit $f'(\xi) = 0$.

Beweis: O.b.d.A. sei f nicht konstant (sonst ist die Behauptung sicherlich richtig: Für jedes $\xi \in (a, b)$ gilt $f'(\xi) = 0$). Da f differenzierbar ist, ist f auch stetig. Nach Satz 4.25 gibt es ein Minimum oder ein Maximum ξ von f im Inneren des Intervalls (liegen sowohl das Minimum als auch das Maximum am Rand, müsste die Funktion konstant sein). Sei o.B.d.A. ξ ein Maximum (sonst betrachte $-f$). Mit $f(\xi + h) \leq f(\xi)$ für jedes h folgt für die einseitigen Grenzwerte

$$\lim_{h \rightarrow 0+0} \frac{f(\xi + h) - f(\xi)}{h} \leq 0, \quad \lim_{h \rightarrow 0-0} \frac{f(\xi + h) - f(\xi)}{h} \geq 0.$$

Es folgt $f'(\xi) = \lim_{h \rightarrow 0} \frac{f(\xi + h) - f(\xi)}{h} = 0$.

Q.E.D.

Satz 6.20: (Der Mittelwertsatz)

Sei $f : [a, b] \mapsto \mathbb{R}$ differenzierbar auf dem Intervall $[a, b]$. Dann gibt es ein $\xi \in (a, b)$ mit

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

Beweis: Betrachte $g(x) = f(x) - (f(a) - f(b)) \cdot \frac{x-b}{a-b}$. Diese Funktion erfüllt $g(a) = g(b) = f(b)$. Nach Satz 6.19 existiert $\xi \in (a, b)$ mit

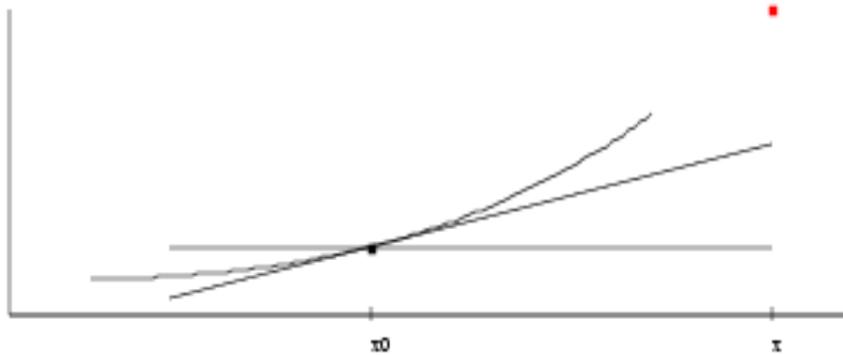
$$g'(\xi) = f'(\xi) - \frac{f(a) - f(b)}{a - b} = 0.$$

Q.E.D.

6.3 Taylor-Reihen

27.11.03↓

Betrachte folgende Funktion, die nur in einer kleinen Umgebung eines Punktes x_0 bekannt ist (genauer: es sind $f(x_0)$, $f'(x_0)$, $f''(x_0)$ etc. bekannt). Man interessiert sich für den Funktionswert an einem Punkt x in der Nähe von x_0 :



In allereinfachster Näherung würde man (für x dicht bei x_0)

$$f(x) \approx f(x_0)$$

setzen. Die nächstbessere Approximation besteht darin, der Tangente am Punkt x_0 zu folgen:

$$f(x) \approx f(x_0) + f'(x_0) \cdot (x - x_0).$$

Im obigen Fall ist deutlich, dass der Funktionswert oberhalb der Tangente zu suchen ist (die Funktion ist „gebogen“: es gilt $f''(x_0) > 0$). Es bietet sich an, einen quadratischen Term hinzuzufügen, um eine bessere Approximation zu erreichen:

$$f(x) \approx f(x_0) + f'(x_0) \cdot (x - x_0) + c \cdot (x - x_0)^2.$$

Wie sollte die Konstante c gewählt werden, wie geht es weiter?

Vorüberlegung zu Taylor-Polynomen 6.21:

Zu einer mehrfach differenzierbaren Funktion f finde ein Polynom

$$T_n(x) = c_0 + c_1 \cdot (x - x_0) + \cdots + c_n \cdot (x - x_0)^n,$$

dass sich an einem Punkt x_0 „möglichst eng an den Graphen von f anschmiegt“. D.h., es soll gelten:

$$f(x_0) = T_n(x_0), \quad f'(x_0) = T_n'(x_0), \quad \dots, \quad f^{(n)}(x_0) = T_n^{(n)}(x_0).$$

Hierdurch ist das Polynom eindeutig bestimmt als

$$T_n(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2!} \cdot (x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!} \cdot (x - x_0)^n.$$

Begründung: Die k -te Ableitung von T_n an der Stelle x_0 ist

$$\begin{aligned} f^{(k)}(x_0) &\stackrel{(!)}{=} T_n^{(k)}(x_0) \\ &= c_k \cdot k! \cdot (x - x_0)^0 + c_{k+1} \cdot (k+1) \cdot k \cdot \dots \cdot 2 \cdot (x - x_0)^1 \Big|_{x=x_0} + \cdots = c_k \cdot k! \\ \Rightarrow c_k &= \frac{f^{(k)}(x_0)}{k!}. \end{aligned}$$

Definition 6.22: (Taylor-Polynome und -Reihen)

Sei f mehrfach am Punkt x_0 differenzierbar. Das Polynom

$$T_n(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k$$

heißt „**Taylor-Polynom**“ n -ten Grades von f am **Entwicklungspunkt** x_0 . Die unendliche Reihe

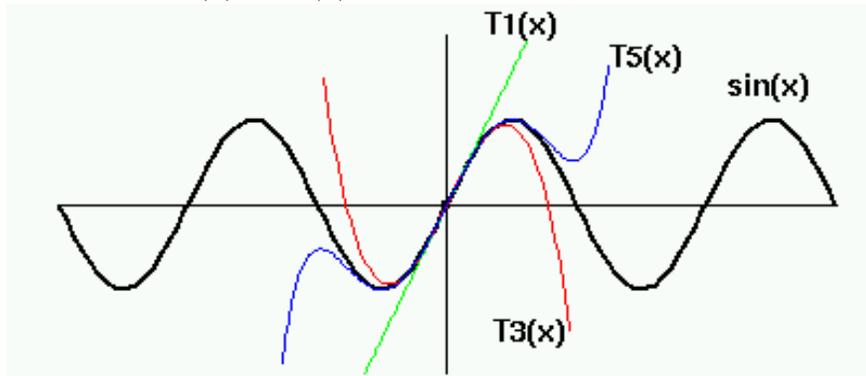
$$T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k$$

heißt „**Taylor-Reihe**“ von f am Entwicklungspunkt x_0 .

Wozu Taylor–Polynome? Taylor–Polynome dienen dazu, komplizierte Funktionen in unmittelbarer Umgebung eines Punktes x_0 durch einfache Funktionen, nämlich Polynome, zu approximieren. Dadurch kann man oft das Verhalten der Funktion in der Nähe spezieller Punkte einfach studieren. Es gilt die

Faustregel: Taylor–Polynome nähern die Funktion an für Werte x , die dicht beim Entwicklungspunkt x_0 liegen: $T_n(x) \approx f(x)$. Je höher n und je kleiner der Abstand $x - x_0$, um so besser ist die Approximation.

Genauer liefert später Satz 6.27. Hier eine Graphik einiger Taylor–Polynome der Funktion $f(x) = \sin(x)$ um den Punkt $x_0 = 0$:



Eine erste Taylor–Reihenberechnung:

Beispiel 6.23: Wir berechnen die Taylor–Reihe von $f(x) = e^x$ um $x_0 = 0$. Wegen $f(x_0) = f'(x_0) = f''(x_0) = \dots = e^{x_0} = e^0 = 1$ ist die Taylor–Reihe

$$e^x = 1 + \frac{1}{1!} \cdot (x - 0) + \frac{1}{2!} \cdot (x - 0)^2 + \dots = 1 + x + \frac{x^2}{2} + \dots$$

Die in Beispiel 3.24 vorgestellte Reihendarstellung der Exponentialfunktion ist also nichts anderes als die Taylor–Entwicklung um den Nullpunkt! Das selbe gilt für die Reihendarstellung der trigonometrischen Funktionen aus Definition 5.10: Mit

$$f(x) = \sin(x), \quad f'(x) = \cos(x), \quad f''(x) = -\sin(x), \quad f^{(3)}(x) = -\cos(x), \quad f^{(4)}(x) = \sin(x)$$

folgt

$$\begin{aligned} f^{(0)}(0) &= f^{(4)}(0) = f^{(8)}(0) = \dots = 0, \\ f^{(1)}(0) &= f^{(5)}(0) = f^{(9)}(0) = \dots = 1, \\ f^{(2)}(0) &= f^{(6)}(0) = f^{(10)}(0) = \dots = 0, \\ f^{(3)}(0) &= f^{(7)}(0) = f^{(11)}(0) = \dots = -1 \end{aligned}$$

$$\Rightarrow \sin(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \cdot x^k = x - \frac{x^3}{3!} + \frac{x^5}{5!} \mp \dots$$

Analog für $f(x) = \cos(x)$:

$$\cos(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \cdot x^k = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \mp \dots$$

Nun eine Anwendung der Taylor-Entwicklung:

Beispiel 6.24: (Vergleiche auch mit Beispiel 4.15) Betrachte die Funktion

$$f(x) = \begin{cases} \frac{1 - \cos(x)}{x^2} & \text{für } x \neq 0, \\ \frac{1}{2} & \text{für } x = 0. \end{cases}$$

Wir behaupten, dass f auch an der Stelle $x = 0$ stetig ist. Wir approximieren $\cos(x)$ durch die Taylor-Entwicklung um den Punkt $x_0 = 0$. Für $x \neq 0$ gilt

$$f(x) = \frac{1 - \cos(x)}{x^2} = \frac{1 - \left(1 - \frac{x^2}{2} + O(x^4)\right)}{x^2} = \frac{\frac{x^2}{2} + O(x^4)}{x^2} = \frac{1}{2} + \frac{O(x^4)}{x^2} = \frac{1}{2} + O(x^2).$$

Hiermit ist nun klar: $\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 0} \left(\frac{1}{2} + O(x^2)\right) = \frac{1}{2}$.

Beispiel 6.25: In MuPAD ist die Funktion `taylor` dafür zuständig, den Beginn einer Taylor-Entwicklung zu berechnen:

```
>> taylor(exp(x), x = 0)
```

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + O(x^6)$$

Die Taylor-Entwicklung von $f(x) = \frac{1}{1-x}$ um $x_0 = 0$ ist die geometrische Reihe aus Beispiel 3.3. Es werden 10 Terme berechnet:

```
>> taylor(1/(1 - x), x = 0, 10)
```

$$1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + O(x^{10})$$

Der folgende Befehl berechnet eine Taylor-Entwicklung um $x_0 = \pi$:

```
>> taylor(2 + sin(x)*cos(x), x = PI)
```

$$2 + (x - \text{PI}) - \frac{(x - \text{PI})^3}{3} + \frac{(x - \text{PI})^5}{15} + O((x - \text{PI})^6)$$

Beispiel 6.26: Betrachte $f(x) = 1 - \sqrt{1-x} = 1 - (1-x)^{\frac{1}{2}}$. Wie kann man Werte $f(x)$ für kleines x ohne technische Hilfsmittel ausrechnen? Zunächst die Berechnung der ersten Taylor-Polynome. Als Entwicklungspunkt wählen wir $x_0 = 0$, da wir uns für **kleine** Werte von x interessieren. Man braucht Ableitungen von $f(x)$ am Entwicklungspunkt $x_0 = 0$:

$$\begin{aligned} f(x) &= 1 - (1-x)^{\frac{1}{2}}, & f(0) &= 0, \\ f'(x) &= \frac{1}{2} \cdot (1-x)^{-\frac{1}{2}}, & f'(0) &= \frac{1}{2}, \\ f''(x) &= \frac{1}{4} \cdot (1-x)^{-\frac{3}{2}}, & f''(0) &= \frac{1}{4}, \\ &\dots \end{aligned}$$

Hiermit folgt die Entwicklung

$$\begin{aligned} f(x) = 1 - \sqrt{1-x} &\approx f(0) + f'(0) \cdot (x - x_0) + \frac{f''(0)}{2!} \cdot (x - x_0)^2 + \dots \\ &= 0 + \frac{x}{2} + \frac{x^2}{8} + \dots \end{aligned}$$

Nun ja, die Terme der Entwicklung sind in der Tat so alle berechenbar, aber das ist ziemlich mühselig. Bequemer mit MuPAD:

```
>> taylor(1 - sqrt(1 - x), x)
```

$$\begin{aligned} & x^2 & x^3 & x^4 & x^5 & & \\ & - + \frac{x^2}{2} + \frac{x^3}{8} + \frac{x^4}{16} + \frac{x^5}{128} + \frac{x^6}{256} + O(x^6) \end{aligned}$$

Aus diesen Taylor-Approximationen bekommt man z.B. für $x = 0.1$:

$$\begin{aligned} f(0.1) &= \frac{0.1}{2} + \frac{0.1^2}{8} + \frac{0.1^3}{16} + \dots \\ &= 0.05 \\ &\quad + 0.00125 \\ &\quad + 0.0000625 \\ &\quad + \dots \\ &= \frac{\quad}{0.05131\dots} \end{aligned}$$

Man sieht der Entwicklung geradezu an, dass die noch nicht berücksichtigten Terme der Entwicklung die angegebenen Dezimalstellen nicht mehr beeinflussen, d.h., die ersten 3 bis 4 Ziffern sind korrekt. Probe mit MuPAD:

```
>> 1 - sqrt(0.9)
```

```
0.05131670195
```

Für Taylor-Polynome endlichen Grades ist es zumindestens intuitiv klar, dass sie eine Approximation der Funktion liefern, wenn nur x dicht genug beim Entwicklungspunkt x_0 liegt. Es verbleibt jedoch zu klären, ob die unendliche Reihe gegen $f(x)$ konvergiert (bzw., wie weit entfernt x von x_0 liegen darf, damit $f(x)$ durch die Taylor-Reihe dargestellt wird).

Satz 6.27: (Restgliedformel der Taylor-Approximation)

Sei $f(x)$ in einer Umgebung des Punktes x_0 $(n + 1)$ -fach stetig differenzierbar. Sei x aus dieser Umgebung. Dann existiert ein Punkt ξ im offenen Intervall zwischen x und x_0 , so dass gilt:

$$f(x) = \underbrace{\sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k}_{\text{Taylor-Polynom vom Grad } n} + \underbrace{\frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot (x - x_0)^{n+1}}_{\text{„Restglied“}}$$

Beweis: (für technisch Interessierte) Wir halten x fest und fassen das Taylor-Polynom als Funktion des Entwicklungspunkts x_0 auf:

$$T_n(t) = \sum_{k=0}^n \frac{f^{(k)}(t)}{k!} \cdot (x - t)^k.$$

Die Ableitung dieser Funktion ist eine Teleskopsumme:

$$\begin{aligned} \frac{d}{dt} T_n(t) &= \sum_{k=0}^n \frac{f^{(k+1)}(t)}{k!} \cdot (x - t)^k - \sum_{k=0}^n \frac{f^{(k)}(t)}{k!} \cdot k \cdot (x - t)^{k-1} \\ &= \sum_{k=0}^n \frac{f^{(k+1)}(t)}{k!} \cdot (x - t)^k - \sum_{k=1}^n \frac{f^{(k)}(t)}{(k-1)!} \cdot (x - t)^{k-1} \\ &= \sum_{k=0}^n \frac{f^{(k+1)}(t)}{k!} \cdot (x - t)^k - \sum_{k=0}^{n-1} \frac{f^{(k+1)}(t)}{k!} \cdot (x - t)^k \\ &= \frac{f^{(n+1)}(t)}{n!} \cdot (x - t)^n. \end{aligned}$$

Betrachte die Hilfsfunktion

$$g(t) = (x - x_0)^{n+1} \cdot T_n(t) + (x - t)^{n+1} \cdot (f(x) - T_n(x_0))$$

mit festem x und x_0 , für die

$$\begin{aligned} g(x_0) &= (x - x_0)^{n+1} \cdot T_n(x_0) + (x - x_0)^{n+1} \cdot (f(x) - T_n(x_0)) \\ &= (x - x_0)^{n+1} \cdot f(x), \\ g(x) &= (x - x_0)^{n+1} \cdot T_n(x) = (x - x_0)^{n+1} \cdot f(x) \end{aligned}$$

gilt, also $g(x) = g(x_0)$. Nach dem Satz von Rolle 6.19 gibt es ein ξ im offenen Intervall zwischen x und x_0 , wo die Ableitung

$$\begin{aligned} \frac{d}{dt} g(t) &= (x - x_0)^{n+1} \cdot \frac{d}{dt} T_n(t) - (n+1) \cdot (x-t)^n \cdot (f(x) - T_n(x_0)) \\ &= (x - x_0)^{n+1} \cdot \frac{f^{(n+1)}(t)}{n!} \cdot (x-t)^n - (n+1) \cdot (x-t)^n \cdot (f(x) - T_n(x_0)) \\ &= (x-t)^n \cdot \left((x-x_0)^{n+1} \cdot \frac{f^{(n+1)}(t)}{n!} - (n+1) \cdot (f(x) - T_n(x_0)) \right) \end{aligned}$$

verschwindet:

$$\begin{aligned} 0 &= (x - \xi)^n \cdot \left((x - x_0)^{n+1} \cdot \frac{f^{(n+1)}(\xi)}{n!} - (n+1) \cdot (f(x) - T_n(x_0)) \right) \\ &\Rightarrow f(x) - T_n(x_0) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot (x - x_0)^{n+1}. \end{aligned}$$

Q.E.D.

2.11.03↓

Interpretation 6.28:

Das Restglied

$$\frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot (x - x_0)^{n+1} = f(x) - \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k$$

ist die Differenz zwischen der Funktion $f(x)$ und dem n -ten Taylor-Polynom um x_0 . Die Funktion wird genau dann durch die unendliche Taylor-Reihe dargestellt, wenn das Restglied bei festem x , x_0 für $n \rightarrow \infty$ gegen 0 konvergiert. Sind z.B. alle Ableitungen $f^{(n)}$ beschränkt bzgl. n , so ist dies für beliebiges x und x_0 der Fall, denn $n!$ wächst schneller gegen ∞ als $|x - x_0|^n$ für jeden Wert von $|x - x_0|$. Dies erklärt z.B., dass die trigonometrischen Funktionen \sin und \cos , deren Ableitungen nur Werte in $[-1, 1]$ annehmen, global durch ihre Taylor-Reihen dargestellt werden (wir haben sie in Definition 5.10 ja auch über diese Reihen eingeführt). Dies gilt auch für \exp , da $f^{(n)}(x) = e^x$ gar nicht von n abhängt.

Beispiel 6.29: Wir betrachten die Taylor-Entwicklung von $f(x) = \ln(1+x)$ um den Punkt $x_0 = 0$:

$$\begin{aligned} f(x) &= \ln(1+x), \quad f'(x) = \frac{1}{1+x}, \quad f''(x) = -\frac{1}{(1+x)^2}, \\ f^{(3)}(x) &= \frac{2}{(1+x)^3}, \quad \dots, \quad f^{(k)}(x) = \frac{(-1)^{k-1} \cdot (k-1)!}{(1+x)^k}. \end{aligned}$$

Mit $f^{(k)}(0) = (-1)^{k-1} \cdot (k-1)!$ folgt als Taylor-Reihe

$$\ln(1+x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \cdot x^k = \sum_{k=1}^{\infty} (-1)^{k-1} \cdot \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} \mp \dots,$$

die die Funktion darstellt, solange die Restglieder

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

gegen 0 konvergieren. Dies ist für positives $x \leq 1$ mit $0 < \xi < x \leq 1$ offensichtlich der Fall:

$$\frac{x^{n+1}}{(1+\xi)^{n+1} \cdot (n+1)} \leq \frac{x^{n+1}}{n+1} \leq \frac{1}{n+1} \xrightarrow{n \rightarrow \infty} 0.$$

Speziell für $x = 1$ ergibt sich der Wert der alternierenden harmonischen Reihe:

$$\ln(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \dots$$

Für negatives $x \geq -\frac{1}{2}$ gilt $-\frac{1}{2} \leq x < \xi < 0$:

$$\left| \frac{x^{n+1}}{(1+\xi)^{n+1} \cdot (n+1)} \right| = \frac{|x|^{n+1}}{(1+\xi)^{n+1} \cdot (n+1)} \leq \frac{(1/2)^{n+1}}{(1/2)^{n+1} \cdot (n+1)} = \frac{1}{n+1} \xrightarrow{n \rightarrow \infty} 0,$$

d.h., auch hier konvergiert das Restglied gegen 0. Weiterhin konvergiert die Taylor-Reihe auch für $-1 < x < -\frac{1}{2}$ gegen $\ln(1+x)$, was wir aus unserer Restgliedformel allerdings nicht herausbekommen (es gibt alternative Restgliedformeln, die dieses Resultat liefern). Zusammengefasst:

$$\boxed{\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \pm \dots \text{ für } x \in (-1, 1].}$$

Für $|x| > 1$ sowie für $x = -1$ divergiert die Taylor-Reihe.

6.4 Monotonie, Extremwerte

↓3.12.03

Eine der wichtigsten Anwendungen der Differentiation ist das Auffinden von Extremwerten. Dazu stellen wir zunächst fest, dass Ableitungswerte (= Tangentensteigungen) auf ansteigendes oder abfallendes Verhalten der Funktion hinweisen:

Satz 6.30: (Ableitungen weisen auf Monotonie hin)

Sei f differenzierbar, die Ableitungsfunktion f' sei stetig. Gilt $f'(x_0) > 0$, so ist f auf einer Umgebung von x_0 streng monoton steigend. Gilt $f'(x_0) < 0$, so ist f auf einer Umgebung von x_0 streng monoton fallend.

Beweis: Da f' stetig ist, gilt für $f'(x_0) > 0$, dass f' auch noch auf einer Umgebung von x_0 positiv ist. Für x, y aus dieser Umgebung von x_0 mit $x < y$ liefert der Mittelwertsatz 6.20

$$f(y) - f(x) = f'(\xi) \cdot (y - x) > 0$$

mit einem Zwischenwert ξ zwischen x und y . Damit ist $f(x)$ streng monoton steigend auf einer Umgebung des Punktes x , auf der für den Zwischenwert $f'(\xi) > 0$ gilt.

Analog folgt, dass $f(x)$ streng monoton fallend ist, wenn mit $f'(x_0) < 0$ die Ableitung auf einer Umgebung von x_0 negative Werte annimmt.

Q.E.D.

Intuitiv: mit der Interpretation der Ableitung 6.4 ist dies unmittelbar klar. Für kleines Δx gilt:

$$f(x_0 + \Delta x) \approx f(x_0) + f'(x_0) \cdot \Delta x.$$

Extrema sind die Stellen, wo die Funktion „auf der einen Seite“ steigend, „auf der anderen Seite“ fallend ist:

Satz 6.31: (An Extremstellen verschwindet die Ableitung)

Sei f differenzierbar. Ist die Stelle x_0 ein (lokales) Maximum oder Minimum, so gilt $f'(x_0) = 0$.

Man findet also alle Kandidaten für Extremstellen einer Funktion f , indem man die Nullstellen von f' sucht.

Beweis: Genau wie im Beweis des Satzes von Rolle 6.19.

Q.E.D.

Beispiel 6.32: Betrachte $f(x) = 2 \cdot x - x^2$:

$$\frac{d}{dx} f(x) = \frac{d}{dx} (2 \cdot x - x^2) = 2 - 2 \cdot x \stackrel{!}{=} 0 \implies x = 1.$$

Damit ist $x_0 = 1$ der einzige Punkt, an dem (möglicherweise) ein Extremum vorliegen kann.

Es gibt allerdings Stellen x_0 mit $f'(x_0) = 0$, die keine Extremstellen (sondern sogenannte „Sattelpunkte“) sind. Beispiel: die Funktion $f(x) = x^3$ ist streng monoton steigend. Am Punkt $x_0 = 0$ gilt $f'(x_0) = 3 \cdot x_0^2 = 0$, aber x_0 ist kein Extremum.

Satz 6.33: (Hinreichende Kriterien für Extrema)

Sei f mehrfach differenzierbar. Gilt an einer Stelle x_0

$$f'(x_0) = 0, \quad f''(x_0) < 0,$$

so ist x_0 ein lokales Maximum. Gilt

$$f'(x_0) = 0, \quad f''(x_0) > 0,$$

so ist x_0 ein lokales Minimum.

„**Beweis**“: Approximiere $f(x)$ in einer Umgebung von x_0 durch das Taylor-Polynom zweiten Grades:

$$f(x) \approx f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2} \cdot (x - x_0)^2.$$

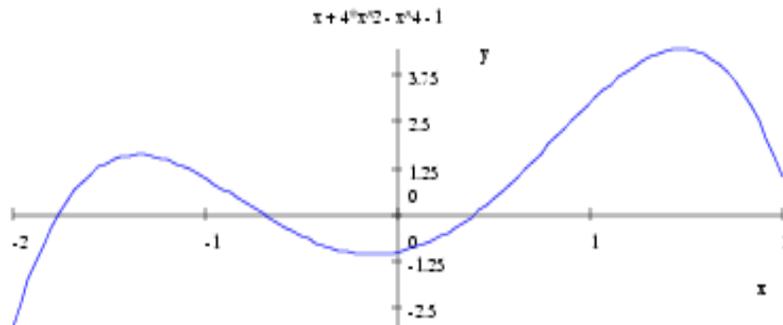
An einem Punkt x_0 mit $f'(x_0) = 0$ gilt näherungsweise:

$$f(x) \approx f(x_0) + \frac{f''(x_0)}{2} \cdot (x - x_0)^2.$$

Da $(x - x_0)^2 > 0$ für $x \neq x_0$ ist, sind die Funktionswerte in der Umgebung größer als $f(x_0)$, wenn $f''(x_0) > 0$ gilt (Minimum). Für $f''(x_0) < 0$ sind die Funktionswerte in der Umgebung kleiner als $f(x_0)$ (Maximum).

Beispiel 6.34: Betrachte $f(x) = x + 4x^2 - x^4 - 1$:

```
>> f:= x -> x + 4*x^2 - x^4 - 1:
>> plotfunc2d(f(x), x = -2..2)
```



Um die Kandidaten für die Extrema zu finden, werden (numerische Approximationen der) Lösungen der Gleichung $f'(x) = 0$ berechnet. Für numerische Lösungen sind die MuPAD-Funktionen `numeric::solve` oder auch `numeric::fsolve` zuständig. Für polynomiale Gleichungen wird eine Menge aller Lösungen geliefert. Die einzelnen Lösungen lassen sich durch „indizierten Zugriff“ `Kandidaten[1]` etc. auswählen:

```
>> Kandidaten:= numeric::solve(f'(x) = 0, x)
      {-1.346997409, -0.1260001926, 1.472997601}
```

Diese Werte werden in die 2-te Ableitung von f eingesetzt:

```
>> f''(Kandidaten[1])
      -13.77282422
>> f''(Kandidaten[2])
      7.809487418
>> f''(Kandidaten[3])
      -18.0366632
```

Nach Satz 6.33 ist der erste Kandidat ein Maximum, der zweite Kandidat ein Minimum, der dritte Kandidat ein Maximum. Die Graphik bestätigt dies.

6.5 Die de l'Hospitalsche Regel

In $\frac{0}{0}$ -Situationen kann man durch Ableiten auch Grenzwerte bestimmen.

Satz 6.35: (de l'Hospitalsche Regel)

Seien f und g differenzierbar, es gelte $f(x_0) = g(x_0) = 0$. Dann gilt

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)},$$

falls der rechte Grenzwert existiert.

„**Beweis:**“ Intuitiv: Approximiere Zähler und Nenner durch das Taylor-Polynom ersten Grades:

$$\frac{f(x)}{g(x)} \approx \frac{f(x_0) + f'(x_0) \cdot (x - x_0)}{g(x_0) + g'(x_0) \cdot (x - x_0)} = \frac{f'(x_0) \cdot (x - x_0)}{g'(x_0) \cdot (x - x_0)} = \frac{f'(x_0)}{g'(x_0)}.$$

Für eine saubere Durchführung des Beweises benutze man den Mittelwertsatz 6.20 (unter der Zusatzannahme, dass f' und g' stetig seien. Die Regel gilt aber auch ohne diese Stetigkeit.)

Q.E.D.

Beispiel 6.36: Betrachte erneut die Funktion

$$f(x) = \begin{cases} \frac{e^x - 1}{x} & \text{für } x \neq 0, \\ 1 & \text{für } x = 0 \end{cases}$$

aus Beispiel 4.15. Für den Punkt $x_0 = 0$ liegt eine $\frac{0}{0}$ -Situation vor. Mit de l'Hospital folgt

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = \lim_{x \rightarrow 0} \frac{\frac{d}{dx}(e^x - 1)}{\frac{d}{dx}x} = \lim_{x \rightarrow 0} \frac{e^x}{1} = \lim_{x \rightarrow 0} e^x = e^0 = 1,$$

wobei in jedem Schritt die Existenz des jeweils rechts stehenden Grenzwerts vorausgesetzt wird (was gerechtfertigt ist, sobald man ganz rechts angekommen ist).

Die de l'Hospitalsche Regel kann auch mehrfach hintereinander angewendet werden:

Beispiel 6.37: Betrachte $\lim_{x \rightarrow 0} \frac{e^{2 \cdot x} - 1 - 2 \cdot x}{x^2}$. Nach einer Anwendung von de l'Hospital trifft man beim Quotienten der Ableitungen wieder auf eine $\frac{0}{0}$ -Situation und kann de l'Hospital erneut anwenden:

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{e^{2 \cdot x} - 1 - 2 \cdot x}{x^2} &= \lim_{x \rightarrow 0} \frac{\frac{d}{dx}(e^{2 \cdot x} - 1 - 2 \cdot x)}{\frac{d}{dx}x^2} = \lim_{x \rightarrow 0} \frac{2 \cdot e^{2 \cdot x} - 2}{2 \cdot x} \\ &= \lim_{x \rightarrow 0} \frac{e^{2 \cdot x} - 1}{x} = \lim_{x \rightarrow 0} \frac{\frac{d}{dx}(e^{2 \cdot x} - 1)}{\frac{d}{dx}x} = \lim_{x \rightarrow 0} \frac{2 \cdot e^{2 \cdot x}}{1} = 2 \cdot e^{2 \cdot 0} = 2. \end{aligned}$$

Bemerkung 6.38: Die de l'Hospitalsche Regel

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$$

gilt auch für $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = \infty$.

Beispiel 6.39: Mit $f(x) = \ln(x + 1)$, $g(x) = \ln(x)$, $f'(x) = \frac{1}{x+1}$, $g'(x) = \frac{1}{x}$:

$$\lim_{x \rightarrow \infty} \frac{\ln(x + 1)}{\ln(x)} = \lim_{x \rightarrow \infty} \frac{\frac{1}{x+1}}{\frac{1}{x}} = \lim_{x \rightarrow \infty} \frac{x}{x + 1} \stackrel{(*)}{=} \lim_{x \rightarrow \infty} \frac{1}{1} = 1,$$

wobei in (*) de l'Hospital ein zweites Mal angewendet wurde.

Beispiel 6.40: Mit kleinen Tricks bekommt man eine de l'Hospital-Technik auch sofort für Situationen wie z.B. $0 \cdot \infty$ oder auch 1^∞ .

Für $0 \cdot \infty$ ist der Standardtrick, ∞ als $1/0$ (oder manchmal 0 als $1/\infty$) zu schreiben. Z.B.:

$$\lim_{x \rightarrow \infty} x \cdot (e^{1/x} - 1) = \lim_{x \rightarrow \infty} \frac{e^{1/x} - 1}{1/x} = \lim_{x \rightarrow \infty} \frac{\frac{d}{dx}(e^{1/x} - 1)}{\frac{d}{dx} \frac{1}{x}}$$

$$= \lim_{x \rightarrow \infty} \frac{-\frac{1}{x^2} \cdot e^{1/x}}{-\frac{1}{x^2}} = \lim_{x \rightarrow \infty} e^{1/x} = 1.$$

Hierbei wurde die ursprüngliche $\infty \cdot 0$ -Situation durch das Umschreiben $x = \frac{1}{1/x}$ in eine $\frac{0}{0}$ -Situation verwandelt, auf die de l'Hospital anwendbar ist.

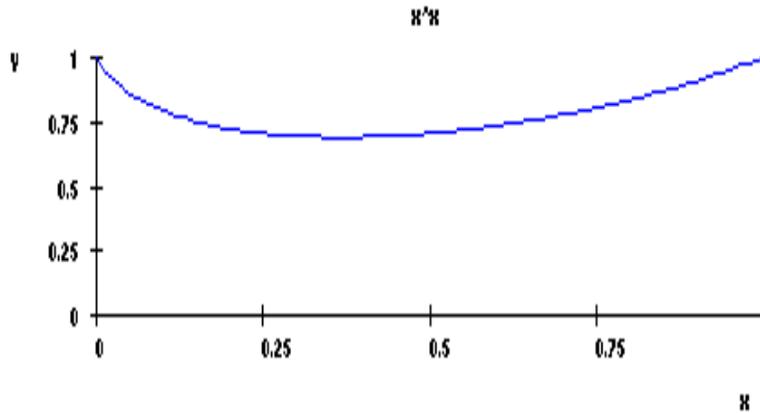
Für eine 1^∞ -Situation ist der Standardtrick, die identische Abbildung in der Form $y = \exp(\ln(y))$ einzubringen, was die 1^∞ -Situation in ein $0 \cdot \infty$ -Problem verwandelt (welches dann wie oben zu behandeln ist). Beispiel:

$$\lim_{x \rightarrow 0+0} x^x = \lim_{x \rightarrow 0+0} e^{\ln(x^x)} = \lim_{x \rightarrow 0+0} e^{x \cdot \ln(x)} = e^{\lim_{x \rightarrow 0+0} (x \cdot \ln(x))}.$$

Hier ist das $0 \cdot (-\infty)$ -Problem $\lim_{x \rightarrow 0+0} x \cdot \ln(x)$ entstanden, was wie oben per de l'Hospital gelöst wird, indem es in ein $\frac{\infty}{\infty}$ -Problem (genauer: in ein $\frac{-\infty}{\infty}$ -Problem) umgeschrieben wird:

$$\begin{aligned} \lim_{x \rightarrow 0+0} x \cdot \ln(x) &= \lim_{x \rightarrow 0+0} \frac{\ln(x)}{1/x} = \lim_{x \rightarrow 0+0} \frac{\frac{d}{dx} \ln(x)}{\frac{d}{dx} \frac{1}{x}} = \lim_{x \rightarrow 0+0} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim_{x \rightarrow 0+0} (-x) = 0 \\ \Rightarrow \lim_{x \rightarrow 0+0} x^x &= e^{\lim_{x \rightarrow 0+0} x \cdot \ln(x)} = e^0 = 1. \end{aligned}$$

Der Grenzwert wird durch die folgende MuPAD-Graphik bestätigt:



Kapitel 7

Potenzreihen

7.1 Der Konvergenzradius

Definition 7.1: (Komplexe Potenzreihen)

Eine „Potenzreihe um den Punkt $z_0 \in \mathbb{C}$ “ ist eine Reihe der Form

$$\sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k, \quad a_k, z, z_0 \in \mathbb{C}.$$

Dort, wo die Reihe konvergiert, definiert sie eine Funktion von z , deren Eigenschaften untersucht werden sollen.

Satz 7.2: (Konvergenz von Potenzreihen)

Zur Potenzreihe $\sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k$ existiert ein $r \geq 0$ (der „**Konvergenzradius**“), so dass

- a) die Reihe für alle z mit $|z - z_0| < r$ **absolut** konvergiert,
- b) die Reihe für kein z mit $|z - z_0| > r$ konvergiert.

Beweis: Gibt es einen Punkt Z , in dem die Reihe $\sum_k a_k \cdot (Z - z_0)^k$ konvergiert, so bildet $a_k \cdot (Z - z_0)^k$ eine Nullfolge. Es gilt also $|a_k \cdot (Z - z_0)^k| \leq 1$ für hinreichend grosses $k \geq k_0$. Mit

$$\sum_{k=k_0}^{\infty} |a_k \cdot (z - z_0)^k| = \sum_{k=k_0}^{\infty} \underbrace{|a_k \cdot (Z - z_0)^k|}_{\leq 1} \cdot \left| \frac{z - z_0}{Z - z_0} \right|^k$$

ist also für jedes z mit $|z - z_0| < |Z - z_0|$ die geometrische Reihe $\sum_k \left| \frac{z - z_0}{Z - z_0} \right|^k$ eine konvergente Majorante, d.h., $\sum_k a_k \cdot (z - z_0)^k$ konvergiert absolut. Damit

ist

$$r = \sup \left\{ |Z - z_0|; \sum_k a_k \cdot (Z - z_0)^k \text{ konvergiert} \right\},$$

falls die Menge der Konvergenzpunkte beschränkt ist. Konvergiert die Reihe für alle $Z \in \mathbb{C}$, setzt man formal $r = \infty$. Für jedes z mit $|z - z_0| > r$ muss die Reihe nach dieser Konstruktion von r divergieren.

Q.E.D.

Bemerkung 7.3: Der Konvergenzbereich einer Potenzreihe besteht also prinzipiell aus einer Kreisscheibe um den Entwicklungspunkt z_0 . Der Radius r kann allerdings 0 sein (d.h., die Potenzreihe konvergiert nur am Punkt $z = z_0$). Im Folgenden interessieren natürlich nur Potenzreihen mit einem Konvergenzradius $r > 0$. Über den Rand des Konvergenzkreises $\{z \in \mathbb{C}; |z - z_0| = r\}$ kann man keine allgemeine Aussagen machen. In folgendem Beispiel konvergiert die Reihe für keinen der Randpunkte (denn für $|z| = 1$ ist z^k keine Nullfolge):

$$\sum_{k=0}^{\infty} z^k \quad (\text{Konvergenz für } |z| < r = 1.)$$

In folgendem Beispiel konvergiert die Reihe für alle Randpunkte mit $|z| = 1$ außer für $z = 1$ ($z = 1$ führt auf die divergente harmonische Reihe):

$$\sum_{k=0}^{\infty} \frac{z^k}{k} \quad (\text{Konvergenz für } |z| \leq r = 1, z \neq 1).$$

Im folgenden Beispiel konvergiert die Reihe für alle Randpunkte:

$$\sum_{k=0}^{\infty} \frac{z^k}{k^2} \quad (\text{Konvergenz für } |z| \leq r = 1.)$$

Beispiel 7.4: Die geometrische Reihe

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1 - z}$$

hat den Konvergenzradius 1. Die Reihe

$$\sum_{k=0}^{\infty} (-1)^k \cdot z^{2 \cdot k} = \sum_{k=0}^{\infty} (-z^2)^k = \frac{1}{1 - (-z^2)} = \frac{1}{1 + z^2}$$

hat ebenfalls den Konvergenzradius 1.

Der Konvergenzradius einer Potenzreihe kann unmittelbar aus den Koeffizienten a_k der Reihe bestimmt werden:

Satz 7.5: (Cauchy-Hadamard-Formel für den Konvergenzradius)

Der Konvergenzradius der Reihe $\sum_{k=1}^{\infty} a_k \cdot (z - z_0)^k$ ist $r = \frac{1}{\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|}}$,
wenn $\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|}$ existiert. Gilt $\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|} = 0$, so setze $r = \infty$. Gilt
 $\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|} = \infty$, so setze $r = 0$.

Beweis: Der Grenzwert $\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|}$ existiere. Betrachte ein beliebiges z mit $|z - z_0| < r$. Dann gilt $c := \lim_{k \rightarrow \infty} \sqrt[k]{|a_k|} \cdot |z - z_0| < 1$. Es folgt, dass für alle $k \geq N$ (mit geeignetem N)

$$\sqrt[k]{|a_k|} \cdot |z - z_0| \leq C < 1 \quad \text{mit } C := \frac{1+c}{2} \in [\frac{1}{2}, 1)$$

gelten muss. Damit kann die Reihe durch eine konvergente geometrische Reihe abgeschätzt werden:

$$\left| a_k \cdot (z - z_0)^k \right| = \left(\sqrt[k]{|a_k|} \cdot |z - z_0| \right)^k \leq C^k \quad \forall k \geq N.$$

Nach Satz 3.16 konvergiert $\sum_k |a_k| \cdot |z - z_0|^k$, da $\sum_k C^k$ mit $C < 1$ eine konvergente Majorante ist. Dies ist die absolute Konvergenz von $\sum_k a_k \cdot (z - z_0)^k$. Das Argument greift auch für den Grenzfall $r = \infty$.

Betrachte nun ein beliebiges z mit $|z - z_0| > r$. Diesmal gilt

$$\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|} \cdot |z - z_0| > 1,$$

also

$$\left| a_k \cdot (z - z_0)^k \right| = \left(\sqrt[k]{|a_k|} \cdot |z - z_0| \right)^k > 1$$

für große k , d.h., die Reihenglieder von $\sum_k a_k \cdot (z - z_0)^k$ bilden keine Nullfolge. Nach Satz 3.8 kann die Reihe $\sum_k a_k \cdot (z - z_0)^k$ nicht konvergieren. Das Argument greift auch für den Grenzfall $r = 0$.

Q.E.D.

Bemerkung 7.6: Die allgemeine Form der Cauchy-Hadamard-Formel ist

$$\text{Konvergenzradius} = \frac{1}{\limsup_{k \rightarrow \infty} \sqrt[k]{|a_k|}},$$

wo \limsup („Limes superior“) der größte Häufungspunkt der Folge $\sqrt[k]{|a_k|}$ ist. Der Limes superior existiert für jede reelle Folge (bei unbeschränkten Folgen definiert man ihn formal als ∞).

7.2 Eigenschaften von Potenzreihen

Auf dem Inneren des Konvergenzkreises sind durch Potenzreihen dargestellte Funktionen „beliebig harmlos und angenehm“. Sie sind automatisch unendlich oft diff'bar und werden durch ihre Taylor-Entwicklung dargestellt (Funktionen, die durch ihre Taylor-Reihen dargestellt werden, nennt man **analytisch**). Dies darf nicht verwundern, denn (gliedweise Differenzierbarkeit mal vorausgesetzt):

$$\begin{aligned} \frac{d^n}{dz^n} \Big|_{z=z_0} \underbrace{\sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k}_{f(z)} &= \left(n! \cdot a_n + (\dots) \cdot (z - z_0) + (\dots) \cdot (z - z_0)^2 + \dots \right)_{z=z_0} \\ &= n! \cdot a_n, \end{aligned}$$

also $a_n = f^{(n)}(z_0)/n!$.

Damit ist die Potenzreihe ihre eigene Taylor-Reihe:

$$f(z) = \sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(z_0)}{k!} \cdot (z - z_0)^k.$$

Diese heuristische Überlegung gilt in der Tat:

Satz 7.7: (Potenzreihen stellen analytische Funktionen dar)

Auf dem Inneren des Konvergenzkreises $\{z \in \mathbb{C}; |z - z_0| < r\}$ einer Potenzreihe mit Konvergenzradius $r > 0$ stellt

$$f(z) = \sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k$$

eine unendlich oft differenzierbare Funktion dar. Es gilt

$$\frac{d^n}{dz^n} f(z) = \sum_{k=n}^{\infty} a_k \cdot k \cdot (k-1) \cdot \dots \cdot (k-n+1) \cdot (z - z_0)^{k-n},$$

d.h., die Potenzreihe kann gliedweise differenziert werden. Der Konvergenzradius der abgeleiteten Reihen ist wiederum r . Speziell gilt $a_k = f^{(k)}(z_0)/k!$.

Beweis: Für die Potenzreihe ist der Differenzenquotient

$$\frac{f(z+h) - f(z)}{h} = \sum_{k=1}^{\infty} a_k \cdot \frac{(z+h-z_0)^k - (z-z_0)^k}{h}.$$

Mit der Binomialentwicklung

$$\begin{aligned} \frac{(z+h-z_0)^k - (z-z_0)^k}{h} &= \frac{1}{h} \cdot \sum_{j=1}^k \binom{k}{j} \cdot h^j \cdot (z-z_0)^{k-j} \\ &= k \cdot (z-z_0)^{k-1} + h \cdot \sum_{j=2}^k \binom{k}{j} \cdot h^{j-2} \cdot (z-z_0)^{k-j} \end{aligned}$$

folgt

$$\frac{f(z+h) - f(z)}{h} = \sum_{k=1}^{\infty} a_k \cdot k \cdot (z-z_0)^{k-1} + h \cdot \underbrace{\sum_{k=2}^{\infty} \sum_{j=2}^k a_k \cdot \binom{k}{j} \cdot h^{j-2} \cdot (z-z_0)^{k-j}}_{g(h)}.$$

Die die Funktion $g(h)$ definierende Reihe ist dabei wohldefiniert, da die linke Seite der Gleichung für hinreichend kleines h definiert ist ($z+h$ muss im Konvergenzradius von $f(z)$ liegen) und die Reihe $\sum_k a_k \cdot k \cdot (z-z_0)^{k-1}$ konvergiert (in den Übungen wird gezeigt, dass $f(z) = \sum_k a_k (z-z_0)^k$ und $f'(z) = \sum_k a_k \cdot k \cdot (z-z_0)^{k-1}$ den selben Konvergenzradius haben). Die Funktion $g(h)$ ist beschränkt in h , denn für $|h| \leq |h_0|$ (mit noch zu wählendem h_0) gilt:

$$\begin{aligned} |g(h)| &= \left| \sum_{k=2}^{\infty} \sum_{j=2}^k a_k \cdot \binom{k}{j} \cdot h^{j-2} \cdot (z-z_0)^{k-j} \right| \leq \sum_{k=2}^{\infty} \sum_{j=2}^k |a_k| \cdot \binom{k}{j} \cdot |h_0|^{j-2} \cdot |z-z_0|^{k-j} \\ &\leq \sum_{k=2}^{\infty} |a_k| \sum_{j=2}^k \binom{k}{j} \cdot |h_0|^{j-2} \cdot |z-z_0|^{k-j} \leq \frac{1}{|h_0|^2} \sum_{k=2}^{\infty} |a_k| \sum_{j=0}^k \binom{k}{j} \cdot |h_0|^j \cdot |z-z_0|^{k-j} \\ &= \frac{1}{|h_0|^2} \sum_{k=2}^{\infty} |a_k| (|z-z_0| + |h_0|)^k. \end{aligned}$$

Ist $r > 0$ der Konvergenzradius von $\sum_k a_k \cdot (z-z_0)^k$ (und damit auch von $\sum_k |a_k| \cdot |z-z_0|^k$), so wähle $|h_0| = (r - |z-z_0|)/2$, womit $|z-z_0| + |h_0| < r$ gilt, d.h., die obige Reihe konvergiert und liefert eine Schranke für $|g(h)|$.

Damit folgt $h \cdot g(h) = O(h)$:

$$\begin{aligned} f'(z) &= \lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h} = \sum_{k=1}^{\infty} a_k \cdot k \cdot (z-z_0)^{k-1} + \lim_{h \rightarrow 0} O(h) \\ &= \sum_{k=1}^{\infty} a_k \cdot k \cdot (z-z_0)^{k-1}. \end{aligned}$$

Damit ist gezeigt, dass die Potenzreihe einmal differenzierbar ist. Die Ableitung ist wieder als Potenzreihe dargestellt.

Die höheren Ableitungen folgen nun sofort per Induktion nach der Ableitungsordnung.

Q.E.D.

Bemerkung 7.8: Die Formel

$$\frac{d^n}{dz^n} \sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k = \sum_{k=n}^{\infty} a_k \cdot k \cdot (k-1) \cdot \dots \cdot (k-n+1) \cdot (z - z_0)^{k-n}$$

ist leicht zu merken. Sie besagt lediglich, dass man Differentiation und Summation vertauschen darf:

$$\begin{aligned} \frac{d^n}{dz^n} \sum_{k=0}^{\infty} a_k \cdot (z - z_0)^k &= \sum_{k=0}^{\infty} a_k \cdot \frac{d^n}{dz^n} (z - z_0)^k \\ &= \sum_{k=n}^{\infty} a_k \cdot k \cdot (k-1) \cdot \dots \cdot (k-n+1) \cdot (z - z_0)^{k-n}. \end{aligned}$$

Beispiel 7.9: Die in Beispiel 3.24 bzw. Definition 5.10 eingeführten Funktionen

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad \sin(z) = \sum_{k=0}^{\infty} \frac{(-1)^k \cdot z^{2 \cdot k + 1}}{(2 \cdot k + 1)!}, \quad \cos(z) = \sum_{k=0}^{\infty} \frac{(-1)^k \cdot z^{2 \cdot k}}{(2 \cdot k)!}$$

haben den Konvergenzradius $r = \infty$ und sind damit auf ganz \mathbb{C} differenzierbar. Aus diesen Darstellungen erhält man sofort:

$$\frac{d}{dz} e^z = e^z, \quad \frac{d}{dz} \sin(z) = \cos(z), \quad \frac{d}{dz} \cos(z) = -\sin(z).$$

Bemerkung 7.10: Die (reelle) Reihe

$$\frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{k=0}^{\infty} (-1)^k \cdot x^{2 \cdot k}$$

um $x_0 = 0$ hat offensichtlich den Konvergenzradius 1. Für $|x| < 1$ stellt sie nach Konstruktion die Funktion $1/(1+x^2)$ da, die längs der reellen Achse eine „harmlose“ Funktion darstellt (überall stetig, beliebig oft differenzierbar). Warum konvergiert die Reihe aber nur für Werte von x mit $|x| < 1$, wo doch $1/(1+x^2)$ für alle $x \in \mathbb{R}$ wohldefiniert ist?

Im Komplexen ist die Theorie der Differentiation und Taylor-Entwickelbarkeit wesentlich weitreichender (und einfacher) als im Reellen. Man muss sich in der Tat $1/(1+x^2)$ im Komplexen vorstellen. Die Funktion $z \in \mathbb{C} \rightarrow 1/(1+z^2) \in \mathbb{C}$ hat Singularitäten (Polstellen) bei $z = \pm i$. In der Tat ist der Konvergenzradius 1 bei Entwicklung um den Nullpunkt der Abstand vom Entwicklungspunkt 0 zur nächsten Singularität! Im Komplexen ist der „kleine“ Konvergenzradius 1 der Reihe daher unmittelbar verständlich. Im Reellen sieht man die komplexen Singularitäten nicht und wundert sich, dass die nette Funktion $1/(1+x^2)$ nicht überall durch ihre Taylor-Reihe dargestellt wird.

Kapitel 8

Banach-Iteration

Aus Zeitgründen ist dieses Kapitel nicht „offizieller“ Bestandteil der Vorlesung (sprich: ist nicht prüfungsrelevant). Es wird jedoch wärmstens empfohlen, dieses Kapitel in Heimarbeit zu lesen und zu verstehen. Speziell gibt es kein Physikerdasein ohne Kenntnis des Newton-Verfahrens (Abschnitt 8.2).

8.1 Der Banachsche Fixpunktsatz

Wir definieren den Begriff einer kontrahierenden Abbildung:

Definition 8.1:

Eine Abbildung $\Phi : A \subset \mathbb{C} \rightarrow \mathbb{C}$ heißt „**Kontraktion in A** “, wenn eine „**Kontraktionskonstante**“ $k \in [0, 1)$ existiert, so dass für alle $x, y \in A$ gilt:

$$|\Phi(x) - \Phi(y)| \leq k \cdot |x - y|.$$

Zur Namensgebung: der Abstand zweier Bildpunkte $|\Phi(x) - \Phi(y)|$ einer Kontraktion ist stets kleiner als der Abstand der Urbildpunkte $|x - y|$.

Bemerkung 8.2: Kontraktionen sind automatisch stetig: Für jede konvergierende Folge (z_n) konvergiert $\Phi(z_n)$, und es gilt

$$\lim_{n \rightarrow \infty} \Phi(z_n) = \Phi\left(\lim_{n \rightarrow \infty} z_n\right).$$

Dies ist leicht einzusehen. Sei z^* der Grenzwert von (z_n) . Die Kontraktionseigenschaft liefert

$$0 \leq |\Phi(z_n) - \Phi(z^*)| < |z_n - z^*|.$$

Wegen der Konvergenz $z_n \rightarrow z^*$ ist die rechte Seite eine reelle Nullfolge. Mit Satz 2.17 folgt, dass $|\Phi(z_n) - \Phi(z^*)|$ eine Nullfolge ist, was die Konvergenz $\Phi(z_n) \rightarrow \Phi(z^*)$ bedeutet.

Der folgende wichtige Fixpunktsatz für kontrahierende Abbildungen geht auf Stefan Banach (polnischer Mathematiker, 1892 – 1945) zurück. Er setzt neben der Kontraktionseigenschaft als wichtige Annahme voraus, dass die Kontraktion ihren Definitionsbereich in sich selbst abbildet:

Satz 8.3: („BFS“: der Banachsche Fixpunktsatz)

Sei $\Phi : A \rightarrow A$ eine Kontraktion in einer abgeschlossenen Menge $A \subset \mathbb{C}$ mit einer Kontraktionskonstanten $k < 1$. Dann

- a) existiert ein eindeutig bestimmter Fixpunkt $z^* = \Phi(z^*) \in A$,
- b) konvergiert jede Folge $z_{n+1} = \Phi(z_n)$ mit beliebigem Startwert $z_0 \in A$ gegen z^* ,
- c) gelten für jede solche Folge die Abschätzungen

$$|z_n - z^*| \leq \underbrace{\frac{k}{1-k} \cdot |z_n - z_{n-1}|}_{\text{„a posteriori“}} \leq \underbrace{\frac{k^n}{1-k} \cdot |z_1 - z_0|}_{\text{„a priori“}}.$$

Vom (wiederum sehr technischen) Beweis braucht man eigentlich nur zu wissen, dass man zeigen kann, dass die per $z_{n+1} = \Phi(z_n)$ konstruierten Folgen Cauchy-Folgen sind. Der Fixpunkt ergibt sich dann über den Grenzwert der Folgen, dessen Existenz mittels Satz 2.28 gesichert ist. Aus der Kontraktionseigenschaft folgt die Eindeutigkeit (alle Folgen konvergieren gegen denselben Grenzwert). Für technisch Interessierte ist der Beweis der Vollständigkeit halber angegeben:

Beweis: Zeige: (z_n) ist Cauchy-Folge.

$$\begin{aligned} |z_{n+m} - z_n| &= |z_{n+m} - z_{n+m-1} + z_{n+m-1} - \dots - z_{n+1} + z_{n+1} - z_n| \\ &\leq |z_{n+m} - z_{n+m-1}| + |z_{n+m-1} - z_{n+m-2}| + \dots + |z_{n+1} - z_n| \end{aligned}$$

für jedes $n, m \geq 0$. Aus $|z_j - z_{j-1}| = |\Phi(z_{j-1}) - \Phi(z_{j-2})| \leq k \cdot |z_{j-1} - z_{j-2}|$, d.h.,

$$|z_j - z_{j-1}| \leq k \cdot |z_{j-1} - z_{j-2}| \leq k^2 \cdot |z_{j-2} - z_{j-3}| \leq \dots$$

folgt

$$\begin{aligned} |z_{n+m} - z_n| &\leq k^{m-1} \cdot |z_{n+1} - z_n| + k^{m-2} \cdot |z_{n+1} - z_n| + \dots + |z_{n+1} - z_n| \\ &= (1 + k + k^2 + \dots + k^{m-1}) \cdot |z_{n+1} - z_n| = \frac{1 - k^m}{1 - k} \cdot |z_{n+1} - z_n| \\ &\leq \frac{|z_{n+1} - z_n|}{1 - k} \stackrel{(\#)}{\leq} \frac{k |z_n - z_{n-1}|}{1 - k} \leq \frac{k^2 |z_{n-1} - z_{n-2}|}{1 - k} \leq \dots \stackrel{(\#\#)}{\leq} \frac{k^n |z_1 - z_0|}{1 - k}. \end{aligned}$$

Mit $k^n \rightarrow 0$ folgt die Cauchy-Eigenschaft. Es existiert somit ein Grenzwert z^* . Mit $z_0 \in A$ und $\Phi : A \rightarrow A$ folgt $z_n \in A \Rightarrow z^*$ ist Häufungspunkt von $A \Rightarrow z^* \in A$ (abgeschlossen). Da Φ im Sinne von Bemerkung 8.2 stetig ist:

$$z^* = \lim_{n \rightarrow \infty} z_{n+1} = \lim_{i \rightarrow \infty} \Phi(z_n) = \Phi(\lim_{n \rightarrow \infty} z_n) = \Phi(z^*).$$

Eindeutigkeit: für einen weiteren Fixpunkt $z^{**} \neq z^*$ folgt der Widerspruch

$$|z^* - z^{**}| = |\Phi(z^*) - \Phi(z^{**})| \leq k \cdot |z^* - z^{**}| < |z^* - z^{**}|.$$

Die Abschätzungen c) ergeben sich aus (#) und (##):

$$\lim_{m \rightarrow \infty} |z_{n+m} - z_n| = |z^* - z_n| \stackrel{(\#)}{\leq} \frac{k}{1-k} |z_n - z_{n-1}| \stackrel{(\#\#)}{\leq} \frac{k^n}{1-k} |z_1 - z_0|.$$

Q.E.D.

Interpretation und Anwendung 8.4:

Eine Gleichung $f(z) = 0$ sei zu lösen. Der BFS gibt ein Rezept, wie man Näherungen für die Lösung konstruieren kann:

- 1) Formuliere die Gleichung $f(z) = 0$ in ein äquivalentes Fixpunktproblem $z = \Phi(z)$ um.
- 2) Ist Φ kontrahierend auf einer Umgebung A der Lösung und bildet Φ diese Menge A auf sich selbst ab, so läßt sich der BFS anwenden: wähle einen beliebigen Punkt $z_0 \in A$ und iteriere $z_{n+1} = \Phi(z_n)$. Diese Folge konvergiert gegen eine Lösung des Fixpunktproblems und damit gegen eine Lösung des Ausgangsproblems $f(z) = 0$.
- 3) Hat man durch Abschätzungen eine Kontraktionskonstante k für die Menge A gefunden, kann man mit den a-priori- bzw. a-posteriori-Abschätzungen bestimmen, wie weit man noch von der Lösung entfernt ist und abbrechen, sobald eine vorgegebene Zielgenauigkeit erreicht ist.

Mit der a-priori-Abschätzung kann man aus dem Startpunkt z_0 und dem nächsten Punkt z_1 sofort ermitteln, wie oft man höchstens iterieren muß, um die Zielgenauigkeit zu erreichen (die Iterationswerte werden dafür nicht benötigt). Nachdem die Iteration durchgeführt worden ist und Zahlenwerte für z_n vorliegen, kann man a-posteriori abschätzen, welche Approximationsgenauigkeit nun wirklich erreicht ist (die a-posteriori-Abschätzung ist prinzipiell genauer als die a-priori-Abschätzung).

8.2 Das Newton-Verfahren

Hier eine für die Praxis besonders wichtige Anwendung des BFS. Es geht darum, numerische Approximationen von Nullstellen x^* einer Funktion zu konstruieren. Die prinzipielle Idee ist, eine Zahlenfolge (x_n) zu konstruieren, die gegen einen Grenzwert x^* mit $f(x^*) = 0$ konvergiert. Für hinreichend großes n bricht man die Iteration ab und akzeptiert x_n als eine Approximation von x^* . Zunächst hat man die Aufgabe, ein Nullstellenproblem $f(x^*) = 0$ in ein Fixpunktproblem $x^* = \Phi(x^*)$ umzuformulieren, um den BFS anwenden zu können.

Bemerkung 8.5: *Es gibt viele Wege, eine gegebene Gleichung $f(x) = 0$ in eine Fixpunktgleichung $x = \Phi(x)$ umzuformen, z.B.*

$$\Phi(x) = x - g(x) \cdot f(x)$$

mit einer (praktisch beliebig wählbaren) Funktion $g(x)$. Ist $f(x)$ differenzierbar und ist die Lösung im Sinne von Definition 1.24 eine einfache Nullstelle, so ist $g(x) = 1/f'(x)$ eine ausgezeichnete Wahl. Die Iteration lautet dann

$$x_{n+1} = \Phi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)} \quad (\text{das „Newton-Verfahren“}).$$

Es gibt folgende einfache **geometrische Interpretation** des Newton-Verfahrens. Sei x_n ein Approximation einer Nullstelle von $f(x)$. Ersetze $f(x)$ in einer Umgebung von x_n durch die Tangente

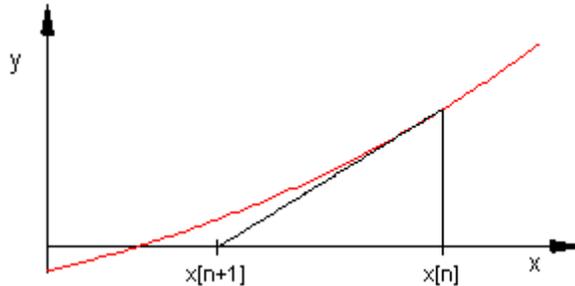
$$t(x) = f(x_n) + (x - x_n) \cdot f'(x_n).$$

Die Nullstelle der Tangente ist: $t(x) = 0 \Rightarrow x = x_n - \frac{f(x_n)}{f'(x_n)}$, also:

Der Newton-Nachfolger $x_{n+1} = x_n - f(x_n)/f'(x_n)$ von x_n ist die Nullstelle der Tangente von f am Punkt x_n .

Hier eine Beispiel-Graphik (mit MuPAD 3.0) zur Demonstration:

```
>> f:= x -> x^4 - 2: x0:= 1.5:
>> plot(plot::Function2d(f(x), x = 1.1..1.6, Color = RGB::Red),
        plot::Function2d(f(x0) + (x - x0)*f'(x0),
                          x = x0 - f(x0)/f'(x0) .. x0,
                          Color = RGB::Black),
        plot::Line2d([x0, 0], [x0, f(x0)], Color = RGB::Black),
        TicksNumber = None,
        XTicksAt = [x0 = "x[n]", x0 - f(x0)/f'(x0) = "x[n+1]"],
        Height = 4*unit::cm, Width = 8*unit::cm)
```



Hier einige numerische Werte bei der Berechnung von $\sqrt[4]{2}$ durch Lösen von $f(x) = x^4 - 2 = 0$:

```
>> f:= x -> float(x^4 - 2): x0:= 1.5:
>> Phi:= x -> x - f(x)/f'(x):
>> x0 := 1.5
                                1.5
>> x1 := Phi(x0)
                                1.273148148
>> x2 := Phi(x1)
                                1.19714982
>> x3 := Phi(x2)
                                1.189285812
>> x4 := Phi(x3)
                                1.189207123
>> x5 := Phi(x4)
                                1.189207115
```

Bemerkung 8.6: Man kann zeigen, dass es immer eine (eventuell kleine) Umgebung A einer Nullstelle von f gibt, auf der Φ eine Kontraktion ist und für die $\Phi(A) \subset A$ gilt. Damit gilt der BFS auf einer (leider oft nicht konkret bekannten) Umgebung einer Lösung, und es gilt:

Für hinreichend genaue Startwerte x_0 dicht bei einer Lösung x^* von $f(x) = 0$ konvergiert die Newton-Folge $x_{n+1} = x_n - f(x_n)/f'(x_n)$ gegen x^* .

Bei einfachen Nullstellen ist die Konvergenz beim Newton-Verfahren sehr schnell, da die Kontraktionskonstanten auf kleinen Umgebungen der Lösung prinzipiell sehr klein sind. Gilt $f'(x^*) \neq 0$ an der Nullstelle x^* , so erhält man durch Taylor-Entwicklung um x^* :

$$f(x) = \underbrace{f(x^*)}_{=0} + f'(x^*) \cdot (x - x^*) + \frac{f''(x^*)}{2} \cdot (x - x^*)^2 + \dots,$$

$$\begin{aligned}
f'(x) &= f'(x^*) + f''(x^*) \cdot (x - x^*) + \dots, \\
\Rightarrow x_{n+1} - x^* &= x_n - x^* - \frac{f'(x^*) \cdot (x_n - x^*) + \frac{f''(x^*)}{2} \cdot (x_n - x^*)^2 + \dots}{f'(x^*) + f''(x^*) \cdot (x_n - x^*) + \dots} \\
&= (x_n - x^*) \cdot \left(1 - \frac{f'(x^*) + \frac{f''(x^*)}{2} \cdot (x_n - x^*) + O((x - x^*)^2)}{f'(x^*) + f''(x^*) \cdot (x_n - x^*) + \dots}\right) \\
&= (x_n - x^*) \cdot \frac{-\frac{f''(x^*)}{2} \cdot (x_n - x^*) + O((x - x^*)^2)}{f'(x^*) + f''(x^*) \cdot (x_n - x^*) + \dots} \\
&= (x_n - x^*)^2 \cdot \frac{-\frac{f''(x^*)}{2} + O(x_n - x^*)}{f'(x^*) + O(x_n - x^*)} \\
&\approx - (x_n - x^*)^2 \cdot \frac{f''(x^*)}{2 \cdot f'(x^*)}.
\end{aligned}$$

Ist der Approximationsfehler $x_n - x^*$ erst einmal klein, so wird er durch Quadrieren sehr schnell sehr klein (man nennt dies „**quadratische Konvergenz**“).

Beispiel 8.7: In einer Softwareumgebung gebe es die Grundarithmetik, aber keine Wurzelfunktion. Um diese zu implementieren, soll die Gleichung $y^2 = b$ für gegebenes positives $b \in \mathbb{R}$ nach y gelöst werden. Mittels Division durch eine geeignete 4er-Potenz kann b auf das Intervall $[1, 4]$ transformiert werden (in Binärdarstellung kostet dies nichts). Sei nun $a = b/4^n \in [1, 4]$. Ist $x \in [1, 2]$ eine Lösung von $x^2 = a$, so ist $y = 2^n \cdot x$ die gesuchte Lösung des Ausgangsproblems $y^2 = b$.

Das verbleibende Problem ist also, ein $x \in [1, 2]$ zu finden, das das Nullstellenproblem $f(x) = x^2 - a = 0$ mit $a \in [1, 4]$ erfüllt. Hierzu soll das Newton-Verfahren benutzt werden. Betrachte also

$$\Phi(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^2 - a}{2 \cdot x} = \frac{x^2 + a}{2 \cdot x} = \frac{1}{2} \cdot \left(x + \frac{a}{x}\right).$$

Die entsprechende Iteration lautet also

$$x_{n+1} = \frac{1}{2} \cdot \left(x_n + \frac{a}{x_n}\right).$$

Das Verfahren konvergiert sehr schnell gegen $x^* = \sqrt{a}$. Beispiel: $a = 2$, $x_0 = 1.5$:

$$x_0 = 1.5, \quad x_1 = 1.416666666\dots, \quad x_2 = 1.414215686\dots,$$

$$x_3 = 1.414213562\dots, \quad x_4 = 1.414213562\dots, \quad \dots$$

Die folgende Analyse ist wiederum sehr technisch und an technisch Interessierte adressiert:

Analyse: betrachte das Intervall $A = [\sqrt{a}, \frac{1+a}{2}]$, das die Lösung \sqrt{a} enthält. (Dieses Intervall fällt hier vom Himmel.) Die folgenden Rechnungen zeigen, dass dieses Intervall

in der Tat so ist, dass der BFS angewendet werden kann.

Es ist zunächst zu zeigen, dass $\Phi(A) \subset A$ gilt. In Erinnerung an die Schule berechnen wir dazu

$$\Phi'(x) = \frac{1}{2} \cdot \left(1 - \frac{a}{x^2}\right) = \frac{x^2 - a}{2 \cdot x^2} \geq 0$$

für $x \in [\sqrt{a}, \frac{1+a}{2}]$. Die Funktion ist also in diesem Bereich monoton steigend, und damit gilt

$$\Phi(\sqrt{a}) = \sqrt{a} \leq \Phi(x) \leq \Phi\left(\frac{1+a}{2}\right) = \frac{1+a}{4} + \frac{a}{1+a} \leq \frac{1+a}{2}$$

für alle $x \in [\sqrt{a}, \frac{1+a}{2}]$. Als Kontraktionskonstante auf A schätzt man ab:

$$\begin{aligned} |\Phi(x) - \Phi(y)| &= \frac{1}{2} \cdot \left|x + \frac{a}{x} - y - \frac{a}{y}\right| = \frac{1}{2} \cdot \left|x - y - \frac{a \cdot (x - y)}{x \cdot y}\right| \\ &= \frac{|x - y|}{2} \cdot \left|1 - \frac{a}{x \cdot y}\right| = \frac{|x - y|}{2} \cdot \left(1 - \frac{a}{x \cdot y}\right) \leq \frac{|x - y|}{2} \cdot \left(1 - \frac{a}{\left(\frac{a+1}{2}\right)^2}\right) \\ &= |x - y| \cdot \frac{1}{2} \cdot \left(\frac{a-1}{a+1}\right)^2. \end{aligned}$$

Also, in Abhängigkeit von a ist

$$k = \frac{1}{2} \cdot \left(\frac{a-1}{a+1}\right)^2$$

eine Kontraktionskonstante für Φ über A . Für alle $a \in [1, 4]$ gilt $k \leq 0.18$, d.h., Φ ist auf A in der Tat eine Kontraktion.

Speziell, für $a = 2$ ist $k = 1/18 \approx 0.0555 \dots$. Starten wir mit $x_0 = 3/2$, so ergibt sich $x_1 = 17/12$ und die a-priori-Abschätzung liefert

$$|x_n - \sqrt{2}| \leq \frac{k^n}{1-k} \cdot |x_0 - x_1| = \frac{3}{34 \cdot 18^n}.$$

Nach $n = 5$ Schritten ergibt sich beispielsweise

$$|x_5 - \sqrt{2}| \leq 4.7 \cdot 10^{-8},$$

d.h., x_5 beschreibt garantiert die ersten 7 bis 8 Dezimalstellen von $\sqrt{2}$ korrekt. (In Wirklichkeit ist x_5 schon wesentlich genauer, aber mehr gibt die a-priori-Abschätzung nicht her.)

Bemerkung 8.8: Das letzte Beispiel hat gezeigt, dass das Abschätzen von Kontraktionskonstanten mühselig ist. Es geht aber auch einfacher! Für stetig differenzierbares Φ gilt, dass

$$k = \sup \{|\Phi'(x)|; x \in A\}$$

die bestmögliche (weil kleinste) Kontraktionskonstante über einem Intervall $A \subset \mathbb{R}$ ist. Um dies einzusehen, brauchen wir nur den Mittelwertsatz der Differentialrechnung 6.20 zu betrachten. Es gilt

$$\frac{\Phi(x) - \Phi(y)}{x - y} = \Phi'(\xi)$$

mit ξ im Intervall zwischen x und y . Es folgt

$$|\Phi(x) - \Phi(y)| = |\Phi'(\xi)| \cdot |x - y| \leq k \cdot |x - y|,$$

wobei k das obige Supremum ist.

Bemerkung 8.9: Im Fall skalarer Gleichungen $f(x) = 0$ hat das Newton-Verfahren angenehme lokale Eigenschaften: Der Erfolg ist garantiert, wenn nur der Startpunkt x_0 hinreichend gut ist. Dies gilt sogar bei p -fach entarteten Nullstellen (bei denen für $p > 1$ die quadratische Konvergenz dann allerdings zu linearer Konvergenz mit der asymptotischen Kontraktionskonstanten $k = 1 - 1/p$ wird).

Die globalen Eigenschaften des Newton-Verfahrens sind deutlich weniger glorreich. Gerät die Newton-Folge in die Nähe eines lokalen Extremums, wird die Iteration $x \rightarrow x - f(x)/f'(x)$ in der Nähe einer Nullstelle von f' ausgewertet, und die Iterationspunkte können weit weg landen.

Kapitel 9

Integration

↓9.12.03

9.1 Stammfunktionen: das unbestimmte Integral

Die Integration ist die Umkehrung der Differentiation: zu einer gegebenen Funktion $f(x)$ sucht man eine Funktion $F(x)$, deren Ableitung $f(x)$ ist.

9.1.1 Definitionen, Grundintegrale

Definition 9.1: (Stammfunktion)

$F(x)$ heißt „**Stammfunktion**“ einer (hinreichend glatten) Funktion $f(x)$, wenn $\frac{d}{dx}F(x) = f(x)$ gilt. Alternativ nennt man $F(x)$ auch das „**unbestimmte Integral über $f(x)$** “ und benutzt auch die Notation $F(x) = \int f(x) dx$. Die Funktion $f(x)$ unter dem Integralzeichen wird als „**Integrand**“ bezeichnet.

Bemerkung 9.2: Stammfunktionen sind nicht eindeutig bestimmt. Da die Ableitung einer konstanten Funktion überall 0 ist, kann man zu einer Stammfunktion eine beliebige Konstante hinzuaddieren, wobei man eine neue Stammfunktion erhält. Andererseits, hat $f(x)$ keine Singularitäten (Polstellen etc.), so sind Stammfunktionen stetig und die Differenz zweier stetiger Stammfunktionen ist immer eine Konstante.

Beispiel 9.3: Zu $f(x) = x$ sind $F_1(x) = \frac{x^2}{2}$ und $F_2(x) = \frac{x^2}{2} + 17$ Stammfunktionen. Die beliebige additive Konstante in Stammfunktionen (die „**Integrationskonstante**“) wird folgendermaßen ausgedrückt:

$$\int x dx = \frac{x^2}{2} + c.$$

Damit ist gemeint: $\int f(x) dx$ stellt die Klasse **aller** Stammfunktionen dar (d.h., in der Schreibweise $\int f(x) dx$ steckt die additive Konstante sozusagen im \int -Symbol und

braucht nicht explizit hingeschrieben zu werden). Sobald das Integralzeichen durch einen konkreten Repräsentanten dieser Klasse (hier $\frac{x^2}{2}$) ersetzt wird, schreiben wir die beliebige additive Konstante explizit dazu.

Bemerkung 9.4: Mit dieser Konvention gilt trivialerweise für jede differenzierbare Funktion $F(x)$:

$$\int F'(x) dx = F(x) + c .$$

Grundintegrale 9.5:

Aus der in Satz 6.6 gegebenen (kleinen) Liste von Ableitungen erhält man eine (kleine) Liste von Stammfunktionen für die einfachen Grundfunktionen:

$$\int x^n dx = \frac{x^{n+1}}{n+1} + c, \quad (n \neq 0)$$

$$\int \frac{1}{x} dx = \ln(|x|) + c, \quad (\text{Beispiel 6.18})$$

$$\int e^x dx = e^x + c,$$

$$\int \sin(x) dx = -\cos(x) + c,$$

$$\int \cos(x) dx = \sin(x) + c.$$

Beispiel 9.6: In MuPAD ist die Funktion `int` (engl.: integrate) für die Integration zuständig. Für die Integrationskonstante wird dabei vom System automatisch ein „besonders einfacher“ Wert gewählt:

```
>> int(cos(x), x)
```

```
sin(x)
```

```
>> int(x*sin(x)*exp(x), x)
```

$$\frac{\cos(x) \exp(x)}{2} - \frac{x \cos(x) \exp(x)}{2} + \frac{x \sin(x) \exp(x)}{2}$$

Für aus den einfachen Grundfunktionen aufgebaute Funktionen würde man gern per Rechenregeln die Integration komplizierter Funktionen auf die Integration einfacher Funktionen zurückführen. Leider ist das nicht so einfach. In der Tat entspricht jeder Rechenregel der Differentiation (Satz 6.6, Satz 6.12) eine Regel für's Integrieren. Die sich ergebenden Regeln sind aber nicht so, dass man damit automatisch alle Integrationen auf Grundintegrale zurückführen kann. Zunächst die einfachsten Regeln:

Satz 9.7: (Summenregel)

Für beliebige Konstanten a, b und Funktionen $f(x), g(x)$ gilt

$$\int (a \cdot f(x) + b \cdot g(x)) dx = a \cdot \int f(x) dx + b \cdot \int g(x) dx.$$

Das ist durch Differenzieren beider Seiten dieser Gleichung unmittelbar klar.

Merke:

Konstante Faktoren können stets aus dem Integralzeichen herausgezogen werden. Das Integral einer Summe ist die Summe der Integrale.

Beispiel 9.8:

$$\begin{aligned} \int \left(2 \cdot e^x + \frac{1}{\sqrt{2}x} \right) dx &= 2 \cdot \int e^x dx + \int \frac{1}{\sqrt{2}} x^{-\frac{1}{2}} dx = 2 \cdot e^x + c_1 + \frac{1}{\sqrt{2}} \int x^{-\frac{1}{2}} dx \\ &= 2 \cdot e^x + c_1 + \frac{1}{\sqrt{2}} \cdot \frac{x^{-\frac{1}{2}+1}}{-\frac{1}{2}+1} + c_2 = 2 \cdot e^x + \frac{1}{\sqrt{2}} \cdot \frac{x^{\frac{1}{2}}}{\frac{1}{2}} + \underbrace{c_1 + c_2}_c \\ &= 2 \cdot e^x + \frac{2}{\sqrt{2}} \cdot \sqrt{x} + c = 2 \cdot e^x + \sqrt{2} \cdot \sqrt{x} + c. \end{aligned}$$

Hierbei wurden die einzelnen Integrationskonstanten c_1, c_2 zu einer neuen beliebigen Konstanten $c = c_1 + c_2$ zusammengefasst.

9.1.2 Partielle Integration

Aus der Produktregel

$$\frac{d}{dx} (f(x) \cdot g(x)) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

der Differentiation gewinnt man durch Integration

$$f(x) \cdot g(x) + c = \int f'(x) \cdot g(x) dx + \int f(x) \cdot g'(x) dx.$$

Diese Gleichung liefert eine Integrationsregel, die man „**partielle Integration**“ nennt:

Satz 9.9: (Partielle Integration)

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

Bemerkung 9.10: Diese Regel ist in folgender Situation anwendbar:

- Der Integrand muss das Produkt zweier Funktionen sein.
- Von einem Faktor ($g'(x)$) muss man die Stammfunktion $g(x)$ kennen.

Ein Integral (über $f(x) \cdot g'(x)$) wird in ein anderes Integral (über $f'(x) \cdot g(x)$) überführt, es verbleibt also die Aufgabe, eine Stammfunktion zu finden. Allerdings ist manchmal das Produkt $f'(x) \cdot g(x)$ einfacher zu integrieren als das Ausgangsprodukt $f(x) \cdot g'(x)$:

- Sinnvoll ist partielle Integration meist, wenn die Ableitung $f'(x)$ „einfacher“ ist als $f(x)$ und $g(x)$ nicht wesentlich „komplizierter“ als $g'(x)$.

Beispiel 9.11: Im Integral $\int x \cdot \ln(x) dx$ ist $f(x) = \ln(x)$ eine „unangenehme“ Funktion, während $f'(x) = \frac{1}{x}$ als rationale Funktion wesentlich angenehmer ist:

$$\begin{aligned} \int \underbrace{x}_{g'(x)} \cdot \underbrace{\ln(x)}_{f(x)} dx &= \underbrace{\ln(x)}_{f(x)} \cdot \underbrace{\frac{x^2}{2}}_{g(x)} - \int \underbrace{\frac{1}{x}}_{f'(x)} \cdot \underbrace{\frac{x^2}{2}}_{g(x)} dx \\ &= \ln(x) \cdot \frac{x^2}{2} - \int \frac{x}{2} dx = \ln(x) \cdot \frac{x^2}{2} - \frac{x^2}{4} + c. \end{aligned}$$

Probe:

$$\frac{d}{dx} \left(\ln(x) \cdot \frac{x^2}{2} - \frac{x^2}{4} + c \right) = \frac{1}{x} \cdot \frac{x^2}{2} + \ln(x) \cdot x - \frac{x}{2} = \ln(x) \cdot x.$$

Es gibt keine allgemeine Regel, was „einfach“ und was „kompliziert“ ist. Im obigen Fall war $f'(x) = \frac{1}{x}$ einfacher als $f(x) = \ln(x)$. Im folgenden Beispiel ist $f(x) = x$ „kompliziert“, zumindestens „komplizierter“ als $f'(x) = 1$:

Beispiel 9.12:

$$\begin{aligned} \int \underbrace{x}_{f(x)} \cdot \underbrace{e^x}_{g'(x)} dx &= \underbrace{x}_{f(x)} \cdot \underbrace{e^x}_{g(x)} - \int \underbrace{1}_{f'(x)} \cdot \underbrace{e^x}_{g(x)} dx \\ &= x \cdot e^x - \int e^x dx = x \cdot e^x - e^x + c = (x - 1) \cdot e^x + c. \end{aligned}$$

Manchmal braucht man einfach Erfahrung um zu sehen, dass partielle Integration hilfreich ist:

Beispiel 9.13:

$$\begin{aligned} \int \sin(x)^2 dx &= \int \underbrace{\sin(x)}_{f(x)} \cdot \underbrace{\sin(x)}_{g'(x)} dx = \underbrace{\sin(x)}_{f(x)} \cdot \underbrace{(-\cos(x))}_{g(x)} - \int \underbrace{\cos(x)}_{f'(x)} \cdot \underbrace{(-\cos(x))}_{g(x)} dx \\ &= -\sin(x) \cdot \cos(x) + \int \cos(x)^2 dx. \end{aligned}$$

Das war bislang nicht sehr erfolgreich: $\int \sin(x)^2 dx$ wurde durch $\int \cos(x)^2 dx$ ausgedrückt. Allerdings gilt $\sin(x)^2 + \cos(x)^2 = 1$, sodass das verbleibende Integral wiederum durch das Ausgangsintegral ausgedrückt werden kann:

$$\int \cos(x)^2 dx = \int 1 dx - \int \sin(x)^2 dx = x - \int \sin(x)^2 dx.$$

Dies liefert eine Gleichung für $\int \sin(x)^2 dx$:

$$\begin{aligned} \int \sin(x)^2 dx &= -\sin(x) \cdot \cos(x) + \int \cos(x)^2 dx \\ &= -\sin(x) \cdot \cos(x) + x - \int \sin(x)^2 dx \\ \Rightarrow 2 \cdot \int \sin(x)^2 dx &= x - \sin(x) \cdot \cos(x) + c \\ \Rightarrow \int \sin(x)^2 dx &= \frac{1}{2} \cdot (x - \sin(x) \cdot \cos(x)) + \tilde{c} \end{aligned}$$

(mit einer neuen Integrationskonstante $\tilde{c} = c/2$).

9.1.3 Substitution

Aus der Kettenregel der Differentiation (mit $y = g(x)$)

$$\frac{d}{dx} F(g(x)) = \left(\frac{d}{dy} F(y) \right) \cdot \left(\frac{d}{dx} g(x) \right) = F'(g(x)) \cdot g'(x)$$

gewinnt man durch Integration

$$F(g(x)) + c = \int F'(g(x)) \cdot g'(x) dx.$$

Diese Gleichung liefert mit $f = F'$ eine Integrationsregel, die man „**Integration durch Substitution**“ nennt:

Satz 9.14: (Substitution)

Sei $F(y)$ eine Stammfunktion von $f(y)$. Mit $y = g(x)$ gilt

$$\int f(g(x)) \cdot \underbrace{g'(x)}_{dy} dx = \int f(y) dy = F(y) + c = F(g(x)) + c.$$

Hierbei läuft die Substitution auf Folgendes hinaus. Aus $y = g(x)$ folgt $\frac{dy}{dx} = g'(x)$, also formal

$$dy = g'(x) dx.$$

Eine Substitution bietet sich auf jeden Fall an, wenn der Integrand einen Faktor $g'(x)$ enthält, der die Ableitung eines Teilausdrucks $g(x)$ im anderen Faktor ist:

Beispiel 9.15: In $\int \cos(x) \cdot e^{\sin(x)} dx$ bietet es sich an, $y = g(x) = \sin(x)$ zu substituieren, denn die Ableitung $g'(x) = \cos(x)$ taucht als Faktor im Integranden auf. Es ergibt sich

$$\int \cos(x) \cdot e^{\sin(x)} dx = \int e^{\overbrace{\sin(x)}^{y=g(x)}} \underbrace{\cos(x) dx}_{g'(x) \cdot dx=dy} = \int e^y dy = e^y + c = e^{\sin(x)} + c.$$

Beispiel 9.16: Wir kennen $\int \frac{1}{y} dy = \ln(|y|)$. Wie steht es mit $\int \frac{1}{a \cdot x + b} dx$? Dies ist ein Fall für die Substitution. Wir setzen $y = g(x) = a \cdot x + b$ (also $dy = a dx$) und erweitern mit a , sodass $dx = \frac{1}{a} \cdot a dx = \frac{1}{a} dy$ auftaucht:

$$\begin{aligned} \int \frac{1}{a \cdot x + b} dx &= \frac{1}{a} \cdot \int \frac{1}{a \cdot x + b} \cdot \overbrace{\frac{dy}{a}}^{dy} dx = \frac{1}{a} \int \frac{1}{y} dy \\ &= \frac{1}{a} \cdot \ln(|y|) + c = \frac{1}{a} \cdot \ln(|a \cdot x + b|) + c. \end{aligned}$$

Beispiel 9.17: In $\int \frac{g'(x)}{g(x)} dx$ bietet sich die Substitution $y = g(x)$ an:

$$\boxed{\int \frac{g'(x)}{g(x)} dx = \int \frac{1}{y} dy = \ln(|y|) + c = \ln(|g(x)|) + c.}$$

Bemerkung 9.18: Es bietet sich allgemein an, eine Substitution $y = g(x)$ in einem Integral $\int h(x) dx$ technisch folgendermaßen durchzuführen:

- Setze $y = g(x)$ und berechne die Ableitung $\frac{dy}{dx} = g'(x)$. Formal gilt $dy = g'(x) dx$.
- Ersetze dx durch $\frac{dy}{g'(x)}$. Drücke im neuen Integranden $h(x) dx = \frac{h(x)}{g'(x)} dy$ jedes x durch y aus.

- Es entsteht ein Ausdruck

$$\int h(x) dx = \int \underbrace{h(x(y)) \cdot \frac{1}{g'(x(y))}}_{f(y)} dy = \int f(y) dy.$$

Versuche, eine Stammfunktion $F(y) = \int f(y) dy$ zu finden.

- **„Rücksubstitution“**: Setze $y = g(x)$ in $F(y)$ ein. Die gesuchte Stammfunktion des ursprünglichen Ausdrucks ist $F(g(x))$.

Manchmal ist es nicht offensichtlich, was man substituieren sollte. Hier hilft nur Erfahrung oder ein guter Hinweis:

Beispiel 9.19: Substituiere $y = \sqrt{x}$, $\frac{dy}{dx} = \frac{1}{2} \cdot \frac{1}{\sqrt{x}}$ ($\Rightarrow dy = \frac{1}{2} \cdot \frac{1}{\sqrt{x}} dx$) in

$$\int \sqrt{x} \cdot e^{\sqrt{x}} dx = \int y \cdot e^y \cdot \underbrace{2 \cdot \sqrt{x} dy}_{dx} = 2 \cdot \int y^2 \cdot e^y dy.$$

Das verbleibende Integral in y kann durch zweifache partielle Integration gelöst werden:

$$\begin{aligned} 2 \cdot \int \underbrace{y^2}_{f(y)} \cdot \underbrace{e^y}_{g'(y)} dy &= 2 \cdot \underbrace{y^2}_{f(y)} \cdot \underbrace{e^y}_{g(y)} - 2 \cdot \int \underbrace{2 \cdot y}_{f'(y)} \cdot \underbrace{e^y}_{g(y)} dy \\ &= 2 \cdot y^2 \cdot e^y - 4 \cdot \int \underbrace{y}_{F(y)} \cdot \underbrace{e^y}_{G'(y)} dy = 2 \cdot y^2 \cdot e^y - 4 \cdot \underbrace{y}_{F(y)} \cdot \underbrace{e^y}_{G(y)} + 4 \cdot \int \underbrace{1}_{F'(y)} \cdot \underbrace{e^y}_{G(y)} dy \\ &= 2 \cdot y^2 \cdot e^y - 4 \cdot y \cdot e^y + 4 \cdot e^y + c. \end{aligned}$$

Rücksubstitution $y = \sqrt{x}$ liefert letztlich:

$$\int \sqrt{x} \cdot e^{\sqrt{x}} dx = 2 \cdot x \cdot e^{\sqrt{x}} - 4 \cdot \sqrt{x} \cdot e^{\sqrt{x}} + 4 \cdot e^{\sqrt{x}} + c.$$

9.1.4 Rationale Integranden: Partialbruchzerlegung

Rationale Integranden lassen sich über die Technik der „Partialbruchzerlegung“ immer so umformulieren, dass man eine Stammfunktion bestimmen kann. Hier der Spezialfall, wenn das Nennerpolynom nur einfache Nullstellen hat:

Satz 9.20: (Partialbruchzerlegung)

Betrachte $f(x) = \frac{p(x)}{q(x)}$ mit Polynomen $p(x)$ und $q(x)$, wobei $\text{grad}(p(x)) < \text{grad}(q(x))$ gelte. Hat das Nennerpolynom $q(x)$ nur einfache Nullstellen x_1, \dots, x_n , so gibt es Konstanten c_1, \dots, c_n , sodass

$$\frac{p(x)}{q(x)} = \frac{c_1}{x - x_1} + \dots + \frac{c_n}{x - x_n}.$$

Damit folgt dann

$$\int \frac{p(x)}{q(x)} dx = c_1 \cdot \ln(|x - x_1|) + \dots + c_n \cdot \ln(|x - x_n|) + c.$$

Beispiel 9.21: Die technische Durchführung geschieht folgendermaßen:

1) Ansatz:

$$\frac{3 \cdot x + 4}{(x - 1) \cdot (x + 2)} = \frac{c_1}{x - 1} + \frac{c_2}{x + 2}.$$

2) Bringe die rechte Seite auf den Hauptnenner:

$$\frac{c_1}{x - 1} + \frac{c_2}{x + 2} = \frac{c_1 \cdot (x + 2) + c_2 \cdot (x - 1)}{(x - 1) \cdot (x + 2)}$$

3) Ordne den Zähler nach Potenzen von x :

$$\frac{c_1 \cdot (x + 2) + c_2 \cdot (x - 1)}{(x - 1) \cdot (x + 2)} = \frac{(c_1 + c_2) \cdot x + (2 \cdot c_1 - c_2)}{(x - 1) \cdot (x + 2)}.$$

4) Der Ansatz lautet nun:

$$\frac{3 \cdot x + 4}{(x - 1) \cdot (x + 2)} = \frac{(c_1 + c_2) \cdot x + (2 \cdot c_1 - c_2)}{(x - 1) \cdot (x + 2)}.$$

Die Nenner stimmen nach Konstruktion überein. Es verbleibt, die Konstanten c_1, c_2 so zu bestimmen, dass auch die Zähler für alle x übereinstimmen. Vergleiche dazu im Zähler die Koeffizienten vor jeder x -Potenz:

$$3 = c_1 + c_2, \quad 4 = 2 \cdot c_1 - c_2.$$

4) Löse das entstandene lineare Gleichungssystem für die unbekanntenen Koeffizienten:

$$c_1 = \frac{7}{3}, \quad c_2 = \frac{2}{3}.$$

Ergebnis:

$$\int \frac{3 \cdot x + 4}{(x - 1) \cdot (x + 2)} dx = \int \left(\frac{\frac{7}{3}}{x - 1} + \frac{\frac{2}{3}}{x + 2} \right) dx = \frac{7}{3} \ln(|x - 1|) + \frac{2}{3} \ln(|x + 2|) + c.$$

Beispiel 9.22: In MuPAD ist die Funktion `partfrac` (engl.: partial fraction) für die Partialbruchzerlegung zuständig:

```
>> partfrac((3*x + 4) / ((x - 1)*(x + 2)), x)
```

$$\frac{7}{3(x-1)} + \frac{2}{3(x+2)}$$

Bemerkung 9.23: Die Partialbruchzerlegung haben wir schon früher beim Summieren rationaler Ausdrücke kennengelernt: siehe Beispiel 3.31.

Bemerkung 9.24: Hat man einen rationalen Integranden $\frac{p(x)}{q(x)}$, bei dem der Grad des Zählerpolynoms nicht kleiner ist als der Grad des Nennerpolynoms (dies wird in Satz 9.20 vorausgesetzt), so ist dies auch kein Problem. Durch Polynomdivision kann man einen polynomialen Anteil abspalten, z.B.:

$$\frac{2 \cdot x^3 + x^2 + 2}{x^2 - 1} = 2 \cdot x + 1 + \frac{2 \cdot x + 3}{x^2 - 1}.$$

Die Division wird dabei wie mit Zahlen durchgeführt (man zieht sukzessiv den „führenden Term“ durch ein geeignetes Vielfaches des Nenners ab):

$$\begin{array}{r} 2 \cdot x^3 + x^2 + 2 \quad : \quad x^2 - 1 \quad = \quad 2 \cdot x + 1 \\ \underline{2 \cdot x^3 - 2 \cdot x} \\ x^2 + 2 \cdot x + 2 \\ \underline{x^2 - 1} \\ 2 \cdot x + 3 \quad (\text{der Rest}) \end{array}$$

Der verbleibende Rest kann durch Partialbruchzerlegung additiv zerlegt werden, das Ergebnis ist:

```
>> partfrac((2*x^3 + x^2 + 2)/(x^2 - 1), x)
```

$$2x + \frac{5}{2(x-1)} - \frac{1}{2(x+1)} + 1$$

Es folgt

$$\begin{aligned} \int \frac{2 \cdot x^3 + x^2 + 2}{(x^2 - 1)} dx &= \int \left(2 \cdot x + 1 + \frac{\frac{5}{2}}{x-1} - \frac{\frac{1}{2}}{x+1} \right) dx \\ &= x^2 + x + \frac{5}{2} \ln(|x-1|) - \frac{1}{2} \ln(|x+1|) + c. \end{aligned}$$

Probe mit MuPAD:

>> int((2*x^3 + x^2 + 2)/(x^2 - 1), x)

$$x + x^2 + \frac{5 \ln(x - 1)}{2} - \frac{\ln(x + 1)}{2}$$

(MuPAD verzichtet darauf, innerhalb des \ln Betragszeichen einzutragen, denn MuPAD kann mit komplexen Zahlen umgehen. Für positives x gilt $\ln(-x) = \sqrt{-1} \cdot \pi + \ln(x)$, d.h., $\ln(-x)$ und $\ln(x)$ stimmen bis auf eine additive (komplexe) Konstante überein. Diese kann in die Integrationskonstante absorbiert werden).

Bemerkung 9.25: Für die Partialbruchzerlegung braucht man die Faktorisierung $q(x) = (x - x_1) \cdot \dots \cdot (x - x_n)$ des Nennerpolynoms, d.h., man muss die Nullstellen x_1, \dots, x_n von $q(x)$ finden.

9.2 Das bestimmte Integral

10.12.03↓

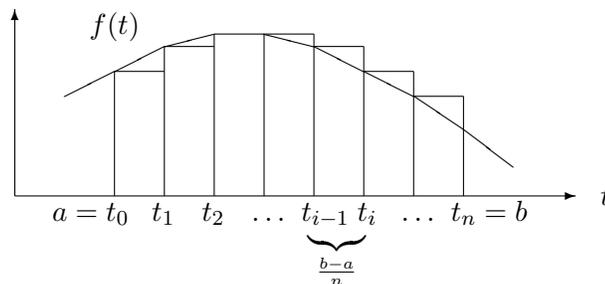
Die grundsätzliche Idee des Integrals ist geometrischer Natur:

Berechne die Fläche unter einem Funktionsgraphen $f(t)$.

Betrachte für die folgende Motivation positive Funktionen $f(t)$. Zerlege ein Intervall $[a, b]$ auf der t -Achse in n Teilintervalle $[t_i, t_{i+1}]$. Dann approximiere den Flächeninhalt durch die Flächen der durch die Punkte

$$(t_{i-1}, 0), \quad (t_{i-1}, f(t_{i-1})), \quad (t_i, f(t_{i-1})), \quad (t_i, 0)$$

gegebenen Rechtecke (mit der Breite $\frac{b-a}{n}$):



Die Summe der n Rechteckflächen ist $\frac{b-a}{n} \cdot \sum_{i=0}^{n-1} f(t_i)$. Im Grenzwert $n \rightarrow \infty$ liefert dies die Fläche unter dem Graphen.

Um zu einem mathematisch sauberen Begriff zu kommen, sollte man die Approximation der Fläche durch Summen von Rechteckflächen nicht nur auf äquidistante Zerlegungen des Integrationsintervalls einschränken. Stattdessen betrachtet man beliebige Zerlegungen in Teilintervalle, deren Feinheit letztlich gegen 0 streben soll.

Definition 9.26: (Partition eines Intervalls)

Eine **Partition** $P = \{t_0, t_1, \dots, t_n\}$ eines endlichen Intervalls $[a, b] \subset \mathbb{R}$ ist eine monoton angeordnete Menge von Zwischenpunkten

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b.$$

Die **Feinheit** der Partition ist $\delta(P) = \max_{i=1..n} (t_i - t_{i-1})$.

Statt der in der obigen Motivation betrachteten Summation über Auswertungen der Funktion f an den Zwischenpunkten t_i betrachtet man Abschätzungen nach oben und unten:

Definition 9.27: (Ober- und Untersummen)

Sei $f : [a, b] \rightarrow \mathbb{R}$ eine beschränkte Funktion, sei $P = \{t_0, t_1, \dots, t_n\}$ eine Partition des Intervalls $[a, b]$. Definiere

$$U_{f,P} = \sum_{i=1}^n (t_i - t_{i-1}) \cdot \inf_{t_{i-1} \leq t \leq t_i} f(t) \quad (\text{„Untersumme“}),$$

$$O_{f,P} = \sum_{i=1}^n (t_i - t_{i-1}) \cdot \sup_{t_{i-1} \leq t \leq t_i} f(t) \quad (\text{„Obersumme“}).$$

Bemerkung 9.28:

a) Für jede Partition P eines gegebenen Intervalls gilt offensichtlich

$$U_{f,P} \leq O_{f,P}.$$

b) Eine Partition $\tilde{P} = \{\tilde{t}_0, \dots, \tilde{t}_n\}$ heißt **Verfeinerung** von $P = \{t_0, \dots, t_n\}$, wenn $P \subset \tilde{P}$ gilt (d.h., \tilde{P} besteht aus den Gitterpunkten von P plus weiteren Gitterpunkten).

c) Für eine Verfeinerung \tilde{P} von P gilt offensichtlich

$$U_{f,P} \leq U_{f,\tilde{P}} \quad \text{und} \quad O_{f,\tilde{P}} \leq O_{f,P},$$

denn z.B. gilt für die Summanden der Obersummen:

$$\begin{aligned} & (t_i - \tilde{t}_j) \cdot \sup_{\tilde{t}_j \leq t \leq t_i} f(t) \quad + \quad (\tilde{t}_j - t_{i-1}) \cdot \sup_{t_{i-1} \leq t \leq \tilde{t}_j} f(t) \\ & \leq (t_i - \tilde{t}_j) \cdot \sup_{t_{i-1} \leq t \leq t_i} f(t) \quad + \quad (\tilde{t}_j - t_{i-1}) \cdot \sup_{t_{i-1} \leq t \leq \tilde{t}_j} f(t) \\ & = (t_i - t_{i-1}) \cdot \sup_{t_{i-1} \leq t \leq t_i} f(t), \end{aligned}$$

wo t_{i-1}, t_i gemeinsame Gitterpunkte von P und \tilde{P} sind und \tilde{t}_j ein zusätzlicher Gitterpunkt von \tilde{P} zwischen t_{i-1} und t_i ist.

d) Insgesamt folgt hieraus: für jedes Pärchen P_1, P_2 von Partitionen gilt

$$U_{f,P_1} \leq O_{f,P_2},$$

denn mit der Verfeinerung $P_{12} = P_1 \cup P_2$ gilt

$$U_{f,P_1} \leq U_{f,P_{12}} \leq O_{f,P_{12}} \leq O_{f,P_2}.$$

Die Untersummen sind damit durch die Obersummen beliebiger (anderer) Partitionen nach oben beschränkt, die Obersummen sind durch die Untersummen beliebiger (anderer) Partitionen nach unten beschränkt. Damit existieren

$$\sup \left\{ U_{f,P}; P \text{ ist Partition von } [a, b] \right\} \leq \inf \left\{ O_{f,P}; P \text{ ist Partition von } [a, b] \right\}.$$

Definition 9.29: (Das bestimmte Integral)

Eine beschränkte Funktion $f : [a, b] \rightarrow \mathbb{R}$ heißt **Riemann-integrierbar**, wenn es zu jedem $\epsilon > 0$ eine Partition P des Intervalls $[a, b]$ gibt mit

$$O_{f,P} - U_{f,P} \leq \epsilon.$$

In diesem Fall gilt

$$\sup \left\{ U_{f,P}; P \text{ ist Partition von } [a, b] \right\} = \inf \left\{ O_{f,P}; P \text{ ist Partition von } [a, b] \right\}.$$

Dieser Wert wird dann das „**bestimmte (Riemann-)Integral**“ von f genannt:

$$\begin{aligned} \int_a^b f(t) dt &:= \sup \{ U_{f,P}; P \text{ ist Partition von } [a, b] \} \\ &= \inf \{ O_{f,P}; P \text{ ist Partition von } [a, b] \}. \end{aligned}$$

Die Funktion $f(t)$ heißt „**Integrand**“.

Diese Definition als (abstraktes) Supremum von Untersummen/Infimum von Obersummen über *alle* Partitionen ist wenig konstruktiv. Der folgende Satz besagt, dass dieses Supremum/Infimum speziell von den Partitionen mit Feinheiten erzeugt werden, die gegen 0 konvergieren:

Satz 9.30: (Das Integral über Partitionen mit gegen 0 konvergierender Feinheit)

Sei $f : [a, b] \rightarrow \mathbb{R}$ beschränkt und Riemann-integrierbar. Dann gilt für jede Folge P_n von Partitionen, deren Feinheit $\delta(P_n)$ gegen Null konvergiert:

$$\int_a^b f(t) dt = \lim_{n \rightarrow \infty} O_{f,P_n} = \lim_{n \rightarrow \infty} U_{f,P_n}.$$

Beweis: (für technisch Interessierte)

Vorüberlegung: Zu zwei Partitionen P_1 und P_2 betrachten wir die Verfeinerung $P_1 \cup P_2$. Sei $|P_2|$ die Anzahl der Zerlegungspunkte von P_2 . Seien I_k die von der Zerlegung P_1 erzeugten Teilintervalle von $[a, b]$, in deren Innerem Punkte aus P_2 liegen (diese sind die Teilintervalle von P_1 , die durch die Verfeinerung mittels P_2 weiter zerlegt werden). Jedes dieser Intervalle I_k hat maximal die Länge $\delta(P_1)$, wo $\delta(P_1)$ die Feinheit von P_1 ist. Insgesamt kann es maximal $|P_2|$ solcher Intervalle I_k geben. Also gilt für die Summe der Intervalllängen $|I_k|$:

$$\sum_k |I_k| \leq |P_2| \cdot \delta(P_1).$$

Sei K_f eine Schranke für den Integranden f , also $|f(t)| \leq K_f$ für alle $t \in [a, b]$. Es folgt für beliebige Zerlegungen P_1, P_2 :

$$O_{f,P_1} - O_{f,P_1 \cup P_2} \leq \sum_k K_f \cdot |I_k| \leq K_f \cdot |P_2| \cdot \delta(P_1),$$

$$U_{f,P_1 \cup P_2} - U_{f,P_1} \leq \sum_k K_f \cdot |I_k| \leq K_f \cdot |P_2| \cdot \delta(P_1).$$

Ende der Vorüberlegung.

Sei nun $\epsilon > 0$ gegeben. Wird f als Riemann-integrierbar vorausgesetzt, gibt es eine Partition P mit

$$O_{f,P} - U_{f,P} \leq \epsilon.$$

Nach der obigen Vorüberlegung gilt

$$\begin{aligned} O_{f,P_n} - U_{f,P_n} &= \underbrace{O_{f,P_n} - O_{f,P \cup P_n}}_{\leq K_f \cdot |P| \cdot \delta(P_n)} + \underbrace{O_{f,P \cup P_n} - U_{f,P \cup P_n}}_{\leq O_{f,P} - U_{f,P} \leq \epsilon} + \underbrace{U_{f,P \cup P_n} - U_{f,P_n}}_{\leq K_f \cdot |P| \cdot \delta(P_n)} \\ &\leq K_f \cdot |P| \cdot \delta(P_n) + \epsilon + K_f \cdot |P| \cdot \delta(P_n). \end{aligned}$$

Da die Feinheiten $\delta(P_n)$ gegen 0 konvergieren, gibt es (bei fixiertem P) ein n_0 , so dass $K_f \cdot |P| \cdot \delta(P_n) \leq \epsilon$ für alle $n \geq n_0$ gilt, also

$$O_{f,P_n} - U_{f,P_n} \leq \epsilon + \epsilon + \epsilon = 3 \cdot \epsilon.$$

Also folgt $O_{f,P_n} - U_{f,P_n} \rightarrow 0$ für $n \rightarrow \infty$. Hieraus folgt unmittelbar:

$$O_{f,P_n} - \int_a^b f(t) dt \leq O_{f,P_n} - U_{f,P_n} \xrightarrow{n \rightarrow \infty} 0,$$

$$\int_a^b f(t) dt - U_{f,P_n} \leq O_{f,P_n} - U_{f,P_n} \xrightarrow{n \rightarrow \infty} 0.$$

Q.E.D.

Bemerkung 9.31: Gilt für irgendeine Zerlegungsfolge P_n von $[a, b]$

$$\lim_{n \rightarrow \infty} O_{f, P_n} = \lim_{n \rightarrow \infty} U_{f, P_n},$$

so folgt hieraus bereits die Riemann-Integrierbarkeit von f gemäß Definition 9.29, denn $\lim_{n \rightarrow \infty} (O_{f, P_n} - U_{f, P_n}) = 0$, d.h., zu jedem $\epsilon > 0$ gibt eine Partition P_n mit $O_{f, P_n} - U_{f, P_n} \leq \epsilon$.

Satz 9.32: (Integrabilitätskriterium)

Eine beschränkte Funktion $f : [a, b] \rightarrow \mathbb{R}$ mit höchstens endlich vielen Unstetigkeitspunkten in $[a, b]$ ist Riemann-integrierbar über $[a, b]$.

Beweisskizze: (technisch) Wir betrachten zunächst den Fall, dass f keinen Unstetigkeitspunkt in $[a, b]$ hat. Man kann zeigen, dass eine stetige Funktion auf einem abgeschlossenen Intervall $[a, b]$ gleichmäßig stetig ist, d.h., zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ mit $|f(x) - f(y)| \leq \epsilon$ für alle $x, y \in [a, b]$ mit $|x - y| \leq \delta$. Wähle n so, dass $(b - a)/n \leq \delta$ gilt. Für die äquidistante Partition $P_n = \{t_0, \dots, t_n\}$ mit $t_i = a + (b - a) \cdot \frac{i}{n}$ gilt dann

$$\begin{aligned} O_{f, P_n} - U_{f, P_n} &= \sum_{i=1}^n (t_i - t_{i-1}) \cdot \max\{f(x) - f(y); x, y \in [t_{i-1}, t_i]\} \\ &\leq \sum_{i=1}^n \frac{b - a}{n} \cdot \epsilon = (b - a) \cdot \epsilon. \end{aligned}$$

Nach Definition 9.29 ist f damit Riemann-integrierbar.

Betrachte nun den Fall, dass es einen Unstetigkeitspunkt x von f in $[a, b]$ gibt. Zerlege $[a, b]$ in die drei Teilintervalle $[a, x - \frac{1}{n}]$, $[x - \frac{1}{n}, x + \frac{1}{n}]$, $[x + \frac{1}{n}, b]$. Für die Teilintervalle $[a, x - \frac{1}{n}]$ und $[x + \frac{1}{n}, b]$, auf denen f stetig ist, läßt sich analog zu den obigen Überlegungen die Differenz zwischen Ober- und Untersumme beliebig klein machen, indem man diese Teilintervalle hinreichend fein äquidistant zerlegt. Für den Beitrag des Intervalls $[x - \frac{1}{n}, x + \frac{1}{n}]$ gilt mit $|f(t)| \leq K$:

$$\sup \left\{ f(t); t \in [x - \frac{1}{n}, x + \frac{1}{n}] \right\} - \inf \left\{ f(t); t \in [x - \frac{1}{n}, x + \frac{1}{n}] \right\} \leq \frac{4 \cdot K}{n}.$$

Dieser Beitrag zur Differenz zwischen Ober- und Untersumme ist ebenfalls beliebig klein zu machen, in dem man n hinreichend gross wählt. Nach Definition 9.29 ist f damit Riemann-integrierbar.

Allgemeiner: Die Funktion f habe N Unstetigkeitspunkte. Zerlege $[a, b]$ in kleine Teilintervalle der Länge $\frac{2}{n}$ um die Unstetigkeitspunkte herum, zerlege den Rest des Intervalls äquidistant. Analog zu oben ist die Differenz zwischen Ober- und

Untersumme beliebig klein zu machen, in dem man einerseits n hinreichend gross wählt und andererseits den Rest des Intervalls hinreichend fein unterteilt.

Q.E.D.

Bemerkung 9.33: Man kann für integrierbares f das Integral als Grenzwert spezieller Riemann-Summen bekommen, indem man das Intervall $[a, b]$ speziell äquidistant unterteilt:

$$\int_a^b f(t) dt = \lim_{n \rightarrow \infty} \frac{b-a}{n} \cdot \sum_{i=0}^{n-1} f(t_i^{(n)}), \quad t_i^{(n)} = \frac{a+i \cdot (b-a)}{n}.$$

Beweis: Für die äquidistanten Zerlegungen $P_n = \{t_0^{(n)}, \dots, t_n^{(n)}\}$ mit den obigen Zwischenstellen $t_i^{(n)}$ gilt mit

$$R_{f, P_n} = \frac{b-a}{n} \cdot \sum_{i=0}^{n-1} f(t_i^{(n)})$$

offensichtlich $U_{f, P_n} \leq R_{f, P_n} \leq O_{f, P_n}$. Nach Satz 9.30 gilt

$$\int_a^b f(t) dt = \lim_{n \rightarrow \infty} O_{f, P_n} = \lim_{n \rightarrow \infty} U_{f, P_n}$$

und damit auch

$$\int_a^b f(t) dt = \lim_{n \rightarrow \infty} R_{f, P_n}.$$

Q.E.D.

↓11.12.03

Definition 9.29 und auch die Konstruktionen nach Satz 9.30 oder auch Bemerkung 9.33 sind zur konkreten Berechnung des Integrals völlig ungeeignet. Die wirkliche Berechnung geschieht über Stammfunktionen von $f(t)$, sobald der Zusammenhang zwischen dem bestimmten Integral und dem unbestimmten Integral geklärt ist (nächster Abschnitt).

Bemerkung 9.34: Das bestimmte Integral kann auch negative Werte annehmen (z.B., wenn überall $f(t) < 0$ gilt). Die Interpretation als „Fläche unter dem Graphen“ gilt nur für positive Funktionen. Aus der Darstellung in Bemerkung 9.33 ergibt sich z.B., dass das Integral negativ ist, wenn überall $f(t) < 0$ gilt.

Bemerkung 9.35: Aus der Darstellung 9.33 des Integrals als Grenzwert von Riemann-Summen R_{f,P_n} folgen unmittelbar einige elementare Eigenschaften des Integrals:

a) Linearität:

$$\int_a^b (\alpha \cdot f(t) + \beta \cdot g(t)) dt = \alpha \cdot \int_a^b f(t) dt + \beta \cdot \int_a^b g(t) dt, \quad \alpha, \beta \in \mathbb{R},$$

b) Monotonie:

$$\int_a^b f(t) dt \leq \int_a^b g(t) dt,$$

wenn $f(t) \leq g(t)$ gilt $\forall t \in [a, b]$,

c) Ist f integrierbar, so auch $|f|$ und es gilt:

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt.$$

Zu c): Die Integrierbarkeit von $|f|$ folgt unmittelbar durch die umgekehrte Dreiecksungleichung

$$\begin{aligned} O_{|f|,P} - U_{|f|,P} &= \sum \left(\sup\{|f(t)|, t \in [t_{i-1}, t_i]\} - \inf\{|f(t)|, t \in [t_{i-1}, t_i]\} \right) \cdot (t_i - t_{i-1}) \\ &= \sum \left(\sup\{||f(x)| - |f(y)||, x, y \in [t_{i-1}, t_i]\} \right) \cdot (t_i - t_{i-1}) \\ &\leq \sum \left(\sup\{|f(x) - f(y)|, x, y \in [t_{i-1}, t_i]\} \right) \cdot (t_i - t_{i-1}) = O_{f,P} - U_{f,P} \end{aligned}$$

mittels Definition 9.29 aus der Integrierbarkeit von f . Die Abschätzung

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt$$

ergibt sich dann aus der Dreiecksungleichung angewendet auf die Riemann-Summen aus Bemerkung 9.33.

Satz 9.36: (Der Mittelwertsatz der Integralrechnung)

Ist $f : [a, b] \rightarrow \mathbb{R}$ stetig auf $[a, b]$, so gibt es einen Punkt $\xi \in (a, b)$ mit

$$\int_a^b f(t) dt = f(\xi) \cdot (b - a).$$

Beweis: Sei $m = \min\{f(t); t \in [a, b]\}$, $M = \max\{f(t); t \in [a, b]\}$ (für stetige Funktionen werden auf einem abgeschlossenen Intervall nach Satz 4.25 an gewissen Stellen t_{min} bzw. t_{max} Minimum bzw. Maximum angenommen). Nach Bemerkung 9.35.b) gilt

$$m \cdot (b - a) = \int_a^b m \, dt \leq \int_a^b f(t) \, dt \leq \int_a^b M \, dt = M \cdot (b - a),$$

also

$$m \leq \frac{1}{b - a} \cdot \int_a^b f(t) \, dt \leq M.$$

Nach dem Zwischenwertsatz 4.22 angewendet auf das Intervall zwischen t_{min} und t_{max} nimmt f alle Werte zwischen m und M an, d.h., es gibt ein ξ (zwischen t_{min} und t_{max}) mit

$$f(\xi) = \frac{1}{b - a} \cdot \int_a^b f(t) \, dt.$$

Q.E.D.

Bestimmte Integrale können additiv zerlegt werden. Man stelle sich dazu eine positive Funktion $f(t)$ vor, d.h., das Integral von a bis b ist die Fläche unter dem Graphen von $t = a$ bis $t = b$. Diese Fläche setzt sich zusammen aus der Fläche unter dem Graphen von $t = a$ bis $t = c$ und der Fläche von $t = c$ bis $t = b$, wobei der Zerlegungspunkt c beliebig gewählt werden kann:

Satz 9.37: (Zerlegung bestimmter Integrale)

Für beliebiges a, b, c gilt:

$$\int_a^c f(t) \, dt + \int_c^b f(t) \, dt = \int_a^b f(t) \, dt.$$

Beweisskizze: Zerlegungen von $[a, c]$ und $[c, b]$ lassen sich offensichtlich zu einer Zerlegung des Gesamtintervalls $[a, b]$ zusammensetzen. Eine Riemann-Summe für \int_a^c zusammen mit einer Riemann-Summe für \int_c^b ergibt eine Riemann-Summe für \int_a^b .

Q.E.D.

Konvention 9.38:

Wir setzen

$$\int_b^a f(t) \, dt = - \int_a^b f(t) \, dt,$$

womit wir in $\int_a^b f(t) \, dt$ nun auch $b < a$ zulassen können. Speziell gilt

$$\int_a^a f(t) \, dt = - \int_a^a f(t) \, dt = 0.$$

Mit dieser Konvention gilt Satz 9.37 auch für Zerlegungspunkte c , die außerhalb des Intervalls $[a, b]$ liegen.

Bemerkung 9.39: In MuPAD ist die Funktion `int` sowohl für bestimmte als auch für unbestimmte Integrale zuständig:

```
>> int(exp(-2*x), x)
```

$$-\frac{1}{2 \exp(x)}$$

```
>> int(exp(-2*t), t = 0..5)
```

$$1/2 - \frac{1}{2 \exp(5)}$$

```
>> float(%)
```

0.4999773

Bemerkung 9.40: Man beachte, dass das unbestimmte Integral $\int f(x) dx$ eine Funktion in x ist, während das bestimmte Integral $\int_a^b f(t) dt$ für konkrete Zahlenwerte a, b einen Zahlenwert darstellt. Diesen kann man numerisch approximieren, indem man z.B. die in der Bemerkung 9.33 gegebene Summe für großes n ausrechnet. Alternativ zur „Riemann-Summe“

$$\int_a^b f(t) dt \approx \frac{b-a}{n} \cdot \sum_{i=0}^{n-1} f\left(a + i \cdot \frac{b-a}{n}\right) \approx \frac{b-a}{n} \cdot \sum_{i=1}^n f\left(a + i \cdot \frac{b-a}{n}\right)$$

ist es günstiger, stattdessen die „Trapez-Summe“

$$\int_a^b f(t) dt \approx \frac{b-a}{n} \left(\frac{f(a)}{2} + \sum_{i=1}^{n-1} f\left(a + i \cdot \frac{b-a}{n}\right) + \frac{f(b)}{2} \right)$$

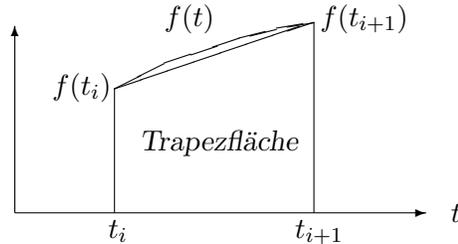
zu berechnen, die sich mit $t_i = a + i \cdot \frac{b-a}{n}$ auch als

$$\frac{b-a}{n} \cdot \sum_{i=0}^{n-1} \frac{f(t_i) + f(t_{i+1})}{2}$$

schreiben lässt. Hierbei ist $\frac{b-a}{n} \cdot \frac{f(t_i) + f(t_{i+1})}{2}$ die Fläche des durch die 4 Punkte

$$(t_i, 0), \quad (t_i, f(t_i)), \quad (t_{i+1}, f(t_{i+1})), \quad (t_{i+1}, 0)$$

definierten Trapezes (d.h., die Fläche unter dem Graphen von $f(t)$ wird nicht durch Rechtecke, sondern durch Trapeze angenähert).



Bemerkung 9.41: In MuPAD ist die Funktion `numeric::int` für die numerische Berechnung von bestimmten Integralen zuständig. Sie arbeitet auch dann, wenn der symbolische Integrator kein Ergebnis liefert (weil er keine Stammfunktion findet):

```
>> int(exp(sqrt(t))*sqrt(t), t = 0..10)
```

```
      1/2      1/2
int(t      exp(t      ), t = 0..10)
```

```
>> numeric::int(exp(sqrt(t))*sqrt(t), t = 0..10)
```

```
264.1573027
```

9.3 Der Hauptsatz: Zusammenhang zwischen bestimmtem und unbestimmtem Integral

Es verbleibt das Problem, wie man effektiv bestimmte Integrale $\int_a^b f(t) dt$ ohne den garstigen Grenzwert von Riemann-Summen berechnen kann. Hier kommt die wesentliche Beobachtung ins Spiel, dass man mit unbestimmten Integralen (Stammfunktionen) bestimmte Integrale ausrechnen kann.

Satz 9.42: (Der Hauptsatz der Differential- und Integralrechnung, Version 1)

Betrachte

$$F_a(x) = \int_a^x f(t) dt.$$

Für stetiges f ist F_a differenzierbar, und es gilt

$$\frac{d}{dx} F_a(x) = f(x),$$

d.h., $F_a(x)$ ist eine Stammfunktion von $f(x)$.

Beweis: Es gilt

$$\Delta F_a = F_a(x+h) - F_a(x) = \int_a^{x+h} f(t) dt - \int_a^x f(t) dt \stackrel{(9.37)}{=} \int_x^{x+h} f(t) dt.$$

Nach dem Mittelwertsatz 9.36 gilt

$$\Delta F_a = f(\xi) \cdot h$$

mit einem Zwischenwert ξ zwischen x und $x+h$. Damit gilt

$$\frac{d}{dx} F_a(x) = \lim_{h \rightarrow 0} \frac{\Delta F_a}{h} = \lim_{h \rightarrow 0} f(\xi) = f(x),$$

da für $h \rightarrow 0$ der Zwischenwert ξ zwischen x und $x+h$ gegen x streben muss.

Q.E.D.

Bemerkung 9.43: *Stammfunktionen sind nur bis auf additive Konstanten bestimmt. Dies wird in der Darstellung einer Stammfunktion über $F_a(x) = \int_a^x f(t) dt$ dadurch deutlich, dass die untere Grenze a beliebig wählbar ist. Die Konstante ist hier durch die Bedingung $F_a(a) = \int_a^a f(t) dt = 0$ festgelegt. Bei unterschiedlicher Wahl der unteren Grenze ist die Differenz der entsprechenden Stammfunktionen in der Tat eine Konstante:*

$$\begin{aligned} F_{a_1}(x) - F_{a_2}(x) &= \int_{a_1}^x f(t) dt - \int_{a_2}^x f(t) dt \\ \stackrel{(9.37)}{=} &\left(\int_{a_1}^{a_2} f(t) dt + \int_{a_2}^x f(t) dt \right) - \int_{a_2}^x f(t) dt = \underbrace{\int_{a_1}^{a_2} f(t) dt}_{\text{unabhängig von } x} \end{aligned}$$

Bestimmte Integrale sind also Stammfunktionen, wenn man sie als Funktion der oberen Grenze auffasst. Umgekehrt, kennt man ein Stammfunktion, so liefert sie ein bestimmtes Integral, denn alle Stammfunktionen $F(x)$ von $f(x)$ unterscheiden sich nur um eine additive Konstante, d.h., es muss gelten

$$F_a(x) = \int_a^x f(t) dt = F(x) + c.$$

Es verbleibt nur, die Integrationskonstante c zu identifizieren. Für $x = a$ folgt

$$0 = \int_a^a f(t) dt = F(a) + c \quad \Rightarrow \quad c = -F(a),$$

also

$$\int_a^x f(t) dt = F(x) - F(a).$$

Dies liefert nun eine effektive Methode, bestimmte Integrale auszurechnen, indem man sich zunächst eine Stammfunktion des Integranden verschafft:

Satz 9.44: (Der Hauptsatz der Differential- und Integralrechnung, Version 2)

Sei $F(x)$ eine beliebige stetige Stammfunktion von $f(x)$. Dann gilt

$$\int_a^b f(t) dt = F(b) - F(a).$$

Die additive Konstante der Stammfunktion fällt dabei bei Differenzbildung heraus.

↓16.12.03

Bemerkung 9.45: Man beachte die Voraussetzung, dass die Stammfunktion $F(x)$ stetig zu sein hat! Bei unstetigen Stammfunktionen (falls der Integrand eine Singularität hat), muss man vorsichtig sein! Gegenbeispiel: Zum Integranden $f(t) = 1/t$ gehört die Stammfunktion $F(x) = \ln(|x|)$, die jedoch am Nullpunkt unstetig ist. In der Tat:

$$\int_{-1}^2 \frac{dx}{x} \neq \ln(|2|) - \ln(|-1|) = \ln(2).$$

(Das Integral $\int_{-1}^2 \frac{dx}{x}$ existiert gar nicht.)

Beispiel 9.46: Zur Berechnung von $\int_1^2 \ln(t) dt$ berechnet man zunächst eine Stammfunktion von $\ln(x)$. Analog zu Beispiel 9.11 ergibt sich durch partielle Integration:

$$\begin{aligned} \int \ln(x) dx &= \int \underbrace{\ln(x)}_{f(x)} \cdot \underbrace{1}_{g'(x)} dx = \underbrace{\ln(x)}_{f(x)} \cdot \underbrace{x}_{g(x)} - \int \underbrace{\frac{1}{x}}_{f'(x)} \cdot \underbrace{x}_{g(x)} dx \\ &= x \cdot \ln(x) - \int 1 dx = x \cdot \ln(x) - x + c. \end{aligned}$$

Mit der Stammfunktion $F(x) = x \cdot \ln(x) - x + c$ ergibt sich das bestimmte Integral

$$\int_1^2 \ln(t) dt = F(2) - F(1) = (2 \cdot \ln(2) - 2 + c) - (1 \cdot \ln(1) - 1 + c) = 2 \cdot \ln(2) - 1.$$

Bemerkung 9.47: Aus dem Zusammenhang mit dem unbestimmten Integral folgt sofort, dass die Rechenregeln aus Abschnitt 9.1 auch für bestimmte Integrale gelten, z.B. (Satz 9.7):

$$\int_a^b (c_1 \cdot f_1(t) + c_2 \cdot f_2(t)) dt = c_1 \cdot \int_a^b f_1(t) dt + c_2 \cdot \int_a^b f_2(t) dt.$$

Partielle Integration gilt in der folgenden Form:

$$\int_a^b f(t) \cdot g'(t) dt = \left[f(t) \cdot g(t) \right]_{t=a}^{t=b} - \int_a^b f'(t) \cdot g(t) dt,$$

wobei $\left[f(t) \cdot g(t) \right]_{t=a}^{t=b}$ als Abkürzung für

$$\left[f(t) \cdot g(t) \right]_{t=a}^{t=b} = f(b) \cdot g(b) - f(a) \cdot g(a)$$

dient. Substitution gilt in der folgenden Form:

$$\int_a^b f(g(t)) \cdot g'(t) dt = \int_{g(a)}^{g(b)} f(y) dy.$$

Beispiel 9.48: Partielle Integration:

$$\begin{aligned} \int_0^1 \underbrace{t}_{f(t)} \cdot \underbrace{\cos(t)}_{g'(t)} dt &= \left[\underbrace{t}_{f(t)} \cdot \underbrace{\sin(t)}_{g(t)} \right]_{t=0}^{t=1} - \int_0^1 \underbrace{1}_{f'(t)} \cdot \underbrace{\sin(t)}_{g(t)} dt \\ &= \left[t \cdot \sin(t) \right]_{t=0}^{t=1} - \left[-\cos(t) \right]_{t=0}^{t=1} \\ &= 1 \cdot \sin(1) - 0 \cdot \sin(0) + \cos(1) - \cos(0) = \sin(1) + \cos(1) - 1. \end{aligned}$$

Beispiel 9.49: Substitution $y = t^2$, $dy = 2t dt$:

$$\begin{aligned} \int_0^{\sqrt{\pi}} t \cdot \cos(t^2) dt &= \frac{1}{2} \cdot \int_0^{\sqrt{\pi}} \cos(t^2) \cdot \underbrace{2 \cdot t dt}_{dy} = \frac{1}{2} \cdot \int_0^{\pi} \cos(y) dy \\ &= \frac{1}{2} \cdot \left[\sin(y) \right]_{y=0}^{y=\pi} = \frac{1}{2} \cdot (\sin(\pi) - \sin(0)) = 0. \end{aligned}$$

Man beachte hierbei, wie sich im Substitutionsschritt die Grenzen ändern: Für $t = 0$ folgt $y = t^2 = 0$, für $t = \sqrt{\pi}$ folgt $y = t^2 = \pi$.

9.4 Uneigentliche Integrale

Bestimmte Integrale $\int_a^b f(t) dt$ sind zunächst nur für endliche Intervalle $[a, b]$ definiert. Wir erweitern die Definition:

Definition 9.50: (Uneigentliche Integrale)

$$\int_a^\infty f(t) dt = \lim_{b \rightarrow \infty} \int_a^b f(t) dt,$$

$$\int_{-\infty}^b f(t) dt = \lim_{a \rightarrow -\infty} \int_a^b f(t) dt,$$

$$\int_{-\infty}^\infty f(t) dt = \lim_{a \rightarrow -\infty} \lim_{b \rightarrow \infty} \int_a^b f(t) dt$$

(falls die Grenzwerte existieren).

Beispiel 9.51:

$$\int_0^\infty e^{-t} dt = \lim_{b \rightarrow \infty} \int_0^b e^{-t} dt = \lim_{b \rightarrow \infty} \left[-e^{-t} \right]_{t=0}^{t=b} = \lim_{b \rightarrow \infty} (-e^{-b} + 1) = 1 - \lim_{b \rightarrow \infty} e^{-b} = 1.$$

Beispiel 9.52: Substitution $y = -\sqrt{t}$, $\frac{dy}{dt} = -\frac{1}{2\sqrt{t}} = \frac{1}{2 \cdot y}$, $dt = 2 \cdot y \cdot dy$:

$$\frac{1}{2} \cdot \int_0^\infty e^{-\sqrt{t}} dt = \frac{1}{2} \cdot \int_0^{-\infty} e^y \cdot 2 \cdot y dy \stackrel{(9.38)}{=} - \int_{-\infty}^0 e^y \cdot y dy.$$

Man achte hierbei auf die Transformation der Grenzen: $t = 0$ entspricht $y = -\sqrt{t} = 0$, $t = \infty$ entspricht $y = -\sqrt{t} = -\infty$. Das verbleibende Integral war bereits in Beispiel 9.12 gelöst worden:

$$\begin{aligned} - \int_{-\infty}^0 e^y \cdot y dy &= - \lim_{a \rightarrow -\infty} \left[(y-1) \cdot e^y \right]_{y=a}^{y=0} \\ &= - \lim_{a \rightarrow -\infty} \left(-1 - (a-1) \cdot e^a \right) = 1 - \lim_{a \rightarrow -\infty} \left((1-a) \cdot e^a \right). \end{aligned}$$

Der verbleibende Grenzwert ist 0:

$$\lim_{a \rightarrow -\infty} \left((1-a) \cdot e^a \right) \stackrel{(b=-a)}{=} \lim_{b \rightarrow \infty} \left((1+b) \cdot e^{-b} \right) = \lim_{b \rightarrow \infty} \left(\frac{b+1}{e^b} \right).$$

Da mit $e^b = 1 + b + \frac{b^2}{2} + \dots$ die Exponentialfunktion für $b \rightarrow \infty$ stärker steigt als jedes Polynom, ist der Grenzwert 0. Endergebnis:

$$\frac{1}{2} \cdot \int_0^\infty e^{-\sqrt{t}} dt = 1.$$

Man geht ähnlich vor, wenn der Integrand eine Singularität hat:

Definition 9.53: (Uneigentliche Integrale bei singulären Integranden)

Hat der Integrand $f(t)$ an der Stelle a oder b eine Singularität, so definiert man

$$\int_a^b f(t) dt = \lim_{\epsilon \rightarrow 0+0} \int_a^{b-\epsilon} f(t) dt,$$

bzw.

$$\int_a^b f(t) dt = \lim_{\epsilon \rightarrow 0+0} \int_{a+\epsilon}^b f(t) dt$$

(falls die Grenzwerte existieren).

Beispiel 9.54: Im folgenden Fall existiert das uneigentliche Integral:

$$\begin{aligned} \int_0^1 \frac{1}{\sqrt{t}} dt &= \lim_{\epsilon \rightarrow 0+0} \int_{\epsilon}^1 t^{-\frac{1}{2}} dt = \lim_{\epsilon \rightarrow 0+0} \left[\frac{t^{\frac{1}{2}}}{\frac{1}{2}} \right]_{t=\epsilon}^{t=1} \\ &= \lim_{\epsilon \rightarrow 0+0} \left[2 \cdot \sqrt{t} \right]_{t=\epsilon}^{t=1} = 2 \cdot \lim_{\epsilon \rightarrow 0+0} (1 - \sqrt{\epsilon}) = 2. \end{aligned}$$

Beispiel 9.55: Im folgenden Fall existiert das uneigentliche Integral nicht (bzw. ist ∞):

$$\int_0^1 \frac{1}{t} dt = \lim_{\epsilon \rightarrow 0+0} \int_{\epsilon}^1 \frac{1}{t} dt = \lim_{\epsilon \rightarrow 0+0} \left[\ln(t) \right]_{t=\epsilon}^{t=1} = \lim_{\epsilon \rightarrow 0+0} (0 - \ln(\epsilon)) = \infty.$$

9.5 Die Diracsche Delta-Funktion

Physiker und Ingenieure betrachten oft Signale, die “kurze aber intensive Impulse” zu einem Zeitpunkt t_0 darstellen:

$$\text{Impuls}(t - t_0) = \begin{cases} 0 & \text{für } t - t_0 < -\frac{\epsilon}{2}, \\ \frac{1}{\epsilon} & \text{für } -\frac{\epsilon}{2} \leq t - t_0 \leq \frac{\epsilon}{2}, \\ 0 & \text{für } t - t_0 > \frac{\epsilon}{2}. \end{cases}$$

Für $0 < \epsilon \ll 1$ ist dies ein Signal mit der grossen Amplitude $1/\epsilon$, aber kurzer Lebensdauer ϵ . Die durch

$$\int_{-\infty}^{\infty} \text{Impuls}(t - t_0) dt$$

gegebene "Gesamtenergie" des Signals ist im obigen Fall 1. Betrachtet man den Grenzwert $\epsilon \rightarrow 0$, so ergibt sich punktweise die Grenzfunktion

$$\delta(t - t_0) = \begin{cases} 0 & \text{für } t \neq t_0, \\ \infty & \text{für } t = t_0. \end{cases}$$

Wegen des Funktionswerts ∞ bei $t = t_0$ ist dies nicht wirklich ein Funktion von \mathbb{R} nach \mathbb{R} , sondern eine sogenannte „**verallgemeinerte Funktion**“ oder auch „**Distribution**“. Ihre wesentliche Eigenschaft ist:

$$\int_{-\infty}^{\infty} \delta(t - t_0) dt = 1.$$

Die „Diracsche δ -Funktion“ macht eigentlich nur unter einem Integralzeichen Sinn, wo für stetiges $f(t)$ ihr Verhalten durch

$$\int_a^b \delta(t - t_0) \cdot f(t) dt = \begin{cases} f(t_0) & \text{falls } t_0 \in (a, b), \\ 0 & \text{falls } t_0 \notin [a, b] \end{cases}$$

definiert wird. (Für $t_0 = a$ oder $t_0 = b$ kann man sich selbst aussuchen, was das Ergebnis sein soll. Dieser Grenzfall ist für die Praxis nicht interessant.) Diese Vereinbarung führt zu

$$\text{Heaviside}(x - t_0) = \int_{-\infty}^x \delta(t - t_0) dt = \begin{cases} 0 & \text{für } x < t_0, \\ 1 & \text{für } x > t_0. \end{cases}$$

Damit ist die Heavisidesche Sprungfunktion formal die Stammfunktion der δ -Funktion. Rein formal ergibt dies: die δ -Funktion ist die Ableitung der Sprungfunktion:

$$\frac{d}{dx} \text{Heaviside}(x - x_0) = \delta(x - x_0).$$

In vielen physikalischen Anwendungen gibt es Formeln, die Feldgrößen als Integral darstellen, z.B.

$$\text{Gravitationspotential}(x) = \int_{-\infty}^{\infty} \frac{\rho(t)}{|x - t|} dt,$$

wo $\rho(t)$ eine Massenverteilung darstellt. Wie stellt man nun eine Punktmasse am Punkt x_0 dar? Als „normale Funktion“ mit $\rho(t) = 0$ für alle $t \neq x_0$ würde das obige Integral immer 0 ergeben.

Mittels $\rho(t) = \text{Masse} \cdot \delta(t - x_0)$ ergibt sich das gewünschte Graviationspotential

$$\text{Gravitationspotential}(x) = \int_{-\infty}^{\infty} \frac{\rho(t)}{|x - t|} dt = \frac{\text{Masse}}{|x - x_0|}.$$

Bemerkung 9.56: In MuPAD stehen δ -Funktion und Sprungfunktion als `dirac` und `heaviside` zur Verfügung:

```
>> int(dirac(t - t0)*f(t), t = -infinity .. infinity)
```

`f(t0)`

```
>> diff(heaviside(x - x0), x)
```

`dirac(x - x0)`

9.6 Einige spezielle Anwendungen

17.12.03↓

Satz 9.57: (logarithmische Divergenz der harmonischen Reihe)

Die Folge $\sum_{k=1}^n \frac{1}{k} - \ln(n)$ konvergiert monoton fallend gegen einen Grenzwert $C \approx 0.5772\dots$ (die „Eulersche Konstante“):

$$\sum_{k=1}^n \frac{1}{k} \approx \ln(n) + C.$$

Beweis: Sei $C_n = \sum_{k=1}^n \frac{1}{k} - \ln(n)$. Mit

$$\int_k^{k+1} \frac{1}{x} dx = \ln(k+1) - \ln(k)$$

gilt

$$\begin{aligned} C_n &> \sum_{k=1}^n \frac{1}{k} - \ln(n+1) = \sum_{k=1}^n \frac{1}{k} - \int_1^{n+1} \frac{1}{x} dx = \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \int_k^{k+1} \frac{1}{x} dx \\ &= \sum_{k=1}^n \left(\frac{1}{k} - \int_k^{k+1} \frac{1}{x} dx \right) = \sum_{k=1}^n \int_k^{k+1} \left(\frac{1}{k} - \frac{1}{x} \right) dx > 0, \end{aligned}$$

denn für die monoton fallende Funktion $\frac{1}{x}$ gilt $\frac{1}{k} \geq \frac{1}{x}$ auf dem Intervall $[k, k+1]$. Weiterhin gilt

$$C_n - C_{n+1} = \left(\ln(n+1) - \ln(n) \right) - \frac{1}{n+1} = \int_n^{n+1} \frac{1}{x} dx - \frac{1}{n+1}$$

$$= \int_n^{n+1} \left(\frac{1}{x} - \frac{1}{n+1} \right) dx \geq 0,$$

denn es gilt $\frac{1}{x} \geq \frac{1}{n+1}$ für $x \in [n, n+1]$. Damit ist die Folge (C_n) monoton fallend und nach unten durch 0 beschränkt. Sie konvergiert also gegen einen Grenzwert C .

Q.E.D.

Als weitere „Anwendung“ der Integration versuchen wir, realistische Abschätzungen von $n!$ für $n \gg 1$ zu ermitteln. Zunächst beobachtet man

$$\ln(n!) = \ln(1 \cdot 2 \cdot 3 \cdots n) = \ln(1) + \ln(2) + \ln(3) + \cdots + \ln(n) = \sum_{k=1}^n \ln(k) = \sum_{k=2}^n \ln(k).$$

Diese Summe lässt sich als Riemann-Summe interpretieren, die bei

$$\int_1^n \ln(x) dx = \left[x \cdot (\ln(x) - 1) \right]_{x=1}^{x=n} = n \cdot \ln(n) - n + 1$$

anfällt, wenn man das Integrationsintervall $[1, n]$ in die $n-1$ Teilintervalle $[1, 2]$, $[2, 3]$, \dots , $[n-1, n]$ zerlegt. Wegen der Monotonie von $\ln(x)$ gilt

$$\sum_{k=1}^{n-1} \ln(k) \leq \int_1^n \ln(x) dx = \sum_{k=1}^{n-1} \int_k^{k+1} \ln(x) dx \leq \sum_{k=2}^n \ln(k),$$

also

$$\ln((n-1)!) \leq n \cdot \ln(n) - n + 1 \leq \ln(n!),$$

also

$$(n-1)! \leq n^n \cdot e^{-n+1} \leq n!,$$

also (in der linken Ungleichung wird n durch $n+1$ ersetzt):

$$n^n \cdot e^{-n+1} \leq n! \leq (n+1)^{n+1} \cdot e^{-n},$$

also

$$\boxed{e \cdot \left(\frac{n}{e}\right)^n \leq n! \leq e \cdot \left(\frac{n+1}{e}\right)^{n+1}}.$$

Hiermit ist das Wachstumsverhalten von $n!$ charakterisiert. Diese Abschätzung lässt sich jedoch noch wesentlich verfeinern:

Satz 9.58: (Die Stirling-Formel)

Für alle $n \in \mathbb{N}$ gilt folgende Abschätzung von $n!$:

$$\sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{4n}}.$$

Für großes n gilt $e^{\frac{1}{4n}} = 1 + \frac{1}{4n} + \dots \approx 1$, d.h., das Verhältnis der oberen Schranke zur unteren Schranke ist für großes n dicht bei 1 (d.h., die führenden Stellen der oberen und unteren Schranke sind gleich und stimmen mit den führenden Stellen von $n!$ überein).

Merke:

$$n! \approx \sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n.$$

Diese „**Stirling-Approximation**“ für $n!$ ist schon ab $n = 4$ auf etwa 2 Prozent genau! Beispiel:

n	$\sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n$	$n!$	$\sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{4n}}$
2	1.9...	2	2.1...
5	118.0...	120	124.0...
10	3598695.6...	3628800	3689797.0...

Beweis: (für technisch Interessierte)

Es ist zu zeigen:

$$\sqrt{2 \cdot \pi} \leq \frac{n!}{\sqrt{n} \cdot \left(\frac{n}{e}\right)^n} = \frac{n!}{\underbrace{n^{n+\frac{1}{2}} \cdot e^{-n}}_{a_n}} \leq \sqrt{2 \cdot \pi} \cdot e^{\frac{1}{4n}}$$

Wir zeigen zunächst, dass die Folge

$$a_n = \frac{n!}{n^{n+\frac{1}{2}} \cdot e^{-n}}$$

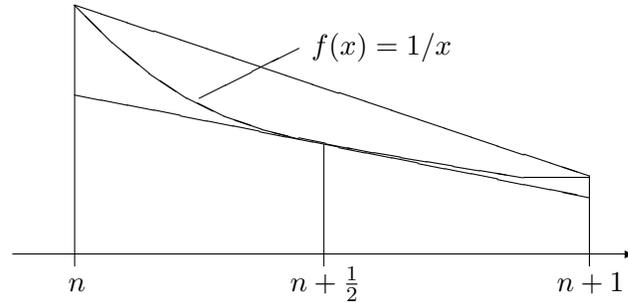
konvergiert. Für die Quotienten aufeinander folgender Elemente bekommt man

$$\frac{a_n}{a_{n+1}} = \frac{1}{e} \cdot \left(1 + \frac{1}{n}\right)^{n+\frac{1}{2}}$$

und damit

$$\ln\left(\frac{a_n}{a_{n+1}}\right) = \left(n + \frac{1}{2}\right) \cdot \ln\left(1 + \frac{1}{n}\right) - 1. \quad (\#)$$

Betrachte die Integration der Funktion $f(x) = 1/x$ über dem Intervall $[n, n+1]$:



Das Integral wird nach oben abgeschätzt durch das Trapez durch die Punkte $(n, f(n))$ und $(n+1, f(n+1))$. Die Trapezfläche ist *Breite* \times *mittlere Höhe* = $\frac{1}{2} \cdot (f(n) + f(n+1))$:

$$\int_n^{n+1} \frac{1}{x} dx = \ln(n+1) - \ln(n) = \ln\left(1 + \frac{1}{n}\right) \leq \frac{1}{2} \cdot \left(\frac{1}{n} + \frac{1}{n+1}\right).$$

Das Integral wird nach unten abgeschätzt durch das Trapez, dessen obere Kante die Tangente an $f(x)$ im Mittelpunkt ist. Die Trapezfläche ist *Breite* \times *mittlere Höhe* = $f(n + \frac{1}{2})$:

$$\int_n^{n+1} \frac{1}{x} dx = \ln(n+1) - \ln(n) = \ln\left(1 + \frac{1}{n}\right) \geq \frac{1}{n + \frac{1}{2}}.$$

Diese Abschätzungen liefern die Ungleichungskette

$$\frac{1}{n + \frac{1}{2}} \leq \ln\left(1 + \frac{1}{n}\right) \leq \frac{1}{2} \cdot \left(\frac{1}{n} + \frac{1}{n+1}\right),$$

also

$$1 \leq \left(n + \frac{1}{2}\right) \cdot \ln\left(1 + \frac{1}{n}\right) \leq \left(n + \frac{1}{2}\right) \cdot \frac{1}{2} \cdot \left(\frac{1}{n} + \frac{1}{n+1}\right),$$

also

$$\begin{aligned} 0 &\leq \left(n + \frac{1}{2}\right) \cdot \ln\left(1 + \frac{1}{n}\right) - 1 \leq \left(n + \frac{1}{2}\right) \cdot \frac{1}{2} \cdot \left(\frac{1}{n} + \frac{1}{n+1}\right) - 1 \\ &= \frac{1/4}{n \cdot (n+1)} = \frac{1}{4} \cdot \left(\frac{1}{n} - \frac{1}{n+1}\right). \end{aligned}$$

Eingesetzt in (#) erhält man:

$$0 \leq \ln\left(\frac{a_n}{a_{n+1}}\right) \leq \frac{1}{4} \cdot \left(\frac{1}{n} - \frac{1}{n+1}\right),$$

also

$$1 \leq \frac{a_n}{a_{n+1}} \leq e^{\frac{1}{4n}} \cdot e^{-\frac{1}{4(n+1)}}.$$

Es folgt

$$\begin{aligned}
 1 &\leq \frac{a_n}{a_{n+k}} = \frac{a_n}{a_{n+1}} \cdot \frac{a_{n+1}}{a_{n+2}} \cdot \dots \cdot \frac{a_{n+k-1}}{a_{n+k}} \\
 &\leq e^{\frac{1}{4 \cdot n}} \cdot e^{-\frac{1}{4 \cdot (n+1)}} \cdot e^{\frac{1}{4 \cdot (n+1)}} \cdot e^{-\frac{1}{4 \cdot (n+2)}} \cdot \dots \cdot e^{\frac{1}{4 \cdot (n+k-1)}} \cdot e^{-\frac{1}{4 \cdot (n+k)}} \\
 &= e^{\frac{1}{4 \cdot n}} \cdot e^{-\frac{1}{4 \cdot (n+k)}} \leq e^{\frac{1}{4 \cdot n}}
 \end{aligned}$$

für alle $k \geq 1$. Für fixiertes n ist die Folge (a_{n+k}) (im Folgenindex k) damit monoton fallend und nach unten beschränkt, d.h., es existiert der Grenzwert $a^* = \lim_{k \rightarrow \infty} a_{n+k} = \lim_{n \rightarrow \infty} a_n$, für den gilt:

$$1 \leq \frac{a_n}{a^*} \leq e^{\frac{1}{4 \cdot n}} \quad \Rightarrow \quad a^* \leq a_n \leq a^* \cdot e^{\frac{1}{4 \cdot n}}.$$

Es verbleibt damit lediglich, $a^* = \sqrt{2} \cdot \pi$ zu zeigen. Das ist wesentlich aufwendiger, und wir verzichten hier darauf.

Q.E.D.

Kapitel 10

Gewöhnliche Differentialgleichungen

10.1 DGLen 1-ter Ordnung

10.1.1 Definitionen

Definition 10.1: (Gewöhnliche Differentialgleichungen)

Eine Gleichung der Form

$$\frac{d}{dx} y(x) = f(x, y(x))$$

heißt (explizite) „gewöhnliche Differentialgleichung erster Ordnung“ für die Funktion $y(x)$, kurz „DGL“.

Beispiel 10.2: Einige Beispiele von DGLen für $y = y(x)$:

$$\frac{dy}{dx} = y, \quad \frac{dy}{dx} = x \cdot y^2, \quad \frac{dy}{dx} = \frac{y - x}{y + x}.$$

Auch

$$x \cdot y \cdot \frac{dy}{dx} = \sin(y),$$

ist eine DGL, wir werden sie aber immer in der nach $\frac{dy}{dx}$ aufgelösten Form

$$\frac{dy}{dx} = \frac{\sin(y)}{x \cdot y}$$

betrachten. Als Kurzform schreibt man auch $y' = \frac{\sin(y)}{x \cdot y}$.

Beispiel 10.3: Wir betrachten die Differentialgleichung

$$\frac{dy}{dx} = \lambda \cdot y$$

mit einer Konstanten λ . Wir raten zunächst eine Lösung:

$$y(x) = c \cdot e^{\lambda \cdot x},$$

wobei c eine beliebige Konstante ist. In der Tat ist dies für jeden Wert von c eine Lösung der DGL:

$$\frac{dy}{dx} = \frac{d}{dx} c \cdot e^{\lambda \cdot x} = c \cdot \frac{d}{dx} e^{\lambda \cdot x} = c \cdot \lambda \cdot e^{\lambda \cdot x} = \lambda \cdot \underbrace{(c \cdot e^{\lambda \cdot x})}_{y(x)} = \lambda \cdot y(x).$$

Wir können die Konstante c durch den „Anfangswert“ $y(0)$ ausdrücken:

$$y(0) = c \cdot e^{\lambda \cdot 0} = c.$$

Also: die Lösung von $\frac{dy}{dx} = \lambda \cdot y$ ist

$$y(x) = y(0) \cdot e^{\lambda \cdot x}.$$

In Beispiel 10.20 zeigen wir, dass dies in der Tat die allgemeinste Lösung der DGL $y' = \lambda \cdot y$ ist.

Merke 10.4:

Die allgemeine Lösung einer DGL erster Ordnung hat immer eine beliebige freie Konstante. Die Lösung mit dieser freien Konstanten heißt „allgemeine Lösung“. Lösungen für spezielle Werte dieser Konstante heißen „spezielle Lösungen“.

Beispiel 10.5: In MuPAD werden DGLen mittels `ode` (engl: `ode` = ordinary differential equation = gewöhnliche DGL) erzeugt, z.B. $y' = y + x$:

```
>> DGL:= ode({y'(x) = x + y(x)}, {y(x)}):
```

Ein solches Objekt kann dann mittels `solve` gelöst werden:

```
>> solve(DGL)
```

```
{exp(x - C1) - x - 1}
```

Man kann Anfangsbedingungen direkt in das `ode`-Objekt einbauen, woraufhin `solve` die freie Konstante so wählt, dass die Anfangsbedingung erfüllt ist:

```
>> DGL:= ode({y'(x) = x + y(x), y(0) = 1}, {y(x)}):
```

```
>> solve(DGL)
```

```
{exp(x + ln(2)) - x - 1}
```

18.12.03↓

10.1.2 Graphische Lösung

Dieses graphische Verfahren ist zwar ungenau (halt graphisch), aber funktioniert dafür für beliebige DGLen $y' = f(x, y)$. Es ist äußerst nützlich, um sich schnell einen *qualitativen* Überblick über das Verhalten der Lösungen zu verschaffen.

Graphisches Rezept 10.6:

Löse eine beliebige DGL $\frac{dy}{dx} = f(x, y)$:

- Zeichne in die x - y -Ebene an vielen Punkten (x, y) die Vektoren des sogenannten „**Richtungsfeldes**“

$$\vec{v}(x, y) = \begin{pmatrix} 1 \\ f(x, y) \end{pmatrix}$$

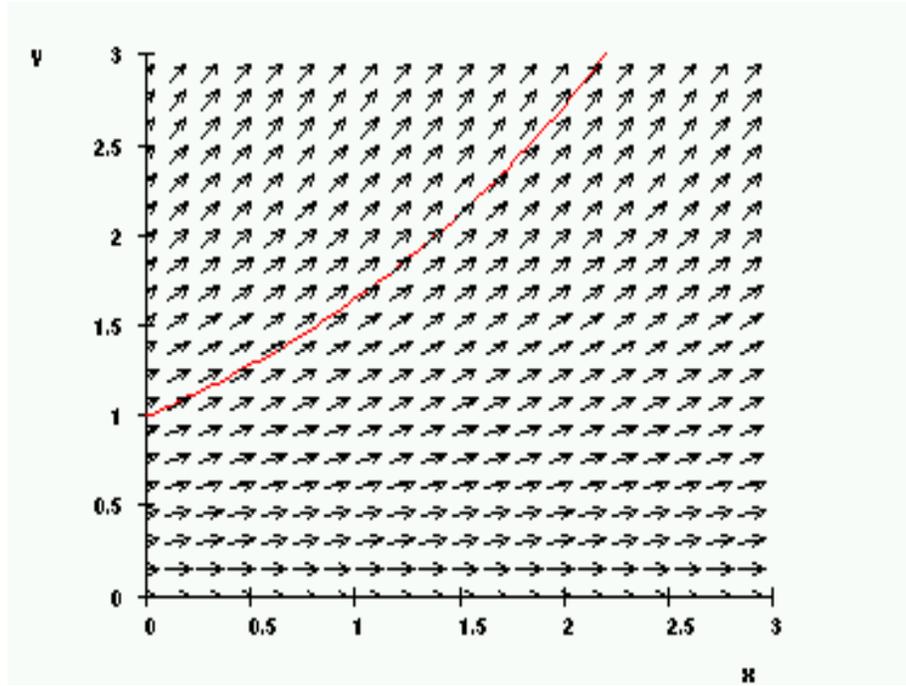
ein. Es interessieren nur die Richtungen der Vektoren, nicht die Längen.

- Wähle einen Startpunkt (x_0, y_0) und folge den Pfeilen! D.h., zeichne eine Kurve, deren Tangente an jedem Punkt mit den vorgegebenen Richtungsvektoren $\vec{v}(x, y)$ übereinstimmt.

Die so vom Startpunkt ausgehende Kurve ist der Graph derjenigen Lösung, für die $y(x_0) = y_0$ gilt.

Beispiel 10.7: Wir zeichnen das Richtungsfeld $\vec{v}(x, y) = \begin{pmatrix} 1 \\ \frac{y}{2} \end{pmatrix}$ der DGL $y' = \frac{y}{2}$. Vom Punkt $(x_0, y_0) = (0, 1)$ startend wird eine Lösungskurve eingezeichnet. Wir benutzen hier, dass wir die Lösung $y(x) = e^{\frac{x}{2}}$ in Beispiel 10.3 (mit $\lambda = \frac{1}{2}$, $y(0) = 1$) bereits berechnet haben:

```
>> plot(
  plot::Function2d(exp(x/2), x = 0..2.9, Color = RGB::Red),
  plot::VectorField2d([1, y/2], x = 0..2.9, y = 0..2.9,
    Color = RGB::Black),
  ViewingBox = [0..3, 0..3]
)
```

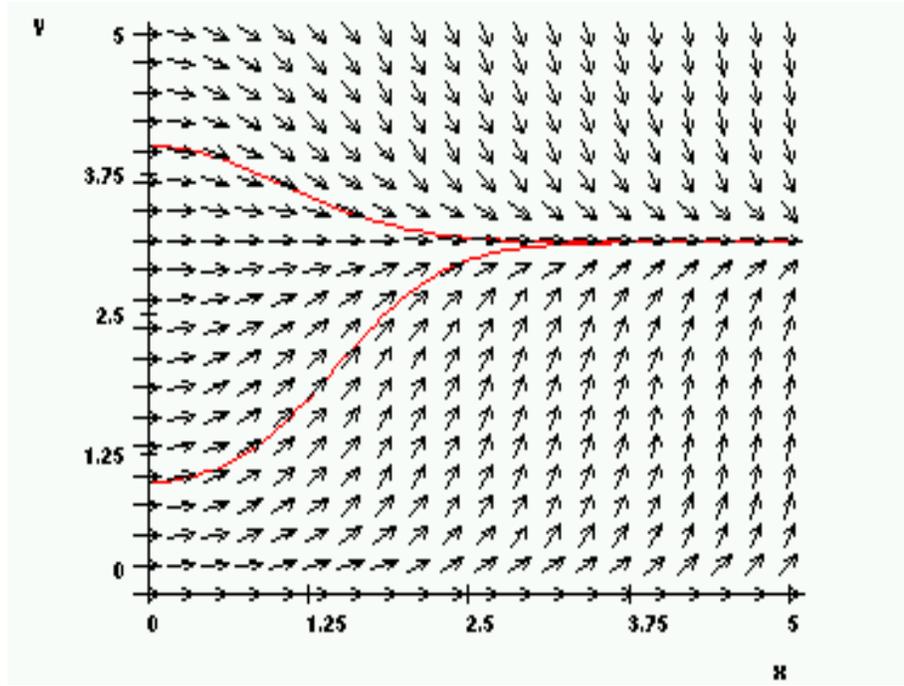


Beispiel 10.8: Ein weiteres Beispiel:

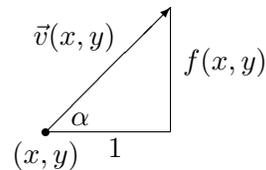
$$\frac{dy}{dx} = x \cdot \sin(y), \quad \vec{v}(x, y) = \begin{pmatrix} 1 \\ x \cdot \sin(y) \end{pmatrix}.$$

Die Lösungen durch die Punkte $(x_0, y_0) = (0, 1)$ und $(x_0, y_0) = (0, 4)$ werden eingezeichnet. Hier wird der numerische DGL-Löser `numeric::odesolve2` zum Einzeichnen der Lösungskurven eingesetzt (engl: ode = ordinary differential equation = gewöhnliche DGL):

```
>> Loesung1:= numeric::odesolve2((x, y) -> [x*sin(y[1])], 0, [1]):
>> Loesung2:= numeric::odesolve2((x, y) -> [x*sin(y[1])], 0, [4]):
>> plot(
  plot::Function2d(Loesung1(x)[1], x = 0..5, Color = RGB::Red),
  plot::Function2d(Loesung2(x)[1], x = 0..5, Color = RGB::Red),
  plot::VectorField2d([1, x*sin(y)], x = 0..5, y = 0..5,
    Color = RGB::Black)
)
```



Bemerkung 10.9: Worauf basiert das Verfahren? Die Tangente der eingezeichneten Kurve ist parallel zum Richtungsfeld $\vec{v}(x, y)$. Die Steigung der Kurventangente (der Tangens des Steigungswinkels α) ist y' , die Steigung des Richtungsvektors ist $\tan(\alpha) = f(x, y)$:



Damit gilt $y' = \tan(\alpha) = f(x, y)$.

Bemerkung 10.10: Die Tatsache, dass man an irgendeinem Punkt (x_0, y_0) starten kann, d.h., in der Lösung eine „Startbedingung“ $y(x_0) = y_0$ vorgeben kann, entspricht der Tatsache, dass die allgemeine Lösung eine beliebige Konstante enthält.

10.1.3 Separation (Trennung der Variablen)

Dies ist eine Technik, die nur für DGLn der speziellen Form $y' = f(x) \cdot g(y)$ funktioniert: die rechte Seite der DGL muss ein Produkt von Funktionen jeweils einer Variablen sein.

Beispiel 10.11: Betrachte die DGL

$$\frac{dy}{dx} = x \cdot y.$$

Die Idee ist, alle Ausdrücke in y (inclusive dy) auf der einen Seite der Gleichung, die Ausdrücke in x (inclusive dx) auf der anderen Seite zu sammeln („Trennung der Variablen“):

$$\frac{1}{y} dy = x \cdot dx.$$

Nun trage auf beiden Seiten ein Integralzeichen ein und bestimme die Stammfunktionen:

$$\int \frac{1}{y} dy = \int x \cdot dx \quad \Rightarrow \quad \ln(|y|) = \frac{x^2}{2} + \tilde{c}$$

(mit einer Integrationskonstanten \tilde{c}). Diese Gleichung definiert y in Abhängigkeit von x . Löse nach y auf:

$$|y| = e^{\frac{x^2}{2} + \tilde{c}} = e^{\tilde{c}} \cdot e^{\frac{x^2}{2}} \quad \Rightarrow \quad y = \pm e^{\tilde{c}} \cdot e^{\frac{x^2}{2}}.$$

Zur Vereinfachung wird die Konstante $\pm e^{\tilde{c}}$ als neue Konstante c geschrieben:

$$y = c \cdot e^{\frac{x^2}{2}}.$$

Dies ist die allgemeine Lösung der DGL $y' = x \cdot y$. Probe:

$$y' = \frac{d}{dx} c \cdot e^{\frac{x^2}{2}} = c \cdot \frac{d}{dx} e^{\frac{x^2}{2}} = c \cdot x \cdot e^{\frac{x^2}{2}} = x \cdot \underbrace{c \cdot e^{\frac{x^2}{2}}}_{y(x)} = x \cdot y(x).$$

Bemerkung 10.12: Diese Technik kann nur dann funktionieren, wenn man es schafft, die Variablen x und y auf verschiedene Seiten der Gleichung zu bringen. Für die allgemeine DGL $y' = h(x, y)$ ist dies genau dann der Fall, wenn die Funktion $h(x, y)$ das Produkt einer Funktion in x und einer Funktion in y ist:

$$\begin{aligned} \frac{dy}{dx} = f(x) \cdot g(y) &\quad \Rightarrow \quad \frac{1}{g(y)} dy = f(x) dx \\ \Rightarrow \int \frac{1}{g(y)} dy = \int f(x) \cdot dx &\quad \Rightarrow \quad G(y) = F(x) + c, \end{aligned}$$

wobei $G(y)$ eine Stammfunktion von $\frac{1}{g(y)}$ und $F(x)$ eine Stammfunktion von $f(x)$ ist.

Diese Separationstechnik läßt sich folgendermaßen zusammenfassen:

Satz 10.13: (Separation der Variablen)

Sei $G(y)$ eine Stammfunktion von $\frac{1}{g(y)}$, sei $F(x)$ eine Stammfunktion von $f(x)$. Die allgemeine Lösung der DGL

$$\frac{dy}{dx} = f(x) \cdot g(y)$$

erfüllt die Gleichung

$$G(y) = F(x) + c.$$

Durch Auflösen dieser Gleichung nach y ergibt sich die explizite Form der Lösung

$$y(x) = G^{-1}(F(x) + c)$$

mit der Umkehrfunktion G^{-1} von G .

„**Beweis**“: Sei $y(x)$ eine Funktion, mit der für alle x die Gleichung

$$G(y(x)) = F(x) + c$$

mit einer Konstanten c erfüllt ist. Durch Ableiten nach x erhält man

$$G'(y(x)) \cdot y'(x) = F'(x) \Rightarrow \frac{1}{g(y(x))} \cdot y'(x) = f(x) \Rightarrow y'(x) = f(x) \cdot g(y).$$

Wir haben also eine Lösung mit einer freien Konstante c vor uns.

Q.E.D.

Beispiel 10.14: Betrachte die DGL $y' = y^2$. Separation:

$$\begin{aligned} \frac{dy}{dx} = y^2 &\Rightarrow \frac{1}{y^2} dy = 1 \cdot dx \Rightarrow \int \frac{1}{y^2} dy = \int 1 \cdot dx \\ &\Rightarrow -\frac{1}{y} = x + c \Rightarrow y = \frac{1}{-x - c}. \end{aligned}$$

Damit ist die allgemeine Lösung von $y' = y^2$ bestimmt. Es ist schöner, die beliebige Konstante c durch eine Anfangsbedingung $y(x_0)$ auszudrücken. Z.B., für $x_0 = 0$:

$$y(0) = \frac{1}{-0 - c} = -\frac{1}{c} \Rightarrow c = -\frac{1}{y(0)}.$$

Damit ergibt sich die Lösung zur Anfangsbedingung $y(0)$ in der folgenden Form:

$$y(x) = \frac{1}{-x + \frac{1}{y(0)}} = \frac{y(0)}{1 - x \cdot y(0)}.$$

Probe:

$$\frac{dy}{dx} = \frac{d}{dx} \frac{y(0)}{1 - x \cdot y(0)} = \frac{y(0)^2}{(1 - x \cdot y(0))^2} = y(x)^2.$$

Weiterhin ergibt sich für $x = 0$ in der Tat:

$$y(x)|_{x=0} = \frac{y(0)}{1 - 0 \cdot y(0)} = y(0).$$

Beispiel 10.15: Betrachte eine **lineare** DGL

$$\frac{dy}{dx} = f(x) \cdot y.$$

„Linear“ heißt, dass mit irgendeiner speziellen Lösung $y_s(x)$ auch jedes Vielfache $y(x) = \tilde{c} \cdot y_s(x)$ mit einer beliebigen Konstanten \tilde{c} eine Lösung ist:

$$\frac{dy}{dx} = \frac{d}{dx} \tilde{c} \cdot y_s(x) = \tilde{c} \cdot \frac{d}{dx} y_s(x) = \tilde{c} \cdot f(x) \cdot y_s(x) = f(x) \cdot \underbrace{\tilde{c} \cdot y_s(x)}_{y(x)} = f(x) \cdot y(x).$$

Damit ist klar: die freie Konstante muss multiplikativ eingehen. Man kann die allgemeine Lösung per Separation immer durch eine einzige Integration ermitteln:

$$\frac{1}{y} dy = f(x) dx \quad \Rightarrow \quad \ln(|y|) = \int f(x) dx \quad \Rightarrow \quad y = \pm e^{\int f(x) dx}.$$

Wo ist nun die multiplikative Konstante? Sei $F(x)$ irgendeine konkrete Stammfunktion von $f(x)$ (ohne Integrationskonstante), so gilt $\int f(x) dx = F(x) + \tilde{c}$ mit einer beliebigen Integrationskonstanten \tilde{c} , also

$$y = \pm e^{F(x) + \tilde{c}} = \underbrace{\pm e^{\tilde{c}}}_c \cdot e^{F(x)} = c \cdot e^{F(x)},$$

wobei c wiederum eine beliebige Konstante ist. Vergleiche auch das spezielle Beispiel 10.11.

Merke: Die Lösung von $y' = f(x) \cdot y$ ist

$$y(x) = c \cdot e^{F(x)},$$

wo $F(x)$ eine Stammfunktion von $f(x)$ ist.

Beispiel 10.16: Man kann die Startbedingung auch direkt in den Integrationsschritt der Separation einbauen. Dazu folgendes Beispiel:

$$\frac{dy}{dx} = \cos(x) \cdot e^y, \quad y(x_0) = y_0.$$

Separation:

$$\frac{1}{e^y} dy = \cos(x) dx \quad \Rightarrow \quad e^{-y} dy = \cos(x) dx.$$

Rezept:

Anstatt links und rechts allgemeine Stammfunktionen mit einer beliebigen Integrationskonstanten zu bestimmen, werden **bestimmte** Integrale startend bei y_0 bzw. x_0 berechnet.

Wir integrieren bis Y bzw. X , wobei Y die Lösung $Y = y(X)$ für $x = X$ darstellt. In diesem Beispiel:

$$\begin{aligned} \int_{y_0}^Y e^{-y} dy &= \int_{x_0}^X \cos(x) dx \\ \Rightarrow [-e^{-y}]_{y=y_0}^{y=Y} &= [\sin(x)]_{x=x_0}^{x=X} \\ \Rightarrow -e^{-Y} + e^{-y_0} &= \sin(X) - \sin(x_0). \end{aligned}$$

Durch Auflösen nach Y erhält man die Lösung $Y = y(X)$, welche für $X = x_0$ den Wert y_0 annimmt:

$$Y(X) = \ln(e^{-y_0} + \sin(x_0) - \sin(X)).$$

Mit der Umbenennung $X \rightarrow x$, $Y \rightarrow y$ erhält man die Lösung

$$y(x) = -\ln(e^{-y_0} + \sin(x_0) - \sin(x)).$$

Probe:

$$\begin{aligned} \frac{dy}{dx} &= \frac{\cos(x)}{e^{y_0} + \sin(x_0) - \sin(x)} = \frac{\cos(x)}{e^{\ln(e^{y_0} + \sin(x_0) - \sin(x))}} = \frac{\cos(x)}{e^{-y(x)}} = \cos(x) \cdot e^{y(x)}, \\ y(x_0) &= -\ln(e^{-y_0} + \sin(x_0) - \sin(x_0)) = -\ln(e^{-y_0}) = y_0. \end{aligned}$$

10.1.4 Variation der Konstanten

↓6.1.04

In diesem Abschnitt wird die Lösungstechnik für DGLen der Form

$$\frac{dy}{dx} = f(x) \cdot y + g(x) \quad (1)$$

vorgelegt. Die DGL

$$\frac{dy}{dx} = f(x) \cdot y \quad (2)$$

ohne den Term $g(x)$ heißt „**homogene DGL**“, mit dem Term $g(x)$ heißt die DGL „**inhomogen**“. Die Lösung $y_h(x)$ der homogenen DGL (2) haben wir in Beispiel 10.15 kennengelernt:

$$y_h(x) = c \cdot e^{F(x)},$$

wobei $F(x)$ eine Stammfunktion von $f(x)$ ist.

Rezept „Variation der Konstanten“ 10.17:

Zur Lösung der inhomogenen DGL (1) ersetze die freie Konstante c in der Lösung $y_h(x)$ der homogenen DGL (2) durch eine Funktion $c(x)$. Einsetzen dieses „Ansatzes“ in die DGL (1) liefert eine DGL für $c(x)$, die immer durch einfache Integration gelöst werden kann.

Ansatz für die Lösung der inhomogenen DGL (1):

$$y(x) = c(x) \cdot e^{F(x)}.$$

Einsetzen in die inhomogene DGL (2). Die linke Seite:

$$\frac{dy}{dx} = \frac{d}{dx} \left(c(x) \cdot e^{F(x)} \right) = \boxed{\frac{dc(x)}{dx} \cdot e^{F(x)}} + c(x) \cdot F'(x) \cdot e^{F(x)}.$$

Die rechte Seite:

$$f(x) \cdot y(x) + g(x) = f(x) \cdot c(x) \cdot e^{F(x)} + \boxed{g(x)}.$$

Mit $F'(x) = f(x)$ heben sich beim Gleichsetzen der linken und rechten Seite zwei Terme weg, und es verbleibt folgende DGL für $c(x)$:

$$\frac{dc}{dx} \cdot e^{F(x)} = g(x), \quad \text{also} \quad \frac{dc}{dx} = g(x) \cdot e^{-F(x)}.$$

Mit anderen Worten, $c(x)$ ist eine Stammfunktion von $g(x) \cdot e^{-F(x)}$:

$$c(x) = \int g(x) \cdot e^{-F(x)} dx.$$

Damit haben wir eine allgemeine Darstellung der Lösung $y(x) = c(x) \cdot e^{F(x)}$ der inhomogenen DGL (1):

Satz 10.18: (Lösungsformel)

Die allgemeine Lösung der DGL $\frac{dy}{dx} = f(x) \cdot y + g(x)$ ist

$$\boxed{y(x) = e^{F(x)} \cdot \int g(x) \cdot e^{-F(x)} dx,}$$

wobei $F(x)$ eine Stammfunktion von $f(x)$ ist.

Diese Lösungsformel wird man sich nicht merken können. Man merke sich daher lieber das Konstruktionsprinzip 10.17 der „Variation der Konstanten“.

Beispiel 10.19: Finde die allgemeine Lösung der DGL

$$\frac{dy}{dx} = y + x.$$

Schritt 1: Die Lösung der homogenen DGL $y' = y$ ist nach Beispiel 10.15 (mit $f(x) = 1$, $F(x) = x$):

$$y_h(x) = c \cdot e^x.$$

Schritt 2: Variationsansatz $y(x) = c(x) \cdot e^x$ für die inhomogene DGL $y' = y + x$.

Schritt 3: Einsetzen in die inhomogene DGL. Die linke Seite der DGL:

$$\frac{dy}{dx} = \frac{d}{dx} (c(x) \cdot e^x) = \frac{dc}{dx} \cdot e^x + c(x) \cdot e^x.$$

Die rechte Seite der DGL:

$$y + x = c(x) \cdot e^x + x.$$

Durch Vergleich folgt

$$\frac{dc}{dx} \cdot e^x = x \quad \Rightarrow \quad \frac{dc}{dx} = x \cdot e^{-x}.$$

Schritt 4: Bestimme $c(x)$ durch Integration:

$$c(x) = \int x \cdot e^{-x} dx.$$

Dieses Integral wird analog zu Beispiel 9.12 durch partielle Integration bestimmt:

$$\begin{aligned} c(x) &= \int \underbrace{x}_{f(x)} \cdot \underbrace{e^{-x}}_{g'(x)} dx = \underbrace{x}_{f(x)} \cdot \underbrace{(-e^{-x})}_{g(x)} - \int \underbrace{1}_{f'(x)} \cdot \underbrace{(-e^{-x})}_{g(x)} dx \\ &= -x \cdot e^{-x} + \int e^{-x} dx = -x \cdot e^{-x} - e^{-x} + \tilde{c} = -(x+1) \cdot e^{-x} + \tilde{c} \end{aligned}$$

mit einer beliebigen Integrationskonstanten \tilde{c} .

Schritt 5: Das Endergebnis (die allgemeine Lösung von $y' = y + x$) ist

$$\begin{aligned} y(x) &= c(x) \cdot e^x = \left(-(x+1) \cdot e^{-x} + \tilde{c} \right) \cdot e^x = -(x+1) \cdot \underbrace{e^{-x} \cdot e^x}_1 + \tilde{c} \cdot e^x \\ &= -x - 1 + \tilde{c} \cdot e^x. \end{aligned}$$

Schritt 6: Probe:

$$\frac{dy}{dx} - (y + x) = \underbrace{-1 + \tilde{c} \cdot e^x}_{y'} - \left(\underbrace{-x - 1 + \tilde{c} \cdot e^x}_y + x \right) = 0. \quad (\text{OK})$$

Alternativ: Wir wenden die Lösungsformel 10.18 direkt an. Es wird betrachtet $y' = y + x$, d.h., $f(x) = 1$ und $g(x) = x$. Mit der Stammfunktion $F(x) = x$ von $f(x) = 1$ ergibt sich sofort

$$y(x) = \underbrace{e^x}_{e^{F(x)}} \cdot \int \underbrace{x}_{g(x)} \cdot \underbrace{e^{-x}}_{e^{-F(x)}} dx.$$

Damit landen wir unmittelbar bei dem oben bereits gelösten Integrationsproblem

$$\int x \cdot e^{-x} dx = -(x+1) \cdot e^{-x} + \tilde{c},$$

und es ergibt sich wiederum die Lösung

$$y(x) = e^x \cdot \left(-(x+1) \cdot e^{-x} + \tilde{c} \right) = -x - 1 + \tilde{c} \cdot e^x.$$

Beispiel 10.20: Wir zeigen, dass die Lösung

$$y(x) = c \cdot e^{\lambda \cdot x}$$

in der Tat die *allgemeinste* Lösung der DGL

$$\frac{dy}{dx} = \lambda \cdot y$$

ist. Dazu wenden wir das Prinzip der Variation der Konstanten auf diese homogene DGL an, d.h., wir versuchen, die allgemeinste Lösung in der Form $y(x) = c(x) \cdot e^{\lambda \cdot x}$ zu bestimmen. Einsetzen dieses Ansatzes in die DGL liefert auf der linken Seite:

$$\frac{dy}{dx} = \frac{dc}{dx} \cdot e^{\lambda \cdot x} + c(x) \cdot \lambda \cdot e^{\lambda \cdot x}.$$

Die rechte Seite ist

$$\lambda \cdot y(x) = \lambda \cdot c(x) \cdot e^{\lambda \cdot x}.$$

Durch Vergleich folgt

$$\frac{dc}{dx} = 0.$$

Die allgemeinste Lösung dieser „DGL“ für $c(x)$ ist offensichtlich eine konstante Funktion:

$$c(x) = \int 0 dx = 0 + \tilde{c} = \text{eine beliebige Konstante.}$$

10.2 Lineare DGLen höherer Ordnung

Wir betrachten hier speziell **lineare DGLen** n -ter Ordnung

$$a_n(x) \cdot y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_1(x) \cdot y'(x) + a_0(x) \cdot y(x) = f(x)$$

mit gegebener „**Inhomogenität**“ $f(x)$. Wir können o.B.d.A. durch $a_n(x)$ teilen und die Koeffizientenfunktionen $a_i(x)$ und die Inhomogenität $f(x)$ umdefinieren, so dass wir nur die „normierte“ Form

$$y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_1(x) \cdot y'(x) + a_0(x) \cdot y(x) = f(x)$$

zu betrachten brauchen. Als Anfangsbedingungen können wir die n Werte

$$y(x_0), y'(x_0), \dots, y^{(n-1)}(x_0)$$

an einem Punkt x_0 vorgeben (die höheren Ableitungen an dieser Stelle liegen dann wegen $y^{(n)}(x_0) = f(x_0) - a_{n-1}(x_0) \cdot y^{(n-1)}(x_0) - \dots - a_0(x_0) \cdot y(x_0)$ fest).

Die allgemeine Lösung einer DGL n -ter Ordnung besitzt n freie Konstanten.

Für lineare DGLen gilt:

Satz 10.21: (Superpositionsprinzip)

Seien $y_1(x)$ und $y_2(x)$ Lösungen der homogenen linearen DGL

$$y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_0(x) \cdot y(x) = 0.$$

Dann ist $y(x) = c_1 \cdot y_1(x) + c_2 \cdot y_2(x)$ mit beliebigen Konstanten c_1, c_2 wieder eine Lösung.

Beweis: trivial.

So trivial das obige Superpositionsprinzip mathematisch auch sein mag, es ist ein fundamentales Grundprinzip beim Suchen von Lösungen. Es besagt, dass man sich zunächst auf einzelne spezielle Lösungen konzentrieren darf, die man zuletzt per Linearkombination zur Gesamtlösung zusammensetzen kann.

Beispiel 10.22: Wir betrachten die homogene lineare DGL zweiter Ordnung

$$x^2 \cdot y''(x) - 2 \cdot x \cdot y'(x) + 2 \cdot y(x) = 0.$$

Zum Auffinden von Lösungen machen wir den Ansatz $y(x) = x^\lambda$. Einsetzen in die DGL liefert

$$\begin{aligned} & x^2 \cdot \underbrace{\lambda \cdot (\lambda - 1) \cdot x^{\lambda-2}}_{y''(x)} - 2 \cdot x \cdot \underbrace{\lambda \cdot x^{\lambda-1}}_{y'(x)} + 2 \cdot \underbrace{x^\lambda}_{y(x)} \\ &= (\lambda^2 - 3 \cdot \lambda + 2) \cdot x^\lambda = (\lambda - 1) \cdot (\lambda - 2) \cdot x^\lambda = 0. \end{aligned}$$

Offensichtlich ist diese Gleichung genau dann für alle x erfüllt, wenn $\lambda = 1$ oder $\lambda = 2$ gilt. Damit haben wir zwei „Grundlösungen“

$$y_1(x) = x, \quad y_2(x) = x^2$$

gefunden. Das Superpositionsprinzip besagt, dass

$$y(x) = c_1 \cdot y_1(x) + c_2 \cdot y_2(x) = c_1 \cdot x + c_2 \cdot x^2$$

mit beliebigen Konstanten c_1, c_2 wieder eine Lösung ist. Wir haben damit eine Lösung mit 2 freien Konstanten gefunden. Der folgende Satz 10.26 wird uns garantieren, dass dies die allgemeine Lösung ist.

Definition 10.23:

Ein Satz von Funktionen $y_1(x), \dots, y_n(x)$ heißt **linear unabhängig**, wenn die Linearkombination

$$f(x) = c_1 \cdot y_1(x) + \dots + c_n \cdot y_n(x)$$

nur für die Konstanten $c_1 = 0, \dots, c_n = 0$ identisch 0 ist.

Beispiel 10.24: Die Monome $y_0(x) = 1, y_1(x) = x, \dots, y_n(x) = x^n$ sind linear unabhängig, denn die Polynomgleichung

$$c_0 + c_1 \cdot x + \dots + c_n \cdot x^n = 0$$

hat nur endlich viele Lösungen für x , kann also nicht für alle x erfüllt sein, wenn nicht alle Konstanten c_0, \dots, c_n verschwinden.

Beispiel 10.25: Die Funktionen $y_1(x) = \sin(x)^2, y_2(x) = \cos(x)^2$ sind unabhängig. Sei

$$c_1 \cdot \sin(x)^2 + c_2 \cdot \cos(x)^2 = 0 \quad \forall x.$$

Für $x = 0$ folgt $c_2 = 0$, für $x = \pi/2$ folgt $c_1 = 0$.

Die Funktionen $y_1(x) = \sin(x)^2, y_2(x) = \cos(x)^2, y_3(x) = 1$ hingegen sind abhängig. Es gilt speziell

$$y_1(x) + y_2(x) - y_3(x) = \sin(x)^2 + \cos(x)^2 - 1 = 0 \quad \forall x.$$

Der folgende Satz besagt, dass es bei homogenen linearen DGLen n -ter Ordnung ausreicht, n spezielle linear unabhängige Lösungen zu finden. Diese lassen sich dann per Superposition zur allgemeinen Lösung zusammensetzen:

Satz 10.26: (Fundamentalsystem: die allgemeine Lösung)

Seien $y_1(x), \dots, y_n(x)$ linear unabhängige Lösungen der homogenen DGL n -ter Ordnung

$$y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_0(x) \cdot y(x) = 0.$$

Dann ist die allgemeine Lösung

$$y(x) = c_1 \cdot y_1(x) + \dots + c_n \cdot y_n(x)$$

mit beliebigen Konstanten c_1, \dots, c_n . Die Lösungen $y_1(x), \dots, y_n(x)$ werden ein „**Fundamentalsystem**“ der DGL genannt.

Beweis: (nur verdaulich, wenn man schon Lineare Algebra kann)

Es ist zu zeigen, dass ein Fundamentalsystem y_1, \dots, y_n die allgemeine homogene Lösung aufspannt. Wir haben hier noch nicht den technischen Apparat (Determinanten), daher ist das folgende nur für diejenigen verdaulich, die das Konzept von Determinanten aus der Schule mitbringen. Die Beobachtung ist, dass Funktionen f_1, \dots, f_m genau dann linear unabhängig sind, wenn die sogenannte **Wronski-Determinante** verschwindet:

$$W(f_1, \dots, f_m) = \det \begin{pmatrix} f_1 & f_2 & \dots & f_m \\ f_1' & f_2' & \dots & f_m' \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(m-1)} & f_2^{(m-1)} & \dots & f_m^{(m-1)} \end{pmatrix} = 0.$$

Dies liegt daran, dass

$$c_1 \cdot f_1(x) + \dots + c_m \cdot f_m(x) = 0 \quad \forall x$$

die linearen Gleichungen

$$\begin{array}{ccccccc} c_1 \cdot f_1(x) & + & \dots & + & c_m \cdot f_m(x) & = & 0, \\ c_1 \cdot f_1'(x) & + & \dots & + & c_m \cdot f_m'(x) & = & 0, \\ \vdots & & \ddots & & \ddots & & \vdots \\ c_1 \cdot f_1^{(m-1)}(x) & + & \dots & + & c_m \cdot f_m^{(m-1)}(x) & = & 0 \end{array}$$

impliziert. In Matrixschreibweise:

$$\begin{pmatrix} f_1 & f_2 & \dots & f_m \\ f_1' & f_2' & \dots & f_m' \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(m-1)} & f_2^{(m-1)} & \dots & f_m^{(m-1)} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Es gibt nur dann einen nicht-trivialen Lösungsvektor (c_1, \dots, c_m) , wenn die Determinante der Koeffizientenmatrix verschwindet (dies ist die Wronski-Determinante von f_1, \dots, f_m).

Sei nun y_1, \dots, y_n ein Fundamentalsystem der homogenen DGL, sei y eine weitere Lösung. Dann verschwindet die Wronski-Determinante der $n+1$ Funktionen y_1, \dots, y_n, y :

$$W(y_1, \dots, y_n, y) = \det \begin{pmatrix} y_1 & y_2 & \cdots & y_n & y \\ y_1' & y_2' & \cdots & y_n' & y' \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_1^{(n)} & y_2^{(n)} & \cdots & y_n^{(n)} & y^{(n)} \end{pmatrix} = 0.$$

Um dies zu sehen, brauchen nur die in der n -ten Zeile stehenden n -ten Ableitungen über die DGL als Linearkombination der ersten $n-1$ Zeilen geschrieben zu werden. Die Determinante einer Matrix mit linear abhängigen Zeilen ist 0. Also sind die Funktionen y_1, \dots, y_n, y linear abhängig. Da die ersten $n-1$ Spalten als linear unabhängig vorausgesetzt sind (Fundamentalsystem), läßt sich die letzte Spalte als Linearkombination der ersten $n-1$ Spalten schreiben, also

$$y = c_1 \cdot y_1 + \dots + c_n \cdot y_n.$$

Q.E.D.

Satz 10.27: (Die allgemeine Lösung linearer DGLen mit Inhomogenität)

Die allgemeine Lösung der inhomogenen DGL n -ter Ordnung

$$y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_0(x) \cdot y(x) = f(x)$$

ist

$$y(x) = y_p(x) + c_1 \cdot y_1(x) + \dots + c_n \cdot y_n(x),$$

wobei $y_1(x), \dots, y_n(x)$ ein Fundamentalsystem der homogenen DGL

$$y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_0(x) \cdot y(x) = 0$$

ist und $y_p(x)$ irgendeine spezielle Lösung der inhomogenen DGL ist ($y_p(x)$ wird „partikuläre (spezielle) Lösung“ genannt).

Merke: Allgemeine inhomogene Lösung = spezielle inhomogene Lösung + allgemeine homogene Lösung.

Beweis: Zur Abkürzung führen wir ein

$$L(y) = y^{(n)} + a_{n-1} \cdot y^{(n-1)} + \dots + a_0 \cdot y.$$

Haben wir irgendeine Lösung y_p von $L(y) = f$, so erfüllt $y = y_p + u$ wegen $L(y_p + u) = L(y_p) + L(u)$ offensichtlich genau dann $L(y) = f$, wenn $L(u) = 0$ gilt. Damit besteht die allgemeine inhomogene Lösung y aus einer partikulären Lösung y_p plus der allgemeinen homogenen Lösung u .

Q.E.D.

Bemerkung 10.28: Hier noch eine weitere nützliche allgemeine Beobachtung für reelle homogene lineare DGLen: Sind die Koeffizientenfunktionen $a_i(x)$ in

$$y^{(n)}(x) + a_{n-1}(x) \cdot y^{(n-1)}(x) + \dots + a_0(x) \cdot y(x) = 0$$

reell, so gilt:

Mit jeder komplexen Lösung $y(x)$ sind auch der Realteil $\Re(y(x))$ und der Imaginärteil $\Im(y(x))$ Lösungen.

(Für jeden Summanden der DGL gilt $\Re(a_i(x) \cdot y(x)) = a_i(x) \cdot \Re(y(x))$ bzw. $\Im(a_i(x) \cdot y(x)) = a_i(x) \cdot \Im(y(x))$.) Damit können wir uns komplexe Lösungen verschaffen (was technisch oft einfacher ist als auf reellen Lösungen zu bestehen) und dann daraus per \Re und \Im reelle Lösungen extrahieren.

Beispiel 10.29: (Der gedämpfte harmonische Oszillator) Die lineare DGL

$$y''(x) + 2 \cdot \delta \cdot y'(x) + \omega_0^2 \cdot y(x) = 0$$

mit der reellen „Dämpfungskonstante“ $\delta > 0$ und der reellen „Grundfrequenz“ $\omega_0 > 0$ soll durch den Ansatz $y(x) = e^{\lambda \cdot x}$ gelöst werden, wobei λ geeignet zu bestimmen ist. Einsetzen des Ansatzes in die DGL liefert die Gleichung

$$\underbrace{\lambda^2 \cdot e^{\lambda \cdot x}}_{y''(x)} + 2 \cdot \delta \cdot \underbrace{\lambda \cdot e^{\lambda \cdot x}}_{y'(x)} + \omega_0^2 \cdot \underbrace{e^{\lambda \cdot x}}_{y(x)} = \underbrace{(\lambda^2 + 2 \cdot \delta \cdot \lambda + \omega_0^2)}_{=0} \cdot e^{\lambda \cdot x} = 0.$$

Dies führt zu

$$\lambda = -\delta \pm \sqrt{\delta^2 - \omega_0^2}.$$

Für kleine Dämpfungswerte $|\delta| < \omega_0$ schreiben wir intuitiver

$$\lambda = -\delta \pm i \cdot \sqrt{\omega_0^2 - \delta^2}.$$

Wir erhalten als komplexe Fundamentallösung der DGL die gedämpften Schwingungen

$$y_{\pm}(x) = e^{-\delta \cdot x} \cdot e^{\pm i \cdot \omega \cdot x}$$

mit

$$\omega = \sqrt{\omega_0^2 - \delta^2},$$

also als allgemeine Lösung der DGL:

$$y(x) = e^{-\delta \cdot x} \cdot \left(c_+ \cdot e^{i \cdot \omega \cdot x} + c_- \cdot e^{-i \cdot \omega \cdot x} \right).$$

Reelle Lösungen ergeben sich durch Real- und Imaginärteil. Nimmt man c_{\pm} als reell an:

$$\begin{aligned} \Re(y(x)) &= e^{-\delta \cdot x} \cdot \overbrace{(c_+ + c_-)}^{c_1} \cdot \cos(\omega \cdot x), \\ \Im(y(x)) &= e^{-\delta \cdot x} \cdot \overbrace{(c_+ - c_-)}^{c_2} \cdot \sin(\omega \cdot x) \end{aligned}$$

mit neuen beliebigen Konstanten c_1, c_2 .

Bemerkung 10.30: Wir gehen hier nicht weiter auf lineare DGLen höherer Ordnung mit beliebigen Koeffizientenfunktion $a_0(x), \dots, a_n(x)$ ein, da sich –abhängig von den $a_i(x)$ – die Lösung meist nur durch spezielle Funktionen (Bessel-Funktionen, Airy-Funktionen, hypergeometrische Funktionen etc., etc., etc.) ausdrücken lassen.

Eine allgemeine Lösungstheorie gibt es nur für den Fall, dass die Koeffizientenfunktionen konstant sind. Dieser Fall wird im folgenden Abschnitt diskutiert.

10.3 Lineare DGLen mit konstanten Koeffizienten

Wir betrachten nun die spezielle Klasse der linearen DGLen

$$y^{(n)}(x) + a_{n-1} \cdot y^{(n-1)}(x) + \dots + a_1 \cdot y'(x) + a_0 \cdot y(x) = f(x),$$

in denen die Koeffizienten a_{n-1}, \dots, a_0 Konstanten (unabhängig von x) sind.

Definition 10.31: (das charakteristische Polynom einer linearen DGL)

Der homogenen linearen DGL mit konstanten Koeffizienten

$$y^{(n)}(x) + a_{n-1} \cdot y^{(n-1)}(x) + \dots + a_1 \cdot y'(x) + a_0 \cdot y(x) = 0$$

wird das „**charakteristische Polynom**“

$$P(\lambda) = \lambda^n + a_{n-1} \cdot \lambda^{n-1} + \dots + a_1 \cdot \lambda + a_0$$

zugeordnet.

Machen wir den Ansatz $y(x) = e^{\lambda \cdot x}$, so ergibt sich mit

$$y = e^{\lambda \cdot x}, \quad y' = \lambda \cdot e^{\lambda \cdot x} = \lambda \cdot y, \quad y'' = \lambda^2 \cdot e^{\lambda \cdot x} = \lambda^2 \cdot y \text{ etc.}$$

aus der DGL das charakteristische Polynom:

$$\begin{aligned} & y^{(n)} + a_{n-1} \cdot y^{(n-1)} + \dots + a_1 \cdot y' + a_0 \cdot y \\ &= \lambda^n \cdot y + a_{n-1} \cdot \lambda^{n-1} \cdot y + \dots + a_1 \cdot \lambda \cdot y + a_0 \cdot y \\ &= \left(\lambda^n + a_{n-1} \cdot \lambda^{n-1} + \dots + a_1 \cdot \lambda + a_0 \right) \cdot y = P(\lambda) \cdot y = 0. \end{aligned}$$

Ergebnis:

$$y(x) = e^{\lambda \cdot x} \text{ löst die DGL genau dann, wenn } \lambda \text{ Nullstelle des charakteristischen Polynoms ist.}$$

Damit haben wir linear unabhängige Grundlösungen $e^{\lambda_1 \cdot x}$, $e^{\lambda_2 \cdot x}$ etc., wo λ_i die unterschiedlichen Nullstellen des charakteristischen Polynoms sind. Um ein Fundamentalsystem zu erhalten, brauchen wir n unabhängige Funktionen, wo n der Grad der DGL ist. Sind einige der Nullstellen λ_i entartet, fehlen also noch einige Grundlösungen. Diese sind

$$x^j \cdot e^{\lambda_i \cdot x},$$

wo j von 0 bis $n_i - 1$ laufen darf (n_i ist die Vielfachheit von λ_i):

Satz 10.32: (Die Lösung homogener linearer DGLn mit konstanten Koeffizienten)

Seien $\lambda_1, \dots, \lambda_k$ die Nullstellen mit Vielfachheiten n_1, \dots, n_k des charakteristischen Polynoms

$$P(\lambda) = \lambda^n + a_{n-1} \cdot \lambda^{n-1} + \dots + a_1 \cdot \lambda + a_0$$

der homogenen linearen DGL mit konstanten Koeffizienten

$$y^{(n)}(x) + a_{n-1} \cdot y^{(n-1)}(x) + \dots + a_1 \cdot y'(x) + a_0 \cdot y(x) = 0. \quad (\#)$$

Dann bilden die $n_1 + \dots + n_k = n$ Funktionen

$$y_{ij}(x) = x^j \cdot e^{\lambda_i \cdot x}, \quad i = 1, \dots, k, \quad j = 0, \dots, n_i - 1$$

ein Fundamentalsystem für (#), d.h., die allgemeine Lösung ist

$$y(x) = \sum_{i=1}^k \sum_{j=0}^{n_i-1} c_{ij} \cdot x^j \cdot e^{\lambda_i \cdot x}$$

mit beliebigen Koeffizienten c_{ij} .

Beweis: Wir ersetzen im Polynom $P(\lambda)$ den Wert λ durch den Operator $\frac{d}{dx}$:

$$P\left(\frac{d}{dx}\right) = \frac{d^n}{dx^n} + a_{n-1} \cdot \frac{d^{n-1}}{dx^{n-1}} + \dots + a_1 \cdot \frac{d}{dx} + a_0,$$

so dass die homogene DGL durch $P\left(\frac{d}{dx}\right) y = 0$ beschrieben wird. Mit der Produktregel

$$\frac{d^k}{dx^k} f \cdot g = \sum_{m=0}^k \binom{k}{m} \cdot \frac{d^m f}{dx^m} \cdot \frac{d^{k-m} g}{dx^{k-m}}$$

ergibt sich speziell für $f \cdot g = x \cdot g$:

$$\frac{d^k}{dx^k} x \cdot g = \sum_{m=0}^k \binom{k}{m} \cdot \frac{d^m x}{dx^m} \cdot \frac{d^{k-m} g}{dx^{k-m}} = \sum_{m=0}^1 \binom{k}{m} \cdot \frac{d^m x}{dx^m} \cdot \frac{d^{k-m} g}{dx^{k-m}} = x \cdot g^{(k)} + k \cdot g^{(k-1)}.$$

Es folgt

$$\begin{aligned} P\left(\frac{d}{dx}\right) x \cdot g &= \left(\sum_{k=0}^n a_k \cdot \frac{d^k}{dx^k} \right) x \cdot g = x \cdot \underbrace{\sum_{k=0}^n a_k \cdot g^{(k)}}_{P\left(\frac{d}{dx}\right) g} + \underbrace{\sum_{k=0}^n a_k \cdot k \cdot g^{(k-1)}}_{P'\left(\frac{d}{dx}\right) g} \\ &= x \cdot P\left(\frac{d}{dx}\right) g + P'\left(\frac{d}{dx}\right) g. \end{aligned}$$

Induktiv folgt

$$\begin{aligned} P\left(\frac{d}{dx}\right) x^j \cdot g &= P\left(\frac{d}{dx}\right) x \cdot (x^{j-1} \cdot g) \\ &= x \cdot P\left(\frac{d}{dx}\right) x^{j-1} \cdot g + P'\left(\frac{d}{dx}\right) x^{j-1} \cdot g \\ &= x^2 \cdot P\left(\frac{d}{dx}\right) x^{j-2} \cdot g + 2 \cdot x \cdot P'\left(\frac{d}{dx}\right) x^{j-2} \cdot g + P''\left(\frac{d}{dx}\right) x^{j-2} \cdot g \\ &= \dots = \left(\sum_{m=0}^j \binom{j}{m} \cdot x^{j-m} \cdot P^{(m)}\left(\frac{d}{dx}\right) \right) g. \end{aligned}$$

Für $g = e^{\lambda x}$ ergibt sich mit $P\left(\frac{d}{dx}\right) e^{\lambda x} = P(\lambda) \cdot e^{\lambda x}$:

$$P\left(\frac{d}{dx}\right) x^j \cdot e^{\lambda x} = \left(\sum_{m=0}^j \binom{j}{m} \cdot x^{j-m} \cdot P^{(m)}(\lambda) \right) \cdot e^{\lambda x}$$

Ist $\lambda = \lambda_i$ eine n_i -fache Nullstelle des charakteristischen Polynoms, verschwinden alle Ableitungen $P^{(m)}(\lambda_i)$ für $m = 0, \dots, n_i - 1$, und die Funktionen $x^j \cdot e^{\lambda_i x}$ mit $j = 0, \dots, n_i - 1$ erfüllen die homogene DGL $P\left(\frac{d}{dx}\right) x^j \cdot e^{\lambda_i x} = 0$.

Q.E.D.

Beispiel 10.33: Betrachte die homogene DGL

$$y''' - 7 \cdot y'' + 16 \cdot y' - 12 \cdot y = 0.$$

Das charakterische Polynom

$$P(\lambda) = \lambda^3 - 7 \cdot \lambda^2 + 16 \cdot \lambda - 12 = (\lambda - 2)^2 \cdot (\lambda - 3)$$

hat eine doppelte Nullstelle $\lambda_1 = 2$ und eine einfache Nullstelle $\lambda_2 = 3$. Die allgemeine Lösung ist damit

$$y(x) = c_{10} \cdot e^{2 \cdot x} + c_{11} \cdot x \cdot e^{2 \cdot x} + c_{20} \cdot e^{3 \cdot x}.$$

Bemerkung 10.34: *Bislang haben wir nur die homogene lineare DGL mit konstanten Koeffizienten betrachtet. Was ist nun mit der inhomogenen DGL*

$$y^{(n)}(x) + a_{n-1} \cdot y^{(n-1)}(x) + \dots + a_0 \cdot y(x) = f(x)?$$

Nach Satz 10.27 reicht es, irgendeine spezielle Lösung y_p zu finden, die dann zusammen mit der oben beschriebenen allgemeinen Lösung der homogenen DGL die allgemeine Lösung der inhomogenen DGL liefert.

Leider gibt es kein allgemeingültiges Rezept, bei gegebenem $f(x)$ systematisch eine Lösung zu ermitteln. (Gewisse Halbsystematiken wie z.B. Fourier- oder Laplace-Transformation werden (in späteren Semestern?) zur Verfügung stehen. Bei Spezialfällen wie z.B.

$$f(x) = f_0 \cdot e^{\lambda \cdot x}$$

kommt man mit dem Ansatz $y_p(x) = c \cdot e^{\lambda \cdot x}$ hin. Sei P das charakteristische Polynom der homogenen DGL, d.h., es ist (mit der Bezeichnung des Beweises von 10.32) die inhomogene DGL

$$P\left(\frac{d}{dx}\right) y = f$$

zu lösen. Analog zum Beweis von 10.32 erhält man für den Ansatz $y_p(x) = c \cdot e^{\lambda \cdot x}$ die Gleichung

$$P\left(\frac{d}{dx}\right) y_p = c \cdot P(\lambda) \cdot e^{\lambda \cdot x} = f_0 \cdot e^{\lambda \cdot x} \quad \Rightarrow \quad c = \frac{f_0}{P(\lambda)},$$

also

$$y_p = \frac{f_0}{P(\lambda)} \cdot e^{\lambda \cdot x},$$

falls λ nicht gerade Nullstelle des charakteristischen Polynoms ist.

„**Resonanz**“: Ist λ eine Nullstelle der Vielfachheit v , so liefert der Ansatz $y_p(x) = c \cdot x^v \cdot e^{\lambda \cdot x}$ dem Beweis von 10.32 folgend

$$\begin{aligned} P\left(\frac{d}{dx}\right) y_p &= c \cdot P\left(\frac{d}{dx}\right) x^v \cdot e^{\lambda \cdot x} = c \cdot \left(\sum_{m=0}^v \binom{v}{m} \cdot x^{v-m} \cdot P^{(m)}(\lambda) \right) \cdot e^{\lambda \cdot x} \\ &= c \cdot P^{(v)}(\lambda) \cdot e^{\lambda \cdot x} = f_0 \cdot e^{\lambda \cdot x}, \end{aligned}$$

also

$$y_p(x) = \frac{f_0}{P^{(v)}(\lambda)} \cdot x^v \cdot e^{\lambda \cdot x}.$$

Bemerkung 10.35: Allgemein gilt für

$$y^{(n)}(x) + a_{n-1} \cdot y^{(n-1)}(x) + \dots + a_0 \cdot y(x) = f_1(x) + f_2(x) + \dots$$

ein Superpositionsprinzip: hat man spezielle Lösungen y_{p1}, y_{p2} etc. zu den rechten Seiten f_1, f_2 etc., so liefert $y_{p1} + y_{p2} + \dots$ eine spezielle Lösung zur rechten Seite $f_1 + f_2 + \dots$

Beispiel 10.36: Die n freien Konstanten in der allgemeinen Lösung einer DGL n -ter Ordnung lassen sich z.B. dazu verwenden, die allgemeine Lösung an n Anfangsbedingungen

$$y(x_0), \quad y'(x_0), \quad \dots, \quad y^{(n-1)}(x_0)$$

anzupassen. Als Beispiel betrachten wir das Anfangswertproblem

$$y''(x) + y(x) = f_0 \cdot \sin(x), \quad y(0) = 1, \quad y'(x) = 0.$$

Die allgemeine Lösung der homogenen Gleichung $y'' + y = 0$ ist mit den Nullstellen $\lambda = \pm i$ des charakteristischen Polynoms:

$$y_{hom}(x) = c_+ \cdot e^{i \cdot x} + c_- \cdot e^{-i \cdot x} = c_1 \cdot \cos(x) + c_2 \cdot \sin(x)$$

mit $c_1 = c_+ + c_-$ und $c_2 = i \cdot (c_+ - c_-)$. Die Lösung u der inhomogenen DGL

$$u''(x) + u(x) = f_0 \cdot e^{i \cdot x} \quad (\#)$$

liefert per $y_p = \Im(u)$ eine Lösung der interessierenden DGL

$$y''(x) + y(x) = \Im(f_0 \cdot e^{i \cdot x}) = f_0 \cdot \sin(x). \quad (\#\#)$$

Da die „Anregung“ $f_0 \cdot \sin(x)$ eine Grundlösung der homogenen DGL ist (Resonanz: $\lambda = i$ ist einfache Nullstelle des charakteristischen Polynoms), wird nach den Rezepten von Bemerkung 10.34 der Ansatz

$$u(x) = c \cdot x \cdot e^{i \cdot x}$$

gemacht. Einsetzen in (#) liefert

$$\underbrace{-c \cdot x \cdot e^{i \cdot x} + 2 \cdot c \cdot i \cdot e^{i \cdot x}}_{u''(x)} + \underbrace{c \cdot x \cdot e^{i \cdot x}}_{u(x)} \stackrel{(!)}{=} f_0 \cdot e^{i \cdot x} \Rightarrow c = \frac{f_0}{2 \cdot i} = -\frac{i \cdot f_0}{2}.$$

Eine spezielle Lösung der inhomogenen DGL (##) ist damit

$$y_p = \Im(u) = \Im\left(-\frac{i \cdot f_0}{2} \cdot x \cdot e^{i \cdot x}\right) = -\frac{f_0}{2} \cdot x \cdot \cos(x).$$

Die allgemeine Lösung der inhomogenen DGL ist damit

$$y(x) = y_{hom}(x) + y_p(x) = c_1 \cdot \cos(x) + c_2 \cdot \sin(x) - \frac{f_0}{2} \cdot x \cdot \cos(x).$$

Wir passen die freien Konstanten an die Anfangsbedingungen $y(0)$, $y'(0)$ an:

$$y(0) = \left(c_1 \cdot \cos(x) + c_2 \cdot \sin(x) - \frac{f_0}{2} \cdot x \cdot \cos(x)\right)_{x=0} = c_1,$$

also $\boxed{c_1 = y(0)}$,

$$y'(0) = \left(-c_1 \cdot \sin(x) + c_2 \cdot \cos(x) - \frac{f_0}{2} \cdot \cos(x) + \frac{f_0}{2} \cdot x \cdot \sin(x)\right)_{x=0} = c_2 - \frac{f_0}{2},$$

also $\boxed{c_2 = y'(0) + f_0/2}$. Die Lösung der inhomogenen DGL ist damit

$$\boxed{y(x) = y(0) \cdot \cos(x) + \left(y'(0) + \frac{f_0}{2}\right) \cdot \sin(x) - \frac{f_0}{2} \cdot x \cdot \cos(x)}.$$

Für die speziellen Werte $y(0) = 1$, $y'(0) = 0$:

$$y(x) = \cos(x) + \frac{f_0}{2} \cdot \sin(x) - \frac{f_0}{2} \cdot x \cdot \cos(x).$$

Allgemein gilt:

Bei linearen DGLen führt das Anpassen von Anfangsbedingungen stets auf ein lösbares lineares Gleichungssystem für die freien Linearfaktoren (Konstanten) eines Fundamentalsystems.

Kapitel 11

Fourier-Entwicklungen

11.1 Hilbert-Raumstrukturen

11.1.1 Definitionen: Skalarprodukt und Norm

Gegeben sei eine $2 \cdot \pi$ -periodische Funktion $f(x)$, d.h., es gelte $f(x + 2 \cdot \pi) = f(x)$ für alle $x \in \mathbb{R}$. Man kann sich solche Funktionen als über dem „Grundintervall“ $x \in [-\pi, \pi]$ definiert vorstellen, aber es ist für die Intuition meist nützlicher, sich solche Funktionen über ganz \mathbb{R} definiert mit der obigen Periodizität vorzustellen. Die einfachsten $2 \cdot \pi$ -periodischen Funktionen sind

$$1, \sin(x), \cos(x), \sin(2 \cdot x), \cos(2 \cdot x), \sin(3 \cdot x), \cos(3 \cdot x) \text{ etc.}$$

bzw. (mit der Euler-Identität $e^{i \cdot x} = \cos(x) + i \cdot \sin(x)$) die Funktionen

$$1, e^{i \cdot x}, e^{-i \cdot x}, e^{i \cdot 2 \cdot x}, e^{-i \cdot 2 \cdot x}, e^{i \cdot 3 \cdot x}, e^{-i \cdot 3 \cdot x} \text{ etc.}$$

Die grundsätzliche Idee ist nun, eine beliebige $2 \cdot \pi$ -periodische Funktion $f(x)$ zu zerlegen in die obigen Basisfunktionen, d.h., zu versuchen, die Funktion als Linearkombination

$$\begin{aligned} f(x) &= a_0 + a_1 \cdot \cos(x) + a_2 \cdot \cos(2 \cdot x) + \dots \\ &\quad + b_1 \cdot \sin(x) + b_2 \cdot \sin(2 \cdot x) + \dots \end{aligned}$$

bzw.

$$f(x) = c_0 + c_1 \cdot e^{i \cdot x} + c_{-1} \cdot e^{-i \cdot x} + c_2 \cdot e^{i \cdot 2 \cdot x} + c_{-2} \cdot e^{-i \cdot 2 \cdot x} + \dots$$

mit geeigneten Koeffizienten a_i, b_i, c_i zu schreiben. In der Tat kann man dies für nicht allzu wilde Funktionen immer machen. Eine dabei besonders angenehme Klasse sind die „quadratintegralen“ Funktionen:

Definition 11.1:

Zu einem Intervall $[a, b] \subset \mathbb{R}$ wird der „**Hilbert–Raum der quadratintegralen Funktionen**“

$$L_2([a, b]) = \left\{ f : [a, b] \rightarrow \mathbb{C}; \int_a^b |f(x)|^2 dx < \infty \right\}$$

mit dem „**Skalarprodukt**“

$$\langle f, g \rangle = \int_a^b \overline{f(x)} \cdot g(x) dx$$

definiert. Wir betrachten im Folgenden speziell $[a, b] = [-\pi, \pi]$ und bezeichnen den Hilbert-Raum $L_2([-\pi, \pi])$ kurz als L_2 .

Dieses Skalarprodukt hat offensichtlich folgende Eigenschaften:

Eigenschaften 11.2:

Für die Abbildung $\langle \cdot, \cdot \rangle : L_2 \times L_2 \rightarrow \mathbb{C}$ gilt:

a) *Linearität:*

$$\langle f, \alpha \cdot g_1 + \beta \cdot g_2 \rangle = \alpha \cdot \langle f, g_1 \rangle + \beta \cdot \langle f, g_2 \rangle$$

für alle $f, g_1, g_2 \in L_2, \alpha, \beta \in \mathbb{C}$.

b) $\langle f, g \rangle = \overline{\langle g, f \rangle}$ (komplexe Konjugation) für alle $f, g \in L_2$.

c) $\langle f, f \rangle \geq 0$ für alle $f \in L_2$. (Beachte: b) impliziert $\langle f, f \rangle \in \mathbb{R}$.)

d) $\langle f, f \rangle = 0 \Leftrightarrow f = 0$.

Hierbei muss man bei d) etwas aufpassen: Das Integral $\langle f, f \rangle = \int_{-\pi}^{\pi} |f(x)|^2 dx$ merkt nichts davon, wenn man Funktionen an einzelnen Stellen undefiniert. Wir identifizieren daher einfach Funktionen, die bei Integration das selbe Verhalten zeigen, d.h., zwei Funktionen f_1, f_2 werden als die selben Elemente des L_2 angesehen, wenn

$$\int_{-\pi}^{\pi} |f_1(x) - f_2(x)|^2 dx = 0$$

gilt. Damit ist dann d) trivialerweise per Definition erfüllt, da man z.B. die Funktionen $f(x)$, die überall bis auf endlich viele Stellen x_0, x_1 etc. verschwinden, als Nullfunktion ansieht.

Bemerkung 11.3: Nach b) ergibt das Vertauschen der Einträge im Skalarprodukt den komplex konjugierten Wert. Nach a) kann man einen skalaren Faktor aus dem rechten Eintrag des Skalarprodukts herausziehen. **Achtung:** Beim Herausziehen eines skalaren Faktors aus der linken Seite taucht dieser wegen b) komplex konjugiert vor dem Skalarprodukt auf:

$$\langle \alpha \cdot f, g \rangle = \overline{\langle g, \alpha \cdot f \rangle} = \overline{\alpha \cdot \langle g, f \rangle} = \bar{\alpha} \cdot \overline{\langle g, f \rangle} = \bar{\alpha} \cdot \langle f, g \rangle.$$

Analog ergibt sich die Linearität bezüglich des linken Eintrags im Skalarprodukt in der folgenden Form:

$$\langle \alpha \cdot f_1 + \beta \cdot f_2, g \rangle = \bar{\alpha} \cdot \langle f_1, g \rangle + \bar{\beta} \cdot \langle f_2, g \rangle.$$

Das Konzept eines Skalarproduktes ist auf endlich-dimensionalen Räumen (dem \mathbb{R}^n bzw. \mathbb{C}^n) aus der Schule bereits wohlbekannt. Es geht aber auch auf unendlich dimensionale Räume (Funktionsräume) über, wobei die Summen des Euklidischen Skalarprodukts zu Integralen werden. Einem Skalarprodukt ist stets ein Längen- und damit ein Abstandsbegriff zugeordnet (eine „Norm“). Wir verwenden hier schon den Begriff „Vektorraum“, der erst in Definition 12.1 axiomatisch eingeführt werden wird. An dieser Stelle sei er einfach nur als „Menge von Objekten, die man addieren und mit skalaren Faktoren multiplizieren kann“ zu lesen:

Definition und Satz 11.4:

Sei L ein Vektorraum mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$. Dann ist $\|f\|_2 := \sqrt{\langle f, f \rangle}$ eine „Norm“ auf L , d.h., die Abbildung $\|\cdot\|_2 : L \rightarrow [0, \infty)$ erfüllt:

- a) $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$ für alle $f, g \in L$ („Dreiecksungleichung“).
- b) $\|\alpha \cdot f\|_2 = |\alpha| \cdot \|f\|_2$ für alle $f \in L, \alpha \in \mathbb{C}$ („Homogenität“).
- c) $\|f\|_2 = 0 \Leftrightarrow f = 0$.

Die dem Skalarprodukt $\langle f, g \rangle = \int_{-\pi}^{\pi} \overline{f(x)} \cdot g(x) dx$ entstammende Norm auf L_2 wird als „ L_2 -Norm“ bezeichnet.

Beweis: Die Eigenschaften a) – c) folgen unmittelbar aus den geforderten Eigenschaften a) – d) der Definition 11.2.

Die Dreiecksungleichung a) ergibt sich folgendermaßen:

$$\begin{aligned} \|f + g\|_2^2 &= \langle f + g, f + g \rangle = \langle f, f \rangle + \langle f, g \rangle + \langle g, f \rangle + \langle g, g \rangle \\ &= \|f\|_2^2 + \langle f, g \rangle + \overline{\langle f, g \rangle} + \|g\|_2^2 \leq \|f\|_2^2 + 2 \cdot |\langle f, g \rangle| + \|g\|_2^2. \end{aligned}$$

Mit einer Anleihe beim nächsten Satz (der Cauchy-Schwarzschen Ungleichung $|\langle f, g \rangle| \leq \|f\|_2 \cdot \|g\|_2$) folgt

$$\|f + g\|_2^2 \leq \|f\|_2^2 + 2 \cdot \|f\|_2 \cdot \|g\|_2 + \|g\|_2^2 = (\|f\|_2 + \|g\|_2)^2.$$

b) ergibt sich durch $\|\alpha \cdot f\|_2^2 = \langle \alpha \cdot f, \alpha \cdot f \rangle = \alpha \cdot \underline{\alpha} \cdot \langle f, f \rangle = |\alpha|^2 \cdot \|f\|_2^2$.
c) ist nichts anderes als Definition 11.2.d).

Q.E.D.

Die folgende Aussage ist eher technischer Natur (ein nützliches Hilfsmittel in Beweisen):

Satz 11.5: (Die Cauchy-Schwarzsche Ungleichung)

Für alle $f, g \in L$ gilt $|\langle f, g \rangle|^2 \leq \langle f, f \rangle \cdot \langle g, g \rangle$, d.h., $|\langle f, g \rangle| \leq \|f\|_2 \cdot \|g\|_2$.
Das Gleichheitszeichen gilt genau dann, wenn f und g bis auf eine multiplikative Konstante übereinstimmen.

Beweis: Übungsaufgabe.

Bemerkung 11.6: Mit der Norm hat man einen Abstandsbegriff: Als Abstand zweier Elemente $f, g \in L$ betrachtet man $\|f - g\|_2$. Mit 11.4.d) haben 2 Elemente genau dann den Abstand 0, wenn sie übereinstimmen.

Definition 11.7: (Konvergenz)

Eine Folge (f_n) von Elementen in L heißt „**konvergent**“ gegen den „**Grenzwert**“ $f^* \in L$, wenn die Abstände zum Grenzwert gegen 0 konvergieren:

$$\lim_{n \rightarrow \infty} \|f_n - f^*\|_2 = 0.$$

Es wird die Notation $\lim_{n \rightarrow \infty} f_n = f^*$ benutzt.

Man nennt auf dem Hilbert-Raum $L = L_2$ die Konvergenz bezüglich der L_2 -Norm auch „**Konvergenz im quadratischen Mittel**“.

Beispiel 11.8: Betrachte $L = \{f : [-1, 1] \rightarrow \mathbb{R}; f \text{ stetig}\}$ mit der L_2 -Norm $\|f\|_2 = \int_{-1}^1 |f(x)|^2 dx$. Betrachte die folgenden Funktionen, die im Bereich $[0, 1/n]$ linear von 0 bis 1 wachsen:

$$f_n(x) = \begin{cases} 0 & \text{für } -1 \leq x \leq 0, \\ n \cdot x & \text{für } 0 \leq x \leq \frac{1}{n}, \\ 1 & \text{für } \frac{1}{n} \leq x \leq 1. \end{cases}$$

Bei vorgegebenem x konvergieren diese Werte offensichtlich punktweise gegen

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

Mit

$$\begin{aligned}\|f_n - f^*\|_2 &= \int_{-1}^1 (f_n(x) - f^*(x))^2 dx = \int_0^{1/n} (f_n(x) - f^*(x))^2 dx \\ &= \int_0^{1/n} (n \cdot x - 1)^2 dx = \frac{1}{3 \cdot n}\end{aligned}$$

konvergiert die Funktionenfolge f_n auch bezüglich der L_2 -Norm gegen f^* . Aber (Problem!): f^* ist nicht stetig, also $f^* \notin L$.

Das Problem, dass die Grenzfunktion einer Folge stetiger Funktionen nicht mehr stetig zu sein braucht, soll nun etwas erläutert werden. Dazu zunächst eine mehr mathematisch-technische Definition (die in späteren Anwendungen dann nicht mehr so interessant sein wird):

Definition 11.9: (Vollständigkeit)

Eine Folge (f_n) heißt „**Cauchy-Folge**“ bezüglich einer Norm $\|\cdot\|$, wenn zu jedem $\epsilon > 0$ ein $N(\epsilon)$ existiert, sodass $\|f_n - f_m\| < \epsilon$ gilt für alle $n, m \geq N(\epsilon)$.

Ein Raum mit einer Norm heißt „**vollständig**“, wenn jede Cauchy-Folge in diesem Raum einen Grenzwert besitzt, der wieder in diesem Raum liegt.

Definition 11.10: (Hilberträume)

Ein Vektorraum mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$ heißt „**Hilbert-Raum**“, wenn er bezüglich der vom Skalarprodukt stammenden Norm $\|f\|_2 = \sqrt{\langle f, f \rangle}$ vollständig ist.

Bemerkung 11.11: Beispiel 11.8 zeigt, dass der Raum der stetigen Funktionen bezüglich der L_2 -Norm nicht vollständig ist. (Die Funktionenfolge f_n in diesem Beispiel ist eine Cauchy-Folge. Der Grenzwert f^* (es kann nur einen geben) ist aber unstetig.) Es ist daher im Folgenden nicht sinnvoll, den Raum der stetigen Funktionen weiter zu betrachten. Stattdessen sucht man nach dem kleinsten bzgl. der L_2 -Norm vollständigen Raum, der die stetigen Funktionen umfasst. Dieser stellt sich als der Raum der quadratintegrablen Funktionen L_2 heraus (Definition und Satz 11.1).

b) Damit die oben definierten L_2 -Räume wirklich vollständig sind, muss man statt des Riemann-Integrals einen verallgemeinerten Integralbegriff benutzen: das sogenannte **Lebesgue-Integral**. Außer für sehr pathologische (irreguläre) Funktionen stimmt das Lebesgue-Integral aber mit dem Riemann-Integral überein. Wir werden uns daher mit diesen sehr mathematisch-technischen Details nicht abmühen. Einen exakten Beweis für die Vollständigkeit des L_2 -Raums können wir folglich hier nicht führen.

c) Zu jeder L_2 -Funktion f gibt es zu jedem $\epsilon > 0$ eine stetige Funktion g , die f approximiert: $\|f - g\|_2 < \epsilon$. Das liefert eine einfache Beweistechnik im L_2 :

zeige die behauptete Aussage zunächst nur für stetige Funktionen. Betrachte dann eine konvergente Folge stetiger Funktionen und zeige, dass die Aussage auf den (eventuell unstetigen) Grenzwert in L_2 übergeht.

d) Für endliche Intervalle $[a, b]$ liegen alle auf $[a, b]$ stetigen Funktionen in $L_2([a, b])$. Unstetige Funktionen mit Sprungstellen können ebenfalls integriert werden und liegen im L_2 -Raum. Sogar Funktionen mit (milden) Singularitäten sind quadratintegrabel, z.B.:

$$f(x) = x^{-\frac{1}{3}} \in L_2, \text{ denn } \int_{-\pi}^{\pi} (x^{-1/3})^2 dx = 3 \cdot \pi^{2/3} < \infty.$$

e) Für quadratintegrale Funktionen existiert das Skalarprodukt selbst dann, wenn die Funktionen singulär sind. Dies liegt an der Cauchy-Schwarzschen Ungleichung 11.5. Diese Ungleichung garantiert weiterhin, dass die Summe zweier L_2 -Funktionen wieder eine L_2 -Funktion ist (d.h., L_2 ist in der Tat ein Vektorraum).

f) Die folgende Theorie von (verallgemeinerten) Fourier-Reihen gilt für beliebige L_2 -Funktionen, also auch für unstetige und sogar (mild-)singuläre Funktionen. Gerade diese Allgemeinheit macht die Fourier-Analyse für Anwendungen interessant.

11.1.2 Bestapproximation und Konvergenz

Wie aus der Schule bekannt, ist im \mathbb{R}^n mit dem Euklidischen Skalarprodukt (den wir nun als Hilbert-Raum ansehen) die Orthogonalität zweier Vektoren dadurch charakterisiert, dass das Skalarprodukt verschwindet. Wir benutzen die selbe Notation für allgemeine Hilbert-Räume.

Definition 11.12: (Orthogonalität)

Zwei Elemente f, g eines Hilbert-Raums mit $\langle f, g \rangle = 0$ heißen „**orthogonal**“. Eine Familie $\{f_n\}$ von Hilbert-Raumelementen nennt man ein „**orthogonales System**“, wenn alle Paare f_n, f_m mit $m \neq n$ orthogonal sind. Gilt zusätzlich $\langle f_n, f_n \rangle = 1$, so heißt das System „**orthonormal**“.

Beispiel 11.13: Hier ist das entscheidende Beispiel, das uns im Folgenden am meisten interessieren wird:

Die Familie der komplexen Exponentialfunktionen $\{e^{i \cdot n \cdot x}\}, i = \sqrt{-1}, n \in \mathbb{Z}$ ist ein Orthogonalsystem auf $L_2([-\pi, \pi])$.

Dies ist leicht nachgerechnet. Für $n \neq m$ gilt:

$$\langle e^{i \cdot m \cdot x}, e^{i \cdot n \cdot x} \rangle = \int_{-\pi}^{\pi} \overline{e^{i \cdot m \cdot x}} \cdot e^{i \cdot n \cdot x} dx = \int_{-\pi}^{\pi} e^{-i \cdot m \cdot x} \cdot e^{i \cdot n \cdot x} dx$$

$$= \int_{-\pi}^{\pi} e^{i \cdot (n-m) \cdot x} dx = \left[\frac{e^{i \cdot (n-m) \cdot x}}{i \cdot (n-m)} \right]_{x=-\pi}^{x=\pi} = \frac{(-1)^{n-m} - (-1)^{n-m}}{i \cdot (n-m)} = 0.$$

Für $m = n$:

$$\langle e^{i \cdot n \cdot x}, e^{i \cdot n \cdot x} \rangle = \int_{-\pi}^{\pi} e^{-i \cdot n \cdot x} \cdot e^{i \cdot n \cdot x} dx = \int_{-\pi}^{\pi} 1 dx = 2 \cdot \pi.$$

Mit der Euler-Identität

$$e^{i \cdot n \cdot x} = \cos(n \cdot x) + i \cdot \sin(n \cdot x)$$

bzw.

$$\cos(n \cdot x) = \frac{e^{i \cdot n \cdot x} + e^{-i \cdot n \cdot x}}{2}, \quad \sin(n \cdot x) = \frac{e^{i \cdot n \cdot x} - e^{-i \cdot n \cdot x}}{2 \cdot i}$$

ergibt sich für die Familie reeller Funktionen $\{1, \cos(x), \sin(x), \cos(2 \cdot x), \sin(2 \cdot x), \dots\}$:

$$\begin{aligned} \langle \cos(m \cdot x), \cos(n \cdot x) \rangle &= 0, & m \neq n, \\ \langle \sin(m \cdot x), \sin(n \cdot x) \rangle &= 0, & m \neq n, \\ \langle \cos(m \cdot x), \sin(n \cdot x) \rangle &= 0 & (\text{auch für } m = n). \end{aligned}$$

Die konstante Funktion 1 ist dabei $\cos(m \cdot x)$ mit $m = 0$. Für $m = n$ gilt:

$$\begin{aligned} \langle 1, 1 \rangle &= 2 \cdot \pi, \\ \langle \cos(n \cdot x), \cos(n \cdot x) \rangle &= \pi & (n > 0), \\ \langle \sin(n \cdot x), \sin(n \cdot x) \rangle &= \pi & (n > 0). \end{aligned}$$

Die trigonometrischen Funktionen $\{1, \cos(x), \sin(x), \cos(2 \cdot x), \sin(2 \cdot x), \dots\}$ bilden ein Orthogonalsystem auf $L_2([-\pi, \pi])$.

Will man Entwicklungen nach „einfachen Basisfunktionen“ f_0, f_1, \dots vornehmen, so ist es technisch sehr vorteilhaft, wenn diese ein Orthogonalsystem bilden. Ist dies nicht der Fall, so kann die Folge f_0, f_1, \dots mit einem sehr einfachen Algorithmus durch Linearkombinationen in ein Orthogonalsystem verwandelt werden („Gram-Schmidt-Orthogonalisierung“, siehe Übungen).

Bemerkung 11.14: Aus einem Orthogonalsystem $\{f_0, f_1, \dots\}$ kann man sofort ein Orthonormalsystem $\{\tilde{f}_0, \tilde{f}_1, \dots\}$ machen, indem man $\tilde{f}_n = f_n / \|f_n\|_2$ setzt.

Wir betrachten ein Orthogonalsystem $\{f_0, f_1, \dots\}$ und versuchen zunächst, ein Hilbert-Raumelement f durch eine endliche Linearkombination

$$f \approx \sum_{k=0}^n c_k \cdot f_k$$

zu approximieren. Welche Koeffizienten c_k sollte man wählen, damit der Abstand zwischen f und der Summe im L_2 -Sinne möglichst klein wird? Als

Vorüberlegung betrachten wir zunächst, wie man für ein exakt darstellbares f die Entwicklungskoeffizienten c_k aus f ermittelt:

$$f = \sum_{k=0}^n c_k \cdot f_k \quad \Rightarrow \quad \langle f_m, f \rangle = \langle f_m, \sum_{k=0}^n c_k \cdot f_k \rangle = \sum_{k=0}^n c_k \cdot \langle f_m, f_k \rangle.$$

Wegen der Orthogonalität verbleibt für $m \in \{0, \dots, n\}$ von der Summe nur der Term mit $k = m$:

$$\langle f_m, f \rangle = c_m \cdot \langle f_m, f_m \rangle, \quad \text{also} \quad c_m = \frac{\langle f_m, f \rangle}{\langle f_m, f_m \rangle}.$$

Definition 11.15: (Verallgemeinerte Fourier-Koeffizienten)

Für eine L_2 -Funktion f heißen

$$c_k = \frac{\langle f_k, f \rangle}{\langle f_k, f_k \rangle}, \quad k = 0, 1, 2, \dots$$

die (verallgemeinerten) „**Fourier-Koeffizienten bezüglich des Orthogonalsystems** $\{f_0, f_1, \dots\}$ “. Die Linearkombination

$$S_n = \sum_{k=0}^n c_k \cdot f_k$$

heißt (verallgemeinerte) „**Fourier-Approximation von f bezüglich des Orthogonalsystems**“.

Das folgende elementare Ergebnis zeigt, dass diese Fourier-Koeffizienten in der Tat die beste Koeffizientenwahl darstellen, um den Abstand zwischen f und einer Linearkombination $\sum_k \alpha_k \cdot f_k$ im L_2 -Sinne möglichst klein zu machen:

Lemma 11.16: (Besselsche (Un-)Gleichung)

Sei $\{f_0, \dots, f_n\}$ ein Orthogonalsystem.

a) Für jede Wahl von Koeffizienten $\alpha_k \in \mathbb{C}$ gilt die „**allgemeine Besselsche Gleichung**“

$$\left\| f - \sum_{k=0}^n \alpha_k \cdot f_k \right\|_2^2 = \|f\|_2^2 - \sum_{k=0}^n |c_k|^2 \cdot \|f_k\|_2^2 + \sum_{k=0}^n |\alpha_k - c_k|^2 \cdot \|f_k\|_2^2,$$

wo die c_k die Fourier-Koeffizienten aus Definition 11.15 sind.

b) Für die Wahl $\alpha_k = c_k$ folgt die „**spezielle Besselsche Gleichung**“

$$\left\| f - \sum_{k=0}^n c_k \cdot f_k \right\|_2^2 = \|f\|_2^2 - \sum_{k=0}^n |c_k|^2 \cdot \|f_k\|_2^2.$$

c) Es folgt die „**Besselsche Ungleichung**“

$$\sum_{k=0}^n |c_k|^2 \cdot \|f_k\|_2^2 \leq \|f\|_2^2.$$

Beweis: Die allgemeine Besselsche Gleichung a) wird als Übungsaufgabe nachgerechnet. Hieraus folgen sofort b) und c).

Q.E.D.

Interpretation 11.17:

Die allgemeine Besselsche Gleichung liefert sofort, dass der Abstand

$$\left\| f - \sum_{k=0}^n \alpha_k \cdot f_k \right\|_2$$

durch die Wahl der Koeffizienten $\alpha_k = c_k$ minimiert wird. Die Fourier-Approximation von f wird daher als „**Bestapproximation von f im quadratischen Mittel**“ bezüglich des Orthogonalsystems bezeichnet. Der Abstand zwischen f und dieser Bestapproximation (der „**Approximationsfehler**“) ist durch

$$\left\| f - \sum_{k=0}^n c_k \cdot f_k \right\|_2^2 = \|f\|_2^2 - \sum_{k=0}^n |c_k|^2 \cdot \|f_k\|_2^2$$

gegeben. Eine Besonderheit dieser L_2 -Bestapproximation ist, dass die Entwicklungskoeffizienten c_k nicht von der Anzahl n der benutzten Entwicklungselemente f_0, \dots, f_n abhängen! Man kann den Approximationsfehler kleiner und kleiner machen, indem man einfach immer mehr Elemente im Orthogonalsystem hinzunimmt (also n anwachsen läßt), ohne dass man die Koeffizienten anpassen muss!

Nun geht es um die Frage, was passiert, wenn man den Grenzwert $n \rightarrow \infty$ betrachtet. Konvergiert die Fourier-Approximation (im L_2 -Sinne) gegen f ? Dies kann nicht allgemein beantwortet werden, sondern hängt vom betrachteten Orthogonalsystem $\{f_0, f_1, \dots\}$ ab:

Definition 11.18: (Vollständige Orthonormalsysteme)

Ein Orthogonalsystem $\{f_0, f_1, \dots\}$ heißt „**vollständig**“, wenn für jedes f im Hilbert-Raum die Fourier-Approximation im L_2 -Sinne gegen f konvergiert:

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n c_k \cdot f_k = f,$$

d.h.,

$$\lim_{n \rightarrow \infty} \left\| f - \sum_{k=0}^n c_k \cdot f_k \right\|_2^2 = 0.$$

Mit der Besselschen Gleichung ergibt sich sofort die folgende Charakterisierung vollständiger Orthogonalsysteme, die mathematisch aber recht trivial ist und keinerlei effektives Hilfsmittel liefert, die Vollständigkeit eines Orthogonalsystems zu überprüfen:

Satz 11.19: (Parsevalsche Gleichung)

Ein Orthogonalsystem $\{f_0, f_1, \dots\}$ ist genau dann vollständig, wenn für jedes Hilbert-Raumelement f die folgende „Parsevalsche Gleichung“ gilt:

$$\sum_{k=0}^{\infty} |c_k| \cdot \|f_k\|_2^2 = \|f\|_2^2.$$

Beweis: Die Gleichung folgt sofort aus Lemma 11.16.b).

Q.E.D.

Das einzig für uns Interessante an dieser Aussage ist die folgende Interpretation:

Physikalische Interpretation 11.20:

In vielen physikalischen Anwendungen stellt das „Signal“ f ein Feld (Strom, Spannung, elektrisches oder magnetisches Feld) dar, dessen Quadrat als eine physikalische Leistung oder als lokale Energiedichte zu interpretieren ist. Das die L_2 -Norm definierende Integral hat dann die Interpretation einer Energie, genauer, einer über eine Periode des Signals „gemittelten Energie“.

Bei einer Fourier-Entwicklung $f = \sum_{k=0}^{\infty} c_k \cdot f_k$ sprechen Physiker auch von einer „Zerlegung des Signals“ in die „Fourier-Moden“ f_k .

Die Parsevalsche Gleichung besagt, dass sich die Energie des Signals **vollständig** auf die Fourier-Moden verteilen muss. Ist dies nicht der Fall (d.h., ist das Orthogonalsystem nicht vollständig), so muss es weitere Anteile des Signals geben, die durch die Fourier-Moden nicht erfasst werden.

11.2 Trigonometrische Fourier-Reihen

14.1.04↓

Die allgemeinen Hilbert-Raumstrukturen der Fourier-Reihen wurde im letzten Abschnitt in etwas abstrakter Art und Weise zusammengetragen. Nun wird

es endlich konkret. Wir betrachten das in Beispiel 11.13 diskutierte spezielle Orthogonalsystem

$$\{e^{i \cdot k \cdot x}\}_{k \in \mathbb{Z}} \quad (i = \sqrt{-1})$$

auf dem Hilbert-Raum $L_2 = L_2([-\pi, \pi])$. Ziel ist es, eine auf dem Intervall $[-\pi, \pi]$ gegebene Funktion nach dem obigen Orthogonalsystem zu zerlegen und zu studieren, ob und wie die Fourier-Approximationen gegen die Funktion konvergieren.

Rechentechnisch ist es meist einfacher, die komplexen Exponentialfunktionen zu verwenden. Äquivalenterweise kann man die trigonometrischen Funktionen

$$\{1, \cos(x), \sin(x), \cos(2 \cdot x), \sin(2 \cdot x), \dots\}$$

betrachten, die für reelle Funktionen zu reellen Formeln führen und daher im Anwendungskontext oft physikalisch direkter zu interpretieren sind.

11.2.1 Definition der trigonometrischen Fourier-Reihen

Wir fassen zunächst die Ergebnisse von Beispiel 11.13 zusammen. Der Unterschied zwischen der „komplexen“ und der „reellen“ Darstellung besteht lediglich darin, dass statt des komplexen Pärchens $(e^{i \cdot k \cdot x}, e^{-i \cdot k \cdot x})$ in der reellen Darstellung das Pärchen $(\cos(k \cdot x), \sin(k \cdot x))$ benutzt wird, das mit der Euler-Formel

$$e^{i \cdot k \cdot x} = \cos(k \cdot x) + i \cdot \sin(k \cdot x)$$

bzw.

$$\cos(k \cdot x) = \frac{e^{i \cdot k \cdot x} + e^{-i \cdot k \cdot x}}{2}, \quad \sin(k \cdot x) = \frac{e^{i \cdot k \cdot x} - e^{-i \cdot k \cdot x}}{2 \cdot i}$$

dem komplexen Pärchen äquivalent ist:

Definition 11.21: (Die trigonometrische Fourier-Entwicklung)

Zu einer Funktion $f \in L_2([-\pi, \pi])$ definiere die „**Fourier-Koeffizienten**“

$$c_k = \frac{\langle e^{i \cdot k \cdot x}, f \rangle}{\langle e^{i \cdot k \cdot x}, e^{i \cdot k \cdot x} \rangle} = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(x) \cdot e^{-i \cdot k \cdot x} dx$$

mit $k = 0, \pm 1, \pm 2, \dots$. Alternativ definiere

$$a_k = \frac{\langle \cos(k \cdot x), f \rangle}{\langle \cos(k \cdot x), \cos(k \cdot x) \rangle} = \begin{cases} \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(x) dx, & k = 0, \\ \frac{1}{\pi} \cdot \int_{-\pi}^{\pi} f(x) \cdot \cos(k \cdot x) dx, & k = 1, 2, \dots \end{cases}$$

sowie

$$b_k = \frac{\langle \sin(k \cdot x), f \rangle}{\langle \sin(k \cdot x), \sin(k \cdot x) \rangle} = \frac{1}{\pi} \cdot \int_{-\pi}^{\pi} f(x) \cdot \sin(k \cdot x) dx, \quad k = 1, 2, \dots$$

Die entsprechenden „Fourier-Approximationen“ sind

$$S_n(x) = \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} = a_0 + \sum_{k=1}^n \left(a_k \cdot \cos(k \cdot x) + b_k \cdot \sin(k \cdot x) \right)$$

mit $n = 0, 1, 2, \dots$.

Mit der Euler-Formel ergibt sich folgender Zusammenhang zwischen den beiden Darstellungen:

$$\begin{aligned} \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} &= c_0 + \sum_{k=1}^n \left(c_k \cdot e^{i \cdot k \cdot x} + c_{-k} \cdot e^{-i \cdot k \cdot x} \right) \\ &= c_0 + \sum_{k=1}^n \left(c_k \cdot \left(\cos(k \cdot x) + i \cdot \sin(k \cdot x) \right) + c_{-k} \cdot \left(\cos(k \cdot x) - i \cdot \sin(k \cdot x) \right) \right) \\ &= \underbrace{c_0}_{a_0} + \sum_{k=1}^n \left(\underbrace{(c_k + c_{-k})}_{a_k} \cdot \cos(k \cdot x) + \underbrace{i \cdot (c_k - c_{-k})}_{b_k} \cdot \sin(k \cdot x) \right). \end{aligned}$$

Zusammenhang 11.22:

Der Zusammenhang zwischen den komplexen Fourier-Koeffizienten c_k und den reellen Koeffizienten a_k, b_k ist gegeben durch:

$a_k = c_k + c_{-k}$	$c_k = \frac{1}{2} \cdot (a_k - i \cdot b_k)$
$b_k = i \cdot (c_k - c_{-k})$	$c_{-k} = \frac{1}{2} \cdot (a_k + i \cdot b_k)$

mit $k = 1, 2, \dots$ und zusätzlich $a_0 = c_0$.

Bemerkung 11.23: Ist die Funktion f reell, so sind die Fourier-Koeffizienten a_k, b_k reell. In der komplexen Entwicklung entspricht dies der Bedingung

$$c_{-k} = \overline{c_k}, \quad k = 0, 1, 2, \dots$$

Bemerkung 11.24: Für die Definition der Fourier-Koeffizienten a_k, b_k bzw. c_k sind nur die Werte der Funktion f auf dem Intervall $[-\pi, \pi]$ von Belang. Man sollte sich aber gleich angewöhnen, sich die Funktion f als $2 \cdot \pi$ -periodisch auf ganz \mathbb{R} fortgesetzt vorzustellen, denn die Fourier-Approximationen

$$S_n(x) = \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} = a_0 + \sum_{k=1}^n \left(a_k \cdot \cos(k \cdot x) + b_k \cdot \sin(k \cdot x) \right)$$

sind ja ebenfalls $2 \cdot \pi$ -periodisch. Das liefert gleich die richtige Sichtweise für die im Abschnitt 11.2.3 folgenden Betrachtungen zur punktweisen Konvergenz.

Bemerkung 11.25: Ist die Funktion f gerade oder ungerade, so lassen sich die Integrale $\int_{-\pi}^{\pi}$, die die Fourier-Koeffizienten bestimmen, vereinfachen. Die Teilintegrale $\int_{-\pi}^0$ stimmen (eventuell bis auf ein Vorzeichen) mit den Teilintegralen \int_0^{π} überein und verdoppeln diesen Wert oder löschen ihn zu 0 aus:

Für gerade Funktionen f (also $f(x) = f(-x)$) gilt

$$a_0 = \frac{1}{\pi} \int_0^{\pi} f(x) dx, \quad a_k = \frac{2}{\pi} \int_0^{\pi} f(x) \cdot \cos(k \cdot x) dx, \quad b_k = 0$$

mit $k = 1, 2, \dots$.

Für ungerade Funktionen f (also $f(x) = -f(-x)$) gilt

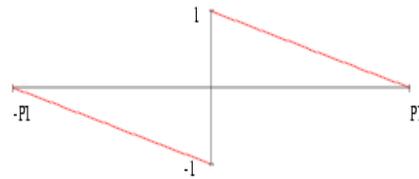
$$a_0 = 0, \quad a_k = 0, \quad b_k = \frac{2}{\pi} \int_0^{\pi} f(x) \cdot \sin(k \cdot x) dx$$

mit $k = 1, 2, \dots$.

Beispiel 11.26: Einige Beispiele von Fourier-Reihen:

a) Die („Sägezahn“-) Funktion

$$f(x) = \begin{cases} -1 - \frac{x}{\pi} & \text{für } x \in [-\pi, 0), \\ 1 - \frac{x}{\pi} & \text{für } x \in [0, \pi] \end{cases}$$



ist ungerade, also $a_k = 0 \forall k = 0, 1, 2, \dots$

$$\begin{aligned} b_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cdot \sin(k \cdot x) dx = \frac{2}{\pi} \int_0^{\pi} \left(1 - \frac{x}{\pi}\right) \cdot \sin(k \cdot x) dx \\ &= \frac{1}{k} - \frac{\sin(k \cdot \pi)}{\pi \cdot k^2} = \frac{1}{k} \quad \forall k = 1, 2, \dots, \end{aligned}$$

also

$$f(x) \sim \sum_{k=1,2,\dots} \frac{\sin(k \cdot x)}{k}.$$

b) Die Funktion $f(x) = |x|$ ist gerade, also $b_k = 0 \forall k = 1, 2, \dots$

$$a_0 = \frac{1}{2 \cdot \pi} \int_{-\pi}^{\pi} f(x) dx = \frac{1}{\pi} \int_0^{\pi} x dx = \frac{\pi}{2},$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cdot \cos(k \cdot x) dx = \frac{2}{\pi} \int_0^{\pi} x \cdot \cos(k \cdot x) dx$$

$$= \frac{2}{\pi} \cdot \left[\frac{\cos(k \cdot x)}{k^2} + \frac{x \cdot \sin(k \cdot x)}{k} \right]_{x=0}^{x=\pi} = \frac{2}{\pi} \cdot \frac{(-1)^k - 1}{k^2} = \begin{cases} -\frac{4}{\pi \cdot k^2}, & k = 1, 3, 5, \dots, \\ 0, & k = 2, 4, 6, \dots, \end{cases}$$

also

$$|x| \sim \frac{\pi}{2} - \frac{4}{\pi} \cdot \sum_{k=1,3,5,\dots} \frac{\cos(k \cdot x)}{k^2}.$$

11.2.2 Konvergenz im quadratischen Mittel

Satz 11.27: (Die trigonometrischen Funktionen sind vollständig)

Die trigonometrischen Funktionen $\{e^{i \cdot k \cdot x}\}_{k \in \mathbb{Z}}$ bzw.

$$\{1, \cos(x), \sin(x), \cos(2 \cdot x), \sin(2 \cdot x), \dots\}$$

bilden ein vollständiges Orthogonalsystem in $L_2([-\pi, \pi])$ im Sinne von Definition 11.18. Also: für jedes $f \in L_2([-\pi, \pi])$ gilt im L_2 -Sinne

$$f = \lim_{n \rightarrow \infty} S_n,$$

wo S_n die endlichen Fourier-Approximationen

$$S_n(x) = \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} = a_0 + \sum_{k=1}^n \left(a_k \cdot \cos(k \cdot x) + b_k \cdot \sin(k \cdot x) \right)$$

sind.

Wir haben hier nicht den technischen Apparat, einen sauberen Beweis zu führen, daher nur die grobe Beweisidee:

Beweisidee: Wir werden im Abschnitt 11.2.3 die punktweise Konvergenz von $S_n(x)$ gegen $f(x)$ betrachten und herausbekommen, dass

$$f(x) = \lim_{n \rightarrow \infty} S_n(x)$$

an allen Stellen $x \in [-\pi, \pi]$ gilt, wenn f nur genügend glatt ist (z.B. stetig und beidseitig differenzierbar, siehe Satz 11.34). Diese punktweise Konvergenz impliziert natürlich die L_2 -Konvergenz von S_n gegen f . Die Funktionen in L_2 lassen sich durch solche glatten Funktionen im L_2 -Sinne beliebig genau approximieren, woraus (nach einigen technischen Abschätzungen) dann folgt, dass die L_2 -Konvergenz für die glatten Funktionen auch auf die weniger glatten Funktionen in L_2 übergeht.

Q.E.D.

11.2.3 Punktweise Konvergenz

Die Vollständigkeit der trigonometrischen Funktionen nach Satz 11.27 garantiert uns die Konvergenz der trigonometrischen Fourier-Reihen S_n gegen die Ausgangsfunktion f im L_2 -Sinne. Intuitiv besagt dies, dass die Fourier-Approximationen $S_n(x)$ für „die meisten Punkte“ $x \in [-\pi, \pi]$ gegen den Wert $f(x)$ konvergieren muss, es bleibt aber hinreichend viel Spielraum für Stellen x ,

an denen $S_n(x)$ nicht gegen $f(x)$ konvergiert. In diesem Abschnitt soll der Frage nach der punktweisen Konvergenz genauer nachgegangen werden: an welchen Punkten x kann garantiert werden, dass $\lim_{n \rightarrow \infty} S_n(x) = f(x)$ gilt? Zunächst eine allgemeine Aussage zur Konvergenz von Funktionenreihen:

Satz 11.28: (Weierstraßsches Majorantenkriterium)

Sei (f_k) eine Folge stetiger Funktionen über $[a, b] \subset \mathbb{R}$. Konvergiert die Reihe $\sum_{k=0}^{\infty} y_k$ mit $y_k := \sup\{|f_k(x)|; x \in [a, b]\}$, so konvergiert

$$S_n(x) = \sum_{k=0}^n f_k(x)$$

für jedes $x \in [a, b]$. Die punktweise durch $S^*(x) = \lim_{n \rightarrow \infty} S_n(x)$ definierte Grenzfunktion ist wiederum stetig.

Beweis: technisch.

Beispiel 11.29: Da $\sum_k \frac{1}{k^2}$ konvergiert, erhalten wir sofort für das Beispiel 11.26.b), dass die Reihe

$$S^*(x) = \frac{\pi}{2} - \frac{4}{\pi} \cdot \sum_{k=1,3,5,\dots} \frac{\cos(k \cdot x)}{k^2} \stackrel{(k=2 \cdot m+1)}{=} \frac{\pi}{2} - \frac{4}{\pi} \cdot \sum_{m=0}^{\infty} \frac{\cos((2 \cdot m + 1) \cdot x)}{(2 \cdot m + 1)^2}$$

für jedes $x \in [\pi, \pi]$ konvergiert und dass die Grenzfunktion $S^*(x)$ stetig ist. Aus der punktweisen Konvergenz folgt auch die Konvergenz von S_n gegen S^* im L_2 -Sinne, denn

$$\begin{aligned} \lim_{n \rightarrow \infty} \|S_n - S^*\|_2 &= \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} (S_n(x) - S^*(x))^2 dx \\ &\stackrel{(*)}{=} \int_{-\pi}^{\pi} \lim_{n \rightarrow \infty} (S_n(x) - S^*(x))^2 dx = \int_{-\pi}^{\pi} 0 dx = 0. \end{aligned}$$

(Mathematisch ist der Schritt $(*)$ (Vertauschen von Limes und Integral) alles andere als trivial. Hier müssten wir den Begriff der „gleichmäßigen Konvergenz“ einführen. Als Physiker machen wir das einfach skrupellos.)

Dem Vollständigkeitsatz 11.27 nach konvergiert S_n im L_2 -Sinne gegen die Ausgangsfunktion $f(x) = |x|$, welche die Fourier-Entwicklung

$$\frac{\pi}{2} - \frac{4}{\pi} \cdot \sum_{k=1,3,5,\dots} \frac{\cos(k \cdot x)}{k^2}$$

erzeugte. Andererseits ist auch S^* der Grenzwert von S_n im L_2 -Sinne. Da Grenzwerte eindeutig sind, folgt $f = S^*$ im L_2 -Sinne (d.h., „fast überall“)¹. Da weiterhin sowohl

¹Der Begriff „fast überall“ sollte hier intuitiv als „überall bis auf endliche viele Ausnahmestellen“ verstanden werden. Es gibt eine exakte maßtheoretische Definition dieses Begriffes, der jedoch recht technischer Natur ist und hier nicht weiter vorgestellt werden soll.

f als auch S^* stetige Funktionen sind, bedeutet „ $f = S^*$ im L_2 -Sinne“, dass auch punktweise $f(x) = S^*(x)$ für alle $x \in [-\pi, \pi]$ gelten muss, also

$$|x| = \frac{\pi}{2} - \frac{4}{\pi} \cdot \sum_{m=0}^{\infty} \frac{\cos((2 \cdot m + 1) \cdot x)}{(2 \cdot m + 1)^2} \quad \text{für alle } x \in [-\pi, \pi].$$

Wir haben damit ein erstes Beispiel, in dem die (unendliche) Fourier-Reihe punktweise die Ausgangsfunktion darstellt.

Zunächst technisches Vorgeplänkel:

Definition 11.30: (Die integrierbaren Funktionen)

Die Funktionen des Raums

$$L_1([a, b]) = \left\{ f : [a, b] \rightarrow \mathbb{C}; \int_a^b |f(x)| dx < \infty \right\}$$

nennt man die „über $[a, b]$ integrierbaren Funktionen“. Die trigonometrischen Fourier-Koeffizienten

$$c_k = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(x) \cdot e^{-i \cdot k \cdot x} dx$$

existieren wegen

$$\left| \int_{-\pi}^{\pi} f(x) \cdot e^{-i \cdot k \cdot x} dx \right| \leq \int_{-\pi}^{\pi} |f(x) \cdot e^{-i \cdot k \cdot x}| dx = \int_{-\pi}^{\pi} |f(x)| dx < \infty$$

für jedes $f \in L_1([-\pi, \pi])$.

Bemerkung 11.31: Der Raum $L_2([a, b])$ der quadratintegrierbaren Funktionen 11.1 ist ein Unterraum von $L_1([a, b])$, wenn $[a, b]$ ein endliches Intervall ist. Dies liegt an der Cauchy-Schwarzschen Ungleichung 11.5. Für endliche Intervalle liegt nämlich

$$g(x) = \begin{cases} \frac{f(x)}{|f(x)|} & \text{für } f(x) \neq 0, \\ 1 & \text{für } f(x) = 0 \end{cases}$$

in $L_2([a, b])$, da $|g(x)| \equiv 1$ gilt:

$$\|g\|_2 = \sqrt{\int_a^b |g(x)|^2 dx} = \sqrt{b-a}.$$

Per Cauchy-Schwarz folgt mit $\overline{f(x)} \cdot g(x) = \overline{f(x)} \cdot f(x)/|f(x)| = |f(x)|$:

$$\int_a^b |f(x)| dx = \int_a^b \overline{f(x)} \cdot g(x) dx = \langle f, g \rangle \leq \|f\|_2 \cdot \|g\|_2 = \sqrt{b-a} \cdot \|f\|_2 < \infty.$$

Lemma 11.32: (Riemann–Lebesgue–Lemma)

Die Fourier–Koeffizienten jeder integrierbaren Funktion $f \in L_1([-\pi, \pi])$ bilden eine Nullfolge:

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} f(x) \cdot \cos(k \cdot x) \, dx = \lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} f(x) \cdot \sin(k \cdot x) \, dx = 0.$$

Beweisskizze (technisch): Wir stellen uns f als $2 \cdot \pi$ –periodisch fortgesetzt vor. Für $2 \cdot \pi$ –periodische Funktionen $g(x)$ gilt

$$\int_{-\pi+\alpha}^{\pi+\alpha} g(x) \, dx = \int_{-\pi}^{\pi} g(x) \, dx$$

für jedes $\alpha \in \mathbb{R}$. Mit der Substitution $x = y - \pi/k$ folgt:

$$\begin{aligned} c_k &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(x) \cdot e^{-i \cdot k \cdot x} \, dx = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi+\frac{\pi}{k}}^{\pi+\frac{\pi}{k}} f\left(y - \frac{\pi}{k}\right) \cdot e^{-i \cdot k \cdot \left(y - \frac{\pi}{k}\right)} \, dy \\ &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f\left(y - \frac{\pi}{k}\right) \cdot e^{-i \cdot k \cdot y} \cdot \underbrace{e^{i \cdot k \cdot \frac{\pi}{k}}}_{-1} \, dy = -\frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f\left(y - \frac{\pi}{k}\right) \cdot e^{-i \cdot k \cdot y} \, dy. \end{aligned}$$

Es folgt

$$c_k = \frac{1}{2} \cdot (c_k + c_k) = \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left(f(x) - f\left(x - \frac{\pi}{k}\right) \right) \cdot e^{-i \cdot k \cdot x} \, dx,$$

also

$$|c_k| \leq \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left| f(x) - f\left(x - \frac{\pi}{k}\right) \right| \, dx. \quad (\#)$$

Für stetiges f ist nun alles klar:

$$\begin{aligned} \lim_{k \rightarrow \infty} |c_k| &\leq \lim_{k \rightarrow \infty} \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left| f(x) - f\left(x - \frac{\pi}{k}\right) \right| \, dx \\ &\stackrel{(*)}{=} \int_{-\pi}^{\pi} \lim_{k \rightarrow \infty} \left| f(x) - f\left(x - \frac{\pi}{k}\right) \right| \, dx = 0. \end{aligned}$$

(Für einen mathematisch sauberen Beweis ist die Vertauschung von Limes und Integral im Schritt $(*)$ natürlich genauer zu untersuchen.) Für eine unstetige integrierbare Funktion benutzt man, dass man integrierbare Funktionen beliebig genau durch eine stetige Funktion approximieren kann, d.h., zu jedem $\epsilon > 0$ existiert eine stetige Funktion $s(x)$ mit

$$\int_{-\pi}^{\pi} |f(x) - s(x)| \, dx < \epsilon.$$

Mit (#) folgt

$$\begin{aligned}
 |c_k| &\leq \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left| \underbrace{f(x) - s(x) + s(x)}_0 - \underbrace{s\left(x - \frac{\pi}{k}\right) + s\left(x - \frac{\pi}{k}\right)}_0 - f\left(x - \frac{\pi}{k}\right) \right| dx \\
 &\leq \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} |f(x) - s(x)| dx + \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left| s(x) - s\left(x - \frac{\pi}{k}\right) \right| dx \\
 &\quad + \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left| s\left(x - \frac{\pi}{k}\right) - f\left(x - \frac{\pi}{k}\right) \right| dx \\
 &\stackrel{(**)}{=} \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} |f(x) - s(x)| dx + \frac{1}{4 \cdot \pi} \cdot \int_{-\pi}^{\pi} \left| s(x) - s\left(x - \frac{\pi}{k}\right) \right| dx \\
 &\leq \epsilon + \text{Nullfolge}_k.
 \end{aligned}$$

(Zu Schritt (**): das dritte Integral stimmt mit dem ersten überein, da der Integrand $2 \cdot \pi$ -periodisch ist).

Zu jedem ϵ kann man damit für alle hinreichend großen k garantieren, dass $|c_k| \leq \epsilon + \epsilon$ gilt. Damit ist $|c_k|$ eine Nullfolge.

Q.E.D.

Um die punktweise Konvergenz genauer zu untersuchen, brauchen wir zunächst ein starkes technisches Hilfsmittel. Die endlichen Fourier-Reihen stellen sich als die Wirkung eines Integraloperators auf die Ausgangsfunktion dar:

Satz 11.33: (Integraldarstellung per Dirichlet-Kern)

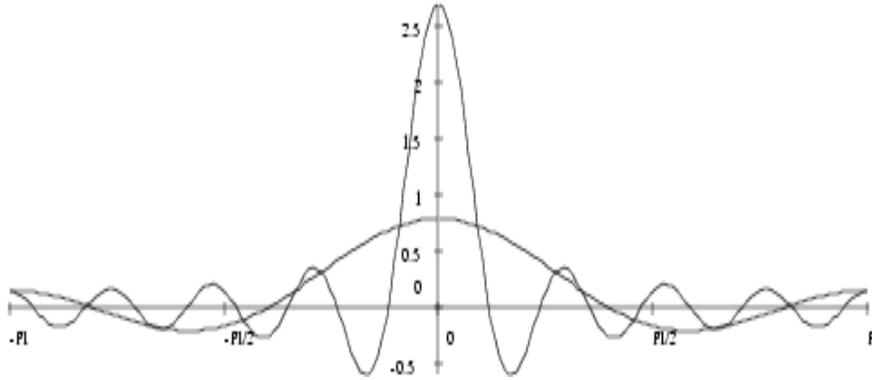
Interpretiere eine auf dem Intervall $[-\pi, \pi]$ gegebene Funktion $f \in L_1([-\pi, \pi])$ als $2 \cdot \pi$ -periodisch auf \mathbb{R} fortgesetzt (also $f(x + 2 \cdot \pi) = f(x)$). Für die Fourier-Approximation von f gilt

$$\begin{aligned}
 S_n(x) &= \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} = \int_{-\pi}^{\pi} D_n(x-t) \cdot f(t) dt = \int_{-\pi}^{\pi} D_n(t) \cdot f(x+t) dt \\
 &= \int_0^{\pi} D_n(t) \cdot (f(x+t) + f(x-t)) dt
 \end{aligned}$$

mit dem sogenannten „Dirichlet-Kern“

$$D_n(y) = \begin{cases} \frac{1}{2 \cdot \pi} \frac{\sin\left(\left(n + \frac{1}{2}\right) \cdot y\right)}{\sin\left(\frac{y}{2}\right)} & \text{für } y \neq 0 \\ \frac{n + \frac{1}{2}}{\pi} & \text{für } y = 0. \end{cases}$$

Es gilt $\int_{-\pi}^{\pi} D_n(y) dy = 1$. Hier eine Graphik von D_n mit $n = 2$ und $n = 8$:



Beweis:

$$\begin{aligned} S_n(x) &= \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} = \sum_{k=-n}^n \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(t) \cdot e^{-i \cdot k \cdot t} dt \cdot e^{i \cdot k \cdot x} \\ &= \int_{-\pi}^{\pi} f(t) \cdot \underbrace{\frac{1}{2 \cdot \pi} \cdot \sum_{k=-n}^n e^{i \cdot k \cdot (x-t)}}_{D_n(x-t)} dt. \end{aligned}$$

Es gilt

$$\begin{aligned} D_n(y) &= \frac{1}{2 \cdot \pi} \cdot \sum_{k=-n}^n e^{i \cdot k \cdot y} = \frac{1}{2 \cdot \pi} \cdot \left(\sum_{k=0}^n e^{i \cdot k \cdot y} + \sum_{k=0}^n e^{-i \cdot k \cdot y} - 1 \right) \\ &= \frac{1}{2 \cdot \pi} \cdot \left(\sum_{k=0}^n (e^{i \cdot y})^k + \sum_{k=0}^n (e^{-i \cdot y})^k - 1 \right). \end{aligned}$$

Mit der Summenformel $\sum_{k=0}^n x^k = \frac{x^{n+1} - 1}{x - 1}$ für geometrische Reihen erhält man:

$$\begin{aligned} D_n(y) &= \frac{1}{2 \cdot \pi} \cdot \left(\frac{e^{i \cdot (n+1) \cdot y} - 1}{e^{i \cdot y} - 1} + \frac{e^{-i \cdot (n+1) \cdot y} - 1}{e^{-i \cdot y} - 1} - 1 \right) \\ &= \frac{1}{2 \cdot \pi} \cdot \frac{2 \cdot \cos(n \cdot y) - 2 \cdot \cos((n+1) \cdot y)}{2 - 2 \cdot \cos(y)} \\ &\stackrel{(*)}{=} \frac{1}{2 \cdot \pi} \cdot \frac{4 \cdot \sin((n + \frac{1}{2}) \cdot y) \cdot \sin(\frac{1}{2} \cdot y)}{4 \cdot \sin^2(\frac{y}{2})} = \frac{1}{2 \cdot \pi} \cdot \frac{\sin((n + \frac{1}{2}) \cdot y)}{\sin(\frac{y}{2})}. \end{aligned}$$

Hierbei werden in (*) die üblichen Additionstheoreme auf

$$\cos(n \cdot y) - \cos((n+1) \cdot y) = \cos((n + \frac{1}{2}) \cdot y - \frac{1}{2} \cdot y) - \cos((n + \frac{1}{2}) \cdot y + \frac{1}{2} \cdot y)$$

angewendet. Weiterhin gilt mit der Substitution $\xi = x - t$

$$\int_{-\pi}^{\pi} D_n(x-t) \cdot f(t) dt = \int_{x-\pi}^{x+\pi} D_n(\xi) \cdot f(x+\xi) d\xi = \int_{-\pi}^{\pi} D_n(\xi) \cdot f(x+\xi) d\xi,$$

wobei im letzten Schritt die $2 \cdot \pi$ -Periodizität von f und D_n verwendet wird. Letztlich folgt mit $D_n(-\xi) = D_n(\xi)$:

$$\begin{aligned} \int_{-\pi}^{\pi} D_n(\xi) \cdot f(x+\xi) d\xi &= \int_{-\pi}^0 D_n(\xi) \cdot f(x+\xi) d\xi + \int_0^{\pi} D_n(\xi) \cdot f(x+\xi) d\xi \\ &= \int_0^{\pi} D_n(-\xi) \cdot f(x-\xi) d\xi + \int_0^{\pi} D_n(\xi) \cdot f(x+\xi) d\xi \\ &= \int_0^{\pi} D_n(\xi) \cdot (f(x-\xi) + f(x+\xi)) d\xi \end{aligned}$$

und

$$\int_{-\pi}^{\pi} D_n(y) dy = \int_{-\pi}^{\pi} \frac{1}{2 \cdot \pi} \cdot \sum_{k=-n}^n e^{i \cdot k \cdot y} dy = \frac{1}{2 \cdot \pi} \cdot \sum_{k=-n}^n \int_{-\pi}^{\pi} e^{i \cdot k \cdot y} dy = 1$$

(alle Terme außer $k = 0$ ergeben 0).

Q.E.D.

Hier nun der entscheidende Satz, der für die in praktischen Anwendungen relevanten Funktionen die Frage der punktweisen Konvergenz klärt:

Satz 11.34: (Das Dirichlet-Kriterium für punktweise Konvergenz)

Sei $f \in L_1([-\pi, \pi])$. Gilt an einer Stelle $x_0 \in (-\pi, \pi)$, dass sowohl der links- als auch der rechtsseitige Grenzwert

$$f(x_0 - 0) := \lim_{\substack{h \rightarrow 0 \\ h > 0}} f(x_0 - h) \quad f(x_0 + 0) := \lim_{\substack{h \rightarrow 0 \\ h > 0}} f(x_0 + h)$$

existiert und ebenfalls die links- und rechtsseitigen Ableitungen

$$\lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{f(x_0 - 0) - f(x_0 - h)}{h}, \quad \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{f(x_0 + h) - f(x_0 + 0)}{h},$$

so konvergieren die Fourier-Approximationen gegen den Mittelwert des links- und rechtsseitigen Grenzwerts:

$$\lim_{n \rightarrow \infty} S_n(x_0) = \boxed{\sum_{k=-\infty}^{\infty} c_k \cdot e^{i \cdot k \cdot x_0} = \frac{f(x_0 - 0) + f(x_0 + 0)}{2}}.$$

Dies gilt auch für $x_0 = -\pi$ und $x_0 = \pi$, wenn die $2 \cdot \pi$ -periodisch fortgesetzte Funktion f die obigen Glattheitseigenschaften hat.

Beweis (technisch): Mit $D_n(-t) = D_n(t)$ und $\int_{-\pi}^{\pi} D_n(t) dt = 1$ (aus Symmetriegründen folgt $\int_0^{\pi} D_n(t) dt = \frac{1}{2}$) ergibt sich

$$\begin{aligned}
& S_n(x_0) - \frac{f(x_0 - 0) + f(x_0 + 0)}{2} \\
&= \int_0^{\pi} D_n(t) \cdot (f(x_0 + t) + f(x_0 - t)) dt - \int_0^{\pi} D_n(t) \cdot (f(x_0 - 0) + f(x_0 + 0)) dt \\
&= \int_0^{\pi} D_n(t) \cdot ((f(x_0 + t) - f(x_0 + 0)) + (f(x_0 - t) - f(x_0 - 0))) dt \\
&= \int_0^{\pi} \frac{1}{2 \cdot \pi} \frac{\sin((n + \frac{1}{2}) \cdot t)}{\sin(\frac{t}{2})} \cdot ((f(x_0 + t) - f(x_0 + 0)) + (f(x_0 - t) - f(x_0 - 0))) dt \\
&= \int_0^{\pi} \sin((n + \frac{1}{2}) \cdot t) \cdot \underbrace{\frac{1}{\pi} \cdot \frac{t}{2} \cdot \left(\frac{f(x_0 + t) - f(x_0 + 0)}{t} + \frac{f(x_0 - t) - f(x_0 - 0)}{t} \right)}_{g(t)} dt \\
&= \int_0^{\pi} \left(\sin(n \cdot t) \cdot \cos(\frac{1}{2} \cdot t) + \cos(n \cdot t) \cdot \sin(\frac{1}{2} \cdot t) \right) \cdot g(t) dt \\
&= \frac{1}{2} \cdot \int_{-\pi}^{\pi} \sin(n \cdot t) \cdot \cos(\frac{1}{2} \cdot t) \cdot (g(t) - g(-t)) dt \\
&\quad + \frac{1}{2} \cdot \int_{-\pi}^{\pi} \cos(n \cdot t) \cdot \sin(\frac{1}{2} \cdot t) \cdot (g(t) - g(-t)) dt.
\end{aligned}$$

Die entscheidende Beobachtung hier ist, dass wegen der vorausgesetzten Existenz der einseitigen Ableitungen die Funktion $g(t)$ keine Singularität bei $t = 0$ hat und damit integrierbar ist. Mit dem Riemann–Lebesgue–Lemma 11.32 ergibt sich, dass die obige Differenz zwischen $S^*(x)$ und dem Mittelwert der einseitigen Grenzwerte eine Nullfolge ist.

Q.E.D.

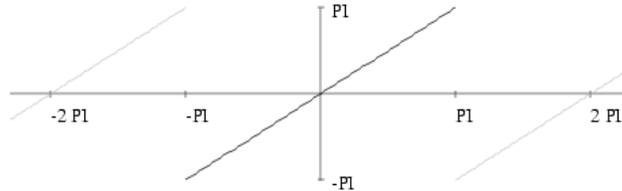
↓15.1.04

Merke 11.35:

An allen Stellen, wo die Fourier-entwickelte Funktion stetig und differenzierbar ist, wird sie durch die unendliche Fourier-Reihe dargestellt. Werden glatte (stetig und differenzierbare) Funktionen aneinander gestückt, so konvergiert die unendliche Fourier-Reihe gegen den Mittelpunkt der Grenzwerte von links- und rechts. Wenn die Stückelung stetig ist, wird auch die Nahtstelle von der Fourier-Reihe korrekt dargestellt.

Vorsicht 11.36:

Die Stetigkeit bezieht sich auf die $2 \cdot \pi$ -periodisch fortgesetzte Funktion!
 Die Funktion $f(x) = x$ ist nicht stetig an den Stellen $x = \pm\pi, \pm 3 \cdot \pi, \pm 5 \cdot \pi, \dots$, wenn man sie $2 \cdot \pi$ -periodisch fortgesetzt betrachtet!



Beispiel 11.37: In Beispiel 11.26.a) hatten wir für die „Sägezahnfunktion“

$$f(x) = \begin{cases} -1 - \frac{x}{\pi} & \text{für } x \in [-\pi, 0), \\ 1 - \frac{x}{\pi} & \text{für } x \in [0, \pi] \end{cases}$$

die Fourier-Entwicklung

$$f(x) \sim \sum_{k=1,2,\dots} \frac{\sin(k \cdot x)}{k}$$

gefunden. An allen Stetigkeitsstellen ist die Funktion differenzierbar, so dass die Fourier-Reihe an diesen Stellen gegen $f(x)$ konvergiert. An den Unstetigkeitsstellen $x = 0, \pm 2 \cdot \pi, \pm 4 \cdot \pi, \dots$ ist das Dirichlet-Kriterium aber immer noch erfüllt: Die Fourier-Reihe konvergiert gegen den Mittelwert des links- und rechtsseitigen Limes (also 0):

$$\sum_{k=1,2,\dots} \frac{\sin(k \cdot x)}{k} = \begin{cases} -1 - \frac{x}{\pi} & \text{für } x \in [-\pi, 0), \\ 0 & \text{für } x = 0 \\ 1 - \frac{x}{\pi} & \text{für } x \in (0, \pi]. \end{cases}$$

Außer an der Sprungstelle (wo wir willkürlich $f(0) = 1$ definiert hatten), wird $f(x)$ also überall punktweise durch die unendliche Fourier-Reihe dargestellt.

Bemerkung 11.38: Wir haben mit Satz 11.34 ein entscheidendes Teilargument für den Beweis von Satz 11.27 über die Vollständigkeit der trigonometrischen Funktionen im L_2 -Sinne nachgeliefert. Man kann alle Funktionen in $L_2([-\pi, \pi])$ beliebig genau durch stetige Funktionen approximieren, die das Dirichlet-Kriterium erfüllen und somit punktweise über ihre Fourier-Reihen dargestellt werden.

Bemerkung 11.39: Es ergibt sich ein verblüffendes „Lokalitätsprinzip“:

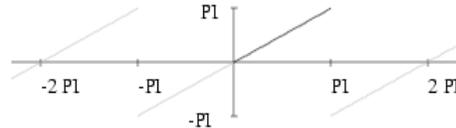
Die punktweise Konvergenz der Fourier-Reihe hängt nur von lokalen Eigenschaften der Funktion in der Umgebung des betrachteten Punktes ab!

Z.B. reichen Stetigkeit und beidseitige Differenzierbarkeit an einem Punkt für die lokale Darstellbarkeit als Fourier-Reihe an diesem Punkt. Dies steht in krassem Gegensatz zu der Tatsache, dass die Fourier-Koeffizienten global von der Funktion bestimmt werden.

Hat man vor, eine Funktion f auf einem Teilintervall $(a, b) \subset [-\pi, \pi]$ durch eine Fourier-Reihe darzustellen, so kann das auf viele verschiedene Arten und Weisen geschehen, indem man f von (a, b) **irgendwie** auf den Rest des Intervalles $[-\pi, \pi]$ fortsetzt. Abhängig von dieser Fortsetzung ergeben sich völlig unterschiedliche Fourier-Reihen, die jedoch innerhalb des Intervalls (a, b) alle gegen dieselbe Funktion konvergieren!

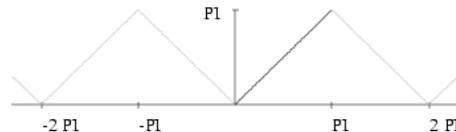
Beispiel: Wir wollen die Funktion $f(x) = x$ auf dem Intervall $[0, \pi] \subset [-\pi, \pi]$ durch eine Fourier-Reihe darstellen. Einige Möglichkeiten:

1) ungerade Fortsetzung; unstetig:



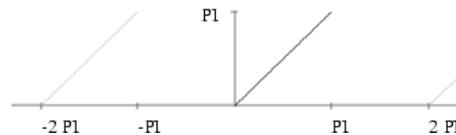
$$f_1(x) = 2 \cdot \left(\sin(x) - \frac{\sin(2 \cdot x)}{2} + \frac{\sin(3 \cdot x)}{3} - \dots \right).$$

2) gerade Fortsetzung; stetig:



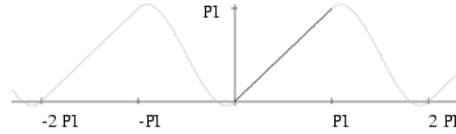
$$f_2(x) = \frac{\pi}{2} - \frac{4}{\pi} \cdot \left(\cos(x) + \frac{\cos(3 \cdot x)}{3^2} + \frac{\cos(5 \cdot x)}{5^2} + \dots \right).$$

3) unstetige Fortsetzung:



$$f_3(x) = \frac{f_1(x) + f_2(x)}{2}.$$

4) glatte Fortsetzung:



$f_4(x)$ = eine schnell konvergierende Fourier-Reihe.

11.2.4 Abklingverhalten der Fourier-Koeffizienten

Nach dem Riemann–Lebesgue–Lemma 11.32 bilden die Fourier-Koeffizienten prinzipiell eine Nullfolge. Es gilt das Prinzip: je glatter die Funktion, um so schneller fallen die Fourier-Koeffizienten. Zunächst eine einfache Beobachtung:

Satz 11.40:

Für eine $2 \cdot \pi$ -periodische differenzierbare Funktion f mit integrierbarer Ableitung f' gilt

$$c_k(f') = i \cdot k \cdot c_k(f),$$

wo $c_k(f')$ bzw. $c_k(f)$ die Fourier-Koeffizienten von f' bzw. f sind.

Beweis: Es gilt $f(-\pi) = f(\pi)$ und $e^{i \cdot k \cdot \pi} = (-1)^k = e^{-i \cdot k \cdot \pi}$. Durch partielle Integration folgt:

$$\begin{aligned} c_k(f') &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f'(x) \cdot e^{-i \cdot k \cdot x} dx \\ &= \frac{1}{2 \cdot \pi} \cdot \left[f(x) \cdot e^{-i \cdot k \cdot x} \right]_{x=-\pi}^{x=\pi} - \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(x) \cdot (-i \cdot k) \cdot e^{-i \cdot k \cdot x} dx \\ &= \frac{f(\pi) \cdot e^{-i \cdot k \cdot \pi} - f(-\pi) \cdot e^{i \cdot k \cdot \pi}}{2 \cdot \pi} + i \cdot k \cdot c_k(f) = i \cdot k \cdot c_k(f). \end{aligned}$$

Q.E.D.

Diese Aussage ist leicht zu merken: stellt man sich die Funktion direkt als Fourier-Reihe gegeben vor, so kann man die Summanden der Reihe einzeln differenzieren:

$$\frac{d}{dx} \underbrace{\sum_{k=-\infty}^{\infty} c_k \cdot e^{i \cdot k \cdot x}}_{f(x)} = \sum_{k=-\infty}^{\infty} c_k \cdot \frac{d}{dx} e^{i \cdot k \cdot x} = \sum_{k=-\infty}^{\infty} \underbrace{c_k \cdot (i \cdot k)}_{c_k(f')} \cdot e^{i \cdot k \cdot x}.$$

Bemerkung 11.41: So trivial der Zusammenhang zwischen den Fourier-Koeffizienten von f und f' mathematisch auch sein mag, er ist für Anwendungen von zentraler Bedeutung:

Das Ableiten wird für die Fourier-Koeffizienten zu einer algebraischen Operation (Multiplikation mit $i \cdot k$). Fourier-Ansätze verwandeln daher Differentialgleichungen in algebraische Gleichungen für die Fourier-Koeffizienten.

Satz 11.42: (Abklingverhalten der Fourier-Koeffizienten)

Ist eine $2 \cdot \pi$ -periodische Funktion p -fach differenzierbar und ist die p -te Ableitung integrierbar, so gilt für die Fourier-Koeffizienten der Funktion:

$$|c_k| = o\left(\frac{1}{|k|^p}\right).$$

Zur Erinnerung: Das „Landau-Symbol“ o bedeutet $\lim_{|k| \rightarrow \infty} |k|^p \cdot |c_k| = 0$.

Merke: Je glatter die Funktion (d.h., je öfter differenzierbar), um so schneller konvergieren die Fourier-Koeffizienten gegen 0.

Beweis: Seien $c_k(f^{(p)})$ die Fourier-Koeffizienten der p -ten Ableitung. Nach Satz 11.40 gilt

$$c_k(f^{(p)}) = i \cdot k \cdot c_k(f^{(p-1)}) = (i \cdot k)^2 \cdot c_k(f^{(p-2)}) = \dots = (i \cdot k)^p \cdot c_k(f).$$

Nach dem Riemann-Lebesgue-Lemma 11.32 bilden die Fourier-Koeffizienten $c_k(f^{(p)})$ der p -ten Ableitung eine Nullfolge für $|k| \rightarrow \infty$, d.h., die Fourier-Koeffizienten $c_k = c_k(f)$ fallen schneller ab als $1/|k|^p$:

$$c_k(f) = \frac{c_k(f^{(p)})}{(i \cdot k)^p} = o\left(\frac{1}{|k|^p}\right).$$

Q.E.D.

Bemerkung 11.43: In Satz 11.42 wird das Abklingverhalten der Fourier-Koeffizienten mittels des Landau-Symbols o recht unspezifisch als „schneller abfallend als $1/|k|^p$ “ beschrieben. Die in der Praxis auftretenden Funktionen sind meist stückweise aus glatten Anteilen zusammengesetzt, wodurch sich in der Praxis meist „schneller abfallend als $1/|k|^p$ “ in der Form

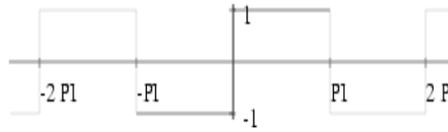
„um eine k -Potenz schneller abfallend als $1/|k|^p$ “

ergibt. Es gilt die Faustregel:

$$\begin{aligned}
 f \text{ unstetig:} & \quad c_k \sim \frac{1}{k}. \\
 f \text{ stetig, } f' \text{ unstetig:} & \quad c_k \sim \frac{1}{k^2}. \\
 f' \text{ stetig, } f'' \text{ unstetig:} & \quad c_k \sim \frac{1}{k^3}. \\
 f^{(p)} \text{ stetig, } f^{(p+1)} \text{ unstetig:} & \quad c_k \sim \frac{1}{k^{p+2}}.
 \end{aligned}$$

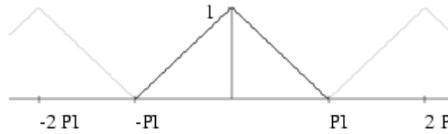
Beispiele:

f_1 unstetig:



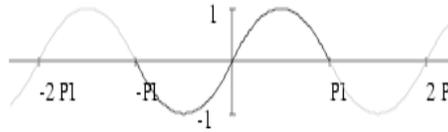
$$f_1(x) = \frac{4}{\pi} \cdot \left(\sin(x) + \frac{\sin(3 \cdot x)}{3} + \frac{\sin(5 \cdot x)}{5} + \dots \right).$$

f_2 stetig, f_2' unstetig:



$$f_2(x) = 1 + \frac{4}{\pi^2} \cdot \left(\cos(x) + \frac{\cos(3 \cdot x)}{3^2} + \frac{\cos(5 \cdot x)}{5^2} + \dots \right).$$

f_3' stetig, f_3'' unstetig:



$$f_3(x) = \frac{4}{\pi^2} \cdot x \cdot (\pi - |x|) = \frac{32}{\pi^3} \cdot \left(\frac{\sin(x)}{1^3} + \frac{\sin(3 \cdot x)}{3^3} + \frac{\sin(5 \cdot x)}{5^3} + \dots \right).$$

Bemerkung 11.44: Für eine durch ihre Fourier-Reihe dargestellte Funktion gilt für den Approximationsfehler durch endliche Fourier-Approximationen

$$\begin{aligned}
 f(x) - \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} &= \sum_{|k|>n} c_k \cdot e^{i \cdot k \cdot x} \\
 \Rightarrow \left| f(x) - \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot x} \right| &\leq \sum_{|k|>n} |c_k \cdot e^{i \cdot k \cdot x}| = \sum_{|k|>n} |c_k|.
 \end{aligned}$$

Je schneller die Fourier-Koeffizienten abfallen, umso schneller die punktweise Konvergenz an allen Stellen!

11.2.5 Das Gibbs'sche Phänomen

Wir betrachten das Konvergenzverhalten der endlichen Fourier-Transformation für die „Sprungfunktion“ („Heaviside-Funktion“):

$$H(x) = \begin{cases} 0 & \text{für } x \in [-\pi, 0), \\ 1 & \text{für } x \in [0, \pi). \end{cases}$$

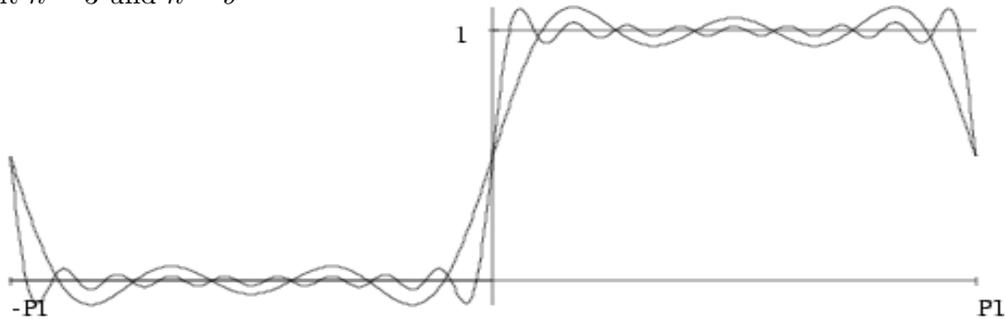
Als zugeordnete Fourier-Reihe berechnet man

$$H(x) = \frac{1}{2} + \frac{2}{\pi} \cdot \left(\sin(x) + \frac{\sin(3 \cdot x)}{3} + \frac{\sin(5 \cdot x)}{5} + \dots \right),$$

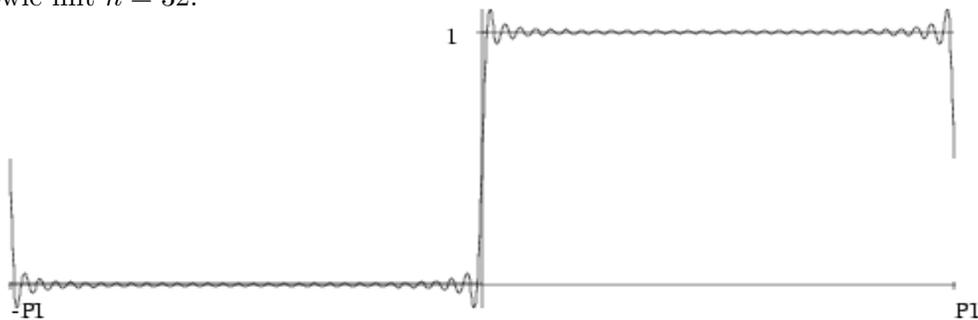
welche in der Tat $H(x)$ überall außer an den Sprungstellen punktweise darstellt. Hier die endlichen Fourier-Approximationen

$$S_{2 \cdot n - 1}(x) = \frac{1}{2} + \frac{2}{\pi} \cdot \sum_{m=0}^{n-1} \frac{\sin((2 \cdot m + 1) \cdot x)}{2 \cdot m + 1}$$

mit $n = 3$ und $n = 9$



sowie mit $n = 32$:



Das Gibbs'sche² Phänomen ist, dass die Fourier-Approximationen an Sprungstellen ein Überschwingverhalten zeigen, dass mit wachsendem n nicht verschwindet: das erste Maximum von $S_n(x)$ für $x > 0$ liegt für jedes n deutlich über der Grenzfunktion $H(x) \equiv 1$ (für $x > 0$). Dies lässt sich unschwer explizit nachrechnen. Das betrachtete Maximum ist durch die erste positive Nullstelle der Ableitung bestimmt:

$$\begin{aligned} \frac{d}{dx} S_{2 \cdot n - 1}(x) &= \frac{2}{\pi} \cdot \sum_{m=0}^{n-1} \cos((2 \cdot m + 1) \cdot x) = \frac{2}{\pi} \cdot \Re \left(\sum_{m=0}^{n-1} e^{i \cdot (2 \cdot m + 1) \cdot x} \right) \\ &= \frac{2}{\pi} \cdot \Re \left(e^{i \cdot x} \cdot \sum_{m=0}^{n-1} (e^{i \cdot 2 \cdot x})^m \right) = \frac{2}{\pi} \cdot \Re \left(e^{i \cdot x} \cdot \frac{e^{i \cdot 2 \cdot n \cdot x} - 1}{e^{i \cdot 2 \cdot x} - 1} \right) = \frac{2}{\pi} \cdot \Re \left(\frac{e^{i \cdot 2 \cdot n \cdot x} - 1}{e^{i \cdot x} - e^{-i \cdot x}} \right) \\ &= \frac{2}{\pi} \cdot \Re \left(\frac{e^{i \cdot 2 \cdot n \cdot x} - 1}{2 \cdot i \cdot \sin(x)} \right) = \frac{1}{\pi \cdot \sin(x)} \cdot \Im(e^{i \cdot 2 \cdot n \cdot x} - 1) = \frac{\sin(2 \cdot n \cdot x)}{\pi \cdot \sin(x)}. \end{aligned}$$

Die erste positive Nullstelle der Ableitung ist demnach $\xi_0 = \pi/(2 \cdot n)$. Der Wert von $S_{2 \cdot n - 1}$ an dieser Stelle ist

$$S_{2 \cdot n - 1}(\xi_0) = \frac{1}{2} + \frac{2}{\pi} \cdot \sum_{m=0}^{n-1} \frac{\sin((2 \cdot m + 1) \cdot \frac{\pi}{2 \cdot n})}{2 \cdot m + 1} = \frac{1}{2} + \frac{1}{n} \cdot \sum_{m=0}^{n-1} \frac{\sin(\frac{(2 \cdot m + 1) \cdot \pi}{2 \cdot n})}{\frac{(2 \cdot m + 1) \cdot \pi}{2 \cdot n}}.$$

Die Summe lässt sich leicht interpretieren. Betrachte dazu das Integral

$$\frac{1}{\pi} \cdot \int_0^\pi \frac{\sin(y)}{y} dy \approx 0.5895.$$

Zerlege das Integrationsintervall $[0, \pi]$ durch die n Stützstellen $y_m = \frac{(2 \cdot m + 1) \cdot \pi}{2 \cdot n}$, ($m = 0, \dots, n - 1$) in äquidistante Intervalle der Länge π/n . Für $n \rightarrow \infty$ konvergieren die Riemann-Summen gegen das Integral:

$$\frac{1}{n} \cdot \sum_{m=0}^{n-1} \frac{\sin(\frac{(2 \cdot m + 1) \cdot \pi}{2 \cdot n})}{\frac{(2 \cdot m + 1) \cdot \pi}{2 \cdot n}} = \frac{1}{\pi} \cdot \sum_{m=0}^{n-1} \frac{\sin(y_m)}{y_m} \cdot \frac{\pi}{n} \xrightarrow{(n \rightarrow \infty)} \frac{1}{\pi} \cdot \int_0^\pi \frac{\sin(y)}{y} dy.$$

Die Höhe des ersten positiven Maximums der Fourier-Approximation konvergiert damit für $n \rightarrow \infty$:

$$S_{2 \cdot n - 1}(\xi_0) \approx \frac{1}{2} + \frac{1}{\pi} \cdot \int_0^\pi \frac{\sin(y)}{y} dy \approx 1.0895 \dots$$

²J.W. Gibbs, 1839 – 1903, amerikanischer Mathematiker/Physiker. Ein Zitat von ihm:

„A mathematician may say anything he pleases, but a physicist must be at least partially sane.“

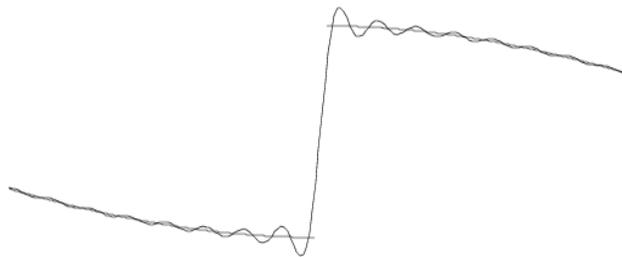
Betrachte noch einmal $H(x)$ und beachte, dass $H(x) - \frac{1}{2}$ eine ungerade Funktion ist. Aus Symmetriegründen nehmen daher alle Fourier-Approximationen $S_{2 \cdot n - 1}$ als letztes negatives Minimum vor der Sprungstelle bei 0 den Wert $\frac{1}{2} - 0.5895 = -0.0895$ an. Damit hat die Differenz des ersten Maximums rechts vom Sprung zum letzten Minimum links vom Sprung den Wert $1 + 2 \cdot 0.0895 = 1.179$, wobei die Sprunghöhe 1 ist. Multipliziert man die Sprungfunktion mit einem Faktor, so skalieren sich die Fourier-Koeffizienten mit diesem Faktor und die Fourier-Approximationen ebenfalls. Damit gilt:

Das Gibbs'sche Phänomen: An einer Sprungstelle ist für alle Fourier-Approximationen die Differenz zwischen dem benachbarten Minimum und Maximum, welche den Sprung einschließen, um etwa 17.9% größer als die Sprunghöhe.

In der Tat gilt dies nicht nur für die hier betrachtete Heaviside-Funktion: eine beliebige Funktion, die an einer Stelle x_0 eine Sprung macht, kann immer in der Form

$$\text{Funktion}(x) = \text{bei } x_0 \text{ stetige Funktion}(x) + \text{Sprunghöhe} \cdot H(x - x_0)$$

zerlegt werden. Ist die stetige Funktion hinreichend glatt, konvergiert die Fourier-Approximation des stetigen Anteils deutlich schneller gegen den stetigen Anteil als die Fourier-Approximation des Sprunganteils gegen den Sprunganteil. Der stetige Anteil kann daher im Approximationsfehler vernachlässigt werden gegen den Approximationsfehler von $H(x - x_0)$. Das Gibbs'sche Phänomen für H vererbt sich damit auf die Sprungstellen beliebig zusammengestückelter glatter Funktionen:



11.2.6 Beliebige Perioden

Bislang hatten wir die Fourier-Zerlegung von auf $[-\pi, \pi]$ gegebenen Funktionen betrachtet, die man sich als $2 \cdot \pi$ -periodisch fortgesetzt vorstellen sollte. Die Einschränkung auf das Standardintervall $[-\pi, \pi]$ geschah dabei lediglich aus schreibtechnischen Gründen, da alle Formeln hierfür besonders einfach werden.

Wir betrachten nun ein beliebiges Intervall $[-\frac{L}{2}, \frac{L}{2}]$, auf dem eine L -periodische

Funktion $f(x) = f(x+L)$ gegeben ist. In Analogie zur Definition 11.21 betrachtet man das Orthogonalsystem

$$\left\{ e^{i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x} \right\}_{k \in \mathbb{Z}},$$

die entsprechenden Fourier-Koeffizienten

$$c_k = \frac{\langle e^{i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x}, f \rangle}{\langle e^{i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x}, e^{i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x} \rangle} = \frac{1}{L} \cdot \int_{-\frac{L}{2}}^{\frac{L}{2}} f(x) \cdot e^{-i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x} dx$$

und die entsprechenden Fourier-Approximationen

$$S_n(x) = \sum_{k=-n}^n c_k \cdot e^{i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x}.$$

Diese Definitionen lassen sich folgendermassen interpretieren:

Zu $f : [-\frac{L}{2}, \frac{L}{2}] \rightarrow \mathbb{C}$ betrachte die Funktion $g : [-\pi, \pi] \rightarrow \mathbb{C}$, die durch eine Transformation $x \rightarrow \frac{L}{2 \cdot \pi} \cdot x$ aus f entsteht:

$$g(x) = f\left(\frac{L}{2 \cdot \pi} \cdot x\right), \quad f(x) = g\left(\frac{2 \cdot \pi}{L} \cdot x\right).$$

Diese Funktion ist $2 \cdot \pi$ -periodisch:

$$g(x + 2 \cdot \pi) = f\left(\frac{L}{2 \cdot \pi} \cdot (x + 2 \cdot \pi)\right) = f\left(\frac{L}{2 \cdot \pi} \cdot x + L\right) = f\left(\frac{L}{2 \cdot \pi} \cdot x\right) = g(x).$$

Zerlegt man g den vorangegangenen Kapiteln gemäß in eine Fourier-Reihe über dem Grundintervall $[-\pi, \pi]$, so ergeben sich nach Rücktransformation auf das Intervall $[-\frac{L}{2}, \frac{L}{2}]$ die obigen Darstellungen der Fourier-Koeffizienten und Fourier-Approximationen für f .

Damit ist unmittelbar klar, dass alle Strukturaussagen (Vollständigkeit bzgl. der L_2 -Norm, punktweise Konvergenz bei Dirichlet-Bedingungen, Gibbs'sches Phänomen etc.) auch für allgemeine Perioden gelten.

Bemerkung 11.45: Läßt man die Länge des Intervalls $[-\frac{L}{2}, \frac{L}{2}]$ gegen ∞ laufen, erhält man statt der Fourier-Koeffizienten

$$c_k = \frac{1}{L} \cdot \int_{-\frac{L}{2}}^{\frac{L}{2}} f(x) \cdot e^{-i \cdot k \cdot \frac{2 \cdot \pi}{L} \cdot x} dx$$

im Grenzwert $L \rightarrow \infty$ eine Funktion $\hat{f}\left(\frac{k \cdot 2 \cdot \pi}{L}\right) = \frac{L}{\sqrt{2 \cdot \pi}} \cdot c_k$ des kontinuierlichen „Frequenzparameters“ ω , der immer dichter beieinander liegenden Punkten $\omega = k \cdot 2 \cdot \pi / L$ (mit $k \in \mathbb{Z}$) entspricht. Diese Funktion

$$\hat{f}(\omega) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\infty} f(x) \cdot e^{-i \cdot \omega \cdot x} dx$$

nennt man die „**Fourier-Integraltransformation**“ von f . Ist f hinreichend glatt (z.B. Dirichlet-Bedingungen wie in Satz 11.34), so ist f aus der Fourier-Transformierten $\hat{f}(\omega)$ rekonstruierbar. Die Summen der Fourier-Approximation kann man dabei als Riemann-Summen interpretieren, aus denen im Grenzwert $L \rightarrow \infty$ ein Integral wird:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\infty} \hat{f}(\omega) \cdot e^{i\omega \cdot x} d\omega.$$

Kapitel 12

Lineare Algebra

12.1 Vektorräume

↓20.1.04

Definition 12.1: (Vektorräume)

Ein **Vektorraum** V über dem **Skalarenkörper** K ist eine nichtleere Menge von Objekten (**Vektoren** genannt). Es muss eine Addition

$$v_1, v_2 \in V \Rightarrow v_1 + v_2 \in V$$

mit den Eigenschaften

$$v_1 + v_2 = v_2 + v_1 \quad (\text{Kommutativität}),$$

$$(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3) \quad (\text{Assoziativität})$$

für alle $v_i \in V$ gegeben sein, es muss ein neutrales Element 0 (der Nullvektor) mit $v + 0 = v \forall v \in V$ existieren, zu jedem $v \in V$ muss es einen negativen Vektor $-v$ mit $v + (-v) = 0$ geben.

Weiterhin muss jedes $v \in V$ mit jedem „Skalar“ $\lambda \in K$ multiplizierbar sein, so dass $\lambda \cdot v$ wieder in V liegt und

$$\lambda_1 \cdot (\lambda_2 \cdot v) = (\lambda_1 \cdot \lambda_2) \cdot v \quad (\text{Assoziativität}),$$

$$1 \cdot v = v \quad (\text{neutrales Element der Multiplikation}),$$

$$(\lambda_1 + \lambda_2) \cdot v = \lambda_1 \cdot v + \lambda_2 \cdot v \quad (\text{Distributivität})$$

für alle $\lambda_1, \lambda_2 \in K$ und $v \in V$ gilt. Hierbei sei 1 das neutrale Element der Multiplikation im Körper K .

Bemerkung 12.2: Als Skalarenkörper K interessiert uns hier nur \mathbb{R} oder \mathbb{C} , wir sprechen dann von einem „reellen“ oder einem „komplexen“ Vektorraum.

Beispiel 12.3: 1) Der Raum¹

$$\mathbb{R}^n := \left\{ (x_1, x_2, \dots, x_n)^T ; x_1, \dots, x_n \in \mathbb{R} \right\}$$

bzw.

$$\mathbb{C}^n := \left\{ (x_1, x_2, \dots, x_n)^T ; x_1, \dots, x_n \in \mathbb{C} \right\}$$

mit den Operationen

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

und

$$\lambda \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda \cdot x_1 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix}$$

ist ein reeller bzw. komplexer Vektorraum.

2) Die Menge

$$\left\{ \lambda_1 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} ; \lambda_1, \lambda_2 \in \mathbb{R} \right\} \subset \mathbb{R}^3$$

mit den Operationen des \mathbb{R}^3 (Addition und skalare Multiplikation) ist ein in den \mathbb{R}^3 eingebetteter Vektorraum. Man nennt diese auch einen „**Unterraum**“ des \mathbb{R}^3 . Es handelt sich um eine den Ursprung enthaltende Ebene im \mathbb{R}^3 . Man beachte, dass die Summe zweier Ebenenvektoren wieder in der Ebene liegt.

3) Der Raum der stetigen Funktionen

$$C([0, 1]) = \{ f : [0, 1] \rightarrow \mathbb{R}; f \text{ ist eine stetige Funktion} \}$$

mit den Operationen der punktweisen Addition

$$f + g : x \rightarrow f(x) + g(x)$$

und der Multiplikation

$$\lambda \cdot f : x \rightarrow \lambda \cdot f(x)$$

¹Vektoren des \mathbb{R}^n bzw. \mathbb{C}^n sind bei uns immer **Spalten**. Aus Platzgründen verwendet man im Text lieber eine Zeilendarstellung (x_1, \dots, x_n) . Damit die Bezeichnungen konsistent sind, wird das Symbol T (Transposition)

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

benutzt, das anzeigt, dass dieser Vektor als Spalte zu interpretieren ist.

ist ein reeller Vektorraum (man beachte, dass die Summe stetiger Funktionen wieder stetig ist und dass mit f auch $\lambda \cdot f$ stetig ist).

Notation 12.4:

Für allgemeine Vektorräume V benutzen wir die Notation $v \in V$ ohne Vektorpfeil über dem Vektor v . Für $V = \mathbb{R}^n$ bzw. $V = \mathbb{C}^n$ schreiben wir Vektoren mit einem Vektorpfeil (Physikernotation), also $\vec{v} \in \mathbb{R}^n$ bzw. $\vec{v} \in \mathbb{C}^n$.

Definition 12.5: (lineare Unabhängigkeit)

Eine Menge von Vektoren v_1, \dots, v_n eines Vektorraums heißt „**linear unabhängig**“, wenn eine Linearkombination $\sum_{i=1}^n \lambda_i \cdot v_i$ nur dann der Nullvektor sein kann, wenn alle „Linearkoeffizienten“ λ_i verschwinden:

$$\sum_{i=1}^n \lambda_i \cdot v_i = 0 \quad \Rightarrow \quad \lambda_1 = \dots = \lambda_n = 0.$$

Sind Vektoren nicht linear unabhängig, nennt man sie „**linear abhängig**“.

Bemerkung 12.6: In einer Menge linear unabhängiger Vektoren $\{v_1, \dots, v_n\}$ kann keiner der Vektoren 0 sein. Angenommen, $v_1 = 0$. Dann wäre

$$\lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n = 0$$

für beliebiges λ_1 und $\lambda_2 = \dots = \lambda_n = 0$ erfüllt.

Beispiel 12.7: Die Vektoren

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

sind linear unabhängige Vektoren im \mathbb{R}^3 , denn aus

$$\sum_{i=1}^3 \lambda_i \cdot \vec{v}_i = \lambda_1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \lambda_3 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \stackrel{(!)}{=} \vec{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

folgt sofort $\lambda_1 = \lambda_2 = \lambda_3 = 0$. Die Vektoren

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

hingegen sind nicht linear unabhängig, denn z.B. die Linearkombination

$$\vec{v}_1 + \vec{v}_2 - \vec{v}_3 = \vec{0}$$

verschwindet, obwohl die Linearfaktoren $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -1$ nicht verschwinden.

Definition 12.8: (von Vektoren aufgespannte Räume)

Seien v_1, \dots, v_n Vektoren eines Vektorraums über dem Skalarenkörper K .
Die Menge der Vektoren

$$\text{span}(v_1, \dots, v_n) = \left\{ \sum_{i=1}^n \lambda_i \cdot v_i ; \lambda_1, \dots, \lambda_n \in K \right\}$$

nennt man „den von den Vektoren v_1, \dots, v_n aufgespannten (Unter-)Raum“. Bezeichnung: die Vektoren v_1, \dots, v_n bilden ein „Erzeugendensystem“ des (Unter-)Raums.

Definition 12.9: (Basis eines Vektorraums)

Eine Menge $\{v_1, \dots, v_n\}$ von Vektoren eines Vektorraums V heißt „Basis“, wenn die Vektoren linear unabhängig sind und $V = \text{span}(v_1, \dots, v_n)$ gilt.

Satz und Definition 12.10:

Besitzt ein Vektorraum eine Basis mit endlich vielen Elementen, so hat jede alternative Basis die selbe Anzahl von Elementen. Diese Anzahl von Basiselementen wird die „Dimension“ des Vektorraums genannt.

Beweis: ist technisch (z.B. über den „Steinitzchen Austauschatz“). Siehe Bücher über Lineare Algebra.

Satz 12.11:

- a) Seien $\{v_1, \dots, v_m\}$ linear unabhängig. Für jeden Vektor v , der sich nicht als Linearkombination von v_1, \dots, v_m schreiben läßt, sind die Vektoren $\{v_1, \dots, v_m, v\}$ linear unabhängig.
- b) In einem Raum der Dimension n bildet jede Menge mit n linear unabhängigen Vektoren eine Basis.
- c) In einem Raum der Dimension n gibt es maximal n linear unabhängige Vektoren.

Beweis: (für technisch Interessierte)

a) Sei $\{v_1, \dots, v_m\}$ eine beliebige Menge linear unabhängiger Vektoren. Sei v ein Vektor, der nicht Linearkombination von v_1, \dots, v_m ist, d.h., für jede Wahl von skalaren Linearfaktoren μ_1, \dots, μ_m gilt

$$v - \sum_{i=1}^m \mu_i \cdot v_i \neq 0. \quad (\#)$$

Betrachte nun die Vektorgleichung

$$\lambda_0 \cdot v + \sum_{i=1}^m \lambda_i \cdot v_i = 0. \quad (\#\#)$$

Es folgt, dass $\lambda_0 = 0$ gelten muss, denn sonst würde

$$v - \sum_{i=1}^m \left(-\frac{\lambda_i}{\lambda_0} \right) \cdot v_i = 0$$

folgen, was im Widerspruch zu $(\#)$ steht. Damit wird aus $(\#\#)$ die Gleichung $\sum_{i=1}^m \lambda_i \cdot v_i = 0$, woraus wegen der linearen Unabhängigkeit von v_1, \dots, v_m folgt, dass $\lambda_1 = \dots = \lambda_m = 0$ gilt. Also impliziert $(\#\#)$ insgesamt $\lambda_0 = \lambda_1 = \dots = \lambda_m = 0$, d.h., die $m + 1$ Vektoren v_1, \dots, v_m, v sind linear unabhängig.

b) Seien v_1, \dots, v_n linear unabhängige Vektoren. Sei $\{e_1, \dots, e_n\}$ irgendeine Basis. Seien e_{j_1}, \dots, e_{j_m} die Basisvektoren, die sich nicht als Linearkombination von v_1, \dots, v_n schreiben lassen. Nach a) sind $\{v_1, \dots, v_n, e_{j_1}, \dots, e_{j_m}\}$ linear unabhängig. Jeder Vektor v des Vektorraums kann als Linearkombination der Basiselemente e_i geschrieben werden. Jedes e_i läßt sich aber nach Konstruktion als Linearkombination von $v_1, \dots, v_n, e_{j_1}, \dots, e_{j_m}$ ausdrücken, so dass sich insgesamt jedes v als Linearkombination der Vektoren $v_1, \dots, v_n, e_{j_1}, \dots, e_{j_m}$ darstellen läßt. Damit ist $\text{span}(v_1, \dots, v_n, e_{j_1}, \dots, e_{j_m})$ der gesamte Vektorraum, also ist $\{v_1, \dots, v_n, e_{j_1}, \dots, e_{j_m}\}$ eine Basis mit $n + m$ Elementen. Dies darf nach Satz 12.10 aber nur für $m = 0$ sein, d.h., die n linear unabhängigen Vektoren v_1, \dots, v_n bildeten schon eine Basis.

c) Von m linear unabhängigen Vektoren (mit $m > n$) bilden die ersten n Vektoren nach b) schon eine Basis, der $(n + 1)$ -te Vektor kann damit als Linearkombination der ersten n geschrieben werden und ist damit nicht linear unabhängig von den ersten n .

Q.E.D.

Beispiel 12.12: 1) Der reelle Raum

$$\mathbb{R}^n = \left\{ (x_1, x_2, \dots, x_n)^T ; x_1, \dots, x_n \in \mathbb{R} \right\}$$

und der komplexe Raum

$$\mathbb{C}^n = \left\{ (x_1, x_2, \dots, x_n)^T ; x_1, \dots, x_n \in \mathbb{C} \right\}$$

haben jeweils die Dimension n . Die „**Standardbasis**“ besteht jeweils aus den Vektoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Beachte dazu

$$\sum_{i=1}^n x_i \cdot \vec{e}_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Jede Menge von n linear unabhängigen Vektoren bildet jeweils eine alternative Basis.

2) Seien \vec{v}_1, \vec{v}_2 zwei linear unabhängige Vektoren im \mathbb{R}^n (mit $n \geq 2$). Sie bilden eine Basis der 2-dimensionalen Ebene

$$E = \text{span}(\vec{v}_1, \vec{v}_2) = \left\{ \lambda_1 \cdot \vec{v}_1 + \lambda_2 \cdot \vec{v}_2 ; \lambda_1, \lambda_2 \in \mathbb{R} \right\}.$$

Eine alternative Basis des Vektorraums E (Unterraum des \mathbb{R}^n) ist z.B.

$$\vec{v}'_1 = \vec{v}_1 + \vec{v}_2, \vec{v}'_2 = \vec{v}_1 - \vec{v}_2,$$

denn \vec{v}'_1, \vec{v}'_2 sind wieder linear unabhängig und jeder von \vec{v}_1, \vec{v}_2 aufgespannte Vektor läßt sich auch von \vec{v}'_1, \vec{v}'_2 aufspannen:

$$\vec{v} = \lambda_1 \cdot \vec{v}_1 + \lambda_2 \cdot \vec{v}_2 = \left(\frac{\lambda_1}{2} + \frac{\lambda_2}{2} \right) \cdot \vec{v}'_1 + \left(\frac{\lambda_1}{2} - \frac{\lambda_2}{2} \right) \cdot \vec{v}'_2.$$

3) Sei V die Menge aller Polynome vom Grad $\leq n$. In Blatt 13 ist zu zeigen, dass die Monome $\{1, x, x^2, \dots, x^n\}$ eine Basis dieses $n + 1$ dimensionalen Vektorraums sind.

4) Sei V die Menge aller Polynome (mit beliebigem Grad). Die Dimension dieses Raums ist ∞ , denn es gibt beliebig viele linear unabhängige Elemente (z.B. die Monome x^n mit beliebig großem n).

Für analytische Betrachtungen auf dem \mathbb{R}^n bzw. dem \mathbb{C}^n braucht man einen Längenbegriff, mit dem man den Abstand $\|\vec{v} - \vec{w}\|$ zwischen Vektoren messen kann:

Definition 12.13: (Norm)

Sei V ein Vektorraum über dem Skalarenkörper K . Eine „Norm“ auf V ist eine Abbildung $\|\cdot\| : V \rightarrow [0, \infty)$ mit den Eigenschaften:

- a) $\|v + w\| \leq \|v\| + \|w\|$ für alle $v, w \in V$ („Dreiecksungleichung“).
- b) $\|\lambda \cdot v\| = |\lambda| \cdot \|v\|$ für alle $v \in V$, $\lambda \in K$ („Homogenität“).
- c) $\|v\| = 0 \Leftrightarrow v = 0$.

Beispiel 12.14: Auf dem \mathbb{R}^n bzw. dem \mathbb{C}^n existieren die sogenannten p -Normen:

$$\left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad p \in [1, \infty).$$

Die Norm-Eigenschaften b) und c) sind offensichtlich, der Beweis der Dreiecksungleichung

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p$$

ist jedoch nicht-trivial. Zusätzlich kann man die Norm

$$\left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_\infty = \max_{i=1, \dots, n} |x_i|$$

betrachten. Die Namensgebung entstammt der Tatsache, dass

$$\|\vec{x}\|_\infty = \lim_{p \rightarrow \infty} \|\vec{x}\|_p$$

für jedes $\vec{x} \in \mathbb{R}^n$ bzw. $\vec{x} \in \mathbb{C}^n$ gilt.

Eine wichtige Struktur auf dem \mathbb{R}^n bzw. dem \mathbb{C}^n ist das Skalarprodukt:

Definition 12.15: (Skalarprodukt)

Ein „Skalarprodukt“ auf einem Vektorraum V über dem Skalarenkörper K ist eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$ mit den Eigenschaften

a) *Linearität:*

$$\begin{aligned} \langle v, \lambda_1 \cdot w_1 + \lambda_2 \cdot w_2 \rangle &= \lambda_1 \cdot \langle v, w_1 \rangle + \lambda_2 \cdot \langle v, w_2 \rangle, \\ \langle \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2, w \rangle &= \overline{\lambda_1} \cdot \langle v_1, w \rangle + \overline{\lambda_2} \cdot \langle v_2, w \rangle, \end{aligned}$$

für alle $v, v_1, v_2, w, w_1, w_2 \in V$, $\lambda_1, \lambda_2 \in K$.

b) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ (komplexe Konjugation) für alle $v, w \in V$.

c) $\langle v, v \rangle \geq 0$ für alle $v \in V$. (Beachte: b) impliziert $\langle v, v \rangle \in \mathbb{R}$.)

d) $\langle v, v \rangle = 0 \Leftrightarrow v = 0$.

Beispiel 12.16: 1) Auf dem \mathbb{R}^n ist das interessanteste Skalarprodukt

$$\vec{x} \cdot \vec{y} = \left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{i=1}^n x_i \cdot y_i.$$

Offensichtlich sind alle geforderten Eigenschaften a) – d) des Skalarprodukts erfüllt. Die komplexe Konjugation ist dabei irrelevant.

2) Auf dem \mathbb{C}^n ist das interessanteste Skalarprodukt

$$\vec{x} \cdot \vec{y} = \left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{i=1}^n \bar{x}_i \cdot y_i.$$

Offensichtlich sind alle geforderten Eigenschaften a) – d) des Skalarprodukts erfüllt, wobei die komplexe Konjugation wesentlich ist.

Definition 12.17: (Orthogonalität)

Zwei Vektoren \vec{x}, \vec{y} im \mathbb{R}^n bzw. \mathbb{C}^n heißen „**orthogonal**“, wenn $\langle \vec{x}, \vec{y} \rangle = 0$ gilt.

Bemerkung 12.18: Offensichtlich erzeugt das Skalarprodukt die p -Norm aus Bemerkung 12.14 mit $p = 2$:

$$\|\vec{x}\|_2 = \sqrt{\langle \vec{x}, \vec{x} \rangle}.$$

Es gilt die Cauchy-Schwarzsche Ungleichung

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|_2 \cdot \|\vec{y}\|_2,$$

die wir schon in Satz 11.5 kennengelernt haben (für den Beweis siehe Aufgabe 99 auf Blatt 12). Das Gleichheitszeichen gilt in der Cauchy-Schwarzschen Ungleichung genau dann, wenn die Vektoren \vec{x} und \vec{y} parallel (= linear abhängig) sind.

12.2 Etwas Geometrie im \mathbb{R}^3

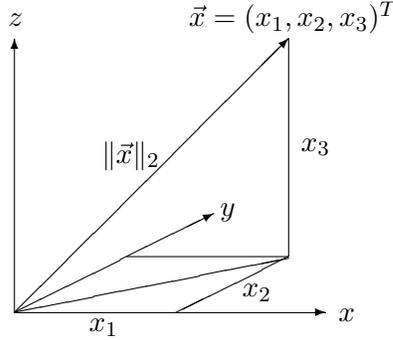
12.2.1 Skalar-, Vektor- und Spatprodukt

Auf dem \mathbb{R}^3 (und dem darin eingebetteten \mathbb{R}^2) gelten einige wichtige geometrische Interpretationen der eingeführten Strukturen.

Satz 12.19: (Geometrische Interpretation der 2-Norm)

Die 2-Norm $\|\vec{x}\|_2$ eines Vektors $\vec{x} \in \mathbb{R}^2$ bzw. $\vec{x} \in \mathbb{R}^3$ ist seine geometrische Länge.

Beweis:



Pythagoras liefert als geometrische Länge $\sqrt{x_1^2 + x_2^2 + x_3^2} = \|\vec{x}\|_2$.

Q.E.D.

Satz 12.20: (Geometrische Interpretation des Skalarprodukts)

Für beliebige Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^2$ bzw. $\vec{x}, \vec{y} \in \mathbb{R}^3$ gilt

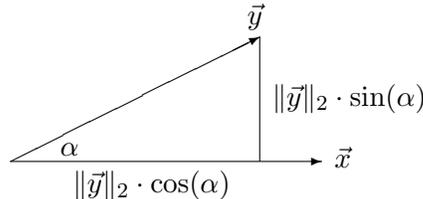
$$\langle \vec{x}, \vec{y} \rangle = \|\vec{x}\|_2 \cdot \|\vec{y}\|_2 \cdot \cos(\alpha),$$

wo α der Winkel zwischen den Vektoren \vec{x} und \vec{y} ist.

Beweis: Das Skalarprodukt ist invariant unter Drehungen, d.h., es gilt

$$\langle \vec{x}, \vec{y} \rangle = \langle \vec{x}', \vec{y}' \rangle,$$

wenn $\vec{x}' = O \vec{x}$ $\vec{y}' = O \vec{y}$ durch die selbe Drehung (beschrieben durch eine orthogonale Matrix O) aus \vec{x} und \vec{y} hervorgehen (wir erläutern dies, sobald wir orthogonale Matrizen eingeführt haben). Damit können wir o.B.d.A. die Vektoren so drehen, dass $\vec{x} = (\|\vec{x}\|_2, 0, 0)^T$ und $\vec{y} = (y_1, y_2, 0)^T = (\|\vec{y}\|_2 \cdot \cos(\alpha), \|\vec{y}\|_2 \cdot \sin(\alpha), 0)^T$ gilt:



Die Definition des Skalarprodukts liefert dann sofort

$$\langle \vec{x}, \vec{y} \rangle = \left\langle \begin{pmatrix} \|\vec{x}\|_2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \|\vec{y}\|_2 \cdot \cos(\alpha) \\ \|\vec{y}\|_2 \cdot \sin(\alpha) \\ 0 \end{pmatrix} \right\rangle = \|\vec{x}\|_2 \cdot \|\vec{y}\|_2 \cdot \cos(\alpha).$$

Q.E.D.

Speziell auf dem \mathbb{R}^3 ist ein interessantes Produkt zweier Vektoren definiert, das einen Vektor liefert, der auf den zwei gegebenen Vektoren senkrecht steht:

Definition 12.21: (Das Kreuzprodukt auf dem \mathbb{R}^3)

Das Kreuzprodukt $\vec{x} \times \vec{y}$ zweier Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^3$ ist der Vektor

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2 \cdot y_3 - x_3 \cdot y_2 \\ x_3 \cdot y_1 - x_1 \cdot y_3 \\ x_1 \cdot y_2 - x_2 \cdot y_1 \end{pmatrix}.$$

Bemerkung 12.22: a) Wie leicht nachzurechnen ist, gilt

$$\langle \vec{x}, \vec{x} \times \vec{y} \rangle = \langle \vec{y}, \vec{x} \times \vec{y} \rangle = 0,$$

d.h., das Kreuzprodukt zweier Vektoren steht senkrecht auf jedem der beiden Vektoren.

b) Das Kreuzprodukt verschwindet genau dann, wenn die beiden Vektoren parallel sind (oder einer der Vektoren verschwindet).

c) Das Kreuzprodukt ist weder kommutativ noch assoziativ. Es gilt

$$\vec{x} \times \vec{y} = -\vec{y} \times \vec{x}$$

und die „Jacobi-Identität“

$$\vec{x} \times (\vec{y} \times \vec{z}) + \vec{y} \times (\vec{z} \times \vec{x}) + \vec{z} \times (\vec{x} \times \vec{y}) = 0$$

für alle $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$.

Satz 12.23: (Geometrische Interpretation des Kreuzprodukts)

Für beliebige Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^3$ gilt

$$\|\vec{x} \times \vec{y}\|_2 = \|\vec{x}\|_2 \cdot \|\vec{y}\|_2 \cdot |\sin(\alpha)|,$$

wo α der Winkel zwischen den Vektoren \vec{x} und \vec{y} ist. Die Länge des Kreuzprodukts ist damit die Fläche des von \vec{x}, \vec{y} aufgespannten Parallelogramms.

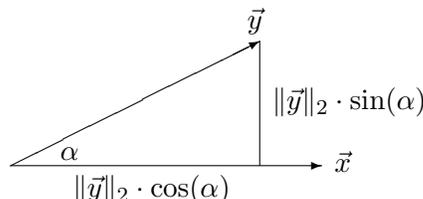
Beweis: Wir argumentieren ähnlich wie bei Satz 12.20. Das Kreuzprodukt ist invariant unter Drehungen: es gilt²

$$(O\vec{x}) \times (O\vec{y}) = \pm O(\vec{x} \times \vec{y}),$$

wobei O eine orthogonale Drehmatrix ist (das Vorzeichen hängt davon ab, ob O eine Spiegelung enthält). Da Drehungen die Längen der Vektoren und den Winkel zwischen den Vektoren nicht ändern, können wir o.B.d.A. die Vektoren so drehen, dass

$$\vec{x} = (\|\vec{x}\|_2, 0, 0)^T \quad \text{und} \quad \vec{y} = (y_1, y_2, 0)^T = (\|\vec{y}\| \cdot \cos(\alpha), \|\vec{y}\| \cdot \sin(\alpha), 0)^T$$

gilt:



Die Definition des Kreuzprodukts liefert dann sofort

$$\vec{x} \times \vec{y} = \begin{pmatrix} \|\vec{x}\|_2 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} \|\vec{y}\|_2 \cdot \cos(\alpha) \\ \|\vec{y}\|_2 \cdot \sin(\alpha) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \|\vec{x}\|_2 \cdot \|\vec{y}\|_2 \cdot \sin(\alpha) \end{pmatrix}.$$

Die Fläche des von \vec{x}, \vec{y} aufgespannten Parallelogramms ist die Seitenlänge $\|\vec{x}\|_2$ mal der Höhe $\|\vec{y}\|_2 \cdot \sin(\alpha)$.

Q.E.D.

Bemerkung 12.24: Die Beobachtung, dass $\vec{x} \times \vec{y}$ senkrecht auf \vec{x} und \vec{y} steht, legt die Richtung von $\vec{x} \times \vec{y}$ bis auf ein Vorzeichen geometrisch fest. Der letzte Satz legt die Länge von $\vec{x} \times \vec{y}$ geometrisch fest. Das Vorzeichen von $\vec{x} \times \vec{y}$ ist geometrisch durch die „**Rechte-Hand-Regel**“ bestimmt: das aus $\vec{x}, \vec{y}, \vec{x} \times \vec{y}$ bestehende Dreibein entspricht Daumen, Zeigefinger und (gekrümmtem) Mittelfinger der rechten Hand. Dies legt in der Tat die Richtung von $\vec{x} \times \vec{y}$ fest, wenn man sich nicht den Mittelfinger brechen will.

Und noch ein interessantes Produkt auf dem \mathbb{R}^3 :

Definition 12.25: (Das Spatprodukt auf dem \mathbb{R}^3)

Das Spatprodukt $|\langle \vec{x}, \vec{y}, \vec{z} \rangle|$ dreier Vektoren $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ ist der skalare Wert

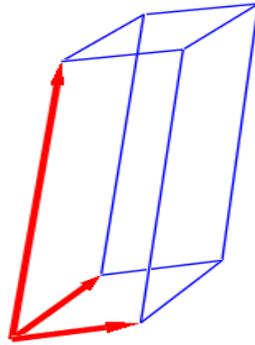
$$|\langle \vec{x}, \vec{y}, \vec{z} \rangle| = \langle \vec{x}, \vec{y} \times \vec{z} \rangle = \langle \vec{y}, \vec{z} \times \vec{x} \rangle = \langle \vec{z}, \vec{x} \times \vec{y} \rangle.$$

Ein „Dreibein“ $\vec{x}, \vec{y}, \vec{z}$ heißt „**rechtshändig**“, wenn $|\langle \vec{x}, \vec{y}, \vec{z} \rangle| > 0$ gilt. Für $|\langle \vec{x}, \vec{y}, \vec{z} \rangle| < 0$ heißt das Dreibein „**linkshändig**“.

²Dies ist etwas mühselig nachzurechnen.

Satz 12.26: (Geometrische Interpretation des Spatprodukts)

Für beliebige Vektoren $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ gilt $|\langle \vec{x}, \vec{y}, \vec{z} \rangle| = \pm V$, wo V das Volumen des von den Vektoren $\vec{x}, \vec{y}, \vec{z}$ aufgespannten Spats ist:



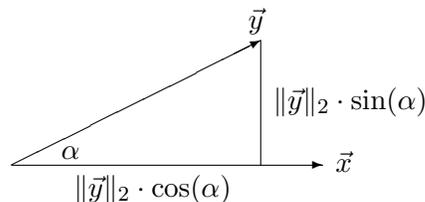
Beweis: Wir argumentieren ähnlich wie bei Satz 12.20 und Satz 12.23. Das Spatprodukt ist invariant unter Drehungen: es gilt

$$|\langle \vec{x}, \vec{y}, \vec{z} \rangle| = \pm |\langle O\vec{x}, O\vec{y}, O\vec{z} \rangle|,$$

wobei O eine orthogonale Drehmatrix ist (das Vorzeichen hängt davon ab, ob O eine Spiegelung enthält). O.B.d.A. können wir die Vektoren so drehen, dass

$$\vec{x} = (\|\vec{x}\|_2, 0, 0)^T \quad \text{und} \quad \vec{y} = (y_1, y_2, 0)^T = (\|\vec{y}\|_2 \cdot \cos(\alpha), \|\vec{y}\|_2 \cdot \sin(\alpha), 0)^T$$

gilt:



Die Definition des Kreuzprodukts liefert dann sofort

$$\begin{aligned} |\langle \vec{x}, \vec{y}, \vec{z} \rangle| &= \langle \vec{z}, \vec{x} \times \vec{y} \rangle = \left\langle \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}, \begin{pmatrix} \|x\|_2 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} \|y\|_2 \cdot \cos(\alpha) \\ \|y\|_2 \cdot \sin(\alpha) \\ 0 \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \|x\|_2 \cdot \|y\|_2 \cdot \sin(\alpha) \end{pmatrix} \right\rangle = z_3 \cdot \|x\|_2 \cdot \|y\|_2 \cdot \sin(\alpha). \end{aligned}$$

Hierbei ist $\|x\|_2 \cdot \|y\|_2 \cdot \sin(\alpha)$ die Fläche des von \vec{x}, \vec{y} aufgespannten Parallelogramms (die Grundfläche des Spats) und z_3 ist die Höhe des Spats (die

Grundfläche liegt hier in der x - y -Ebene). Das Spatvolumen ist Grundfläche mal Höhe.

Q.E.D.

12.2.2 Geraden und Ebenen

Definition 12.27: (Geraden)

Zu gegebenen $\vec{x}_0 \in \mathbb{R}^2$ bzw. \mathbb{R}^3 und $0 \neq \vec{d} \in \mathbb{R}^2$ bzw. \mathbb{R}^3 definieren wir die **Gerade durch \vec{x}_0 mit dem Richtungsvektor \vec{d}** als die Menge aller Punkte der Form

$$\left\{ \vec{x}_0 + \lambda \cdot \vec{d}; \lambda \in \mathbb{R} \right\}.$$

Definition 12.28: (Ebenen)

Zu gegebenem $\vec{x}_0 \in \mathbb{R}^3$ und linear unabhängigen Vektoren $\vec{d}_1, \vec{d}_2 \in \mathbb{R}^3$ definieren wir die **Ebene durch \vec{x}_0 mit den Tangentialvektoren \vec{d}_1, \vec{d}_2** als die Menge aller Punkte der Form

$$\left\{ \vec{x}_0 + \lambda_1 \cdot \vec{d}_1 + \lambda_2 \cdot \vec{d}_2; \lambda_1, \lambda_2 \in \mathbb{R} \right\}.$$

In dieser Definition werden 9 Parameter (jeweils 3 Komponenten der Vektoren $\vec{x}_0, \vec{d}_1, \vec{d}_2$) zur Definition der Ebene vorgegeben. Alternativ kann man die Ebene auch durch insgesamt nur 3 Parameter festlegen, denn eine Ebene ist als die Menge aller Punkte $(x_1, x_2, x_3)^T$ bestimmt, die eine lineare Gleichung

$$\left\langle \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right\rangle = n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 = d$$

bestimmen, wobei o.B.d.A. durch Multiplikation der Gleichung mit einer geeigneten Zahl der Wert von d auf 0 oder 1 festgelegt werden kann. Sei

$$\vec{n} = \vec{d}_1 \times \vec{d}_2$$

der „Normalenvektor“ der Ebene, der senkrecht auf den beiden Tangentenvektoren \vec{d}_1, \vec{d}_2 steht. In der Tat folgt aus der Parameterform

$$\vec{x} = \vec{x}_0 + \lambda_1 \cdot \vec{d}_1 + \lambda_2 \cdot \vec{d}_2$$

der Ebenenpunkte durch skalare Multiplikation mit \vec{n} :

$$\langle \vec{n}, \vec{x} \rangle = \langle \vec{n}, \vec{x}_0 \rangle + \lambda_1 \cdot \underbrace{\langle \vec{n}, \vec{d}_1 \rangle}_{=0} + \lambda_2 \cdot \underbrace{\langle \vec{n}, \vec{d}_2 \rangle}_{=0} = \langle \vec{n}, \vec{x}_0 \rangle,$$

d.h., $\langle \vec{n}, \vec{x} \rangle$ hat für alle Ebenenpunkte den selben Wert $\langle \vec{n}, \vec{x}_0 \rangle$ ($=: d$). Die Gleichung

$$\langle \vec{n}, \vec{x} \rangle = d$$

heißt „**Hessesche Normalform**“ der Ebene. Der Vektor \vec{n} ist dabei der **Normalenvektor** der Ebene. Der Parameter d hat folgende geometrische Interpretation: Wir starten am Nullpunkt und laufen orthogonal auf die Ebene zu, d.h., wir betrachten die Punkte auf der Geraden $\vec{x} = \lambda \cdot \vec{n}$. Für den Durchstoßpunkt durch die Ebene muß

$$\langle \vec{n}, \vec{x} \rangle = \langle \vec{n}, \lambda \cdot \vec{n} \rangle = \lambda \cdot \langle \vec{n}, \vec{n} \rangle = d \quad \Rightarrow \quad \lambda = \frac{d}{\langle \vec{n}, \vec{n} \rangle}$$

gelten, der Durchstoßpunkt ist also

$$\vec{x} = \frac{d}{\langle \vec{n}, \vec{n} \rangle} \cdot \vec{n}.$$

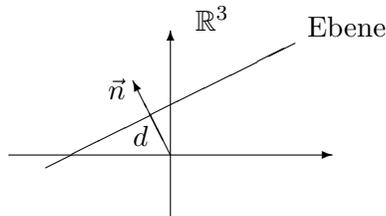
Der Betrag dieses Vektors

$$\|\vec{x}\|_2 = \frac{d}{\langle \vec{n}, \vec{n} \rangle} \cdot \|\vec{n}\|_2 = \frac{d}{\|\vec{n}\|_2}$$

ist der Abstand der Ebene zum Nullpunkt:

Für $\|\vec{n}\|_2 = 1$ ist der Parameter d der durch $\langle \vec{n}, \vec{x} \rangle = d$ gegebenen Ebene der Abstand der Ebene vom Nullpunkt.

Ist $d > 0$, zeigt der Normalenvektor vom Ursprung in Richtung der Ebene:



12.3 Matrizen

Betrachten wir den Fall eines Vektors $\vec{x} = (x_1, \dots, x_n)^T$, der mit der Standardbasis

$$\vec{e}_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

als $\vec{x} = \sum_j x_j \cdot \vec{e}_j$ gegeben ist. Angenommen, wir haben eine alternative Basis \vec{e}'_i gegeben, die durch die Linearkombinationen

$$\vec{e}_j = \sum_{i=1}^n \vec{e}'_i \cdot a_{ij}, \quad j = 1, \dots, n$$

mit der Standardbasis verknüpft ist. In der neuen Basis gilt $\vec{x} = \sum_i x'_i \cdot \vec{e}'_i$, wobei sich die neuen Koeffizienten x'_i durch

$$\vec{x} = \sum_{i=1}^n x'_i \cdot \vec{e}'_i = \sum_{j=1}^n x_j \cdot \vec{e}_j = \sum_{j=1}^n x_j \cdot \sum_{i=1}^n \vec{e}'_i \cdot a_{ij} = \sum_{i=1}^n \underbrace{\left(\sum_{j=1}^n a_{ij} \cdot x_j \right)}_{x'_i} \cdot \vec{e}'_i$$

ergeben, also

$$x'_i = \sum_{j=1}^n a_{ij} \cdot x_j, \quad i = 1, \dots, n.$$

Wir schreiben dies formal als

$$\begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

wobei wir die Koeffizienten (a_{ij}) in ein rechteckiges Schema („Matrix“) einordnen. Wir führen solche Rechteck-Schemata nun formal ein und vereinbaren, wie wir damit rechnen.

Definition 12.29: (Matrizen)

Sei K ein Skalarenkörper (also für uns $K = \mathbb{R}$ bzw. $K = \mathbb{C}$). Die Menge

$$K^{n \times m} = \left\{ \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}; a_{11}, \dots, a_{nm} \in K \right\}$$

wird als **Raum der $(n \times m)$ -Matrizen** bezeichnet. Wir benutzen die Notation $A = (a_{ij}) \in K^{n \times m}$. Der Raum $K^{n \times m}$ wird mit den Operationen

$$(A + B)_{ij} = a_{ij} + b_{ij}, \quad (\lambda \cdot A)_{ij} = \lambda \cdot a_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

für $A = (a_{ij}) \in K^{n \times m}$, $B = (b_{ij}) \in K^{n \times m}$, $\lambda \in K$ zu einem Vektorraum über K .

Neben der Addition von Matrizen und der Multiplikation mit Skalaren definieren wir eine Multiplikation von Matrizen:

$$(n \times p)\text{-Matrix} \times (p \times m)\text{-Matrix} = (n \times m)\text{-Matrix.}$$

Definition 12.30: (Matrizenmultiplikation)

Für $A = (a_{ij}) \in K^{n \times p}$, $B = (b_{ij}) \in K^{p \times m}$ definieren wir das Produkt $C = AB$ als die Matrix $C = (c_{ij}) \in K^{n \times m}$ mit

$$c_{ij} = \sum_{k=1}^p a_{ik} \cdot b_{kj}, \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

Spaltenvektoren entsprechen Matrizen mit nur einer Spalte

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^{n \times 1},$$

Zeilenvektoren entsprechen Matrizen mit nur einer Zeile:

$$(x_1, \dots, x_n) \in K^{1 \times n}.$$

Das Matrixprodukt einer Zeile mit einer Spalte entspricht dem Skalarprodukt:

$$(a_1, \dots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{k=1}^n a_k \cdot b_k.$$

Bemerkung 12.31: Das Matrixprodukt läßt sich damit folgendermassen verinnerlichen:

$$\begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{ip} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{p1} & \dots & b_{pj} & \dots & b_{pm} \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1m} \\ \vdots & c_{ij} & \vdots \\ c_{n1} & \dots & c_{nm} \end{pmatrix},$$

wobei

$$\begin{aligned} c_{ij} &= \text{Skalarprodukt}((i\text{-te Zeile von } A) \times (j\text{-te Spalte von } B)) \\ &= (a_{i1}, \dots, a_{ip}) \begin{pmatrix} b_{1j} \\ \vdots \\ b_{pj} \end{pmatrix} = \sum_{k=1}^p a_{ik} \cdot b_{kj}. \end{aligned}$$

Bemerkung 12.32: Das Matrixprodukt (mit Matrizen A, B, C zulässiger Dimension) ist assoziativ:

$$(AB)C = A(BC),$$

denn

$$\begin{aligned} ((AB)C)_{ij} &= \sum_{k_2} \left(\sum_{k_1} a_{ik_1} \cdot b_{k_1k_2} \right) \cdot c_{k_2j} = \sum_{k_1} \sum_{k_2} a_{ik_1} \cdot b_{k_1k_2} \cdot c_{k_2j} \\ &\quad \parallel \\ (A(BC))_{ij} &= \sum_{k_1} a_{ik_1} \left(\sum_{k_2} b_{k_1k_2} \cdot c_{k_2j} \right) = \sum_{k_1} \sum_{k_2} a_{ik_1} \cdot b_{k_1k_2} \cdot c_{k_2j}. \end{aligned}$$

Warnung: Das Matrixprodukt ist nicht kommutativ, i.A. gilt $AB \neq BA$. Beispiel:

$$\begin{aligned} AB &= \begin{matrix} & A & & B \\ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} & = & \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \\ BA &= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}. \end{matrix}$$

Wir führen weiterhin ein:

Definition 12.33: (Transposition)

- a) Die **Transponierte** $A^T = (b_{ij})$ einer $(n \times m)$ -Matrix $A = (a_{ij})$ ist die $(m \times n)$ -Matrix

$$b_{ij} = a_{ji}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

- b) Eine Matrix heißt **symmetrisch** (oder auch **selbstadjungiert**), wenn $A = A^T$ gilt, also $a_{ij} = a_{ji}$.

Sie heißt **antisymmetrisch** (oder auch **schief-symmetrisch**), wenn $A = -A^T$ gilt, also $a_{ij} = -a_{ji}$.

- c) Die **hermitesche Transponierte** $A^H = (c_{ij})$ einer (komplexen) $(n \times m)$ -Matrix $A = (a_{ij})$ ist die $(m \times n)$ -Matrix

$$c_{ij} = \overline{a_{ji}}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

(Transposition plus komplexe Konjugation).

- d) Eine Matrix heißt **hermitesch**, wenn $A = A^H$ gilt, also $a_{ij} = \overline{a_{ji}}$. Sie heißt **antihhermitesch** (oder auch **schiefhermitesch**), wenn $A = -A^H$ gilt, also $a_{ij} = -\overline{a_{ji}}$.

Bemerkung 12.34: Für das reelle Skalarprodukt $\langle \vec{x}, \vec{y} \rangle = \sum_k x_k \cdot y_k$ auf dem \mathbb{R}^n gilt offensichtlich

$$\langle \vec{x}, A \vec{y} \rangle = \langle A^T \vec{x}, \vec{y} \rangle = \sum_{ij} x_i \cdot a_{ij} \cdot y_j$$

für jedes $\vec{x}, \vec{y} \in \mathbb{R}^n$ und jedes $A \in \mathbb{R}^{n \times n}$.

Für das komplexe Skalarprodukt $\langle \vec{x}, \vec{y} \rangle = \sum_k \bar{x}_k \cdot y_k$ auf dem \mathbb{C}^n gilt offensichtlich

$$\langle \vec{x}, A \vec{y} \rangle = \langle A^H \vec{x}, \vec{y} \rangle = \sum_{ij} \bar{x}_i \cdot a_{ij} \cdot y_j$$

für jedes $\vec{x}, \vec{y} \in \mathbb{C}^n$ und jedes $A \in \mathbb{C}^{n \times n}$.

Durch (hermitesche) Transposition kann eine Matrix von einer Seite des Skalarprodukts auf die andere Seite bewegt werden.

Definition 12.35: (Einheitsmatrix)

Die Matrix

$$\mathbb{1} = (\delta_{ij}) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

auf dem $K^{n \times n}$ heißt $(n \times n)$ -**Einheitsmatrix**. Es gilt

$$\mathbb{1} \vec{x} = \vec{x}, \quad \mathbb{1} A = A, \quad B \mathbb{1} = B$$

für alle $\vec{x} \in K^n$, alle $A \in K^{n \times m}$ und alle $B \in K^{m \times n}$.

Definition 12.36: (Inverse einer Matrix)

Existiert zu einer quadratischen Matrix $A \in K^{n \times n}$ eine Matrix $A^{-1} \in K^{n \times n}$ mit

$$A^{-1} A = \mathbb{1},$$

so heißt A^{-1} **Inverse von A**. Es gilt dann automatisch auch $AA^{-1} = \mathbb{1}$.

Wegen der Nichtkommutativität des Matrixprodukts ist dabei kurz zu überlegen, dass die „Linksinverse“ mit der Eigenschaft $A^{-1}A = \mathbb{1}$ auch automatisch eine „Rechtsinverse“ mit $AA^{-1} = \mathbb{1}$ ist: Angenommen, eine Rechtsinverse B mit $AB = \mathbb{1}$ existiere. Multiplizieren wir von links mit A^{-1} , erhalten wir $A^{-1}AB = B = A^{-1} \mathbb{1} = A^{-1}$, also: $B = A^{-1}$.

Es ist eine fundamentale Aufgabe, Kriterien zu suchen, wann eine Matrix invertierbar ist. Wir werden sehen, dass z.B. eine Matrix genau dann invertierbar ist, wenn die Determinante von A nicht verschwindet. Weiterhin ist es wichtig, einen Algorithmus zur Berechnung der Inversen anzugeben. Dies wird in den Abschnitten 12.3 und 12.5 geschehen.

12.4 Determinanten

Definition 12.37: (Permutation)

Eine invertierbare Abbildung $p : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ heißt „**Permutation**“ der Zahlen $1, \dots, n$.

Die „**Inversionszahl**“ einer Permutation ist

$$I_p = \left| \left\{ (j, k); j, k \in \{1, \dots, n\}, j < k, p(j) > p(k) \right\} \right|$$

(also: I_p ist die Anzahl aller Paare (j, k) mit $j < k$, bei denen die Permutation die Anordnung $j < k$ zu $p(j) > p(k)$ vertauscht).

Das „**Signum**“ einer Permutation ist $\text{sign}(p) = (-1)^{I_p}$.

Satz 12.38:

Es gibt $n!$ verschiedene Permutationen der Zahlen $1, \dots, n$.

Beweis: Für $p(1)$ gibt es n verschiedene Möglichkeiten, für $p(2)$ gibt es $n - 1$ verschiedene Möglichkeiten (da $p(2) \neq p(1)$ gelten muss), für $p(3)$ gibt es $n - 2$ verschiedene Möglichkeiten (da $p(3) \neq p(1)$ und $p(3) \neq p(2)$ gelten muss), usw. Damit gibt es insgesamt $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 1 = n!$ Möglichkeiten, eine Permutation von $1, \dots, n$ zu konstruieren.

Q.E.D.

↓27.1.04

Konstruiert man alle Permutationen, indem man Schritt für Schritt einzelne Zahlenpaare in einer Permutation austauscht, kann man leicht das Signum nachhalten: mit jeder Vertauschung ändert das Signum sein Vorzeichen.

Satz 12.39:

Unterscheiden sich zwei Permutationen p und \tilde{p} nur durch den Austausch eines einzelnen Paares (j_0, k_0) , also

$$\tilde{p}(j_0) = p(k_0), \quad \tilde{p}(k_0) = p(j_0),$$

$$\tilde{p}(i) = p(i) \quad \forall i \in \{1, \dots, n\} \setminus \{j_0, k_0\},$$

so gilt

$$\text{sign}(p) = -\text{sign}(\tilde{p}).$$

Beweis: (für technisch Interessierte)

Es gelte o.B.d.A. $j_0 < k_0$ und $p(k_0) > p(j_0)$ (für $p(k_0) < p(j_0)$ sind in den folgenden Betrachtungen die Mengen A und \tilde{A} auszutauschen).

Die Paare (j, k) mit $j, k \in \{1, \dots, n\} \setminus \{j_0, k_0\}$ sind jeweils in beiden Mengen

$$A = \left\{ (j, k); j, k \in \{1, \dots, n\}, j < k, p(j) > p(k) \right\}$$

und

$$\tilde{A} = \left\{ (j, k); j, k \in \{1, \dots, n\}, j < k, \tilde{p}(j) > \tilde{p}(k) \right\}$$

gemeinsam vorhanden oder nicht vorhanden, da für diese Zahlen $\tilde{p}(j) = p(j)$ und $\tilde{p}(k) = p(k)$ gilt. Für $j \in \{j_0, k_0\}$ bzw. $k \in \{j_0, k_0\}$ enthält A die Teilmenge

$$\begin{aligned} & \left\{ (k_0, k); k \in \{k_0 + 1, \dots, n\}, p(k_0) > p(k) \right\} \\ \cup & \left\{ (j_0, k); k \in \{k_0 + 1, \dots, n\} \setminus \{k_0\}, p(j_0) > p(k) \right\} \\ \cup & \left\{ (j, k_0); j \in \{1, \dots, k_0 - 1\} \setminus \{j_0\}, p(j) > p(k_0) \right\} \\ \cup & \left\{ (j, j_0); j \in \{1, \dots, j_0 - 1\}, p(j) > p(j_0) \right\}, \end{aligned}$$

während \tilde{A} die Teilmenge

$$\begin{aligned} & \left\{ (j_0, k_0) \right\} \cup \left\{ (k_0, k); k \in \{k_0 + 1, \dots, n\}, \underbrace{\tilde{p}(k_0)}_{=p(j_0)} > \underbrace{\tilde{p}(k)}_{=p(k)} \right\} \\ & \cup \left\{ (j_0, k); k \in \{k_0 + 1, \dots, n\} \setminus \{k_0\}, \underbrace{\tilde{p}(j_0)}_{=p(k_0)} > \underbrace{\tilde{p}(k)}_{=p(k)} \right\} \\ & \cup \left\{ (j, k_0); j \in \{1, \dots, j_0 - 1\} \setminus \{j_0\}, \underbrace{\tilde{p}(j)}_{=p(j)} > \underbrace{\tilde{p}(k_0)}_{=p(j_0)} \right\} \\ & \cup \left\{ (j, k_0); j \in \{1, \dots, j_0 - 1\}, \underbrace{\tilde{p}(j)}_{=p(j)} > \underbrace{\tilde{p}(j_0)}_{=p(k_0)} \right\} \end{aligned}$$

enthält. Unter der Annahme $p(k_0) > p(j_0)$ folgt damit:

$$\begin{aligned} \tilde{A} \setminus A = & \left\{ (j_0, k_0) \right\} \cup \left\{ (j_0, k); k \in \{j_0 + 1, \dots, n\}, p(j_0) < p(k) < p(k_0) \right\} \\ & \cup \left\{ (j, k_0); j \in \{1, \dots, k_0 - 1\}, p(j_0) < p(j) < p(k_0) \right\} \end{aligned}$$

und

$$\begin{aligned} A \setminus \tilde{A} = & \left\{ (k_0, k); k \in \{k_0 + 1, \dots, n\}, p(j_0) < p(k) < p(k_0) \right\} \\ & \cup \left\{ (j, j_0); j \in \{1, \dots, j_0 - 1\}, p(j_0) < p(j) < p(k_0) \right\}. \end{aligned}$$

Mit $A = (A \cap \tilde{A}) \cup (A \setminus \tilde{A})$ und $\tilde{A} = (A \cap \tilde{A}) \cup (\tilde{A} \setminus A)$ folgt

$$|\tilde{A}| - |A| = 1 + 2 \cdot \left| \left\{ j; j \in \{j_0 + 1, \dots, k_0 - 1\}, p(j_0) < p(j) < p(k_0) \right\} \right|.$$

Damit unterscheiden sich $|\tilde{A}|$ und $|A|$ um eine ungerade ganze Zahl, womit $\text{sign}(\tilde{p}) = (-1)^{|\tilde{A}|} = -(-1)^{|A|} = -\text{sign}(p)$ folgt.

Q.E.D.

Beispiel 12.40: Wir schreiben eine Permutation in der Form

$$\frac{1 \quad 2 \quad \dots \quad n}{p(1) \quad p(2) \quad \dots \quad p(n)}.$$

Für die $3! = 6$ möglichen Permutationen der Zahlen 1, 2, 3 gilt:

$$\begin{aligned} p_1 &: \frac{1 \quad 2 \quad 3}{1 \quad 2 \quad 3}, & \text{sign}(p_1) &= 1, \\ p_2 &: \frac{1 \quad 2 \quad 3}{1 \quad 3 \quad 2}, & \text{sign}(p_2) &= -1, \\ p_3 &: \frac{1 \quad 2 \quad 3}{3 \quad 1 \quad 2}, & \text{sign}(p_3) &= 1, \\ p_4 &: \frac{1 \quad 2 \quad 3}{3 \quad 2 \quad 1}, & \text{sign}(p_4) &= -1, \\ p_5 &: \frac{1 \quad 2 \quad 3}{2 \quad 3 \quad 1}, & \text{sign}(p_5) &= 1, \\ p_6 &: \frac{1 \quad 2 \quad 3}{2 \quad 1 \quad 3}, & \text{sign}(p_6) &= -1. \end{aligned}$$

Definition 12.41: (Determinante einer Matrix)

Die Determinante einer quadratischen Matrix $A = (a_{i,j}) \in K^{n \times n}$ ist

$$\det(A) = \sum_p \text{sign}(p) \cdot a_{1p(1)} \cdot a_{2p(2)} \cdot \dots \cdot a_{np(n)},$$

wobei sich diese Summe über alle $n!$ Permutationen p der Zahlen $1, \dots, n$ erstreckt.

Beispiel 12.42: Für die Fälle $n = 1, 2, 3$ ergibt sich

$$\begin{aligned} \det((a_{11})) &= a_{11}, \\ \det\left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right) &= a_{11} \cdot a_{22} - a_{12} \cdot a_{21}, \end{aligned}$$

$$\det \left(\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \right) = a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\ - a_{11} \cdot a_{23} \cdot a_{32} - a_{12} \cdot a_{21} \cdot a_{33} - a_{13} \cdot a_{22} \cdot a_{31}.$$

Bezeichnung 12.43:

Seien

$$\vec{a}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \dots, \vec{a}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{pmatrix}$$

Vektoren in K^n . Die aus diesen Spalten aufgebaute Matrix wird mit $[\vec{a}_1, \dots, \vec{a}_n]$ bezeichnet:

$$[\vec{a}_1, \dots, \vec{a}_n] = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

Bemerkung 12.44: Die Definition der Matrixmultiplikation ist so, dass

$$A [\vec{b}_1, \dots, \vec{b}_n] = [A \vec{b}_1, \dots, A \vec{b}_n]$$

gilt, d.h., in einem Produkt AB ergeben sich die Spalten des Produkts als A wirkend auf die einzelnen Spalten von B .

Satz 12.45: (Linearität der Determinante bzgl. der Spalten)

Die Determinante ist linear in jeder Matrixspalte:

$$\det([\vec{a}_1, \dots, \lambda \cdot \vec{a}_i + \lambda' \cdot \vec{a}'_i, \dots, \vec{a}_n]) = \\ \lambda \cdot \det([\vec{a}_1, \dots, \vec{a}_i, \dots, \vec{a}_n]) + \lambda' \cdot \det([\vec{a}_1, \dots, \vec{a}'_i, \dots, \vec{a}_n]).$$

Beweis: Dies folgt unmittelbar aus der Definition

$$\det(A) = \sum_p \text{sign}(p) \cdot a_{1p(1)} \cdot a_{2p(2)} \cdot \dots \cdot a_{np(n)},$$

denn in jedem Summanden taucht genau eine Komponente jeder Spalte genau einmal auf.

Q.E.D.

Satz 12.46: (Antisymmetrie der Determinante bzgl. der Spalten)

Vertauscht man zwei Spalten, so wechselt die Determinante das Vorzeichen:

$$\det([\vec{a}_1, \dots, \vec{a}_i, \dots, \vec{a}_j, \dots, \vec{a}_n]) = -\det([\vec{a}_1, \dots, \vec{a}_j, \dots, \vec{a}_i, \dots, \vec{a}_n]).$$

Beweis: Dies folgt unmittelbar aus der Definition

$$\det(A) = \sum_p \text{sign}(p) \cdot a_{1p(1)} \cdot a_{2p(2)} \cdot \dots \cdot a_{np(n)}.$$

Sei A die Ausgangsmatrix, sei \tilde{A} die Matrix, die durch Austausch zweier Spalten entsteht. Sei τ die Permutation, die die beiden Spaltenindizes austauscht und alle anderen Indizes beibehält. Es gilt

$$\det(\tilde{A}) = \sum_p \text{sign}(p) \cdot \tilde{a}_{1p(1)} \cdot \tilde{a}_{2p(2)} \cdot \dots \cdot \tilde{a}_{np(n)},$$

wobei $\tilde{a}_{ij} = a_{i\tau(j)}$. Damit folgt

$$\det(\tilde{A}) = \sum_p \text{sign}(p) \cdot a_{1\tau(p(1))} \cdot a_{2\tau(p(2))} \cdot \dots \cdot a_{n\tau(p(n))}.$$

Läuft p über alle Permutationen, läuft auch $\tilde{p} = \tau \circ p$ über alle Permutationen, wobei nach Satz 12.39 $\text{sign}(\tilde{p}) = -\text{sign}(p)$ gilt:

$$\det(\tilde{A}) = \sum_{\tilde{p}} (-\text{sign}(\tilde{p})) \cdot a_{1\tilde{p}(1)} \cdot a_{2\tilde{p}(2)} \cdot \dots \cdot a_{n\tilde{p}(n)} = -\det(A).$$

Q.E.D.

Folgerung 12.47:

- a) Stimmen in einer Matrix zwei Spalten überein, so verschwindet ihre Determinante.
- b) Addiert man zu einer Spalte ein beliebiges Vielfaches einer anderen Spalte hinzu, so verändert sich der Wert der Determinante nicht.

Beweis: a) Vertauscht man die beiden übereinstimmenden Spalten, muss die Determinante das Vorzeichen wechseln, andererseits aber gleich bleiben. Das kann nur sein, wenn die Determinante 0 ist.

b) Mit der Linearität nach Satz 12.45 gilt

$$\begin{aligned} & \det([\vec{a}_1, \dots, \vec{a}_i, \dots, \vec{a}_j + \lambda \cdot \vec{a}_i, \dots, \vec{a}_n]) \\ &= \det([\vec{a}_1, \dots, \vec{a}_i, \dots, \vec{a}_j, \dots, \vec{a}_n]) + \lambda \cdot \det([\vec{a}_1, \dots, \vec{a}_i, \dots, \vec{a}_i, \dots, \vec{a}_n]), \end{aligned}$$

wobei die 2-te Determinante nach a) verschwindet.

Q.E.D.

Satz 12.48: (Invarianz unter Transposition)

Für jedes $A \in K^{n \times n}$ gilt $\det(A^T) = \det(A)$.

Beweis: Zu jeder Permutation p gibt es die inverse Permutation p^{-1} mit $p \circ p^{-1} = p^{-1} \circ p = \text{identische Abbildung} : i \rightarrow i$, wobei $\text{sign}(p^{-1}) = \text{sign}(p)$ gilt. Läuft p durch alle Permutationen, läuft auch p^{-1} durch alle Permutationen. Damit gilt

$$\begin{aligned} \det(A) &= \sum_p \text{sign}(p) \cdot a_{1p(1)} \cdot a_{2p(2)} \cdot \dots \cdot a_{np(n)}, \\ &= \sum_p \text{sign}(p^{-1}) \cdot a_{1p^{-1}(1)} \cdot a_{2p^{-1}(2)} \cdot \dots \cdot a_{np^{-1}(n)}. \end{aligned}$$

Für $j = p^{-1}(i)$ ergibt sich mit $a_{ip^{-1}(i)} = a_{p(j)j}$ nach geeigneter Umsortierung der Reihenfolge der Faktoren im Produkt:

$$\det(A) = \sum_p \underbrace{\text{sign}(p^{-1})}_p \cdot a_{p(1)1} \cdot a_{p(2)2} \cdot \dots \cdot a_{p(n)n}. \quad (\#)$$

Dies ist mit $a_{ji} = (A^T)_{ij}$ die Formel für $\det(A^T)$.

Q.E.D.

Folgerung 12.49:

Die Rolle von Zeilen und Spalten sind bei der Determinantenberechnung austauschbar.

Beispielsweise verschwindet analog zu Folgerung 12.47.a) die Determinante, wenn zwei Zeilen übereinstimmen. Man kann ein Vielfaches einer Zeile zu einer anderen addieren, ohne die Determinante zu verändern usw.

Definition 12.50: (Minoren)

Zu $A \in K^{n \times n}$ und $i, j \in \{1, \dots, n\}$ definiere den „Minor“

$$m_{ij} = \det \left(\left(\begin{array}{ccc|ccc} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{array} \right) \right),$$

d.h., die (Unter-)Determinante, die durch Streichen der Zeile i und der Spalte j entsteht.

Der folgende Satz liefert eine rekursive Berechnungsmöglichkeit für Determinanten. Durch „Entwicklung nach einer Spalte bzw. Zeile“ ergibt sich eine $(n \times n)$ -Determinante als Summe über Minoren (also $((n-1) \times (n-1))$ -Determinanten), so dass man rekursiv die Matrixdimension reduzieren kann, bis sich eine große Summe von (1×1) -Determinanten ergibt, die durch den entsprechenden Matrixeintrag gegeben sind.

Satz 12.51: (Laplacescher Entwicklungssatz)

Seien m_{ij} die Minoren der Matrix $A \in K^{n \times n}$. Für jede Spalte j gilt

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} \cdot a_{ij} \cdot m_{ij}.$$

Man nennt diese Formel „**Entwicklung der Determinante nach der j -ten Spalte**“.

Für jede Zeile i gilt

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} \cdot a_{ij} \cdot m_{ij}.$$

Man nennt diese Formel „**Entwicklung der Determinante nach der i -ten Zeile**“.

Beweisskizze: (Technisch) Wir entwickeln zunächst nach der letzten Spalte, wobei wir die Darstellung (#) aus dem Beweis von Satz 12.48 verwenden. In

$$\det(A) = \sum_p \text{sign}(p) \cdot a_{p(1)1} \cdot a_{p(2)2} \cdot \dots \cdot a_{p(n)n}$$

fixieren wir beim Durchlauf durch alle Permutationen zunächst den Bildwert für n , d.h., wir geben $p(n) = i$ mit $i \in \{1, \dots, n\}$ vor:

$$\det(A) = \sum_{i=1}^n a_{in} \cdot \sum_{p_i} \text{sign}(p_i) \cdot a_{p_i(1)1} \cdot \dots \cdot a_{p_i(n-1),n-1},$$

wobei im i -ten Summanden die (Teil-)Permutation p_i alle invertierbaren Abbildungen $\{1, 2, \dots, n-1\} \rightarrow \{1, \dots, n\} \setminus \{i\}$ zu durchlaufen hat. (In $\text{sign}(p_i)$ ist p_i als Permutation aller Zahlen $\{1, \dots, n\}$ mit vorgebenem $p_i(n) = i$ aufzufassen.) Offensichtlich enthält

$$\sum_{p_i} \text{sign}(p_i) \cdot a_{p_i(1)1} \cdot \dots \cdot a_{p_i(n-1),n-1}$$

alle Terme, die in der Definition des Minors m_{in} auftauchen (der Zeilenindex nimmt alle Werte in $\{1, \dots, n\} \setminus \{i\}$ an, der Spaltenindex n taucht im Produkt $a_{p_i(1)1} \cdot \dots \cdot a_{p_i(n-1),n-1}$ nicht auf). Es verbleibt, die Vorzeichen $\text{sign}(p_i)$ nachzuhalten. Sei

$$\tilde{p}_i(k) = \begin{cases} p_i(k) & \text{falls } p_i(k) \in \{1, \dots, i-1\}, \\ i & \text{falls } p_i(k) = i, \text{ also } k = n, \\ p_i(k) - 1 & \text{falls } p_i(k) \in \{i+1, \dots, n\}. \end{cases}$$

Diese Permutation unterscheidet sich von p_i durch die $n-i$ Vertauschungen $i \leftrightarrow i+1, i+1 \leftrightarrow i+2, \dots, n-1 \leftrightarrow n$, also gilt $\text{sign}(p_i) = (-1)^{n-i} \cdot \text{sign}(\tilde{p}_i)$:

$$\sum_{p_i} \text{sign}(p_i) \cdot a_{p_i(1)1} \cdot \dots \cdot a_{p_i(n-1),n-1} = (-1)^{n-i} \sum_{\tilde{p}_i} \text{sign}(\tilde{p}_i) \cdot \tilde{a}_{\tilde{p}_i(1)1} \cdot \dots \cdot \tilde{a}_{\tilde{p}_i(n-1),n-1},$$

wobei (\tilde{a}_{ij}) die Matrix ist, die durch Streichen der i -ten Zeile aus $A = (a_{ij})$ entsteht (die ursprüngliche i -te Zeile kann man sich als n -te Zeile von (\tilde{a}_{ij}) vorstellen).

Die Permutation $\tilde{p}_i : \{1, \dots, n-1\} \rightarrow \{1, \dots, n-1\}$ ist als Permutation der Zahlen $\{1, \dots, n-1\}$ interpretierbar. Damit ist die obige Summe über die \tilde{p}_i nichts anderes als die Determinante der Ausgangsmatrix A , aus der die i -te Zeile und die n -te Spalte gestrichen wurde (also der Minor m_{in} der Matrix (a_{ij})):

$$\sum_{p_i} \text{sign}(p_i) \cdot a_{p_i(1)1} \cdot \dots \cdot a_{p_i(n-1),n-1} = (-1)^{n-i} \cdot m_{in} = (-1)^{n+i} \cdot m_{in}.$$

Insgesamt erhält man also die „Entwicklung nach der n -ten Spalte“

$$\det(A) = \sum_{i=1}^n (-1)^{i+n} \cdot a_{in} \cdot m_{in}.$$

Will man nicht nach der n -ten Spalte sondern nach der j -ten Spalte entwickeln, lösche man in der Ausgangsmatrix die j -te Spalte, lasse die Spalten $j+1$ bis n nach links rücken und stecke die ursprünglich j -te Spalte in die n -te Spalte (dies entspricht den $n-j$ Spaltenvertauschungen $j \leftrightarrow j+1$, dann $j+1 \leftrightarrow j+2$ usw. bis $n-1 \leftrightarrow n$). Entwickle nach der neuen n -ten Spalte (dies ist die alte j -te Spalte). Nach den obigen Überlegungen gilt

$$\det(A') = \sum_{i=1}^n (-1)^{i+n} \cdot a'_{in} \cdot m'_{in},$$

wo $A' = (a'_{ij})$ durch die obigen Spaltenvertauschungen aus A entsteht. Speziell gilt also $a'_{in} = a_{ij}$. Für die entsprechenden Minoren gilt $m'_{in} = m_{ij}$. Nach $n-j$ Spaltenvertauschungen gilt

$$\det(A') = (-1)^{n-j} \cdot \det(A) = (-1)^{j-n} \cdot \det(A),$$

also

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} \cdot a_{ij} \cdot m_{ij}.$$

Damit haben wir die „Entwicklung nach der j -ten Spalte“.

Da nach Satz 12.48 die Rolle von Zeilen und Spalten gegeneinander ausgetauscht werden können, kann man statt nach Spalten auch nach Zeilen entwickeln.

Q.E.D.

Beispiel 12.52: Wir entwickeln eine (2×2) -Determinante nach der ersten Spalte:

$$\det \left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right) = a_{11} \cdot \det((a_{22})) - a_{21} \cdot \det((a_{12})) = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}.$$

Entwicklung nach der zweiten Zeile:

$$\det \left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right) = -a_{21} \cdot \det((a_{12})) + a_{22} \cdot \det((a_{11})) = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}.$$

Beispiel 12.53: Wir entwickeln eine (3×3) -Determinante nach der zweiten Spalte:

$$\begin{aligned} \det \left(\begin{pmatrix} 1 & 2 & 1 \\ 0 & 3 & 2 \\ 0 & 4 & 5 \end{pmatrix} \right) &= \\ -2 \cdot \det \left(\begin{pmatrix} \cancel{1} & \cancel{2} & \cancel{1} \\ 0 & \cancel{3} & 2 \\ 0 & \cancel{4} & 5 \end{pmatrix} \right) &+ 3 \cdot \det \left(\begin{pmatrix} 1 & \cancel{2} & 1 \\ \emptyset & \cancel{3} & \cancel{2} \\ 0 & \cancel{4} & 5 \end{pmatrix} \right) - 4 \cdot \det \left(\begin{pmatrix} 1 & \cancel{2} & 1 \\ 0 & \cancel{3} & 2 \\ \emptyset & \cancel{4} & \cancel{5} \end{pmatrix} \right) \\ &= -2 \cdot \det \left(\begin{pmatrix} 0 & 2 \\ 0 & 5 \end{pmatrix} \right) + 3 \cdot \det \left(\begin{pmatrix} 1 & 1 \\ 0 & 5 \end{pmatrix} \right) - 4 \cdot \det \left(\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \right) \\ &= -2 \cdot 0 + 3 \cdot 5 - 4 \cdot 2 = 7. \end{aligned}$$

Entwicklung nach der ersten Spalte:

$$\det \left(\begin{pmatrix} 1 & 2 & 1 \\ 0 & 3 & 2 \\ 0 & 4 & 5 \end{pmatrix} \right) = 1 \cdot \det \left(\begin{pmatrix} \cancel{1} & \cancel{2} & \cancel{1} \\ \emptyset & 3 & 2 \\ \emptyset & 4 & 5 \end{pmatrix} \right) - 0 \cdot (\dots) + 0 \cdot (\dots) = 1 \cdot (15 - 8) = 7.$$

Beobachtung: man spart sich Rechenarbeit, wenn man nach Zeilen oder Spalten mit möglichst vielen Nullen entwickelt, da die entsprechenden Minoren gar nicht ausgerechnet werden brauchen (sie werden sowieso mit 0 multipliziert und tragen zur Determinante nichts bei).

Es ist rechentechnisch günstig, nach Zeilen oder Spalten mit möglichst vielen Nullen zu entwickeln.

↓28.1.04

Mit Folgerung 12.47.b) hat man eine Möglichkeit, Nullen in der Matrix zu erzeugen, indem man geeignete Vielfache einer Zeile/Spalte von einer anderen Zeile/Spalte abzieht. Hier ein Beispiel. In

$$\det \left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \right)$$

subtrahieren wir zunächst die zweite Zeile von der dritten:

$$\det \left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \right) = \det \left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 3 & 4 \end{pmatrix} \right).$$

Wir subtrahieren die erste Zeile von der zweiten:

$$\det \left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 3 & 4 \end{pmatrix} \right) = \det \left(\begin{pmatrix} 1 & 2 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 4 \end{pmatrix} \right).$$

Wir subtrahieren die erste Spalte von der zweiten:

$$\det \left(\begin{pmatrix} 1 & 2 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 4 \end{pmatrix} \right) = \det \left(\begin{pmatrix} 1 & 1 & 3 \\ 3 & 0 & 3 \\ 3 & 0 & 4 \end{pmatrix} \right).$$

Entwicklung nach der zweiten Spalte liefert:

$$\det \left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \right) = \det \left(\begin{pmatrix} 1 & 1 & 3 \\ 3 & 0 & 3 \\ 3 & 0 & 4 \end{pmatrix} \right) = -1 \cdot \det \left(\begin{pmatrix} 3 & 3 \\ 3 & 4 \end{pmatrix} \right) = -3.$$

Definition 12.54: (Dreiecksmatrizen)

Eine Matrix der Form

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}$$

heißt „**obere Dreiecksmatrix**“. Eine Matrix der Form

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix}$$

heißt „**untere Dreiecksmatrix**“.

Satz 12.55:

Die Determinante einer Dreiecksmatrix ist das Produkt der Diagonalelemente.

Beweis: Entwickle bei einer oberen Dreiecksmatrix die Determinanten rekursiv jeweils nach der ersten Spalte:

$$\begin{aligned} & \det \left(\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \right) \\ &= a_{11} \cdot \det \left(\begin{pmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ 0 & a_{33} & \dots & a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \right) \\ &= a_{11} \cdot a_{22} \cdot \det \left(\begin{pmatrix} a_{33} & a_{34} & \dots & a_{3n} \\ 0 & a_{44} & \dots & a_{4n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \right) \end{aligned}$$

usw. Bei unteren Dreiecksmatrizen entwickelt man analog rekursiv nach der ersten Zeile.

Q.E.D.

Nun ein fundamentaler Satz, der das eindeutige Indiz dafür ist, dass der Determinantenbegriff vernünftig mit Matrizen und ihrer Multiplikation zusammenspielt:

Satz 12.56: (Determinante eines Matrixprodukts)

Für $A, B \in K^{n \times n}$ gilt

$$\det(AB) = \det(A) \cdot \det(B).$$

Beweis: Wir benutzen die Spaltenzerlegung $A = [\vec{a}_1, \dots, \vec{a}_n]$. Mit

$$(AB)_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}$$

folgt die Spaltenzerlegung

$$AB = \left[\sum_{k_1=1}^n \vec{a}_{k_1} \cdot b_{k_1 1}, \sum_{k_2=1}^n \vec{a}_{k_2} \cdot b_{k_2 2}, \dots, \sum_{k_n=1}^n \vec{a}_{k_n} \cdot b_{k_n n} \right].$$

Da die Determinante linear in jeder Spalte ist, folgt

$$\det(AB) = \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_n=1}^n \det([\vec{a}_{k_1}, \vec{a}_{k_2}, \dots, \vec{a}_{k_n}]) \cdot b_{k_1 1} \cdot b_{k_2 2} \cdots b_{k_n n}.$$

In dieser n -fachen Summe (mit n^n Termen) sind nur diejenigen Terme vorhanden, in denen die Summationsindizes k_1, \dots, k_n alle verschieden sind (bei den anderen Termen taucht in der Determinante der Spalten von A mindestens eine Spalte mehrfach auf und liefert 0). Die Mehrfachsumme reduziert sich damit auf die $n!$ Summanden, in denen k_1, \dots, k_n eine Permutation von $\{1, \dots, n\}$ darstellt:

$$\det(AB) = \sum_p \det([\vec{a}_{p(1)}, \vec{a}_{p(2)}, \dots, \vec{a}_{p(n)}]) \cdot b_{p(1)1} \cdot b_{p(2)2} \cdots b_{p(n)n}.$$

Durch Vertauschen der Spalten in der Determinante der Spalten von A erhält man

$$\det([\vec{a}_{p(1)}, \vec{a}_{p(2)}, \dots, \vec{a}_{p(n)}]) = \text{sign}(p) \cdot \det([\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n]) = \text{sign}(p) \cdot \det(A),$$

so dass sich ergibt:

$$\begin{aligned} \det(AB) &= \det(A) \cdot \sum_p \text{sign}(p) \cdot b_{p(1)1} \cdot b_{p(2)2} \cdots b_{p(n)n} \\ &= \det(A) \cdot \det(B^T) = \det(A) \cdot \det(B). \end{aligned}$$

Q.E.D.

Definition 12.57: (Adjunktenmatrix)

Einer Matrix $A \in K^{n \times n}$ mit den Minoren m_{ij} gemäß Definition 12.50 wird die “**Adjunktenmatrix**“ $A_{adj} \in K^{n \times n}$ zugeordnet:

$$(A_{adj})_{ij} = (-1)^{i+j} \cdot m_{ji}.$$

Satz 12.58:

Für jedes $A \in K^{n \times n}$ mit der Adjunktenmatrix A_{adj} gilt:

$$A_{adj}A = A A_{adj} = \det(A) \cdot \mathbb{I}.$$

Beweis: Sei $A = (a_{ij}) = [\vec{a}_1, \dots, \vec{a}_n]$, $A_{adj} = (b_{ij}) = ((-1)^{i+j} \cdot m_{ji})$. Es gilt

$$(A_{adj}A)_{ij} = \sum_{k=1}^n b_{ik} \cdot a_{kj} = \sum_{k=1}^n (-1)^{i+k} \cdot m_{ki} \cdot a_{kj}.$$

Diese Summe ist die Entwicklung der Determinante

$$\det([\vec{a}_1, \dots, \overset{i}{\downarrow} \vec{b}, \dots, \vec{a}_n])$$

mit $\vec{b} = \vec{a}_j$ nach der i -ten Zeile. Für $i \neq j$ taucht in dieser Determinante zweimal die Spalte \vec{a}_j auf und liefert 0. Für $i = j$ ergibt sich $\det([\vec{a}_1, \dots, \vec{a}_n]) = \det(A)$:

$$(A_{adj}A)_{ij} = \begin{cases} \det(A) & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Um zu zeigen, dass auch $AA_{adj} = \det(A) \cdot \mathbb{1}$ gilt, betrachte man $A_{adj}^T A^T$. Nach Zerlegung von A^T in Spalten ergibt sich analog zu oben $A_{adj}^T A^T = \det(A^T) \cdot \mathbb{1}$. Durch Transposition folgt dann

$$(A_{adj}^T A^T)^T = AA_{adj} = (\det(A^T) \cdot \mathbb{1})^T = \det(A) \cdot \mathbb{1}.$$

Q.E.D.

Folgerung 12.59:

Eine quadratische Matrix A ist genau dann invertierbar, wenn $\det(A) \neq 0$ gilt. Mit der Adjunktenmatrix A_{adj} gilt:

$$A^{-1} = \frac{1}{\det(A)} \cdot A_{adj}.$$

Satz 12.60: (Die Cramersche Regel)

Wir betrachten das System

$$\begin{array}{ccccccc} a_{11} \cdot x_1 & + & \cdots & + & a_{1n} \cdot x_n & = & b_1, \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} \cdot x_1 & + & \cdots & + & a_{nn} \cdot x_n & = & b_n \end{array}$$

von n linearen Gleichungen für die n Unbekannten x_1, \dots, x_n . Mit $A = (a_{ij}) \in K^{n \times n}$, $\vec{x} = (x_1, \dots, x_n)^T$, $\vec{b} = (b_1, \dots, b_n)^T$ ergibt sich die kompakte Schreibweise

$$A\vec{x} = \vec{b}.$$

Es gibt genau dann eine eindeutige Lösung $\vec{x} = A^{-1}\vec{b}$, wenn $\det(A) \neq 0$ gilt. Die Komponenten der Lösung sind gegeben durch

$$x_i = \frac{\det \left(\begin{array}{cccc} & & \begin{array}{c} i \\ \downarrow \\ b_1 \end{array} & \\ \left(\begin{array}{cccc} a_{11} & \cdots & & a_{1n} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & b_n & \cdots & a_{nn} \end{array} \right) & & & \end{array} \right)}{\det \left(\begin{array}{ccc} \left(\begin{array}{ccc} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{array} \right) & & \end{array} \right)}$$

mit $i = 1, \dots, n$.

Beweis: Wir benutzen $(A^{-1})_{ij} = \frac{1}{\det(A)} \cdot (-1)^{i+j} \cdot m_{ji}$ mit den Minoren m_{ij} von A . Es folgt

$$x_i = \left(A^{-1}\vec{b} \right)_i = \sum_{j=1}^n (A^{-1})_{ij} \cdot b_j = \frac{1}{\det(A)} \cdot \sum_{j=1}^n (-1)^{i+j} \cdot m_{ji} \cdot b_j.$$

Hierbei ist $\sum_{j=1}^n (-1)^{i+j} \cdot m_{ji} \cdot b_j$ die Entwicklung von

$$\det([\vec{a}_1, \dots, \vec{b}, \dots, \vec{a}_n])$$

nach der i -ten Spalte, wo $\vec{a}_1, \dots, \vec{a}_n$ die Spalten von A sind und die Spalte \vec{a}_i durch die Spalte \vec{b} ersetzt wurde.

Q.E.D.

12.5 Lineare Gleichungssysteme

Im letzten Abschnitt haben wir ein System $A\vec{x} = \vec{b}$ von n linearen Gleichungen für n Unbekannte $\vec{x} = (x_1, \dots, x_n)^T$ gelöst, vorausgesetzt dass die Koeffizientenmatrix A invertierbar ist:

$$\vec{x} = A^{-1}\vec{b}.$$

Hierbei wurden explizite Lösungsformeln gefunden: die Inverse A^{-1} läßt sich über die Adjunktenmatrix von A mittels Determinanten bestimmen, die Lösung \vec{x} ist direkt über die Cramersche Regel 12.60 mittels Determinanten auszudrücken. Dies ist jedoch noch nicht die endgültige Antwort auf die Aufgabenstellung „Lösung linearer Gleichungssysteme“. Es verbleiben die Probleme:

- Was passiert bei nicht-invertierbaren Koeffizientenmatrizen? Was ist, wenn n Gleichungen für m Unbekannte (mit $n \neq m$) zu lösen sind?

- Die Determinantenformeln sind ineffektiv: die bisher vorgestellte Berechnung von Determinanten (z.B. rekursiv über Laplace-Entwicklung nach Zeilen oder Spalten) braucht viel zu viele Rechenschritte. (Eine $(n \times n)$ -Determinante besteht aus einer Summe von $n!$ Termen. Die Stirling-Näherung 9.58 sagt z.B. für $n = 100$ etwa $n! \approx 9.3 \times 10^{157}$ Summanden für die Determinante voraus).

Das Determinantenkalkül des letzten Abschnitts ist in der Tat mehr als allgemeine Strukturtheorie anzusehen, deren wesentliches Ergebnis die Aussage

$A\vec{x} = \vec{b}$ mit invertierbarem A hat genau dann eine eindeutige Lösung $\vec{x} = A^{-1}\vec{b}$, wenn $\det(A) \neq 0$ gilt.

In der Praxis geht man anders vor. Gleichungssysteme mit z.B. $n = 100$ Unbekannten lassen sich in der Tat in Sekundenbruchteilen lösen, wobei Determinanten keinerlei Rolle spielen.

12.5.1 Quadratische Dreieckssysteme

Ist die Koeffizientenmatrix A in einem linearen Gleichungssystem $A\vec{x} = \vec{b}$ von Dreiecksform, so ist die Berechnung der Lösung sehr einfach (und mit wenig Rechenaufwand verbunden):

Satz 12.61: (Rücksubstitution)

Ein lineares System $A\vec{x} = \vec{b}$

$$\begin{array}{rcccc} a_{11} \cdot x_1 & & & & = & b_1, \\ a_{21} \cdot x_1 & + & a_{22} \cdot x_2 & & = & b_2, \\ \vdots & & \vdots & & \ddots & \vdots \\ a_{n1} \cdot x_1 & + & a_{n2} \cdot x_2 & + & \cdots & + & a_{nn} \cdot x_n & = & b_n \end{array}$$

mit einer invertierbaren unteren $(n \times n)$ -Dreiecksmatrix $A = (a_{ij})$ wird durch „Rücksubstitution“

$$\begin{array}{l} x_1 := b_1 / a_{11}; \\ \text{for } i := 2 \text{ to } n \text{ do } x_i := \left(b_i - \sum_{j=1}^{i-1} a_{ij} \cdot x_j \right) / a_{ii}; \end{array}$$

mit

$$1 + \sum_{i=2}^n i = \frac{n(n+1)}{2} = \frac{n^2}{2} \left(1 + O\left(\frac{1}{n}\right) \right)$$

Multiplikationen/Divisionen (und ähnlich vielen Additionen) gelöst.

Bei einem System

$$\begin{array}{ccccccc} a_{11} \cdot x_1 + \cdots + & a_{1,n-1} \cdot x_{n-1} + & a_{1n} \cdot x_n & = & b_1 \\ & \vdots & \vdots & & \vdots \\ & a_{n-1,n-1} \cdot x_{n-1} + & a_{n-1,n} \cdot x_n & = & b_{n-1} \\ & & a_{nn} \cdot x_n & = & b_n \end{array}$$

mit invertierbarer oberer ($n \times n$)-Dreiecksmatrix $A = (a_{ij})$ lautet die Rücksubstitution:

$$\begin{array}{l} x_n := b_n / a_{nn} ; \\ \text{for } i := n - 1 \text{ downto } 1 \text{ do } x_i := \left(b_i - \sum_{j=i+1}^n a_{ij} \cdot x_j \right) / a_{ii} ; \end{array}$$

Beweis: Im Fall eines unteren Dreieckssystems liefert die erste Gleichung sofort $x_1 = b_1/a_{11}$. Eingesetzt in die zweite Gleichung erhält man $a_{22} \cdot x_2 = b_2 - a_{21} \cdot x_1$, wodurch sich sofort x_2 ergibt. Allgemein: hat man x_1, \dots, x_{i-1} bereits bestimmt, liefert die i -te Gleichung

$$a_{ii} \cdot x_i = b_i - \sum_{j=1}^{i-1} a_{ij} \cdot x_j$$

den Wert für x_i . Beachte: die Koeffizientenmatrix A ist genau dann invertierbar, wenn keines der Diagonalelemente verschwindet, man kann also durch die a_{ii} teilen!

Bei einem oberen Dreieckssystemen startet man mit der letzten Gleichung, um x_n zu bestimmen, setzt diesen Wert dann in die vorletzte Gleichung ein, aus der dann x_{n-1} bestimmbar ist usw.

Q.E.D.

Bemerkung 12.62: Der Rechenaufwand von etwa $n^2/2$ Multiplikationen entspricht exakt dem Aufwand, eine Multiplikation einer $(n \times n)$ -Dreiecksmatrix mit einem Vektor mit n Komponenten durchzuführen. Die Lösung von $A\vec{x} = \vec{b}$ bekommt man bei Dreiecksmatrizen also genauso schnell wie die Multiplikation $A\vec{x}$.

29.1.04↓

12.5.2 Der Gauß-Algorithmus

Nun wird der Standardalgorithmus zur Lösung linearer Gleichungssysteme vorgestellt, der

- die allgemeine Lösung von n Gleichungen mit m Unbekannten liefert (auch für $n \neq m$),
- mit einer akzeptablen Anzahl von Rechenoperationen auskommt. Faustregel: für n Gleichungen mit n Unbekannten braucht man $O(n^3)$ Rechenoperationen.

Es handelt sich dabei um den „**Gauss-Algorithmus**“, der je nach Variante als „**Gauss-**“, „**Gauss-Jordan-**“, „**Entzerrungsalgorithmus**“ oder auch als „**LR-Zerlegung**“ bezeichnet wird.

Gegeben seien n lineare Gleichungen für m Unbekannte

$$\begin{array}{ccccccc} a_{11} \cdot x_1 & + & \cdots & + & a_{1m} \cdot x_m & = & b_1, \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} \cdot x_1 & + & \cdots & + & a_{nm} \cdot x_m & = & b_n, \end{array}$$

die wir bequemer als $(n \times (m + 1))$ -Matrix

$$[A \mid \vec{b}] = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1m} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} & b_n \end{array} \right)$$

kodieren (die „**erweiterte Koeffizientenmatrix**“). Jede Zeile

$$(a_{i1} \quad \cdots \quad a_{im} \mid b_i)$$

ist dabei jeweils als Gleichung zu lesen:

$$a_{i1} \cdot x_1 + \cdots + a_{im} \cdot x_m = b_i.$$

Elementaroperationen 12.63:

Wir wollen das Gleichungssystem so umschreiben (vereinfachen), dass die Lösung möglichst direkt abgelesen werden kann. Dabei sollen lediglich folgende **Elementaroperationen** eingesetzt werden:

- i) Vertauschen von Gleichungen (= Zeilen von $[A \mid \vec{b}]$)
- ii) Multiplikation einer Gleichung (= Zeile von $[A \mid \vec{b}]$) mit einer Zahl $\neq 0$,

Mit der Forderung $\lambda \neq 0$ in ii) haben alle Elementarmatrizen nicht-verschwindende Determinanten, sind also invertierbar. Die Anwendung dieser Matrizen entspricht also einer Äquivalenztransformation, die die Lösung des Gleichungssystems nicht verändert.

Mit diesen Operationen soll das Ausgangsschema

$$[A \mid \vec{b}] = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1m} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} & b_n \end{array} \right)$$

in eine neue (einfachere) Form $[A' \mid \vec{b}']$ gebracht werden. Welche Struktur sollte die vereinfachte Koeffizientenmatrix haben? Wenn es uns gelingt, das System $A\vec{x} = \vec{b}$ in ein äquivalentes System $A'\vec{x} = \vec{b}'$ mit $A' = \mathbb{I}$ umzuschreiben, dann sind wir fertig. Nach Abschnitt 12.5.1 reicht es jedoch, als Endform $A' =$ Dreiecksmatrix anzusteuern, da das verbleibende Dreieckssystem $A'\vec{x} = \vec{b}'$ schnell per Rücksubstitution 12.61 gelöst werden kann.

Hier nun der **Gauß-Algorithmus** zur Lösung von n Gleichungen

$$\begin{array}{cccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & \cdots & + & a_{1m} \cdot x_m & = & b_1 \\ a_{21} \cdot x_1 & + & a_{22} \cdot x_2 & + & \cdots & + & a_{2m} \cdot x_m & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{n1} \cdot x_1 & + & a_{n2} \cdot x_2 & + & \cdots & + & a_{nm} \cdot x_m & = & b_n \end{array}$$

für m Unbekannte:

1.ter Eliminationsschritt: ziehe geeignete Vielfache der ersten Zeile von den anderen ab, um Nullen unter a_{11} zu erzeugen:

$$\begin{array}{cccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & \cdots & + & a_{1m} \cdot x_m & = & b_1 \\ \hline a_{21}/a_{11} & & a'_{22} \cdot x_2 & + & \cdots & + & a'_{2m} \cdot x_m & = & b'_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{n1}/a_{11} & & a'_{n2} \cdot x_2 & + & \cdots & + & a'_{nm} \cdot x_m & = & b'_n \end{array}$$

mit

$$a'_{ik} = a_{ik} - \frac{a_{i1}}{a_{11}} \cdot a_{1k}, \quad b'_i = b_i - \frac{a_{i1}}{a_{11}} \cdot b_1, \quad i = 2, \dots, n, k = 2, \dots, m.$$

Bezeichnung: $a_{i1}/a_{11} =$ “**Eliminationsfaktoren**”. Im Prinzip stehen unter a_{11} Nullen, d.h., diese Speicherplätze sind frei geworden. Diese werden mit den Eliminationsfaktoren belegt, was nichts kostet, sich aber im Abschnitt 12.6.2

als äußerst wertvoll herausstellen wird.

2.ter Eliminationsschritt: erzeuge Nullen unter a'_{22} :

$$\begin{array}{rcccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & a_{13} \cdot x_3 & + \cdots & + & a_{1n} \cdot x_m & = & b_1 \\ \frac{a_{21}}{a_{11}} & & a'_{22} \cdot x_2 & + & a'_{23} \cdot x_3 & + \cdots & + & a'_{2n} \cdot x_m & = & b'_2 \\ \vdots & & a'_{32}/a'_{22} & & a''_{33} \cdot x_3 & + \cdots & + & a''_{3m} \cdot x_m & = & b''_3 \\ \vdots & & \vdots & & \vdots & \ddots & & \vdots & & \vdots \\ a_{n1}/a_{11} & & a'_{n2}/a'_{22} & & a''_{n3} \cdot x_3 & + \cdots & + & a''_{nm} \cdot x_m & = & b''_n \end{array}$$

mit

$$a''_{ik} = a'_{ik} - \frac{a'_{i2}}{a'_{22}} \cdot a'_{2k}, \quad b''_i = b'_i - \frac{a'_{i2}}{a'_{22}} \cdot b'_2, \quad i = 3, \dots, n, \quad k = 3, \dots, m.$$

j .ter Eliminationsschritt: ziehe Vielfaches der j .ten Zeile von den restlichen Zeilen ab, um unter dem j .ten Diagonalelement Nullen zu erzeugen.

Einzig mögliches Problem: das j .te Diagonalelement verschwindet.

1. Fall: Darunter stehen nur Nullen. Es ist nichts mehr zu tun, gehe zum $(j+1)$.ten Schritt. Dieser Fall tritt genau dann irgendwann ein, wenn A nicht invertierbar ist.

2. Fall: Es gibt mindestens einen nichtverschwindenden Eintrag unter dem j .ten Diagonalelement. Wähle irgendeinen dieser Einträge (das **“Pivotelement”**), bringe es in die j .te Zeile, indem die gesamte Pivotzeile (inklusive der abzuspeichernden Eliminationsfaktoren!) mit der j .ten Zeile ausgetauscht wird. Dann erzeuge die gewünschten Nullen.

Algorithmus 12.64: (Gauß-Elimination = LR -Zerlegung, siehe Satz 12.75)

Erweitere $A = (a_{ij}) \in K^{n \times m}$ zum $n \times (m+1)$ -Schema:

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1m} & a_{1,m+1} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} & a_{n,m+1} \end{array} \right) \quad \text{mit } a_{i,n+1} = b_i, \quad i = 1, \dots, n.$$

```

for  $j := 1$  to  $\min(m, n)$  do begin
  falls  $a_{jj} = a_{j+1,j} = \dots = a_{nj} = 0$ , so gehe zum  $(j + 1)$ .ten Schritt ;
   $p := \mathbf{Pivotindex}(j) \in \{j, j + 1, \dots, n\}$  ;
  Vertauschung der Zeilen  $j$  und  $p$  im  $n \times (m + 1)$ -Schema ;
  for  $i := j + 1$  to  $n$  do begin
     $a_{ij} := a_{ij}/a_{jj}$  ;    (* Abspeichern der Eliminationsfaktoren *)
    for  $k := j + 1$  to  $m + 1$  do  $a_{ik} := a_{ik} - a_{ij} a_{jk}$  ;
  end ;
end ;

```

Hierbei ist $\mathbf{Pivotindex}(j)$ der Aufruf einer Prozedur, welche unter den Werten $\{a_{jj}, a_{j+1,j}, \dots, a_{nj}\}$ ein Element sucht, das nicht verschwindet, und den entsprechenden Zeilenindex p zurückgibt. Im Prinzip kann irgendein nicht-verschwindendes Pivotelement gewählt werden. Aus Gründen der numerischen Stabilität (minimiere den Einfluss von Rundungsfehlern bei Gleitpunktarithmetik) wählt man meist $p = \mathbf{Pivotindex}(j)$ so, dass

$$|a_{pj}| = \max(|a_{jj}|, |a_{j+1,j}|, \dots, |a_{nj}|)$$

gilt (d.h., das Pivotelement a_{pj} ist das betragsgrößte Element im unteren Teil der j -ten Spalte).

Resultat: zuletzt ist das zu $A\vec{x} = \vec{b}$ äquivalente obere Dreieckssystem $R\vec{x} = \vec{c}$

$$\left(\begin{array}{ccc|c} r_{11} & \cdots & r_{1n} & c_1 \\ & \ddots & \vdots & \vdots \\ * & & r_{nn} & c_n \end{array} \right) \quad (\text{für } n = m)$$

bzw.

$$\left(\begin{array}{ccccc|c} r_{11} & \cdots & r_{1n} & \cdots & r_{1m} & c_1 \\ & \ddots & \vdots & \ddots & \vdots & \vdots \\ * & & r_{nn} & \cdots & r_{nm} & c_n \end{array} \right) \quad (\text{für } n < m)$$

bzw.

$$\left(\begin{array}{ccc|c} r_{11} & \cdots & r_{1m} & c_1 \\ * & \ddots & \vdots & \vdots \\ \vdots & \ddots & r_{mm} & c_m \\ \vdots & \ddots & * & c_{m+1} \\ \vdots & \ddots & \vdots & \vdots \\ * & \ddots & * & c_n \end{array} \right) \quad (\text{für } n > m)$$

im $(n \times (m + 1))$ -Schema gespeichert. Mit * sind dabei die abgespeicherten Eliminationsfaktoren angedeutet, die erst im Abschnitt 12.6.2 von Bedeutung sein werden.

Rechenaufwand zur Erzeugung des oberen Dreieckssystems:

$$\sum_{j=1}^{\min(m,n)} \sum_{i=j+1}^n \left(1 + \sum_{k=j+1}^{m+1} 1\right) = \begin{cases} \frac{n \cdot (n-1) \cdot (3 \cdot m - n + 5)}{6} & \text{für } n \leq m \\ \frac{m \cdot (3 \cdot m \cdot n - m \cdot n - 6 \cdot m + 9 \cdot n - 5)}{6} & \text{für } n \geq m \end{cases}$$

Multiplikationen (und ähnlich viele Additionen). Für den in der Praxis wichtigsten Fall $m = n$ sind dies

$$\frac{n(n-1)(2 \cdot n + 5)}{6} = \frac{n^3}{3} \cdot \left(1 + O\left(\frac{1}{n}\right)\right)$$

Multiplikationen.

Bemerkung 12.65: Die Umformung $A\vec{x} = \vec{b} \iff R\vec{x} = \vec{c}$ funktioniert immer, selbst wenn A nicht invertierbar ist (wobei dann auch R nicht invertierbar ist).

Bemerkung 12.66: Wir diskutieren den Fall, dass die Anzahl n der Gleichungen nicht mit der Anzahl m der Unbekannten übereinstimmt:

Fall $n < m$ (weniger Gleichungen als Unbekannte):

Das vom Gauß-Algorithmus gelieferte obere Dreieckssystem hat die Form

$$\left(\begin{array}{cccccc|c} r_{11} & \cdots & r_{1n} & \cdots & \cdots & r_{1m} & c_1 \\ & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots \\ & & r_{nn} & \cdots & \cdots & r_{nm} & c_n \end{array} \right).$$

Durch Hinzufügen von $m - n$ triviale Gleichungen $0 = 0$

$$\left(\begin{array}{cccccc|c} r_{11} & \cdots & r_{1n} & \cdots & \cdots & r_{1m} & c_1 \\ & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots \\ & & r_{nn} & \cdots & \cdots & r_{nm} & c_n \\ & & & 0 & \cdots & 0 & 0 \\ & & & & \ddots & \vdots & \vdots \\ & & & & & 0 & 0 \end{array} \right)$$

ergibt sich ein quadratisches oberes $(m \times m)$ -System, dessen allgemeine Lösung im nächsten Abschnitt 12.5.3 beschrieben wird.

Fall $n > m$ (mehr Gleichungen als Unbekannte):

Das vom Gauß-Algorithmus gelieferte obere Dreieckssystem hat die Form

$$\left(\begin{array}{ccc|c} r_{11} & \cdots & r_{1m} & c_1 \\ & \ddots & \vdots & \vdots \\ & & r_{mm} & c_m \\ & & & c_{m+1} \\ & & & \vdots \\ & & & c_n \end{array} \right).$$

Ist irgendeiner der Werte c_{m+1}, \dots, c_n ungleich 0, ist das Gleichungssystem nicht lösbar (dort stehen die Gleichungen $0 = c_p$ mit $p = m + 1, \dots, n$). Sei nun $c_{m+1} = \dots = c_n = 0$. Man wirft diese trivialen Gleichungen fort und erhält ein $(m \times m)$ -Gleichungssystem

$$\left(\begin{array}{ccc|c} r_{11} & \cdots & r_{1m} & c_1 \\ & \ddots & \vdots & \vdots \\ & & r_{mm} & c_m \end{array} \right),$$

dessen allgemeine Lösung im nächsten Abschnitt 12.5.3 beschrieben wird. Es sollte im Fall $n > m$ erwähnt werden, dass man in der Praxis nicht wirklich den Gauß-Algorithmus verwendet, da man wegen der Rundungsfehler bei Gleitpunktarithmetik die Entscheidung, ob $c_{m+1} = \dots = c_n = 0$ gilt, nicht fundiert treffen kann (es stehen dort kleine Zahlenwerte, die eventuell reine Rundungsartefakte sein können). Stattdessen benutzt man sogenannte „Singularwertzerlegungen“ zum Lösen überbestimmter Gleichungssysteme.

12.5.3 Die allgemeine Lösung des Dreieckssystems

↓3.2.04

Der Gauß-Algorithmus hat aus den ursprünglichen n Gleichungen $A\vec{x} = \vec{b}$ für m Unbekannte ein äquivalentes oberes Dreieckssystem $R\vec{x} = \vec{c}$ gemacht. Hierbei haben wir gemäß Bemerkung 12.66 für $n < m$ triviale Gleichungen hinzugenommen bzw. für $n > m$ triviale Gleichungen entfernt, um zu einem quadratischen System zu kommen (genausoviele Gleichungen wie Unbekannte). Wir haben nun also noch die Situation eines quadratischen Systems ($m = n$) in oberer Dreiecksform

$$\begin{array}{ccccccc} r_{11} \cdot x_1 & + & \dots & + & r_{1n} \cdot x_n & = & c_1 \\ & & & & \vdots & & \vdots \\ & & & & r_{nn} \cdot x_n & = & c_n \end{array}$$

Diagonalelement des Schemas $\neq 0$, so ziehe ein geeignetes Vielfaches der j -ten Zeile von der i -ten Zeile ab, damit das j -te Element der i -ten Zeile verschwindet.

- Wiederhole die obigen Schritte für die i -te Zeile, bis diese eine Nullzeile ist.

Das Gleichungssystem hat nun „**obere Treppenform**“:

$$\left(\begin{array}{cccccccc|c} * & & & & & & & & * \\ & \ddots & & & & & & & \vdots \\ & & * & * & & & & & * \\ & & & \ddots & \ddots & & & & \vdots \\ & & & & \ddots & \ddots & & & * \\ & & & & & \ddots & \ddots & & \vdots \\ & & & & & & \ddots & \ddots & * \\ & & & 0 & 0 & 0 & \dots & 0 & 0 & * \\ & & & & * & * & \dots & \dots & * & * \\ & & & & & 0 & 0 & \dots & 0 & * \\ & & & & & & * & & & * \\ & & & & & & & \ddots & & \vdots \\ & & & & & & & & \ddots & \vdots \end{array} \right) \quad (\#)$$

Rechts neben den Nullen auf der Diagonale können jetzt nur noch Nullen stehen. Steht in der letzten Spalte des Schemas (eine der Komponenten d_i der rechten Seite \vec{d} von $R\vec{x} = \vec{d}$) rechts neben einer Nullzeile ein nicht-verschwindender Eintrag, so ist das Gleichungssystem insgesamt nicht lösbar: diese Gleichung besagt $0 = d_i \neq 0$.

Elimination „nach oben“: Laufe von rechts unten nach links oben durch die Diagonale:

- Ist das Diagonalelement ungleich Null, mache es zu 1, indem die gesamte Zeile durch das Diagonalelement geteilt wird. Dann erzeuge über dem Diagonalelement Nullen, indem geeignete Vielfache dieser Zeile von den oberen Zeilen abgezogen werden.

Es verbleiben längs der Diagonalen nur noch Nullen und Einsen. Über den Einsen stehen nur noch Nullen, über den Nullen auf der Diagonale verbleiben irgendwelche Einträge:

$$\left(\begin{array}{cccc|ccc|c} 1 & 0 & & 0 & & 0 & & d_1 \\ & \ddots & & \vdots & & \vdots & & \vdots \\ & & 1 & \vdots & & \vdots & & d_{i_1-1} \\ & & & \vdots & & \vdots & & 0 \\ & & & 1 & 0 & \vdots & & d_{i_1+1} \\ & & & & \ddots & \vdots & & \vdots \\ & & & & & 1 & & d_{i_2-1} \\ & & & & & & 1 & 0 \\ & & & & & & & d_{i_2+1} \\ & & & & & & & \vdots \\ & & & & & & & \vdots \\ & & & & & & & \vdots \end{array} \right). \quad (\#\#)$$

Entzerrung:

- Ersetze die Nullen auf der Diagonale durch -1 .

Das Schema hat nun die Endform

$$\left(\begin{array}{cccc|ccc|c} 1 & 0 & & 0 & & 0 & & d_1 \\ & \ddots & & \vdots & & \vdots & & \vdots \\ & & 1 & \vdots & & \vdots & & d_{i_1-1} \\ & & & \vdots & & \vdots & & 0 \\ & & & 1 & 0 & \vdots & & d_{i_1+1} \\ & & & & \ddots & \vdots & & \vdots \\ & & & & & 1 & & d_{i_2-1} \\ & & & & & & 1 & 0 \\ & & & & & & & d_{i_2+1} \\ & & & & & & & \vdots \\ & & & & & & & \vdots \\ & & & & & & & \vdots \end{array} \right), \quad (\#\#\#)$$

an dem sich die allgemeine Lösung unmittelbar ablesen läßt:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_{i_1-1} \\ 0 \\ d_{i_1+1} \\ \vdots \\ d_{i_2-1} \\ 0 \\ d_{i_2+1} \\ \vdots \end{pmatrix} + \lambda_1 \cdot \begin{pmatrix} * \\ \vdots \\ * \\ \boxed{-1} \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} * \\ \vdots \\ * \\ 0 \\ * \\ \vdots \\ * \\ \boxed{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots$$

mit beliebigen freien Konstanten λ_1, λ_1 etc.

Definition 12.67: (Kernvektoren einer Matrix)

Ein Vektor $\vec{k} \in K^m$ einer Matrix $A \in K^{n \times m}$ mit $A\vec{k} = 0$ heißt „**Kernvektor**“ von A . Die Menge aller Kernvektoren bilden entweder die leere Menge oder einen linearen Unterraum im K^m (den „**Kern**“ von A). Die Dimension des Kerns heißt „**Defekt**“ der Matrix. Der „**Rang**“ einer $(n \times m)$ -Matrix ist

$$rg(A) = \min(n, m) - \text{Defekt}(A).$$

Satz 12.68: (Die allgemeine Lösung linearer Systeme)

Die allgemeine Lösung eines lösbaren linearen Gleichungssystems $A\vec{x} = \vec{b}$ ist

$$\vec{x} = \vec{x}_0 + \sum_j \lambda_j \cdot \vec{k}_j,$$

wobei \vec{x}_0 irgendeine spezielle Lösung von $A\vec{x} = \vec{b}$ ist, \vec{k}_1, \vec{k}_2 etc. eine Basis des Kerns von A bilden und λ_1, λ_2 etc. beliebige freie Parameter sind.

Beweis: Sei \vec{x} eine beliebige Lösung von $A\vec{x} = \vec{b}$, sei \vec{x}_0 eine spezielle Lösung. Die Differenz $\vec{k} = \vec{x} - \vec{x}_0$ erfüllt dann $A\vec{k} = A(\vec{x} - \vec{x}_0) = A\vec{x} - A\vec{x}_0 = \vec{b} - \vec{b} = 0$, ist also ein Kernvektor.

Q.E.D.

Jedes markierte Diagonalelement -1 in ($\#\#\#$) entspricht einem Kernvektor der Ausgangsmatrix, der sich als die entsprechende Spalte des Endschemas ablesen läßt. Existieren mehrere solcher eingefügten Diagonalelemente -1 , sind die entsprechenden Kernvektoren automatisch linear unabhängig, denn sie haben als untersten Eintrag jeweils -1 an unterschiedlichen Stellen:

Die Dimension des Kerns der Matrix ist durch die Anzahl der Diagonal-Nullen in (#) bzw. (##) gegeben.

Fazit:

Der Gauß-Algorithmus wie oben beschrieben liefert nach Entzerrung des oberen Gleichungssystems sowohl eine spezielle Lösung $\vec{x}_0 = (d_1, d_2, \dots)^T$ (mit 0 in den Komponenten, die den Kernindizes entsprechen) sowie einen Satz von linear unabhängigen Kernvektoren (mit -1 in den Komponenten, die den Kernindizes entsprechen). Diese Kernvektoren bilden eine Basis des Kerns von A . Wir haben damit die allgemeine Lösung des Ausgangssystems $A\vec{x} = \vec{b}$ gefunden.

Beispiel 12.69: Wir berechnen die allgemeine Lösung des linearen Gleichungssystems (4 Gleichungen für 5 Unbekannte):

$$\begin{pmatrix} 0 & 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 3 & 1 \\ 1 & 2 & 3 & 5 & 2 \\ 1 & 2 & 3 & 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \\ 16 \\ 16 \end{pmatrix}. \quad (\#)$$

Wir bilden die erweiterte Koeffizientenmatrix:

$$\left(\begin{array}{ccccc|c} 0 & 1 & 1 & 2 & 1 & 6 \\ 1 & 1 & 2 & 3 & 1 & 10 \\ 1 & 2 & 3 & 5 & 2 & 16 \\ 1 & 2 & 3 & 6 & 1 & 16 \end{array} \right).$$

Gauß-Elimination der ersten Spalte. Zunächst vertauschen der Zeilen 1 und 2:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 3 & 1 & 10 \\ 0 & 1 & 1 & 2 & 1 & 6 \\ 1 & 2 & 3 & 5 & 2 & 16 \\ 1 & 2 & 3 & 6 & 1 & 16 \end{array} \right).$$

Erzeugen der Nullen in der ersten Spalte:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 3 & 1 & 10 \\ 0 & 1 & 1 & 2 & 1 & 6 \\ 0 & 1 & 1 & 2 & 1 & 6 \\ 0 & 1 & 1 & 3 & 0 & 6 \end{array} \right).$$

Erzeuge Nullen in der zweiten Spalte:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 3 & 1 & 10 \\ 0 & 1 & 1 & 2 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{array} \right).$$

Obere Dreiecksform ist erreicht. Hinzufügen einer trivialen Gleichung, damit ein quadratisches (5×5)-System entsteht:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 3 & 1 & 10 \\ 0 & 1 & 1 & 2 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right).$$

Reduktion auf Treppenstufenform (die Zeilen mit 0 auf der Diagonalen sind zu Nullzeilen zu machen). Es ist hier nichts zu tun. Wir markieren lediglich die Nullen längs der Diagonale:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 3 & 1 & 10 \\ 0 & 1 & 1 & 2 & 1 & 6 \\ 0 & 0 & \boxed{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{0} & 0 \end{array} \right).$$

Erzeuge Nullen über den nicht-verschwindenden Diagonalelementen. Zunächst die Nullen in der 4-ten Spalte. Ziehe dazu das 2-fache der 4-ten Zeile von der ersten und das 3-fache der 4-ten Zeile von der zweiten ab:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 0 & 4 & 10 \\ 0 & 1 & 1 & 0 & 3 & 6 \\ 0 & 0 & \boxed{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{0} & 0 \end{array} \right).$$

Erzeuge Nullen in der 2-ten Spalte. Subtrahiere dazu die zweite Zeile von der ersten:

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & 1 & 4 \\ 0 & 1 & 1 & 0 & 3 & 6 \\ 0 & 0 & \boxed{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{0} & 0 \end{array} \right).$$

Die 3-te und die 5-te Spalte haben 0 auf der Diagonale und liefern damit 2 Kernvektoren, indem man -1 auf der Diagonale einträgt:

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & 1 & 4 \\ 0 & 1 & 1 & 0 & 3 & 6 \\ 0 & 0 & \boxed{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{-1} & 0 \end{array} \right).$$

Die allgemeine Lösung des Ausgangssystems (#) hat damit zwei beliebige freie Parameter λ_1 und λ_2 :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_1 \cdot \begin{pmatrix} 1 \\ 1 \\ \boxed{-1} \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 1 \\ 3 \\ 0 \\ -1 \\ \boxed{-1} \end{pmatrix}.$$

12.6 Inverse von Matrizen

Zu einer invertierbaren $n \times n$ -Matrix A soll die Inverse A^{-1} berechnet werden. Zwar läßt sich dies mit der im nächsten Abschnitt vorgestellten Gauß-Jordan-Variante des Gauß-Algorithmus leicht bewerkstelligen, in der Praxis wird eine explizite Inverse von A jedoch selten benötigt (man hat in der Regel $A\vec{x} = \vec{b}$ mit einer konkret vorgegebenen rechten Seite \vec{b} zu lösen). In einigen Fällen hat man immer wieder $A\vec{x} = \vec{b}$ zu lösen, wobei die Matrix A unverändert bleibt und sich im Laufe einer Iteration die rechte Seite \vec{b} immer wieder neu ergibt. In diesem Fall mag man versucht sein, A^{-1} explizit zu berechnen und zu speichern, um $A\vec{x} = \vec{b}$ durch eine Matrix-Vektor-Multiplikation $A^{-1}\vec{b}$ mit $O(n^2)$ Operationen berechnen zu können, anstatt immer wieder erneut den Gauß-Algorithmus mit einem Aufwand von $O(n^3)$ Operationen auf $A\vec{x} = \vec{b}$ loszulassen. In Abschnitt 12.6.2 zeigt es sich jedoch, dass mit einem Durchlauf des Gauß-Algorithmus eine sogenannte LR -Zerlegung vorliegt, die genauso gut ist wie eine explizite Inverse A^{-1} .

12.6.1 Die Gauß-Jordan-Variante

Statt einer einzelnen Gleichung $A\vec{x} = \vec{b}$ mit $A \in K^{n \times n}$ und einer einzelnen rechten Seite $\vec{b} \in K^n$ lösen wir simultan $A\vec{x} = \vec{e}_i$ mit $i = 1, \dots, n$, wobei \vec{e}_i die Standardbasis des K^n darstellt:

Bezeichnung 12.70:

Die Standardbasis des K^n wird mit $\vec{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ ($i = 1, \dots, n$) bezeichnet.

Dies läuft darauf hinaus, dass wir das erweiterte Koeffizientenschema $[A, \vec{e}_1, \dots, \vec{e}_n] = [A, \mathbf{I}]$ betrachten. Wir lassen den Gauß-Algorithmus 12.64 auf das Schema los, wobei die Elementaroperationen auf das Gesamtschema angewendet werden (d.h., die Schleife über k in 12.64 ist bis zur Obergrenze $2 \cdot n$ (statt $m + 1$) auszudehnen). Als Zwischenergebnis erhält man das Schema

$$[R, B]$$

mit einer oberen Dreiecksmatrix R . Verschwindet mindestens ein Diagonalelement von R , so ist R nicht invertierbar (und damit ist auch die Ausgangsmatrix A nicht invertierbar), und wir können die Berechnung von A^{-1} abbrechen. Seien also nun alle Diagonalelemente von R ungleich 0. Wir dividieren nun jede Zeile von $[R, B]$ durch das entsprechende Diagonalelement von R , so dass längs der Diagonale nur Einsen stehen. Dann eliminieren wir mit der letzten Spalte von R

beginnend „nach oben“, indem wir durch Subtraktion geeigneter Vielfache einer Zeile von den darüber liegenden Zeilen Nullen über den Diagonalelementen erzeugen. Diese Elementaroperationen sind dabei jeweils auf das gesamte Schema anzuwenden. Zuletzt erhalten wir im linken Teil des Schemas die Einheitsmatrix:

$$[\mathbf{I}, C].$$

Satz 12.71:

Die Matrix C im rechten Teils des Schemas ist die Inverse A^{-1} der Ausgangsmatrix A .

Beweis: Alle Schritte des Gauß-Algorithmus zur Erzeugung der oberen Dreiecksform R als auch die Normierung der Diagonalelemente auf 1 und die folgende „Elimination nach oben“ sind durch Multiplikation mit Elementarmatrizen 12.63 E_1, E_2 etc. zu beschreiben, die auch auf den rechten Teil des Schemas angewendet werden:

$$[\mathbf{I}, C] = [E_k \cdots E_2 E_1 A, E_k \cdots E_2 E_1 \mathbf{I}].$$

Vergleich der linken Teilschemata liefert

$$\mathbf{I} = E_k \cdots E_2 E_1 A,$$

d.h., es gilt $E_k \cdots E_2 E_1 = A^{-1}$. Vergleich der rechten Teilschemata liefert $C = E_k \cdots E_2 E_1 \mathbf{I} = A^{-1}$.

Q.E.D.

12.6.2 LR-Zerlegung

Aus Zeitgründen wird dieser Abschnitt über die LR -Zerlegung in der Vorlesung nicht vorgestellt und hiermit als „nicht prüfungsrelevant“ deklariert. ;-)

Wir betrachten noch einmal die Zeilenvertauschungen, die im Gauß-Algorithmus durchzuführen sind. Diese sind als Matrixmultiplikationen mit mit

$$P = \begin{pmatrix} \mathbf{I} & & & \\ & 0 & 1 & \\ & & \mathbf{I} & \\ & 1 & 0 & \\ & & & \mathbf{I} \end{pmatrix} \begin{matrix} \leftarrow i \\ \\ \leftarrow j \\ \\ \end{matrix} = \mathbf{I} - \vec{e}_i \vec{e}_i^T - \vec{e}_j \vec{e}_j^T + \vec{e}_i \vec{e}_j^T + \vec{e}_j \vec{e}_i^T$$

zu interpretieren: $A \rightarrow PA$: Vertauschung der Zeilen i und j ,
 $A \rightarrow AP$: Vertauschung der Spalten i und j ,

Definition 12.72:

Eine **Permutationsmatrix** ist eine quadratische Matrix, die in jeder Zeile und jeder Spalte mit genau einer 1 und ansonsten mit 0 besetzt ist.

Lemma 12.73:

- a) Produkte von Permutationsmatrizen sind wieder Permutationsmatrizen.
 b) Permutationsmatrizen sind orthogonal: $PP^T = P^T P = \mathbf{1}$.
 c) Linksmultiplikation mit einer Permutationsmatrix entspricht einer Zeilenvertauschung, Rechtsmultiplikation einer Spaltenvertauschung.

Beweis: offensichtliche Übung.

Beispiel 12.74: Sei $P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, $A = \begin{pmatrix} \vec{z}_1^T \\ \vec{z}_2^T \\ \vec{z}_3^T \end{pmatrix} = (\vec{s}_1, \vec{s}_2, \vec{s}_3)$ (Zerlegung von A nach Zeilen bzw. Spalten). Es gilt $PA = \begin{pmatrix} \vec{z}_2^T \\ \vec{z}_3^T \\ \vec{z}_1^T \end{pmatrix}$, $AP = (\vec{s}_3, \vec{s}_1, \vec{s}_2)$.

Satz 12.75: (LR-Zerlegung)

Das System $A\vec{x} = \vec{b}$ wird im Gauß-Algorithmus 12.64 auf die Endform

$$\left(\begin{array}{cccc|c} r_{11} & \cdots & \cdots & r_{1n} & c_1 \\ l_{21} & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ l_{n1} & \cdots & l_{n,n-1} & r_{nn} & c_n \end{array} \right)$$

gebracht (der untere Dreiecksanteil (l_{ij}) entsteht aus den abgespeicherten Eliminationsfaktoren). Es seien P_1, \dots, P_{n-1} (eventuell mit $P_j = \mathbf{1}$) die bei der Pivotierung durchgeführten Zeilenvertauschungen. Mit

$$L = \begin{pmatrix} 1 & & & 0 \\ l_{11} & \ddots & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{pmatrix}, \quad R = \begin{pmatrix} r_{11} & \cdots & \cdots & r_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \vdots \\ 0 & & & r_{nn} \end{pmatrix}$$

und $P = P_{n-1} \cdots P_1$ gilt: $PA = LR$ (“LR-Faktorisierung”).

Beweis: (für technisch Interessierte)

Zu Beginn des j .ten Schrittes liegt das Zwischenschema

$$\left(\begin{array}{cccc|ccc} r_{11} & \cdots & \cdots & r_{1,j-1} & r_{1j} & \cdots & r_{1n} \\ l_{21} & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ l_{j-1,j} & \cdots & l_{j-1,j-2} & r_{j-1,j-1} & r_{j-1,j} & \cdots & r_{j-1,n} \\ \hline l_{j1}^{(j)} & \cdots & \cdots & l_{j,j-1}^{(j)} & a_{jj}^{(j)} & \cdots & a_{jn}^{(j)} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ l_{n1}^{(j)} & \cdots & \cdots & l_{n,j-1}^{(j)} & a_{nj}^{(j)} & \cdots & a_{nn}^{(j)} \end{array} \right) = L^{(j)} + R^{(j)}$$

mit

$$L^{(j)} = \left(\begin{array}{ccc|c} 0 & & 0 & \\ l_{21} & \ddots & & 0 \\ \vdots & \ddots & 0 & \\ \hline l_{j1}^{(j)} & \ddots & l_{j,j-1}^{(j)} & \\ \vdots & \ddots & \vdots & 0 \\ l_{n1}^{(j)} & \cdots & l_{n,j-1}^{(j)} & \end{array} \right)$$

und

$$R^{(j)} = \left(\begin{array}{ccc|ccc} r_{11} & \cdots & \cdots & \cdots & \cdots & r_{1n} \\ & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & r_{j-1,j-1} & \cdots & \cdots & r_{j-1,n} \\ \hline & & & a_{jj}^{(j)} & \cdots & a_{jn}^{(j)} \\ & & 0 & \vdots & \ddots & \vdots \\ & & & a_{nj}^{(j)} & \cdots & a_{nn}^{(j)} \end{array} \right)$$

vor. Die nächste Zeilenvertauschung P_j liefert $\tilde{L}^{(j)} = P_j L^{(j)}$ und

$$\tilde{R}^{(j)} = P_j R^{(j)} = \left(\begin{array}{ccc|ccc} r_{11} & \cdots & \cdots & \cdots & \cdots & r_{1n} \\ & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & r_{j-1,j-1} & \cdots & \cdots & r_{j-1,n} \\ \hline & & & r_{jj} & \cdots & r_{jn} \\ & & & \tilde{a}_{j+1,j}^{(j)} & \cdots & \tilde{a}_{j+1,n}^{(j)} \\ & & 0 & \vdots & \ddots & \vdots \\ & & & \tilde{a}_{nj}^{(j)} & \cdots & \tilde{a}_{nn}^{(j)} \end{array} \right).$$

Die neuen Eliminationsfaktoren $\vec{l}_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \tilde{a}_{j+1,j}^{(j)}/r_{jj} \\ \vdots \\ \tilde{a}_{n,j}^{(j)}/r_{jj} \end{pmatrix}$ werden als j .te Spalte

an $\tilde{L}^{(j)}$ angefügt:

$$L^{(j+1)} = \tilde{L}^{(j)} + \vec{l}_j \vec{e}_j^T .$$

Vielfache der j .ten Zeile $\vec{e}_j^T \tilde{R}^{(j)} = (0, \dots, 0, r_{jj}, \dots, r_{jn})$ von $\tilde{R}^{(j)}$ werden von den unteren Zeilen von $\tilde{R}^{(j)}$ abgezogen:

$$R^{(j+1)} = \tilde{R}^{(j)} - \vec{l}_j \vec{e}_j^T \tilde{R}^{(j)} .$$

Beachte hierzu

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ l_{j+1} \\ \vdots \\ l_n \end{pmatrix} (0, \dots, 0, r_{jj}, \dots, r_{jn}) = \left(\begin{array}{c|ccc} 0 & & & 0 \\ \hline & l_{j+1}r_{jj} & \cdots & l_{j+1}r_{jn} \\ & 0 & \ddots & \vdots \\ & l_n r_{jj} & \cdots & l_n r_{jn} \end{array} \right) .$$

Ergebnis: der Gauß-Algorithmus wird beschrieben durch

$$L^{(j+1)} = P_j L^{(j)} + \vec{l}_j \vec{e}_j^T , \quad R^{(j+1)} = P_j R^{(j)} - \vec{l}_j \vec{e}_j^T P_j R^{(j)}$$

mit $L^{(1)} = 0$, $R^{(1)} = A$.

Behauptung: In jedem Schritt gilt $(\mathbf{I} + L^{(j)})R^{(j)} = P_{j-1} \cdots P_1 A$.

(Mit $L = \mathbf{I} + L^{(n)}$, $R = R^{(n)}$, $P = P_{n-1} \cdots P_1$ ist der Satz damit bewiesen).

Beweis: Induktion nach j :

$j = 1$: $(\mathbf{I} + L^{(1)})R^{(1)} = A$ gilt offensichtlich.

Schritt $j \rightarrow j + 1$:

$$\begin{aligned} & (\mathbf{I} + L^{(j+1)}) R^{(j+1)} \\ &= (\mathbf{I} + P_j L^{(j)} + \vec{l}_j \vec{e}_j^T) (\mathbf{I} - \vec{l}_j \vec{e}_j^T) P_j R^{(j)} \\ &= (\mathbf{I} - \cancel{\vec{l}_j \vec{e}_j^T} + P_j L^{(j)} - P_j L^{(j)} \vec{l}_j \vec{e}_j^T + \cancel{\vec{l}_j \vec{e}_j^T} - \vec{l}_j \vec{e}_j^T \vec{l}_j \vec{e}_j^T) P_j R^{(j)} \\ &= (\mathbf{I} + P_j L^{(j)}) P_j R^{(j)} \quad (\text{beachte } L^{(j)} \vec{l}_j = 0, \vec{e}_j^T \vec{l}_j = 0) \\ &= P_j (\mathbf{I} + L^{(j)} P_j) R^{(j)} \stackrel{(*)}{=} P_j (\mathbf{I} + L^{(j)}) R^{(j)} = P_j (P_{j-1} \cdots P_1 A) . \end{aligned}$$

Für (*) beachte $L^{(j)}P_j = L^{(j)}$, da Rechtsmultiplikation mit P_j die Spalten j, \dots, n von $L^{(j)}$ vertauscht, die alle gleich (nämlich 0) sind.

Q.E.D.

Bemerkung 12.76: $L = \begin{pmatrix} 1 & & \\ * & \ddots & \\ * & * & 1 \end{pmatrix}$ ist immer invertierbar. Damit ist

$R = L^{-1}PA$ genau dann invertierbar, wenn A invertierbar ist.

Bemerkung 12.77: Zu jeder quadratischen Matrix existiert damit eine Zeilenvertauschung P , so daß PA in der angegebenen Form LR faktorisiert werden kann. Bei invertierbarem A und fixiertem P sind die LR -Faktoren eindeutig:

$$LR = \tilde{L}\tilde{R} \implies \underbrace{\tilde{L}^{-1}L}_{\text{untere Dreiecksmatrix}} = \underbrace{\tilde{R}R^{-1}}_{\text{obere Dreiecksmatrix}} = \text{Diagonalmatrix } D.$$

Mit $L = \tilde{L}D$ folgt $D = \mathbb{1}$, da die Diagonalen von L und \tilde{L} mit 1 besetzt sind.

Bemerkung 12.78: Es gibt Matrizen, für die keine Faktorisierung $A = LR$ existiert, man braucht i.a. P . Beispiel:

$$\begin{pmatrix} 0 & * \\ 1 & * \end{pmatrix} \stackrel{(?)}{=} \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} \\ l_{21} \cdot r_{11} & l_{21} \cdot r_{12} + r_{22} \end{pmatrix}.$$

Aus der ersten Spalte ergeben sich die Gleichungen

$$0 \stackrel{(!)}{=} r_{11}, \quad 1 \stackrel{(!)}{=} l_{21} \cdot r_{11},$$

die keine Lösung für l_{21} und r_{11} zulassen. In der Tat braucht der Gauß-Algorithmus für die Matrix $\begin{pmatrix} 0 & * \\ 1 & * \end{pmatrix}$ gleich im ersten Schritt wegen der 0 links oben eine Zeilenvertauschung, ohne die der Algorithmus zum Erliegen kommen würde.

Bemerkung 12.79: P ist die Gesamtheit aller im Gauß-Algorithmus durchgeführten Zeilenvertauschungen. Man speichert P nicht als Matrix, sondern als "Buchhaltungsvektor" \vec{p} . Dazu initialisiert man die Spalte $\vec{p} = (1, 2, \dots, n)^T$ und wendet die Zeilenvertauschungen auch auf \vec{p} an. Zuletzt hat man $\vec{p} = (p_1, \dots, p_n)^T$. Interpretation: die ursprüngliche Zeile p_i ist in die i .te Zeile gewandert. Hiermit ergibt sich für beliebiges \vec{b} dann $\vec{c} = P\vec{b}$ durch

for $i := 1$ to n do $c_i := b_{p_i}$;

Bemerkung 12.80: Mit $\det(L) = 1$ und leicht berechenbarem

$$\det(R) = \text{Produkt der Diagonalelemente}$$

kann $\det(A)$ aus $\det(P)\det(A) = \det(R)$ berechnet werden. Mit $\det(P) = \det(P_{n-1}) \cdots \det(P_1)$ und

$$\det(P_j) = \begin{cases} 1 & , \text{ falls } P_j = \mathbb{I} \\ -1 & , \text{ falls } P_j = \text{Austausch zweier Zeilen} \end{cases}$$

braucht nur die Anzahl der nichttrivialen Zeilenvertauschungen im Gauß-Algorithmus mitgezählt zu werden.

Bemerkung 12.81: Die Faktorisierungsdaten P, L, R sind genauso gut wie eine explizite Inverse A^{-1} ! Die Lösung von

$$A\vec{x} = \vec{b} \implies PA\vec{x} = P\vec{b}, \quad \text{d.h., } L \underbrace{R\vec{x}}_{\vec{y}} = P\vec{b}$$

ist durch 2 Rücksubstitutionen

$$\begin{aligned} & \text{bestimme } \vec{y} \text{ aus } L\vec{y} = P\vec{b}, \\ \text{dann} & \text{ bestimme } \vec{x} \text{ aus } R\vec{x} = \vec{y} \end{aligned}$$

schnell zu berechnen. Der Aufwand (etwa n^2 Multiplikationen) entspricht dem der Matrix-Vektor-Multiplikation $A^{-1}\vec{b}$ bei explizit vorliegendem A^{-1} .

Merke: Der Gauß-Algorithmus liefert mit $\approx n^3/3$ Multiplikationen die Zerlegung $PA = LR$ einer $n \times n$ -Matrix. Liegen P, L, R vor, dann ist $A\vec{x} = \vec{b}$ durch doppelte Rücksubstitution mit etwa n^2 Multiplikationen schnell zu lösen.

12.7 Eigenwerte und -vektoren

4.2.04↓

Die „Spektralanalyse“ = „Diagonalisierung“ linearer Abbildungen (das Bestimmen von Eigenwerten und -vektoren) ist speziell für physikalische Anwendungen von zentraler Bedeutung.

12.7.1 Definitionen

Definition 12.82: (Eigenwerte und -vektoren)

Zu einer quadratischen Matrix $A \in K^{n \times n}$ gebe es $\lambda \in \mathbb{C}$ und $\vec{x} \in \mathbb{C}^n$ mit $\vec{x} \neq 0$, so dass

$$A\vec{x} = \lambda \cdot \vec{x}.$$

Der Wert λ heißt „**Eigenwert**“ von A , der Vektor \vec{x} wird als „**Eigenvektor**“ zum Eigenwert λ bezeichnet. Die Vektormenge

$$E_\lambda = \left\{ \vec{x} \in \mathbb{C}^n; A\vec{x} = \lambda \cdot \vec{x} \right\}$$

heißt „**Eigenraum**“ zum Eigenvektor λ .

Bemerkung 12.83: Man überzeugt sich leicht davon, dass zu einem gegebenen Eigenwert λ von A der Eigenraum E_λ ein Unterraum des \mathbb{C}^n ist. Speziell heißt dies, dass Eigenvektoren nicht eindeutig sind. Jedes Vielfache eines Eigenvektors ist wieder ein Eigenvektor (falls das Vielfache nicht gerade 0 ist). Ist die Dimension des Eigenraums E_λ größer als 1, können auch linear unabhängige Eigenvektoren zu λ zu einem neuen Eigenvektor kombiniert werden.

Definition 12.84: (Das charakteristische Polynome einer Matrix)

Zu $A = (a_{ij}) \in K^{n \times n}$ definieren wir das „**charakteristische Polynom**“

$$p_A(\lambda) = \det(A - \lambda \cdot \mathbf{I}) = \det \left(\begin{pmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} - \lambda \end{pmatrix} \right)$$

Man hat sich hierbei schnell davon überzeugt, dass das charakteristische Polynom

$$p_A(\lambda) = (-1)^n \cdot \lambda^n + c_{n-1} \cdot \lambda^{n-1} + \dots + c_1 \cdot \lambda + c_0$$

in der Tat ein Polynom vom Grad n ist (um den führenden Koeffizienten $(-1)^n$ einzusehen, betrachte man einfach $A = 0$). Einige weitere der Koeffizienten c_i des Polynoms sind ebenfalls leicht der Matrix zuzuordnen. Z.B. gilt:

$$c_{n-1} = (-1)^{n-1} \cdot \underbrace{\sum_{i=1}^n a_{ii}}_{\text{Spur}(A)}, \quad c_0 = \det(A).$$

Zu c_0 : Werte einfach das charakteristische Polynom an der Stelle $\lambda = 0$ aus, wodurch sich nach Definition $c_0 = p_A(0) = \det(A - 0 \cdot \mathbf{I}) = \det(A)$ ergibt.

Satz 12.85: (Charakterisierung der Spektraldaten einer Matrix)

Die Eigenwerte einer Matrix $A \in K^{n \times n}$ sind die Nullstellen des charakteristischen Polynoms. Die zum Eigenwert λ gehörigen Eigenvektoren sind die nicht-verschwindenden Kernvektoren der Matrix $A - \lambda \cdot \mathbb{I}$.

Beweis: Die Gleichung $A\vec{x} = \lambda \cdot \vec{x}$ kann als $(A - \lambda \cdot \mathbb{I})\vec{x} = 0$ geschrieben werden. Ist $A - \lambda \cdot \mathbb{I}$ invertierbar, gibt es nur die eindeutige Lösung $\vec{x} = 0$, die wir nicht als Eigenvektor zulassen. Da $A - \lambda \cdot \mathbb{I}$ genau dann nicht invertierbar ist, wenn $p_A(\lambda) = \det(A - \lambda \cdot \mathbb{I})$ verschwindet, sind die Eigenwerte die Nullstellen von $p_A(\lambda)$. Für nicht-invertierbares $A_\lambda := A - \lambda \cdot \mathbb{I}$ existiert ein Kern der Matrix A_λ , der aus Eigenvektoren von A besteht.

Q.E.D.

Wir haben damit unmittelbar eine algorithmische Möglichkeit, sowohl die Eigenwerte als auch die dazugehörigen Eigenvektoren zu bestimmen:

- a) Bestimme die Eigenwerte als die Nullstellen λ_1, λ_2 etc. des charakteristischen Polynoms.
- b) Berechne zum Eigenwert λ_i eine Basis des Kerns von $A - \lambda_i \cdot \mathbb{I}$. Dies kann beispielsweise durch den Gauß-Algorithmus angewendet auf die Gleichung $(A - \lambda \cdot \mathbb{I})\vec{x} = 0$ geschehen, wobei die Kernvektoren sich gemäß Abschnitt 12.5.3 algorithmisch ergeben. Dies liefert eine Basis des Eigenraums E_{λ_i} .

Beispiel 12.86: Zur Matrix

$$A = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}$$

ergibt sich das charakteristische Polynom

$$p_A(\lambda) = \det \left(\begin{pmatrix} -\lambda & 1 \\ -2 & 3 - \lambda \end{pmatrix} \right) = -\lambda \cdot (3 - \lambda) + 2 = \lambda^2 - 3 \cdot \lambda + 2$$

mit den Nullstellen (Eigenwerten)

$$\lambda_1 = 1, \quad \lambda_2 = 2.$$

Die allgemeine Lösung von $(A - \lambda_1 \cdot \mathbb{I})\vec{x} = 0$ ist der Kern der Matrix

$$A - \lambda_1 \cdot \mathbb{I} = \begin{pmatrix} -1 & 1 \\ -2 & 2 \end{pmatrix}.$$

Über den Entzerrungsalgorithmus ergibt sich die allgemeine Lösung $\vec{x} = \alpha \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ mit beliebigen $\alpha \in \mathbb{C}$. Wir wählen als Basis des 1-dimensionalen Kerns den Eigenvektor

$$\vec{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Die allgemeine Lösung von $(A - \lambda_2 \cdot \mathbf{1}) \vec{x} = 0$ ist der Kern der Matrix

$$A - \lambda_2 \cdot \mathbf{1} = \begin{pmatrix} -2 & 1 \\ -2 & 1 \end{pmatrix}.$$

Über den Entzerrungsalgorithmus ergibt sich die allgemeine Lösung $\vec{x} = \alpha \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ mit beliebigen $\alpha \in \mathbb{C}$. Wir wählen als Basis des 1-dimensionalen Kerns den Eigenvektor

$$\vec{x}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Wir gehen nun der Frage nach, wieviele Eigenvektoren man zu einer $n \times n$ -Matrix A finden kann. Speziell: kann ich die Eigenvektoren als Basis des \mathbb{C}^n benutzen?

Satz 12.87:

Eigenvektoren zu unterschiedlichen Eigenwerten sind linear unabhängig.

Beweis: Seien $\vec{x}_1, \dots, \vec{x}_k$ Eigenvektoren von $A \in K^{n \times n}$ zu den Eigenwerten $\lambda_1, \dots, \lambda_k$, die alle paarweise verschieden seien. Wir betrachten die Linearkombination

$$\sum_{i=1}^k c_i \cdot \vec{x}_i = 0 \quad (\#)$$

und müssen hieraus schliessen, dass $c_1 = \dots = c_k = 0$ folgt.

Wir machen dazu Induktion nach k . Für $k = 1$ folgt sicherlich $c_1 = 0$, da \vec{x}_1 als Eigenvektor nicht verschwindet. Wir machen nun den Induktionsschritt $k - 1 \rightarrow k$. Seien also $\vec{x}_1, \dots, \vec{x}_{k-1}$ linear unabhängige Eigenvektoren zu $\lambda_1, \dots, \lambda_{k-1}$, sei \vec{x}_k ein zusätzlicher Eigenvektor zum Eigenwert λ_k . Aus (#) folgt durch Anwendung von A

$$\sum_{i=1}^k c_i \cdot \lambda_i \cdot \vec{x}_i = 0$$

und durch Multiplikation von (#) mit λ_k

$$\sum_{i=1}^k c_i \cdot \lambda_k \cdot \vec{x}_i = 0.$$

Durch Differenzbildung folgt

$$\sum_{i=1}^{k-1} c_i \cdot (\lambda_i - \lambda_k) \cdot \vec{x}_i = 0,$$

wobei der Term mit \vec{v}_k herausfällt. Nach Induktionsvoraussetzung folgt

$$c_i \cdot (\lambda_i - \lambda_k) = 0$$

wegen der linearen Unabhängigkeit von $\vec{v}_1, \dots, \vec{v}_{k-1}$ und damit $c_1 = \dots = c_{k-1} = 0$, da $\lambda_i - \lambda_k \neq 0$. Damit reduziert sich (#) zu $\lambda_k \cdot \vec{v}_k = 0$, woraus auch $c_k = 0$ folgt.

Q.E.D.

Die Frage, ob man den gesamten \mathbb{C}^n durch Eigenvektoren einer Matrix aufspannen kann, läßt sich damit für den Spezialfall leicht beantworten, dass alle Eigenwerte unterschiedlich sind:

Folgerung 12.88:

Hat eine $n \times n$ -Matrix n unterschiedliche Eigenwerte, so bilden die entsprechenden Eigenvektoren eine Basis des \mathbb{C}^n .

Beweis: Nach Satz 12.87 gehören zu den n unterschiedlichen Eigenwerten n linear unabhängige Eigenvektoren. Mit Satz 12.11.b) bilden diese Vektoren eine Basis des \mathbb{C}^n und spannen den gesamten Raum auf.

Q.E.D.

Die Sache wird deutlich komplizierter, wenn es „entartete Eigenwerte“ gibt.

Definition 12.89: (mehrfache Eigenwerte)

*Sei $\lambda \in \mathbb{C}$ ein Eigenwert von $A \in K^{n \times n}$. Die Vielfachheit der Nullstelle λ des charakteristischen Polynoms von A im Sinne von Definition 1.24 heißt „**algebraische Vielfachheit**“ des Eigenwerts. Im Gegensatz dazu wird die Dimension des Kerns von $A - \lambda \cdot \mathbb{1}$ als „**geometrische Vielfachheit**“ bezeichnet.*

Satz 12.90:

Für jeden Eigenwert einer Matrix gilt:

$$1 \leq \text{geometrische Vielfachheit} \leq \text{algebraische Vielfachheit}.$$

Beweis: Die Vielfachheit einer Polynomnullstelle (algebraische Vielfachheit des Eigenwerts) ist nach Definition ≥ 1 . Für jeden Eigenwert λ_0 einer Matrix A ist $A - \lambda_0 \cdot \mathbb{1}$ nicht invertierbar und hat damit einen nicht-trivialen Kern. Damit gilt, dass auch für die geometrische Vielfachheit eines Eigenwerts stets ≥ 1 gilt. Nun zur Abschätzung

$$\text{geometrische Vielfachheit} \leq \text{algebraische Vielfachheit}.$$

Zunächst beobachten wir, dass eine Matrix $A \in K^{n \times n}$ und $T^{-1}AT$ (mit einer beliebigen invertierbaren Matrix $T \in K^{n \times n}$) das selbe charakteristische Polynom haben:

$$\begin{aligned} \det(T^{-1}AT - \lambda \cdot \mathbf{1}) &= \det(T^{-1}(A - \lambda \cdot \mathbf{1})T) \\ &= \det(T^{-1}) \cdot \det(T) \cdot \det(A - \lambda \cdot \mathbf{1}) \\ &= \det(T^{-1}T) \cdot \det(A - \lambda \cdot \mathbf{1}) = \det(A - \lambda \cdot \mathbf{1}). \end{aligned}$$

Seien $\vec{k}_1, \dots, \vec{k}_g$ Basisvektoren des Kerns von $A - \lambda_0 \cdot \mathbf{1}$ für einen Eigenwert λ_0 mit geometrischer Vielfachheit g . Wir wählen irgendwelche linear unabhängigen Vektoren $\vec{v}_{g+1}, \dots, \vec{v}_n$, die zusammen mit den Kernvektoren eine Basis des \mathbb{C}^n bilden. Betrachte nun

$$\begin{aligned} \tilde{A} &= \underbrace{[\vec{k}_1, \dots, \vec{k}_g, \vec{v}_{g+1}, \dots, \vec{v}_n]^{-1}}_{T^{-1}} A \underbrace{[\vec{k}_1, \dots, \vec{k}_g, \vec{v}_{g+1}, \dots, \vec{v}_n]}_T \\ &= T^{-1}[A \cdot \vec{k}_1, \dots, A \cdot \vec{k}_g, A\vec{v}_{g+1}, \dots, A\vec{v}_n] \\ &= T^{-1}[\lambda_0 \cdot \vec{k}_1, \dots, \lambda_0 \cdot \vec{k}_g, A\vec{v}_{g+1}, \dots, A\vec{v}_n] \\ &= [\lambda_0 \cdot T^{-1}\vec{k}_1, \dots, \lambda_0 \cdot T^{-1}\vec{k}_g, T^{-1}A\vec{v}_{g+1}, \dots, T^{-1}A\vec{v}_n]. \end{aligned}$$

Nach Konstruktion gilt

$$T^{-1} \underbrace{[\vec{k}_1, \dots, \vec{k}_g, \vec{v}_{g+1}, \dots, \vec{v}_n]}_T = [T^{-1}\vec{k}_1, \dots, T^{-1}\vec{k}_g, T^{-1}\vec{v}_{g+1}, \dots, T^{-1}\vec{v}_n]$$

$$= \left(\begin{array}{ccc|ccc} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \ddots & 1 & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \end{array} \right).$$

Betrachtet man die ersten g Spalten, folgt $T^{-1}\vec{k}_i = \vec{e}_i$ (Standardbasis), also

$$\tilde{A} = [\lambda_0 \cdot T^{-1}\vec{k}_1, \dots, \lambda_0 \cdot T^{-1}\vec{k}_g, \vec{*}, \dots, \vec{*}] = \left(\begin{array}{ccc|cc} \lambda_0 & 0 & \dots & * & \dots \\ 0 & \ddots & \ddots & \vdots & \ddots \\ \vdots & \ddots & \lambda_0 & \vdots & \ddots \\ \vdots & \ddots & 0 & \vdots & \ddots \\ \vdots & \ddots & \vdots & \vdots & \ddots \\ 0 & \dots & 0 & * & \dots \end{array} \right).$$

Das charakteristische Polynom von \tilde{A} ergibt sich durch rekursive Entwicklung von $\det(\tilde{A} - \lambda \cdot \mathbb{1})$ nach den ersten Spalten zu

$$\det(\tilde{A} - \lambda \cdot \mathbb{1}) = (\lambda_0 - \lambda)^g \cdot \tilde{p}(\lambda)$$

mit irgendeinem Faktorpolynom, das durch den rechten unteren Block von \tilde{A} bestimmt wird. Das obige charakteristische Polynom von \tilde{A} , das mit dem von A übereinstimmt, hat den Eigenwert λ_0 mit der Vielfachheit von mindestens g .

Q.E.D.

Seien $\lambda_1, \dots, \lambda_k$ die unterschiedlichen Nullstellen des charakteristischen Polynoms vom Grad n , ihre algebraischen Vielfachheiten seien n_1, \dots, n_k . Nach dem Fundamentalsatz der Algebra 1.29 addieren sich die algebraischen Vielfachheiten zu Dimension n des Raums auf: $n_1 + \dots + n_k = n$. Wenn dies auch für die geometrischen Vielfachheiten g_1, \dots, g_k gilt, so läßt sich aus den Eigenvektoren eine Basis bilden. Man wähle dazu aus jedem Kern von $A - \lambda_i \cdot \mathbb{1}$ jeweils g_i Basisvektoren, die insgesamt ein System von $g_1 + \dots + g_k$ linear unabhängigen Vektoren im \mathbb{C}^n bilden. Gilt $g_1 + \dots + g_k = n$, so haben wir eine Basis des \mathbb{C}^n vor uns. Die entscheidende Frage ist also:

Gilt immer geometrische Vielfachheit = algebraische Vielfachheit?

Die Antwort ist leider:

Nein!

Beispiel 12.91: Die Matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

hat das charakteristische Polynom

$$p_A(\lambda) = \det \left(\begin{pmatrix} -\lambda & 1 \\ 0 & -\lambda \end{pmatrix} \right) = \lambda^2.$$

Der Wert $\lambda = 0$ ist eine doppelte Nullstelle dieses Polynoms, damit ist $\lambda = 0$ ein Eigenwert mit der algebraischen Vielfachheit 2. Die Kernvektoren $\vec{k} = (k_1, k_2)^T$ von $A - 0 \cdot \mathbb{1} = A$ sind durch

$$A\vec{k} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

charakterisiert, also durch die einzelne Gleichung $k_2 = 0$. Der Kern von A besteht also aus dem 1-dimensionalen Raum

$$\left\{ \begin{pmatrix} k_1 \\ 0 \end{pmatrix}; k_1 \in \mathbb{C} \right\}.$$

Man findet also die algebraische Vielfachheit 2 von $\lambda = 0$, aber die geometrische Vielfachheit 1. Mit dem einzigen Eigenvektor $(1, 0)^T$, der den Kern aufspannt, kann man nicht den \mathbb{C}^2 aufspannen.

12.7.2 Symmetrische/hermitesche Matrizen

↓5.2.04

Zunächst einige einfache Beobachtungen zur den Eigenwerten und -vektoren zur (hermitesch) transponierten Matrix A^H (Definition 12.33) einer Matrix A :

Satz 12.92:

Ist λ Eigenwert von A , so ist $\bar{\lambda}$ Eigenwert von A^H .

Beweis: Wegen

$$\overline{\det(A - \lambda \cdot \mathbf{1})} = \overline{\det(A^T - \lambda \cdot \mathbf{1})} = \det(A^H - \bar{\lambda} \cdot \mathbf{1})$$

stimmen die charakteristischen Polynome von A und A^H bis auf komplexe Konjugation überein: $\overline{p_A(\lambda)} = p_{A^H}(\bar{\lambda})$.

Q.E.D.

Satz 12.93:

Sei \vec{x} ein Eigenvektor von A zum Eigenwert λ , sei \vec{y} ein Eigenvektor von A^H zum Eigenwert $\mu \neq \bar{\lambda}$. Dann gilt $\langle \vec{y}, \vec{x} \rangle = 0$.

Beweis: Wir betrachten das komplexe Skalarprodukt $\langle \vec{a}, \vec{b} \rangle = \sum_i \bar{a}_i \cdot b_i$ auf dem \mathbb{C}^n . Es gilt

$$\begin{aligned} \langle \vec{y}, A\vec{x} \rangle &= \langle \vec{y}, \lambda \cdot \vec{x} \rangle = \lambda \cdot \langle \vec{y}, \vec{x} \rangle \\ &= \langle A^H \vec{y}, \vec{x} \rangle = \langle \mu \cdot \vec{y}, \vec{x} \rangle = \mu \cdot \langle \vec{y}, \vec{x} \rangle. \end{aligned}$$

Es folgt

$$(\lambda - \mu) \cdot \langle \vec{y}, \vec{x} \rangle = 0,$$

womit das Skalarprodukt zwischen \vec{x} und \vec{y} für $\lambda \neq \bar{\mu}$ verschwinden muss.

Q.E.D.

Mit der selben Argumentation wie im obigen Beweis bekommt man im Fall hermitescher Matrizen die wichtige Aussage, dass alle Eigenwerte reell sind:

Satz 12.94:

Alle Eigenwerte hermitescher Matrizen $A = A^H$ sind reell.

Beweis: Wir betrachten das komplexe Skalarprodukt $\langle \vec{a}, \vec{b} \rangle = \sum_i \bar{a}_i \cdot b_i$ auf dem \mathbb{C}^n . Mit $A\vec{x} = \lambda \cdot \vec{x}$ folgt

$$\begin{aligned} \langle \vec{x}, A\vec{x} \rangle &= \langle \vec{x}, \lambda \cdot \vec{x} \rangle = \lambda \cdot \langle \vec{x}, \vec{x} \rangle \\ &= \langle A^H \vec{x}, \vec{x} \rangle = \langle \lambda \cdot \vec{x}, \vec{x} \rangle = \bar{\lambda} \cdot \langle \vec{x}, \vec{x} \rangle, \end{aligned}$$

also

$$(\lambda - \bar{\lambda}) \cdot \langle \vec{x}, \vec{x} \rangle = 0,$$

wobei $\langle \vec{x}, \vec{x} \rangle \neq 0$ gilt, da ein Eigenvektor nicht 0 sein kann. Aus $\lambda = \bar{\lambda}$ folgt, dass λ reell ist.

Q.E.D.

Satz 12.95:

Eigenvektoren hermitescher Matrizen $A = A^H$ zu verschiedenen Eigenwerten sind orthogonal.

Beweis: Dies ist der Spezialfall von Satz 12.93 für hermitesche Matrizen. Sei $A\vec{x} = \lambda \cdot \vec{x}$, $A\vec{y} = \mu \cdot \vec{y}$ mit reellen Eigenwerten $\lambda \neq \mu$. Wir betrachten das komplexe Skalarprodukt $\langle \vec{a}, \vec{b} \rangle = \sum_i \bar{a}_i \cdot b_i$ auf dem \mathbb{C}^n . Es gilt

$$\begin{aligned} \langle \vec{y}, A\vec{x} \rangle &= \langle \vec{y}, \lambda \cdot \vec{x} \rangle = \lambda \cdot \langle \vec{y}, \vec{x} \rangle \\ &= \langle A^H \vec{y}, \vec{x} \rangle = \langle \mu \cdot \vec{y}, \vec{x} \rangle = \bar{\mu} \cdot \langle \vec{y}, \vec{x} \rangle = \mu \cdot \langle \vec{y}, \vec{x} \rangle. \end{aligned}$$

Es folgt

$$(\lambda - \mu) \cdot \langle \vec{y}, \vec{x} \rangle = 0,$$

womit das Skalarprodukt zwischen \vec{x} und \vec{y} für $\lambda \neq \mu$ verschwinden muss.

Q.E.D.

Hier die zentrale Besonderheit des Eigenwertproblems für hermitesche Matrizen:

Satz 12.96:

a) Für jeden Eigenwert einer hermiteschen Matrix gilt

$$1 \leq \text{geometrische Vielfachheit} = \text{algebraische Vielfachheit}.$$

b) Hermitesche Matrizen besitzen eine Basis von orthogonalen Eigenvektoren.

Beweis: Wir beweisen a) und b) simultan für hermitesches $A \in K^{n \times n}$. Seien dazu $\lambda_1, \dots, \lambda_k$ die paarweise unterschiedlichen Eigenwerte mit den geometrischen Vielfachheiten g_1, \dots, g_k . Seien $\vec{v}_j^{(i)}$ mit $j = 1, \dots, g_i$ eine orthogonale Basis des Eigenraums E_{λ_i} (nehme z.B. irgendeine Basis des Kerns von $A - \lambda_i \cdot \mathbb{I}$ und orthogonalisiere sie per Gram-Schmidt (Aufgabe 110, Blatt 13)). Da nach Satz 12.95 die Eigenvektoren zu unterschiedlichen Eigenwerten orthogonal sind, haben wir insgesamt eine Sammlung von $g = \sum_{i=1}^k g_i$ orthogonalen Eigenvektoren

$$\underbrace{\vec{v}_1^{(1)}, \dots, \vec{v}_{g_1}^{(1)}}_{\text{Eigenwert } \lambda_1}, \underbrace{\vec{v}_1^{(2)}, \dots, \vec{v}_{g_2}^{(2)}}_{\text{Eigenwert } \lambda_2}, \dots, \underbrace{\vec{v}_1^{(k)}, \dots, \vec{v}_{g_k}^{(k)}}_{\text{Eigenwert } \lambda_k},$$

wobei

$$g = \sum_{i=1}^k g_i \leq \sum_{i=1}^k n_i = n$$

gilt. Wir wählen weitere $n - g$ Vektoren $\vec{w}_1, \dots, \vec{w}_{n-g}$ hinzu, die zusammen mit den Eigenvektoren den K^n aufspannen. Nach Orthogonalisierung per Gram-Schmidt (Aufgabe 110, Blatt 13), können wir sie als paarweise orthogonal und orthogonal zu den Eigenvektoren $\vec{v}_j^{(i)}$ annehmen.

Aus diesen insgesamt n Vektoren wird spaltenweise die Matrix

$$T = [\vec{v}_1^{(1)}, \dots, \vec{v}_{g_1}^{(1)}, \vec{v}_1^{(2)}, \dots, \vec{v}_{g_2}^{(2)}, \dots, \vec{v}_1^{(k)}, \dots, \vec{v}_{g_k}^{(k)}, \vec{w}_1, \dots, \vec{w}_{n-g}]$$

aufgebaut, die wegen der Orthogonalität der Spalten $T^H T = \mathbb{I}$ erfüllt (also eine orthogonale Matrix mit $T^{-1} = T^H$ ist). Analog zum Beweis von Satz 12.90 betrachten wir die Matrix

$$\tilde{A} = T^{-1} A T,$$

die mit den selben Argumenten wie im Beweis von Satz 12.90 von der Form

$$\tilde{A} = \begin{pmatrix} & g_1 & & g & g+1 & & n \\ & \downarrow & & \downarrow & \downarrow & & \downarrow \\ \lambda_1 & 0 & \dots & \dots & \dots & \dots & 0 & * & \dots & * \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \lambda_1 & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \lambda_k & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 & \lambda_k & * & \dots & * \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & * & \dots & * \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & * & \dots & * \end{pmatrix}.$$

sein muss. Wegen $\tilde{A}^H = T^H A^H (T^{-1})^H = T^{-1} A T = \tilde{A}$ ist \tilde{A} wieder hermitesch, damit muss \tilde{A} speziell von folgender Form sein:

$$\tilde{A} = \begin{pmatrix} & g_1 & & g & g+1 & & n \\ & \downarrow & & \downarrow & \downarrow & & \downarrow \\ \lambda_1 & 0 & \dots & \dots & \dots & \dots & 0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \lambda_1 & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \lambda_k & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 & \lambda_k & 0 & \dots & 0 \\ \hline 0 & \dots & \dots & \dots & \dots & \dots & 0 & * & \dots & * \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & * & \dots & * \end{pmatrix}.$$

Das charakteristische Polynom von \tilde{A} (das mit dem von A übereinstimmt, siehe den Beweis von Satz 12.90) ist

$$p_{\tilde{A}}(\lambda) = (\lambda_1 - \lambda)^{g_1} \cdot \dots \cdot (\lambda_k - \lambda)^{g_k} \cdot \det(S - \lambda \cdot \mathbb{I}),$$

wo S der (symmetrische) Matrixblock der Dimension $(n - g) \times (n - g)$ in der rechten unteren Ecke von \tilde{A} ist. Vergleich mit

$$p_A(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdot \dots \cdot (\lambda_k - \lambda)^{n_k}$$

liefert

$$\det(S - \lambda \cdot \mathbb{I}) = (\lambda_1 - \lambda)^{n_1 - g_1} \cdot \dots \cdot (\lambda_k - \lambda)^{n_k - g_k},$$

d.h., dieser Block hat einen Eigenwert λ_i , falls für mindestens einen Index $g_i < n_i$ gilt. Dazu gehört ein Eigenvektor $(x_1, \dots, x_{n-g})^T \in K^{n-g}$ des Blocks S , der zu einem Eigenvektor

$$\vec{x} = (0, \dots, 0, x_1, \dots, x_{n-g})^T \in K^n$$

von \tilde{A} zum Eigenwert λ_i erweitert werden kann. Hieraus wird ein Eigenvektor $T\vec{x}$ von A , denn

$$A(T\vec{x}) = T\tilde{A}T^{-1}(T\vec{x}) = T\tilde{A}\vec{x} = T\lambda_i \cdot \vec{x} = \lambda_i \cdot (T\vec{x}).$$

Man überzeugt sich leicht davon, dass der Eigenvektor \vec{x} linear unabhängig von den Kernvektoren $\vec{v}_1^{(i)}, \dots, \vec{v}_{g_i}^{(i)}$ ist, denn nach Konstruktion gilt $T^{-1}\vec{v}_{g_i}^{(i)} = (0, \dots, 0, 1, 0, \dots, 0)^T$, wo die 1 unter den ersten g Komponenten des Vektors zu finden sind. (Im Gegensatz dazu hat \vec{x} mindestens einen von 0 verschiedenen Eintrag in den letzten $n-g$ Komponenten.) Damit haben wir einen Widerspruch zu $g_i < n_i$, da wir nun einen weiteren linear unabhängigen Eigenvektor zu λ_i gefunden haben.

Q.E.D.

12.7.3 Diagonalisierung und Funktionalkalkül

Der wesentliche Grund, warum man sich für die Eigenwerte und -vektoren von Matrizen interessiert, ist die Tatsache, dass man mit Hilfe dieser „spektralen Daten“ eine Matrix durch eine „Ähnlichkeitstransformation“ $T^{-1}AT$ auf Diagonalgestalt bringen kann, was das Hantieren mit der Matrix einfacher macht. Dies läuft darauf hinaus, statt der Standardbasis des K^n eine Basis aus Eigenvektoren der Matrix A zu benutzen, um die Vektoren des K^n darzustellen.

Satz 12.97: (Diagonalisierung von Matrizen)

Sei A eine reelle oder komplexe $n \times n$ -Matrix mit den (eventuell komplexen) Eigenwerten $\lambda_1, \dots, \lambda_n$ und entsprechenden Eigenvektoren $\vec{x}_1, \dots, \vec{x}_n$, also $A \vec{x}_k = \lambda_k \vec{x}_k$. Sei $T = [\vec{x}_1, \dots, \vec{x}_n]$ die Matrix, deren Spalten aus diesen Eigenvektoren besteht. Es gilt

$$AT = TD \quad \text{mit} \quad D = \text{diag}(\lambda_1, \dots, \lambda_n) := \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Sind die Eigenvektoren linear unabhängig, so ist T invertierbar, und es folgt

$$A = TDT^{-1}.$$

Beweis: Nach Definition der Matrixmultiplikation gilt

$$AT = A[\vec{x}_1, \dots, \vec{x}_n] = [A\vec{x}_1, \dots, A\vec{x}_n].$$

Mit

$$TD = [\vec{x}_1, \dots, \vec{x}_n] \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = [\lambda_1 \cdot \vec{x}_1, \dots, \lambda_n \cdot \vec{x}_n]$$

und $A\vec{x}_k = \lambda_k \cdot \vec{x}_k$ folgt $AT = TD$.

Q.E.D.

Folgerung 12.98:

Mit einer Diagonalisierung $A = TDT^{-1}$ gilt

$$A^n = \underbrace{TDT^{-1}}_A \underbrace{TDT^{-1}}_A \dots \underbrace{TDT^{-1}}_A = TDD \dots DT^{-1} = TD^nT^{-1}.$$

Die Potenzen von A sind damit auf Potenzen der Diagonalform D von A zurückgeführt, wobei D^n sich ohne große Rechnung durch Potenzieren der Diagonalelemente ergibt:

$$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{pmatrix}^n = \begin{pmatrix} \lambda_1^n & & \\ & \lambda_2^n & \\ & & \ddots \end{pmatrix}.$$

Bemerkung 12.99: Hat man einen Eigenwert λ_k der zu diagonalisierenden Matrix gefunden, so kann man nach 12.97 **irgendeinen** dazugehörigen Eigenvektor \vec{x}_k benutzen, die entsprechende Spalte von T zu besetzen. Nun sind Eigenvektoren aber nicht eindeutig: ist \vec{x}_k ein Eigenvektor, so ist jedes Vielfache $\vec{y}_k = c_k \vec{x}_k$ wieder ein Eigenvektor zum selben Eigenwert λ_k . Es gibt damit viele unterschiedliche Transformationsmatrizen T , während die Diagonalmatrix bis auf Umnummerierung der Eigenwerte eindeutig ist. Wie kann das sein? Antwort: die Transformationsmatrizen unterscheiden sich nur um eine Diagonalmatrix:

$$\tilde{T} = [\vec{y}_1, \dots, \vec{y}_n] = [c_1 \vec{x}_1, \dots, c_n \vec{x}_n] = \underbrace{[\vec{x}_1, \dots, \vec{x}_n]}_T \underbrace{\begin{pmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_n \end{pmatrix}}_C = T C.$$

In der Diagonalisierung fällt die diagonale „Skalierungsmatrix“ C heraus, da die Multiplikation von Diagonalmatrizen kommutativ ist:

$$\begin{aligned} A &= \tilde{T} D \tilde{T}^{-1} = T C D (T C)^{-1} = T C D C^{-1} T^{-1} \\ &= T D C C^{-1} T^{-1} = T D T^{-1}. \end{aligned}$$

Bemerkung 12.100: Da bei der Diagonalisierung die Invertierbarkeit von T wichtig ist, nützt es überhaupt nichts, triviale Eigenvektoren $\vec{x}_k = 0$ zu betrachten (die ja auch strenggenommen per Definition von Eigenvektoren gar nicht als Eigenvektoren zulässig sind).

Die Aufgabenstellung der Diagonalisierung läuft darauf hinaus, die Eigenvektoren zu *allen* Eigenwerten zu finden. Sobald man eine Basis von linear unabhängigen Eigenvektoren gefunden hat, hat man die invertierbare Transformationsmatrix T gefunden, welche die betrachtete Matrix auf Diagonalform bringt. Sind alle Eigenwerte verschieden, so existiert immer eine Basis von linear unabhängigen Eigenvektoren. Bei entarteten Eigenwerten (mehrfachen Nullstellen des charakteristischen Polynoms) ist dies jedoch nicht garantiert, und in der Tat gibt es nicht-diagonalisierbare Matrizen (siehe Beispiel 12.91). Symmetrische reelle bzw. hermitesche komplexe Matrizen sind immer diagonalisierbar, selbst wenn die (bei Symmetrie/Hermitizität der Matrix automatisch reellen) Eigenwerte entartet sind.

Beispiel 12.101: Mittels Diagonalisierung können wir für

$$A = \begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 1 \end{pmatrix}$$

eine explizite Darstellung von A^n für jedes $n \in \mathbb{Z}$ ermitteln. Nach Satz 12.97 sind alle Eigenwerte und -vektoren zu bestimmen. Die Eigenwerte sind die Nullstellen des charakteristischen Polynoms

$$\det(\lambda \mathbb{I} - A) = \det \begin{pmatrix} \lambda - 1 & \frac{1}{2} \\ -2 & \lambda - 1 \end{pmatrix} = \lambda^2 - 2\lambda + 2$$

(hier ist \mathbb{I} die 2×2 -Einheitsmatrix). Nach der Standardformel für die Nullstellen quadratischer Polynome ergeben sich die komplex konjugierten Eigenwerte

$$\lambda_{1,2} = 1 \pm \sqrt{-1} = 1 \pm i.$$

Die Eigenvektoren zu λ_1 bzw. λ_2 findet man, indem man die Kernvektoren von $A - \lambda_k \mathbb{I}$ ermittelt, d.h., die Gleichungen $(A - \lambda_k \mathbb{I}) \vec{x}_k = A \vec{x}_k - \lambda_k \vec{x}_k = \vec{0}$ löst. Mit dem Gauß-Algorithmus 12.64 gefolgt vom Entzerrungsalgorithmus aus Abschnitt 12.5.3 findet man die Lösungen

$$\vec{x}_1 = \begin{pmatrix} i \\ 2 \end{pmatrix} \quad \text{bzw.} \quad \vec{x}_2 = \begin{pmatrix} 1 \\ 2 \cdot i \end{pmatrix}.$$

Die Transformationsmatrix T wird spaltenweise aus diesen Eigenvektoren aufgebaut, die Inverse wird berechnet:

$$T = \begin{pmatrix} i & 1 \\ 2 & 2 \cdot i \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} -\frac{i}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{i}{4} \end{pmatrix}.$$

Damit folgt die Diagonalisierung

$$A = \begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} i & 1 \\ 2 & 2 \cdot i \end{pmatrix}}_T \underbrace{\begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}}_D \underbrace{\begin{pmatrix} -\frac{i}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{i}{4} \end{pmatrix}}_{T^{-1}}.$$

Dies liefert

$$\begin{aligned} A^n &= \underbrace{\begin{pmatrix} i & 1 \\ 2 & 2 \cdot i \end{pmatrix}}_T \underbrace{\begin{pmatrix} (1+i)^n & 0 \\ 0 & (1-i)^n \end{pmatrix}}_{D^n} \underbrace{\begin{pmatrix} -\frac{i}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{i}{4} \end{pmatrix}}_{T^{-1}} \\ &= \begin{pmatrix} i & 1 \\ 2 & 2 \cdot i \end{pmatrix} \begin{pmatrix} -\frac{i}{2} \cdot (1+i)^n & \frac{1}{4} \cdot (1+i)^n \\ \frac{1}{2} \cdot (1-i)^n & -\frac{i}{4} \cdot (1-i)^n \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} \cdot (1+i)^n + \frac{1}{2} \cdot (1-i)^n & \frac{i}{4} \cdot (1+i)^n - \frac{i}{4} \cdot (1-i)^n \\ -i \cdot (1+i)^n + i \cdot (1-i)^n & \frac{1}{2} \cdot (1+i)^n + \frac{1}{2} \cdot (1-i)^n \end{pmatrix}. \end{aligned}$$

Das Ergebnis muß als Potenz der reellen Matrix A reell sein. Dies ist in dieser Darstellung leider alles andere als offensichtlich. Über die Polardarstellungen

$$1 \pm i = \sqrt{2} \cdot e^{\pm i \cdot \frac{\pi}{4}}$$

ergibt sich mit $|1 \pm i| = \sqrt{2}$ und den Polarwinkeln $\pm \frac{\pi}{4}$:

$$(1 \pm i)^n = 2^{n/2} \cdot e^{\pm n \cdot i \cdot \frac{\pi}{4}} = 2^{n/2} \cdot \left(\cos\left(n \cdot \frac{\pi}{4}\right) \pm i \cdot \sin\left(n \cdot \frac{\pi}{4}\right) \right).$$

Damit folgt

$$\begin{aligned} (A^n)_{11} &= \frac{1}{2} \cdot (1+i)^n + \frac{1}{2} \cdot (1-i)^n = 2^{n/2} \cdot \cos\left(n \cdot \frac{\pi}{4}\right), \\ (A^n)_{12} &= \frac{i}{4} \cdot (1+i)^n - \frac{i}{4} \cdot (1-i)^n = -2^{n/2-1} \cdot \sin\left(n \cdot \frac{\pi}{4}\right), \\ (A^n)_{21} &= -i \cdot (1+i)^n + i \cdot (1-i)^n = 2^{n/2+1} \cdot \sin\left(n \cdot \frac{\pi}{4}\right), \\ (A^n)_{22} &= \frac{1}{2} \cdot (1+i)^n + \frac{1}{2} \cdot (1-i)^n = 2^{n/2} \cdot \cos\left(n \cdot \frac{\pi}{4}\right). \end{aligned}$$

Insgesamt erhalten wir also in der Tat eine explizite (und nun recht einfache) reelle Darstellung beliebiger Potenzen von A :

$$\begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 1 \end{pmatrix}^n = \begin{pmatrix} 2^{\frac{n}{2}} \cdot \cos\left(n \cdot \frac{\pi}{4}\right) & -2^{\frac{n}{2}-1} \cdot \sin\left(n \cdot \frac{\pi}{4}\right) \\ 2^{\frac{n}{2}+1} \cdot \sin\left(n \cdot \frac{\pi}{4}\right) & 2^{\frac{n}{2}} \cdot \cos\left(n \cdot \frac{\pi}{4}\right) \end{pmatrix}.$$

Satz 12.102: (Hamilton-Cayley-Theorem)

Das charakteristische Polynom von $A \in K^{n \times n}$ sei

$$p_A(\lambda) = (-1)^n \cdot \lambda^n + c_{n-1} \cdot \lambda^{n-1} + \dots + c_1 \cdot \lambda + c_0.$$

Dann gilt die Matrix-Gleichung

$$p_A(A) = (-1)^n \cdot A^n + c_{n-1} \cdot A^{n-1} + \dots + c_1 \cdot A + c_0 \cdot \mathbf{I} = 0 \in K^{n \times n}.$$

Beweis: Für diagonalisierbares $A = TDT^{-1}$ mit $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ ist der Beweis einfach. Mit $A^k = TD^kT^{-1}$ und $D^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$ folgt

$$\begin{aligned} p_A(A) &= p_A(TDT^{-1}) = Tp_A(D)T^{-1} = T \begin{pmatrix} p_A(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & p_A(\lambda_n) \end{pmatrix} T^{-1} \\ &= T \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ 0 & & 0 \end{pmatrix} T^{-1} = \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ 0 & & 0 \end{pmatrix}. \end{aligned}$$

Die Aussage gilt auch für nicht-diagonalisierbare Matrizen A , wobei der Beweis aber deutlich aufwendiger wird (das lassen wir hier).

Q.E.D.

Für Diagonalmatrizen $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ gilt $D^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$. Man nimmt dies zum Anlass, für beliebige Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ bzw. $f : \mathbb{C} \rightarrow \mathbb{C}$ zu definieren: $f(D) = \text{diag}(f(\lambda_1), \dots, f(\lambda_n))$.

Definition 12.103: (Funktionalkalkül)

Für eine Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ bzw. $f : \mathbb{C} \rightarrow \mathbb{C}$ definieren wir die Matrix-Auswertung dieser Funktion für eine diagonalisierbare Matrix $A = TDT^{-1} \in K^{n \times n}$ mit $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ mittels

$$f(A) = Tf(D)T^{-1} = T \begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix} T^{-1}.$$

Beispiel 12.104: Wir betrachten $f(x) = e^x$. Stellt man sich die exp-Funktion als die Potenzreihe

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

vor, so liegt es nahe, die Matrixauswertung für $A \in K^{n \times n}$ durch

$$e^A = \mathbf{1} + A + \frac{1}{2!} \cdot A^2 + \frac{1}{3!} \cdot A^3 + \dots \quad (\#)$$

zu definieren. Für diagonalisierbares $A = TDT^{-1}$ mit $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ stimmt dies mit der obigen Definition 12.103 überein:

$$\begin{aligned} e^A &= \mathbf{1} + TDT^{-1} + \frac{1}{2!} \cdot (TDT^{-1})^2 + \frac{1}{3!} \cdot (TDT^{-1})^3 + \dots \\ &= T \left(\mathbf{1} + D + \frac{1}{2!} \cdot D^2 + \frac{1}{3!} \cdot D^3 + \dots \right) T^{-1} \\ &= T \begin{pmatrix} 1 + \lambda_1 + \frac{\lambda_1^2}{2!} + \dots & & 0 \\ & \ddots & \\ 0 & & 1 + \lambda_n + \frac{\lambda_n^2}{2!} + \dots \end{pmatrix} T^{-1} \\ &= T \begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix} T^{-1}. \end{aligned}$$

Die Definition (#) über die Potenzreihe ist dabei auch für nicht-diagonalisierbares A sinnvoll.

Hier eine Anwendung:

Satz 12.105: (dynamische Systeme)

Sei $A \in K^{n \times n}$ eine konstante Matrix. Die Lösung des Differentialgleichungssystems

$$\frac{d}{dt} \vec{x}(t) = A \vec{x}(t)$$

ist

$$\vec{x}(t) = e^{t \cdot A} \vec{x}(0).$$

Beweis: Es gilt

$$\begin{aligned} \frac{d}{dt} e^{tA} &= \frac{d}{dt} \left(\mathbb{1} + t \cdot A + \frac{t^2}{2!} \cdot A^2 + \frac{t^3}{3!} \cdot A^3 + \dots \right) = A + t \cdot A^2 + \frac{t^2}{2!} \cdot A^3 + \dots \\ &= A \left(\mathbb{1} + t \cdot A + \frac{t^2}{2!} \cdot A^2 + \frac{t^3}{3!} \cdot A^3 + \dots \right) = A e^{tA}. \end{aligned}$$

Damit ist $\vec{x}(t) = e^{t \cdot A} \vec{x}(0)$ in der Tat eine Lösung der DGL:

$$\frac{d}{dt} \vec{x}(t) = \frac{d}{dt} e^{t \cdot A} \vec{x}(0) = A e^{tA} \vec{x}(0) = A \vec{x}(t).$$

Q.E.D.

Beispiel 12.106: Betrachte das DGL-System

$$\frac{dx_1}{dt} = x_1 - \frac{x_2}{2}, \quad \frac{dx_2}{dt} = 2 \cdot x_1 + x_2,$$

das als

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

geschrieben wird. In Beispiel 12.101 war die Diagonalisierung

$$\begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} i & 1 \\ 2 & 2 \cdot i \end{pmatrix}}_T \underbrace{\begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}}_D \underbrace{\begin{pmatrix} -\frac{i}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{i}{4} \end{pmatrix}}_{T^{-1}}$$

gefunden worden. Die Lösung der DGL ist damit

$$\begin{aligned} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} &= \underbrace{\begin{pmatrix} i & 1 \\ 2 & 2 \cdot i \end{pmatrix}}_T \underbrace{\begin{pmatrix} e^{t \cdot (1+i)} & 0 \\ 0 & e^{t \cdot (1-i)} \end{pmatrix}}_{e^{t \cdot D}} \underbrace{\begin{pmatrix} -\frac{i}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{i}{4} \end{pmatrix}}_{T^{-1}} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} \\ &= \begin{pmatrix} \cos(t) \cdot e^t & -\frac{1}{2} \cdot \sin(t) \cdot e^t \\ 2 \cdot \sin(t) \cdot e^t & \cos(t) \cdot e^t \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix}, \end{aligned}$$

also

$$x_1(t) = \cos(t) \cdot e^t \cdot x_1(0) - \frac{x_2(0)}{2} \cdot \sin(t) \cdot e^t,$$
$$x_2(t) = 2 \cdot \sin(t) \cdot e^t \cdot x_1(0) + x_2(0) \cdot \cos(t) \cdot e^t.$$
