Skript zur Vorlesung

Mathematik I für Inf, WInf

Wintersemester 15/16

Robert Haller-Dintelmann

20. Juli 2015

Inhaltsverzeichnis

Teil I.

Mathematik I

1. Grundbegriffe

1.1. Aussagen

1.1.1. Aussagen

Eine *Aussage* ist ein in verständlicher Sprache formulierter Satz, der entweder wahr (w) oder falsch (f) ist.

Beispiel 1.1.1. Hier sind fünf Aussagen:

A_1 : 3 ist eine ungerade Zahl. (w)

A_2 : Die Erde ist eine Scheibe. (f)

A_3 : Es regnet gerade in Madrid. (?)

A_4 : Jede natürliche Zahl ist gerade. (f)

A_5 : 3 ist eine Primzahl. (w)

Keine Aussage ist: „Guten Morgen.“

1.1.2. Aussageformen

Eine *Aussageform* ist ein Satz mit einer oder mehreren Variablen, der bei Belegung der Variablen durch einen konkreten Wert eine Aussage wird.

Beispiel 1.1.2. Hier sind vier Aussageformen:

$E_1(x)$: $x + 10 = 5$.

$E_2(x)$: $x^2 \geq 0$.

$E_3(n)$: n ist gerade.

$E_4(x, y)$: $3x - 4y \neq 10$.

1.1.3. All- und Existenzquantor

Sei $E(x)$ eine Aussageform und M eine Menge von möglichen x. Dann bedeutet

$$\forall x \in M : E(x) \quad \text{„Für alle } x \text{ aus } M \text{ ist } E(x) \text{ wahr“.}$$

Man nennt $\forall$ den *Allquantor*.
Weiter bedeutet

$$\exists x \in M : E(x) \quad \text{„Es existiert ein } x \text{ aus } M, \text{ für das } E(x) \text{ wahr ist“.}$$

Man nennt $\exists$ den *Existenzquantor*.

Man beachte, dass durch das Vorstellen eines Quantors auf diese Weise aus einer Aussageform eine Aussage wird. Hat die Aussageform mehrere Variablen braucht es natürlich auch mehrere Quantoren.

Beispiel 1.1.3. Aus obigen Aussageformen können wir z.B. die folgenden Aussagen machen:

(a) $\forall x \in \mathbb{R} : E_2(x)$, d.h. $\forall x \in \mathbb{R} : x^2 \geq 0$. (w)

(b) $\forall n \in \mathbb{N} : E_3(n)$ entspricht genau A_4. (f)

(c) $\exists n \in \mathbb{N} : E_3(n)$, d.h. es gibt eine gerade natürliche Zahl. (w)

Warnung 1.1.4. „Es existiert ein x“ bedeutet nicht „Es existiert genau ein x“. Ist die Aussage $\exists x \in M : E(x)$ wahr, so kann es durchaus mehrere x geben, für die $E(x)$ wahr wird!

1.1.4. Verknüpfung von Aussagen

Seien A und B zwei Aussagen. Dann können wir daraus verschiedene neue Aussagen machen.

Konjunktion („und“): Zeichen: $\wedge$

$$A \wedge B : \quad \text{Sowohl } A \text{ als auch } B \text{ sind wahr.}$$

Disjunktion („oder“): Zeichen: $\vee$

$$A \vee B : \quad A \text{ ist wahr oder } B \text{ ist wahr.}$$

Negation („nicht“): Zeichen: $\neg$

$$\neg A : \quad A \text{ gilt nicht.}$$

Implikation („wenn . . . , dann“): Zeichen: $\Longrightarrow$

$$A \Longrightarrow B: \quad \begin{array}{l} \text{Wenn } A \text{ gilt dann auch } B. \\ \text{Aus } A \text{ folgt } B. \\ A \text{ impliziert } B. \end{array}$$

Äquivalenz („genau dann, wenn“): Zeichen: $\Longleftrightarrow$

$$A \Longleftrightarrow B: \quad \begin{array}{l} A \text{ gilt genau dann, wenn } B \text{ gilt.} \\ A \text{ und } B \text{ sind äquivalent.} \end{array}$$

Warnung 1.1.5. (a) $\vee$ ist *nicht* „entweder . . . oder “, d.h. $A \vee B$ ist auch wahr, wenn sowohl A als auch B wahr sind.

(b) Gewöhnungsbedürftig ist zunächst folgendes: Wenn A falsch ist, dann ist $A \Longrightarrow B$ in jedem Fall wahr. Anders ausgedrückt: Aus einer falschen Aussage kann man alles folgern. Man sieht das auch an der Wahrheitstafel der Implikation

A	B	$A \Longrightarrow B$
w	w	w
w	f	f
f	w	w
f	f	w

Bemerkung 1.1.6. Als kleine Übung machen wir uns noch klar, dass die Aussage $C := (A \Longrightarrow B) \Longleftrightarrow (\neg B \Longrightarrow \neg A)$ immer wahr ist:

A	B	$A \Longrightarrow B$	$\neg A$	$\neg B$	$\neg B \Longrightarrow \neg A$	C
w	w	w	f	f	w	w
w	f	f	f	w	f	w
f	w	w	w	f	w	w
f	f	w	w	w	w	w

Das bedeutet, dass der Wahrheitsgehalt der Aussagen $A \Longrightarrow B$ und $\neg B \Longrightarrow \neg A$ immer der selbe ist. Zum Nachweis von „$A \Longrightarrow B$ ist wahr“ kann man also gleichbedeutend auch „$\neg B \Longrightarrow \neg A$ ist wahr“ beweisen. Das ist dann ein sogenannter Beweis durch Kontraposition und ist manchmal einfacher als ein direkter Beweis, vgl. Abschnitt 1.5.

1.2. Mengen

Beispiele von Mengen sind: Die Menge aller Studierenden in einem Hörsaal, ein Dreieck (als Punktmenge der Ebene), die Menge aller Dreiecke in der Ebene oder

die Mengen $\mathbb{N}$,[1] $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, also die Mengen der natürlichen, ganzen, rationalen, bzw. reellen Zahlen. Den Begriff der Menge definieren wir hier nicht, sondern legen ihn naiv zu Grunde; wir stellen uns damit auf den Standpunkt der naiven (und nicht der axiomatischen) Mengenlehre.
Wenn wir Mengen bilden, ist unser Ausgangspunkt immer eine gegebene, unter Umständen sehr großen Grundmenge G, aus der Elemente ausgesondert und zu neuen Mengen zusammengefasst werden. Auf diese Weise vermeidet man Bildungen wie die „Menge aller Mengen“, die zu Widersprüchen führen.
Mengen kann man, solange sie klein genug sind, einfach durch das Aufzählen ihrer Elemente angeben, z.B.

$$M_1 = \{0, 1, 2, 3, 4, 5\}.$$

Es ist aber häufig angenehmer, sie durch die Angabe einer definierenden Eigenschaft, die genau für die Elemente der Menge, und nur für diese, wahr ist, zu beschreiben. Für unsere Menge M_1 könnte das so aussehen:

$$M_1 = \{x \in \mathbb{N} : x < 6\} \quad \text{oder} \quad M_1 = \{x \in \mathbb{N} : x - 6 \text{ ist keine natürliche Zahl}\}.$$

Allgemein schreibt man

$$M = \{x \in G : E(x)\},$$

wobei G die Grundmenge ist, aus der die Elemente der Menge M ausgesondert werden sollen und $E(x)$ eine Aussageform.

Definition 1.2.1. *Seien M und N Mengen. Wir schreiben $a \in M$, falls a ein Element von M ist und, falls dem nicht so ist, $a \notin M$.*
Ist jedes Element von N auch in M enthalten, so schreiben wir $N \subseteq M$ und sagen N ist eine Teilmenge *von M. Weiter nennt man in diesem Fall M eine* Obermenge *von N und schreibt $M \supseteq N$. Solche Teilmengenbeziehungen werden oft auch als* Inklusion *bezeichnet.*
Schlussendlich schreiben wir $\emptyset$ für die leere Menge*, d.h. die Menge, die kein Element enthält.*

Bemerkung 1.2.2. Für zwei Mengen M und N gilt $M = N$ genau dann, wenn $M \subseteq N$ und $N \subseteq M$ gilt.

Definition 1.2.3. *Seien M und N Mengen in einer Grundmenge G. Dann ist*

(a) $M \cup N := \{x \in G : x \in M \vee x \in N\}$	Vereinigung *von M und N*,
(b) $M \cap N := \{x \in G : x \in M \wedge x \in N\}$	Schnitt *von M und N*,
(c) $M^c := \{x \in G : x \notin M\}$	Komplement *von M in G*,
(d) $M \setminus N := \{x \in M : x \notin N\}$	Mengendifferenz *von M und N*,
(e) $M \times N := \{(x, y) : x \in M, y \in N\}$	kartesisches Produkt *von M und N*.

[1] In dieser Vorlesung ist $\mathbb{N} := \{0, 1, 2, 3, \dots\}$. Für die natürlichen Zahlen ohne Null schreiben wir $\mathbb{N}^* := \{1, 2, 3, 4, \dots\}$.

Bemerkung 1.2.4. Damit ist ebenfalls für eine endliche Anzahl von Mengen $A_1, A_2, \ldots, A_n$ das n-fache kartesische Produkt

$$A_1 \times A_2 \times \cdots \times A_n = \big\{(a_1, a_2, \ldots, a_n) : a_j \in A_j \text{ für } j = 1, 2, \ldots, n\big\}.$$

definiert.

Satz 1.2.5. *Seien A, B und C Mengen. Dann gilt*

(a) $A \cup B = B \cup A$ *und* $A \cap B = B \cap A$. (Kommutativgesetze)

(b) $(A \cup B) \cup C = A \cup (B \cup C)$ *und* $(A \cap B) \cap C = A \cap (B \cap C)$. (Assoziativgesetze)

(c) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ *und* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivgesetze),

(d) $(A \cup B)^c = A^c \cap B^c$ *und* $(A \cap B)^c = A^c \cup B^c$ (Regeln von De Morgan).

Beweis. Wir behandeln hier das erste Distributivgesetz und die erste Regel von De Morgan, die weiteren verbleiben als Übungsaufgabe.
Für das Distributivgesetz zeigen wir zuerst (vgl. Bemerkung 1.2.2)

$$A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C),$$

und zwar folgendermaßen: Sei $x \in A \cup (B \cap C)$. Dann ist also $x \in A$ oder $x \in B \cap C$. Betrachten wir zunächst den Fall $x \in A$. Dann gilt natürlich auch $x \in A \cup B$ und $x \in A \cup C$, denn diese Mengen sind ja größer als A. Also ist $x \in (A \cup B) \cap (A \cup C)$ und wir sind fertig. Betrachten wir also den Fall $x \in B \cap C$. Dann ist $x \in B$ und $x \in C$, also gilt wieder $x \in A \cup B$ und $x \in A \cup C$, dieses Mal, weil x sowohl in B als auch in C liegt. Daraus folgt wieder $x \in (A \cup B) \cap (A \cup C)$ und wir haben $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$ gezeigt.
Um die im ersten Distributivgesetz behauptete Gleichheit zu zeigen, müssen wir nun noch die umgekehrte Inklusion

$$(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$$

zeigen. Dazu sei $x \in (A \cup B) \cap (A \cup C)$. Dann ist x sowohl in $A \cup B$, als auch in $A \cup C$. Wir betrachten die beiden Fälle $x \in A$ und $x \notin A$. (Man beachte, dass wir dann alle denkbaren Fälle $x \in G$ berücksichtigt haben!) Ist $x \in A$, so haben wir sofort auch $x \in A \cup (B \cap C)$, was unser Ziel war. Es bleibt also der Fall $x \notin A$. Da dann x in $A \cup B$ ist, ohne in A zu sein, muss x zwangsläufig in B sein, denn wie sollte es sonst da hineinkommen? Genauso folgt $x \in C$ aus $x \in A \cup C$. Also ist x in $B \cap C$ und damit auch $x \in A \cup (B \cap C)$ und wir haben auch die zweite Inklusion und damit die Gleichheit

$$(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$$

gezeigt.

Für die erste De Morgan'sche Regel zeigen wir wieder zuerst

$$(A \cup B)^c \subseteq A^c \cap B^c.$$

Sei dazu $x \in (A \cup B)^c$. Dann ist $x \notin (A \cup B)$, d.h. x ist nicht in der Vereinigung von A und B. Damit kann x weder in A noch in B sein, denn sonst würde es ja in dieser Vereinigung liegen. Es ist also $x \notin A$ und $x \notin B$, d.h. $x \in A^c$ und $x \in B^c$, was schließlich $x \in A^c \cap B^c$ nach sich zieht.
Die zweite Inklusion

$$(A \cup B)^c \supseteq A^c \cap B^c$$

geht folgendermaßen: Es sei $x \in A^c \cap B^c$. Dann ist $x \in A^c$ und $x \in B^c$. Also ist x nicht in A und nicht in B, es ist also auch nicht in der Vereinigung von A und B, was gerade $x \in (A \cup B)^c$ bedeutet. □

Definition 1.2.6. *Eine Menge M heißt* endlich*, falls sie endlich viele Elemente besitzt. In diesem Fall schreiben wir $|M|$ für die Anzahl der Elemente von M.*

Bemerkung 1.2.7. Seien A und B endliche Mengen, dann gilt $|A \times B| = |A| \cdot |B|$. Warum? Es gilt $A \times B = \{(a, b) : a \in A, b \in B\}$. Für die Wahl der $a \in A$ in der ersten Komponente hat man $|A|$ Möglichkeiten. Ist dann $a \in A$ gewählt, so gibt es für jede dieser Wahlen wieder $|B|$ Möglichkeiten ein $b \in B$ zuzulosen. Zusammen ergibt das $|A| \cdot |B|$ Möglichkeiten, d.h. es gilt $|A \times B| = |A| \cdot |B|$.

Übungsaufgabe 1.2.8. Es seien A und B endliche Mengen. Zeigen Sie:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Definition 1.2.9. *Ist M eine Menge, so heißt*

$$\mathcal{P}(M) := \{N : N \text{ Teilmenge von } M\}$$

Potenzmenge *von M.*

Beispiel 1.2.10. Es ist $\mathcal{P}(\{0, 1\}) = \big\{\emptyset, \{0\}, \{1\}, \{0, 1\}\big\}$.

1.3. Relationen

Definition 1.3.1. *Sei X eine Menge. Eine Teilmenge $R \subseteq X \times X$ heißt* (zweistellige) Relation *auf X. Man schreibt xRy, falls das Tupel $(x, y) \in R$ liegt und sagt „x steht in Relation zu y".*

Beispiel 1.3.2. (a) $\leq$ in $\mathbb{N}$: Dann ist $R = \{(n, m) \in \mathbb{N} \times \mathbb{N} : n \leq m\}$ und x steht genau dann mit y in Relation, wenn $x \leq y$ gilt.

(b) Nehmen Sie als X die Menge aller Internetseiten, so können Sie durch die Setzung $R := \{(x, y) : x \text{ verlinkt nach } y\}$ eine Relation auf X definieren, die die Verlinkungsstruktur codiert.

Definition 1.3.3. *Sei X eine Menge. Eine Relation R auf X heißt*

(a) reflexiv*, falls xRx für jedes $x \in X$ gilt.*

(b) symmetrisch*, falls für alle $x, y \in X$ mit xRy auch yRx gilt.*

(c) antisymmetrisch*, falls für alle $x, y \in X$, für die xRy und yRx gilt, $x = y$ folgt.*

(d) transitiv*, falls für alle $x, y, z \in X$ mit xRy und yRz auch xRz gilt.*

(e) Äquivalenzrelation*, falls R reflexiv, symmetrisch und transitiv ist. In diesem Fall schreibt man meist „$\sim$“ statt „R“.*

(f) Ordnungsrelation*, falls R reflexiv, antisymmetrisch und transitiv ist. Man schreibt dann meist „$\leq$“ statt „R“.*

Ist $\leq$ eine Ordnungsrelation auf X, so heißt X partiell geordnet*.*

1.3.1. Ordnungsrelationen

Beispiel 1.3.4. Ordnungsrelationen sind z.B.

(a) „$\leq$“ in $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$.

(b) die lexikographische Ordnung.

(c) Ist M eine Menge, so ist $\subseteq$ eine Ordnungsrelation auf $\mathcal{P}(M)$.

Bemerkung 1.3.5. (a) Hat man eine Ordnungsrelation $\leq$ auf einer Menge X, so kann es immer noch sein, dass es Elemente $x, y \in X$ gibt, die unvergleichbar sind, für die also weder $x \leq y$ noch $y \leq x$ gilt, vgl. z.B. Beispiel 1.3.4 (c). Gilt für eine Ordnungsrelation zusätzlich

$$\text{Für alle } x, y \in X \text{ gilt } x \leq y \text{ oder } y \leq x,$$

so heißt $\leq$ eine *Totalordnung* und die Menge X dann *total geordnet*.

(b) Ist $(X, \leq)$ eine partiell (total) geordnete Menge, so ist auch jede Teilmenge Y von X durch $\leq$ partiell (total) geordnet.

(c) Sei $(X, \leq)$ eine partiell geordnete Menge. Wir schreiben

$$x \geq y, \text{ falls } y \leq x,$$
$$x < y, \text{ falls } x \leq y \text{ und } x \neq y,$$
$$x > y, \text{ falls } y < x.$$

Definition 1.3.6. *Sei $(X, \leq)$ eine partiell geordnete Menge und $Y \subseteq X$*

(a) $g \in X$ *heißt* größtes Element *von X, falls $x \leq g$ für alle $x \in X$.*

$k \in X$ *heißt* kleinstes Element *von X, falls $k \leq x$ für alle $x \in X$.*

(b) $s \in X$ *heißt* obere Schranke *von Y, falls $y \leq s$ für alle $y \in Y$.*

$t \in X$ *heißt* untere Schranke *von Y, falls $t \leq y$ für alle $y \in Y$.*

Satz 1.3.7. *Sei $(X, \leq)$ eine partiell geordnete Menge. Dann hat X höchstens ein größtes und höchstens ein kleinstes Element.*

Beweis. Seien g_1 und g_2 größte Elemente von X. Da g_1 größtes Element ist, gilt $g_2 \leq g_1$. Da aber auch g_2 ein größtes Element ist, haben wir auch $g_1 \leq g_2$. Also ist wegen der Antisymmetrie von Ordnungsrelationen $g_1 = g_2$.
Für die kleinsten Elemente führt ein analoges Argument zum Ziel. □

Definition 1.3.8. *Es sei $(X, \leq)$ eine partiell geordnete Menge und $Y \subseteq X$.*

(a) Hat $S := \{s \in X : s$ obere Schranke von $Y\}$ ein kleinstes Element s_0, so heißt $\sup Y := s_0$ Supremum *von Y.*

Hat $T := \{t \in X : t$ untere Schranke von $Y\}$ ein größtes Element t_0, so heißt $\inf Y := t_0$ Infimum *von Y.*

(b) Gilt $s_0 = \sup(Y) \in Y$, so heißt s_0 Maximum *von Y; Bezeichnung* $\max Y$.

Gilt $t_0 = \inf(Y) \in Y$, so heißt t_0 Minimum *von Y; Bezeichnung* $\min Y$.

Merkregel:

Das Supremum ist die kleinste obere Schranke.
Das Infimum ist die größte untere Schranke.

Beispiel 1.3.9. (a) $\mathbb{Q}_+ := \{x \in \mathbb{Q} : x > 0\}$ hat in $\mathbb{Q}$, versehen mit der üblichen Ordnung, kein größtes und kein kleinstes Element. Wohl hat diese Menge aber untere Schranken, z.B. -7, -43 oder 0. Die größte untere Schranke und damit $\inf \mathbb{Q}_+$ ist 0. Dieses ist aber kein Minimum, denn $0 \notin \mathbb{Q}_+$.

(b) $\{x \in \mathbb{Q} : x^2 < 2\}$ hat in $\mathbb{Q}$ obere Schranken, z.B. 2 oder 37, aber kein Supremum, denn die Menge der oberen Schranken ist $\{q \in \mathbb{Q} : q \geq \sqrt{2}\}$ und diese Menge hat kein kleinstes Element, denn $\sqrt{2} \notin \mathbb{Q}$.

(c) In $\mathbb{N}$ mit der üblichen Ordnung hat jede nicht-leere Teilmenge ein Minimum und jede nihct-leere, endliche Teilmenge ein Maximum.

(d) In $(\mathcal{P}(\{0,1,2\}), \subseteq)$ hat die Teilmenge $\mathcal{M} := \{\emptyset, \{0\}\}$ obere Schranken, z.B. $\{0\}$, $\{0,1\}$ und $\{0,2\}$. Dabei ist $\{0\}$ die kleinste obere Schranke, also das Supremum, das in diesem Fall, wegen $\{0\} \in \mathcal{M}$, auch das Maximum ist.

Hat $\mathcal{N} := \big\{\emptyset, \{0\}, \{1\}\big\}$ ein Supremum, Infimum, Maximum, bzw. Minimum?

1.3.2. Äquivalenzrelationen

Beispiel 1.3.10. (a) „=“ in $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$

(b) gleicher Vorname, Pulloverfarbe, Armlänge in Menschengruppen

(c) Verwandtschaftsbeziehungen

Beispiel 1.3.11. Sei $n \in \mathbb{N}^*$ fest gewählt. Wir definieren die Relation $\sim_n$ auf $\mathbb{Z}$ durch

$$a \sim_n b \iff a - b \text{ ist Vielfaches von } n \iff \exists k \in \mathbb{Z} : a - b = k \cdot n, \quad a, b \in \mathbb{Z}.$$

Wir werden gleich zeigen, dass $\sim_n$ eine Äquivalenzrelation auf $\mathbb{Z}$ ist. Vorher sei noch vermerkt, dass man statt $a \sim_n b$ oft $a \equiv b \pmod{n}$ schreibt, gelesen: „a ist kongruent b modulo n“.
Beispielsweise ist

$$\begin{aligned} &19 \equiv 9 \pmod 5, \quad \text{denn } 19 - 9 = 10 \text{ ist Vielfaches von } 5, \\ &23 \equiv 1 \pmod 2, \\ &17 \equiv 3 \pmod 7. \end{aligned}$$

Nun zum Nachweis, dass $\sim_n$ Äquivalenzrelation ist:

1. Reflexivität: Sei $a \in \mathbb{Z}$. Dann ist $a - a = 0 = 0 \cdot n$ ein Vielfaches von n, also gilt $a \sim_n a$.
2. Symmetrie: Seien $a, b \in \mathbb{Z}$ mit $a \sim_n b$. Dann gibt es ein $k \in \mathbb{Z}$ mit $a - b = k \cdot n$. Also gilt $b - a = (-k) \cdot n$. Nun ist auch $-k \in \mathbb{Z}$. Also gibt es ein $\ell \in \mathbb{Z}$ mit $b - a = \ell \cdot n$, d.h. $b \sim_n a$.
3. Transitivität: Seien $a, b, c \in \mathbb{Z}$ mit $a \sim_n b$ und $b \sim_n c$. Das bedeutet, dass es zwei Zahlen $k, \ell \in \mathbb{Z}$ gibt mit $a - b = k \cdot n$ und $b - c = \ell \cdot n$. Damit ist

$$a - c = a - b + b - c = k \cdot n + \ell \cdot n = (k + \ell) \cdot n.$$

Da auch $k + \ell \in \mathbb{Z}$ ist, folgt damit $a \sim_n c$.

Satz 1.3.12. *Sei $\sim$ eine Äquivalenzrelation auf einer Menge $X \neq \emptyset$. Wir definieren für jedes $a \in X$ die* Äquivalenzklasse $\tilde{a}$ *als*

$$\tilde{a} := \{x \in X : x \sim a\}.$$

Dann gilt

(a) $\tilde{a} \neq \emptyset$ für jedes $a \in X$.

(b) Für alle $a, b \in X$ mit $\tilde{a} \neq \tilde{b}$ gilt $\tilde{a} \cap \tilde{b} = \emptyset$.

(c) $\bigcup_{a \in X} \tilde{a} = X$, d.h. die Vereinigung aller Äquivalenzklassen ist gleich X.

Beweis. (a) Sei $a \in X$. Wegen der Reflexivität von $\sim$, gilt $a \sim a$, also ist $a \in \tilde{a}$.

(b) Wir beweisen die Aussage per Kontraposition (vgl. Bemerkung 1.1.6), d.h. *wir zeigen:* $\tilde{a} \cap \tilde{b} \neq \emptyset \Longrightarrow \tilde{a} = \tilde{b}$.

Wenn $\tilde{a} \cap \tilde{b} \neq \emptyset$ ist, so gibt es ein Element x aus dieser Menge. Für dieses x gilt dann sowohl $x \sim a$, als auch $x \sim b$. Wegen der Symmetrie von $\sim$, haben wir also $a \sim x$ und $x \sim b$ und damit folgt aus der Transitivität $a \sim b$.

Sei nun $y \in \tilde{a}$. Dann ist $y \sim a$ und da wir auch $a \sim b$ haben, folgt wieder mit der Transitivität von $\sim$ die Beziehung $y \sim b$. Das bedeutet $y \in \tilde{b}$ und wir haben damit $\tilde{a} \subseteq \tilde{b}$ gezeigt.

Startet man mit einem $z \in \tilde{b}$, so zeigt man genauso $z \in \tilde{a}$ und bekommt $\tilde{b} \subseteq \tilde{a}$.

Zusammen ist also $\tilde{a} = \tilde{b}$ und wir sind fertig.

(c) Zunächst gilt für alle $a \in X$ natürlich $\tilde{a} \subseteq X$, also ist auch $\bigcup_{a \in X} \tilde{a} \subseteq X$. Wir müssen nur noch die umgekehrte Inklusion zeigen.

Sei $b \in X$. Dann ist $b \in \tilde{b}$ nach (a), also ist auch $b \in \bigcup_{a \in X} \tilde{a}$ und wir haben die umgekehrte Inklusion. □

Bemerkung 1.3.13. Satz 1.3.12 bedeutet, dass die Äquivalenzrelation $\sim$ eine Zerlegung von X in die Äquivalenzklassen erzeugt, die die Elemente von X nach der durch $\sim$ beschriebenen Eigenschaft sortiert.
Die Menge

$$X/_\sim := \{\tilde{a} : a \in X\}$$

aller Äquivalenzklassen heißt *Faktormenge* von X bezüglich $\sim$.

Beispiel 1.3.14. Als Beispiel betrachten wir wieder $\sim_n$ auf $\mathbb{Z}$ aus Beispiel 1.3.11. Dann ist für jedes $a \in \mathbb{Z}$

$$\begin{aligned}\tilde{a} &= \{b \in \mathbb{Z} : a \sim_n b\} = \{b \in \mathbb{Z} : \exists k \in \mathbb{Z} \text{ mit } a - b = n \cdot k\} \\ &= \{b \in \mathbb{Z} : b = a - nk \text{ für ein } k \in \mathbb{Z}\} = \{b \in \mathbb{Z} : b = a + nk \text{ für ein } k \in \mathbb{Z}\} \\ &= \{a + nk : k \in \mathbb{Z}\} =: a + n \cdot \mathbb{Z}.\end{aligned}$$

Für $n = 3$ gilt also beispielsweise

$$\begin{aligned}\tilde{0} &= 0 + 3 \cdot \mathbb{Z} = \{\dots, -6, -3, 0, 3, 6, \dots\} && \text{(durch 3 teilbare Zahlen)} \\ \tilde{1} &= 1 + 3 \cdot \mathbb{Z} = \{\dots, -5, -2, 1, 4, 7, \dots\} && \text{(Rest 1 beim teilen durch 3)} \\ \tilde{2} &= 2 + 3 \cdot \mathbb{Z} = \{\dots, -4, -1, 2, 5, 8, \dots\} && \text{(Rest 2 beim teilen durch 3)} \\ \tilde{3} &= 3 + 3 \cdot \mathbb{Z} = \{\dots, -3, 0, 3, 6, 9, \dots\} = \tilde{0}.\end{aligned}$$

Also ist

$$\mathbb{Z}/_{\sim_3} = \{\tilde{0}, \tilde{1}, \tilde{2}\}$$

eine Zerlegung von $\mathbb{Z}$, die die ganzen Zahlen nach ihrem Rest beim Teilen durch drei sortiert. Genauso enthält $\mathbb{Z}/_{\sim_n}$ die n Elemente $\tilde{0}$, $\tilde{1}$, …, $\widetilde{n-1}$ und in $\tilde{a}$ sind jeweils alle die ganzen Zahlen enthalten, die beim Teilen durch n den Rest a haben.
Man schreibt meist kurz $\mathbb{Z}_n$ statt $\mathbb{Z}/_{\sim_n}$.
Wir werden uns diesem Thema im Abschnitt 2.1 noch genauer widmen.

1.4. Abbildungen

Definition 1.4.1. *Seien A und B Mengen und jedem Element $a \in A$ sei genau ein Element $f(a) \in B$ zugeordnet. Diese Zuordnung heißt* Abbildung *oder* Funktion *f. Man schreibt*

$$f : \begin{cases} A \to B \\ a \mapsto f(a). \end{cases}$$

und nennt A den Definitionsbereich*, B den* Zielbereich*, sowie $a \mapsto f(a)$ die* Funktionsvorschrift *von f.*
Weiter heißt die Menge $f(A) := \{f(a) : a \in A\} \subseteq B$ das Bild *und die Menge $\{(a, f(a)) : a \in A\} \subseteq A \times B$ der* Graph *von f.*
Ist schließlich $C \subseteq B$, so bezeichnet man mit $f^{-1}(C) := \{a \in A : f(a) \in C\} \subseteq A$ das Urbild *von C unter f.*

Beispiel 1.4.2. (a) Bekannt sind Funktionen wie

$$f : \begin{cases} \mathbb{R} \to \mathbb{R} \\ x \mapsto x^2 \end{cases} \qquad \text{oder} \qquad g : \begin{cases} [0, \infty) \to [0, \infty) \\ x \mapsto \sqrt{x}. \end{cases}$$

(b) Auch $+ : \begin{cases} \mathbb{N} \times \mathbb{N} \to \mathbb{N} \\ (a, b) \mapsto a + b \end{cases}$, d.h. die Addition in $\mathbb{N}$, ist eine Abbildung.

(c) Auf jeder Menge A kann man die *Identität*, d.h. $id : \begin{cases} A \to A \\ a \mapsto a \end{cases}$ definieren.

(d) Ist X eine Menge mit einer auf X erklärten Äquivalenzrelation $\sim$, so ist $\nu : \begin{cases} X \to X/_\sim \\ a \mapsto \tilde{a} \end{cases}$ die sogenannte *kanonische Abbildung.*

Definition 1.4.3. *Seien A, B, C Mengen und $f : A \to B$, sowie $g : B \to C$ Funktionen. Dann heißt*

$$g \circ f : \begin{cases} A \to C \\ a \mapsto (g \circ f)(a) := g(f(a)) \end{cases}$$

Verkettung *von f und g. Man liest $g \circ f$ als „g nach f“.*

Definition 1.4.4. *Eine Funktion $f : A \to B$ heißt*

(a) surjektiv*, wenn $f(A) = B$.*

(b) injektiv*, wenn für alle $x, y \in A$ aus $f(x) = f(y)$ schon $x = y$ folgt.*

(c) bijektiv*, wenn f surjektiv und injektiv ist.*

Satz 1.4.5. *Eine Funktion $f : A \to B$ ist genau dann bijektiv, wenn für jedes $b \in B$ genau ein $a \in A$ existiert mit $f(a) = b$. In diesem Fall existiert eine Abbildung $f^{-1} : B \to A$, so dass*

$$f^{-1}(f(a)) = a \quad \text{für alle } a \in A \quad \text{und} \quad f(f^{-1}(b)) = b \quad \text{für alle } b \in B$$

gilt.

Beweis. 1. Schritt: Wir zeigen: f bijektiv $\Longrightarrow$ für alle $b \in B$ existiert genau ein $a \in A$ mit $f(a) = b$.

Da f surjektiv ist, gibt es zu jedem $b \in B$ mindestens ein $a \in A$ mit $f(a) = b$. Nehmen wir an, es gäbe mehr als eins, d.h. es gäbe $a_1, a_2 \in A$ mit $f(a_1) = f(a_2) = b$, so folgt aus der Injektivität von f sofort $a_1 = a_2$, es kann also nur genau ein solches $a \in A$ geben.

2. Schritt: Wir zeigen: Für alle $b \in B$ existiert genau ein $a \in A$ mit $f(a) = b \Longrightarrow f$ bijektiv.

Nach Voraussetzung sind alle $b \in B$ in $f(A)$ enthalten, also ist f surjektiv. Seien nun $a_1, a_2 \in A$ mit $f(a_1) = f(a_2)$ gegeben. Da jedes $b \in B$ nur genau ein Urbild hat, muss dann $a_1 = a_2$ sein, d.h. f ist auch injektiv.

3. Schritt: Wir zeigen: f bijektiv $\Longrightarrow$ es existiert $f^{-1} : B \to A$ mit $f^{-1}(f(a)) = a$ für alle $a \in A$ und $f(f^{-1}(b)) = b$ für alle $b \in B$.

Für jedes $b \in B$ definieren wir $f^{-1}(b) := a$, wobei $a \in A$ das nach dem ersten Schritt eindeutig bestimmte Element mit $f(a) = b$ ist. Dann ist $f^{-1}(f(a))$ das Element von A, das in f eingesetzt $f(a)$ ergibt, also $f^{-1}(f(a)) = a$ für alle $a \in A$. Sei nun $b \in B$. Dann ist $f^{-1}(b)$ das Element von A mit $f(f^{-1}(b)) = b$ und wir sind fertig. □

Definition 1.4.6. *Es seien A, B zwei Mengen und $f : A \to B$ bijektiv. Dann heißt die Abbildung f^{-1} aus Satz 1.4.5* Umkehrfunktion *von f.*

Beispiel 1.4.7. Die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = x^2$ (vgl. Beispiel 1.4.2 (a)) ist nicht injektiv, denn es gilt $f(1) = 1^2 = 1 = (-1)^2 = f(-1)$, aber $1 \neq -1$. Sie ist auch nicht surjektiv, denn $-1 \notin f(\mathbb{R})$.
Betrachtet man $\hat{f} : \mathbb{R} \to [0, \infty)$ mit $\hat{f}(x) = x^2$, so ist diese nun surjektiv, denn $f(\mathbb{R}) = \{x^2 : x \in \mathbb{R}\} = [0, \infty)$, aber genau so wie oben nicht injektiv.
Geht man jedoch zu $\hat{\hat{f}} : [0, \infty) \to [0, \infty)$ mit $\hat{\hat{f}}(x) = x^2$ über, so ist diese injektiv und surjektiv, d.h. bijektiv. Die nach Satz 1.4.5 existierende Umkehrfunktion $\hat{\hat{f}}^{-1}$ ist genau die Funktion g aus Beispiel 1.4.2 (a), d.h. die Wurzelfunktion.

Definition 1.4.8. *Sei $f : A \to B$ eine Funktion und $M \subseteq A$. Dann heißt*

$$f|_M : \begin{cases} M \to B \\ x \mapsto f(x) \end{cases}$$

die Einschränkung *von f auf M.*

Übungsaufgabe 1.4.9. Beweisen Sie: Sind $f : A \to B$ und $g : B \to C$ bijektive Funktionen, so ist auch $g \circ f : A \to C$ bijektiv.

Übungsaufgabe 1.4.10. Unter welchen Voraussetzungen an die Menge A ist die Abbildung $id : A \to A$ bijektiv? Bestimmen Sie in diesem Fall id^{-1}.

1.5. Beweisprinzipien

In einem Beweis ist die Aufgabe aus einer Aussage A, der *Voraussetzung*, eine Aussage B, die *Behauptung*, zu folgern. Anders ausgedrückt: Man muss nachweisen, dass die Aussage $A \Longrightarrow B$ wahr ist. Selbst, wenn der Satz, der zu beweisen ist, eine Äquivalenz, d.h. eine Aussage der Form $A \iff B$ postuliert, wird der Beweis fast immer in die Teilbeweise $A \Longrightarrow B$ und $B \Longrightarrow A$ aufgeteilt, vgl. den Beweis von Satz 1.4.5.
Dieser Abschnitt stellt mögliche Beweismethoden zusammen und liefert jeweils ein kurzes Beispiel. Einige davon haben wir in den vorherigen Kapiteln schon gesehen, einige sind neu.

1.5.1. Der direkte Beweis

Der *direkte Beweis* hat folgende Form:

Voraussetzung:	Aussage A
Behauptung:	Aussage B
Beweis:	Sei A erfüllt. Dann gilt ... bla bla bla und deswegen .... Also gilt auch B.

Bisherige Beispiele für direkte Beweise waren die Beweise von (a) und (c) aus Satz 1.3.12 und die von Satz 1.2.5. Hier ist ein weiteres:

Beispiel 1.5.1. *Voraussetzung*: Seien $n, m \in \mathbb{N}$ gerade Zahlen.
Behauptung: Dann ist auch $n + m$ gerade.
Beweis: Seien n und m gerade Zahlen. Dann gibt es $\ell, k \in \mathbb{N}$ mit $n = 2\ell$ und $m = 2k$. Mit diesen ℓ, k gilt dann $n + m = 2\ell + 2k = 2(\ell + k)$. Mit ℓ und k ist auch $p := \ell + k \in \mathbb{N}$. Also haben wir gezeigt, dass es ein $p \in \mathbb{N}$ mit $n + m = 2p$ gibt. Damit ist $n + m$ gerade. □

1.5.2. Beweis durch Kontraposition

Der *indirekte Beweis* oder auch *Beweis durch Kontraposition* hat folgende Form:

Voraussetzung:	Aussage A
Behauptung:	Aussage B
Beweis:	Es gelte $\neg B$. Dann gilt ... bla bla bla und deswegen .... Also ist auch A falsch.

Durch diese Beweisführung $\neg B \Longrightarrow \neg A$ ist auch die Aussage $A \Longrightarrow B$ wahr, vgl. Bemerkung 1.1.6. Ein Beispiel für einen Beweis durch Kontraposition haben wir bereits bei Satz 1.3.12 (b) gesehen. Ein weiteres kurzes Beispiel ist folgendes:

Beispiel 1.5.2. *Voraussetzung*: Sei $n \in \mathbb{N}$ mit n^2 gerade.
Behauptung: Dann ist auch n gerade.
Beweis: Sei $n \in \mathbb{N}$ so, dass die Behauptung nicht gilt, d.h. n sei ungerade. Dann gibt es ein $k \in \mathbb{N}$ mit $n = 2k + 1$ und es gilt $n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$. Damit haben wir $\ell := 2k^2 + 2k \in \mathbb{N}$ gefunden mit $n^2 = 2\ell + 1$. Dann ist n^2 ebenfalls ungerade und die Aussage „n^2 gerade“ falsch. □

1.5.3. Beweis durch Widerspruch

Der *Beweis durch Widerspruch* ist eng verwandt mit der Kontraposition. Seine übliche Form ist die folgende:

Voraussetzung:	Aussage A
Behauptung:	Aussage B
Beweis:	Es gelte A. Angenommen B wäre falsch. Dann gilt ... bla bla bla und deswegen .... Also ergäbe sich ein Widerspruch. Damit war die Annahme falsch und es gilt B.

Eines der typischen ersten Beispiele für diese Beweistechnik ist der Beweis, dass $\sqrt{2}$ irrational ist.

Beispiel 1.5.3. *Behauptung*: Die Zahl $\sqrt{2}$ ist nicht rational.
Beweis: *Annahme*: $\sqrt{2}$ ist rational.
Dann gibt es $n, m \in \mathbb{N}$ mit $\sqrt{2} = n/m$. Außerdem können wir annehmen, dass dieser Bruch bereits maximal gekürzt ist, d.h. wir können die Zahlen n und m teilerfremd wählen. Es gilt $2 = \sqrt{2}^2 = n^2/m^2$, d.h.

$$n^2 = 2m^2. \tag{1.1}$$

Aus dieser Gleichheit bekommen wir jetzt insbesondere, dass die Zahl n^2 eine gerade Zahl ist und nach Beispiel 1.5.2 ist dann auch n gerade. Also gibt es ein

$k \in \mathbb{N}$ mit $n = 2k$ und wir erhalten wieder mit (1.1), dass $2m^2 = (2k)^2 = 4k^2$, d.h. $m^2 = 2k^2$ ist.
Also ist auch m^2 gerade und damit wie oben m gerade und wir haben einen Widerspruch, denn nun sind n und m teilerfremd und beide gerade.
Also war die Annahme falsch und $\sqrt{2}$ ist nicht rational. □

1.5.4. Vollständige Induktion über $\mathbb{N}$

Die *vollständige Induktion* ist ein Beweisverfahren, das dazu dient, die Richtigkeit einer Aussageform $E(n)$ für alle natürlichen Zahlen n nachzuweisen. Es sieht so aus:

Voraussetzung:	Aussage A
Behauptung:	Für alle $n \in \mathbb{N}$ gilt $E(n)$
Beweis:	*Induktionsanfang*: Es gilt A und bla bla bla, also gilt auch $E(0)$.
	Induktionsvoraussetzung: Für ein $n \in \mathbb{N}$ gelte $E(n)$.
	Induktionsschluss: $E(n)$ und A sind wahr, also ist ... bla bla bla und deswegen .... Damit gilt auch $E(n+1)$.

Bemerkung 1.5.4. Das Verfahren funktioniert allgemeiner auch um zu zeigen, dass eine Aussage $E(n)$ für alle $n \geq n_0$, für ein $n_0 \in \mathbb{N}$, also ab einem gewissen n_0 für alle größeren n gilt. Dann muss der Induktionsanfang den Nachweis erbringen, dass $E(n_0)$ wahr ist.
Außerdem werden Sie in der Vorlesung „Formale Grundlagen der Informatik“ noch weitere Verallgemeinerungen dieser Methodik auf andere Strukturen als $\mathbb{N}$ kennen lernen.

Beispiel 1.5.5. *Behauptung*: Für alle $n \in \mathbb{N}^*$ gilt $1 + 2 + \cdots + n = \dfrac{n(n+1)}{2}$.
Beweis: *Induktionsanfang*: Auf der linken Seite der behaupteten Gleichheit steht für $n = 1$ einfach 1 und auf der rechten Seite steht $1 \cdot (1+1)/2 = 2/2 = 1$. Also stimmt diese für $n = 1$.
Induktionsvoraussetzung: Für ein $n \in \mathbb{N}^*$ gelte $1 + 2 + \cdots + n = \dfrac{n(n+1)}{2}$.
Induktionsschritt: Es ist $1 + 2 + \cdots + (n+1) = (1 + 2 + \cdots + n) + (n+1)$, also erhalten wir mit der Induktionsvoraussetzung

$$\begin{aligned} 1 + 2 + \cdots + (n+1) &= \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2} \\ &= \frac{(n+2)(n+1)}{2} = \frac{(n+1)\big((n+1)+1\big)}{2}, \end{aligned}$$

was die behauptete Gleichheit für $n+1$ zeigt. □

Da das Prinzip der Induktion bisher noch nicht vorkam, hier noch ein Beispiel.

Beispiel 1.5.6. *Behauptung*: Für jede endliche Menge M gilt $|\mathcal{P}(M)| = 2^{|M|}$.
Beweis: Wir führen eine Induktion nach der Mächtigkeit der Menge M.
Induktionsanfang: Ist $|M| = 0$, so muss $M = \emptyset$ sein. Dann ist $\mathcal{P}(M) = \mathcal{P}(\emptyset) = \{\emptyset\}$ und wir haben $|\mathcal{P}(M)| = 1 = 2^0 = 2^{|M|}$. Die Behauptung stimmt also für $|M| = 0$.
Induktionsvoraussetzung: Für ein $n \in \mathbb{N}$ gilt für alle Mengen mit n Elementen $|\mathcal{P}(M)| = 2^{|M|}$.
Induktionsschritt: Sei M eine Menge mit $n+1$ Elementen. Dann hat M mindestens ein Element. Sei also ein $x \in M$ fest gewählt. Wir betrachten nun $N := M \setminus \{x\}$. Dann hat N genau n Elemente, nach der Induktionsvoraussetzung gilt also $|\mathcal{P}(N)| = 2^{|N|} = 2^n$.
Es gilt aber

$$\mathcal{P}(M) = \mathcal{P}(N) \cup \big\{A \cup \{x\} : A \in \mathcal{P}(N)\big\}. \tag{1.2}$$

Um das einzusehen, beweisen wir zunächst die Inklusion „$\subseteq$“. Sei also $B \in \mathcal{P}(M)$. Dann ist entweder $x \in B$ oder $x \notin B$. Im zweiten Fall ist B auch Teilmenge von N also in $\mathcal{P}(N)$ und damit in der Menge auf der rechten Seite in (1.2). Ist $x \in B$, so ist $\hat{B} := B \setminus \{x\} \in \mathcal{P}(N)$ und damit $B = \hat{B} \cup \{x\}$ wiederum in dem Mengensystem auf der rechten Seite von (1.2) enthalten. Die Inklusion „$\supseteq$“ ist klar, denn wegen $x \in M$ und $N \subseteq M$ ist jedes Element des rechten Mengensystems eine Teilmenge von M.
Mit der Hilfe von (1.2) sind wir nun bald am Ziel. Wichtig ist noch die Beobachtung, dass wegen $x \notin N$

$$\mathcal{P}(N) \cap \big\{A \cup \{x\} : A \in \mathcal{P}(N)\big\} = \emptyset$$

gilt, denn damit folgt mit Übungsaufgabe 1.2.8 und der Induktionsvoraussetzung

$$\begin{aligned} |\mathcal{P}(M)| &= \big|\mathcal{P}(N) \cup \big\{A \cup \{x\} : A \in \mathcal{P}(N)\big\}\big| \\ &= |\mathcal{P}(N)| + \big|\big\{A \cup \{x\} : A \in \mathcal{P}(N)\big\}\big| \\ &= |\mathcal{P}(N)| + |\mathcal{P}(N)| = 2|\mathcal{P}(N)| = 2 \cdot 2^{|N|} = 2 \cdot 2^n = 2^{n+1} \end{aligned}$$

und wir haben die Behauptung mit $|M| = n+1$ gezeigt. □

2. Algebraische Strukturen: Gruppen, Ringe, Körper

Nach dem vorhergehenden Abschnitt, der vor allem die Sprache der Mathematik einführen sollte, wollen wir nun „richtig“ anfangen. Ein häufiges Missverständnis über Mathematik, ist „Mathematik = Rechnen“, das werden Sie schon gemerkt haben, so viel gerechnet haben wir bisher nicht. Eher mathematisch ist die folgende Frage: Was ist das überhaupt: „rechnen“? Was machen wir, wenn wir rechnen? Was ist die dahinterliegende allgemeine Struktur? Dieser Frage wollen wir ein bisschen nachgehen und uns verschiedene Rechenstrukturen anschauen. Der Weg wird nicht ganz geradlinig sein, sondern wir werden den einen oder anderen Abstecher, z.B. zum RSA-Algorithmus aus der Public-Key-Verschlüsselung machen, aber die Grundfrage dieses Abschnitts ist obiges „was ist rechnen?“ Beginnen wollen wir auf vertrautem Grund, dem Rechnen mit ganzen Zahlen.

2.1. Rechnen in $\mathbb{Z}$, Primzahlen und Teiler

Definition 2.1.1. *Es seien $a, b \in \mathbb{Z}$ und $p \in \mathbb{N}$.*

(a) Man sagt p teilt *a und schreibt $p|a$, falls ein $m \in \mathbb{Z}$ existiert mit $a = m \cdot p$.*

(b) Eine natürliche Zahl $p > 1$ heißt Primzahl*, wenn p nur durch p und 1 teilbar ist.*

(c) Die Zahl $\mathrm{ggT}(a, b) := \max\{q \in \mathbb{N} : q|a \text{ und } q|b\}$ *heißt* größter gemeinsamer Teiler *von a und b.*

Satz 2.1.2 (Division mit Rest)**.** *Seien $a \in \mathbb{Z}$ und $b \in \mathbb{N}^*$. Dann gibt es eindeutig bestimmte Zahlen $q \in \mathbb{Z}$ und $r \in \{0, 1, \ldots, b-1\}$ mit $a = q \cdot b + r$.*

Beweis. Wir betrachten nur den Fall $a \geq 0$, der Beweis im Fall $a < 0$ verläuft analog. Zu vorgegebenen a und b betrachten wir die Menge

$$M := \{s \in \mathbb{N} : s \cdot b \leq a\}.$$

Dann ist $M \subseteq \{0, 1, \ldots, a\}$, denn für alle $s \in M$ gilt $s = s \cdot 1 \leq s \cdot b \leq a$. Damit existiert $q := \max M$ als größte ganze Zahl, für die noch $q \cdot b \leq a$ gilt. Mit diesem q setzen wir nun $r := a - q \cdot b$. Dann ist in jedem Fall $a = q \cdot b + r$.

Zum Nachweis, dass $r \in \{0, 1, \ldots, b-1\}$ gilt, überlegen wir uns zunächst, dass wegen $q \cdot b \leq a$ auch $r = a - q \cdot b \geq 0$ gilt. Wir müssen also noch zeigen, dass $r < b$ ist. Nehmen wir an, es wäre $r \geq b$, so folgt

$$(q+1) \cdot b = qb + b \leq qb + r = a$$

und damit wäre $q + 1 \in M$, was im Widerspruch zur Konstruktion von q als größtem Element von M steht.
Es bleibt noch die Eindeutigkeit zu zeigen. Seien dazu $q_1, q_2 \in \mathbb{Z}$ und $r_1, r_2 \in \{0, 1, \ldots, b-1\}$ mit $a = q_1 \cdot b + r_1 = q_2 \cdot b + r_2$. Dann gilt

$$(q_1 - q_2) \cdot b = r_2 - r_1.$$

Insbesondere teilt damit b die Zahl $r_2 - r_1$. Nun liegt aber $r_2 - r_1$ zwischen $-(b-1)$ und $b-1$ und die einzige Zahl in diesem Bereich, die durch b teilbar ist, ist Null. Also gilt $r_2 - r_1 = 0$, d.h. $r_2 = r_1$. Damit ist aber $q_1 \cdot b = q_2 \cdot b$ und wegen $b \neq 0$ erhalten wir auch $q_1 = q_2$ und sind fertig. □

Definition 2.1.3. *Seien $a \in \mathbb{Z}$, $b \in \mathbb{N}^*$ und q und r seien die eindeutig bestimmten Zahlen aus Satz 2.1.2. Dann heißt q* Quotient *und r* Rest *der Division von a und b. Man schreibt*

$$q = \left\lfloor \frac{a}{b} \right\rfloor \qquad \text{und} \qquad r = a \bmod b.$$

Übungsaufgabe 2.1.4. Zeigen Sie:

(a) Für jedes $a \in \mathbb{Z}$ und $b \in \mathbb{N}^*$ ist die Zahl $a \bmod b$ das eindeutige $r \in \{0, 1, \ldots, b-1\}$ mit $b|(a-r)$.

(b) $a|b \iff b \bmod a = 0$.

2.1.1. Modulare Arithmetik

In diesem ganzen Abschnitt sei $n \in \mathbb{N}^*$ eine feste Zahl.

Satz 2.1.5. *Für alle $a, b \in \mathbb{Z}$ gilt*

(a) $(a+b) \bmod n = \big((a \bmod n) + (b \bmod n)\big) \bmod n$.

(b) $(a \cdot b) \bmod n = \big((a \bmod n) \cdot (b \bmod n)\big) \bmod n$.

(c) $a^b \bmod n = (a \bmod n)^b \bmod n$.

Beweis. Seien $k = \lfloor a/n \rfloor$ und $\ell = \lfloor b/n \rfloor$, d.h.

$$a = kn + a \bmod n \qquad \text{und} \qquad b = \ell n + b \bmod n.$$

(a) Mit obiger Notation gilt

$$a + b = kn + a \bmod n + \ell n + b \bmod n = (k + \ell)n + a \bmod n + b \bmod n.$$

Also ist

$$(a + b) \bmod n = \big((a \bmod n) + (b \bmod n)\big) \bmod n,$$

denn $(k + \ell)n$ ist durch n teilbar.

(b) Übung

(c) Unter Verwendung von (b) gilt

$$a^b \bmod n = (\underbrace{a \cdot a \cdot \ldots \cdot a}_{b \text{ Mal}}) \bmod n = (a \bmod n)^b \bmod n. \qquad \square$$

Bemerkung 2.1.6. Obiges Resultat bedeutet, dass man, wann immer am Ende einer Rechnung nur der Rest modulo n interessiert, auch nach jedem Rechenschritt schon die Zwischenergebnisse modulo n reduzieren kann. Das ist insbesondere im Zusammenhang mit Fragen der Effizienz von Algorithmen und damit für die Geschwindigkeit von Computerprogrammen von Interesse.

Beispiel 2.1.7. Wir wählen mal $n = 7$ und berechnen

$$\begin{aligned}
&\big((9 - 15) \cdot 23 + 705\big)^{322} \bmod 7 \\
&= \big((9 \bmod 7 - 15 \bmod 7) \cdot (23 \bmod 7) + 705 \bmod 7\big)^{322} \bmod 7 \\
&= \big((2 - 1) \cdot 2 + 5\big)^{322} \bmod 7 = 7^{322} \bmod 7 \\
&= 0^{322} \bmod 7 = 0.
\end{aligned}$$

Wir haben also (einfach und von Hand!) herausgefunden, dass $((9 - 15) \cdot 23 + 705)^{322}$ durch 7 teilbar ist.

Als Übung können Sie zeigen, dass die Zahl $3^{444} + 4^{333}$ durch 5 teilbar ist.

2.1.2. Der Euklidische Algorithmus

Das erste Ziel dieses Abschnittes ist die algorithmische Bestimmung von $\mathrm{ggT}(a, b)$ für gegebene $a, b \in \mathbb{N}^*$.

Lemma 2.1.8. *Seien $a, b \in \mathbb{N}^*$ mit $a \geq b$. Dann gilt*

(a) $\mathrm{ggT}(a, b) = \mathrm{ggT}(b, a \bmod b)$

(b) $b|a \Longrightarrow \mathrm{ggT}(a, b) = b.$

Beweis. (a) Sei $d := \mathrm{ggT}(a,b)$. Dann gilt nach Definition $d|a$ und $d|b$, also ist $a \bmod d = 0$ und $b \bmod d = 0$, vgl. Übungsaufgabe 2.1.4 Damit gilt nach Definition der Division mit Rest und unter Mitwirkung von Satz 2.1.5

$$\begin{aligned}(a \bmod b) \bmod d &= \left[a - \left\lfloor \frac{a}{b} \right\rfloor b\right] \bmod d \\ &= a \bmod d - \left(\left\lfloor \frac{a}{b} \right\rfloor \bmod d\right)(b \bmod d) = 0.\end{aligned}$$

Wir finden also $d|(a \bmod b)$. Damit ist d schon mal gemeinsamer Teiler von b und $a \bmod b$. Es bleibt noch zu zeigen, dass d der größte solche ist.

Sei also c ein gemeinsamer Teiler von b und $a \bmod b$. Dann gilt wegen $c|b$ und $c|(a \bmod b)$ auch $c|(kb+a \bmod b)$ für jedes $k \in \mathbb{Z}$. Insbesondere können wir $k = \lfloor a/b \rfloor$ wählen. Dann ist $kb + a \bmod b = a$ und wir bekommen $c|a$. Also ist c auch ein gemeinsamer Teiler von a und b. Dieser kann nicht größer sein als $d = \mathrm{ggT}(a,b)$, also gilt $c \leq d$ und wir sind fertig.

(b) Es gilt immer $b|b$ und da b nach Voraussetzung auch a teilt, ist b schon mal ein gemeinsamer Teiler von a und b. Sei c ein weiterer gemeinsamer Teiler von a und b. Dann gilt wegen $c|b$ sofort $c \leq b$, womit b der größte gemeinsame Teiler von a und b ist. □

Beispiel 2.1.9. Wir bestimmen den größten gemeinsamen Teiler von 128 und 36. Es ist

$$\begin{aligned}128 \bmod 36 = 20,\ \text{da } 3 \cdot 36 = 108,\ \text{also } \mathrm{ggT}(128,36) &\overset{(a)}{=} \mathrm{ggT}(36,20) \\ 36 \bmod 20 = 16,\ \text{da } 1 \cdot 20 = 20,\ \text{also } \mathrm{ggT}(128,36) &\overset{(a)}{=} \mathrm{ggT}(20,16) \\ 20 \bmod 16 = 4,\ \text{da } 1 \cdot 16 = 16,\ \text{also } \mathrm{ggT}(128,36) &\overset{(a)}{=} \mathrm{ggT}(16,4) \\ 16 \bmod 4 = 0,\ \text{da } 4 \cdot 4 = 16,\ \text{also } \mathrm{ggT}(128,36) &\overset{(a)}{=} \mathrm{ggT}(4,0) \overset{(b)}{=} 4.\end{aligned}$$

Schematisch:

a	b
128	36
36	20
20	16
16	4
4	0

Beim Übergang von einer Zeile zur nächsten überträgt man jeweils die rechte Zahl in die linke Spalte und füllt die rechte Spalte mit dem Rest der beim Teilen mit Rest übrigbleibt. Man wiederholt dieses bis in der rechten Spalte einer Zeile Null steht, die Zahl in der linken Spalte dieser Zeile ist dann der größte gemeinsame Teiler.

Dieses Verfahren ist sehr wichtig und hat darum auch einen eigenen Namen.

Satz 2.1.10 (Euklidischer Algorithmus)**.** *Seien $a, b \in \mathbb{N}^*$ mit $a > b$. Der Algorithmus*

```
Euklid(a, b)
    IF b = 0 THEN return a
             ELSE return Euklid(b, a mod b)
```

terminiert nach endlich vielen Schritten und liefert $\mathrm{ggT}(a, b)$.

Beweis. Für jede Ausgangswahl von $a, b \in \mathbb{N}^*$ gilt $0 \leq a \bmod b < b$, also wird das zweite Argument des Aufrufs in jedem Schritt echt kleiner. Damit muss es in endlich vielen Schritten nach Null kommen, d.h. der Algorithmus terminiert.
Seien a_n, b_n die Eingangswerte beim n-ten Aufruf von Euklid. Terminiert der Algorithmus nach n Schritten, d.h. ist $b_n = 0$, so bedeutet das $a_{n-1} \bmod b_{n-1} = 0$, d.h. $b_{n-1} | a_{n-1}$. Damit ist nach Lemma 2.1.8 (b)

$$\mathrm{ggT}(a_{n-1}, b_{n-1}) = b_{n-1} = a_n = \mathrm{Euklid}(a, b).$$

Andererseits ist nach (a) des selben Lemmas

$$\mathrm{ggT}(a_{n-1}, b_{n-1}) = \mathrm{ggT}(a_{n-2}, b_{n-2}) = \cdots = \mathrm{ggT}(a_0, b_0) = \mathrm{ggT}(a, b). \qquad \square$$

Satz 2.1.11 (Erweiterter Euklidischer Algorithmus)**.** *Seien $a, b \in \mathbb{N}^*$ mit $a > b$. Der Algorithmus*

```
Erw-Euklid(a, b)
    IF b = 0 THEN return (a, 1, 0)
             ELSE DO
                      (d, x, y) := Erw-Euklid(b, a mod b)
                      return (d, y, x − ⌊a/b⌋ · y)
             OD
```

terminiert nach endlich vielen Schritten und liefert $(d, k, \ell) =$ Erw-Euklid(a, b) mit $d = \mathrm{ggT}(a, b)$ und die Zahlen k und ℓ erfüllen die Beziehung

$$d = \mathrm{ggT}(a, b) = ka + \ell b.$$

Der erweiterte Euklidsche Algorithmus liefert damit im Spezialfall von Zahlen mit ggT Eins insbesondere die folgende wichtige Erkenntnis.

Korollar 2.1.12. *Es seien $a, b \in \mathbb{N}^*$ mit $\mathrm{ggT}(a, b) = 1$. Dann gibt es $k, \ell \in \mathbb{Z}$ mit $ka + \ell b = 1$.*

Wir wollen Satz 2.1.11 hier nicht beweisen, sondern nur kurz erwähnen, dass der Algorithmus zur Bestimmung von d genau der selbe ist wie im einfachen Euklidschen Algorithmus, man also den Beweis, dass dies der ggT ist und dass der

Algorithmus nach endlich vielen Schritten terminiert von oben übernehmen kann. Die zweite Aussage über die Zahlen k und ℓ beweist man schließlich per Induktion nach der Anzahl der rekursiven Aufrufe, aber das soll hier nicht ausgeführt werden.

Beispiel 2.1.13. Wir starten den erweiterten Euklid mit $a = 141$ und $b = 9$. Wie oben bekommen wir ein Schema:

a	b	$\lfloor a/b \rfloor$	k	ℓ
141	9	15	-1	$16 = 1 - 15 \cdot (-1)$
9	6	1	1	$-1 = 0 - 1 \cdot 1$
6	3	2	0	$1 = 1 - 2 \cdot 0$
3	0		1	0

Man rechnet dabei zunächst die ersten zwei Spalten wie in Beispiel 2.1.9 und protokolliert zusätzlich in der dritten Spalte jeweils den Quotienten der ersten beiden Spalten mit. Damit bekommt man schon mal den ggT in der linken unteren Ecke, hier ist er 3.
Nun schreibt man 1 und 0 in die unterste Zeile der k- und ℓ-Spalte und rechnet von unten wieder rauf. Dabei kommt in die k-Spalte jeweils der Wert aus der ℓ-Spalte in der Zeile drunter und in die ℓ-Spalte kommt $x - q \cdot y$, wobei x der Wert aus der k-Spalte der Zeile darunter, q der Quotient aus der aktuellen Zeile und y der Wert aus der ℓ-Spalte der Zeile drunter ist.
Das Endergebnis besteht nun aus dem ggT der beiden Zahlen und aus den Einträgen bei k und ℓ in der ersten Zeile. Für diese beiden Zahlen gilt dann (vgl. den Satz) $\mathrm{ggT}(a, b) = ka + \ell b$.
In obigem Beispiel haben wir tatsächlich

$$ka + \ell b = (-1) \cdot 141 + 16 \cdot 9 = -141 + 144 = 3 = \mathrm{ggT}(141, 9).$$

Beispiel 2.1.14. Erw-Euklid liefert auch eine Methode, um Gleichungen zu lösen der Form: Gegeben a und n mit $\mathrm{ggT}(a, n) = 1$, finde x mit $ax \equiv 1 (\mathrm{mod}\ n)$. Ist nämlich $(d, k, \ell) = \text{Erw-Euklid}(n, a)$, so gilt $d = 1$ und $1 = kn + \ell a$, also

$$1 \bmod n = (kn + \ell a) \bmod n = \ell a \bmod n.$$

Also ist $x = \ell$ eine Lösung.
Als Beispiel betrachte man $93x \equiv 1 \pmod{100}$. Wegen $\text{Erw-Euklid}(100, 93) = (1, 40, -43)$ ist $x = -43 \bmod 100 = 57$ eine Lösung. Tatsächlich ist $57 \cdot 93 = 5301 \equiv 1 \pmod{100}$.

Übungsaufgabe 2.1.15. Es seien $a, b, n \in \mathbb{N}^*$ mit $\mathrm{ggT}(a, n) = 1$. Dann gilt $n|ab \Longrightarrow n|b$.
Daraus folgt, dass für Primzahlen p und $a, b \in \mathbb{N}^*$ aus $p|ab$ folgt, dass $p|a$ oder $p|b$.

2.1.3. Der kleine Satz von Fermat

Satz 2.1.16 (Kleiner Satz von Fermat)**.** *Für alle Primzahlen p und alle $a \in \mathbb{N}$ gilt $a^p \equiv a \pmod{p}$.*

Beweis. Wir führen eine Induktion nach a.
Induktionsanfang: Für $a = 0$ gilt $0^p = 0 \equiv 0 \pmod{p}$.
Induktionsvoraussetzung: Für ein $a \in \mathbb{N}$ gelte $a^p \equiv a \pmod{p}$.
Induktionsschluss: Nach der allgemeinen binomischen Formel gilt

$$(a+1)^p = a^p + \binom{p}{1}a^{p-1} + \binom{p}{2}a^{p-2} + \cdots + \binom{p}{p-1}a + 1, \tag{2.1}$$

wobei für jedes $k \in \{1, 2, \ldots, p-1\}$

$$\binom{p}{k} = \frac{p \cdot (p-1) \cdot \ldots \cdot (p-k+1)}{1 \cdot 2 \cdot \ldots \cdot k} \in \mathbb{N}$$

ist. Also teilt p die natürliche Zahl $\binom{p}{k} \cdot 1 \cdot 2 \cdot \ldots \cdot k$. Da p eine Primzahl ist und somit $\mathrm{ggT}(p, \ell) = 1$ für $\ell \in \{1, 2, \ldots, p-1\}$, folgt mit Übungsaufgabe 2.1.15, dass p die Zahl $\binom{p}{k}$ teilt. Damit liefert (2.1)

$$(a+1)^p \bmod p = \left(a^p + 0 \cdot a^{p-1} + 0 \cdot a^{p-2} + \cdots + 0 \cdot a + 1\right) \bmod p = (a^p + 1) \bmod p.$$

Nach der Induktionsvoraussetzung ist $a^p \bmod p = a \bmod p$, also haben wir schließlich

$$(a+1)^p \bmod p = (a+1) \bmod p,$$

womit die Behauptung für $a + 1$ gezeigt ist. □

Korollar 2.1.17. *Ist p Primzahl und $a \in \mathbb{N}$ eine Zahl, die nicht von p geteilt wird, so gilt $a^{p-1} \equiv 1 \pmod{p}$.*

Beweis. Nach Satz 2.1.16 gilt $a^p \equiv a \pmod{p}$, es gibt also ein $k \in \mathbb{Z}$ mit $a^p = kp + a$, womit $a(a^{p-1} - 1) = kp$ folgt. Damit teilt p das Produkt $a(a^{p-1} - 1)$. Da p eine Primzahl ist, die a nicht teilt, ist $\mathrm{ggT}(a, p) = 1$. Wir erhalten also mit Hilfe von Übungsaufgabe 2.1.15 $p | (a^{p-1} - 1)$, d.h. $(a^{p-1} - 1) \bmod p = 0$. Das liefert $a^{p-1} \bmod p = 1 \bmod p$, also die Behauptung. □

2.2. Die Mathematik hinter Public-Key-Verfahren der Kryptographie

Das Ausgangssituation der Kryptographie ist, dass jemand, der üblicherweise Bob genannt wird, jemand anderes, üblicherweise Alice, eine Nachricht zukommen lassen will, ohne dass diese von anderen Personen gelesen werden kann. Die

Inkarnation des Bösen, die versucht an die Nachricht zu gelangen, wird dabei üblicherweise Eve genannt.
Das Grunddilemma lautet: Bob könnte die Nachricht verschlüsseln, doch dazu müssen sich Bob und Alice zunächst über den Schlüssel einigen und wie macht man das so, dass Eve nicht die Kommunikation über den Schlüssel abfängt? Eine Lösung liefert die modulare Arithmetik; wie, das wollen wir hier kurz anhand des *RSA-Algorithmus* beschreiben. Dieser ist benannt nach den Entwicklern Roland Rivest, Adi Shamir und Leonard Adleman und er ist einer der grundlegenden Public-Key-Verfahren, d.h. Verschlüsselungsverfahren, bei denen Alice den Verschlüsselungsschlüssel einfach öffentlich zugänglich macht.
Die Stärke des Algorithmus steht und fällt damit, dass es kein effizientes Verfahren zum Zerlegen großer natürlicher Zahlen in ihre Primfaktoren gibt.
Der Algorithmus braucht drei Schritte, die einmalig zur Vorbereitung von Alice ausgeführt werden müssen:

1. Alice wählt zwei (große) Primzahlen p und q mit $p \neq q$ und berechnet $n = p \cdot q$ und $N = (p-1) \cdot (q-1)$.

2. Alice wählt ein $e \in \mathbb{N}$ mit $\mathrm{ggT}(e, N) = 1$ und bestimmt dann ein $x \in \mathbb{N}$ mit $ex \equiv 1 \pmod{N}$, vgl. Beispiel 2.1.14.

3. Alice schickt unverschlüsselt und frei zugänglich das Zahlenpaar (n, e) an Bob, das ist ihr sogenannter *Public Key*.

Verschlüsseln und Entschlüsseln einer Nachricht $M \in \mathbb{N}$ mit $M < n$ geht dann so:

Verschlüsseln: Bob rechnet $M' := M^e \bmod n$ und schickt das Ergebnis an Alice.

Entschlüsseln: Alice rechnet $M'' := (M')^x \bmod n$.

Zum Entschlüsseln verwendet Alice ihren sogenannten *Private Key* (n, x). Dieser ist nur ihr bekannt und um x aus dem public key (n, e) zu berechnen, bräuchte man N, d.h. p und q und damit die Primfaktorzerlegung von n.
Es bleibt uns noch zu zeigen, dass Alice auch wirklich Bobs Nachricht lesen kann, d.h. dass $M'' = M$ gilt.

Satz 2.2.1. *Mit obigen Bezeichnungen gilt* $M'' = M^{ex} \bmod n = M$ *für alle* $M < n$.

Beweis. Es ist nach Konstruktion $ex \equiv 1 \pmod{N}$, wobei $N = (p-1)(q-1)$ ist. Also gibt es ein $k \in \mathbb{N}$ mit $ex = 1 + k(p-1)(q-1)$, woraus

$$M^{ex} = M \cdot M^{(p-1)(q-1)k} = M \cdot (M^{p-1})^{(q-1)k}$$

folgt. Nun betrachten wir zwei Fälle: Ist $M \not\equiv 0 \pmod{p}$, so liefert der kleine Satz von Fermat, vgl. Korollar 2.1.17, $M^{p-1} \bmod p = 1$, und damit ist

$$M^{ex} \bmod p = (M \bmod p) \cdot (M^{p-1} \bmod p)^{(q-1)k} \bmod p = M \bmod p.$$

Ist dagegen M ein Vielfaches von p, so gilt ebenfalls $M^{ex} \bmod p = M \bmod p$, denn dann steht auf beiden Seiten der Gleichung Null.
Mit der selben Argumentation für q statt p bekommen wir auch

$$M^{ex} = M \cdot (M^{q-1})^{(p-1)k}$$

und damit

$$M^{ex} \bmod q = (M \bmod q) \cdot (M^{q-1} \bmod q)^{(p-1)k} \bmod q = M \bmod q.$$

Zusammengenommen gibt es also zwei Zahlen $k_1, k_2 \in \mathbb{N}$ mit $M^{ex} = M + k_1 p = M + k_2 q$, woraus insbesondere $k_1 p = k_2 q$ folgt. Nun sind aber p und q zwei verschiedene Primzahlen. Das bedeutet $p | k_2$ und wir bekommen noch ein $k_3 \in \mathbb{N}$ mit $k_2 = k_3 p$. Damit haben wir nun endgültig

$$M^{ex} \bmod n = (M + k_3 pq) \bmod n = (M + k_3 n) \bmod n = M \bmod n = M. \quad \square$$

2.3. Gruppen

In diesem Abschnitt beginnen wir mit der abstrakten Beschreibung des Rechnens. Wir beschränken uns dazu zunächst auf nur eine Rechenoperation. Diese kann ein Plus, ein Mal oder noch etwas anderes sein. Deshalb brauchen wir ein neues nicht mit Assoziationen beladenes Zeichen, als das wir im Folgenden meistens „$*$“ nehmen.

Definition 2.3.1. *Eine* Gruppe *ist eine Menge $G \neq \emptyset$ mit einer Abbildung (*Verknüpfung*) $* : G \times G \to G$, so dass gilt*

(a) *Für alle $a, b, c \in G$ gilt $a * (b * c) = (a * b) * c$. (Assoziativität)*

(n) *Es gibt ein $n \in G$, so dass für alle $a \in G$ gilt $n * a = a$ und $a * n = a$. (Existenz eines* neutralen Elements*)*

(i) *Zu jedem $a \in G$ gibt es ein $a^\sharp \in G$, so dass $a^\sharp * a = n$ und $a * a^\sharp = n$ gilt. (Existenz des* inversen Elements*)*

Gilt zusätzlich noch

(k) *Für alle $a, b \in G$ ist $a * b = b * a$ (Kommutativität),*

so heißt die Gruppe G abelsch*.*

Beispiel 2.3.2. (a) $\mathbb{Z}$ mit der üblichen Addition ist eine abelsche Gruppe, aber nicht $\mathbb{N}$.

(b) $\mathbb{Q}$ mit der üblichen Addition und $\mathbb{Q} \setminus \{0\}$ mit der üblichen Multiplikation sind abelsche Gruppen.

(c) Sei M eine beliebige nichtleere Menge und

$$F := \{f : M \to M \text{ bijektiv}\}.$$

Dann ist F mit der Verkettung „$\circ$“ als Verknüpfung eine Gruppe. Man nennt diese die *Permutationsgruppe* von M.

Man beachte, dass diese Gruppe i.A. nicht abelsch ist. Machen Sie sich das an dem Beispiel $M = \mathbb{R}$ und $f(x) = x^3$, $g(x) = x + 1$ klar.

Um einzusehen, dass $(F, \circ)$ eine Gruppe ist, müssen wir uns zunächst überlegen, dass $\circ$ eine vernünftige Verknüpfung auf F ist, d.h. dass für alle $f, g \in F$ auch $f \circ g \in F$, also bijektiv, ist. Das war aber gerade die Aussage von Übungsaufgabe 1.4.9.

Nun müssen wir noch (a), (n) und (i) zeigen. Zum Nachweis von (a) rechnen wir für drei Funktionen $f, g, h \in F$, dass für alle $x \in M$ gilt

$$\big[f\circ(g\circ h)\big](x) = f\big((g\circ h)(x)\big) = f\big(g(h(x))\big) = (f\circ g)(h(x)) = \big[(f\circ g)\circ h\big](x).$$

Das bedeutet aber gerade, dass $f \circ (g \circ h) = (f \circ g) \circ h$ ist.

Ein neutrales Element können wir explizit angeben, nämlich die Identität $\mathrm{id} : M \to M$ mit $\mathrm{id}(x) = x$ für alle $x \in M$. Es gilt nämlich für alle $f \in F$ und jedes $x \in M$

$$(\mathrm{id} \circ f)(x) = \mathrm{id}(f(x)) = f(x) \quad \text{und} \quad (f \circ \mathrm{id})(x) = f(\mathrm{id}(x)) = f(x).$$

Damit haben wir $\mathrm{id} \circ f = f$ und $f \circ \mathrm{id} = f$ und somit gezeigt, dass id ein neutrales Element in F ist.

Schließlich ist jedes $f \in F$ bijektiv, besitzt also eine Umkehrfunktion f^{-1}, die wiederum bijektiv, also ein Element von F ist. Für diese gilt $f \circ f^{-1} = \mathrm{id}$ und $f^{-1} \circ f = \mathrm{id}$, vgl. Satz 1.4.5, also ist jeweils f^{-1} das inverse Element zu f.

(d) Betrachtet man eine geometrische Figur in der Ebene und alle Spiegelungen und Drehungen der Ebene, die die Figur auf sich selbst abbilden, so erhält man die sogenannte *Symmetriegruppe* der Figur. Nimmt man beispielsweise ein Quadrat Q, vgl. Abbildung 2.1, so ist die Symmetriegruppe gerade gegeben durch

$$G = \big\{\text{Spiegelung an } a,\ \text{Spiegelung an } b,\ \text{Spiegelung an } c,\ \text{Spiegelung an } d,\\ \text{Drehung um } 90^\circ,\ \text{Drehung um } 180^\circ,\ \text{Drehung um } 270^\circ, \mathrm{id}\big\}$$

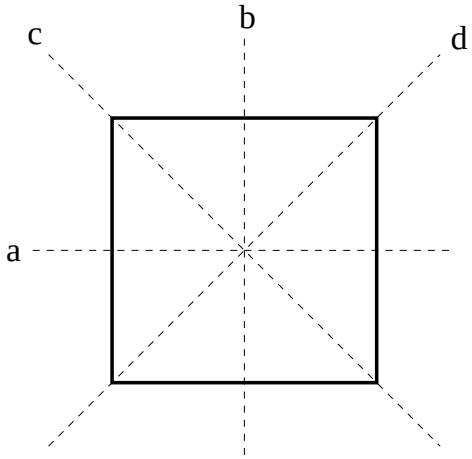

Abbildung 2.1.: Das Quadrat Q mit den Symmetrielinien a, b, c und d

Machen Sie sich klar, was die Verknüpfung in dieser Gruppe ist und dass es sich mit dieser tatsächlich um eine Gruppe handelt. Was ergibt die Spiegelung an a verknüpft mit der Spiegelung an b? Ist diese Gruppe abelsch?

(e) Zu $n \in \mathbb{N}^*$ betrachten wir wieder die Menge $\mathbb{Z}_n = \{\tilde{0}, \tilde{1}, \ldots, \widetilde{n-1}\}$ der Restklassen modulo n. Definieren wir für $\tilde{a}, \tilde{b} \in \mathbb{Z}_n$

$$\tilde{a} + \tilde{b} := \widetilde{a+b}$$

so ist $\mathbb{Z}_n$ mit diesem „+“ eine abelsche Gruppe.

Bevor wir den Nachweis der Axiome (a), (n), (i) und (k) als Übungsaufgabe stehen lassen können, sollte zumindest geklärt werden, dass obiges Plus eine vernünftige Verknüpfung ist, d.h. dass die Definition von den gewählten Repräsentanten a, bzw. b unabhängig ist.

Seien dazu $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ mit $\tilde{a}_1 = \tilde{a}_2$ und $\tilde{b}_1 = \tilde{b}_2$. Dann gilt $a_1 \equiv a_2 \pmod{n}$ und $b_1 \equiv b_2 \pmod{n}$ und wir haben mit Satz 2.1.5 auch

$$\begin{aligned}(a_1 + b_1) \bmod n &= (a_1 \bmod n + b_1 \bmod n) \bmod n \\ &= (a_2 \bmod n + b_2 \bmod n) \bmod n = (a_2 + b_2) \bmod n,\end{aligned}$$

oder anders ausgedrückt $\widetilde{a_1 + b_1} = \widetilde{a_2 + b_2}$.

Satz 2.3.3. *Sei $(G, *)$ eine Gruppe. Dann gilt*

(a) G enthält nur ein neutrales Element.

(b) Zu jedem $a \in G$ gibt es genau ein Inverses.

(c) Für gegebene $a, b, c, d \in G$ sind die Gleichungen $a * x = b$ und $x * c = d$ jeweils eindeutig lösbar.

(d) Für alle $g, h \in G$ gilt $(g * h)^\sharp = h^\sharp * g^\sharp$.

Beweis. (a) Seien $n_1, n_2 \in G$ neutrale Elemente. Dann gilt durch zweimalige Anwendung von (n)

$$n_1 = n_1 * n_2 = n_2.$$

(b) Sei $a \in G$. Die Existenz eines inversen Elements zu a garantiert uns (i), zu zeigen bleibt die Eindeutigkeit. Dazu seien b_1 und b_2 inverse Elemente von a. Dann gilt

$$b_1 \overset{(n)}{=} b_1 * n \overset{(i)}{=} b_1 * (a * b_2) \overset{(a)}{=} (b_1 * a) * b_2 \overset{(i)}{=} n * b_2 \overset{(n)}{=} b_2.$$

(c) Wir betrachten nur die erste Gleichung, das Argument für die zweite verläuft analog. Zum Nachweis der Existenz einer Lösung geben wir einfach eine an, nämlich $x = a^\sharp * b$, denn für dieses x gilt

$$a * x = a * (a^\sharp * b) \overset{(a)}{=} (a * a^\sharp) * b \overset{(i)}{=} n * b \overset{(n)}{=} b.$$

Um Eindeutigkeit zu zeigen, nehmen wir wieder zwei Lösungen x_1 und x_2 her. Dann gilt $a * x_1 = b = a * x_2$ und damit auch $a^\sharp * (a * x_1) = a^\sharp * (a * x_2)$. Mit (a) folgt daraus $(a^\sharp * a) * x_1 = (a^\sharp * a) * x_2$, was, (i) folgend, $n * x_1 = n * x_2$ bedeutet. Werfen wir nun noch Axiom (n) dazu, liefert das $x_1 = x_2$ und wir sind fertig.

(d) Seien $g, h \in G$. Dann gilt

$$(g * h) * (h^\sharp * g^\sharp) \overset{(a)}{=} g * (h * h^\sharp) * g^\sharp \overset{(i)}{=} g * n * g^\sharp \overset{(n)}{=} g * g^\sharp \overset{(i)}{=} n$$

und analog $(h^\sharp * g^\sharp) * (g * h) = n$. Also ist $h^\sharp * g^\sharp$ ein Inverses von $g * h$ und mit Teil (b) dieses Satzes folgt die Behauptung. □

Übungsaufgabe 2.3.4. Es sei $(G, *)$ eine Gruppe mit neutralem Element n. Zeigen Sie:

(a) Für alle $g \in G$ gilt $(g^\sharp)^\sharp = g$.

(b) Es ist $n^\sharp = n$.

(c) Ist G endlich, so gibt es ein $k \in \mathbb{N}$ mit $\underbrace{g * g * \cdots * g}_{k \text{ Mal}} = n$ für alle $g \in G$.

2.3.1. Untergruppen

Beispiel 2.3.5. Wir betrachten die Menge $2\mathbb{Z} = \{\dots, -6, -4, -2, 0, 2, 4, 6, \dots\}$. Dann ist $2\mathbb{Z}$ mit der üblichen Addition aus $\mathbb{Z}$ wieder eine Gruppe, denn für je 2 Elemente aus $2\mathbb{Z}$ ist deren Summe wieder in $2\mathbb{Z}$, das neutrale Element 0 ist in $2\mathbb{Z}$ und zu jedem $z \in 2\mathbb{Z}$ ist auch das inverse Element $-z \in 2\mathbb{Z}$.
Damit ist also $(2\mathbb{Z}, +)$ eine Teilmenge der Gruppe $(\mathbb{Z}, +)$, die selbst wieder eine Gruppe ist. Solche Phänomene gibt es sehr oft und wir wollen dies im Folgenden ein wenig untersuchen.

Definition 2.3.6. *Eine Teilmenge U einer Gruppe $(G, *)$ heißt* Untergruppe *von G, falls auch $(U, *)$ eine Gruppe ist.*

Beispiel 2.3.7. (a) Die Teilmengen G und $\{n\}$ sind Untergruppen einer jeden Gruppe G. Man nennt diese die *trivialen* Untergruppen von G.

(b) Erinnern wir uns noch einmal an die Symmetriegruppe des Quadrates aus Beispiel 2.3.2 (d), so hat diese neben den trivialen Untergruppen noch die Untergruppe, die aus den drei Drehungen zusammen mit der Identität, die man auch als Drehung um 0° auffassen kann, besteht. Indem Sie sich das klar machen, können Sie schon einiges Gefühl für Gruppen gewinnen.

Bilden auch die Spiegelungen eine Untergruppe?

Satz 2.3.8 (Untergruppenkriterium)**.** *Eine Teilmenge U einer Gruppe $(G, *)$ ist genau dann eine Untergruppe von G, wenn*

(UG1) *$U \neq \emptyset$ und*

(UG2) *für alle $a, b \in U$ ist auch $a * b^\sharp \in U$.*

Beweis. **„⇒“** Ist U eine Untergruppe, so muss U eine Gruppe sein und damit zumindest ein neutrales Element enthalten, also ist U nicht leer und wir haben (UG1).

Wir zeigen als nächstes, dass das neutrale Element von U, das wir mit n_U bezeichnen, gleich dem neutralen Element n_G von G ist. Dazu bezeichne ${n_U}^\sharp$ das Inverse zu n_U in G. Dann gilt $n_U = n_U * n_U$, also

$$n_G = n_U * {n_U}^\sharp = (n_U * n_U) * {n_U}^\sharp = n_U * (n_U * {n_U}^\sharp) = n_U * n_G = n_U.$$

Seien nun $a, b \in U$ und sei $b^\sharp$ das inverse Element von b in G. Da U eine Gruppe ist, hat b auch ein inverses Element $\hat{b}$ in U. Sind die beiden wirklich verschieden? Nein, denn wegen $b * \hat{b} = \hat{b} * b = n_U = n_G$ ist $\hat{b}$ auch ein Inverses von b in G und dieses ist in der Gruppe G eindeutig, vgl. Satz 2.3.3 (b). Also gilt $b^\sharp = \hat{b} \in U$. Da U eine Gruppe ist, muss schlussendlich mit a und $b^\sharp$ auch $a * b^\sharp \in U$ sein und wir sind fertig.

„⇐“ Sei $U \subseteq G$ so, dass (UG1) und (UG2) gelten. Wir müssen zeigen, dass U dann eine Gruppe ist, d.h. dass $* : U \times U \to U$ gilt, und dass die Axiome (a), (n) und (i) erfüllt sind.

Zunächst ist $*$ auf U assoziativ, da dies auf G gilt. Wir haben also (a).

Weiter gibt es wegen (UG1) auf jeden Fall irgendein $a \in U$. Wegen (UG2) ist dann auch $a * a^\sharp = n \in U$. Nun ist jedes Element $b \in U$ auch in G und dort gilt $n * b = b$ und $b * n = b$, also gilt das auch in U und wir haben (n) gezeigt.

Zum Nachweis von (i) sei nun $a \in U$ gegeben. Da nach obigen Überlegungen das neutrale Element n von G ebenfalls in U liegen muss, gilt wiederum nach (UG2) nun $n * a^\sharp = a^\sharp \in U$.

Es bleibt zu zeigen, dass $*$ eine vernünftige Verknüpfung auf U ist, die Elemente aus U zu Elementen aus U verknüpft. Seien dazu $a, b \in U$. Wie wir oben schon gezeigt haben, ist dann auch $b^\sharp \in U$ und wegen (UG2) und dem Resultat von Übungsaufgabe 2.3.4 (a) haben wir $a * (b^\sharp)^\sharp = a * b \in U$. □

Lemma 2.3.9. *Sei $(G, *)$ eine Gruppe, I eine beliebige Indexmenge und U_j sei für jedes $j \in I$ eine Untergruppe von G. Dann ist auch der Schnitt all dieser Untergruppen, d.h.*

$$\bigcap_{j \in I} U_j = \big\{x \in G : x \in U_j \text{ für jedes } j \in I\big\},$$

eine Untergruppe von G.

Beweis. Wir wenden das Untergruppenkriterium an. Da alle U_j Untergruppen sind, muss jede dieser Gruppen das neutrale Element n von G enthalten, vgl. den Beweis von Satz 2.3.8, also ist dieses auch im Schnitt aller U_j, $j \in I$, enthalten und wir haben $\bigcap_{j \in I} U_j \neq \emptyset$.
Zum Nachweis von (UG2) seien $a, b \in \bigcap_{j \in I} U_j$. Das bedeutet, dass diese beiden für jedes $j \in I$ in der Untergruppe U_j enthalten sind. Dann liefert aber Satz 2.3.8 sofort $a * b^\sharp \in U_j$ und zwar für jedes $j \in I$. Also haben wir auch $a * b^\sharp \in \bigcap_{j \in I} U_j$ und sind fertig. □

Definition 2.3.10. *Sei G eine Gruppe und $M \subseteq G$. Dann heißt*

$$\langle M \rangle := \bigcap_{\substack{U \text{ Untergruppe von } G \\ U \supseteq M}} U$$

Erzeugnis *von M oder die von M* erzeugte Untergruppe.

Bemerkung 2.3.11. Man beachte, dass nach Lemma 2.3.9 das Erzeugnis $\langle M \rangle$ immer eine Untergruppe von G ist. Tatsächlich ist $\langle M \rangle$ die kleinste Untergruppe von G, in der M ganz enthalten ist.
Insbesondere gilt $M = \langle M \rangle \iff M$ Untergruppe von G.

Beispiel 2.3.12. Betrachten wir in der Gruppe $(\mathbb{Z}, +)$ die Teilmenge $M = \{2\}$, so gilt $\langle M \rangle = 2\mathbb{Z}$, denn zum Einen müssen natürlich die Zahlen 2, -2, 0, $2+2$, $-2-2$, $2+2+2$, $-2-2-2$, ... drin sein, d.h. $2\mathbb{Z} \subseteq \langle M \rangle$. Zum Anderen ist $2\mathbb{Z}$ eine Untergruppe von $\mathbb{Z}$, vgl. Beispiel 2.3.5, also haben wir auch $\langle M \rangle \subseteq 2\mathbb{Z}$ und damit Gleichheit.

Zur Verkürzung der Notation führen wir noch die folgende Schreibweise für ein Gruppenelement g einer Gruppe $(G, *)$ und eine Zahl $k \in \mathbb{Z}$ ein:

$$g^k := \begin{cases} \underbrace{g * g * g * \cdots * g}_{k \text{ Mal}}, & \text{falls } k > 0, \\ n, & \text{falls } k = 0, \\ \underbrace{g^\sharp * g^\sharp * g^\sharp * \cdots * g^\sharp}_{k \text{ Mal}}, & \text{falls } k < 0. \end{cases}$$

Übungsaufgabe 2.3.13. (a) Bestimmen Sie $\langle \{3, 6\} \rangle$ und $\langle \{3, 2\} \rangle$ in $(\mathbb{Z}, +)$.

(b) Zeigen Sie: Ist $(G, *)$ Gruppe und $g \in G$, so gilt $\langle \{g\} \rangle = \{g^k : k \in \mathbb{Z}\}$.

2.3.2. Gruppenhomomorphismen

Definition 2.3.14. *(a) Es seien $(G, *)$ und $(H, \diamond)$ Gruppen. Eine Abbildung $f : G \to H$ heißt* (Gruppen-)Homomorphismus*, falls*

$$f(g_1 * g_2) = f(g_1) \diamond f(g_2) \quad \textit{für alle } g_1, g_2 \in G$$

gilt.

(b) Ein bijektiver Gruppenhomomorphismus heißt (Gruppen-)Isomorphismus*.*

(c) Zwei Gruppen G und H, für die ein Isomorphismus $f : G \to H$ existiert, heißen isomorph*.*

Beispiel 2.3.15. (a) Die Abbildung

$$f : \begin{cases} (\mathbb{Z}, +) & \to (\mathbb{Z}, +) \\ k & \mapsto 4k \end{cases}$$

ist ein Homomorphismus, denn für alle $k, \ell \in \mathbb{Z}$ gilt

$$f(k + \ell) = 4(k + \ell) = 4k + 4\ell = f(k) + f(\ell).$$

Allerdings ist f kein Isomorphismus, denn f ist nicht surjektiv.

Zeigen Sie, dass $f : (\mathbb{Z}, +) \to (4\mathbb{Z}, +)$ ein Isomorphismus ist.

(b) Die Abbildung

$$g : \begin{cases} (\mathbb{R}, +) & \to (\mathbb{R} \setminus \{0\}, \cdot) \\ x & \mapsto 2^x \end{cases}$$

ist ebenfalls ein Homomorphismus, denn für alle $x_1, x_2 \in \mathbb{R}$ gilt

$$g(x_1 + x_2) = 2^{x_1+x_2} = 2^{x_1} \cdot 2^{x_2} = g(x_1) \cdot g(x_2).$$

Ist g ein Isomorphismus?

Übungsaufgabe 2.3.16. Zeigen Sie: Ist $(G, *)$ eine Gruppe und $g \in G$ ein beliebiges Gruppenelement, so ist

$$\varphi_g : \begin{cases} G & \to G \\ h & \mapsto g^\sharp * h * g \end{cases}$$

ein Homomorphismus.
Finden Sie weiter Beispiele von Gruppen G und Elementen $g \in G$, so dass φ_g einmal ein Isomorphismus ist und einmal nicht.

Satz 2.3.17. *Es seien $(G, *)$ und $(H, \diamond)$ Gruppen mit neutralen Elementen n_G bzw. n_H und $f : G \to H$ ein Homomorphismus. Dann gilt*

(a) $f(n_G) = n_H$.

(b) $f(g)^\sharp = f(g^\sharp)$ für jedes $g \in G$.

(c) $f(G)$ ist eine Gruppe, d.h. eine Untergruppe von H.

(d) Ist G abelsch, so ist auch $f(G)$ abelsch.

Beweis. (a) Dank (n) haben wir $n_G = n_G * n_G$. Also ist wegen der Homomorphieeigenschaft von f auch $f(n_G) = f(n_G * n_G) = f(n_G) \diamond f(n_G)$. Weiter hat $f(n_G)$ wie jedes Gruppenelement wegen (i) ein Inverses $f(n_G)^\sharp$ in H. Damit gilt

$$\begin{aligned} n_H &\overset{\text{(i)}}{=} f(n_G)^\sharp \diamond f(n_G) = f(n_G)^\sharp \diamond \big(f(n_G) \diamond f(n_G)\big) \\ &\overset{\text{(a)}}{=} \big(f(n_G)^\sharp \diamond f(n_G)\big) \diamond f(n_G) \overset{\text{(i)}}{=} n_H \diamond f(n_G) \overset{\text{(n)}}{=} f(n_G). \end{aligned}$$

(b) Sei $g \in G$. Dann gilt mit der Homomorphieeigenschaft von f und Teil (a) des Beweises.

$$f(g) \diamond f(g^\sharp) = f(g * g^\sharp) = f(n_G) = n_H,$$

sowie

$$f(g^\sharp) \diamond f(g) = f(g^\sharp * g) = f(n_G) = n_H.$$

Also ist $f(g^\sharp) = f(g)^\sharp$.

(c) Wir wenden das Untergruppenkriterium an. Wegen Teil (a) des Beweises gilt $n_H = f(n_G) \in f(G)$, also ist $f(G) \neq \emptyset$ und wir haben (UG1). Zum Nachweis von (UG2) seien $h_1, h_2 \in f(G)$ gegeben. Dann gibt es $g_1, g_2 \in G$ mit $f(g_1) = h_1$ und $f(g_2) = h_2$. Mit diesen haben wir dank Teil (b)

$$h_1 \diamond {h_2}^\sharp = f(g_1) \diamond f(g_2)^\sharp = f(g_1) \diamond f({g_2}^\sharp) = f(g_1 * {g_2}^\sharp) \in f(G).$$

(d) Sei nun G abelsch und seien $h_1, h_2 \in f(G)$. Dann gibt es wieder $g_1, g_2 \in G$ mit $f(g_1) = h_1$ und $f(g_2) = h_2$ und wir bekommen

$$h_1 \diamond h_2 = f(g_1) \diamond f(g_2) = f(g_1 * g_2) \overset{(k)}{=} f(g_2 * g_1) = f(g_2) \diamond f(g_1) = h_2 \diamond h_1.$$

□

Definition 2.3.18. *Es seien $(G, *)$ und $(H, \diamond)$ Gruppen mit neutralen Elementen n_G bzw. n_H und $f : G \to H$ ein Homomorphismus. Dann heißt* $\ker(f) := \{g \in G : f(g) = n_H\}$ Kern *von f.*

Beispiel 2.3.19. Wir betrachten

$$f : \begin{cases} (\mathbb{Z}_4, +) & \to (\mathbb{Z}_4, +) \\ \tilde{n} & \mapsto \widetilde{2n}. \end{cases}$$

Diese Abbildung ist wegen

$$f(\tilde{n}_1 + \tilde{n}_2) = f(\widetilde{n_1 + n_2}) = 2(\widetilde{n_1 + n_2}) = \widetilde{2n_1} + \widetilde{2n_2} = f(\tilde{n}_1) + f(\tilde{n}_2)$$

ein Homomorphismus mit

$$f(\tilde{0}) = \tilde{0}, \quad f(\tilde{1}) = \tilde{2}, \quad f(\tilde{2}) = \tilde{4} = \tilde{0}, \quad f(\tilde{3}) = \tilde{6} = \tilde{2}.$$

In diesem Fall ist also $\ker(f) = \{\tilde{0}, \tilde{2}\}$.

Satz 2.3.20. *Die Menge* $\ker(f)$ *ist immer eine Untergruppe von G.*

Beweis. Es ist immer $f(n_G) = n_H$, vgl. Satz 2.3.17 (a), also ist $n_G \in \ker(f)$ und damit $\ker(f) \neq \emptyset$. Seien nun $g_1, g_2 \in \ker(f)$. Dann gilt

$$f(g_1 * {g_2}^\sharp) = f(g_1) \diamond f({g_2}^\sharp) = f(g_1) \diamond f(g_2)^\sharp = n_H \diamond {n_H}^\sharp = n_H \diamond n_H = n_H.$$

Also ist auch $g_1 * {g_2}^\sharp \in \ker(f)$ und die Behauptung folgt aus dem Untergruppenkriterium. □

Insbesondere haben wir damit gesehen, dass $\{\tilde{0}, \tilde{2}\}$ eine Untergruppe von $(\mathbb{Z}_4, +)$ ist.

2.4. Ringe und Körper

2.4.1. Ringe

Definition 2.4.1. *(a) Eine Menge R mit zwei Verknüpfungen $+ : R \times R \to R$ und $\cdot : R \times R \to R$ heißt* Ring*, falls die folgenden Bedingungen erfüllt sind:*

- *$(R, +)$ ist eine abelsche Gruppe.*
- *$\forall a, b, c \in R : a \cdot (b \cdot c) = (a \cdot b) \cdot c$, d.h. „$\cdot$“ ist assoziativ.*
- *$\forall a, b, c \in R : a \cdot (b + c) = a \cdot b + a \cdot c$ und $(a + b) \cdot c = a \cdot c + b \cdot c$, d.h. die beiden Verknüpfungen erfüllen die Distributivgesetze.*

(b) Das neutrale Element der Gruppe $(R, +)$ heißt Nullelement*, Symbol:* 0.

(c) Existiert ein Element $1 \in R$ mit $a \cdot 1 = 1 \cdot a = a$ für jedes $a \in R$, so heißt 1 Einselement *von R und man nennt dann R einen* Ring mit Eins.

(d) Ist zusätzlich die Verknüpfung „$\cdot$“ auf R kommutativ, so nennt man R einen kommutativen Ring.

Beispiel 2.4.2. (a) $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$ sind kommutative Ringe mit Eins.

(b) $\mathbb{R}[x]$, d.h. die Menge aller Polynome in einer Variablen über $\mathbb{R}$, ist ein kommutativer Ring mit dem Einselement, das durch das konstante Polynom 1 gegeben ist.

(c) Sei $n \in \mathbb{N}$ mit $n \geq 2$. Definieren wir auf $\mathbb{Z}_n$ eine Multiplikation durch $\tilde{a} \cdot \tilde{b} = \widetilde{a \cdot b}$, so ist diese nach Satz 2.1.5 (b) wohldefiniert.

Wir wollen nun zeigen, dass $(\mathbb{Z}_n, +, \cdot)$ sogar ein kommutativer Ring mit Eins ist. Dazu erinnern wir uns zunächst, dass wir in Beispiel 2.3.2 (e) bereits festgestellt haben, dass $(\mathbb{Z}_n, +)$ eine abelsche Gruppe ist. Die Assoziativität und die Kommutativität von „$\cdot$“, sowie die Distributivgesetze können wir leicht auf die entsprechenden Eigenschaften von $\mathbb{Z}$ zurückspielen. Hier führen wir beispielhaft nur das Assoziativgesetz vor. Für alle $\tilde{k}, \tilde{\ell}, \tilde{m} \in \mathbb{Z}_n$ gilt

$$(\tilde{k} \cdot \tilde{\ell}) \cdot \tilde{m} = (\widetilde{k \cdot \ell}) \cdot \tilde{m} = \widetilde{(k \cdot \ell) \cdot m} = k \cdot \widetilde{(\ell \cdot m)} = \tilde{k} \cdot (\tilde{\ell} \cdot \tilde{m}).$$

Schließlich bleibt uns noch das Einselement zu identifizieren, aber auch das ist nicht sonderlich schwer, denn für alle $\tilde{k} \in \mathbb{Z}_n$ gilt $\tilde{1} \cdot \tilde{k} = \widetilde{1 \cdot k} = \tilde{k}$, also ist $\tilde{1}$ das Einselement von $(\mathbb{Z}_n, +, \cdot)$.

Was bei der Multiplikation in $\mathbb{Z}_n$ für konkrete (kleine) Werte von n passiert, kann man sich gut mit *Multiplikationstafeln* klar machen. Hier sind die für

$n = 5$

	$\tilde{0}$	$\tilde{1}$	$\tilde{2}$	$\tilde{3}$	$\tilde{4}$
$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$
$\tilde{1}$	$\tilde{0}$	$\tilde{1}$	$\tilde{2}$	$\tilde{3}$	$\tilde{4}$
$\tilde{2}$	$\tilde{0}$	$\tilde{2}$	$\tilde{4}$	$\tilde{1}$	$\tilde{3}$
$\tilde{3}$	$\tilde{0}$	$\tilde{3}$	$\tilde{1}$	$\tilde{4}$	$\tilde{2}$
$\tilde{4}$	$\tilde{0}$	$\tilde{4}$	$\tilde{3}$	$\tilde{2}$	$\tilde{1}$

$n = 4$

	$\tilde{0}$	$\tilde{1}$	$\tilde{2}$	$\tilde{3}$
$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$
$\tilde{1}$	$\tilde{0}$	$\tilde{1}$	$\tilde{2}$	$\tilde{3}$
$\tilde{2}$	$\tilde{0}$	$\tilde{2}$	$\tilde{0}$	$\tilde{2}$
$\tilde{3}$	$\tilde{0}$	$\tilde{3}$	$\tilde{2}$	$\tilde{1}$

Schreibweise 2.4.3. *Sei $(R, +, \cdot)$ ein Ring.*

(a) Das zu $r \in R$ additiv inverse Element bezeichnet man mit $-r$ und für $r, s \in R$ schreibt man $r - s$ statt $r + (-s)$.

(b) Oft lässt man das „$\cdot$“ weg und schreibt rs statt $r \cdot s$ für $r, s \in R$.

Satz 2.4.4. *Sei $(R, +, \cdot)$ ein Ring. Dann gelten die folgenden Aussagen:*

(a) Für jedes $r \in R$ gilt $0 \cdot r = r \cdot 0 = 0$.

(b) Für alle $r, s \in R$ gilt $(-r) \cdot s = r \cdot (-s) = -(r \cdot s)$ und $(-r) \cdot (-s) = rs$.

(c) Für jede Wahl von $r, s, t \in R$ gilt $r(s - t) = rs - rt$.

Beweis. (a) Sei $r \in R$. Dann gilt wegen (n) und dank des Distributivgesetzes $r \cdot 0 = r \cdot (0 + 0) = r \cdot 0 + r \cdot 0$. Da $(R, +)$ eine Gruppe ist, besitzt $r \cdot 0$ ein additives Inverses $-(r \cdot 0)$. Mit diesem gilt

$$0 = r \cdot 0 - r \cdot 0 = (r \cdot 0 + r \cdot 0) - r \cdot 0 = r \cdot 0 + (r \cdot 0 - r \cdot 0) = r \cdot 0.$$

(b), (c) Übung. □

Analog zur Situation bei Gruppen definieren wir Homomorphismen und Isomorphismen von Ringen, als die Abbildungen, die die beiden Verknüpfungen von Ringen respektieren.

Definition 2.4.5. *Seien $(R, +, \cdot)$, $(S, \oplus, \odot)$ Ringe.*

(a) Eine Abbildung $f : R \to S$ heißt (Ring-)Homomorphismus*, falls für alle $r, s \in R$ gilt*

$$f(r + s) = f(r) \oplus f(s) \qquad \text{und} \qquad f(r \cdot s) = f(r) \odot f(s). \tag{2.2}$$

(b) Sind R und S Ringe mit Eins und sind 1_R und 1_S die beiden Einselelemente, so fordert man zusätzlich zu (2.2)

$$f(1_R) = 1_S.$$

(c) Einen bijektiven Ringhomomorphismus nennt man (Ring-)Isomorphismus *und sagt in diesem Fall, dass die beiden Ringe R und S* isomorph *sind.*

Bemerkung 2.4.6. Wie bei Gruppen gilt auch für Ringe, dass das Bild eines Rings unter einem Ringhomomorphismus immer wieder ein Ring ist.

2.4.2. Körper

Die letzte klassische Rechenart, die wir jetzt noch nicht untersucht haben, ist das Teilen. Zum Teilen brauchen wir auf jeden Fall eine Eins, wir sollten also mit einem Ring mit Eins R starten und uns überlegen, was wir weiterhin zum Teilen brauchen. Teilen durch ein $r \in R$ bedeutet Multiplizieren mit $1/r$, aber was ist das, $1/r$? Das ist ein Element, das, wenn wir es mit r multiplizieren das Einselement ergibt. Wegen Satz 2.4.4 (a) heißt das von vornherein, dass wir Teilen durch Null komplett vergessen können. Um ansonsten freizügig teilen zu können, fordern wir also, dass es zu jedem $r \in R \setminus \{0\}$ ein Element $r^{-1} \in R$ gibt mit $r \cdot r^{-1} = r^{-1} \cdot r = 1$.
Nun erhebt sich natürlich die Frage: Gibt es solche Ringe? Ja, z.B. $\mathbb{Q}$ und $\mathbb{R}$ sehen gut aus. Wie ist es mit unseren anderen Beispielen?

$\mathbb{Z}$: Nein, sicher nicht, denn es gibt kein $k \in \mathbb{Z}$ mit $k \cdot 2 = 1$.

$\mathbb{R}[x]$: Ebensowenig, denn für welches Polynom P gilt $P(x) \cdot x^2 = 1$?

$\mathbb{Z}_n$: Hier ist die Sache weniger klar und hängt von n ab (wie genau werden wir weiter unten sehen). Wir betrachten die Beispiele $n = 4$ und $n = 5$, vgl. Beispiel 2.4.2 (c).

Für $n = 4$ gilt $\tilde{2} \cdot \tilde{2} = \tilde{4} = \tilde{0}$, was schon mal befremdlich aussieht. Nehmen wir nun an, es gäbe ein Element $\tilde{2}^{-1} \in \mathbb{Z}_4$, also ein $n \in \{0, 1, 2, 3\}$ mit $\tilde{n} = \tilde{2}^{-1}$, so folgt

$$\tilde{0} = \widetilde{0 \cdot n} = \tilde{0} \cdot \tilde{n} = \tilde{2} \cdot \tilde{2} \cdot \tilde{2}^{-1} = \tilde{2},$$

was ein sauberer Widerspruch ist, also kann man in $\mathbb{Z}_4$ nicht durch $\tilde{2}$ teilen.

In $\mathbb{Z}_5$ sieht das schon anders aus. Aus der Multiplikationstabelle in Beispiel 2.4.2 (c) liest man ab:

$$\tilde{1}^{-1} = \tilde{1}, \quad \tilde{2}^{-1} = \tilde{3}, \quad \tilde{3}^{-1} = \tilde{2}, \quad \tilde{4}^{-1} = \tilde{4}.$$

Dem befremdlichen Verhalten von $\mathbb{Z}_4$ oben geben wir zunächst einen Namen.

Definition 2.4.7. *Sei $(R, +, \cdot)$ ein Ring. Gibt es Zahlen $r, s \in R \setminus \{0\}$ mit $rs = 0$, so heißt r ein* linker *und s ein* rechter Nullteiler.

Wir wollen nun den besonders schönen Ringen, in denen wir durch alles außer der Null teilen können, einen eigenen Namen geben.

Definition 2.4.8. *Ein kommutativer Ring mit Eins K, in dem zusätzlich $(K \setminus \{0\}, \cdot)$ eine abelsche Gruppe ist, heißt* Körper.

Beispiele von Körpern sind die rationalen Zahlen $\mathbb{Q}$, sowie die reellen Zahlen $\mathbb{R}$ und wie wir oben gesehen haben $\mathbb{Z}_5$.

Bemerkung 2.4.9. Da es in der Defintion des Körpers etwas versteckt ist, sei an dieser Stelle explizit darauf hingewiesen, dass in jedem Körper $1 \neq 0$ gelten muss, denn die Eins ist das neutrale Element der Gruppe $(K \setminus \{0\}, \cdot)$ und kann damit nicht Null sein.

Wir wollen nun zeigen, dass es in Körpern keine Nullteiler geben kann.

Satz 2.4.10. *Ist K ein Körper, so gilt für alle $x, y \in K$*

$$x \cdot y = 0 \Longrightarrow x = 0 \text{ oder } y = 0.$$

Beweis. Seien $x, y \in K$ mit $x \cdot y = 0$. Ist $x = 0$ sind wir fertig, sei also $x \neq 0$. Dann gibt es das multiplikative Inverse $x^{-1} \in K$ und wir haben mit Unterstützung von Satz 2.4.4 (a)

$$y = 1 \cdot y = (x^{-1} \cdot x) \cdot y = x^{-1} \cdot (x \cdot y) = x^{-1} \cdot 0 = 0.$$

□

Satz 2.4.11. *Der Ring $(\mathbb{Z}_n, +, \cdot)$ ist genau dann ein Körper, wenn n prim ist.*

Beweis. **„⇒“** Sei n nicht prim. Dann gibt es $p, r \in \{2, 3, \ldots, n-1\}$ mit $n = pr$. Das bedeutet aber $\tilde{p} \cdot \tilde{r} = \widetilde{pr} = \tilde{n} = \tilde{0}$ und da $\tilde{p}$ und $\tilde{r}$ beide nicht gleich $\tilde{0}$ sind, hat $\mathbb{Z}_n$ also Nullteiler und kann kein Körper sein.

„⇐“ Zum Nachweis der Rückrichtung beobachten wir zunächst, dass $(\mathbb{Z}_n, +, \cdot)$ nach Beispiel 2.4.2 (c) ein kommutativer Ring mit Eins ist, es bleibt also nur zu zeigen, dass jedes Element von $\mathbb{Z}_n$, das nicht Null ist, ein multiplikatives Inverses besitzt. Sei also $a \in \{1, 2, \ldots, n-1\}$ gegeben.

Da n prim ist und $a < n$ gilt, bekommen wir aus dem kleinen Satz von Fermat, vgl. Korollar 2.1.17, dass $a^{n-1} \equiv 1 \pmod{n}$ ist. Das bedeutet in $\mathbb{Z}_n$ gilt $\widetilde{a^{n-1}} = \tilde{1}$. Betrachten wir nun das Element $\tilde{b} := \widetilde{a^{n-2}} \in \mathbb{Z}_n$, so gilt

$$\tilde{a} \cdot \tilde{b} = \tilde{a} \cdot \widetilde{a^{n-2}} = \widetilde{a \cdot a^{n-2}} = \widetilde{a^{n-1}} = \tilde{1}.$$

Also ist $\tilde{a}^{-1} = \tilde{b} = \widetilde{a^{n-2}}$ das gesuchte inverse Element und wir sind fertig. □

Definition 2.4.12. *Seien $(K, +, \cdot)$ und $(L, \oplus, \odot)$ Körper mit Einselementen 1_K und 1_L.*

(a) Ein Ringhomomorphismus $f : K \to L$ (mit $f(1_K) = 1_L$, vgl. Definition 2.4.5 (b)) heißt (Körper-)Homomorphismus.

(b) Ist f zusätzlich bijektiv, so heißt f (Körper-)Isomorphismus*, und man nennt dann K und L* isomorph.

(c) *Ist schließlich $f : K \to K$ ein Isomorphismus, so nennt man f einen* (Körper-)Automorphismus *von K.*

Bemerkung 2.4.13. Wie bei Gruppen und Ringen gilt auch hier, dass für jeden Körperhomomorphismus $f : K \to L$ die Menge $f(K)$ ein Körper ist.

Beispiel 2.4.14. (a) Jeder Körper K hat den Automorphismus $\mathrm{id} : K \to K$, dies ist der sogenannte *triviale* Körperautomorphismus.

(b) Ein Körperhomomorphismus ist z.B. $\mathrm{id} : \mathbb{Q} \to \mathbb{R}$.

Sind $(K, +, \cdot)$ und $(L, \oplus, \odot)$ Körper und $f : K \to L$ ein Homomorphismus, so ist wenig überraschend, dass $f(0_K) = 0_L$ und $f(1_K) = 1_L$ gilt, denn f ist ja insbesondere auch jeweils ein Gruppenhomomorphismus von $(K, +)$ nach $(L, \oplus)$, bzw. von $(K \setminus \{0\}, \cdot)$ nach $(L \setminus \{0\}, \odot)$. Auf den ersten Blick weniger zu erwarten ist folgendes Resultat.

Satz 2.4.15. *Jeder Körperhomomorphismus $f : K \to L$ ist injektiv.*

Beweis. Wir zeigen zunächst, dass $f^{-1}(\{0_L\}) = \{0_K\}$ ist, d.h. nur das Nullelement von K wird auf das Nullelement von L abgebildet. Dazu nehmen wir an, es gäbe ein $x \in K$ mit $x \neq 0_K$ und $f(x) = 0_L$. Wegen $x \neq 0_K$ gibt es dann $x^{-1} \in K$ mit $x \cdot x^{-1} = 1_K$. Also ist

$$1_L = f(1_K) = f(x \cdot x^{-1}) = f(x) \odot f(x^{-1}) = 0_L \odot f(x^{-1}) = 0_L$$

und das ist in einem Körper nicht möglich.
Seien nun $x_1, x_2 \in K$ mit $f(x_1) = f(x_2)$ gegeben. Dann gilt

$$\begin{aligned} f(x_1 - x_2) &= f\big(x_1 + (-x_2)\big) = f(x_1) \oplus f(-x_2) = f(x_1) \oplus (-f(x_2)) \\ &= f(x_1) \ominus f(x_2) = 0_L. \end{aligned}$$

Also muss nach obigen Erkenntnissen $x_1 - x_2 = 0_K$ und damit $x_1 = x_2$ sein. Das bedeutet aber gerade, dass f injektiv ist. □

Definition 2.4.16. *Ist $(K, +, \cdot)$ ein Körper, auf dem eine Totalordnung „$\leq$“ gegeben ist, so dass*

- $\forall a, b, c \in K : a \leq b \Longrightarrow a + c \leq b + c$ *und*
- $\forall a, b, c \in K : (a \leq b$ *und* $0_K \leq c) \Longrightarrow ac \leq bc$

gelten, so heißt $(K, +, \cdot, \leq)$ angeordneter Körper.

Ein Paradebeispiel für einen angeordneten Körper ist $\mathbb{Q}$.

Übungsaufgabe 2.4.17. Ist $(K, +, \cdot, \leq)$ ein angeordneter Körper, so gilt

(a) Für alle $a \in K$ mit $a > 0$ gilt $-a < 0$.

(b) Für alle $a \in K$ gilt $a^2 \geq 0$.

Wir können nun den Begriff des angeordneten Körpers verwenden, um den Körper der reellen Zahlen zu definieren.

Definition 2.4.18. *Ein angeordneter Körper, der das* Vollständigkeitsaxiom*:*

Jede Teilmenge, die eine obere Schranke besitzt, besitzt auch ein Supremum.

erfüllt, heißt Körper der reellen Zahlen.

Natürlich ist diese Definition erst dann sinnvoll, wenn sie keinen inneren Widerspruch enthält, d.h. die Konstruktion eines solchen Körpers überhaupt möglich ist und außerdem müssten wir noch zeigen, dass es (bis auf Isomorphismen von Körpern) nur einen solchen Körper gibt.

2.5. Der Körper der komplexen Zahlen

Definition 2.5.1. *Wir definieren auf der Menge* $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ *eine Addition* $\oplus$ *und eine Multiplikation* $\odot$*, indem wir für* (x_1, y_1) *und* $(x_2, y_2) \in \mathbb{R}^2$ *setzen:*

$$\begin{aligned}(x_1, y_1) \oplus (x_2, y_2) &= (x_1 + x_2, y_1 + y_2) \quad \textit{und} \\ (x_1, y_1) \odot (x_2, y_2) &= (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2).\end{aligned}$$

Satz 2.5.2. $(\mathbb{R}^2, \oplus, \odot)$ *ist ein Körper.*

Beweis. Zunächst ist festzustellen, dass $\oplus$ und $\odot$ wohldefinierte Verknüpfungen sind, da sie jeweils zwei Elementen von $\mathbb{R}^2$ wieder Elemente von $\mathbb{R}^2$ zuordnen. Wir wenden uns also dem Nachweis zu, dass $(\mathbb{R}^2, \oplus)$ eine abelsche Gruppe ist. Für alle $(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{R}^2$ gilt dank der Assoziativität bzw. Kommutativität von $\mathbb{R}$

$$\begin{aligned}\big((x_1, y_1) \oplus (x_2, y_2)\big) \oplus (x_3, y_3) &= \big((x_1 + x_2) + x_3, (y_1 + y_2) + y_3\big) \\ &= \big(x_1 + (x_2 + x_3), y_1 + (y_2 + y_3)\big) \\ &= (x_1, y_1) \oplus \big((x_2, y_2) \oplus (x_3, y_3)\big)\end{aligned}$$

und

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1 + x_2, y_1 + y_2) = (x_2 + x_1, y_2 + y_1) = (x_2, y_2) \oplus (x_1, y_1).$$

Weiterhin ist $(0, 0)$ das additive neutrale Element, denn für jedes $(x, y) \in \mathbb{R}^2$ gilt

$$(x, y) \oplus (0, 0) = (x + 0, y + 0) = (x, y).$$

Schließlich ist das zu $(x, y) \in \mathbb{R}^2$ additiv inverse Element gegeben durch $(-x, -y)$, denn $(x, y) \oplus (-x, -y) = (x - x, y - y) = (0, 0)$.
Die nächste Etappe ist der Nachweis, dass $(\mathbb{R}^2 \setminus \{(0, 0)\}, \odot)$ eine abelsche Gruppe ist. Die Assoziativität und die Kommutativität findet man wieder durch eine geradlinige Rechnung, die allerdings leicht länglich wird. Wir wollen hier deshalb darauf verzichten (Weniger freundlich ausgedrückt: Der Autor kneift...). Das multiplikative neutrale Element ist in diesem Fall gegeben durch $(1, 0)$, denn für jedes $(x, y) \in \mathbb{R}^2$ gilt nach Definition der Multiplikation

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

Ganz so einfach zu erraten ist das multiplikativ inverse Element nicht (aber Sie werden in wenigen Seiten wissen, wie man es sich merken kann). Wir geben uns also ein $(x, y) \in \mathbb{R}^2$ mit $(x, y) \neq (0, 0)$ vor und behaupten, dass $(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2})$ das Inverse ist. Bevor wir das nachrechnen, beachte man noch, dass dieser Ausdruck tatsächlich für alle $(x, y) \neq (0, 0)$ definiert ist, da der Nenner nur Null wird, wenn x und y beide Null sind. Tatsächlich haben wir

$$(x, y) \odot \Big(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2}\Big) = \Big(\frac{x^2}{x^2 + y^2} - \frac{-y^2}{x^2 + y^2}, \frac{-xy}{x^2 + y^2} + \frac{xy}{x^2 + y^2}\Big) = (1, 0).$$

Damit bleibt uns zum Körperglück nur noch ein Distributivgesetz nachzurechnen. Seien also $(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{R}^2$. Dann gilt

$$\begin{aligned}
&(x_1, y_1) \odot \big((x_2, y_2) \oplus (x_3, y_3)\big) = (x_1, y_1) \odot (x_2 + x_3, y_2 + y_3) \\
&= \big(x_1(x_2 + x_3) - y_1(y_2 + y_3), x_1(y_2 + y_3) + y_1(x_2 + x_3)\big) \\
&= \big(x_1x_2 + x_1x_3 - y_1y_2 - y_1y_3, x_1y_2 + x_1y_3 + x_2y_1 + x_3y_1\big) \\
&= \big(x_1x_2 - y_1y_2, x_1y_2 + x_2y_1\big) \oplus \big(x_1x_3 - y_1y_3, x_1y_3 + x_3y_1\big) \\
&= (x_1, y_1) \odot (x_2, y_2) \oplus (x_1, y_1) \odot (x_3, y_3).
\end{aligned}$$

□

Definition 2.5.3. $(\mathbb{R}^2, \oplus, \odot)$ *heißt* Körper der komplexen Zahlen *und wird üblicherweise mit* $\mathbb{C}$ *bezeichnet.*

Satz 2.5.4. *Die Abbildung* $f : \begin{cases} \mathbb{R} & \to \mathbb{C} \\ x & \mapsto (x, 0) \end{cases}$ *ist ein Körperhomomorphismus.*

Beweis. Es gilt für alle $x, y \in \mathbb{R}$

$$f(x + y) = (x + y, 0) = (x, 0) \oplus (y, 0) = f(x) \oplus f(y)$$

und

$$f(xy) = (xy, 0) = (xy - 0 \cdot 0, x \cdot 0 + 0 \cdot y) = (x, 0) \odot (y, 0).$$

Da schließlich noch $f(1) = (1, 0) = 1_{\mathbb{C}}$ ist, sind wir schon fertig. □

Bemerkung 2.5.5. Dank Satz 2.4.15 ist obiger Homomorphismus $f : \mathbb{R} \to \{(x, y) \in \mathbb{R}^2 : y = 0\}$ bijektiv. Die reellen Zahlen können daher mit der Identifikation $x \mathrel{\widehat{=}} (x, 0)$ als Teilkörper der komplexen Zahlen aufgefasst werden. Wir haben also unseren Zahlraum noch mal erweitert.

Beispiel 2.5.6. Wir berechnen zwei wesentliche Produkte komplexer Zahlen. Es ist

$$(0, 1) \odot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1, 0) \mathrel{\widehat{=}} -1$$

und für alle $y \in \mathbb{R}$ gilt

$$(y, 0) \odot (0, 1) = (y \cdot 0 - 0 \cdot 1, y \cdot 1 + 0 \cdot 0) = (0, y).$$

Bemerkung 2.5.7. Setzt man $\mathrm{i} := (0, 1)$, so gilt nach obiger Rechnung $\mathrm{i}^2 = (-1, 0) \mathrel{\widehat{=}} -1$. Damit können wir eine andere, intuitiver zu verwendende Schreibweise der komplexen Zahlen einführen. Wir machen uns dazu zu nutze, dass für $(x, y) \in \mathbb{C}$ mit obigen Identifikationen gilt

$$(x, y) = (x, 0) \oplus (0, y) = (x, 0) \odot (1, 0) \oplus (y, 0) \odot (0, 1) \mathrel{\widehat{=}} x \cdot 1 + y \cdot \mathrm{i}.$$

Die komplexe Addition und Multiplikation berechnet sich dann wegen

$$\begin{aligned}(x_1 + y_1\mathrm{i}) + (x_2 + y_2\mathrm{i}) &= (x_1 + x_2) + (y_1 + y_2)\mathrm{i} \quad \text{und} \\ (x_1 + y_1\mathrm{i}) \cdot (x_2 + y_2\mathrm{i}) &= x_1x_2 + y_1y_2\mathrm{i}^2 + x_1y_2\mathrm{i} + x_2y_1\mathrm{i} \\ &= (x_1x_2 - y_1y_2) + (x_1y_2 + x_2y_1)\mathrm{i},\end{aligned}$$

indem man wie wir es aus $\mathbb{R}$ gewohnt sind rechnet und unterwegs immer $\mathrm{i}^2 = -1$ beachtet.
Wir werden deshalb in Zukunft auf die Kringel um Plus und Mal verzichten und die gewohnten Symbole verwenden.
Die für die komplexen Zahlen fundamentale Zahl i nennt man auch die *imaginäre Einheit*.

Definition 2.5.8. *Sei $z \in \mathbb{C}$ und seien $x, y \in \mathbb{R}$ so, dass $z = (x, y) = x + y\mathrm{i}$ ist. Dann heißt*

$$\begin{aligned}\mathrm{Re}(z) &:= x \quad \text{Realteil } \textit{von } z \textit{ und} \\ \mathrm{Im}(z) &:= y \quad \text{Imaginärteil } \textit{von } z.\end{aligned}$$

Ist $y = 0$, so nennt man z reell *und ist $x = 0$, so heißt z* rein imaginär.

Bemerkung 2.5.9. (a) Zunächst als Warnung vor einem häufigen Fehler der Hinweis, dass der Imaginärteil einer komplexen Zahl immer reell ist. Es ist z.B. $\mathrm{Im}(3 + 2\mathrm{i}) = 2$ und nicht $2\mathrm{i}$.

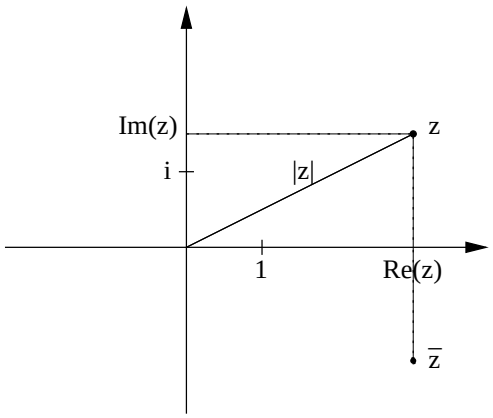

Abbildung 2.2.: Die komplexe Zahlenebene

(b) Da die komplexen Zahlen aus $\mathbb{R}^2$ hervorgehen, kann man sie sich gut in der sogenannten *komplexen Zahlenebene*, auch *Gauß'sche Zahlenebene* genannt, vgl. Abbildung 2.2, veranschaulichen.

Definition 2.5.10. *Sei $z = x + y\mathrm{i} \in \mathbb{C}$ mit $x, y \in \mathbb{R}$. Dann heißt.*

$$\overline{z} := x - y\mathrm{i} \qquad \text{zu } z \text{ konjugiert komplexe Zahl } \textit{und}$$
$$|z| := \sqrt{x^2 + y^2} \qquad \text{Betrag } \textit{von } z$$

Satz 2.5.11. *(a) Für jedes $z \in \mathbb{C}$ gilt $\overline{\overline{z}} = z$.*

(b) Die Abbildung $z \mapsto \overline{z}$ ist ein nichttrivialer Körperautomorphismus von $\mathbb{C}$.

(c) Es ist $z + \overline{z} = 2 \cdot \mathrm{Re}(z)$ und $z - \overline{z} = 2 \cdot \mathrm{Im}(z)\mathrm{i}$.

(d) Ein $z \in \mathbb{C}$ ist genau dann reell, wenn $z = \overline{z}$ gilt.

Beweis. (a) $\overline{\overline{z}} = \overline{\overline{x + y\mathrm{i}}} = \overline{x - y\mathrm{i}} = x + y\mathrm{i} = z$.

(b) Zunächst ist $\overline{1} = 1 = 1_{\mathbb{C}}$. Weiter gilt für alle $z = x + y\mathrm{i}$ und $w = u + v\mathrm{i}$ aus $\mathbb{C}$ mit $x, y, u, v \in \mathbb{R}$

$$\begin{aligned}\overline{z + w} &= \overline{x + y\mathrm{i} + u + v\mathrm{i}} = \overline{x + u + (y + v)\mathrm{i}} = x + u - (y + v)\mathrm{i} \\ &= x - y\mathrm{i} + u - v\mathrm{i} = \overline{z} + \overline{w}\end{aligned}$$

und

$$\begin{aligned}\overline{zw} &= \overline{(x + y\mathrm{i}) \cdot (u + v\mathrm{i})} = \overline{xu - yv + (xv + yu)\mathrm{i}} = xu - yv - (xv + yu)\mathrm{i} \\ &= xu - (-y)(-v) + x(-v)\mathrm{i} + (-y)u\mathrm{i} = (x + (-y)\mathrm{i}) \cdot (u + (-v)\mathrm{i}) \\ &= (x - y\mathrm{i}) \cdot (u - v\mathrm{i}) = \overline{z} \cdot \overline{w}.\end{aligned}$$

Also ist die Konjugation schon mal ein Körperhomomorphismus. Nach Satz 2.4.15 ist dieser auch injektiv, wir brauchen also zum Nachweis, dass es sich um einen Automorphismus handelt, nur noch die Surjektivität. Doch diese folgt direkt aus (a), denn zu jedem $z \in \mathbb{C}$ ist demnach $\overline{z}$ ein Urbild unter der Konjugation.

Schließlich ist die Konjugation nicht der triviale Körperautomorphismus, denn $\overline{\mathrm{i}} = -\mathrm{i} \neq \mathrm{i}$.

(c) $z + \overline{z} = x + y\mathrm{i} + x - y\mathrm{i} = 2x = 2 \cdot \mathrm{Re}(z)$ und $z - \overline{z} = x + y\mathrm{i} - (x - y\mathrm{i}) = 2y\mathrm{i} = 2 \cdot \mathrm{Im}(z)\mathrm{i}$.

(d) Ist $z \in \mathbb{C}$ reell, so gilt $z = x + 0 \cdot \mathrm{i}$ für ein $x \in \mathbb{R}$. Also ist in diesem Fall $\overline{z} = \overline{x + 0 \cdot \mathrm{i}} = x - 0 \cdot \mathrm{i} = x = z$.

Gilt umgekehrt $z = \overline{z}$, so ist mit obiger Rechnung $\mathrm{Im}(z) = (z - \overline{z})/(2\mathrm{i}) = 0$, also ist $z = \mathrm{Re}(z)$ reell. □

Satz 2.5.12. *Für alle $z, z_1, z_2 \in \mathbb{C}$ gilt*

(a) $|z| = |\overline{z}|$.

(b) $z \cdot \overline{z} = |z|^2$.

(c) $z^{-1} = \frac{\overline{z}}{|z|^2}$ *falls* $z \neq 0$.

(d) $\mathrm{Re}(z) \leq |z|$ *und* $\mathrm{Im}(z) \leq |z|$.

(e) $|z| \in \mathbb{R}$ *und* $|z| \geq 0$ *und* $(|z| = 0 \iff z = 0)$.

(f) $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$.

(g) $|z_1 + z_2| \leq |z_1| + |z_2|$. (Dreiecksungleichung)

Beweis. Übung □

Mit dem Wissen aus (c) dieses Satzes erklärt sich nun auch rückwirkend die zunächst unintuitive Wahl des multiplikativen Inversen im Beweis von Satz 2.5.2.

Satz 2.5.13. *Es gibt keine Totalordnung auf $\mathbb{C}$, die $\mathbb{C}$ zu einem angeordneten Körper macht.*

Beweis. Wir nehmen an, es gäbe eine solche Totalordnung „$\leq$“. Nach Übungsaufgabe 2.4.17 (b) gilt dann $z^2 \geq 0$ für jedes $z \in \mathbb{C}$. Speziell für $z = -1$ erhalten wir also $1 = (-1)^2 \geq 0$ und mit $z = \mathrm{i}$ erhalten wir $-1 = \mathrm{i}^2 \geq 0$. Das ist nun ein Widerspruch zu Übungsaufgabe 2.4.17 (a). Also kann es solch eine Totalordnung nicht geben. □

Satz 2.5.14 (Fundamentalsatz der Algebra)**.** *Es sei $n \in \mathbb{N}^*$ und $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$ ein Polynom mit $a_j \in \mathbb{C}$ für $j = 0, 1, \ldots, n$ und $a_n \neq 0$. Dann hat p eine Nullstelle in $\mathbb{C}$.*
Insbesondere zerfällt jedes komplexe Polynom über $\mathbb{C}$ in Linearfaktoren.

Bemerkung 2.5.15. Der Fundamentalsatz der Algebra bedeutet, dass jede polynomiale Gleichung über $\mathbb{C}$ lösbar ist (ja, außer $3 = 5$ und Ähnlichem natürlich...). Der entsprechende Schönheitsfehler von $\mathbb{R}$, wo z.B. die Gleichung $x^2 + 1 = 0$ keine Lösung besitzt, ist damit durch die Zahlerweiterung nach $\mathbb{C}$ behoben. Wir werden diese Eigenschaft von $\mathbb{C}$ noch sehr zu schätzen lernen.

3. Lineare Algebra

3.1. Vektorräume

3.1.1. Das Axiomensystem und Beispiele

Definition 3.1.1. *(a) Sei V eine Menge und K ein Körper. Weiter seien zwei Verknüpfungen*

$$\begin{aligned} + &: V \times V \to V, && \text{(Vektoraddition)} \\ \cdot &: K \times V \to V && \text{(Skalar-Multiplikation)} \end{aligned}$$

gegeben. Die Menge V mit diesen beiden Verknüpfungen heißt dann Vektorraum *über K oder auch K-*Vektorraum*, falls die folgenden Axiome erfüllt sind:*

(V1) $(V, +)$ *ist eine abelsche Gruppe.*

(V2) $\forall v \in V : 1 \cdot v = v.$

(V3) $\forall v \in V\ \forall \alpha, \beta \in K : (\alpha\beta) \cdot v = \alpha \cdot (\beta \cdot v).$

(V4) $\forall v \in V\ \forall \alpha, \beta \in K : (\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v.$

(V5) $\forall v, w \in V\ \forall \alpha \in K : \alpha \cdot (v + w) = \alpha \cdot v + \alpha \cdot w.$

(b) Das neutrale Element der Gruppe $(V, +)$ wird als Nullvektor *bezeichnet und die Elemente des zugrundeliegenden Körpers K nennt man* Skalare.

Ist speziell $K = \mathbb{R}$, bzw. $K = \mathbb{C}$, so spricht man von einem reellen, *bzw.* komplexen *Vektorraum.*

Vektorräume spielen in vielen Bereichen der Mathematik eine fundamentale Rolle. Wir wollen das mit einem Stapel verschiedener Beispiele andeuten.

Beispiel 3.1.2. (a) **Der Raum K^n der n-Tupel**

Sei K ein Körper und $n \in \mathbb{N}^*$. Dann ist

$$K^n := \underbrace{K \times K \times \cdots \times K}_{n \text{ Mal}} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : x_j \in K \text{ für } j = 1, 2, \ldots, n \right\}$$

mit den für $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in K^n$ und $\alpha \in K$ durch

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad \text{und} \quad \alpha \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

gegebenen Verknüpfungen ein K-Vektorraum mit $\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ als Nullvektor. Das Nachrechnen der Axiome ist eine leichte Übung.

Im Falle $K = \mathbb{R}$ ist $\mathbb{R}^n$ der sogenannte reelle *Standardvektorraum*.

Aus Gründen, die erst später klar werden werden, ist es sinnvoll die Elemente von K^n als Spalten zu schreiben. Da das aber in der schriftlichen Darstellung manchmal sehr viel Platz verbraucht, führen wir die Notation

$$(x_1, x_2, \ldots, x_n)^T := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^T := (x_1, x_2, \ldots, x_n)$$

ein. Das „T“ macht also formal aus einem Zeilenvektor einen Spaltenvektor und umgekehrt. Man liest x^T als „x transponiert“.

(b) **Der Raum der $p \times n$-Matrizen**

Seien K ein Körper und $p, n \in \mathbb{N}^*$. Dann ist $K^{p\times n}$ der Vektorraum aller Matrizen mit p Zeilen und n Spalten, d.h.

$$K^{p\times n} := \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \ldots & \alpha_{pn} \end{pmatrix} : \alpha_{j,k} \in K,\ j = 1, \ldots, p,\ k = 1, \ldots, n \right\}.$$

Eine *Matrix* ist also ein rechteckiges Schema aus pn Elementen aus K. Wie kann man nun mit solchen Monstern rechnen und wozu ist das gut? Die Antwort auf die zweite Frage werde ich Ihnen im weiteren Verlauf der Vorlesung näher bringen, zunächst wollen wir für die Matrizen eine Addition und eine Skalar-Multiplikation definieren.

Seien also $A = (\alpha_{jk})_{j=1,\ldots,p,k=1,\ldots,n}$ und $B = (\beta_{jk})_{j=1,\ldots,p,k=1,\ldots,n}$ Matrizen aus $K^{p\times n}$, sowie $\lambda \in K$. Dann definieren wir beide Verknüpfungen im Prinzip

wie in (a) komponentenweise durch

$$A + B = (\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n} + (\beta_{jk})_{j=1,\dots,p,k=1,\dots,n} := (\alpha_{jk} + \beta_{jk})_{j=1,\dots,p,k=1,\dots,n}$$
$$= \begin{pmatrix} \alpha_{11} + \beta_{11} & \alpha_{12} + \beta_{12} & \dots & \alpha_{1n} + \beta_{1n} \\ \alpha_{21} + \beta_{21} & \alpha_{22} + \beta_{22} & \dots & \alpha_{2n} + \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} + \beta_{p1} & \alpha_{p2} + \beta_{p2} & \dots & \alpha_{pn} + \beta_{pn} \end{pmatrix}$$

und

$$\lambda A = \lambda(\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n} := (\lambda\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n}$$
$$= \begin{pmatrix} \lambda\alpha_{11} & \lambda\alpha_{12} & \dots & \lambda\alpha_{1n} \\ \lambda\alpha_{21} & \lambda\alpha_{22} & \dots & \lambda\alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda\alpha_{p1} & \lambda\alpha_{p2} & \dots & \lambda\alpha_{pn} \end{pmatrix}.$$

Der Nullvektor, der hier auch *Nullmatrix* genannt wird, ist die Matrix, deren Einträge alle Null sind. Das Nachrechnen der Axiome ist hier naturgemäß etwas mühsamer als in (a) aber genauso elementar.

Hier ist noch ein konkretes Beispiel für das Rechnen mit Matrizen über $\mathbb{Q}$, bzw. $\mathbb{R}$, bzw. $\mathbb{C}$.

$$\begin{pmatrix} 3 & 1 & 2 \\ -2 & 5 & 1 \\ 1 & -1 & -1 \end{pmatrix} + 2 \begin{pmatrix} -2 & 0 & 3 \\ 1 & -2 & -2 \\ 1 & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 3 & 1 & 2 \\ -2 & 5 & 1 \\ 1 & -1 & -1 \end{pmatrix} + \begin{pmatrix} -4 & 0 & 6 \\ 2 & -4 & -4 \\ 2 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 8 \\ 0 & 1 & -3 \\ 3 & -1 & -1 \end{pmatrix}.$$

(c) **Funktionenräume**

Sei K ein Körper und M eine Menge. Die Menge $\mathrm{Abb}(M, K)$ aller Funktionen von M nach K ist mit den für $f, g \in \mathrm{Abb}(M, K)$ und $\alpha \in K$ definierten Verknüpfungen

$$f + g : \begin{cases} M & \to K \\ x & \mapsto f(x) + g(x) \end{cases} \qquad \text{und} \qquad \alpha f : \begin{cases} M & \to K \\ x & \mapsto \alpha f(x) \end{cases}$$

ein K-Vektorraum.

Dies wollen wir exemplarisch ausführlich beweisen. Zunächst beobachten wir, das die beiden oben definierten Verknüpfungen in dem Sinne wohldefiniert sind, dass sie als Ergebnis jeweils immer wieder ein Element von $\mathrm{Abb}(M, K)$ liefern. Es bleiben also die Axiome (V1) bis (V5) zu zeigen.

($V1$) Zunächst ist die Verknüpfung „+“ assoziativ, denn für je drei Funktionen $f, g, h \in \mathrm{Abb}(M, K)$ gilt dank der Assoziativität der Addition in K für alle $x \in M$

$$\begin{aligned}\big[(f+g)+h\big](x) &= (f+g)(x) + h(x) = \big(f(x)+g(x)\big) + h(x)\\ &= f(x) + \big(g(x)+h(x)\big) = f(x) + (g+h)(x)\\ &= \big[f+(g+h)\big](x).\end{aligned}$$

Also ist $(f+g)+h = f+(g+h)$.

Genauso bekommt man die Kommutativität wegen

$$(f+g)(x) = f(x)+g(x) = g(x)+f(x) = (g+f)(x).$$

Als neutrales Element der Addition identifizieren wir die Nullabbildung $o : M \to K$ mit $o(x) = 0$ für alle $x \in M$. Mit dieser gilt nämlich für alle $f \in \mathrm{Abb}(M, K)$ und alle $x \in M$

$$(f+o)(x) = f(x) + o(x) = f(x) + 0 = f(x),$$

also ist $f + o = f$ für jedes $f \in \mathrm{Abb}(M, K)$.

Schließlich findet man zu jedem $f \in \mathrm{Abb}(M, K)$ das additiv inverse Element $-f : M \to K$ mit $(-f)(x) = -f(x)$, $x \in M$, denn für dieses gilt für jedes $x \in M$

$$(f+(-f))(x) = f(x) + (-f)(x) = f(x) - f(x) = 0 = o(x),$$

und damit haben wir $f + (-f) = o$.

($V2$) Sei $f \in \mathrm{Abb}(M, K)$. Dann gilt $(1 \cdot f)(x) = 1 \cdot f(x) = f(x)$ für alle $x \in M$, also ist $1 \cdot f = f$.

($V3$) Seien $\alpha, \beta \in K$ und $f \in \mathrm{Abb}(M, K)$. Dann gilt für jedes $x \in M$ unter Ausnutzung der Assoziativität der Multiplikation in K

$$\big[(\alpha\beta)\cdot f\big](x) = (\alpha\beta)f(x) = \alpha(\beta f(x)) = \alpha\big((\beta\cdot f)(x)\big) = \big[\alpha\cdot(\beta\cdot f)\big](x)$$

und damit $(\alpha\beta)\cdot f = \alpha\cdot(\beta\cdot f)$.

($V4$) Seien $\alpha, \beta \in K$ und $f \in \mathrm{Abb}(M, K)$. Dann gilt für jedes $x \in M$ unter Ausnutzung des Distributivgesetzes in K

$$\begin{aligned}\big[(\alpha+\beta)\cdot f\big](x) &= (\alpha+\beta)f(x) = \alpha f(x) + \beta f(x)\\ &= (\alpha\cdot f)(x) + (\beta\cdot f)(x) = \big[\alpha\cdot f + \beta\cdot f\big](x).\end{aligned}$$

Das liefert wieder $(\alpha+\beta)\cdot f = \alpha\cdot f + \beta\cdot f$.

($V5$) Es seien $\alpha \in K$ und $f, g \in \mathrm{Abb}(M, K)$. Dann gilt für jedes $x \in M$ wieder dank des Distributivgesetzes

$$\begin{aligned}\big[\alpha \cdot (f+g)\big](x) &= \alpha(f+g)(x) = \alpha\big(f(x)+g(x)\big) = \alpha f(x) + \alpha g(x)\\ &= (\alpha \cdot f)(x) + (\alpha \cdot g)(x) = \big[\alpha \cdot f + \alpha \cdot g\big](x)\end{aligned}$$

und wir bekommen $\alpha \cdot (f+g) = \alpha \cdot f + \alpha \cdot g$ wie gefordert.

(d) **Der Raum aller Folgen in K**

Als Spezialfall von (c) erhalten wir mit $M = \mathbb{N}$ den Raum $F = \mathrm{Abb}(\mathbb{N}, K)$ aller *Folgen* in K. Für ein Element $a \in F$ schreibt man für das Bild von $n \in \mathbb{N}$ unter a statt $a(n)$ üblicherweise a_n und gibt die Abbildung a als „unendliche Liste“ der Bilder $(a_0, a_1, a_2, \dots)$ an. Beispiele für Folgen in $\mathbb{R}$ sind

$$\begin{aligned}\Big(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\Big) &= \Big(\frac{1}{n+1}\Big)_{n\in\mathbb{N}},\\ (1, 2, 4, 8, 16, \dots) &= (2^n)_{n\in\mathbb{N}},\\ (1, -1, 1, -1, 1, -1, 1, \dots) &= \big((-1)^n\big)_{n\in\mathbb{N}}.\end{aligned}$$

In dieser Schreibweise lesen sich die Verknüpfungen aus $F = \mathrm{Abb}(\mathbb{N}, K)$ mit $a, b \in F$ und $\alpha \in K$ so:

$$\begin{aligned}a + b &= (a_n)_{n\in\mathbb{N}} + (b_n)_{n\in\mathbb{N}} = (a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots)\\ &:= (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots) = (a_n + b_n)_{n\in\mathbb{N}} \quad \text{und}\\ \alpha \cdot a &= \alpha \cdot (a_n)_{n\in\mathbb{N}} = \alpha \cdot (a_0, a_1, a_2, \dots) := (\alpha a_0, \alpha a_1, \alpha a_2, \dots) = (\alpha a_n)_{n\in\mathbb{N}}.\end{aligned}$$

(e) **Der Raum aller endlichen Folgen**

Betrachtet man die Teilmenge

$$c_{00} := \{a \in F : a_n \neq 0 \text{ nur für endlich viele } n \in \mathbb{N}\},$$

von F aus (d), so ist auch diese mit den Verknüpfungen aus F ein K-Vektorraum.

Elemente von c_{00} sind z.B. für jedes $k \in \mathbb{N}$ die Folgen $e^{(k)}$ mit $e^{(k)}_j = 1$ für $j = k$ und Null sonst, d.h.

$$\begin{aligned}e^{(0)} &= (1, 0, 0, 0, 0, 0, 0, \dots), & e^{(1)} &= (0, 1, 0, 0, 0, 0, 0, 0, \dots),\\ e^{(2)} &= (0, 0, 1, 0, 0, 0, 0, \dots), & e^{(3)} &= (0, 0, 0, 1, 0, 0, 0, 0, \dots), \quad \text{usw.}\end{aligned}$$

Verwendet man für $j, k \in \mathbb{Z}$ das sogenannte *Kronecker-Delta*, d.h. die Schreibweise

$$\delta_{jk} := \begin{cases} 1, & \text{falls } j = k,\\ 0, & \text{falls } j \neq k,\end{cases}$$

so kann man diese speziellen Elemente kurz beschreiben durch

$$e^{(k)} = (\delta_{jk})_{j\in\mathbb{N}}, \qquad k \in \mathbb{N}.$$

Bemerkung 3.1.3. In jedem Vektorraum V gelten für die abelsche Gruppe $(V, +)$ natürlich alle Ergebnisse aus dem Abschnitt über Gruppen. Insbesondere ist also der Nullvektor und das additive Inverse jeweils eindeutig bestimmt. Ebenso übernehmen wir für $u, v \in G$ die Schreibweise $u - v$ für $u + (-v)$. Schließlich bemerken wir noch, dass nach unseren Erkenntnissen über Gruppen die Gleichung $a + x = b$ für jede Vorgabe von $a, b \in V$ in V eindeutig lösbar ist.

Satz 3.1.4. *Es sei V ein K-Vektorraum mit Nullvektor 0_V. Dann gilt für jedes $\alpha \in K$ und alle $v \in V$*

(a) $\alpha \cdot v = 0_V \iff (\alpha = 0$ *oder* $v = 0_V)$.

(b) $(-\alpha) \cdot v = -(\alpha \cdot v)$, *insbesondere ist* $(-1) \cdot v = -v$.

Beweis. Wir beweisen nur (a), der Teil (b) verbleibt als Übung.
Zum Nachweis von „$\Leftarrow$“ in (a) sei zunächst $\alpha = 0$. Dann gilt

$$v \overset{(V2)}{=} 1 \cdot v = (1 + 0) \cdot v \overset{(V4)}{=} 1 \cdot v + 0 \cdot v \overset{(V2)}{=} v + 0 \cdot v.$$

Nach Bemerkung 3.1.3 hat die Gleichung $v = v + x$ genau eine Lösung in V. Da 0_V nach (V1) eine Lösung ist, muss also $0 \cdot v = 0_V$ sein, wie gewünscht.
Sei nun $v = 0_V$. Wir beobachten, dass für jedes $\alpha \in K$ und $w \in V$ gilt

$$\alpha \cdot w \overset{(V1)}{=} \alpha \cdot (w + 0_V) \overset{(V5)}{=} \alpha \cdot w + \alpha \cdot 0_V.$$

Auch die eindeutig lösbare Gleichung $\alpha \cdot w = \alpha \cdot w + x$ hat wieder $x = 0_V$ als Lösung. Also ist $\alpha \cdot 0_V = 0_V$.
Es bleibt noch „$\Rightarrow$“ zu zeigen. Seien also $\alpha \in K$ und $v \in V$ mit $\alpha \cdot v = 0_V$ gegeben. Ist $\alpha = 0$ so sind wir fertig, wir betrachten also den Fall $\alpha \neq 0$. Dann gilt

$$v \overset{(V2)}{=} 1 \cdot v \overset{\alpha \neq 0}{=} (\alpha^{-1}\alpha) \cdot v \overset{(V3)}{=} \alpha^{-1} \cdot (\alpha \cdot v) = \alpha^{-1} \cdot 0_V \overset{\text{“}\Leftarrow\text{“}}{=} 0_V. \qquad \square$$

3.1.2. Die Summenschreibweise

In Vektorräumen wird viel addiert und wir werden in den nächsten Kapiteln Unmengen Summen mit einer variablen Anzahl, also z.B. n, Summanden haben. Dazu führen wir folgende sehr praktische Notation ein, die Sie sich unbedingt angewöhnen sollten.

Definition 3.1.5. *Sei $n \in \mathbb{N}$ und $a_0, a_1, a_2, \ldots, a_n$ seien Elemente einer kommutativen additiven Struktur, also z.B. eines Vektorraums, Körpers oder Rings, oder auch einer abelschen Gruppe, deren Verknüpfung additiv geschrieben wird. Dann schreibt man*

$$\sum_{j=0}^{n} a_j := a_0 + a_1 + a_2 + \cdots + a_n.$$

Die Variable, die die Summanden hochzählt, in obigem Beispiel j, heißt Summationsindex.

In Erweiterung obiger Definition schreibt man auch

$$\sum_{j=3}^{9} 2^j = 2^3 + 2^4 + 2^5 + \cdots + 2^9 \quad \text{oder} \quad \sum_{j=1}^{\infty} x^j = x + x^2 + x^3 + x^4 + \ldots$$

mit hoffentlich intuitiv klarer Bedeutung. Zumindest sollte jeder/m, die/der schon mal eine Schleife programmiert hat, klar sein was hier passiert.
Im folgenden Beispiel kann man einige oft verwendete Rechenregeln für das Summenzeichen finden.

Beispiel 3.1.6. (a)

$$\sum_{k=3}^{9} (k-3)^5 = 0^5 + 1^5 + 2^5 + \cdots + 6^5 = \sum_{k=0}^{6} k^5 \quad \text{(Indexshift)}$$

(b) Seien $n \in \mathbb{N}^*$, V ein K-Vektorraum und $\alpha \in \mathbb{K}$, sowie $a_1, a_2, \ldots, a_n \in V$. Dann ist

$$\alpha \cdot \sum_{k=1}^{n} a_k = \alpha \cdot (a_1 + a_2 + \cdots + a_n) = \alpha a_1 + \alpha a_2 + \cdots + \alpha a_n = \sum_{k=1}^{n} \alpha a_k.$$

So einfach ist Ausmultiplizieren und Ausklammern mit dem Summenzeichen.

3.2. Untervektorräume, Basis und Dimension

3.2.1. Untervektorräume

Definition 3.2.1. *Sei V ein K-Vektorraum. Eine Teilmenge U von V heißt* Untervektorraum *von V, falls U mit den Verknüpfungen von V ebenfalls ein K-Vektorraum ist.*

Bemerkung 3.2.2. Die Teilmengen $\{0_V\}$ und V sind in jedem Vektorraum V Untervektorräume.

Satz 3.2.3 (Untervektorraumkriterium)**.** *Eine Teilmenge U eines Vektorraums V ist genau dann ein Untervektorraum von V, wenn*

(UVR1) $U \neq \emptyset$ *und*

(UVR2) $\forall a, b \in U\ \forall \lambda, \mu \in K : \lambda a + \mu b \in U$

gelten.

Beweis. **„⇒“** Sei U ein Untervektorraum von V. Da damit U ein Vektorraum ist, muss U nach (V1) zumindest einen Nullvektor enthalten, also gilt (UVR1). Seien zum Nachweis von (UVR2) nun $a, b \in U$ und $\lambda, \mu \in K$. Da U mit den Verknüpfungen aus V ein K-Vektorraum ist, muss $\cdot : K \times U \to U$ und $+ : U \times U \to U$ gelten. Damit sind zunächst λa und μb in U und dann auch $\lambda a + \mu b$.

„⇐“ Wir müssen zeigen, dass wir aus der Information, dass U eine Teilmenge eines K-Vektorraums ist, und dass (UVR1) und (UVR2) gelten, schon nachweisen können, dass U selbst ein K-Vektorraum ist. Setzen wir in (UVR2) $\lambda = \mu = 1$, so erhalten wir, dass für alle $a, b \in U$ gilt $a + b \in U$, also ist $+ : U \times U \to U$ schon mal eine vernünftige Verknüpfung auf U. Setzt man $\mu := 0$ und $b := a$, so erhält man das selbe Resultat für die Skalar-Multiplikation.

Der Nachweis der Assoziativität und der Kommutativität von „+“, sowie von (V2)–(V5) ergibt sich jeweils direkt aus den entsprechenden Eigenschaften von V. Was wirklich zu zeigen bleibt, ist die Existenz eines neutralen Elements und der additiv inversen Elemente in $(U, +)$.

Nach (UVR1) ist $U \neq \emptyset$, also gibt es irgendein $u \in U$. Mit Hilfe von (UVR2) muss dann auch $0 \cdot u = 0_V \in U$ sein und dieses Element ist natürlich auch in U ein neutrales, denn es gilt ja sogar $v + 0_V = v$ für alle $v \in V$.

Sei nun $a \in U$ gegeben. Dann ist wiederum nach (UVR2) auch das Element $-a = (-1) \cdot a \in U$ und wir sind fertig. □

Beispiel 3.2.4. (a) Der Vektorraum c_{00} der endlichen Folgen in K, den wir in Beispiel 3.1.2 (e) kennengelernt haben, ist ein Untervektorraum des Raums aller Folgen aus Beispiel 3.1.2 (d).

(b) In $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ betrachten wir die Menge U aller Funktionen $f : \mathbb{R} \to \mathbb{R}$, für die $f(0) = 0$ gilt und weisen nach, das dies ein Untervektorraum ist. Zunächst hat die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = x^3 - \pi x^2 + 15x$ diese Eigenschaft, also ist U nicht leer, d.h. (UVR1) gilt. Seien nun $f, g \in U$ und $\lambda, \mu \in \mathbb{R}$. Dann gilt

$$(\lambda f + \mu g)(0) = \lambda f(0) + \mu g(0) = \lambda \cdot 0 + \mu \cdot 0 = 0,$$

also ist auch $\lambda f + \mu g \in U$ und wir haben (UVR2).

(c) Im Standardvektorraum $\mathbb{R}^2$ kann man alle Untervektorräume angeben. Neben den immer vorhandenen $\{0\}$ und $\mathbb{R}^2$ sind das genau die durch Ursprungsgeraden beschriebenen Mengen.

Definition 3.2.5. *Seien V ein K-Vektorraum, $n \in \mathbb{N}^*$ und $a_1, a_2, \ldots, a_n \in V$, sowie $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$. Dann heißt*

$$\sum_{j=1}^{n} \alpha_j a_j$$

eine Linearkombination *von $a_1, a_2, \ldots, a_n$.*

Beispiel 3.2.6. In $\mathbb{R}^2$ ist der Vektor $(1,6)^T$ eine Linearkombination von $a_1 = (1,2)^T$, $a_2 = (0,1)^T$ und $a_3 = (0,2)^T$, z.B. mit

$$(1,6)^T = a_1 - 2a_2 + 3a_3 \qquad \text{oder} \qquad (1,6)^T = a_1 + 4a_2 + 0a_3.$$

Definition 3.2.7. *Sei V ein K-Vektorraum und $M \subseteq V$. Dann heißt*

$$\begin{aligned}\langle M\rangle &:= \big\{v \in V : v \text{ ist Linearkombination von Vektoren aus } M\big\} \\ &= \Big\{v \in V : \exists n \in \mathbb{N}^* \; \exists \alpha_1, \ldots, \alpha_n \in K \; \exists m_1, \ldots m_n \in M \text{ mit } v = \sum_{j=1}^{n} \alpha_j m_j\Big\}\end{aligned}$$

lineare Hülle *von M.*
Ist $M = \{m_1, m_2, \ldots, m_n\}$ endlich, so schreibt man auch $\langle m_1, m_2, \ldots, m_n\rangle$ statt $\langle\{m_1, m_2, \ldots, m_n\}\rangle$.
Schließlich setzen wir $\langle\emptyset\rangle = \{0\}$.

Satz 3.2.8. *Sei V ein K-Vektorraum und $M \subseteq V$. Dann ist $\langle M\rangle$ ein Untervektorraum von V.*

Beweis. Ist $M = \emptyset$, so ist $\langle M\rangle = \{0\}$ und damit ein Untervektorraum. Wir betrachten also den Fall $M \neq \emptyset$. Da jedes Element $m \in M$ sich als die Linearkombination $1 \cdot m$ mit Elementen aus M schreiben lässt, ist immer $M \subseteq \langle M\rangle$ und damit insbesondere auch $\langle M\rangle \neq \emptyset$. Zum Nachweis von (UVR2) seien $u, v \in \langle M\rangle$ und $\lambda, \mu \in K$. Dann existieren $n, p \in \mathbb{N}^*$, sowie $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_p \in K$ und $a_1, \ldots, a_n, b_1, \ldots b_p \in M$ mit

$$u = \sum_{j=1}^{n} \alpha_j a_j \qquad \text{und} \qquad v = \sum_{k=1}^{p} \beta_k b_k.$$

Also ist

$$\lambda u + \mu v = \lambda \sum_{j=1}^{n} \alpha_j a_j + \mu \sum_{k=1}^{p} \beta_k b_k = \sum_{j=1}^{n} (\lambda\alpha_j) a_j + \sum_{k=1}^{p} (\mu\beta_k) b_k$$

und dieses ist wiederum eine Linearkombination von endlich vielen Elementen $a_1, \ldots, a_n, b_1, \ldots, b_p \in M$, also ist $\lambda u + \mu v \in \langle M\rangle$ und damit ist (UVR2) erfüllt. □

Übungsaufgabe 3.2.9. Seien V ein K-Vektorraum und M, M_1, M_2 Teilmengen von V. Dann gilt

(a) $M_1 \subseteq M_2 \Longrightarrow \langle M_1 \rangle \subseteq \langle M_2 \rangle$.

(b) $M = \langle M \rangle \iff M$ Untervektorraum von V.

Übungsaufgabe 3.2.10. In Definition 2.3.10 im Abschnitt über Gruppen haben wir das (Gruppen-)Erzeugnis definiert, dessen Idee sehr an das der linearen Hülle erinnert: Finde eine möglichst kleine Unterstruktur, die die gegebene Teilmenge aber komplett enthält. Dort war das Vorgehen allerdings ein anderes. Ziel dieser Aufgabe ist es, zu sehen, dass das dort verwendete Verfahren ebenfalls zur Definition der linearen Hülle verwendet werden kann und das dabei auch genau das selbe herauskommt. Beweisen sie dazu die beiden folgenden Aussagen für einen K-Vektorraum V.

(a) Ist $\mathcal{U} \subseteq \mathcal{P}(V)$ ein Mengensystem von Untervektorräumen von V, so ist auch $\bigcap_{U \in \mathcal{U}} U$ ein Untervektorraum von V.

(b) Sei $M \subseteq V$ und

$$\mathcal{U} := \{U \in \mathcal{P}(V) : M \subseteq U \text{ und } U \text{ Untervektorraum von } V\}.$$

Dann gilt

$$\bigcap_{U \in \mathcal{U}} U = \langle M \rangle.$$

3.2.2. Lineare Unabhängigkeit und Basen

Definition 3.2.11. *Sei V ein K-Vektorraum und $M \subseteq V$.*

(a) Die Menge M heißt linear abhängig*, falls es eine nichttriviale Linearkombination des Nullvektors aus Elementen von M gibt, d.h. wenn es ein $n \in \mathbb{N}^*$, n verschiedene Vektoren $v_1, v_2, \ldots, v_n \in M$ und Koeffizienten $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$ gibt, die nicht alle Null sind, mit $\sum_{j=1}^{n} \alpha_j v_j = 0_V$.*

(b) Ist M nicht linear abhängig, so nennt man M linear unabhängig*.*

Beispiel 3.2.12. (a) Die Teilmenge von $\mathbb{R}^2$, gegeben durch

$$\left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \end{pmatrix} \right\}$$

ist linear abhängig, denn

$$1 \cdot \begin{pmatrix} 5 \\ 1 \end{pmatrix} - 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 2 \cdot \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

(b) Betrachtet man in $\mathbb{R}^2$ hingegen die Menge

$$\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\},$$

so ist diese linear unabhängig, denn aus

$$\alpha \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

folgt $\alpha = 0$ und $\beta = 0$.

(c) Im $\mathbb{R}$-Vektorraum $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ betrachten wir die Menge $\{f, g\} := \{x \mapsto x^2, x \mapsto x\}$. Diese ist linear unabhängig, denn hat man $\alpha, \beta \in \mathbb{R}$ mit $\alpha f + \beta g = o$, so folgt $\alpha x^2 + \beta x = 0$ für jedes $x \in \mathbb{R}$. Setzt man speziell $x = 1$, bzw. $x = -1$ ein, so erhält man $\alpha + \beta = 0$, bzw. $\alpha - \beta = 0$. Addition der beiden Gleichungen liefert dann $2\alpha = 0$ und damit zuerst $\alpha = 0$ und dann $\beta = 0$.

(d) In jedem Vektorraum ist $\{0\}$ linear abhängig, denn $1 \cdot 0 = 0$ ist eine nichttriviale Linearkombination des Nullvektors.

Bemerkung 3.2.13. Linearkombinationen sind immer *endliche* Summen, d.h. eine Teilmenge M eines Vektorraums ist genau dann linear unabhängig, wenn für jede Wahl von $n \in \mathbb{N}^*$ und von Vektoren $v_1, \ldots, v_n \in M$ gilt

$$\sum_{j=1}^{n} \alpha_j v_j = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

Satz 3.2.14. *Sind V ein K-Vektorraum, $n \in \mathbb{N}^*$ und $v_1, \ldots, v_n \in V$, so gilt*

(a) Die Vektoren $v_1, \ldots, v_n$ sind genau dann linear abhängig, wenn einer von ihnen eine Linearkombination der $n-1$ anderen ist.

(b) Ist $p \leq n$ und sind $v_1, \ldots, v_n$ linear unabhängig, so sind auch $v_1, \ldots, v_p$ linear unabhängig.

(c) Ist $p \leq n$ und sind $v_1, \ldots, v_p$ linear abhängig, so sind auch $v_1, \ldots, v_n$ linear abhängig.

(d) Bildet man $n+1$ Linearkombinationen $w_1, \ldots, w_{n+1}$ aus $v_1, \ldots, v_n$, so sind $w_1, \ldots, w_{n+1}$ linear abhängig.

Beweis. (a) Wir beweisen zunächst „$\Longrightarrow$“. Seien dazu $v_1, \ldots, v_n \in V$ linear abhängig. Dann existieren $\alpha_1, \ldots, \alpha_n \in K$, die nicht alle Null sind, so dass

$\sum_{j=1}^{n} \alpha_j v_j = 0$ gilt. Sei $j_0 \in \{1, \ldots, n\}$ so gewählt, dass $\alpha_{j_0} \neq 0$ ist. Dann haben wir

$$\alpha_{j_0} v_{j_0} + \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \alpha_j v_j = 0.$$

Also ergibt sich dank $\alpha_{j_0} \neq 0$

$$v_{j_0} = -\frac{1}{\alpha_{j_0}} \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \alpha_j v_j = \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \frac{-\alpha_j}{\alpha_{j_0}} v_j,$$

was eine Linearkombination der $n-1$ übrigen Vektoren ist.

Zum Nachweis der umgekehrten Implikation „$\Longleftarrow$“, sei j_0 ein Index, für den v_{j_0} eine Linearkombination der übrigen Vektoren ist. Das bedeutet, dass es $\alpha_1, \ldots, \alpha_{j_0-1}, \alpha_{j_0+1}, \ldots, \alpha_n \in K$ gibt mit

$$v_{j_0} = \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \alpha_j v_j, \quad \text{also ist} \quad 1 \cdot v_{j_0} - \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \alpha_j v_j = 0$$

und wir haben wegen $1 \neq 0$ eine nichttriviale Linearkombination des Nullvektors gefunden. Also sind $v_1, \ldots, v_n$ linear abhängig.

(b) Übung

(c) Übung

(d) ohne Beweis □

Definition 3.2.15. *Sei V ein K-Vektorraum. Eine Teilmenge $\mathcal{B} \subseteq V$ heißt* Basis *von V, falls*

(B1) *$\mathcal{B}$ ist linear unabhängig und*

(B2) *$\langle \mathcal{B} \rangle = V$, d.h. B erzeugt V*

gelten.

Beispiel 3.2.16. (a) Die Menge $\{(1,0)^T, (0,1)^T\} \subseteq \mathbb{R}^2$ aus Beispiel 3.2.12 (b) ist eine Basis von $\mathbb{R}^2$, denn erstens ist sie nach diesem Beipiel linear unabhängig und zweitens gilt für alle $(\alpha, \beta) \in \mathbb{R}^2$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

(b) Genauso ist auch die Teilmenge von $\mathbb{R}^n$, die gegeben ist durch

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right\},$$

eine Basis des $\mathbb{R}^n$, die sogenannte *Standardbasis*.

(c) Im Raum c_{00} der endlichen Folgen betrachten wir für $k \in \mathbb{N}$ die Vektoren $e^{(k)} = (\delta_{jk})_{j\in\mathbb{N}}$, vgl. Beispiel 3.1.2 (e). Dann ist $\mathcal{B} := \{e^{(k)} : k \in \mathbb{N}\}$ eine Basis von c_{00}. Um das einzusehen, beobachten wir zunächst, dass man jede endliche Folge $(\alpha_1, \alpha_2, \dots, \alpha_n, 0, 0, \dots)$ als Linearkombination

$$(\alpha_1, \alpha_2, \dots, \alpha_n, 0, 0, \dots) = \sum_{k=1}^{n} \alpha_k e^{(k)}$$

dieser speziellen Folgen schreiben kann. Zum Nachweis der lineaeren Unabhängigkeit seien $n \in \mathbb{N}$, sowie $k_1, k_2, \dots, k_n \in \mathbb{N}$ mit $k_1 < k_2 < k_3 < \dots < k_n$ paarweise verschiedene Indizes und $\alpha_1, \dots, \alpha_n \in K$ so, dass

$$\sum_{\ell=1}^{n} \alpha_\ell e^{(k_\ell)} = 0 = (0)_{j\in\mathbb{N}}$$

gilt. Das heißt

$$(0, 0, 0, \dots) = \big(0, \dots, 0, \alpha_1, 0, \dots, 0, \alpha_2, 0 \dots\dots\dots 0, \alpha_n, 0, 0, \dots\big)$$

und wir sehen $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$.

Man beachte, dass wir hier ein Beispiel einer Basis haben, die unendlich viele Elemente hat.

(d) Der Nullraum $\{0\}$ hat immer die Basis $\emptyset$.

Den folgenden sehr nützlichen Satz wollen wir hier ohne Beweis verwenden.

Satz 3.2.17 (Basissatz, Basisergänzungssatz)**.**

(a) Jeder Vektorraum hat eine Basis.

(b) Ist V ein Vektorraum und $M \subseteq V$ linear unabhängig, so gibt es eine Basis $\mathcal{B}$ von V mit $M \subseteq \mathcal{B}$.

Eine wichtige Anwendung dieses Resultats ist die folgende Beobachtung.

Satz 3.2.18. *Sei V ein Vektorraum und $\mathcal{B}$ eine Basis von V mit $n \in \mathbb{N}$ Elementen. Dann hat jede Basis von V genau n Elemente.*

Beweis. Es sei $\mathcal{B} = \{v_1, \ldots, v_n\}$ und es sei $\mathcal{B}' = \{w_1, \ldots, w_p\}$ eine weitere Basis von V mit $p \in \mathbb{N}$ Elementen. Wir nehmen nun an, es wäre $p > n$ und betrachten die dann vorhandenen Elemente $w_1, \ldots, w_n, w_{n+1} \in \mathcal{B}'$. Da $\mathcal{B}$ eine Basis ist, müssen diese alle Linearkombinationen der Vektoren $v_1, \ldots, v_n$ sein. Das bedeutet aber nach Satz 3.2.14 (d), dass $w_1, \ldots, w_{n+1}$ linear abhängig sein müssen und wir haben einen Widerspruch, da $\mathcal{B}'$ eine Basis von V sein soll.
Nehmen wir umgekehrt $p < n$ an, so sind mit den gleichen Argumenten wie oben die $p + 1$ Elemente $v_1, \ldots, v_{p+1} \in \mathcal{B}$ Linearkombinationen der $w_1, \ldots, w_p$ und deshalb ist $\mathcal{B}$ linear abhängig, was wieder auf einen Widerspruch führt.
Damit bleibt nur $p = n$ übrig. □

Definition 3.2.19. *Es sei V ein Vektorraum. Besitzt V eine n-elementige Basis, so sagt man V hat die* Dimension n *und schreibt* $\dim(V) = n$.
Besitzt V keine endliche Basis, so nennt man V unendlichdimensional.

Wir betrachten zur Illustration noch einmal die Vektorräume aus Beispiel 3.1.2.

Beispiel 3.2.20. (a) Der Standardvektorraum K^n: Es gilt $\dim(K^n) = n$, vgl. Beispiel 3.2.16 (b).

(b) $p \times n$-Matrizen: Hier ist $\dim(K^{p\times n}) = pn$.

(c) Funktionenräume: Wir werden später im Abschnitt 3.6 sehen, dass für endliche Mengen M gilt $\dim(\mathrm{Abb}(M, K)) = |M|$.

(d) Folgenräume: Wie man an Beispiel 3.2.16 (c) sieht, ist c_{00} unendlichdimensional. Selbiges steht dann auch für den Raum aller Folgen zu vermuten.

Können Sie die hinter dieser Vermutung stehende abstrakte Aussage beweisen?

Übungsaufgabe 3.2.21. Ist $n \in \mathbb{N}^*$ und V ein n-dimensionaler Vektorraum, so ist jede linear unabhängige Teilmenge von V mit n Elementen eine Basis von V.

Satz 3.2.22. *Seien $n \in \mathbb{N}^*$, sowie V ein n-dimensionaler K-Vektorraum und $\mathcal{B} = \{b_1, \ldots, b_n\}$ eine Basis von V. Dann gibt es für jedes $v \in V$ eindeutig bestimmte $\alpha_1, \ldots, \alpha_n \in K$ mit $v = \sum_{j=1}^{n} \alpha_j b_j$.*

Beweis. Sei $v \in V$ beliebig vorgegeben. Die Existenz passender $\alpha_1, \ldots, \alpha_n \in K$ mit $v = \sum_{j=1}^{n} \alpha_j b_j$ folgt sofort daraus, dass $\mathcal{B}$ eine Basis von V ist. Zu zeigen bleibt die Eindeutigkeit. Seien also zusätzlich $\beta_1, \ldots, \beta_n \in K$, für die ebenfalls

$$\sum_{j=1}^{n} \beta_j b_j = v = \sum_{j=1}^{n} \alpha_j b_j$$

gilt. Dann ist insbesondere

$$0 = \sum_{j=1}^{n} \beta_j b_j - \sum_{j=1}^{n} \alpha_j b_j = \sum_{j=1}^{n} (\beta_j - \alpha_j) b_j.$$

Da $\mathcal{B}$ eine Basis ist, sind die Vektoren $b_1, \ldots, b_n$ linear unabhängig, so dass aus obiger Gleichheit $\alpha_j - \beta_j = 0$ und damit $\alpha_j = \beta_j$ für alle $j = 1, \ldots, n$ gilt. □

Definition 3.2.23. *Seien $n \in \mathbb{N}^*$, V ein n-dimensionaler K-Vektorraum, $\mathcal{B}$ eine Basis von V und $v \in V$. Die nach Satz 3.2.22 eindeutig bestimmten Elemente $\alpha_1, \ldots, \alpha_n \in K$ mit $v = \sum_{j=1}^{n} \alpha_j b_j$ heißen* Koordinaten von v bezüglich $\mathcal{B}$.
Weiter heißt der Vektor $(\alpha_1, \ldots, \alpha_n)^T \in K^n$ Koordinatenvektor von v bezüglich $\mathcal{B}$. *Wir werden diesen häufiger mit $\vec{v}$, oder, wenn die zugrundeliegende Basis klar betont werden soll, mit $[\vec{v}]_{\mathcal{B}}$, bezeichnen.*

Warnung 3.2.24. Der Koordinatenvektor eines Vektors liegt nur nach der Wahl einer konkreten Basis fest. Ändert man die Basis, ändern sich auch alle Koordinatenvektoren. Die Schreibweise $\vec{v}$ für den Koordinatenvektor von v ist also nur angebracht, wenn aus dem Zusammenhang vollkommen klar ist, bezüglich welcher Basis dieser zu bilden ist. In allen anderen Fällen ist die genauere Schreibweise notwendig.

Beispiel 3.2.25. (a) Im Raum $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ betrachten wir $U := \langle f_1, f_2 \rangle$, wobei $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ mit $f_1(x) = x$ und $f_2(x) = x^2$ gegeben seien. Die beiden Vektoren f_1 und f_2 sind linear unabhängig nach Beispiel 3.2.12 (c), also bilden sie eine Basis von U.

Das Element $g \in U$ mit $g(x) = 3x^2 + x$ für $x \in \mathbb{R}$ hat bezüglich dieser Basis den Koordinatenvektor $\vec{g} = (1, 3)^T \in \mathbb{R}^2$, da $g = 1 \cdot f_1 + 3 \cdot f_2$ gilt.

(b) Wir betrachten $\mathbb{R}^2$ mit der Basis, die durch die beiden Vektoren $b_1 := (1, 1)^T$ und $b_2 := (-1, 1)^T$ gegeben ist. Der Vektor $e_1 = (1, 0)^T \in \mathbb{R}^2$ hat dann bezüglich dieser Basis den Koordinatenvektor $\vec{e_1} = (1/2, -1/2)^T$, denn es gilt

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Bemerkung 3.2.26. Durch die Bestimmung von Koordinaten scheint jeder n-dimensionale K-Vektorraum in gewisser Weise mit dem Raum K^n übereinzustimmen, wenn wir jeden Vektor mit seinem Koordinatenvektor identifizieren. Diese Intuition werden wir im Abschnitt 3.6 mathematisch rigoros machen.

3.3. Der Faktorraum

Lemma 3.3.1. *Es sei V ein K-Vektorraum und U ein Untervektorraum von V. Die Relation, die für $v, w \in V$ gegeben ist durch*

$$v \sim w \iff v - w \in U,$$

ist eine Äquivalenzrelation auf V.

Beweis. **Reflexivität:** Sei $v \in V$. Dann gilt $v - v = 0 \in U$, also ist $v \sim v$.

Symmetrie: Seien $v, w \in V$ mit $v \sim w$ gegeben. Dann ist $v - w \in U$. Da U ein Untervektorraum ist, ist dann auch $w - v = (-1)(v - w) \in U$ und das bedeutet $w \sim v$.

Transitivität: Es seien $v, w, x \in V$ mit $v \sim w$ und $w \sim x$. Das bedeutet $v - w \in U$ und $w - x \in U$. Da U ein Untervektorraum ist, gilt insbesondere dann auch $v - x = (v - w) + (w - x) \in U$. Also haben wir $v \sim x$ und sind fertig. □

Nun erinnern wir uns an ein Ergebnis unserer Betrachtungen in Abschnitt 1.3.2, den Satz 1.3.12:
Ist $\sim$ eine Äquivalenzrelation auf einer Menge V, so bilden die Äquivalenzklassen $\tilde{v}$ von $v \in V$ mit

$$\tilde{v} = \{w \in V : w \sim v\}$$

eine Zerlegung von V, d.h. die Vereingung aller Äquivalenzklassen ist ganz V und je zwei verschiedene solcher Klassen sind disjunkt.

Bemerkung 3.3.2. (a) Wie sieht nun die Äquivalenzklasse eines Elements $v \in V$ bezüglich obiger Äquivalenzrelation aus? Wir überlegen uns folgendes:

$$\begin{aligned} w \in \tilde{v} &\iff w - v \in U \iff \exists u \in U : w - v = u \\ &\iff \exists u \in U : w = v + u \iff w \in v + U, \end{aligned}$$

wobei $v + U := \{v + u : u \in U\}$ ist. Anschaulich ist damit $\tilde{v}$ der um v verschobene Unterraum U.

Am besten kann man sich dies an einem eindimensionalen Unterraum U des $\mathbb{R}^2$ veranschaulichen, vgl. Beispiel 3.3.3 und Abbildung 3.1.

(b) Es ist immer $\tilde{0} = U$, denn zum Einen gilt für jedes $u \in U$ auch $u - 0 = u \in U$ und damit $u \sim 0$, d.h. $u \in \tilde{0}$. Zum Anderen ist für jedes $v \in \tilde{0}$ nach Definition $v = v - 0 \in U$.

Beispiel 3.3.3. In $V = \mathbb{R}^2$ betrachten wir

$$U = \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R} \right\}.$$

Nach obiger Bemerkung ist für jedes $v \in \mathbb{R}^2$ die Äquivalenzklasse $\tilde{v}$ in $\mathbb{R}^2/U$ gegeben durch

$$\tilde{v} = v + U = \left\{ v + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R} \right\},$$

ist also anschaulich die Gerade in Richtung U mit Aufpunkt v.
Beispielsweise ist mit $\mu = \lambda - 1$

$$\widetilde{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 1 \end{pmatrix} : \mu \in \mathbb{R} \right\} = \widetilde{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}.$$

Die Äquivalenzrelation, die durch den Unterraum U gegeben ist, identifiziert also alle Vektoren miteinander, die auf einer gemeinsamen Geraden mit Richtung U liegen.

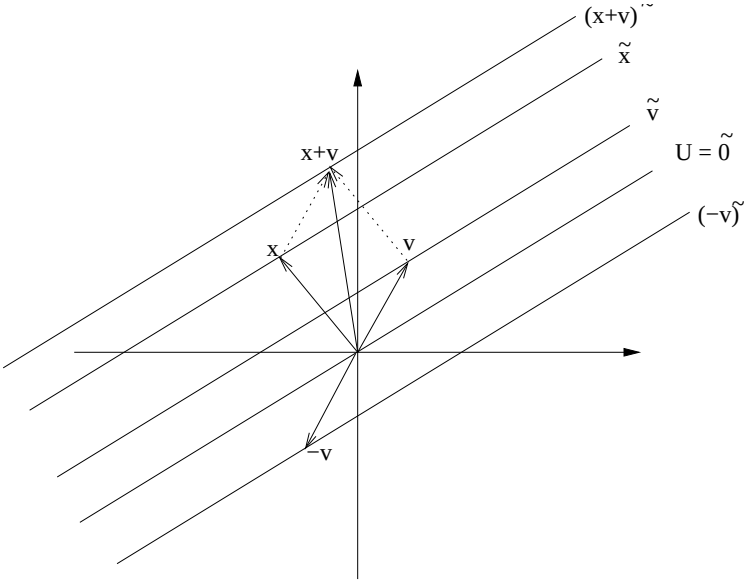

Abbildung 3.1.: Die Äquivalenzklassen in $\mathbb{R}^2/U$

Nun kommt die überraschende Wendung. Wenn wir uns die Menge der durch den obigen Prozess gegebenen Äquivalenzklassen anschauen, so können wir diese auf eine kanonische Weise selbst wieder zu einem Vektorraum machen.

Satz 3.3.4. *Es sei V ein K-Vektorraum, U ein Untervektorraum von V und $\sim$ sei definiert wie in Lemma 3.3.1. Dann ist die Menge*

$$V/U := V/_\sim = \{\tilde{v} : v \in V\}$$

mit den für $\tilde{v}, \tilde{w} \in V/U$ und $\alpha \in K$ durch

$$\tilde{v} + \tilde{w} := \widetilde{v + w} \quad \textit{und} \quad \alpha \cdot \tilde{v} := \widetilde{\alpha v}$$

definierten Verknüpfungen ein K-Vektorraum.

Beweis. Der entscheidende Punkt im Beweis ist die Wohldefiniertheit der beiden Verknüpfungen, das bedeutet in diesem Zusammenhang wir müssen Repräsentantenunabhängigkeit zeigen.
Seien also $\tilde{v}, \tilde{w} \in V/U$ und seien $v_0 \in \tilde{v}$, sowie $w_0 \in \tilde{w}$. Dann ist $v - v_0 \in U$ und $w - w_0 \in U$. Da U ein Untervektorraum ist, bedeutet das insbesondere, dass auch die Summe $v - v_0 + w - w_0$ in U ist und damit auch $(v + w) - (v_0 + w_0) \in U$ gilt. Also ist $v + w \sim v_0 + w_0$, was nach Satz 1.3.12 (b) gerade $\widetilde{v + w} = \widetilde{v_0 + w_0}$ bedeutet. Das liefert nun

$$\tilde{v} + \tilde{w} = \widetilde{v + w} = \widetilde{v_0 + w_0} = \tilde{v}_0 + \tilde{w}_0.$$

Um den entsprechenden Nachweis für die Skalar-Multiplikation zu führen, geben wir ein $\alpha \in K$ und ein $\tilde{v} \in V/U$ vor. Für jedes $v_0 \in \tilde{v}$ gilt dann wieder $v - v_0 \in U$ und da U ein Untervektorraum ist, ist auch $\alpha(v - v_0) \in U$. Das liefert $\alpha v - \alpha v_0 \in U$, was gerade $\alpha v \sim \alpha v_0$ und damit $\widetilde{\alpha v} = \widetilde{\alpha v_0}$ bedeutet. Das liefert zum Abschluss wieder

$$\alpha\tilde{v} = \widetilde{\alpha v} = \widetilde{\alpha v_0} = \alpha\tilde{v}_0.$$

Da nun die Wohldefiniertheit der Verknüpfungen gezeigt ist, müssen wir noch die Vektorraumaxiome nachweisen. Diese lassen sich jedoch ohne Probleme von V auf V/U übertragen. Der Nullvektor in V/U ist dabei $\tilde{0}$ und das Inverse $-\tilde{v}$ zu $\tilde{v} \in V/U$ ist gegeben durch $\widetilde{-v}$. Wir führen die Übertragung exemplarisch anhand der Kommutativität der Vektorraumaddition und (V4) vor.
Seien also für den Nachweis der Kommutativität $\tilde{v}, \tilde{w} \in V/U$. Dann gilt dank der Kommutativität in $(V, +)$

$$\tilde{v} + \tilde{w} = \widetilde{v + w} = \widetilde{w + v} = \tilde{w} + \tilde{v}.$$

Zum Nachweis von (V4) seien $\alpha, \beta \in K$ und $\tilde{v} \in V/U$. Dann haben wir

$$(\alpha + \beta) \cdot \tilde{v} = \widetilde{(\alpha + \beta)v} = \widetilde{\alpha v + \beta v} = \widetilde{\alpha v} + \widetilde{\beta v} = \alpha \cdot \tilde{v} + \beta \cdot \tilde{v}. \qquad \square$$

Definition 3.3.5. *Der Raum V/U in obigem Satz heißt* Faktorraum von V nach U *oder auch* Quotientenraum. *Die Schreibweise V/U wird „V (faktorisiert) nach U“ gelesen.*

Satz 3.3.6. *Sei V ein n-dimensionaler K-Vektorraum und U ein m-dimensionaler Untervektorraum von V. Dann gilt* $\dim(V/U) = n - m$.

Beweis. Nach Satz 3.2.17 hat U eine Basis, sei also $\mathcal{B} = \{b_1, \ldots, b_m\}$ eine solche. Nach dem gleichen Satz können wir diese nun zu einer Basis von V erweitern, es gibt also $b_{m+1}, b_{m+2}, \ldots, b_n \in V$, so dass $\mathcal{B}' = \{b_1, \ldots, b_n\}$ eine Basis von V ist.
Wir zeigen nun, dass $\tilde{\mathcal{B}} := \{\tilde{b}_{m+1}, \ldots, \tilde{b}_n\}$ eine Basis von V/U ist. Dazu überzeugen wir uns zunächst, dass diese Menge ganz V/U erzeugt. Sei also $\tilde{v} \in V/U$. Da

$\mathcal{B}'$ eine Basis von V ist, gibt es $\alpha_1, \ldots, \alpha_n \in K$ mit $v = \sum_{j=1}^{n} \alpha_j b_j$. Damit gilt

$$\tilde{v} = \widetilde{\sum_{j=1}^{n} \alpha_j b_j} = \sum_{j=1}^{n} \widetilde{\alpha_j b_j} = \sum_{j=1}^{n} \alpha_j \tilde{b}_j.$$

Da $b_1, \ldots, b_m \in U$ sind, gilt für alle diese $\tilde{b}_1 = \cdots = \tilde{b}_m = \tilde{0}$ und wir verbleiben mit

$$\tilde{v} = \sum_{j=m+1}^{n} \alpha_j \tilde{b}_j,$$

was gerade bedeutet, dass $\tilde{v}$ in der linearen Hülle von $\tilde{\mathcal{B}}$ liegt.
Zum Nachweis der linearen Unabhängigkeit seien $\alpha_{m+1}, \ldots, \alpha_n \in K$ gegeben mit $\sum_{j=m+1}^{n} \alpha_j \tilde{b}_j = \tilde{0}$. Dann gilt wie oben

$$\tilde{0} = \sum_{j=m+1}^{n} \alpha_j \tilde{b}_j = \widetilde{\sum_{j=m+1}^{n} \alpha_j b_j},$$

d.h. $\sum_{j=m+1}^{n} \alpha_j b_j \in U$.
Nun haben wir mit $\mathcal{B} = \{b_1, \ldots, b_m\}$ eine Basis von U, also gibt es nun Koeffizienten $\alpha_1, \ldots, \alpha_m \in K$ mit

$$\sum_{j=m+1}^{n} \alpha_j b_j = \sum_{j=1}^{m} \alpha_j b_j, \quad \text{d.h.} \quad \sum_{j=m+1}^{n} \alpha_j b_j - \sum_{j=1}^{m} \alpha_j b_j = 0.$$

Nun ist aber die Menge $\{b_1, \ldots, b_n\}$ wiederum eine Basis, diesmal von V, insbesondere sind diese n Vektoren linear unabhängig. Da wir in der letzten Gleichung aus diesen aber den Nullvektor kombiniert haben, muss $\alpha_1 = \cdots = \alpha_m = \alpha_{m+1} = \cdots = \alpha_n = 0$ gelten und wir sind fertig. □

3.4. Normierte Räume

Das Ziel dieses Abschnittes ist das Messen von Längen und Abständen in Vektorräumen. Dazu betrachten wir in diesem Abschnitt nur reelle Vektorräume. Alle Begriffe und Ergebnisse lassen sich auf komplexe Vektorräume übertragen, wobei sie allerdings zum Teil leicht angepasst werden müssen. Der Übersichtlichkeit halber wollen wir hier darauf verzichten.

Definition 3.4.1. *Es sei V ein $\mathbb{R}$-Vektorraum. Eine Abbildung $\|\cdot\| : V \to \mathbb{R}$ heißt* Norm*, falls*

(N1) $\forall v \in V : \|v\| \geq 0$ *und* $(\|v\| = 0 \iff v = 0)$. (Definitheit)

(N2) $\forall \alpha \in \mathbb{R} \, \forall v \in V : \|\alpha v\| = |\alpha| \|v\|$. (Homogenität)

(N3) $\forall v, w \in V : \|v + w\| \leq \|v\| + \|w\|$. (Dreiecksungleichung)

Ein Vektorraum mit einer Norm heißt normierter Raum.

Beispiel 3.4.2. (a) Der Betrag $|\cdot|$ in $\mathbb{R}$ ist eine Norm.

Das ergibt sich aus den Eigenschaften des Betrags in $\mathbb{C}$ in Satz 2.5.12.

(b) Unser Alltagsbegriff von Länge ist der Euklidische Abstand, der z.B. in der Ebene $\mathbb{R}^2$ gegeben ist durch die *Euklidische Norm* oder auch 2-*Norm*

$$\|x\|_2 := \sqrt{x_1^2 + x_2^2}, \qquad x = (x_1, x_2)^T \in \mathbb{R}^2,$$

oder allgemein in $\mathbb{R}^n$ durch

$$\|x\|_2 := \sqrt{\sum_{j=1}^{n} x_j^2}, \qquad x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n.$$

Beim Nachweis, dass dies eine Norm ist, sind die Nachweise von (N1) und (N2) einfach, dagegen stellt sich der Nachweis von (N3) auf direktem Wege als sehr sperrig und rechenintensiv heraus. Versuchen Sie sich ruhig einmal ein bisschen daran. Dass $\|\cdot\|_2$ eine Norm auf $\mathbb{R}^n$ ist, werden wir in Kürze aus einem deutlich allgemeineren Resultat geschenkt bekommen. Wir können also auf die lange Rechnung verzichten.

(c) Es gibt in $\mathbb{R}^n$ aber auch noch andere Normen. Wir betrachten in $\mathbb{R}^2$ die Abbildung $\|\cdot\|_1 : \mathbb{R}^2 \to \mathbb{R}$ mit

$$\|x\|_1 = |x_1| + |x_2|, \qquad x = (x_1, x_2)^T \in \mathbb{R}^2,$$

die sogenannte 1-*Norm*. Diese ist eine Norm, denn

(N1) Für jedes $x = (x_1, x_2)^T \in \mathbb{R}^2$ ist $\|x\|_1 = |x_1| + |x_2| \geq 0$ und ist $x \neq 0$, so muss $x_1 \neq 0$ oder $x_2 \neq 0$ gelten. Also ist in diesem Fall $|x_1| > 0$ oder $|x_2| > 0$ und wir erhalten $\|x\|_1 > 0$.

Damit folgt die Definitheit per Kontraposition.

(N2) Seien $\alpha \in \mathbb{R}$ und $x \in \mathbb{R}^2$. Dann gilt

$$\|\alpha x\|_1 = \|(\alpha x_1, \alpha x_2)^T\|_1 = |\alpha x_1| + |\alpha x_2| = |\alpha|(|x_1| + |x_2|) = |\alpha|\|x\|_1.$$

(N3) Seien $x, y \in \mathbb{R}^2$. Dann gilt mit Hilfe von (a)

$$\begin{aligned}\|x + y\|_1 = \|(x_1 + y_1, x_2 + y_2)^T\|_1 &= |x_1 + y_1| + |x_2 + y_2| \\ &\leq |x_1| + |y_1| + |x_2| + |y_2| = \|x\|_1 + \|y\|_1.\end{aligned}$$

Auch die 1-Norm gibt es für jedes $n \in \mathbb{N}^*$ in $\mathbb{R}^n$:

$$\|x\|_1 = \sum_{j=1}^{n} |x_j|, \qquad x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n.$$

(d) Ein weiteres Beispiel, das als Übungsaufgabe verbleibt, ist die *Maximums-* oder ∞*-Norm* in $\mathbb{R}^n$ gegeben durch

$$\|x\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}, \qquad x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n.$$

Hat man in einem $\mathbb{R}$-Vektorraum V eine Norm $\|\cdot\|$, so kann man damit Abstände messen. Die entsprechenden Begriffe liefert die folgende Definition.

Definition 3.4.3. *Es sei V ein $\mathbb{R}$-Vektorraum mit einer Norm $\|\cdot\|$ und A und B seien nicht-leere Teilmengen von V. Dann heißt*

$$\operatorname{dist}(A, B) := \inf\{\|a - b\| : a \in A,\ b \in B\}$$

Abstand *von A und B. Sind $A = \{a\}$ und/oder $B = \{b\}$ einelementig, so schreibt man auch* $\operatorname{dist}(a, B)$ *oder* $\operatorname{dist}(a, b)$ *statt* $\operatorname{dist}(\{a\}, B)$*, bzw.* $\operatorname{dist}(\{a\}, \{b\})$.

Bemerkung 3.4.4. Der Abstand zwischen zwei Vektoren u und v aus V ist damit gegeben durch $\|u - v\|$.
Nimmt man die euklidische Norm in $\mathbb{R}^2$ oder $\mathbb{R}^3$, so stimmt dieser Abstandsbegriff mit unserem alltäglich gemessenen Abstand überein.
Man beachte auch, dass dank der Homogenität der Norm die für einen Abstand recht sinnige Eigenschaft

$$\operatorname{dist}(u, v) = \|u - v\| = \|(-1)(v - u)\| = |-1|\|v - u\| = \|v - u\| = \operatorname{dist}(v, u)$$

gilt.

Definition 3.4.5. *Es sei V ein $\mathbb{R}$-Vektorraum. Eine Abbildung $(\cdot|\cdot) : V \times V \to \mathbb{R}$ heißt* Skalarprodukt*, falls*

(SP1) $\forall x \in V : (x|x) \geq 0$ *und* $\big((x|x) = 0 \iff x = 0\big)$. (Definitheit)

(SP2) $\forall x, y \in V : (x|y) = (y|x)$. (Symmetrie)

(SP3) $\forall x, y, z \in V\ \forall \alpha, \beta \in \mathbb{R} : (\alpha x + \beta y|z) = \alpha(x|z) + \beta(y|z)$.
(Linearität im ersten Argument)

Bemerkung 3.4.6. Aus (SP3) und (SP2) folgt wegen

$$(x|\alpha y + \beta z) = (\alpha y + \beta z|x) = \alpha(y|x) + \beta(z|x) = \alpha(x|y) + \beta(x|z)$$

für alle $x, y, z \in V$ und alle $\alpha, \beta \in \mathbb{R}$ auch die Linearität im zweiten Argument.

Beispiel 3.4.7. In $\mathbb{R}^n$ erhält man ein Skalarprodukt, das sogenannte *Standardskalarprodukt*, wenn man für $x = (x_1, x_2, \ldots, x_n)^T$ und $y = (y_1, y_2, \ldots, y_n)^T$ aus $\mathbb{R}^n$ setzt

$$(x|y) := \sum_{j=1}^{n} x_j y_j.$$

Das sieht man folgendermaßen:

(SP1) Für jedes $x \in \mathbb{R}^n$ gilt $(x|x) = \sum_{j=1}^{n} x_j^2 \geq 0$. Weiter ist offensichtlich $(0|0) = 0$. Ist schließlich $x \neq 0$ so gibt es einen Index j_0 mit $x_{j_0} \neq 0$ und es ist

$$(x|x) = \sum_{j=1}^{n} x_j^2 \geq x_{j_0}^2 > 0,$$

also insbesondere $(x|x) \neq 0$ in diesem Fall. Damit folgt die Definitheit per Kontraposition.

(SP2) Seien $x, y \in \mathbb{R}^n$. Dann gilt

$$(x|y) = \sum_{j=1}^{n} x_j y_j = \sum_{j=1}^{n} y_j x_j = (y|x).$$

(SP3) Seien $x, y, z \in \mathbb{R}^n$ und $\alpha, \beta \in \mathbb{R}$. Dann ist

$$\begin{aligned}(\alpha x + \beta y|z) &= \sum_{j=1}^{n} (\alpha x_j + \beta y_j) z_j = \sum_{j=1}^{n} (\alpha x_j z_j + \beta y_j z_j) \\ &= \alpha \sum_{j=1}^{n} x_j z_j + \beta \sum_{j=1}^{n} y_j z_j = \alpha(x|z) + \beta(y|z).\end{aligned}$$

Übungsaufgabe 3.4.8. Sei $(\cdot|\cdot)$ ein Skalarprodukt auf einem $\mathbb{R}$-Vektorraum V. Dann gilt $(x|0) = (0|x) = 0$ für alle $x \in V$.

Was hat nun ein Skalarprodukt mit Abständen und Normen zu tun? Eine ganze Menge. Hat man ein Skalarprodukt $(\cdot|\cdot)$ auf einem $\mathbb{R}$-Vektorraum V, so werden wir nun zeigen, dass dann durch

$$\|x\| := \sqrt{(x|x)}, \quad x \in V, \tag{3.1}$$

eine Norm definiert wird. Man beachte zunächst, dass dank der Definitheit des Skalarprodukts diese Setzung wohldefiniert ist, da das Argument der Wurzel nie negativ werden kann.
Zum Nachweis, dass wir so wirklich eine Norm bekommen, benötigen wir zunächst die folgende Ungleichung.

Satz 3.4.9 (Cauchy-Schwarz-Ungleichung)**.** *Es sei $(\cdot|\cdot)$ ein Skalarprodukt auf einem $\mathbb{R}$-Vektorraum V und $\|\cdot\|$ definiert wie in* (3.1)*. Dann gilt für alle $v, w \in V$*

$$\big|(v|w)\big| \leq \|v\| \cdot \|w\|$$

und Gleichheit gilt genau dann, wenn v und w linear abhängig sind.

Beweis. Wir betrachten zunächst den Fall $w = 0$. Dann ist $(w|w) = 0$ und damit auch $\|w\| = 0$. Aus Übungaufgabe 3.4.8 folgt dann

$$(v|w) = 0 = \|v\| \cdot \|w\|$$

und wir können diesen Fall zu den Akten legen.
Sei also nun $w \neq 0$. Dann gilt für alle $\alpha \in \mathbb{R}$ dank der Definitheit aus (SP1)

$$\begin{aligned} 0 \leq (v - \alpha w|v - \alpha w) &\overset{(SP3)}{=} (v|v - \alpha w) - \alpha(w|v - \alpha w) \\ &\overset{3.4.6}{=} (v|v) - \alpha(v|w) - \alpha(w|v) + \alpha^2(w|w) \overset{(SP2)}{=} (v|v) - 2\alpha(v|w) + \alpha^2(w|w). \end{aligned}$$

Da $w \neq 0$ ist, haben wir mit (SP1) $(w|w) \neq 0$ und können nun $\alpha := (v|w)/(w|w)$ setzen. Damit erhalten wir aus obiger Rechnung

$$\begin{aligned} 0 \leq (v|v) - 2\frac{(v|w)}{(w|w)}(v|w) + \frac{(v|w)^2}{(w|w)^2}(w|w) &= (v|v) - 2\frac{(v|w)^2}{(w|w)} + \frac{(v|w)^2}{(w|w)} \\ &= (v|v) - \frac{(v|w)^2}{(w|w)}. \end{aligned}$$

Da $(w|w) > 0$ ist, können wir die Ungleichung mit diesem Wert multiplizieren, ohne dass sich das Relationszeichen umdreht. Das liefert

$$0 \leq (v|v)(w|w) - (v|w)^2.$$

Also ist

$$\big|(v|w)\big|^2 = (v|w)^2 \leq (v|v)(w|w) = \|v\|^2 \cdot \|w\|^2,$$

woraus die behauptete Ungleichung folgt.
Abschließend müssen wir uns noch überlegen, wann Gleichheit gilt. Nach unserer Rechnung gilt Gleichheit genau dann, wenn $(v - \alpha w|v - \alpha w) = 0$ ist. Wegen der Definitheit des Skalarprodukts gilt das aber genau dann, wenn $v - \alpha w = 0$, d.h. $v = \alpha w$ ist. Womit wir bei der linearen Abhängigkeit von v und w sind. □

Satz 3.4.10. *Sei $(\cdot|\cdot)$ ein Skalarprodukt auf einem $\mathbb{R}$-Vektorraum V. Dann ist $\|\cdot\|$ definiert wie in* (3.1) *eine Norm auf V.*

Beweis. Wir rechnen die drei Norm-Axiome nach.

(N1) Zunächst ist $\|v\| = \sqrt{(v|v)} \geq 0$ für jedes $v \in V$ und es gilt $\|0\| = \sqrt{(0|0)} = \sqrt{0} = 0$. Außerdem folgt aus $\|v\| = 0$ auch $\sqrt{(v|v)} = 0$, d.h. $(v|v) = 0$ und das liefert uns mit (SP1) nun $v = 0$.

(N2) Seien $\alpha \in \mathbb{R}$ und $v \in V$. Dann gilt dank der Linearität des Skalarprodukts in beiden Variablen

$$\|\alpha v\| = \sqrt{(\alpha v|\alpha v)} = \sqrt{\alpha^2(v|v)} = \sqrt{\alpha^2}\sqrt{(v|v)} = |\alpha|\|v\|.$$

(N3) Seien $u, v \in V$. Dann gilt wieder mit der Linearität und dieses Mal zusätzlich der Symmetrie des Skalarprodukts

$$\|u+v\|^2 = (u+v|u+v) = (u|u)+(u|v)+(v|u)+(v|v) = \|u\|^2+2(u|v)+\|v\|^2.$$

Mit der Cauchy-Schwarz-Ungleichung liefert das nun

$$\|u+v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = \big(\|u\| + \|v\|\big)^2,$$

woraus $\|u+v\| \leq \|u\| + \|v\|$ folgt. □

Korollar 3.4.11. *Die Euklidische Norm $\|\cdot\|_2$ auf $\mathbb{R}^n$, vgl. Beispiel 3.4.2 (b), ist tatsächlich eine Norm.*

Beweis. Für das Standardskalarprodukt auf $\mathbb{R}^n$, vgl. Beispiel 3.4.7, gilt mit $x \in \mathbb{R}^n$

$$\sqrt{(x|x)} = \sqrt{\sum_{j=1}^{n} x_j^2} = \|x\|_2.$$

Also folgt die Behauptung aus Satz 3.4.10. □

Definition 3.4.12. *Sei V ein $\mathbb{R}$-Vektorraum mit einem Skalarprodukt $(\cdot|\cdot)$.*

(a) Zwei Vektoren $v, w \in V$ heißen senkrecht *oder* orthogonal*, falls $(v|w) = 0$ ist. Man schreibt dann $v \perp w$.*

(b) Eine Basis $\mathcal{B}$ von V heißt Orthogonalbasis*, falls für alle Wahlen von $b_1, b_2 \in \mathcal{B}$ mit $b_1 \neq b_2$ gilt $b_1 \perp b_2$.*

(c) Eine Orthogonalbasis $\mathcal{B}$ von V heißt Orthonormalbasis*, falls $\|b\| = 1$ für alle $b \in \mathcal{B}$ gilt.*

Beispiel 3.4.13. (a) Die Standardbasis ist eine Orthonormalbasis von $\mathbb{R}^n$ mit dem Standardskalarprodukt.

(b) Die Menge $\{\binom{1}{-1}, \binom{1}{1}\}$ ist eine Orthogonalbasis von $\mathbb{R}^2$ mit dem Standardskalarprodukt, aber keine Orthonormalbasis.

Den folgenden Basisergänzungssatz für Orthonormalbasen wollen wir wieder ohne Beweis stehen lassen.

Satz 3.4.14. *Jeder $\mathbb{R}$-Vektorraum mit Skalarprodukt hat eine Orthonormalbasis und jede Menge von normierten und paarweise orthogonalen Vektoren lässt sich zu einer Orthonormalbasis ergänzen.*

Bemerkung 3.4.15. Sei V ein n-dimensionaler $\mathbb{R}$-Vektorraum mit Skalarprodukt $(\cdot|\cdot)$ und $\{e_1, e_2, \ldots, e_n\}$ eine Orthonormalbasis von V. Bezüglich dieser Basis gibt es dann zu einem gegebenen $v \in V$ den Koordinatenvektor

$$\vec{v} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad \text{wobei } v = \sum_{j=1}^{n} \alpha_j e_j$$

ist, vgl. Definition 3.2.23. Das Problem ist, wie bestimmt man ganz konkret diesen Koordinatenvektor?
Bei einer Orthonormalbasis ist das zum Glück einfach. Wir multiplizieren für ein $k \in \{1, 2, \ldots, n\}$ obige Gleichung mit e_k und erhalten dank der Linearität des Skalarprodukts

$$(v|e_k) = \Big(\sum_{j=1}^{n} \alpha_j e_j \Big| e_k\Big) = \sum_{j=1}^{n} \alpha_j (e_j|e_k) = \sum_{j=1}^{n} \alpha_j \delta_{jk} = \alpha_k,$$

wobei wir wieder das Kronecker-Delta, vgl. Bemerkung 3.1.2 (e), verwendet haben.
Zusammen gilt also

$$\vec{v} = \begin{pmatrix} (v|e_1) \\ (v|e_2) \\ \vdots \\ (v|e_n) \end{pmatrix}.$$

Man beachte, dass diese Formel *nur für Orthonormalbasen* gilt!

Übungsaufgabe 3.4.16. (a) Es sei V ein n-dimensionaler $\mathbb{R}$-Vektorraum mit Skalarprodukt und U sei ein Untervektorraum von V. Dann gibt es zu jedem $v \in V$ genau ein $u \in U$, so dass $v - u$ auf allen Vektoren aus U senkrecht steht. Man nennt u die *Orthogonalprojektion* von v auf U.

(b) Ist $B = \{e_1, e_2, \ldots, e_n\}$ eine Orthonormalbasis von V, so dass $\{e_1, e_2, \ldots, e_k\}$ für ein $k \in \{1, \ldots, n\}$ eine Orthonormalbasis von U ist, so berechnet sich die Orthogonalprojektion von v auf U als

$$\sum_{j=1}^{k} (v|e_j) e_j.$$

3.5. Geometrie im $\mathbb{R}^n$

Der uns vertraute Raum und die euklidische Ebene lassen sich leicht mit $\mathbb{R}^3$ bzw. $\mathbb{R}^2$ identifizieren und tragen damit eine Vektorraum-Struktur. Wie diese dazu dienen kann, elementargeometrische Betrachtungen anzustellen, wollen wir in diesem Abschnitt ein wenig anreißen.
Dabei betrachten wir hier durchgängig den reellen Standardvektorraum $\mathbb{R}^n$ mit dem Standardskalarprodukt.

Definition 3.5.1. *(a) Es seien $x, v \in \mathbb{R}^n$ mit $v \neq 0$. Dann heißt*

$$g := \{x + \lambda v : \lambda \in \mathbb{R}\}$$

eine Gerade *mit* Aufpunkt x *und* Richtungsvektor v.

(b) Seien $x, v, w \in \mathbb{R}^n$ und seien v und w linear unabhängig. Dann heißt

$$E := \{x + \lambda v + \mu w : \lambda, \mu \in \mathbb{R}\}$$

Ebene *mit* Aufpunkt x *und* Richtungsvektoren v *und* w.

Bemerkung 3.5.2. Man kann Geraden und Ebenen als um den Aufpunkt verschobene Untervektorräume auffassen. Mit den obigen Notationen also

$$g = x + \langle v \rangle \qquad \text{und} \qquad E = x + \langle v, w \rangle.$$

Ein solcher verschobener Untervektorraum wird auch *affiner Raum* genannt.

Oft werden Geraden durch die Angabe von zwei Punkten angegeben, durch die die Gerade geht. Dass das ein sinnvolles Vorgehen ist, zeigt der folgende Satz.

Satz 3.5.3. *Seien $x, y \in \mathbb{R}^n$ mit $x \neq y$ gegeben. Dann existiert genau eine Gerade g mit $x, y \in g$, nämlich $g = \{x + \lambda(y - x) : \lambda \in \mathbb{R}\}$.*

Beweis. Zunächst erhalten wir mit den speziellen Wahlen $\lambda = 0$, bzw. $\lambda = 1$, dass $x, y \in g$ gilt.
Sei nun $h = \{u + \mu v : \mu \in \mathbb{R}\}$ eine andere Gerade mit $x, y \in h$. Dann gibt es also ein $\mu_x \in \mathbb{R}$ mit $u + \mu_x v = x$ und ein $\mu_y \in \mathbb{R}$, so dass $u + \mu_y v = y$ gilt. Außerdem ist $\mu_x \neq \mu_y$, denn es ist ja $x \neq y$.
Insbesondere haben wir damit $x - \mu_x v = u = y - \mu_y v$, d.h. $x - y = (\mu_x - \mu_y)v$. Damit folgt

$$u = x - \mu_x v = x - \frac{\mu_x}{\mu_x - \mu_y}(x - y)$$

und das liefert schließlich

$$\begin{aligned} h = \{u + \mu v : \mu \in \mathbb{R}\} &= \left\{x - \frac{\mu_x}{\mu_x - \mu_y}(x - y) + \frac{\mu}{\mu_x - \mu_y}(x - y) : \mu \in \mathbb{R}\right\} \\ &= \left\{x + \frac{\mu - \mu_x}{\mu_x - \mu_y}(x - y) : \mu \in \mathbb{R}\right\} = \{x + \lambda(x - y) : \lambda \in \mathbb{R}\} = g. \end{aligned}$$

Dabei haben wir im letzten Schritt $\lambda := (\mu - \mu_x)/(\mu_x - \mu_y)$ gesetzt. Man beachte dabei, dass $\mu \mapsto (\mu - \mu_x)/(\mu_x - \mu_y)$ eine Bijektion von $\mathbb{R}$ nach $\mathbb{R}$ ist. □

Bemerkung 3.5.4. In gleicher Weise liegt eine Ebene durch die Angabe von drei Punkten $x, y, z \in \mathbb{R}^n$, die nicht auf einer Geraden liegen, fest. Die Ebene ist dann gegeben durch

$$\{x + \lambda(y - x) + \mu(z - x) : \lambda, \mu \in \mathbb{R}\}.$$

Die Bedingung, dass die drei Punkte nicht auf einer Geraden liegen dürfen, sorgt dann dafür, dass die beiden Richtungsvektoren $y-x$ und $z-x$ linear unabhängig sind.

Beispiel 3.5.5. In $\mathbb{R}^3$ sei g die Gerade, die die Punkte $(3, -4, 2)^T$ und $(3, 2, 4)^T$ enthält, sowie E die durch

$$E := \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} : \lambda, \mu \in \mathbb{R} \right\}$$

gegebene Ebene. Wir bestimmen die Schnittmenge $g \cap E$.
Nach Satz 3.5.3 ist

$$g = \left\{ \begin{pmatrix} 3 \\ -4 \\ 2 \end{pmatrix} + \gamma \left[\begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix} - \begin{pmatrix} 3 \\ -4 \\ 2 \end{pmatrix} \right] : \gamma \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 3 \\ -4 \\ 2 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 6 \\ 2 \end{pmatrix} : \gamma \in \mathbb{R} \right\}$$

Damit ein Punkt $x \in g \cap E$ existieren kann, muss es $\lambda, \mu, \gamma \in \mathbb{R}$ geben mit

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ 2 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 6 \\ 2 \end{pmatrix}.$$

Das liefert uns das lineare Gleichungssystem

$$\left\{ \begin{array}{rcrcrcr} \lambda & + & \mu & & & = & 2 \\ -\lambda & & & - & 6\gamma & = & -4 \\ \lambda & + & 2\mu & - & 2\gamma & = & 2. \end{array} \right.$$

Die erste Zeile liefert uns $\mu = 2 - \lambda$ und die zweite verrät $\gamma = (-4 + \lambda)/(-6) = 2/3 - 1/6 \cdot \lambda$. Setzt man diese beiden Informationen in die dritte Zeile ein, so erhält man

$$\lambda + 2(2 - \lambda) - 2\Big(\frac{2}{3} - \frac{1}{6}\lambda\Big) = 2 \iff -\frac{2}{3}\lambda + \frac{8}{3} = 2 \iff -\frac{2}{3}\lambda = -\frac{2}{3} \iff \lambda = 1.$$

Das bedeutet abschließend $\mu = 1$ und $\gamma = 1/2$.
Da das Gleichungssystem eine eindeutige Loesung hat, gibt es genau einen Punkt $x \in g \cap E$, und zwar

$$x = \begin{pmatrix} 3 \\ -4 \\ 2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix}.$$

Definition 3.5.6. *Sei U ein $(n-1)$-dimensionaler Untervektorraum des $\mathbb{R}^n$ und $x \in \mathbb{R}^n$. Dann nennt man den affinen Raum $x+U$ eine* Hyperebene *in $\mathbb{R}^n$.*

Beispiel 3.5.7. Geraden sind Hyperebenen in $\mathbb{R}^2$ und Ebenen sind Hyperebenen in $\mathbb{R}^3$. In $\mathbb{R}^4$ wäre z.B.

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} : \lambda, \mu, \gamma \in \mathbb{R} \right\}$$

eine Hyperebene.

Übungsaufgabe 3.5.8. Ist $H = x+U$ eine Hyperebene und $y \in H$ ein Punkt in H, so gilt auch $H = y+U$. Anders formuliert, der Untervektorraum, der H aufspannt, hängt nicht von der speziellen Wahl des Aufpunkts ab.

Satz 3.5.9. *Es sei $H = x+U$ eine Hyperebene in $\mathbb{R}^n$. Dann existiert ein bis auf sein Vorzeichen eindeutiger Vektor $\nu \in \mathbb{R}^n$ mit $\|\nu\|_2 = 1$ und $\nu \perp u$ für alle $u \in U$, ein sogenannter* Normaleneinheitsvektor *von H.*

Beweis. Es sei $\mathcal{B}' = \{e_1, e_2, \ldots, e_{n-1}\}$ eine Orthonormalbasis von U und wir wählen $\nu \in \mathbb{R}^n$ so, dass dieser Vektor die Menge $\mathcal{B}'$ zu einer Orthonormalbasis $\mathcal{B} = \{e_1, e_2, \ldots, e_{n-1}, \nu\}$ von $\mathbb{R}^n$ ergänzt, vgl. Satz 3.4.14. Dann gilt nach Konstruktion $\|\nu\|_2 = 1$ und $\nu \perp e_j$ für jedes $j \in \{1, 2, \ldots, n-1\}$. Ist $u \in U$, so ist aber u eine Linearkombination von $e_1, e_2, \ldots, e_{n-1}$, also ist auch $\nu \perp u$.
Es bleibt die Eindeutigkeit (bis auf ein Vorzeichen) zu zeigen. Sei dazu ein $v \in \mathbb{R}^n$ mit $\|v\|_2 = 1$ und $v \perp u$ für jedes $u \in U$ gegeben. Nun ist $\mathcal{B}$ eine Orthonormalbasis, also gilt nach Bemerkung 3.4.15 und weil $v \perp e_j$ für alle $j = 1, 2, \ldots, n-1$ gilt

$$v = \sum_{j=1}^{n-1} (v|e_j)e_j + (v|\nu)\nu = (v|\nu)\nu.$$

Der Vektor v ist also ein Vielfaches des Vektors ν. Da aber beide Länge Eins haben sollen, muss $v = \nu$ oder $v = -\nu$ gelten und wir sind fertig. □

Satz 3.5.10. *Es sei H eine Hyperebene in $\mathbb{R}^n$ mit Normaleneinheitsvektor ν und es sei $x_0 \in H$. Dann gilt für $d := (x_0|\nu)$*

$$H = \{x \in \mathbb{R}^n : (x|\nu) = d\}.$$

Beweis. Sei U der Untervektorraum von $\mathbb{R}^n$, mit dem $H = x_0 + U$ gilt.

„⊆“ Es sei $x \in H$. Dann gibt es ein $u \in U$ mit $x = x_0 + u$. Also ist dank der Linearität des Skalarprodukts und mit Hilfe der Definition des Normaleneinheitsvektors

$$(x|\nu) = (x_0 + u|\nu) = (x_0|\nu) + (u|\nu) = (x_0|\nu) = d.$$

„⊇“ Sei $x \in \mathbb{R}^n$ so, dass $(x|\nu) = d = (x_0|\nu)$ gilt. Dann ist $(x - x_0|\nu) = 0$, d.h. $x - x_0 \perp \nu$. Ist wieder $\{e_1, e_2, \ldots, e_{n-1}\}$ eine Orthonormalbasis von U, so ist $\{e_1, e_2, \ldots, e_{n-1}, \nu\}$ eine Orthonormalbasis von $\mathbb{R}^n$ und wir haben mit Hilfe von Bemerkung 3.4.15

$$x - x_0 = \sum_{j=1}^{n-1} (x - x_0|e_j)e_j + (x - x_0|\nu)\nu = \sum_{j=1}^{n-1} (x - x_0|e_j)e_j.$$

Das bedeutet aber, dass $x - x_0 \in U$ liegt, d.h. $x \in x_0 + U = H$. □

Definition 3.5.11. *Die Darstellung $H = \{x \in \mathbb{R}^n : (x|\nu) = d\}$ für eine Hyperebene H mit Normaleneinheitsvektor ν heißt* Hesse-Normalform *von H.*

Mit Hilfe der Hesse-Normalform ist die Bestimmung des Abstandes von der Hyperebene recht einfach, wie der folgende Satz zeigt. Zum Begriff des Abstandes sei an Definition 3.4.3 erinnert.

Satz 3.5.12. *Es sei $H \subseteq \mathbb{R}^n$ eine Hyperebene mit Hesse-Normalform $H = \{x \in \mathbb{R}^n : (x|\nu) = d\}$ und $x_0 \in \mathbb{R}^n$. Dann gilt*

$$\operatorname{dist}(x_0, H) = \big|(x_0|\nu) - d\big|.$$

Insbesondere ist $\operatorname{dist}(0, H) = |d|$.

Beweis. Zunächst zeigen wir, dass $y_0 := x_0 - \big((x_0|\nu) - d\big)\nu$ in H liegt. Dazu rechnen wir

$$\begin{aligned}(y_0|\nu) &= \big(x_0 - ((x_0|\nu) - d)\nu\big|\nu\big) = (x_0|\nu) - \big((x_0|\nu) - d\big)(\nu|\nu)\\ &= (x_0|\nu) - (x_0|\nu) + d = d.\end{aligned}$$

Also ist $y_0 \in H$ und wir bekommen

$$\begin{aligned}\operatorname{dist}(x_0, H) &= \inf\{\|x_0 - y\|_2 : y \in H\} \le \|x_0 - y_0\|_2 = \big\|\big((x_0|\nu) - d\big)\nu\big\|\\ &= \big|(x_0|\nu) - d\big|\|\nu\|_2 = \big|(x_0|\nu) - d\big|.\end{aligned}$$

Es bleibt die umgekehrte Ungleichung zu zeigen. Sei dazu U der Untervektorraum von V mit $H = y_0 + U$. Ist nun $z \in H$ beliebig, so gilt $y_0 - z \in U$ und wir haben

$$(x_0 - y_0|y_0 - z) = \big(((x_0|\nu) - d)\nu\big|y_0 - z\big) = \big((x_0|\nu) - d\big)(\nu|y_0 - z) = 0,$$

da $\nu \perp u$ für alle $u \in U$ gilt. Das liefert nun

$$\begin{aligned}\|x_0 - z\|_2^2 &= (x_0 - z|x_0 - z) = (x_0 - y_0 + y_0 - z|x_0 - y_0 + y_0 - z)\\ &= (x_0 - y_0|x_0 - y_0) + (y_0 - z|y_0 - z) + 2(x_0 - y_0|y_0 - z).\end{aligned}$$

Nun ist nach der Vorüberlegung der letzte Summand Null und die ersten beiden lassen sich als Normen schreiben, die beide positiv sind. Das ergibt

$$\|x_0 - z\|_2^2 = \|x_0 - y_0\|_2^2 + \|y_0 - z\|_2^2 \geq \|x_0 - y_0\|_2^2.$$

Also ist $\|x_0 - y_0\|_2 \leq \|x_0 - z\|_2$ für jedes $z \in H$, was uns zum Abschluss

$$\|x_0 - y_0\|_2 \leq \inf\{\|x_0 - z\|_2 : z \in H\} = \operatorname{dist}(x_0, H)$$

liefert. □

Beispiel 3.5.13. Wir wollen das obige Verfahren zur Abstandsberechnung einmal beispielhaft anwenden. Wir betrachten dazu in $\mathbb{R}^3$ die Ebene

$$E := \left\{ \begin{pmatrix}1\\0\\1\end{pmatrix} + \lambda \begin{pmatrix}1\\1\\0\end{pmatrix} + \mu \begin{pmatrix}1\\2\\2\end{pmatrix} : \lambda, \mu \in \mathbb{R} \right\} = \begin{pmatrix}1\\0\\1\end{pmatrix} + \left\langle \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\2\\2\end{pmatrix} \right\rangle.$$

und bestimmen ihren Abstand vom Punkt $(1,1,1)^T$. Dazu ermitteln wir zunächst die Hesse-Normalform von E, d.h. wir suchen einen Normaleneinheitsvektor $\nu = (\nu_1, \nu_2, \nu_3)^T \in \mathbb{R}^3$ mit $\nu \perp (1,1,0)^T$ und $\nu \perp (1,2,2)^T$, sowie $\|\nu\|_2 = 1$.
Aus der ersten Bedingung bekommen wir die Gleichung $\nu_1 + \nu_2 = 0$, d.h. $\nu_1 = -\nu_2$. Die zweite Bedingung liefert $\nu_1 + 2\nu_2 + 2\nu_3 = 0$, d.h. wir erhalten durch Einsetzen der ersten Gleichung $\nu_1 - 2\nu_1 + 2\nu_3 = 0$. Das liefert $\nu_1 = 2\nu_3$. Wir setzen nun $\nu_1 = 2$ und erhalten $\nu_2 = -2$ und $\nu_3 = 1$. Ein Vektor, der die ersten beiden Bedingungen erfüllt, ist also $\hat{\nu} := (2, -2, 1)^T$.
Dieser hat nun noch nicht Länge Eins. Wir beachten, dass mit $\hat{\nu} \perp x$ auch $\alpha\hat{\nu} \perp x$ für jedes $\alpha \in \mathbb{R}$ gilt. Auch die Vektoren $\alpha\hat{\nu}$ erfüllen also weiter die ersten beiden Bedingungen. Es reicht also $\|\hat{\nu}\|_2 = \sqrt{4+4+1} = 3$ zu bestimmen und $\hat{\nu}$ damit zu „normieren“. Dann ist

$$\nu := \frac{1}{3}\hat{\nu} = \frac{1}{3}\begin{pmatrix}2\\-2\\1\end{pmatrix}$$

der gesuchte Normaleneinheitsvektor.
Weiter ist der Vektor $x := (1,0,1)^T$ ein Element von E, womit wir

$$d = (x|\nu) = \left(\begin{pmatrix}1\\0\\1\end{pmatrix} \middle| \frac{1}{3}\begin{pmatrix}2\\-2\\1\end{pmatrix} \right) = \frac{1}{3}(2+1) = 1$$

bestimmen. Also ist die Hesse-Normalform gegeben durch

$$E = \{y \in \mathbb{R}^n : (y|\nu) = 1\} = \left\{ y = (y_1, y_2, y_3)^T \in \mathbb{R}^3 : \frac{2}{3}y_1 - \frac{2}{3}y_2 + \frac{1}{3}y_3 = 1 \right\}.$$

Damit bekommen wir nun sofort mit Satz 3.5.12

$$\operatorname{dist}\big((1,1,1)^T, E\big) = \left| \left(\begin{pmatrix}1\\1\\1\end{pmatrix} \middle| \frac{1}{3}\begin{pmatrix}2\\-2\\1\end{pmatrix} \right) - d \right| = \left| \frac{1}{3}(2-2+1) - 1 \right| = \frac{2}{3}.$$

Das mühsamste an obiger Berechnung war die Bestimmung von $\hat{\nu}$, also eines Vektors, der senkrecht auf dem Untervektorraum steht, der die Ebene aufspannt. Im für das reale Leben wichtigen Spezialfall des dreidimensionalen Raums $\mathbb{R}^3$ gibt es zum Glück eine relativ einfache Möglichkeit einen solchen Vektor zu bestimmen. Diese wollen wir zum Abschluss dieses Abschnitts noch schnell angeben.

Definition 3.5.14. *Es seien $x = (x_1, x_2, x_3)^T$ und $y = (y_1, y_2, y_3)^T$ aus $\mathbb{R}^3$. Dann heißt der Vektor*

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} := \begin{pmatrix} x_2y_3 - y_2x_3 \\ x_3y_1 - y_3x_1 \\ x_1y_2 - y_1x_2 \end{pmatrix}$$

das Kreuzprodukt *von x und y.*

Satz 3.5.15. *Sind $x, y \in \mathbb{R}^3$, so gilt $(x \times y) \perp x$ und $(x \times y) \perp y$.*

3.6. Lineare Abbildungen

Definition 3.6.1. *Es seien V und W zwei K-Vektorräume bezüglich desselben Körpers K.*

(a) Eine Abbildung $\Phi : V \to W$ heißt linear *oder* (Vektorraum-)Homomorphismus*, falls für alle $a, b \in V$ und alle $\alpha \in K$*

$$\Phi(a + b) = \Phi(a) + \Phi(b) \qquad \text{und} \qquad \Phi(\alpha a) = \alpha\Phi(a)$$

gilt.

(b) Ist Φ zusätzlich bijektiv, so heißt Φ (Vektorraum-)Isomorphismus*, und V und W werden dann als* isomorph *bezeichnet, in Zeichen: $V \cong W$.*

Die beiden Bedingungen in der Definition einer linearen Abbildung können zu einer verschmolzen werden:

Satz 3.6.2. *Seien V und W zwei K-Vektorräume. Dann ist $\Phi : V \to W$ genau dann linear, wenn für alle $a, b \in V$ und alle $\alpha, \beta \in K$ gilt*

$$\Phi(\alpha a + \beta b) = \alpha\Phi(a) + \beta\Phi(b).$$

Beweis. „⇒“ Mit den beiden Eigenschaften aus Defintion 3.6.1 (a) folgt nacheinander

$$\Phi(\alpha a + \beta b) = \Phi(\alpha a) + \Phi(\beta b) = \alpha\Phi(a) + \beta\Phi(b).$$

„⇐“ Seien $a, b \in V$. Dann gilt nach Voraussetzung mit $\alpha = \beta = 1$

$$\Phi(a + b) = \Phi(1 \cdot a + 1 \cdot b) = 1 \cdot \Phi(a) + 1 \cdot \Phi(b) = \Phi(a) + \Phi(b).$$

Sind $a \in V$ und $\alpha \in K$, so ist auf die gleiche Weise mit $b = 0$ und $\beta = 0$

$$\Phi(\alpha a) = \Phi(\alpha a + 0 \cdot 0) = \alpha\Phi(a) + 0 \cdot \Phi(0) = \alpha\Phi(a). \qquad \square$$

Übungsaufgabe 3.6.3. (a) Für jeden Vektorraumhomomorphismus $\Phi : V \to W$ gilt $\Phi(0_V) = 0_W$.

(b) Sind $\Phi : V \to W$ und $\Psi : W \to X$ lineare Abbildungen zwischen K-Vektorräumen V, W und X, so ist auch $\Psi \circ \Phi : V \to X$ linear.

(c) Ist $\Phi : V \to W$ ein Isomorphismus, so ist auch $\Phi^{-1} : W \to V$ eine lineare Abbildung, also wieder ein Isomorphismus.

Beispiel 3.6.4. (a) Für zwei beliebige K-Vektorräume V und W ist die *Nullabbildung* $\Omega : V \to W$ mit $\Omega(a) = 0_W$ für jedes $a \in V$ linear, denn für alle $a, b \in V$ und alle $\alpha, \beta \in K$ gilt

$$\Omega(\alpha a + \beta b) = 0_W = \alpha \cdot 0_W + \beta \cdot 0_W = \alpha\Omega(a) + \beta\Omega(b).$$

(b) Sei V ein K-Vektorraum und $\lambda \in K$ fest. Dann ist $\Phi_\lambda : V \to V$ mit $\Phi_\lambda(a) = \lambda a$, $a \in V$, ein Homomorphismus. Für $\lambda \neq 0$ ist das sogar ein Isomorphismus mit $\Phi_\lambda^{-1} = \Phi_{1/\lambda}$.

(c) Zu vorgegebenem $v \neq 0$ aus einem K-Vektorraum V betrachten wir die Abbildung $\Psi_v : V \to V$ mit $\Psi_v(a) = a + v$ für jedes $a \in V$. Dieses ist keine lineare Abbildung, denn es gilt $\Psi_v(0) = 0 + v = v \neq 0$.

Wir wollen uns noch ein paar sehr wichtige lineare Abbildungen im Anschauungsraum zu Gemüte führen.

Beispiel 3.6.5. Die folgenden Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^n$ sind linear:

(a) Die in Übungsaufgabe 3.4.16 definierte Orthogonalprojektion.

(b) Jede Spiegelung an einem $(n-1)$-dimensionalen Untervektorraum.

(c) Die Streckung um einen Faktor $\lambda \in \mathbb{R}$, vgl. Beispiel 3.6.4 (b)

(d) Für $n = 2$ die Drehung der Ebene um einen Winkel α und für $n = 3$ die Drehung des Raums um eine Ursprungsgerade.

(e) Außerdem alle Verkettungen der obigen, vgl. Übungsaufgabe 3.6.3 (b), also alle Drehstreckspiegelungsprojektionen,...

Um den Isomorphie-Begriff zu beleuchten, zeigen wir beispielhaft den folgenden Satz.

Satz 3.6.6. *Es sei K ein Körper und M eine Menge mit $|M| = n \in \mathbb{N}^*$. Dann gilt* $\mathrm{Abb}(M, K) \cong K^n$.

Beweis. Wir benennen $M = \{m_1, m_2, \ldots, m_n\}$ und betrachten die Abbildung

$$\Phi : \begin{cases} \mathrm{Abb}(M,K) & \to K^n \\ f & \mapsto \big(f(m_1), f(m_2), \ldots, f(m_n)\big)^T, \end{cases}$$

von der wir nun zeigen wollen, dass sie ein Isomorphismus ist.
Zum Nachweis der Linearität seien $f, g \in \mathrm{Abb}(M,K)$ und $\alpha, \beta \in K$. Dann gilt

$$\begin{aligned} \Phi(\alpha f + \beta g) &= \begin{pmatrix} (\alpha f + \beta g)(m_1) \\ (\alpha f + \beta g)(m_2) \\ \vdots \\ (\alpha f + \beta g)(m_n) \end{pmatrix} = \begin{pmatrix} \alpha f(m_1) + \beta g(m_1) \\ \alpha f(m_2) + \beta g(m_2) \\ \vdots \\ \alpha f(m_n) + \beta g(m_n) \end{pmatrix} \\ &= \alpha \begin{pmatrix} f(m_1) \\ f(m_2) \\ \vdots \\ f(m_n) \end{pmatrix} + \beta \begin{pmatrix} g(m_1) \\ g(m_2) \\ \vdots \\ g(m_n) \end{pmatrix} = \alpha\Phi(f) + \beta\Phi(g) \end{aligned}$$

und die Linearität von Φ folgt aus Satz 3.6.2.
Es bleibt also noch die Bijektivität von Φ zu zeigen. Seien dazu $f, g \in \mathrm{Abb}(M,K)$ mit $\Phi(f) = \Phi(g)$ gegeben. Dann ist

$$\big(f(m_1), f(m_2), \ldots, f(m_n)\big)^T = \big(g(m_1), g(m_2), \ldots, g(m_n)\big)^T$$

und damit $f(m_j) = g(m_j)$ für jedes $j \in \{1, 2, \ldots, n\}$. Also ist $f = g$ und wir wissen, dass Φ injektiv ist.
Sei weiter $x = (x_1, x_2, \ldots, x_n)^T \in K^n$ gegeben. Für die Funktion $f : M \to K$ mit $f(m_j) = x_j$, $j \in \{1, 2, \ldots, n\}$, gilt dann $\Phi(f) = \big(f(m_1), f(m_2), \ldots, f(m_n)\big)^T = (x_1, x_2, \ldots, x_n)^T = x$. Also ist Φ auch surjektiv und wir sind fertig. □

Übungsaufgabe 3.6.7. (a) Zeigen Sie, dass die Abbildung

$$\Phi : \begin{cases} \mathbb{R}^3 & \to \mathbb{R}^4 \\ (x_1, x_2, x_3)^T & \mapsto (x_2, x_1 + x_2, x_2 + x_3, x_3)^T \end{cases}$$

linear und injektiv, aber nicht surjektiv ist.

(b) Finden Sie eine surjektive lineare Abbildung, die nicht injektiv ist oder zeigen Sie, dass es eine solche nicht geben kann.

Die Menge der linearen Abbildungen bildet selbst wieder einen K-Vektorraum. Wir geben dieser zunächst eine Bezeichnung.

Definition 3.6.8. *Wir setzen*

$$\mathcal{L}(V,W) := \{\Phi : V \to W : \Phi \text{ linear}\}.$$

Ist $V = W$, so schreibt man auch kurz $\mathcal{L}(V)$ statt $\mathcal{L}(V,V)$.

Übungsaufgabe 3.6.9. Es seien V, W zwei K-Vektorräume. Dann ist $\mathcal{L}(V, W)$ mit der Addition und Skalar-Multiplikation aus Beispiel 3.1.2 (c) ein K-Vektorraum.

Satz 3.6.10. *Es seien V und W zwei K-Vektorräume, $n \in \mathbb{N}^*$ und $\Phi \in \mathcal{L}(V, W)$.*

(a) Sind $v_1, v_2, \ldots, v_n \in V$ linear abhängig, so sind auch $\Phi(v_1), \Phi(v_2), \ldots, \Phi(v_n)$ linear abhängig.

(b) Ist Φ injektiv und sind $v_1, v_2, \ldots, v_n \in V$ linear unabhängig, so sind auch $\Phi(v_1), \Phi(v_2), \ldots, \Phi(v_n)$ linear unabhängig.

(c) Ist Φ ein Isomorphismus und $\mathcal{B}$ eine Basis von V, so ist auch $\Phi(\mathcal{B})$ eine Basis von W. Insbesondere gilt $\dim(V) = \dim(W)$.

Beweis. (a) Es sei $\sum_{j=1}^{n} \alpha_j v_j = 0_V$ eine nichttriviale Linearkombination des Nullvektors. Dann gilt wegen Übungsaufgabe 3.6.3 (a) und der Linearität von Φ

$$0_W = \Phi(0_V) = \Phi\Big(\sum_{j=1}^{n} \alpha_j v_j\Big) = \sum_{j=1}^{n} \alpha_j \Phi(v_j)$$

und dieses ist eine nichttriviale Linearkombination des Nullvektors in W. Also sind $\Phi(v_1), \Phi(v_2), \ldots, \Phi(v_n)$ linear abhängig.

(b) Wir nehmen an, dass $\Phi(v_1), \Phi(v_2), \ldots, \Phi(v_n)$ linear abhängig sind. Dann gibt es eine nichttriviale Linearkombination $\sum_{j=1}^{n} \alpha_j \Phi(v_j) = 0_W$. Mit dieser und der Linearität von Φ gilt nun

$$\Phi(0_V) = 0_W = \sum_{j=1}^{n} \alpha_j \Phi(v_j) = \Phi\Big(\sum_{j=1}^{n} \alpha_j v_j\Big).$$

Nun ist Φ nach Voraussetzung injektiv, also haben wir $0_V = \sum_{j=1}^{n} \alpha_j v_j$, was eine nichttriviale Linearkombination des Nullvektors aus $v_1, v_2, \ldots, v_n$ wäre und damit ein Widerspruch zur linearen Unabhängigkeit dieser Vektoren.

(c) Sei Φ bijektiv und $\mathcal{B}$ eine Basis von V. Dann ist dank (b) auch $\Phi(\mathcal{B})$ eine linear unabhängige Teilmenge von W. Wir müssen noch zeigen, dass $\Phi(\mathcal{B})$ ganz W erzeugt. Sei dazu $w \in W$ beliebig vorgegeben. Da Φ surjektiv ist, gibt es ein $v \in V$ mit $\Phi(v) = w$. Weiter ist $\mathcal{B}$ eine Basis von V, also gibt es ein $m \in \mathbb{N}$ und $b_1, b_2, \ldots, b_m \in \mathcal{B}$, sowie $\alpha_1, \alpha_2, \ldots, \alpha_m \in K$ mit $v = \sum_{j=1}^{m} \alpha_j b_j$. Dann ist aber

$$w = \Phi(v) = \Phi\Big(\sum_{j=1}^{m} \alpha_j b_j\Big) = \sum_{j=1}^{m} \alpha_j \Phi(b_j),$$

was bedeutet, dass $w \in \langle \Phi(\mathcal{B}) \rangle$ ist. □

Bemerkung 3.6.11. Damit haben wir im Zusammenspiel mit Satz 3.6.6 insbesondere gezeigt, dass $\dim(\mathrm{Abb}(M,K)) = |M|$ gilt und das entsprechende Versprechen aus Beispiel 3.2.20 (c) ist eingelöst.

Übungsaufgabe 3.6.12. (a) Zeigen Sie, dass jeder n-dimensionale K-Vektorraum isomorph zu K^n ist, vgl. Bemerkung 3.2.26.

(b) Seien V und W zwei endlichdimensionale Vektorräume. Zeigen Sie, dass $V \cong W$ genau dann gilt, wenn $\dim(V) = \dim(W)$ ist.

Wir wollen nun den fundamental wichtigen Satz beweisen, dass eine lineare Abbildung durch die Angabe der Bilder der Basisvektoren festgelegt werden kann.

Satz 3.6.13. *Es seien V, W zwei K-Vektorräume und V sei n-dimensional mit einer Basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$. Für jede Wahl von $w_1, w_2, \ldots, w_n \in W$ gibt es dann genau eine lineare Abbildung $\Phi : V \to W$, für die $\Phi(b_j) = w_j$ für alle $j \in \{1, 2, \ldots, n\}$ gilt.*

Beweis. Wir zeigen zunächst, dass eine solche Abbildung Φ, wenn sie existiert, durch die Angabe der $\Phi(b_j)$, $j = 1, 2, \ldots, n$, eindeutig bestimmt ist. Sei dazu $v \in V$ beliebig. Da $\mathcal{B}$ eine Basis von V ist, gibt es dann eindeutig bestimmte Koeffizienten $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$ mit $v = \sum_{j=1}^{n} \alpha_j b_j$. Wegen der Linearität von Φ muss dann gelten

$$\Phi(v) = \Phi\Big(\sum_{j=1}^{n} \alpha_j b_j\Big) = \sum_{j=1}^{n} \alpha_j \Phi(b_j) = \sum_{j=1}^{n} \alpha_j w_j.$$

Das Bild eines jeden $v \in V$ liegt also durch die Angabe der Werte $w_1, w_2, \ldots, w_n$ bereits fest.
Obiger Eindeutigkeitsbeweis liefert nun auch gleich die Blaupause für die Konstruktion der gesuchten Abbildung. Wir definieren einfach für jedes $v \in V$ die Abbildung Φ durch

$$\Phi(v) := \sum_{j=1}^{n} \alpha_j w_j, \qquad \text{falls} \quad v = \sum_{j=1}^{n} \alpha_j b_j.$$

Wir müssen nun zeigen, dass das so definierte Φ eine lineare Abbildung ist und dass $\Phi(b_j) = w_j$ für jedes $j \in \{1, 2, \ldots, n\}$ gilt.
Zum Nachweis der Linearität seien $a, b \in V$ und $\lambda, \mu \in K$ gegeben. Mit den Koeffizienten $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n \in K$, für die $a = \sum_{j=1}^{n} \alpha_j b_j$ und $b = \sum_{j=1}^{n} \beta_j b_j$ gilt, haben wir dann nach der Definition von Φ

$$\begin{aligned}\Phi(\lambda a + \mu b) &= \Phi\Big(\lambda \sum_{j=1}^{n} \alpha_j b_j + \mu \sum_{j=1}^{n} \beta_j b_j\Big) = \Phi\Big(\sum_{j=1}^{n} (\lambda\alpha_j + \mu\beta_j) b_j\Big) \\ &= \sum_{j=1}^{n} (\lambda\alpha_j + \mu\beta_j) w_j = \lambda \sum_{j=1}^{n} \alpha_j w_j + \mu \sum_{j=1}^{n} \beta_j w_j = \lambda\Phi(a) + \mu\Phi(b).\end{aligned}$$

Schließlich ist für jedes $j \in \{1, 2, \ldots, n\}$ die Basisdarstellung von b_j gegeben durch b_j selbst, also gilt

$$\Phi(b_j) = \sum_{\substack{k=1 \\ k \neq j}}^{n} 0 \cdot w_k + 1 \cdot w_j = w_j. \qquad \square$$

Definition 3.6.14. *Es seien V, W zwei K-Vektorräume und $\Phi \in \mathcal{L}(V, W)$. Dann heißt*

$$\ker(\Phi) := \left\{ v \in V : \Phi(v) = 0_W \right\}$$

der Kern *von Φ.*

Satz 3.6.15. *Es seien V, W zwei K-Vektorräume und $\Phi : V \to W$ linear. Dann gilt*

(a) $\ker(\Phi)$ *ist ein Untervektorraum von V.*

(b) Φ *ist genau dann injektiv, wenn* $\ker(\Phi) = \{0_V\}$.

(c) $\Phi(V)$ *ist ein Untervektorraum von W, der sogenannte* Bildraum *von Φ.*

Beweis. (a) Zum Einen ist immer $0_V \in \ker(\Phi)$, also ist der Kern nicht leer, zum Anderen gilt für alle $a, b \in \ker(\Phi)$ und alle $\alpha, \beta \in K$ dank der Linearität von Φ

$$\Phi(\alpha a + \beta b) = \alpha \Phi(a) + \beta \Phi(b) = \alpha \cdot 0_W + \beta \cdot 0_W = 0_W$$

und damit auch $\alpha a + \beta b \in \ker(\Phi)$. Die Behauptung folgt damit aus dem Untervektorraumkriterium, vgl. Satz 3.2.3.

(b) Ist Φ injektiv, so kann 0_W außer 0_V kein weiteres Urbild haben und es ist $\ker(\Phi) = \{0_V\}$. Ist umgekehrt diese Mengengleichheit gegeben und haben wir $a, b \in V$ mit $\Phi(a) = \Phi(b)$, so gilt mit Hilfe der Linearität von Φ

$$\Phi(a - b) = \Phi(a) - \Phi(b) = 0_W, \quad \text{also} \quad a - b = 0_V,$$

womit $a = b$ gezeigt ist. Das bedeutet aber gerade, dass Φ injektiv ist.

(c) Zum Einen ist $\Phi(V)$ nicht leer und zum anderen gibt es für jede Wahl von $w, x \in \Phi(V)$ Vektoren $u, v \in V$ mit $\Phi(u) = w$ und $\Phi(v) = x$. Damit ist für alle $\alpha, \beta \in K$ auch

$$\alpha w + \beta x = \alpha \Phi(u) + \beta \Phi(v) = \Phi(\alpha u + \beta v) \in \Phi(V),$$

die Behauptung folgt also wieder aus dem Untervektorraumkriterium. $\square$

Definition 3.6.16. *Ist in der Situation von Satz 3.6.15 der Raum $\Phi(V)$ endlichdimensional, so heißt* $\operatorname{Rang}(\Phi) := \dim(\Phi(V))$ *der* Rang *von Φ.*

Wir haben in obigem Satz gezeigt, dass $\ker(\Phi)$ ein Untervektorraum von V ist. In Erinnerung an Abschnitt 3.3 können wir also den Faktorraum $V/\ker(\Phi)$ betrachten. Wie sieht eine Äquivalenzklasse in diesem Raum aus? Erstens gilt $\widetilde{0_V} = \ker(\Phi)$ und allgemein ist nach der Definition des Faktorraums

$$\begin{aligned} a \in \tilde{b} &\iff a - b \in \ker(\Phi) \iff \Phi(a-b) = 0_W \\ &\iff \Phi(a) - \Phi(b) = 0_W \iff \Phi(a) = \Phi(b). \end{aligned}$$

Das bedeutet, dass die Elemente einer Äquivalenzklasse $\tilde{a}$ genau die Elemente von V sind, die unter Φ das selbe Bild wie a haben. Damit ist die Abbildung

$$\tilde{\Phi} : \begin{cases} V/\ker(\Phi) & \to W \\ \tilde{v} & \mapsto \Phi(v) \end{cases} \tag{3.2}$$

wohldefiniert. Diese hat eine wesentliche Bedeutung durch den folgenden Satz.

Satz 3.6.17 (Homomorphiesatz für Vektorräume)**.** *Seien V und W zwei K-Vektorräume, $\Phi : V \to W$ linear und $\tilde{\Phi}$ wie in* (3.2) *definiert. Dann ist $\tilde{\Phi} : V/\ker(\Phi) \to \Phi(V)$ ein Isomorphismus und es gilt $\Phi = \tilde{\Phi} \circ \nu$, wobei $\nu : V \to V/\ker(\Phi)$ die kanonische Abbildung ist.*

Beweis. Wir haben schon festgestellt, dass $\tilde{\Phi}$ eine wohldefinierte Abbildung ist. Wir zeigen also Linearität von $\tilde{\Phi}$. Seien dazu $\tilde{a}, \tilde{b} \in V/\ker(\Phi)$ und $\alpha, \beta \in K$. Dann gilt

$$\tilde{\Phi}(\alpha\tilde{a} + \beta\tilde{b}) = \tilde{\Phi}(\widetilde{\alpha a + \beta b}) = \Phi(\alpha a + \beta b) = \alpha\Phi(a) + \beta\Phi(b) = \alpha\tilde{\Phi}(\tilde{a}) + \beta\tilde{\Phi}(\tilde{b}).$$

Zum Nachweis der Injektivität von $\tilde{\Phi}$ sei $\tilde{a} \in \ker(\tilde{\Phi})$. Dann gilt $\tilde{\Phi}(\tilde{a}) = 0_W$. Also ist $\Phi(a) = 0_W$, was wiederum $a \in \ker(\Phi) = \widetilde{0_V}$ impliziert und schließlich $\tilde{a} = \widetilde{0_V}$ liefert. Diese Überlegungen bedeuten $\ker(\tilde{\Phi}) = \{\widetilde{0_V}\}$ und mit Satz 3.6.15 (b) folgt die Injektivität von Φ.
Da wir den Zielbereich von $\tilde{\Phi}$ auf $\Phi(V)$ eingeschränkt haben, ist $\tilde{\Phi}$ auch surjektiv und wir haben bewiesen, dass $\tilde{\Phi} : V/\ker(\Phi) \to \Phi(V)$ ein Isomorphismus ist.
Abschließend bleibt noch zu bemerken, dass für jedes $a \in V$ gilt

$$(\tilde{\Phi} \circ \nu)(a) = \tilde{\Phi}(\nu(a)) = \tilde{\Phi}(\tilde{a}) = \Phi(a). \qquad \square$$

Korollar 3.6.18. *Seien V und W zwei K-Vektorräume, wobei V endliche Dimension habe. Ist dann $\Phi \in \mathcal{L}(V, W)$, so gilt die* Dimensionsformel

$$\mathrm{Rang}(\Phi) + \dim(\ker(\Phi)) = \dim(V).$$

Beweis. Es gilt nach dem Homomorphiesatz $V/\ker(\Phi) \cong \Phi(V)$. Also erhalten wir mit Übungsaufgabe 3.6.12 (b)

$$\dim(V/\ker(\Phi)) = \dim(\Phi(V)) = \mathrm{Rang}(\Phi).$$

Andererseits liefert Satz 3.3.6

$$\dim(V/\ker(\Phi)) = \dim(V) - \dim(\ker(\Phi))$$

und die Kombination der beiden Gleichungen liefert die Behauptung. □

Satz 3.6.19. *Es sei V ein endlichdimensionaler K-Vektorraum und $\Phi : V \to V$ eine lineare Abbildung. Dann sind die folgenden Aussagen äquivalent:*

(a) Φ *ist bijektiv.*

(b) Φ *ist injektiv.*

(c) $\ker(\Phi) = \{0\}$.

(d) $\mathrm{Rang}(\Phi) = \dim(V)$.

(e) Φ *ist surjektiv.*

Beweis. Wir zeigen (a) ⇒ (b) ⇒ (c) ⇒ (d) ⇒ (e) ⇒ (a). Dann sind alle Äquivalenzen gezeigt, denn man kommt dann von jedem Buchstaben zu jedem Buchstaben.
Die Implikation (a) ⇒ (b) ist klar und (b) ⇒ (c) findet sich in Satz 3.6.15 (b). Wir zeigen also (c) ⇒ (d). Ist $\ker(\Phi) = \{0\}$, so ist die Dimension dieses Raumes Null und wir erhalten mit der Dimensionsformel aus Korollar 3.6.18 $\mathrm{Rang}(\Phi) = \dim(V)$ und damit (d).
Zum Nachweis von (d) ⇒ (e) erinnern wir uns, dass $\mathrm{Rang}(\Phi) = \dim(\Phi(V))$ ist. Gilt also (d), so haben wir $\dim(\Phi(V)) = \dim(V)$. Dann ist $\Phi(V)$ ein Unterraum von V mit gleicher Dimension wie V, es muss also $\Phi(V) = V$ sein und Φ ist surjektiv.
Es bleibt noch (a) aus (e) zu folgern. Ist Φ surjektiv, so gilt $\Phi(V) = V$, also insbesondere

$$\mathrm{Rang}(\Phi) = \dim(\Phi(V)) = \dim(V).$$

Mit der Dimensionsformel sehen wir, dass dann $\dim(\ker(\Phi)) = 0$ sein muss, also ist $\ker(\Phi) = \{0\}$. Das liefert aber nach Satz 3.6.15 (b) die Injektivität von Φ und damit (a). □

Beispiel 3.6.20. Wir betrachten die lineare Abbildung $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ mit

$$\Phi(x) = \begin{pmatrix} x_2 + x_3 \\ -x_1 + 2x_2 + x_3 \\ x_1 - x_2 \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3.$$

und wollen ihren Kern und ihr Bild bestimmen. Zur Bestimmung des Kerns suchen wir alle $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$ mit $\Phi(x) = 0$, also haben wir das Gleichungssystem

$$\left\{ \begin{array}{rcrcrcl} & & x_2 & + & x_3 & = & 0 \\ -x_1 & + & 2x_2 & + & x_3 & = & 0 \\ x_1 & - & x_2 & & & = & 0 \end{array} \right.$$

zu lösen. Aus der letzten Gleichung bekommen wir $x_1 = x_2$ und aus der ersten $x_2 = -x_3$. Also ist auch $-x_1 + 2x_2 + x_3 = -x_2 + 2x_2 - x_2 = 0$ und die zweite Gleichung ist automatisch erfüllt. Die Lösungsmenge des obigen Gleichungssystems ist also

$$\ker(\Phi) = \left\{ \begin{pmatrix} \alpha \\ \alpha \\ -\alpha \end{pmatrix} \in \mathbb{R}^3 : \alpha \in \mathbb{R} \right\} = \left\langle \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\rangle .$$

Da damit $\dim(\ker(\Phi)) = 1$ ist, wissen wir nach der Dimensionsformel schon, dass $\dim(\Phi(V)) = \text{Rang}(\Phi) = 2$ sein muss. Damit reicht es zur Bestimmung des Bildes von Φ aus, wenn wir zwei linear unabhängige Vektoren in $\Phi(V)$ angeben können. Dazu betrachten wir aufs Geratewohl die Bilder

$$\Phi((1,0,0)^T) = (0,-1,1)^T \qquad \text{und} \qquad \Phi((0,0,1)^T) = (1,1,0)^T.$$

Da diese beiden linear unabhängig sind, gilt

$$\Phi(V) = \left\langle \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\rangle .$$

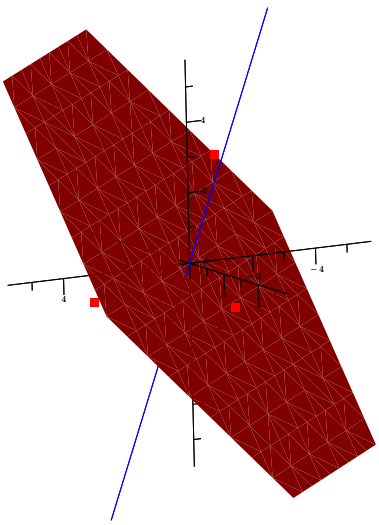

Abbildung 3.2.: Das Bild (Projektionsebene) und der Kern (Projektionsrichtung) der linearen Abbildung aus Beispiel 3.6.20

Bemerkung 3.6.21. Seien V und W endlichdimensionale K-Vektorräume und $\mathcal{B} := \{b_1, b_2, \ldots, b_n\}$ eine Basis von V, sowie $\mathcal{C} := \{c_1, c_2, \ldots, c_p\}$ eine Basis von W. Weiter sei $\Phi : V \to W$ eine lineare Abbildung.
Ein gegebenes $x \in V$ können wir nun bezüglich der Basis $\mathcal{B}$ darstellen und erhalten

$$x = \sum_{k=1}^{n} \xi_k b_k,$$

wobei $[\vec{x}]_{\mathcal{B}} = (\xi_1, \xi_2, \ldots, \xi_n)^T \in K^n$ der Koordinatenvektor von x bezüglich $\mathcal{B}$ ist. Damit gilt nun dank der Linearität von Φ

$$\Phi(x) = \Phi\Big(\sum_{k=1}^{n} \xi_k b_k\Big) = \sum_{k=1}^{n} \xi_k \Phi(b_k).$$

Stellen wir nun wiederum jedes $\Phi(b_k)$ durch seine Koordinaten bzgl. $\mathcal{C}$ dar, also bestimmen wir für jedes $k \in \{1, 2, \ldots, n\}$ Koeffizienten $\alpha_{1,k}, \alpha_{2,k}, \ldots, \alpha_{p,k} \in K$ mit

$$\Phi(b_k) = \sum_{j=1}^{p} \alpha_{j,k} c_j,$$

so erhalten wir zusammen

$$\Phi(x) = \sum_{k=1}^{n} \xi_k \Phi(b_k) = \sum_{k=1}^{n} \xi_k \sum_{j=1}^{p} \alpha_{j,k} c_j = \sum_{j=1}^{p} \Big(\sum_{k=1}^{n} \alpha_{j,k} \xi_k\Big) c_j.$$

Was sagt uns dieser Zeichenwust nun? Wir haben mit

$$[\overrightarrow{\Phi(x)}]_{\mathcal{C}} = \Big(\sum_{k=1}^{n} \alpha_{1,k}\xi_k, \sum_{k=1}^{n} \alpha_{2,k}\xi_k, \ldots, \sum_{k=1}^{n} \alpha_{n,k}\xi_k\Big)^T$$

den Koordinatenvektor von $\Phi(x)$ bezüglich der Basis $\mathcal{C}$ bestimmt, d.h. die Darstellung von $\Phi(x)$ in der Basis $\mathcal{C}$ in W angegeben. Dabei können wir die Koeffizienten $\sum_{k=1}^{n} \alpha_{j,k}\xi_k$ berechnen, wenn wir zum Einen die ξ_k, für jedes k kennen und zum anderen die Koeffizienten $\alpha_{j,k}$. Die ξ_k lassen sich direkt aus x bestimmen, sobald wir die Basis $\mathcal{B}$ haben, diese haben also nichts mit der speziellen Abbildung Φ zu tun. Umgekehrt bestimmen sich die $\alpha_{j,k}$ ausschließlich aus den Vektoren $\Phi(b_1), \Phi(b_2), \ldots, \Phi(b_n)$ und der Basis $\mathcal{C}$. Diese sind also für jedes $x \in V$ die selben. Diese Beobachtungen sind aus verschiedenen Gründen bemerkenswert:

1. Wir haben gesehen, dass es zur Berechnung von $\Phi(x)$ für jedes $x \in V$ ausreicht, wenn wir die n Vektoren $\Phi(b_1), \Phi(b_2), \ldots, \Phi(b_n)$ kennen.

2. Das bedeutet umgekehrt: Gibt uns jemand die gesamte Kollektion der Koeffizienten $\alpha_{j,k}$ für $j \in \{1, 2, \ldots, p\}$ und $k \in \{1, 2, \ldots, n\}$, so kennen wir die lineare Abbildung Φ, die dahinter steht komplett, denn wir können dann jedes $\Phi(x)$ ausrechnen.

3.7. Matrizen und lineare Abbildungen

3.7.1. Matrixrechnung

Wir erinnern noch mal an den Vektorraum der $p \times n$-Matrizen über einem Körper K, vgl. Beispiel 3.1.2 (b). Eine solche Matrix ist ein Schema von Elementen aus K mit p Zeilen und n Spalten:

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pn} \end{pmatrix} = (\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n},$$

wobei die Addition und Skalar-Multiplikation komponentenweise erklärt sind. Wichtig ist außerdem der Spezialfall $n = 1$. In diesem Fall hat die Matrix nur eine Spalte, es hat also ein $A \in K^{p\times 1}$ die Form

$$A = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{p1} \end{pmatrix}.$$

und wir haben $K^{p\times 1} \cong K^p$.
Wir wollen nun eine weitere Rechenoperation einführen, die Matrixmultiplikation.

Definition 3.7.1. *Es seien K ein Körper und $n, p, q \in \mathbb{N}^*$. Weiter seien zwei Matrizen $A = (\alpha_{j\ell})_{j=1,\dots,q,\ell=1,\dots,p} \in K^{q\times p}$, sowie $B = (\beta_{\ell k})_{\ell=1,\dots,p,k=1,\dots,n} \in K^{p\times n}$ gegeben. Dann definieren wir das* Matrixprodukt $A \cdot B = AB \in K^{q\times n}$ *als*

$$AB := \Big(\sum_{\ell=1}^{p} \alpha_{j\ell}\beta_{\ell k}\Big)_{j=1,\dots,q,k=1,\dots,n}.$$

Bemerkung 3.7.2. (a) Man beachte, dass das Produkt zweier Matrizen nur dann definiert ist, wenn die Anzahl der Spalten der ersten Matrix gleich der Anzahl der Zeilen der zweiten ist.

(b) Sieht man die Matrix A als Spaltenvektor ihrer Zeilen und B als Zeilenvektor der Spalten, d.h.

$$A = \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_q^T \end{pmatrix} \quad \text{und} \quad B = (b_1, b_2, \dots, b_n)$$

wobei $a_1, a_2, \ldots, a_q \in K^p$ die Zeilen von A und $b_1, b_2, \ldots, b_n \in K^p$ die Spalten von B sind, so bekommt man den Eintrag γ_{jk} von AB für $j \in \{1, 2, \ldots, q\}$ und $k \in \{1, 2, \ldots, n\}$, indem man das Skalarprodukt der j-ten Zeile von A mit der k-ten Spalte von B berechnet, also

$$\gamma_{jk} = (a_j | b_k) = \sum_{\ell=1}^{p} \alpha_{j\ell}\beta_{\ell k}.$$

Beispiel 3.7.3.

(a) Wir betrachten $A = \begin{pmatrix} 2 & -1 & 0 \\ 3 & 5 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 3}$ und $B = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ -1 & 5 \end{pmatrix} \in \mathbb{R}^{3\times 2}$.

Dann ist $A \cdot B \in \mathbb{R}^{2\times 2}$ mit

$$A \cdot B = \begin{pmatrix} 2 & -1 & 0 \\ 3 & 5 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ -1 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 7 & 11 \end{pmatrix}$$

und $B \cdot A \in \mathbb{R}^{3\times 3}$ mit

$$B \cdot A = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ -1 & 5 \end{pmatrix} \cdot \begin{pmatrix} 2 & -1 & 0 \\ 3 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 9 & 2 \\ 2 & -1 & 0 \\ 13 & 26 & 5 \end{pmatrix}.$$

Man beachte insbesondere, dass damit die Matrixmultiplikation nicht kommutativ ist.

(b) Sei nun $A = \begin{pmatrix} 3 & 1 & 0 \\ 2 & -1 & 1 \\ 0 & 2 & -2 \end{pmatrix} \in \mathbb{R}^{3\times 3}$ und $x = \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} \in \mathbb{R}^3 = \mathbb{R}^{3\times 1}$.

Dann ist auch $Ax \in \mathbb{R}^{3\times 1} = \mathbb{R}^3$ mit

$$Ax = \begin{pmatrix} 3 & 1 & 0 \\ 2 & -1 & 1 \\ 0 & 2 & -2 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 9 \\ 8 \\ -4 \end{pmatrix}.$$

Für die Matrixmultiplikation gelten die folgenden Rechenregeln:

Satz 3.7.4. *Seien $A, D \in K^{q\times p}$ und $B, C \in K^{p\times n}$, sowie $\lambda \in K$. Dann gilt*

(a) $A \cdot (\lambda B) = (\lambda A) \cdot B = \lambda(A \cdot B)$.

(b) $A \cdot (B + C) = A \cdot B + A \cdot C$ *und* $(A + D) \cdot B = A \cdot B + D \cdot B$.

Beweis. Wir beweisen die Aussage in (a), die zweite Aussage verbleibt als Übung. Es sei $A = (\alpha_{j\ell})_{j=1,\dots,q,\ell=1,\dots,p}$ und $B = (\beta_{\ell k})_{\ell=1,\dots,p,k=1,\dots,n}$. Dann gilt $\lambda B = (\lambda\beta_{\ell k})_{\ell=1,\dots,p,k=1,\dots,n}$ und nach der Definition der Matrixmultiplikation

$$A \cdot (\lambda B) = \Big(\sum_{\ell=1}^{p} \alpha_{j\ell}(\lambda\beta_{\ell k})\Big)_{j=1,\dots,q,k=1,\dots,n} = \Big(\sum_{\ell=1}^{p} (\lambda\alpha_{j\ell})\beta_{\ell k}\Big)_{j=1,\dots,q,k=1,\dots,n}.$$

Dieses ist nun zum Einen gleich der Matrix $(\lambda A)\cdot B$ und zum Anderen können wir nun das λ ganz aus der Summe und dann aus der Matrix ziehen und bekommen so

$$A \cdot (\lambda B) = \Big(\lambda \sum_{\ell=1}^{p} \alpha_{j\ell}\beta_{\ell k}\Big)_{j=1,\dots,q,k=1,\dots,n} = \lambda\Big(\sum_{\ell=1}^{p} \alpha_{j\ell}\beta_{\ell k}\Big)_{j=1,\dots,q,k=1,\dots,n} = \lambda(A \cdot B).$$

□

Bemerkung 3.7.5. Insbesondere gilt also für $A, B \in K^{p\times n}$ und $x, y \in K^n$ damit

$$A(x + y) = Ax + Ay \qquad \text{und} \qquad (A + B)x = Ax + Bx.$$

Besonders wichtig ist in der Matrixrechnung der Fall $p = n$. Man spricht dann von einer *quadratischen Matrix*. Hat man zwei gleich große quadratische Matrizen $A, B \in \mathbb{R}^{n\times n}$, so sind beide Produkte AB und BA definiert und wieder Elemente von $\mathbb{R}^{n\times n}$. Tatsächlich gilt sogar der folgende Satz.

Satz 3.7.6. *Sei $n \in \mathbb{N}^*$ und K ein Körper. Dann ist $(K^{n\times n}, +, \cdot)$ ein Ring mit Eins, der für $n \geq 2$ nicht kommutativ ist.*

Beweis. Nach Beispiel 3.1.2 (b) ist $K^{n\times n}$ ein K-Vektorraum, also ist $(K^{n\times n}, +)$ insbesondere eine abelsche Gruppe. Das Distributivgesetz ist genau die Aussage aus Satz 3.7.4 (b). Das Assoziativgesetz für die Multiplikation folgt aus der entsprechenden Eigenschaft von K, denn für drei Matrizen $A, B, C \in K^{n\times n}$ gilt

$$\begin{aligned}
A(BC) &= (\alpha_{j\ell})_{j,\ell=1,\dots,n} \cdot \big[(\beta_{\ell m})_{\ell,m=1,\dots,n} \cdot (\gamma_{mk})_{m,k=1,\dots,n}\big] \\
&= (\alpha_{j\ell})_{j,\ell=1,\dots,n} \cdot \Big(\sum_{m=1}^{n} \beta_{\ell m}\gamma_{mk})\Big)_{\ell,k=1,\dots,n} = \Big(\sum_{\ell=1}^{n} \alpha_{j\ell} \sum_{m=1}^{n} \beta_{\ell m}\gamma_{mk}\Big)_{j,k=1,\dots,n} \\
&= \Big(\sum_{\ell=1}^{n}\sum_{m=1}^{n} \alpha_{j\ell}(\beta_{\ell m}\gamma_{mk})\Big)_{j,k=1,\dots,n} = \Big(\sum_{m=1}^{n}\sum_{\ell=1}^{n} (\alpha_{j\ell}\beta_{\ell m})\gamma_{mk}\Big)_{j,k=1,\dots,n} \\
&= (AB)C,
\end{aligned}$$

wobei man für das letzte Gleichheitszeichen, die davor getätigte Rechnung wieder rückwärts machen muss.

Das Einselement ist die Matrix

$$I := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = (\delta_{jk})_{j,k=1,\dots,n},$$

denn für alle $A = (\alpha_{j\ell})_{j,\ell=1,\dots,n} \in K^{n\times n}$ gilt

$$AI = (\alpha_{j\ell})_{j,\ell=1,\dots,n} \cdot (\delta_{\ell k})_{\ell,k=1,\dots,n} = \Big(\sum_{\ell=1}^{n} \alpha_{j\ell}\delta_{\ell k}\Big)_{j,k=1,\dots,n} = (\alpha_{jk})_{j,k=1,\dots,n} = A$$

und umgekehrt genauso.
Die Nichtkommutativität sieht man schließlich für jedes $n \geq 2$ an

$$\begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

und

$$\begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \qquad \square$$

Definition 3.7.7. *Die Matrix $I = (\delta_{jk})_{j,k=1,\dots,n} \in K^{n\times n}$, d.h. das Einselement von $K^{n\times n}$, wird auch* Einheitsmatrix *genannt.*

Definition 3.7.8. *Sei $A = (\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n} \in K^{p\times n}$ eine Matrix. Dann heißt*

$$A^T := (\alpha_{kj})_{k=1,\dots,n,j=1,\dots,p} \in K^{n\times p}$$

die zu A transponierte *Matrix.*

Beispielsweise ist

$$\begin{pmatrix} 3 & 2 & 5 \\ 1 & 2 & 3 \end{pmatrix}^T = \begin{pmatrix} 3 & 1 \\ 2 & 2 \\ 5 & 3 \end{pmatrix} \quad \text{und} \quad (1\ 3\ 7)^T = \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix},$$

vgl. Beispiel 3.1.2 (a).
Für das Transponieren gelten die folgenden Rechenregeln:

Satz 3.7.9. *(a) Für alle $A, B \in K^{p\times n}$ und alle $\lambda \in K$ gilt*

- $(A+B)^T = A^T + B^T$,
- $(A^T)^T = A$,
- $(\lambda A)^T = \lambda A^T$.

(b) Für alle $A \in K^{q\times p}$ und $B \in K^{p\times n}$ gilt $(AB)^T = B^T A^T$.

Beweis. Wir beweisen nur beispielhaft den dritten Punkt von (a), der Rest verbleibt als Übung.
Sei also $A = (\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n} \in K^{p\times n}$ und $\lambda \in K$. Dann gilt nach der Definition der Skalar-Multiplikation in $K^{p\times n}$

$$(\lambda A)^T = \begin{pmatrix} \lambda\alpha_{11} & \lambda\alpha_{12} & \dots & \lambda\alpha_{1n} \\ \lambda\alpha_{21} & \lambda\alpha_{22} & \dots & \lambda\alpha_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda\alpha_{p1} & \lambda\alpha_{p2} & \dots & \lambda\alpha_{pn} \end{pmatrix}^T = \begin{pmatrix} \lambda\alpha_{11} & \lambda\alpha_{21} & \dots & \lambda\alpha_{p1} \\ \lambda\alpha_{12} & \lambda\alpha_{22} & \dots & \lambda\alpha_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda\alpha_{1n} & \lambda\alpha_{2n} & \dots & \lambda\alpha_{pn} \end{pmatrix} = \lambda A^T.$$

□

Übungsaufgabe 3.7.10. Es sei $A \in \mathbb{R}^{n\times n}$ eine Matrix und $(\cdot|\cdot)$ das Standardskalarprodukt auf $\mathbb{R}^n$. Zeigen Sie, dass dann für alle $x, y \in \mathbb{R}^n$ gilt

$$(Ax|y) = (x|A^T y).$$

Im Lichte dieser Übungsaufgabe ist auch das Vertauschen der Reihenfolge in Satz 3.7.9 (b) zu sehen:

$$\big((AB)^T x|y\big) = (x|ABy) = (A^T x|By) = (B^T A^T x|y).$$

3.7.2. Die Abbildungsmatrix einer linearen Abbildung

Was haben nun Matrizen mit linearen Abbildungen zu tun? Dieser Frage wollen wir jetzt nachgehen. Dazu sei $\Phi : V \to W$ eine lineare Abbildung zwischen zwei endlichdimensionalen K-Vektorräumen, sowie $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$ eine Basis von V und $\mathcal{C} = \{c_1, c_2, \dots, c_p\}$ eine Basis von W.
In dieser Situation haben wir in Bemerkung 3.6.21 folgendes gesehen: Bestimmt man α_{jk} für $j = 1, \dots, p$ und $k = 1, \dots, n$ so, dass $\Phi(b_k) = \sum_{j=1}^{p} \alpha_{jk} c_j$ gilt und stellt man $x \in V$ durch seinen Koordinatenvektor $\vec{x} = [\vec{x}]_{\mathcal{B}} = (\xi_1, \xi_2, \dots, \xi_n)^T \in K^n$ bezüglich $\mathcal{B}$ dar, so gilt

$$\Phi(x) = \sum_{j=1}^{p} \Big(\sum_{k=1}^{n} \alpha_{jk}\xi_k \Big) c_j = \sum_{j=1}^{p} (A\vec{x})_j c_j,$$

mit $A = (\alpha_{j,k})_{j=1,\dots,p,k=1,\dots,n}$, d.h. der Vektor $A\vec{x}$ enthält die Koordinaten des Vektors $\Phi(x)$ bezüglich der Basis $\mathcal{C}$, oder in Formeln $[\overrightarrow{\Phi(x)}]_{\mathcal{C}} = A[\vec{x}]_{\mathcal{B}}$.

Definition 3.7.11. *Es seien V, W, $\mathcal{B}$, $\mathcal{C}$, n, p und $\Phi : V \to W$ wie oben. Die Matrix $A = (\alpha_{jk})_{j=1,\dots,p,k=1,\dots,n} \in K^{p\times n}$ heißt dann* Darstellungsmatrix *oder* Abbildungsmatrix von Φ bezüglich $\mathcal{B}$ und $\mathcal{C}$. *Wir bezeichnen diese mit $A =: M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$.*

Im Kopf haben sollten Sie die folgende

Merkregel: In den Spalten der Abbildungsmatrix stehen die Koordinaten der Bilder der Basisvektoren.

Beispiel 3.7.12. Es sei $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ die Spiegelung an der x_2-Achse. Das ist nach Beispiel 3.6.5 (b) eine lineare Abbildung. Statten wir $\mathbb{R}^2$ mit der Standardbasis $\mathcal{B} = \{b_1, b_2\} = \{\binom{1}{0}, \binom{0}{1}\}$ aus, so können wir die Abbildungsmatrix von Φ bezüglich $\mathcal{B}$ und $\mathcal{B}$ angeben, indem wir die Bilder der Basisvektoren bestimmen:

$$\Phi(b_1) = \Phi\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}-1\\0\end{pmatrix} = -1 \cdot b_1 + 0 \cdot b_2$$

und

$$\Phi(b_2) = \Phi\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix} = 0 \cdot b_1 + 1 \cdot b_2.$$

Also ist $M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}$.
Tatsächlich ist zum Beispiel für $x = \binom{2}{1}$ der Koordinatenvektor $\vec{x}$ bezüglich $\mathcal{B}$ gleich dem Vektor x und wir bekommen

$$M_{\mathcal{B}}^{\mathcal{B}}(\Phi)\vec{x} = \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}-2\\1\end{pmatrix} = \overrightarrow{\Phi(x)}.$$

Beispiel 3.7.13. Wir betrachten die lineare Abbildung $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ mit

$$\Phi\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) = \begin{pmatrix}3x_1 - 2x_2\\ x_1 + x_2\end{pmatrix}.$$

(a) Es sei zunächst $\mathcal{B} = \mathcal{C} = \{\binom{1}{0}, \binom{0}{1}\}$. Dann ist

$$\Phi(b_1) = \Phi\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}3\\1\end{pmatrix} = 3 \cdot c_1 + 1 \cdot c_2$$

und

$$\Phi(b_2) = \Phi\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}-2\\1\end{pmatrix} = -2 \cdot c_1 + 1 \cdot c_2.$$

Also ist

$$M_{\mathcal{C}}^{\mathcal{B}}(\Phi) = M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = \begin{pmatrix}3 & -2\\ 1 & 1\end{pmatrix}.$$

(b) Ändert man die Basen, bekommt man auch eine andere Abbildungsmatrix für die selbe Abbildung. Wir behalten $\mathcal{B} = \{\binom{1}{0}, \binom{0}{1}\}$ wie oben, aber ersetzen $\mathcal{C}$ durch $\mathcal{D} = \{d_1, d_2\} = \{\binom{3}{1}, \binom{-2}{1}\}$. Dann ist

$$\Phi(b_1) = \begin{pmatrix} 3 \\ 1 \end{pmatrix} = d_1 \quad \text{und} \quad \Phi(b_2) = \begin{pmatrix} -2 \\ 1 \end{pmatrix} = d_2.$$

Also ist damit

$$M_{\mathcal{D}}^{\mathcal{B}}(\Phi) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Damit ist $M_{\mathcal{D}}^{\mathcal{B}}(\Phi)\binom{1}{1} = \binom{1}{1}$, aber es ist doch $\Phi((1,1)^T) = (1,2)^T$? Was ist hier schief gegangen?

Nichts! Wir müssen nur richtig interpretieren. Nach unserer Definition der Abbildungsmatrix ist das Ergebnis von $M_{\mathcal{D}}^{\mathcal{B}}(\Phi)\binom{1}{1}$ der Koordinatenvektor $[\overrightarrow{\Phi(x)}]_{\mathcal{D}}$ von $\Phi(x)$ bezüglich $\mathcal{D}$ und nicht $\Phi(x)$ selbst! Da nun $\mathcal{D}$ nicht die Standardbasis ist, sind Vektor und Koordinatenvektor nicht mehr identisch. Aber tatsächlich gilt

$$1 \cdot d_1 + 1 \cdot d_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \Phi\big((1,1)^T\big).$$

Beispiel 3.7.14. Wir wählen $V = W$ und betrachten die Identität $\mathrm{id} : V \to V$ auf V. Diese ist linear, also hat sie zu einer gegebenen Basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ von V eine Abbildungsmatrix bezüglich $\mathcal{B}$ und $\mathcal{B}$, die wir nun bestimmen wollen. Es ist $\mathrm{id}(b_j) = b_j$, also ist der Koordinatenvektor von $\mathrm{id}(b_j)$ der j-te Einheitsvektor. In der Abbildungsmatrix $M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id})$ der Identität steht also in der j-ten Spalte der j-te Einheitsvektor, also

$$M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}) = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix} = I,$$

d.h. die Abbildungsmatrix ist die Einheitsmatrix.

Bemerkung 3.7.15. Es seien V und W endlichdimensionale K-Vektorräume und $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ eine Basis von V und $\mathcal{C} = \{c_1, c_2, \ldots, c_p\}$ eine Basis von W. Dann haben wir nun gesehen, dass es zu jeder linearen Abbildung $\Phi : V \to W$ eine Abbildungsmatrix $M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$ gibt. Aber auch umgekehrt wird ein Schuh daraus: Geben wir eine beliebige Matrix $A = (\alpha_{jk})_{j=1,\ldots,p,k=1,\ldots,n} \in K^{p\times n}$ vor, so gibt es nach Satz 3.6.13 genau eine lineare Abbildung $\Phi_A : V \to W$ mit

$$\Phi_A(b_k) = \sum_{j=1}^{p} \alpha_{jk} c_j.$$

Für diese Abbildung gilt dann $M_{\mathcal{C}}^{\mathcal{B}}(\Phi_A) = A$ (warum?).
Damit haben wir zusammengefasst folgendes gesehen: Wählt man die Basen $\mathcal{B}$ und $\mathcal{C}$ fest, so gibt es eine eins-zu-eins-Beziehung zwischen den linearen Abbildungen von V nach W und den Matrizen aus $K^{p\times n}$. Das bedeutet, dass wir jede lineare Abbildung zwischen endlichdimensionalen K-Vektorräumen, also ein u.U. durchaus unübersichtliches Objekt, durch eine solche Matrix, also etwas recht überschaubares, beschreiben können und dass uns jede Erkenntnis über Matrizen eine Erkenntnis über lineare Abbildungen beschert und umgekehrt.

Definition 3.7.16. *Es sei $A \in K^{p\times n}$ eine Matrix und $a_1, a_2, \ldots, a_n \in K^p$ seien die Spalten von A. Dann heißt*

(a) $\mathrm{Rang}(A) := \dim(\langle a_1, a_2, \ldots, a_n\rangle)$ (Spalten-)rang *von A.*

(b) der Untervektorraum $\ker(\mathrm{A}) := \{x \in K^n : Ax = 0\}$ *von K^n* Kern *von A.*

Der folgende Satz zeigt ein paar Beziehungen zwischen einer linearen Abbildung und ihrer Abbildungsmatrix auf. Er bleibt hier ohne Beweis stehen.

Satz 3.7.17. *Es seien V und W endlichdimensionale K-Vektorräume mit Basen $\mathcal{B}$, bzw. $\mathcal{C}$, $\Phi : V \to W$ eine lineare Abbildung und $A = M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$. Dann gilt*

(a) $\mathrm{Rang}(\Phi) = \mathrm{Rang}(A)$.

(b) $\mathrm{Rang}(A)$ *ist die Maximalanzahl linear unabhängiger Spalten von A.*

(c) $\dim(\ker(\Phi)) = \dim(\ker(A))$.

(d) $\dim(V) = \mathrm{Rang}(A) + \dim(\ker(A))$.

Im Abschnitt über lineare Abbildungen haben wir in Übungsaufgabe 3.6.3 gesehen, dass die Verkettung von linearen Abbildungen und die Umkehrfunktion von Isomorphismen wieder lineare Abbildungen sind. Wie bestimmt man nun deren Abbildungsmatrizen, d.h. wie bekommt man $M_{\mathcal{D}}^{\mathcal{B}}(\Psi\circ\Phi)$ und $M_{\mathcal{B}}^{\mathcal{C}}(\Phi^{-1})$ aus $M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$ und $M_{\mathcal{D}}^{\mathcal{C}}(\Psi)$?
Zuerst zeigen wir, dass die Abbildungsmatrix der Verkettung genau das Matrixprodukt der beiden Abbildungsmatrizen von Ψ und Φ ist.

Satz 3.7.18. *Es seien V, W und X endlichdimensionale K-Vektorräume mit Basen $\mathcal{B}$, $\mathcal{C}$, bzw. $\mathcal{D}$. Weiter seien $\Phi \in \mathcal{L}(V, W)$ und $\Psi \in \mathcal{L}(W, X)$. Dann gilt*

$$M_{\mathcal{D}}^{\mathcal{B}}(\Psi \circ \Phi) = M_{\mathcal{D}}^{\mathcal{C}}(\Psi) \cdot M_{\mathcal{C}}^{\mathcal{B}}(\Phi).$$

Beweis. Es sei $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$, $\mathcal{C} = \{c_1, c_2, \ldots, c_p\}$ und $\mathcal{D} = \{d_1, d_2, \ldots, d_q\}$. Weiterhin seien $M_{\mathcal{D}}^{\mathcal{C}}(\Psi) = (\alpha_{j\ell})_{j=1,\ldots,q,\ell=1,\ldots,p}$ und $M_{\mathcal{C}}^{\mathcal{B}}(\Phi) = (\beta_{\ell k})_{\ell=1,\ldots,p,k=1,\ldots,n}$, sowie $M_{\mathcal{D}}^{\mathcal{B}}(\Psi \circ \Phi) = (\gamma_{jk})_{j=1,\ldots,q,k=1,\ldots,n}$. Um die Abbildungsmatrix von $\Psi \circ \Phi$ zu

bestimmen, müssen wir die Koordinaten von $(\Psi\circ\Phi)(b_k)$ bezüglich der Basis $\mathcal{D}$ für jedes $k = 1, \ldots, n$ bestimmen. Es gilt nach der Definition der Abbildungsmatrix

$$(\Psi \circ \Phi)(b_k) = \Psi(\Phi(b_k)) = \Psi\Big(\sum_{\ell=1}^{p} \beta_{\ell k} c_\ell\Big).$$

Verwenden wir nun die Linearität von Ψ und dann die Abbildungsmatrix von Ψ, so erhalten wir

$$(\Psi \circ \Phi)(b_k) = \sum_{\ell=1}^{p} \beta_{\ell k}\Psi(c_\ell) = \sum_{\ell=1}^{p} \beta_{\ell k} \sum_{j=1}^{q} \alpha_{j\ell} d_j = \sum_{j=1}^{q}\Big(\sum_{\ell=1}^{p} \alpha_{j\ell}\beta_{\ell k}\Big) d_j.$$

Also ist die Abbildungsmatrix von $\Psi \circ \Phi$ gegeben durch

$$\gamma_{jk} = \sum_{\ell=1}^{p} \alpha_{j\ell}\beta_{\ell k}$$

und das ist genau die Definition des Matrixprodukts. □

Entsprechend kann man auch die folgenden Rechenregeln zeigen:

Übungsaufgabe 3.7.19. Es seien V und W endlichdimensionale K-Vektorräume mit Basen $\mathcal{B}$, bzw. $\mathcal{C}$. Sind $\Phi, \Psi \in \mathcal{L}(V, W)$ und $\lambda \in K$, so gilt

(a) $M_{\mathcal{C}}^{\mathcal{B}}(\Phi + \Psi) = M_{\mathcal{C}}^{\mathcal{B}}(\Phi) + M_{\mathcal{C}}^{\mathcal{B}}(\Psi)$ und

(b) $M_{\mathcal{C}}^{\mathcal{B}}(\lambda\Phi) = \lambda M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$.

Beispiel 3.7.20. Die lineare Abbildung in $\mathbb{R}^2$, die man bekommt, wenn man zunächst die Abbildung Φ in Beispiel 3.7.13 ausführt und dann die Spiegelung an der x_2-Achse, vgl. Beispiel 3.7.12, hat also bezüglich der Standardbasis in $\mathbb{R}^2$ die Abbildungsmatrix

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 3 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 2 \\ 1 & 1 \end{pmatrix}.$$

Wir wenden uns der Abbildungsmatrix der Umkehrfunktion zu. Es seien also V und W endlichdimensionale K-Vektorräume mit Basen $\mathcal{B}$, bzw. $\mathcal{C}$ und $\Phi : V \to W$ ein Isomorphismus.
Wir bemerken zunächst, dass Φ nur bijektiv sein kann, wenn die Dimensionen von V und W übereinstimmen, denn nach der Dimensionsformel gilt

$$\dim(W) \overset{\Phi \text{ surj.}}{=} \operatorname{Rang}(\Phi) = \dim(V) - \dim(\ker(\Phi)) \overset{\Phi \text{ inj.}}{=} \dim(V).$$

Also muss die Abbildungsmatrix einer solchen bijektiven Abbildung quadratisch sein.

Es gilt nun $\Phi \circ \Phi^{-1} = \mathrm{id}_W$ und $\Phi^{-1} \circ \Phi = \mathrm{id}_V$, also ist nach Beispiel 3.7.14

$$M_{\mathcal{C}}^{\mathcal{B}}(\Phi) \cdot M_{\mathcal{B}}^{\mathcal{C}}(\Phi^{-1}) = M_{\mathcal{C}}^{\mathcal{C}}(\mathrm{id}_W) = I \quad \text{und} \quad M_{\mathcal{B}}^{\mathcal{C}}(\Phi^{-1}) \cdot M_{\mathcal{C}}^{\mathcal{B}}(\Phi) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = I.$$

Nach Satz 3.7.6 ist I das neutrale Element der Multiplikation im Ring $K^{n\times n}$. Wir haben also gerade gezeigt, dass man einen Vektorraum-Isomorphismus daran erkennt, dass seine Abbildungsmatrix in diesem Ring ein multiplikatives Inverses besitzt. Und dieses multiplikative Inverse ist dann die Abbildungsmatrix der Umkehrfunktion. Wir gießen das in eine Definition.

Definition 3.7.21. *Es sei $n \in \mathbb{N}^*$ und K ein Körper. Eine Matrix $A \in K^{n\times n}$ heißt* invertierbar *oder* regulär*, falls ein $A^{-1} \in K^{n\times n}$ existiert mit $A \cdot A^{-1} = I$ und $A^{-1} \cdot A = I$. Die Matrix A^{-1} heißt dann die* Inverse *von A.*
Ist A nicht regulär, so nennt man A singulär*.*

Bemerkung 3.7.22. (a) Nicht jede von der Nullmatrix verschiedene quadratische Matrix ist invertierbar, d.h. $K^{n\times n}$ ist kein Körper. Z.B. ist

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

d.h. $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ ist ein linker Nullteiler.

(b) Die Inverse ist eindeutig, denn sind A' und A'' zwei Inverse von A, so ist mit dem schon mehrfach bemühten Argument $A' = A'(AA'') = (A'A)A'' = A''$.

Satz 3.7.23. *Es sei $n \in \mathbb{N}^*$, K ein Körper und $A \in K^{n\times n}$. Dann sind die folgenden Aussagen äquivalent:*

(a) A ist invertierbar.

(b) $\mathrm{Rang}(A) = n$.

(c) $\ker(A) = \{0\}$.

Beweis. Nach unseren obigen Überlegungen ist A genau dann invertierbar, wenn die Abbildung Φ_A, vgl. Bemerkung 3.7.15, bijektiv ist. Dies wiederum ist wegen Satz 3.6.19 und Satz 3.7.17 (a) äquivalent dazu, dass $\mathrm{Rang}(A) = \mathrm{Rang}(\Phi_A) = n$ ist. Damit haben wir (a) $\iff$ (b) gezeigt.
Mit Hilfe der Dimensionsformel $n = \mathrm{Rang}(A) + \dim(\ker(A))$ aus Satz 3.7.17 (d) sieht man, dass (b) genau dann gilt, wenn $\dim(\ker(A)) = 0$ ist, was wiederum zu (c) äquivalent ist. □

Beispiel 3.7.24. Die Matrix

$$A = \begin{pmatrix} 2 & 3 \\ -1 & 2 \end{pmatrix} \in \mathbb{R}^{2\times 2} \quad \text{ist invertierbar mit} \quad A^{-1} = \frac{1}{7}\begin{pmatrix} 2 & -3 \\ 1 & 2 \end{pmatrix},$$

denn

$$AA^{-1} = \begin{pmatrix} 2 & 3 \\ -1 & 2 \end{pmatrix} \cdot \frac{1}{7} \begin{pmatrix} 2 & -3 \\ 1 & 2 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix} = I$$

und

$$A^{-1}A = \frac{1}{7} \begin{pmatrix} 2 & -3 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 2 & 3 \\ -1 & 2 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix} = I.$$

Im Kontrast dazu ist

$$B := \begin{pmatrix} \tilde{2} & \tilde{3} \\ \widetilde{-1} & \tilde{2} \end{pmatrix} \in \mathbb{Z}_7^{2\times 2} \quad \text{nicht invertierbar,}$$

denn es ist

$$\tilde{5} \begin{pmatrix} \tilde{2} \\ \widetilde{-1} \end{pmatrix} = \begin{pmatrix} \widetilde{10} \\ \widetilde{-5} \end{pmatrix} = \begin{pmatrix} \tilde{3} \\ \tilde{2} \end{pmatrix}.$$

Das bedeutet, dass die zweite Spalte von B ein Vielfaches der ersten Spalte ist, die beiden sind also linear abhängig. Die maximale Anzahl linear unabhängiger Spaltenvektoren von B ist folglich 1 und Satz 3.7.17(b) liefert, dass $\text{Rang}(B) = 1$ ist. Das bedeutet wiederum nach Satz 3.7.23 das Aus für die Invertierbarkeit.

Bemerkung 3.7.25. (a) Sind $A, B \in K^{n\times n}$ invertierbare Matrizen, so ist auch ihr Produkt AB invertierbar mit $(AB)^{-1} = B^{-1}A^{-1}$, denn

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

und genauso gilt umgekehrt $B^{-1}A^{-1}AB = I$.

Man beachte, dass sich die Reihenfolge der beiden Matrizen umkehrt!

(b) Außerdem gilt für jedes $\lambda \in K \setminus \{0\}$ und jedes invertierbare $A \in K^{n\times n}$

$$(\lambda A)^{-1} = \lambda^{-1}A^{-1}.$$

(c) Schließlich ist mit $A \in K^{n\times n}$ immer auch A^T invertierbar und es gilt

$$(A^{-1})^T = (A^T)^{-1},$$

weshalb man manchmal auch die etwas verwegene Schreibweise A^{-T} sieht.

Übungsaufgabe 3.7.26. Die Menge

$$GL(n, K) := \{A \in K^{n\times n} : A \text{ invertierbar}\}$$

ist mit der Matrixmultiplikation als Verknüpfung eine Gruppe. Diese heißt *allgemeine lineare Gruppe*. Die Abkürzung kommt von der englischen Bezeichnung "general linear group".

3.8. Lineare Gleichungssysteme

In vielen Zusammenhängen stößt man in der Linearen Algebra (und auch anderswo) auf die Problematik lineare Gleichungssysteme lösen zu müssen.
Die Problemstellung ist die folgende: Gegeben Koeffizienten α_{jk}, $j \in \{1, 2, \ldots, p\}$, $k \in \{1, 2, \ldots, n\}$, aus einem Körper K und weitere Elemente $b_1, b_2, \ldots, b_p \in K$, bestimme $x_1, x_2, \ldots, x_n \in K$ mit

$$\begin{aligned}
\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= b_1 \\
\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n &= b_2 \\
\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots & \\
\alpha_{p1}x_1 + \alpha_{p2}x_2 + \cdots + \alpha_{pn}x_n &= b_p.
\end{aligned}$$

Setzt man $b := (b_1, b_2, \ldots, b_p)^T \in K^p$ und $A := (\alpha_{jk})_{j=1,\ldots,p,k=1,\ldots,n} \in K^{p\times n}$, so kann man jedes lineare Gleichungssystem in der *Matrixform*

$$Ax = b$$

schreiben.

Definition 3.8.1. *(a) Ein* lineares Gleichungssystem (LGS) *ist die Gleichung $Ax = b$, wobei $A \in K^{p\times n}$ und $b \in K^p$ gegeben und der Vektor $x \in K^n$ gesucht ist.*

(b) Ein LGS heißt homogen*, falls $b = 0$ ist. Sonst heißt es* inhomogen*.*

(c) Das LGS $Ax = b$ heißt lösbar*, falls ein $x \in K^n$ existiert mit $Ax = b$ und es heißt* eindeutig lösbar*, wenn genau ein solches $x \in K^n$ existiert. Gibt es kein $x \in K^n$ mit dieser Eigenschaft, so nennt man das LGS* unlösbar*.*

3.8.1. Lösbarkeitstheorie

Bemerkung 3.8.2. In folgenden Spezialfällen ist es leicht, das Lösungsverhalten von linearen Gleichungssystemen zu bestimmen:

(a) Ist $p = n$ und $A \in K^{n\times n}$ invertierbar, so ist $Ax = b$ für jedes $b \in K^n$ eindeutig lösbar durch $x = A^{-1}b$, denn mit dieser Wahl gilt

$$Ax = A(A^{-1}b) = (AA^{-1})b = Ib = b$$

und ist $y \in K^n$ eine weitere Lösung, gilt also $Ay = b$, so ist $x = A^{-1}b = A^{-1}Ay = y$.

(b) Ist $A \in K^{p\times n}$ die Nullmatrix, so ist $Ax = b$ nur für $b = 0$ lösbar, dann aber ist jedes $x \in K^n$ eine Lösung.

(c) Das homogene System $Ax = 0$ ist für jede Matrix A lösbar, da der Nullvektor eine Lösung ist. Außerdem bilden alle Lösungen den Untervektorraum $\ker(A)$, vgl. Definition 3.7.16.

Kennt man den Kern von A, so ist zur Lösung des Systems $Ax = b$ schon ein Großteil der Arbeit getan, denn es gilt der folgende Satz über die Struktur der Lösungsmenge eines linearen Gleichungssystems.

Satz 3.8.3. *Es seien $A \in K^{p\times n}$ und $b \in K^p$. Hat das LGS $Ax = b$ eine Lösung $x_s \in K^n$, so sind alle Lösungen des LGS gegeben durch*

$$\{x \in K^n : Ax = b\} = \{x_s + y : y \in \ker(A)\}.$$

Kennt man also eine Lösung, so erhält man alle Lösungen als die Elemente der Äquivalenzklasse dieser einen Lösung in $V/\ker(A)$.

Beweis. Wir beweisen zunächst „$\subseteq$“. Sei dazu $x \in K^n$ eine Lösung von $Ax = b$. Dann gilt

$$A(x - x_s) = Ax - Ax_s = b - b = 0.$$

Also ist $x - x_s \in \ker(A)$ und es gibt ein $y \in \ker(A)$ mit $x = x_s + y$.
Für die umgekehrte Inklusion sei $y \in \ker(A)$. Dann gilt

$$A(x_s + y) = Ax_s + Ay = b + 0 = b.$$

Somit ist $x_s + y$ eine Lösung von $Ax = b$ und wir sind fertig. □

Man nennt die Lösung x_s, deren Existenz man irgendwoher bekommen muss, um obigen Satz anwenden zu können, eine *spezielle Lösung* oder auch *Partikulärlösung* des LGS.
Der folgende Satz bietet ein Kriterium für die grundsätzliche Lösbarkeit eines LGS.

Satz 3.8.4. *Es seien $A \in K^{p\times n}$ und $b \in K^p$. Bezeichnen wir mit $A|b$ die Matrix in $K^{p\times(n+1)}$, die durch anfügen von b als $(n+1)$-te Spalte an A entsteht (man nennt diese auch* erweiterte Koeffizientenmatrix*), so gilt*

(a) Das LGS $Ax = b$ ist genau dann lösbar, wenn $\mathrm{Rang}(A) = \mathrm{Rang}(A|b)$ gilt.

(b) Die folgenden Aussagen sind äquivalent:

i) Das LGS $Ax = b$ ist eindeutig lösbar.

ii) $Ax = b$ ist lösbar und $\ker(A) = \{0\}$.

iii) $\mathrm{Rang}(A) = \mathrm{Rang}(A|b) = n$.

Beweis. (a) Wir bezeichnen mit $a_1, a_2, \ldots, a_n \in K^p$ die Spalten von A. Zum Nachweis von „⇒“ bemerken wir zunächst, dass auf jeden Fall $\text{Rang}(A) \leq \text{Rang}(A|b)$ gilt. Sei nun $x \in K^n$ eine Lösung von $Ax = b$, d.h. es gilt $(a_1\ a_2\ \ldots\ a_n)\cdot x = b$. Dann ist nach der Definition der Matrixmultiplikation

$$\sum_{j=1}^{n} x_j a_j = b,$$

was uns sagt, dass b eine Linearkombination der Vektoren $a_1, a_2, \ldots, a_n$, also der Spalten von A, ist. Damit ist der Rang von $A|b$ sicher nicht größer als der von A und wir haben die behauptete Gleichheit gezeigt.

Wir beweisen „⇐“. $\text{Rang}(A|b) = \text{Rang}(A)$ bedeutet, dass die beiden Untervektorräume $\langle a_1, a_2, \ldots, a_n, b\rangle$ und $\langle a_1, a_2, \ldots, a_n\rangle$ von K^p die selbe Dimension haben. Es muss also schon $b \in \langle a_1, a_2, \ldots, a_n\rangle$ gelten. Das bedeutet, dass es $x_1, x_2, \ldots, x_n \in K$ gibt mit $\sum_{j=1}^{n} x_j a_j = b$. Damit ist $x = (x_1, x_2, \ldots, x_n)^T \in K^n$ eine Lösung des LGS $Ax = b$ und wir sind fertig.

(b) Wir zeigen (b)i)⇒ (b)ii)⇒(b)iii) ⇒(b)i).

Ist $Ax = b$ eindeutig lösbar, so ist das LGS natürlich lösbar und enthielte $\ker(A)$ mehr als die Null, so gäbe es nach Satz 3.8.3 mehr als eine Lösung, somit haben wir (b)i)⇒(b)ii).

Gilt (b)ii), so haben wir mit Teil (a) sofort $\text{Rang}(A|b) = \text{Rang}(A)$. Weiter folgt dank der Dimensionsformel für Matrizen, vgl. Satz 3.7.17 (d), aus $\dim(\ker(A)) = 0$ auch $\text{Rang}(A) = n$.

Zum Nachweis von (b)iii)⇒(b)i) sehen wir zunächst, dass die Lösbarkeit des LGS aus (a) folgt. Es bleibt die Eindeutigkeit der Lösung zu zeigen. Wegen $\text{Rang}(A) = n$ und wiederum der Dimensionsformel bekommen wir $\ker(A) = \{0\}$ und damit liefert Satz 3.8.3 die gewünschte Eindeutigkeit. □

3.8.2. Der Gauß-Algorithmus

Nachdem wir nun im letzten Abschnitt einiges Wissen über die Struktur der Lösungsmenge von linearen Gleichungssystemen gesammelt haben, wollen wir uns nun der Frage zuwenden, wie man konkrete Systeme algorithmisch lösen kann. Das meist verwendete Verfahren heißt *Gauß-Algorithmus*. Dieser soll hier intuitiv anhand von Beispielen behandelt werden, ohne exakte mathematische Begründung.
Ist $A = (\alpha_{jk})_{j=1,\ldots,p,k=1,\ldots,n} \in K^{p\times n}$ und $b = (b_1, \ldots, b_p)^T \in K^p$ so ändern folgende

Umformungen die Menge der Lösungen des linearen Gleichungssystems

$$\begin{aligned}
\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= b_1 \\
\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n &= b_2 \\
\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots & \\
\alpha_{p1}x_1 + \alpha_{p2}x_2 + \cdots + \alpha_{pn}x_n &= b_p.
\end{aligned}$$

nicht:

1. Vertauschen zweier Zeilen (Gleichungen),

2. Multiplizieren einer Zeile (Gleichung) mit einem $\lambda \neq 0$ aus K,

3. Addition des Vielfachen einer Zeile (Gleichung) zu einer anderen Zeile (Gleichung).

Man nennt diese Umformungen *Elementarumformungen*. Der Gauß'sche Algorithmus zur Lösung von linearen Gleichungssystemen beruht nun auf der Anwendung dieser drei Umformungen.

Beispiel 3.8.5. Wir betrachten das LGS:

$$\begin{array}{rrrrl}
x_1 & +x_2 & +x_3 & +x_4 & =5 \\
2x_1 & +x_2 & -2x_3 & -3x_4 & =4 \\
-x_1 & -x_2 & +2x_3 & -x_4 & =1 \\
2x_1 & +x_2 & -x_3 & +2x_4 & =1,
\end{array}$$

bzw. in Matrixform

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -2 & -3 \\ -1 & -1 & 2 & -1 \\ 2 & 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 1 \\ 1 \end{pmatrix}.$$

Wir specken die Notation noch weiter ab und schreiben das LGS als

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 5 \\ 2 & 1 & -2 & -3 & 4 \\ -1 & -1 & 2 & -1 & 1 \\ 2 & 1 & -1 & 2 & 1 \end{array}\right).$$

Als erstes muss man nun durch das Vertauschen von Zeilen dafür sorgen, dass links oben in der Ecke ein Eintrag steht, der nicht Null ist. Das ist hier schon der Fall, so dass wir gleich damit beginnen können die erste Spalte aufzuräumen.

Wir addieren nach Elementarumformung 3. die erste Zeile zur dritten dazu und addieren ihr (-2)-faches zur 2. und 4. Zeile. Das notiert man folgendermaßen:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 5 \\ 2 & 1 & -2 & -3 & 4 \\ -1 & -1 & 2 & -1 & 1 \\ 2 & 1 & -1 & 2 & 1 \end{array}\right) \begin{array}{l} \cdot(-2) \\ \leftarrow \\ \mid \\ \leftarrow \end{array} \begin{array}{c} \cdot 1 \\ \mid \\ \leftarrow \\ \end{array} \rightsquigarrow \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 5 \\ 0 & -1 & -4 & -5 & -6 \\ 0 & 0 & 3 & 0 & 6 \\ 0 & -1 & -3 & 0 & -9 \end{array}\right).$$

Nun arbeiten wir mit der zweiten Zeile. Diese multiplizieren wir zunächst mit (-1), so dass wir wieder eine führende 1 haben. Im Falle, dass dort Null steht, so tauscht man sich wieder eine passende Zeile dorthin. Ist die zweite Spalte sogar in allen Zeilen nach der ersten Null, so geht man gleich zur dritten Spalte über. Also

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 5 \\ 0 & -1 & -4 & -5 & -6 \\ 0 & 0 & 3 & 0 & 6 \\ 0 & -1 & -3 & 0 & -9 \end{array}\right) \cdot(-1) \rightsquigarrow \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 5 \\ 0 & 1 & 4 & 5 & 6 \\ 0 & 0 & 3 & 0 & 6 \\ 0 & -1 & -3 & 0 & -9 \end{array}\right) \begin{array}{l} \leftarrow \\ \cdot(-1) \\ \\ \end{array} \begin{array}{c} \\ \cdot 1 \\ \mid \\ \leftarrow \end{array}$$

$$\rightsquigarrow \left(\begin{array}{cccc|c} 1 & 0 & -3 & -4 & -1 \\ 0 & 1 & 4 & 5 & 6 \\ 0 & 0 & 3 & 0 & 6 \\ 0 & 0 & 1 & 5 & -3 \end{array}\right)$$

Auf diese Weise haben wir dafür gesorgt, dass auch in der zweiten Spalte nur noch eine Eins und sonst nur Nullen stehen.

Nun ist die dritte Spalte dran, wir multiplizieren zunächst die dritte Zeile mit $1/3$ und räumen dann wieder auf:

$$\left(\begin{array}{cccc|c} 1 & 0 & -3 & -4 & -1 \\ 0 & 1 & 4 & 5 & 6 \\ 0 & 0 & 3 & 0 & 6 \\ 0 & 0 & 1 & 5 & -3 \end{array}\right) : 3 \rightsquigarrow \left(\begin{array}{cccc|c} 1 & 0 & -3 & -4 & -1 \\ 0 & 1 & 4 & 5 & 6 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 5 & -3 \end{array}\right) \begin{array}{l} \leftarrow \\ \mid \\ \cdot 3 \\ \\ \end{array} \begin{array}{c} \\ \leftarrow \\ \cdot(-4) \\ \\ \end{array} \begin{array}{c} \\ \\ \cdot(-1) \\ \leftarrow \end{array}$$

$$\rightsquigarrow \left(\begin{array}{cccc|c} 1 & 0 & 0 & -4 & 5 \\ 0 & 1 & 0 & 5 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 5 & -5 \end{array}\right).$$

Nun noch mal das selbe Spielchen in der vierten Spalte mit der 5 in der vierten

Zeile:

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -4 & 5 \\ 0 & 1 & 0 & 5 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 5 & -5 \end{array}\right) : 5 \quad \rightsquigarrow \quad \left(\begin{array}{cccc|c} 1 & 0 & 0 & -4 & 5 \\ 0 & 1 & 0 & 5 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right) \begin{array}{c} \\ \leftarrow \\ | \\ \cdot(-5) \end{array} \begin{array}{c} \leftarrow \\ | \\ | \\ \cdot 4 \end{array}$$

$$\rightsquigarrow \quad \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right).$$

Abschließend haben wir damit das ursprüngliche LGS durch Elemtarumformungen auf die Form

$$\begin{array}{rcl} x_1 & & = 1 \\ x_2 & & = 3 \\ x_3 & & = 2 \\ & x_4 & = -1 \end{array}$$

gebracht und haben damit die Lösung dastehen. Das LGS ist eindeutig lösbar mit $x = (1\ 3\ 2\ -1)^T$.

Bemerkung 3.8.6. Formt man ein LGS $Ax = b$ mit Hilfe von Elementarumformungen um, so ändert sich nichts an der Lösungsmenge und insbesondere auch nichts an allen Eigenschaften der Koeffizientenmatrix und der erweiterten Koeffizientenmatrix, die mit dem Lösungsverhalten des LGS zu tun haben. So bleiben z.B. $\mathrm{Rang}(A)$ und $\mathrm{Rang}(A|b)$ erhalten.

Beispiel 3.8.7. Der Gauß-Algorithmus funktioniert auch für nicht-quadratische lineare Gleichungssysteme. Für $a \in \mathbb{R}$ betrachten wir z.B.

$$\begin{array}{rrrrl} & x_2 & + 3x_3 & + 5x_4 & = a \\ -2x_1 & - x_2 & + x_3 & + x_4 & = 0 \\ x_1 & + x_2 & + x_3 & + 2x_4 & = 1. \end{array}$$

Also

$$\left(\begin{array}{cccc|c} 0 & 1 & 3 & 5 & a \\ -2 & -1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 2 & 1 \end{array}\right) \begin{array}{c} \leftarrow \\ | \\ \leftarrow \end{array} \quad \rightsquigarrow \quad \left(\begin{array}{cccc|c} 1 & 1 & 1 & 2 & 1 \\ -2 & -1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 5 & a \end{array}\right) \begin{array}{c} \cdot 2 \\ \leftarrow \\ \\ \end{array}$$

$$\rightsquigarrow \quad \left(\begin{array}{cccc|c} 1 & 1 & 1 & 2 & 1 \\ 0 & 1 & 3 & 5 & 2 \\ 0 & 1 & 3 & 5 & a \end{array}\right) \begin{array}{c} \leftarrow \\ \cdot(-1) \\ \leftarrow \end{array} \quad \rightsquigarrow \quad \left(\begin{array}{cccc|c} 1 & 0 & -2 & -3 & -1 \\ 0 & 1 & 3 & 5 & 2 \\ 0 & 0 & 0 & 0 & a-2 \end{array}\right).$$

Nun müssen wir zwei Fälle unterscheiden. Ist $a \neq 2$, so gilt $\mathrm{Rang}(A) = 2 \neq 3 = \mathrm{Rang}(A|b)$, also ist in diesem Fall das LGS unlösbar.

Im Fall $a = 2$ ist die letzte Zeile eine komplette Nullzeile und damit sind der Rang der Matrix und der erweiterten Koeffizientenmatrix beide 2. Das LGS ist also nach Satz 3.8.4 (a) lösbar. Nach Satz 3.8.3 ist weiterhin die Lösungsmenge von der Form $x_s + \ker(A)$, wobei x_s eine spezielle Lösung und A die Koeffizientenmatrix des LGS ist.
Da der Rang von A nach obiger Rechnung 2 ist, vgl. Bemerkung 3.8.6, bekommen wir mit dem (b)-Teil von Satz 3.8.4, dass $\dim(\ker(A)) = 2$ ist. Die Lösung des verbleibenden LGS

$$\left(\begin{array}{cccc|c} 1 & 0 & -2 & -3 & -1 \\ 0 & 1 & 3 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

erhalten wir nun so: wir setzen die noch „unbearbeiteten“ Variablen $x_3 = \lambda$ und $x_4 = \mu$. Dann liefern uns die beiden verbliebenen Gleichungen $x_1 = -1 + 2\lambda + 3\mu$ und $x_2 = 2 - 3\lambda - 5\mu$. Also sind alle $x \in \mathbb{R}^4$ mit

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 + 2\lambda + 3\mu \\ 2 - 3\lambda - 5\mu \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ -3 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 3 \\ -5 \\ 0 \\ 1 \end{pmatrix}$$

die Lösungen des LGS. Wir erhalten also als Lösungsmenge

$$\mathbb{L} = \left\{ \begin{pmatrix} -1 \\ 2 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ -3 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 3 \\ -5 \\ 0 \\ 1 \end{pmatrix} : \lambda, \mu \in \mathbb{R} \right\}.$$

Beispiel 3.8.8. Ein weiteres Beispiel über dem Körper $\mathbb{Z}_5$:

$$\begin{pmatrix} \tilde{3} & \tilde{2} & \tilde{1} & \tilde{0} \\ \tilde{1} & \tilde{1} & \tilde{4} & \tilde{1} \\ \tilde{1} & \tilde{3} & \tilde{1} & \tilde{2} \end{pmatrix} \cdot\tilde{2} \quad \rightsquigarrow \quad \begin{pmatrix} \tilde{1} & \tilde{4} & \tilde{2} & \tilde{0} \\ \tilde{1} & \tilde{1} & \tilde{4} & \tilde{1} \\ \tilde{1} & \tilde{3} & \tilde{1} & \tilde{2} \end{pmatrix} \begin{matrix} \cdot\widetilde{(-1)} \\ \leftarrow \\ \leftarrow \end{matrix} \quad \rightsquigarrow \quad \left(\begin{array}{ccc|c} \tilde{1} & \tilde{4} & \tilde{2} & \tilde{0} \\ \tilde{0} & \widetilde{-3} & \tilde{2} & \tilde{1} \\ \tilde{0} & \widetilde{-1} & \widetilde{-1} & \tilde{2} \end{array}\right)$$

$$\rightsquigarrow \quad \begin{pmatrix} \tilde{1} & \tilde{4} & \tilde{2} & \tilde{0} \\ \tilde{0} & \tilde{2} & \tilde{2} & \tilde{1} \\ \tilde{0} & \tilde{4} & \tilde{4} & \tilde{2} \end{pmatrix} \cdot\tilde{3} \quad \rightsquigarrow \quad \begin{pmatrix} \tilde{1} & \tilde{4} & \tilde{2} & \tilde{0} \\ \tilde{0} & \tilde{1} & \tilde{1} & \tilde{3} \\ \tilde{0} & \tilde{4} & \tilde{4} & \tilde{2} \end{pmatrix} \begin{matrix} \leftarrow \\ \cdot\widetilde{(-4)} \\ \leftarrow \end{matrix} \quad \rightsquigarrow \quad \left(\begin{array}{ccc|c} \tilde{1} & \tilde{0} & \widetilde{-2} & \widetilde{-12} \\ \tilde{0} & \tilde{1} & \tilde{1} & \tilde{3} \\ \tilde{0} & \tilde{0} & \tilde{0} & \widetilde{-10} \end{array}\right)$$

$$\rightsquigarrow \quad \left(\begin{array}{ccc|c} \tilde{1} & \tilde{0} & \tilde{3} & \tilde{3} \\ \tilde{0} & \tilde{1} & \tilde{1} & \tilde{3} \\ \tilde{0} & \tilde{0} & \tilde{0} & \tilde{0} \end{array}\right).$$

Der Lösungsraum ist damit eindimensional. Wir setzen $x_3 = \lambda$ und bekommen aus den beiden ersten verbliebenen Gleichungen $x_1 = \tilde{3} - \tilde{3}\lambda$ und $x_2 = \tilde{3} - \lambda$. Also ist

$$\mathbb{L} = \left\{ \begin{pmatrix} \tilde{3} - \tilde{3}\lambda \\ \tilde{3} - \lambda \\ \lambda \end{pmatrix} : \lambda \in \mathbb{Z}_5 \right\} = \left\{ \begin{pmatrix} \tilde{3} \\ \tilde{3} \\ \tilde{0} \end{pmatrix} + \lambda \begin{pmatrix} \tilde{2} \\ \tilde{4} \\ \tilde{1} \end{pmatrix} : \lambda \in \mathbb{Z}_5 \right\}.$$

Bemerkung 3.8.9. Der Gauß-Algorithmus ist auch ein Mittel zur Berechnung von Inversen invertierbarer Matrizen, denn dieses lässt sich folgendermaßen als das Lösen mehrerer linearer Gleichungssysteme auffassen. Ist $A \in K^{n\times n}$ eine invertierbare Matrix und ist $x_j \in K^n$ für jedes $j \in \{1, 2, \ldots, n\}$ die j-te Spalte von A^{-1}, so muss $AA^{-1} = I$ gelten, d.h. es ist Ax_j die j-te Spalte der Einheitsmatrix I, was gerade der Vektor $e_j := (\delta_{jk})_{k=1}^n$ ist.
Die Inversion der Matrix A entspricht also dem Lösen der n linearen Gleichungssysteme $Ax_j = e_j$ für $j = 1, 2, \ldots, n$. Diese kann man mit Hilfe des Gauß'schen Algorithmus simultan lösen.

Beispiel 3.8.10. Es sei $A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \\ 2 & -1 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 3}$.
Wir lösen die drei LGSe simultan:

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 2 & -1 & 3 & 0 & 0 & 1 \end{array}\right) \begin{matrix} \cdot(-2) \\ | \\ \leftarrow \end{matrix} \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 3 & -2 & 0 & 1 \end{array}\right) \begin{matrix} \leftarrow \\ \cdot 1 \\ \end{matrix} \begin{matrix} \\ \cdot(-1) \\ \leftarrow \end{matrix}$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & -1 & 1 \end{array}\right) \begin{matrix} \leftarrow \\ \leftarrow \\ \cdot(-2) \end{matrix} \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 5 & 3 & -2 \\ 0 & 1 & 0 & 4 & 3 & -2 \\ 0 & 0 & 1 & -2 & -1 & 1 \end{array}\right).$$

Also ist $A^{-1} = \begin{pmatrix} 5 & 3 & -2 \\ 4 & 3 & -2 \\ -2 & -1 & 1 \end{pmatrix}$.

3.9. Basiswechsel

Es seien V und W zwei endlichdimensionale K-Vektorräume, $\mathcal{B}$ eine Basis von V und $\mathcal{C}$ eine Basis von W. Ist $\Phi : V \to W$ eine lineare Abbildung, so haben wir zu dieser in Abschnitt 3.7.2 die Abbildungsmatrix $M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$ gefunden. Nun seien $\mathcal{B}'$ eine weitere Basis von V und $\mathcal{C}'$ eine weitere Basis von W und wir wollen uns der Frage zuwenden, wie man die Abbildungsmatrix $M_{\mathcal{C}'}^{\mathcal{B}'}(\Phi)$ aus $M_{\mathcal{C}}^{\mathcal{B}}(\Phi)$ bestimmen kann.
Die Idee dazu ist die Abbildung Φ kompliziert als $\mathrm{id}_W \circ \Phi \circ \mathrm{id}_V$ zu schreiben und dann nach Satz 3.7.18

$$M_{\mathcal{C}'}^{\mathcal{B}'}(\Phi) = M_{\mathcal{C}'}^{\mathcal{B}'}(\mathrm{id}_W \circ \Phi \circ \mathrm{id}_V) = M_{\mathcal{C}'}^{\mathcal{C}}(\mathrm{id}_W) \cdot M_{\mathcal{C}}^{\mathcal{B}}(\Phi) \cdot M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_V)$$

zu berechnen.

Satz 3.9.1. *Es seien V und W zwei endlichdimensionale K-Vektorräume mit Basen $\mathcal{B}$ und $\mathcal{B}'$, bzw. $\mathcal{C}$ und $\mathcal{C}'$ wie oben. Ist dann $\Phi : V \to W$ linear, so existieren invertierbare Matrizen S und T mit*

$$M_{\mathcal{C}'}^{\mathcal{B}'}(\Phi) = T M_{\mathcal{C}}^{\mathcal{B}}(\Phi) S.$$

Beweis. Nach obigen Vorüberlegungen ist nur noch zu zeigen, dass die Matrizen

$$T := M_{\mathcal{C}'}^{\mathcal{C}}(\mathrm{id}_W) \quad \text{und} \quad S := M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_V)$$

invertierbar sind. Dazu überlegen wir uns mit Hilfe von Satz 3.7.18 und Beispiel 3.7.14

$$S \cdot M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}_V) = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_V) \cdot M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}_V) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V \circ \mathrm{id}_V) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = I$$

und genauso

$$M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}_V) \cdot S = M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}_V) \cdot M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_V) = M_{\mathcal{B}'}^{\mathcal{B}'}(\mathrm{id}_V) = I.$$

Die Argumentation für T verläuft analog. □

Besonders wichtig ist der Spezialfall $V = W$, also für lineare Abbildungen $\Phi : V \to V$.

Satz 3.9.2. *Es seien V ein endlichdimensionaler K-Vektorraum, $\mathcal{B}$ und $\mathcal{B}'$ Basen von V und $\Phi : V \to V$ linear. Sind $A := M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ und $A' := M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi)$ die Abbildungsmatrizen von Φ bezüglich $\mathcal{B}$, bzw. $\mathcal{B}'$, so existiert eine invertierbare Matrix S mit*

$$A' = S^{-1}AS.$$

Die Matrix S in obigem Satz, die die Abbildungsmatrizen bezüglich der verschiedenen Basen ineinander übersetzt, heißt *Basiswechselmatrix*.

Beweis. Nach Satz 3.9.1 gilt für $S := M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id})$ und $T := M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id})$ die Beziehung $A' = TAS$. Da außerdem

$$\begin{aligned} TS &= M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}) M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}) = M_{\mathcal{B}'}^{\mathcal{B}'}(\mathrm{id}) = I \quad \text{und} \\ ST &= M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}) M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}) = I \end{aligned}$$

gilt, ist $T = S^{-1}$ und wir sind fertig. □

Bemerkung 3.9.3. Es bleibt natürlich die Frage, wie man denn die Basiswechselmatrix S im konkreten Fall berechnet. Dazu bezeichnen wir $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ und $\mathcal{B}' = \{b_1', b_2', \ldots, b_n'\}$ und erinnern uns, dass $S = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id})$ gilt.
In den Spalten von S stehen also die Koordinaten von $\mathrm{id}(b_1'), \mathrm{id}(b_2'), \ldots, \mathrm{id}(b_n')$, d.h. von $b_1', b_2', \ldots, b_n'$, bezüglich der Basis $\mathcal{B}$. Um die Matrix S zu erhalten, muss man also die Basisvektoren $b_1', b_2', \ldots, b_n'$ in der Basis $\mathcal{B}$ ausdrücken, das bedeutet man hat im schlimmsten Fall n lineare Gleichungssysteme zu lösen.
Ein wichtiger und sehr einfacher Spezialfall ist der, wenn $V = K^n$ und $\mathcal{B}$ die Standardbasis ist, denn dann sind die Koordinaten von $b_1', b_2', \ldots, b_n'$ bezüglich $\mathcal{B}$ einfach die Vektoren $b_1', b_2', \ldots, b_n'$ selbst, d.h. diese bilden dann die Spalten der Basiswechselmatrix S.

Beispiel 3.9.4. Wir betrachten die lineare Abbildung $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ mit

$$\Phi(x) = \begin{pmatrix} x_1 - 4x_2 - 4x_3 \\ 3x_2 + 2x_3 \\ -2x_1 - 7x_2 - 4x_3 \end{pmatrix} = \begin{pmatrix} 1 & -4 & -4 \\ 0 & 3 & 2 \\ -2 & -7 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Ist $\mathcal{B}$ die Standardbasis von $\mathbb{R}^3$, so ist also

$$A = M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = \begin{pmatrix} 1 & -4 & -4 \\ 0 & 3 & 2 \\ -2 & -7 & -4 \end{pmatrix}.$$

Wir wollen nun die Abbildungsmatrix von Φ bezüglich der Basis

$$\mathcal{B}' := \left\{ \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} \right\}$$

berechnen. Da $\mathcal{B}$ die Standardbasis ist, gilt für die Basiswechselmatrix

$$S = \begin{pmatrix} -1 & 4 & 2 \\ 1 & -2 & -1 \\ -1 & 1 & 3 \end{pmatrix}.$$

Deren Inverse bestimmt man mit dem Gauß-Verfahren, vgl. Beispiel 3.8.10, zu

$$S^{-1} = \frac{1}{5} \begin{pmatrix} 5 & 10 & 0 \\ 2 & 1 & -1 \\ 1 & 3 & 2 \end{pmatrix}.$$

Damit haben wir nach Satz 3.9.2

$$\begin{aligned} M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi) = S^{-1}AS &= \frac{1}{5} \begin{pmatrix} 5 & 10 & 0 \\ 2 & 1 & -1 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & -4 & -4 \\ 0 & 3 & 2 \\ -2 & -7 & -4 \end{pmatrix} \begin{pmatrix} -1 & 4 & 2 \\ 1 & -2 & -1 \\ -1 & 1 & 3 \end{pmatrix} \\ &= \frac{1}{5} \begin{pmatrix} 5 & 10 & 0 \\ 2 & 1 & -1 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} -1 & 8 & -6 \\ 1 & -4 & 3 \\ -1 & 2 & -9 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & -15 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix}. \end{aligned}$$

Dieses Beispiel zeigt auch, dass man durch eine angepasste Wahl der Basis die Abbildungsmatrix sehr stark vereinfachen kann. Mit der Frage, wie man eine solche Basis findet, werden wir uns im Abschnitt 3.11 beschäftigen.

Beispiel 3.9.5. Die Technik des Basiswechsels kann auch dazu genutzt werden, die Abbildungsmatrix einer durch geometrische Angaben definierten linearen Abbildung zu ermitteln. Beispielhaft betrachten wir als lineare Abbildung $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ die Projektion auf den Unterraum $\langle \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \rangle$ in Richtung $\langle \begin{pmatrix} -4 \\ -1 \\ 2 \end{pmatrix} \rangle$. Gesucht ist die Abbildungsmatrix dieser Abbildung bezüglich der Standardbasis $\mathcal{B}$. Da es sehr schwierig ist, die Abbildung direkt in dieser Basis zu beschreiben, betrachten wir zunächst die dem Problem angepasste Basis

$$\mathcal{B}' = \left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -4 \\ -1 \\ 2 \end{pmatrix} \right\} = \{b'_1, b'_2, b'_3\}.$$

Nach der Definition von Φ gilt dann $\Phi(b'_1) = b'_1$, $\Phi(b'_2) = b'_2$ und $\Phi(b'_3) = 0$.
Also ist

$$M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Mit der Basiswechselmatrix

$$S = \begin{pmatrix} 0 & 1 & -4 \\ 1 & 2 & -1 \\ 1 & 1 & 2 \end{pmatrix},$$

die den Basiswechsel von $\mathcal{B}$ nach $\mathcal{B}'$ vermittelt, gilt nun $M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi) = S^{-1} M_{\mathcal{B}}^{\mathcal{B}}(\Phi) S$.
Also ist $S M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi) = M_{\mathcal{B}}^{\mathcal{B}}(\Phi) S$ und schließlich $M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = S M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi) S^{-1}$.
Mit dem Gauß-Verfahren kann man wieder

$$S^{-1} = \begin{pmatrix} 5 & -6 & 7 \\ -3 & 4 & -4 \\ -1 & 1 & -1 \end{pmatrix}$$

bestimmen. Also ist die gesuchte Abbildungsmatrix

$$\begin{aligned} M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = S M_{\mathcal{B}'}^{\mathcal{B}'}(\Phi) S^{-1} &= \begin{pmatrix} 0 & 1 & -4 \\ 1 & 2 & -1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 5 & -6 & 7 \\ -3 & 4 & -4 \\ -1 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & -4 \\ 1 & 2 & -1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 5 & -6 & 7 \\ -3 & 4 & -4 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -3 & 4 & -4 \\ -1 & 2 & -1 \\ 2 & -2 & 3 \end{pmatrix}. \end{aligned}$$

Definition 3.9.6. *Zwei Matrizen $A, B \in K^{n \times n}$ heißen* ähnlich, *wenn es eine invertierbare Matrix $S \in K^{n \times n}$ gibt mit $B = S^{-1} A S$.*

Bemerkung 3.9.7. (a) Darstellungsmatrizen einer linearen Abbildung bezüglich verschiedener Basen sind immer zueinander ähnlich.

(b) Die Ähnlichkeit von Matrizen ist eine Äquivalenzrelation. Machen Sie sich das als Übung klar!

(c) Sind alle Matrizen zueinander ähnlich? Nein!

Ist $A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{pmatrix} \in K^{n\times n}$, so heißt $\mathrm{Spur}(A) := \sum_{j=1}^{n} \alpha_{jj}$ die *Spur* von A.

Man kann zeigen: Sind A und B ähnliche Matrizen, dann gilt $\mathrm{Spur}(A) = \mathrm{Spur}(B)$.

Eine weitere charakteristische Größe einer Matrix, die sich beim Basiswechsel nicht ändert, werden wir im Abschnitt 3.10 kennenlernen.

Zum Abschluss dieses Abschnitts betrachten wir noch den Spezialfall, dass die beiden Basen $\mathcal{B}$ und $\mathcal{B}'$, zwischen denen gewechselt wird, jeweils Orthonormalbasen sind.

Lemma 3.9.8. *Sind $\mathcal{B}$ und $\mathcal{B}'$ Orthonormalbasen eines n-dimensionalen $\mathbb{R}$-Vektorraums V mit Skalarprodukt $(\cdot|\cdot)_V$, so gilt für die Basiswechselmatrix $S = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}) \in \mathbb{R}^{n\times n}$, dass ihre Spalten eine Orthonormalbasis des $\mathbb{R}^n$ bezüglich des Standardskalarproduktes $(\cdot|\cdot)_{\mathbb{R}^n}$ bilden.*

Beweis. Es sei $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$ und $\mathcal{B}' = \{b'_1, b'_2, \dots, b'_n\}$.
Die j-te Spalte s_j von S ist nach Bemerkung 3.9.3 für jedes $j = 1, 2, \dots, n$ gegeben durch den Koordinatenvektor $[\vec{b}'_j]_{\mathcal{B}}$ von b'_j bezüglich $\mathcal{B}$. Wie wir in Bemerkung 3.4.15 gesehen haben, ist dieser Koordinatenvektor gegeben durch

$$s_j = \begin{pmatrix} (b'_j|b_1)_V \\ (b'_j|b_2)_V \\ \vdots \\ (b'_j|b_n)_V \end{pmatrix}.$$

Also ist für alle $j, k \in \{1, 2, \dots, n\}$

$$(s_k|s_j)_{\mathbb{R}^n} = \sum_{\ell=1}^{n} (b'_k|b_\ell)_V \cdot (b'_j|b_\ell)_V = \Big(b'_j \Big| \sum_{\ell=1}^{n} (b'_k|b_\ell)_V b_\ell\Big)_V.$$

Die Summe im zweiten Argument ist nun nach Bemerkung 3.4.15 gerade der Vektor b'_k. Also haben wir

$$(s_k|s_j)_{\mathbb{R}^n} = (b'_j|b'_k)_V = \delta_{jk}$$

und sind fertig. □

Definition 3.9.9. *Eine Matrix $A \in \mathbb{R}^{n\times n}$ heißt* orthogonal*, falls die Spalten von A eine Orthonormalbasis bezüglich des Standardskalarproduktes bilden.*

Man beachte, dass eine orthogonale Matrix immer invertierbar ist, da ihre Spalten eine Basis bilden und der Rang somit gleich der Spaltenanzahl ist.

Übungsaufgabe 3.9.10. Es sei $A \in \mathbb{R}^{n\times n}$. Beweisen Sie, dass die folgenden Aussagen äquivalent sind:

(a) A ist orthogonal.

(b) A ist invertierbar und es gilt $A^{-1} = A^T$.

(c) Die Zeilen von A bilden eine Orthonormalbasis bezüglich des Standardskalarproduktes.

(d) A ist invertierbar und A^{-1} ist orthogonal.

(e) A^T ist orthogonal.

Bemerkung 3.9.11. Beim Basiswechsel zwischen Orthonormalbasen ist vor allem die Beziehung $A^{-1} = A^T$ nützlich, denn das Transponieren einer Matrix ist vom Rechenaufwand her viel einfacher als das Invertieren.

Übungsaufgabe 3.9.12. Die Menge

$$O(n,\mathbb{R}) := \{A \in \mathbb{R}^{n\times n} : A \text{ orthogonal}\}$$

ist eine Untergruppe von $\mathrm{GL}(n,\mathbb{R})$, genannt *orthogonale Gruppe*.

3.10. Determinanten

Wie sieht man einer Matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in K^{2\times 2}$ an, ob sie invertierbar ist? Berechnet man allgemein die Inverse, so findet man

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \qquad \text{falls } ad-bc \neq 0.$$

Anscheinend ist also der Wert $ad-bc$ von besonderer Bedeutung. Man nennt ihn die Determinante. Allgemein ist diese folgendermaßen definiert.

Definition 3.10.1. *(a) Es seien $A \in K^{n\times n}$ und $j,k \in \{1,2,\ldots,n\}$. Dann bezeichne $A_{jk} \in K^{(n-1)\times(n-1)}$ die Matrix, die aus A durch Streichen der j-ten Zeile und der k-ten Spalte entsteht.*

(b) Für $A = (\alpha) \in K^{1\times 1}$ definieren wir die Determinante *durch* $\det(A) := \alpha$.

(c) *Für ein* $A = (\alpha_{jk})_{j,k=1}^n \in K^{n\times n}$ *mit* $n > 1$ *erklären wir die* Determinante *als*

$$\det(A) = \sum_{k=1}^{n} (-1)^{1+k} \alpha_{1k} \det(A_{1k}). \quad \text{(Entwicklung nach der ersten Zeile)}$$

(d) *Für die Determinante einer Matrix* $A = (\alpha_{jk})_{j,k=1}^n \in K^{n\times n}$ *schreibt man auch*

$$\det(A) = \begin{vmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{vmatrix}.$$

Beispiel 3.10.2. (a) Im Fall $n = 2$ gilt nach obiger Definition tatsächlich

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = (-1)^2 a \det((d)) + (-1)^3 b \det((c)) = ad - bc.$$

(b) Schon für $n = 3$ wird die Berechnung allerdings ein bisschen mühsamer:

$$\begin{aligned} \begin{vmatrix} 2 & 1 & 3 \\ 4 & 0 & 5 \\ 7 & 6 & 8 \end{vmatrix} &= (-1)^2 \cdot 2 \cdot \begin{vmatrix} 0 & 5 \\ 6 & 8 \end{vmatrix} + (-1)^3 \cdot 1 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} + (-1)^4 \cdot 3 \cdot \begin{vmatrix} 4 & 0 \\ 7 & 6 \end{vmatrix} \\ &= 2 \cdot (0 \cdot 8 - 5 \cdot 6) - 1 \cdot (4 \cdot 8 - 5 \cdot 7) + 3 \cdot (4 \cdot 6 - 0 \cdot 7) \\ &= -60 - 32 + 35 + 72 = 15. \end{aligned}$$

Von großer praktischer Bedeutung ist die folgende Beobachtung.

Beispiel 3.10.3. Es sei $A \in K^{n\times n}$ eine sogenannte *untere Dreiecksmatrix*, d.h.

$$A = \begin{pmatrix} \alpha_{11} & 0 & \dots & 0 \\ * & \alpha_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \dots & * & \alpha_{nn} \end{pmatrix},$$

wobei anstelle der Sterne „$*$“ irgendwelche Elemente aus K stehen. Dann gilt nach der Definition der Determinante

$$\begin{aligned} \det(A) &= \alpha_{11} \cdot \begin{vmatrix} \alpha_{22} & 0 & \dots & 0 \\ * & \alpha_{33} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \dots & * & \alpha_{nn} \end{vmatrix} = \alpha_{11}\alpha_{22} \cdot \begin{vmatrix} \alpha_{33} & 0 & \dots & 0 \\ * & \alpha_{44} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \dots & * & \alpha_{nn} \end{vmatrix} \\ &= \dots = \alpha_{11}\alpha_{22} \cdot \ldots \cdot \alpha_{nn}. \end{aligned}$$

Insbesondere gilt damit immer

$$\det(I) = 1.$$

Im folgenden Satz fassen wir einige grundlegende Rechenregeln für die Determinante ohne Beweis zusammen.

Satz 3.10.4. *Es sei*

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in K^{n\times n},$$

wobei $a_1^T, a_2^T, \ldots, a_n^T \in K^n$ *die Zeilen von* A *seien. Dann gilt*

(a) Vertauscht man zwei beliebige Zeilen der Matrix, so ändert sich das Vorzeichen der Determinante, z.B. ist

$$\begin{vmatrix} a_2 \\ a_1 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{vmatrix} = -\det(A).$$

(b) Die Determinante $\det(A)$ *ist linear in jeder Zeile, d.h. für jeden Vektor* $b \in K^n$*, für alle* $\lambda, \mu \in K$*, sowie jedes* $j \in \{1, 2, \ldots, n\}$ *gilt*

$$\begin{vmatrix} a_1 \\ \vdots \\ a_{j-1} \\ \lambda a_j + \mu b \\ a_{j+1} \\ \vdots \\ a_n \end{vmatrix} = \lambda \det(A) + \mu \begin{vmatrix} a_1 \\ \vdots \\ a_{j-1} \\ b \\ a_{j+1} \\ \vdots \\ a_n \end{vmatrix}.$$

(c) Sei $\lambda \in K$*. Addiert man zu einer Zeile von* A *das* λ*-fache einer anderen Zeile von* A *hinzu, so ändert sich die Determinante nicht.*

(d) Man kann statt nach der ersten Zeile zu entwickeln, vgl. die Definition der Determinante, auch nach der j*-ten Zeile für jedes* $j \in \{1, 2, \ldots, n\}$ *entwickeln. Genauer gesagt gilt für jedes solche* j

$$\det(A) = \sum_{k=1}^{n} (-1)^{j+k} \alpha_{jk} \det(A_{jk}).$$

(e) Es ist $\det(A) = \det(A^T)$.

(f) Ist $B \in K^{n\times n}$ eine weitere Matrix, so ist $\det(AB) = \det(A)\det(B)$.

Korollar 3.10.5. *(a) Die Aussagen in Satz 3.10.4 (a)-(d) gelten auch für Spalten statt Zeilen. Als Formel für das Entwickeln nach der k-ten Spalte bekommt man*

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+k} \alpha_{jk} \det(A_{jk}).$$

(b) Ist $\lambda \in K$ und $A \in K^{n\times n}$, so gilt $\det(\lambda A) = \lambda^n \det(A)$.

Beweis. (a) Das folgt aus Satz 3.10.4 (e).

(b) Mit $A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ wie in Satz 3.10.4 gilt $\lambda A = \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{pmatrix}$, also ist durch n-malige Anwendung von (b) aus Satz 3.10.4

$$\det(\lambda A) = \begin{vmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{vmatrix} = \lambda \begin{vmatrix} a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{vmatrix} = \cdots = \lambda^n \begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{vmatrix} = \lambda^n \det(A). \qquad \square$$

Die praktische Relevanz obiger Rechenregeln liegt unter Anderem darin, dass (a) bis (c) aus Satz 3.10.4 uns sagen, was mit der Determinante unter den Elementarumformungen aus dem Gauß-Verfahren passiert und dass wir dieses Werkzeug damit zur Berechnung von Determinanten nutzen können. Dank Korollar 3.10.5 (a) können wir es sogar sowohl auf die Zeilen als auch auf die Spalten der Matrix anwenden. Wir betrachten die Berechnung von Determinanten anhand zweier Beispiele.

Beispiel 3.10.6. (a) Wir berechnen

$$D := \begin{vmatrix} 1 & -2 & 3 & 5 & 8 \\ 0 & -1 & -1 & 2 & 3 \\ 2 & 4 & -1 & 3 & 1 \\ 0 & 0 & 5 & 0 & 0 \\ 1 & 3 & 0 & 4 & -1 \end{vmatrix}.$$

Zuerst ist es sinnvoll nach der vierten Zeile zu entwickeln, denn dabei sind vier der fünf Summanden Null:

$$D = (-1) \cdot 5 \cdot \begin{vmatrix} 1 & -2 & 5 & 8 \\ 0 & -1 & 2 & 3 \\ 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & -1 \end{vmatrix}.$$

Nun können wir los„gaußen“. Wir machen zunächst ein paar der schon gewohnten Zeilenumformungen. Dabei bleibt die Determinante unverändert.

$$D = -5 \cdot \begin{vmatrix} 1 & -2 & 5 & 8 \\ 0 & -1 & 2 & 3 \\ 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & -1 \end{vmatrix} \begin{matrix} \\ \\ \leftarrow \\ \cdot(-2) \end{matrix} \begin{matrix} \leftarrow \\ | \\ | \\ \cdot(-1) \end{matrix} = -5 \cdot \begin{vmatrix} 0 & -5 & 1 & 9 \\ 0 & -1 & 2 & 3 \\ 0 & -2 & -5 & 3 \\ 1 & 3 & 4 & -1 \end{vmatrix}.$$

Nun haben wir die Sache so weit vereinfacht, dass wir mit Gewinn nach der ersten Spalte entwickeln können, vgl. Korollar 3.10.5:

$$D = -5 \cdot (-1) \cdot 1 \cdot \begin{vmatrix} -5 & 1 & 9 \\ -1 & 2 & 3 \\ -2 & -5 & 3 \end{vmatrix}.$$

Im nächsten Schritt verwenden wir (b) aus Satz 3.10.4 um die dritte Spalte durch drei zu teilen und machen dann wieder ein paar Gauß'sche Zeilenumformungen, um eine neue Spalte mit nur einem von Null verschiedenen Eintrag zu bekommen:

$$D = 5 \cdot 3 \cdot \begin{vmatrix} -5 & 1 & 3 \\ -1 & 2 & 1 \\ -2 & -5 & 1 \end{vmatrix} \begin{matrix} \\ \leftarrow \\ \cdot(-1) \end{matrix} \begin{matrix} \leftarrow \\ | \\ \cdot(-3) \end{matrix} = 15 \cdot \begin{vmatrix} 1 & 16 & 0 \\ 1 & 7 & 0 \\ -2 & -5 & 1 \end{vmatrix}$$

Nun entwickeln wir nach der extra präparierten dritten Spalte und rechnen fertig:

$$D = 15 \cdot 1 \cdot \begin{vmatrix} 1 & 16 \\ 1 & 7 \end{vmatrix} = 15 \cdot (7 - 16) = -15 \cdot 9 = -135.$$

(b) Für jede Wahl von $a, b \in K$ berechnen wir die folgende $(n \times n)$-Determinante

$$D := \begin{vmatrix} a & b & \dots & b \\ b & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \dots & b & a \end{vmatrix} \begin{matrix} \cdot(-1) \\ \leftarrow \\ \vdots \\ \leftarrow \end{matrix} = \begin{vmatrix} a & b & b & \dots & b \\ b-a & a-b & 0 & \dots & 0 \\ b-a & 0 & a-b & \ddots & \vdots \\ b-a & \vdots & \ddots & \ddots & 0 \\ b-a & 0 & \dots & 0 & a-b \end{vmatrix}.$$

Nun addieren wir die zweite bis n-te Spalte zur ersten Spalte und erhalten:

$$D = \begin{vmatrix} a+(n-1)b & b & b & \dots & b \\ 0 & a-b & 0 & \dots & 0 \\ 0 & 0 & a-b & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & a-b \end{vmatrix}.$$

Der Sinn der ganzen Aktion erschließt sich nun: Es ist eine obere Dreiecksmatrix übriggeblieben. Wenn wir diese transponieren, ändert sich nach Satz 3.10.4 (e) die Determinante nicht und wir erhalten eine untere Dreiecksmatrix, deren Determinante wir nach Beispiel 3.10.3 einfach bestimmen können:

$$D = \begin{vmatrix} a+(n-1)b & 0 & 0 & \dots & 0 \\ b & a-b & 0 & \dots & 0 \\ b & 0 & a-b & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ b & 0 & \dots & 0 & a-b \end{vmatrix} = \big(a+(n-1)b\big)(a-b)^{n-1}.$$

Übungsaufgabe 3.10.7. Rechnen Sie die *Formel von Sarrus* nach:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi.$$

Satz 3.10.8. *Ist* $A \in K^{n\times n}$ *mit* $\mathrm{Rang}(A) < n$*, so ist* $\det(A) = 0$.

Beweis. **1. Schritt:** Ist die erste Spalte von A der Nullvektor, so gilt $\det(A) = 0$. Es seien $a_2, a_3, \dots, a_n \in K^n$ die weiteren Spalten von A. Dann ist dank der Linearität der Determinante in jeder Spalte, vgl. Satz 3.10.4 (b) und Korollar 3.10.5 (a)

$$\begin{aligned} \det(A) &= \begin{vmatrix} 0 & a_2 & a_3 & \dots & a_n \end{vmatrix} = \begin{vmatrix} 0+0 & a_2 & a_3 & \dots & a_n \end{vmatrix} \\ &= \begin{vmatrix} 0 & a_2 & a_3 & \dots & a_n \end{vmatrix} + \begin{vmatrix} 0 & a_2 & a_3 & \dots & a_n \end{vmatrix} = \det(A) + \det(A). \end{aligned}$$

Also ist $\det(A) = 0$.
2. Schritt: Beweis des Satzes.
Die Voraussetzung $\mathrm{Rang}(A) < n$ bedeutet, dass die Spaltenvektoren $a_1, \dots, a_n \in K^n$ von A linear abhängig sind. Also existieren $\lambda_1, \lambda_2, \dots, \lambda_n \in K$, die nicht alle Null sind, mit $\sum_{j=1}^n \lambda_j a_j = 0$. Sei $j_0 \in \{1, \dots, n\}$ ein Index mit $\lambda_{j_0} \neq 0$. Dann gilt nach den verschiedenen Rechenregeln für Determinanten und abschließend Schritt 1

$$\begin{aligned} \det(A) &= |a_1\ a_2\ \dots\ a_{j_0}\ \dots\ a_n| = \frac{1}{\lambda_{j_0}} |a_1\ a_2\ \dots\ \lambda_{j_0} a_{j_0}\ \dots\ a_n| \\ &= \frac{1}{\lambda_{j_0}} \Big| a_1\ a_2\ \dots\ \sum_{j=1}^n \lambda_j a_j\ \dots\ a_n \Big| = \frac{1}{\lambda_{j_0}} |a_1\ a_2\ \dots\ 0\ \dots\ a_n| \\ &= -\frac{1}{\lambda_{j_0}} |0\ a_2\ \dots\ a_1\ \dots\ a_n| = 0. \end{aligned}$$

□

Satz 3.10.9. *Eine Matrix* $A \in K^{n\times n}$ *ist genau dann invertierbar, wenn* $\det(A) \neq 0$ *gilt. In diesem Fall ist* $\det(A^{-1}) = \det(A)^{-1}$.

Beweis. Ist A invertierbar, so gilt $A \cdot A^{-1} = I$. Mit Hilfe von Beispiel 3.10.3 und Satz 3.10.4 (f) ist dann

$$1 = \det(I) = \det(A \cdot A^{-1}) = \det(A) \cdot \det(A^{-1}).$$

Damit haben wir sowohl $\det(A) \neq 0$ als auch $\det(A^{-1}) = \det(A)^{-1}$ gezeigt.
Gilt umgekehrt $\det(A) \neq 0$, so muss nach Satz 3.10.8 der Rang von A voll, d.h. gleich n sein. Also ist mit Hilfe von Satz 3.7.23 die Matrix A invertierbar. □

Korollar 3.10.10. *Es seien $A, B \in K^{n\times n}$ zwei ähnliche Matrizen. Dann gilt* $\det(A) = \det(B)$.

Beweis. Nach Definition der Ähnlichkeit gibt es eine invertierbare Matrix $S \in K^{n\times n}$ mit $B = S^{-1}AS$. Also gilt mit vorstehendem Satz

$$\begin{aligned}\det(B) = \det(S^{-1}AS) &= \det(S^{-1})\det(A)\det(S) = \frac{1}{\det(S)}\det(A)\det(S)\\ &= \det(A).\end{aligned}$$

□

Bemerkung 3.10.11. Ist V ein endlichdimensionaler Vektorraum und $\Phi : V \to V$ eine lineare Abbildung, so besagt Korollar 3.10.10, dass der Wert $\det(M_{\mathcal{B}}^{\mathcal{B}}(\Phi))$ für jede Wahl der Basis $\mathcal{B}$ derselbe ist. Die Determinante ist also eine charakteristische Größe der linearen Abbildung und hängt nicht von der speziellen Wahl der Basis und damit der Abbildungsmatrix ab. Man schreibt deshalb auch $\det(\Phi)$ für diesen Wert und spricht von der Determinante der linearen Abbildung.
Anschaulich ist $|\det(\Phi)|$ der Faktor, um den die Abbildung Φ bei ihrer Anwendung Volumina verändert.

Übungsaufgabe 3.10.12. Beweisen oder widerlegen Sie die folgenden Aussagen:

(a) Für jede orthogonale Matrix $A \in \mathbb{R}^{n\times n}$ gilt $|\det(A)| = 1$.

(b) Für alle $A, B \in K^{n\times n}$ gilt $\det(A + B) = \det(A) + \det(B)$.

3.11. Eigenwerttheorie

In diesem abschließenden Kapitel zur linearen Algebra wollen wir der schon in Beispiel 3.9.4 angekündigten Frage nachgehen, wie man zu einer gegebenen linearen Abbildung $\Phi : V \to V$ eine Basis finden kann, in der die Abbildungsmatrix $M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ möglichst einfach wird.

Definition 3.11.1. *Es sei V ein K-Vektorraum und $\Phi : V \to V$ eine lineare Abbildung. Ein $\lambda \in K$ heißt* Eigenwert *von Φ, falls es einen Vektor $v \in V$ gibt mit $v \neq 0$ und $\Phi(v) = \lambda v$. Ein solches v heißt dann* Eigenvektor *von Φ zum Eigenwert λ.*

Beispiel 3.11.2. Ist $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ die Spiegelung an der x_2-Achse, vgl. Beispiel 3.7.12, so gilt für den ersten Standardbasisvektor $e_1 = \binom{1}{0}$ die Beziehung $\Phi(e_1) = -e_1$, dieser ist also ein Eigenvektor von Φ zum Eigenwert -1. Weiter ist der zweite Standardbasisvektor $e_2 = \binom{0}{1}$ ein Eigenvektor zum Eigenwert 1, denn $\Phi(e_2) = e_2$, vgl. Abbildung 3.3.

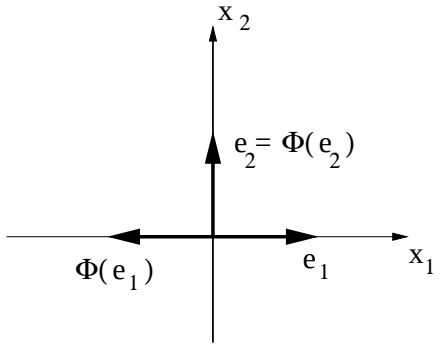

Abbildung 3.3.: Die Spiegelung an der x_2-Achse Φ mit Eigenvektoren

Definition 3.11.3. *Es sei $A \in K^{n\times n}$ eine Matrix. Ein $\lambda \in K$ heißt* Eigenwert *von A, falls es ein $x \in K^n$ mit $x \neq 0$ und $Ax = \lambda x$ gibt. Ein solcher Vektor x heißt dann* Eigenvektor *von A zum Eigenwert λ.*

Wir wollen nun zeigen, dass die beiden Definitionen für Eigenwerte von linearen Abbildungen und Matrizen zusammenpassen. Dazu machen wir uns zunächst klar, dass Eigenwerte bei Basiswechseln erhalten bleiben.

Satz 3.11.4. *Ist $\lambda \in K$ ein Eigenwert von $A \in K^{n\times n}$ und ist $S \in K^{n\times n}$ invertierbar, so ist λ auch ein Eigenwert von $B = S^{-1}AS$.*

Beweis. Es sei $x \in K^n \setminus \{0\}$ ein Eigenvektor von A zum Eigenwert λ. Da S invertierbar ist und $x \neq 0$ gilt, muss auch $y := S^{-1}x \neq 0$ gelten. Für diesen Vektor gilt

$$By = (S^{-1}AS)(S^{-1}x) = S^{-1}A(SS^{-1})x = S^{-1}(Ax) = S^{-1}(\lambda x) = \lambda S^{-1}x = \lambda y,$$

er ist also ein Eigenvektor von B zum Eigenwert λ. Insbesondere hat B den Eigenwert λ. □

Satz 3.11.5. *Es sei V ein endlichdimensionaler K-Vektorraum und $\Phi : V \to V$ eine lineare Abbildung. Dann ist $\lambda \in K$ genau dann ein Eigenwert von Φ, wenn λ für jede Basis $\mathcal{B}$ von V ein Eigenwert von $M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ ist.*

Beweis. „⇒“ Es sei $\mathcal{B}$ eine beliebige Basis von V und $v \in V$ ein Eigenvektor von Φ zum Eigenwert λ. Weiter bezeichnen wir wie üblich mit $\vec{v}$ den Koordinatenvektor von v bezüglich $\mathcal{B}$. Dann ist auch $\vec{v} \neq 0$ und es gilt bezüglich dieser Basis auch

$$\overrightarrow{\Phi(v)} = \overrightarrow{\lambda v} = \lambda\vec{v}.$$

Also ist nach der Definition der Abbildungsmatrix

$$M_{\mathcal{B}}^{\mathcal{B}}(\Phi)\vec{v} = \overrightarrow{\Phi(v)} = \lambda\vec{v}$$

und wir haben gezeigt, dass λ auch ein Eigenwert von $M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ ist.

„⇐“ Es sei $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ eine Basis von V und λ ein Eigenwert von $M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ mit zugehörigem Eigenvektor $x = (x_1, x_2, \ldots, x_n)^T \in K^n$. Dann ist $x \neq 0$ und damit auch $v := \sum_{j=1}^{n} x_j b_j \neq 0_V$. Außerdem ist

$$\overrightarrow{\Phi(v)} = M_{\mathcal{B}}^{\mathcal{B}}(\Phi)\vec{v} = M_{\mathcal{B}}^{\mathcal{B}}(\Phi)x = \lambda x = \lambda\vec{v} = \overrightarrow{\lambda v}.$$

Also ist $\Phi(v) = \lambda v$. □

Bemerkung 3.11.6. Gibt es genügend linear unabhängige Eigenvektoren einer linearen Abbildung, so kann man tatsächlich eine einfache Darstellungsmatrix finden. Wir betrachten auf einem endlichdimensionalen K-Vektorraum V eine lineare Abbildung $\Phi : V \to V$ und setzen voraus, dass es eine Basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ von V gibt, die aus Eigenvektoren von Φ besteht. Wir nennen $\lambda_1, \lambda_2, \ldots, \lambda_n$ die jeweils zugehörigen Eigenwerte, d.h. es gelte $\Phi(b_j) = \lambda_j b_j$ für jedes $j = 1, 2, \ldots, n$. Bezüglich der Basis $\mathcal{B}$ ist dann $\overrightarrow{\Phi(b_j)} = (0, \ldots, 0, \lambda_j, 0, \ldots, 0)^T$ mit dem Eintrag λ_j an der j-ten Stelle. Also ist dann die Abbildungsmatrix

$$M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix}.$$

Solch eine Matrix nennt man eine *Diagonalmatrix*.

Definition 3.11.7. *(a) Es sei V ein endlichdimensionaler Vektorraum. Eine lineare Abbildung $\Phi : V \to V$ heißt* diagonalisierbar*, wenn es eine Basis $\mathcal{B}$ von V gibt, so dass $M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ eine Diagonalmatrix ist.*

(b) Eine Matrix $A \in K^{n\times n}$ heißt diagonalisierbar*, wenn es eine invertierbare Matrix $S \in K^{n\times n}$ gibt, für die $S^{-1}AS$ eine Diagonalmatrix ist.*

Die nächste Frage ist nun wie man die Eigenwerte und Eigenvektoren einer konkret gegebenen linearen Abbildung, bzw. Matrix, bestimmt. Dazu überlegen wir uns für eine Matrix $A \in K^{n\times n}$:

$$\begin{aligned} & \lambda \in K \text{ ist Eigenwert von } A. \\ \iff & \text{Es gibt ein } x \in K^n \setminus \{0\} \text{ mit } Ax = \lambda x. \\ \iff & \text{Es gibt ein } x \in K^n \setminus \{0\} \text{ mit } Ax - \lambda x = 0. \end{aligned}$$

$\iff$ Es gibt ein $x \in K^n \setminus \{0\}$ mit $(A - \lambda I)x = 0$.
$\iff \ker(A - \lambda I) \neq \{0\}$.
$\iff A - \lambda I$ ist nicht invertierbar.
$\iff \det(A - \lambda I) = 0$.

Wir haben also gezeigt:

Satz 3.11.8. *Ein $\lambda \in K$ ist genau dann ein Eigenwert von $A \in K^{n\times n}$, wenn $\det(A - \lambda I) = 0$ ist.*

Ist $A \in K^{n\times n}$, so ist der Ausdruck $\det(A - \lambda I)$ ein Polynom vom Grad n mit Koeffizienten in K. Dieses heißt *charakteristisches Polynom* von A.
Die folgende Übungsaufgabe zeigt, dass nicht nur die Eigenwerte sondern das gesamte charakteristische Polynom bei einem Basiswechsel unverändert bleibt. Man kann also auch vom charakteristischen Polynom einer linearen Abbildung sprechen.

Übungsaufgabe 3.11.9. Sind $A, B \in K^{n\times n}$ ähnliche Matrizen, so gilt $\det(A - \lambda I) = \det(B - \lambda I)$.

Beispiel 3.11.10. Wir betrachten die lineare Abbildung $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ mit

$$\Phi((x_1, x_2, x_3)^T) = \begin{pmatrix} x_2 + x_3 \\ -x_1 + 2x_2 + x_3 \\ x_1 - x_2 \end{pmatrix}$$

aus Beispiel 3.6.20. Außerdem sei $\mathcal{B}$ die Standardbasis. Dann ist

$$A := M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 2 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

und

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} -\lambda & 1 & 1 \\ -1 & 2-\lambda & 1 \\ 1 & -1 & -\lambda \end{vmatrix} \begin{matrix} \leftarrow \\ \cdot 1 \end{matrix} = \begin{vmatrix} -\lambda & 1 & 1 \\ 0 & 1-\lambda & 1-\lambda \\ 1 & -1 & -\lambda \end{vmatrix} \\
&= (1-\lambda) \begin{vmatrix} -\lambda & 1 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & -\lambda \end{vmatrix} \begin{matrix} \leftarrow \\ \cdot(-1) \end{matrix} = (1-\lambda) \begin{vmatrix} -\lambda & 0 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & -\lambda \end{vmatrix} \\
&= (1-\lambda)(-\lambda) \begin{vmatrix} 1 & 1 \\ -1 & -\lambda \end{vmatrix} = -\lambda(1-\lambda)^2.
\end{aligned}$$

Die Eigenwerte der Abbildung sind die Nullstellen dieses Polynoms, also $\lambda_1 = 0$ und $\lambda_2 = 1$.

Zur Bestimmung der zugehörigen Eigenvektoren löst man nun die linearen Gleichungssysteme $(A - \lambda_j I)x = 0$ für $j = 1$ und 2.
Hier ist die Rechnung beispielhaft für $\lambda_1 = 0$. Es ist

$$A - 0 \cdot I = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 2 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Wir lösen also

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ -1 & 2 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{array}\right) \begin{array}{c} \cdot(-1) \\ \leftarrow \\ \\ \end{array} \begin{array}{c} \\ \leftarrow \\ \cdot 1 \end{array} \rightsquigarrow \left(\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{array}\right).$$

Wir haben also $x_1 = x_2 = -x_3$. Alle Eigenvektoren zu $\lambda_1 = 0$ bilden also die Menge

$$\ker(A) \setminus \{0\} = \left\{ \begin{pmatrix} \alpha \\ \alpha \\ -\alpha \end{pmatrix} : \alpha \in \mathbb{R} \setminus \{0\} \right\} = \left\langle \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\rangle \setminus \{0\}.$$

Man vergleiche mit Beispiel 3.6.20.

Satz 3.11.11. *Es sei V ein endlichdimensionaler K-Vektorraum und $\Phi : V \to V$ linear. Weiter sei $A \in K^{n \times n}$. Dann gilt*

(a) Für jedes $\lambda \in K$ ist der Eigenraum *von Φ zum Eigenwert λ*

$$E(\Phi, \lambda) := \{v \in V : \Phi(v) = \lambda v\}$$

ein Untervektorraum von V und der Eigenraum *von A zum Eigenwert λ*

$$E(A, \lambda) := \{x \in K^n : (A - \lambda I)x = 0\} = \ker(A - \lambda I)$$

ein Untervektorraum von K^n.

(b) Eigenvektoren zu verschiedenen Eigenwerten von Φ (bzw. A) sind linear unabhängig.

(c) Sind $\lambda_1, \lambda_2, \ldots, \lambda_r$ verschiedene Eigenwerte von Φ (bzw. A) und $\mathcal{B}_1$, $\mathcal{B}_2$, $\ldots$, $\mathcal{B}_r$ Basen von $E(\Phi, \lambda_1)$, $E(\Phi, \lambda_2)$, $\ldots$, $E(\Phi, \lambda_r)$ (bzw. von $E(A, \lambda_1)$, $E(A, \lambda_2)$, $\ldots$, $E(A, \lambda_r)$), so ist die Menge $\mathcal{B} := \mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_r$ linear unabhängig.

Der Beweis von Teil (a) ist eine direkte Anwendung des Untervektorraumkriteriums, die Aussagen in (b) und (c) wollen wir ohne Beweis hinnehmen.

Übungsaufgabe 3.11.12. Es sei V ein endlichdimensionaler K-Vektorraum mit Basis $\mathcal{B}$, $\Phi : V \to V$ eine lineare Abbildung und $A := M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$. Zeigen Sie, dass der Eigenraum von A zu einem Eigenwert λ genau die Koordinatenvektoren (bezüglich $\mathcal{B}$) der Vektoren aus dem Eigenraum von Φ zum selben λ enthält.

Wir beweisen nun die folgende wichtige Konsequenz aus obigem Satz.

Satz 3.11.13. *Es sei V ein n-dimensionaler K-Vektorraum und $\Phi : V \to V$ eine lineare Abbildung. Dann ist Φ genau dann diagonalisierbar, wenn die Summe der Dimensionen aller Eigenräume gleich n ist.*

Beweis. „⇒“ Ist Φ diagonalisierbar, so gibt es nach Definition eine Basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ von V, für die

$$M_{\mathcal{B}}^{\mathcal{B}}(\Phi) = \begin{pmatrix} \alpha_1 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \alpha_n \end{pmatrix}$$

gilt. Dann gilt für jeden Basisvektor nach der Definition der Abbildungsmatrix

$$\overrightarrow{\Phi(b_j)} = M_{\mathcal{B}}^{\mathcal{B}}(\Phi)\,\overrightarrow{b_j} = \begin{pmatrix} \alpha_1 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \alpha_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \alpha_j \\ \vdots \\ 0 \end{pmatrix} = \alpha_j \overrightarrow{b_j}$$

Also ist $\Phi(b_j) = \alpha_j b_j$, d.h. $\mathcal{B}$ ist eine Basis von Eigenvektoren, was bedeutet, dass tatsächlich die Summe der Dimensionen der Eigenräume gleich n ist.

„⇐“ Es sei nun die Summe der Dimensionen der Eigenräume gleich der Raumdimension n. Wir wählen dann in jedem Eigenraum eine Basis und erhalten so Basen $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_r$ zu jeweils verschiedenen Eigenwerten von Φ. Dann ist nach Satz 3.11.11 (c) die Menge $\mathcal{B} := \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_r$ linear unabhängig und enthält nach Voraussetzung n Vektoren, ist also eine Basis von V. Bezüglich dieser Basis aus Eigenvektoren ist dann $M_{\mathcal{B}}^{\mathcal{B}}(\Phi)$ diagonalisierbar, vgl. Bemerkung 3.11.6. □

Beispiel 3.11.14. Leider gibt es Matrizen und damit auch lineare Abbildungen, die nicht diagonalisierbar sind. Als Beispiel diene $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$. Dann ist

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ 0 & -\lambda \end{vmatrix} = \lambda^2.$$

Also ist nur $\lambda_1 = 0$ ein Eigenwert von A. Dessen Eigenraum berechnet sich als Lösungsmenge des LGS

$$\left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

zu

$$\left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} : \alpha \in \mathbb{R} \right\}.$$

Der Eigenraum hat Dimension 1, es gibt also nur einen linear unabhängigen Eigenvektor.

Bemerkung 3.11.15. Wir haben gesehen, dass die Eigenwerte genau die Nullstellen des charakteristischen Polynoms sind. Das Problem Eigenwerte zu finden, ist also ein Nullstellenproblem für Polynome. Zumindest in theoretischer Hinsicht ist dieses mit dem Fundamentalsatz der Algebra 2.5.14 befriedigend gelöst: Jedes komplexe Polynom zerfällt über $\mathbb{C}$ in Linearfaktoren.
Das ist der Grund, warum Eigenwerttheorie eigentlich immer über $\mathbb{C}$ betrieben wird, selbst wenn alle beteiligten Matrizen rein reell sind.
Auch hier sind alle folgenden Betrachtungen in diesem Sinne zu verstehen.

Satz 3.11.16. *Es sei $n \in \mathbb{N}^*$ und $A \in \mathbb{Q}^{n\times n}$, $\mathbb{R}^{n\times n}$ oder $\mathbb{C}^{n\times n}$. Dann hat A mindestens einen komplexen Eigenwert.*

Beweis. Das charakteristische Polynom von A ist ein Polynom vom Grad n mit Koeffizienten in $\mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$. Die Existenz einer komplexen Nullstelle, und damit eines Eigenwertes, folgt nun aus dem Fundamentalsatz der Algebra 2.5.14. □

Definition 3.11.17. *Eine Matrix $A \in K^{n\times n}$ heißt* symmetrisch, *falls $A = A^T$ gilt.*

Es gilt der folgende Hauptsatz für symmetrische Matrizen, den wir wieder nicht beweisen wollen.

Satz 3.11.18. *Es sei $A \in \mathbb{R}^{n\times n}$ symmetrisch. Dann sind alle Eigenwerte reell und es gibt eine Orthonormalbasis von $\mathbb{R}^n$ aus Eigenvektoren von A. Insbesondere ist jede symmetrische Matrix also diagonalisierbar.*

Beispiel 3.11.19. Es sei $A = \begin{pmatrix} 1 & 4 \\ 4 & -5 \end{pmatrix}$.
Dann ist

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 4 \\ 4 & -5-\lambda \end{vmatrix} = \lambda^2 + 4\lambda - 21$$

und wir erhalten für die Eigenwerte:

$$\lambda_{1/2} = -2 \pm \sqrt{4+21} = -2 \pm 5, \quad \text{also } \lambda_1 = -7, \quad \lambda_2 = 3.$$

Die Eigenvektoren zu $\lambda_1 = -7$ ergeben sich als Lösungen des linearen Gleichungssystems

$$\left(\begin{array}{cc|c} 8 & 4 & 0 \\ 4 & 2 & 0 \end{array}\right) \begin{array}{l} :4 \\ :2 \end{array} \quad \rightsquigarrow \quad \left(\begin{array}{cc|c} 2 & 1 & 0 \\ 2 & 1 & 0 \end{array}\right) \begin{array}{c} \cdot(-1) \\ \leftarrow \end{array} \quad \rightsquigarrow \quad \left(\begin{array}{cc|c} 2 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Also ist $x_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ ein Eigenvektor zum Eigenwert $\lambda_1 = -7$.

Genauso findet man $x_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ als einen Eigenvektor zu $\lambda_2 = 3$.

Damit ist

$$\left\{ \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

eine Orthonormalbasis von $\mathbb{R}^2$ aus Eigenvektoren von A.

Definition 3.11.20. *Es sei $(\cdot|\cdot)$ das Standardskalarprodukt auf $\mathbb{R}^n$. Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ heißt*

(a) positiv definit, *falls $(x|Ax) > 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$ gilt.*

(b) positiv semidefinit, *falls $(x|Ax) \geq 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$ gilt.*

(c) negativ definit, *falls $(x|Ax) < 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$ gilt.*

(d) negativ semidefinit, *falls $(x|Ax) \leq 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$ gilt.*

(e) indefinit, *falls es Vektoren $x, y \in \mathbb{R}^n$ gibt mit $(x|Ax) > 0$ und $(y|Ay) < 0$.*

Bemerkung 3.11.21. Man beachte, dass alle Definitheitsbegriffe von vornherein nur für symmetrische Matrizen A definiert sind. Spricht man von einer positiv oder negativ definiten Matrix, so ist damit immer, auch wenn es nicht dasteht, auch gemeint, dass die Matrix symmetrisch ist.

Im folgenden Satz sind zum Abschluss dieses Abschnitts noch einige Kriterien zum Nachweis von positiver und negativer Defintheit gesammelt. Wir werden diesen Begriffen in einem ganz anderen Zusammenhang in der Matehmatik II noch einmal begegnen.

Satz 3.11.22. *Es sei $A = (\alpha_{jk})_{j,k=1}^n \in \mathbb{R}^{n\times n}$ symmetrisch.*

(a) A ist genau dann positiv definit, wenn $-A$ negativ definit ist.

(b) Die folgenden Aussagen sind äquivalent:

i) A ist positiv definit.

ii) Alle Eigenwerte von A sind größer als Null.

iii) Es gilt für jedes $m \in \{1, 2, \ldots, n\}$, dass $\det(\alpha_{jk})_{j,k=1}^m > 0$.

(c) Die folgenden Aussagen sind äquivalent:

i) A ist negativ definit.

ii) Alle Eigenwerte von A sind kleiner als Null.

iii) Es gilt für jedes $m \in \{1, 2, \ldots, n\}$, dass $(-1)^{m+1} \det(\alpha_{jk})_{j,k=1}^{m} < 0$.

Bemerkung 3.11.23. Die Teildeterminanten $\det(\alpha_{jk})_{j,k=1}^{m}$ in (b)iii) und (c)iii) heißen *Unterminoren* der Determinante. Man kann sich merken: Eine symmetrische Matrix ist genau dann positiv definit, wenn alle Unterminoren positiv sind und genau dann negativ definit, wenn die Unterminoren alternierende Vorzeichen haben, wobei es mit einer negativen Zahl losgehen muss.
Dass es genau so sein muss, macht man sich am besten an Diagonalmatrizen klar.

Wir betrachten noch beispielhaft zwei Matrizen

Beispiel 3.11.24. Es seien

$$A = \begin{pmatrix} 1 & -2 & 2 \\ -2 & 5 & 0 \\ 2 & 0 & 30 \end{pmatrix} \qquad \text{und} \qquad B = \begin{pmatrix} -2 & 3 & 0 \\ 3 & -5 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Dann sind die Unterminoren von A gegeben durch

$$\begin{aligned} \det\big((1)\big) &= 1 > 0 \\ \det \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} &= 5 - 4 = 1 > 0 \\ \det(A) &= 150 - 20 - 120 = 10 > 0, \end{aligned}$$

also ist A positiv definit.
Die Unterminoren von B sind

$$\begin{aligned} \det\big((-2)\big) &= -2 < 0 \\ \det \begin{pmatrix} -2 & 3 \\ 3 & -5 \end{pmatrix} &= 10 - 9 = 1 > 0 \\ \det(B) &= -1 \cdot \begin{vmatrix} -2 & 3 \\ 3 & -5 \end{vmatrix} = -1 \cdot 1 = -1 < 0, \end{aligned}$$

und wir erhalten, dass B negativ definit ist.

Übungsaufgabe 3.11.25. Es sei $(\cdot|\cdot)$ das Standardskalarprodukt auf $\mathbb{R}$ und $A \in \mathbb{R}^{n \times n}$ positiv definit. Zeigen Sie, dass durch

$$(x|y)_A := (x|Ay), \quad x, y \in \mathbb{R}^n,$$

ein Skalarprodukt auf $\mathbb{R}^n$ definiert wird.

4. Ein Ausblick auf die universelle Algebra

4.1. Motivation

Wir haben bislang einige wenige *algebraischer Strukturen* wie Gruppen, Ringe, Körper oder Vektorräume kennengelernt. Sie alle bestehen aus einer oder mehreren Sorten von Objekten und Operationen auf diesen Objekten. Die verschiedenen in der Informatik betrachteten *Datentypen* geben Anlaß zu weiteren algebraischen Strukturen. Deshalb ist es nützlich einen allgemeinen, d.h. "universellen", Begriff algebraischer Strukturen einzuführen und Begriffe wie Homomorphismus, Unteralgebra, Kongruenzrelation etc. auf diese allgemeinen Strukturen zu erweitern.

Oft gibt es mehrere verschiedene Implementierungen ein und derselben Datenstruktur. Diese verschiedenen Implementierungen werden als *Algebren* aufgefaßt und was ihnen gemeinsam ist wird durch ihre *Spezifikation* zum Ausdruck gebracht, d.h. eine Liste von Axiomen wie im Falle der bereits betrachteten traditionellen algebraischen Strukturen.

Als Beispiel betrachten wir folgende Spezifikation des Datentyps der Listen

LIST	$\equiv$
sorts	elem, list
funcs	empty : $\rightarrow$ list cons : elem $\times$ list $\rightarrow$ list head : list $\rightarrow$ elem tail : list $\rightarrow$ list
axioms	$\text{head}(\text{cons}(a, l)) = a$ $\text{tail}(\text{cons}(a, l)) = l$ $\text{tail}(\text{empty}) = \text{empty}$

in der zwei Sorten, die der Elemente und die der Listen, postuliert werden zusammen mit den Operationen empty, cons, head, tail und list, von denen angegeben wird, wieviele Argumente welcher Sorte sie erwarten und von welcher Sorte das Resultat ist. Die intendierte Bedeutung dieser Operationen wird durch Axiome zum Ausdruck gebracht, welche in vorliegendem Fall die spezielle Form von Gleichungen haben. Im allgemeinen reichen aber Gleichungen nicht aus wie z.B. im

Falle von Körpern, wo wir die Existenz eines inversen Elements durch die Formel

$$\forall x \exists y \, xy = 1$$

ausgedrückt haben, in der auch Quantoren vorkommen.

4.2. Signaturen, Algebren und Homomorphismen

Zuerst formalisieren wir den Begriff einer Signatur, in der festgelegt wird, welche Sorten und welche Operationen betrachtet werden. Wir verwenden zu diesem Zweck folgende Notation: für eine Menge A bezeichne A^* die Menge der Wörter über dem Alphabet A.[1]

Definition 4.2.1. *Eine* Signatur *bzw. ein* Strukturtyp *ist ein Tripel*

$$\Sigma = (S, F, \mathrm{ar})$$

wobei S und F Mengen und $\mathrm{ar} : F \to S^* \times S$ *eine Funktion ist. S heißt Menge der* Sortensymbole*, F heißt Menge der* Funktionssymbole *und* ar *heißt* Stelligkeitsfunktion *(englisch* arity function*) der Signatur Σ.*
Statt S, F und ar schreiben wir oft $\Sigma_{\mathrm{sort}}, \Sigma_{\mathrm{func}}, \Sigma_{\mathrm{ar}}$, und für $\mathrm{ar}(f) = (s_1 \dots s_n, s)$ schreiben wir auch $f : s_1 \times \cdots \times s_n \to s$.

Als nächstes definieren wir für beliebige Signaturen Σ den Begriff einer Σ-Algebra.

Definition 4.2.2. *Sei $\Sigma = (S, F, \mathrm{ar})$ eine Signatur. Eine Σ-Algebra A besteht aus der Angabe*

i) *einer Menge A_s für jedes $s \in S$ und*

ii) *einer Funktion $f^A : A_{s_1} \times \cdots \times A_{s_n} \to A_s$ für jedes Funktionssymbol $f \in F$, wobei $\mathrm{ar}(f) = (s_1 \dots s_n, s)$.*

A_s heißt Trägermenge *der Sorte s in A (englisch* carrier set*) und f^A heißt* Interpretation von f in A.

Für $\vec{s} = s_1 \dots s_n \in S^*$ bezeichne $A_{\vec{s}}$ als Abkürzung die Menge $A_{s_1} \times \cdots \times A_{s_n}$. Also schreiben wir $f^A : A_{\vec{s}} \to A_s$ für $\mathrm{ar}(f) = (\vec{s}, s)$.
Falls $\mathrm{ar}(f) = (\varepsilon, s)$, so schreiben wir $f : \to s$. In diesem Fall nennt man f eine *Konstante* der Sorte s, da $A_\varepsilon = \{\emptyset\}$.
Wenn S aus genau einer Sorte besteht, so ist jedes $(\vec{s}, s) \in S^* \times S$ eindeutig durch $n := \mathrm{length}(\vec{s})$ bestimmt und wir können die Stelligkeit eines Funktionssymbols eindeutig durch die *Anzahl* seiner Argumente beschreiben. Allerdings ist in den meisten Informatikanwendungen mehr als eine Sorte vonnöten (wie z.B. bei Listen, Stacks, labelled trees…).
Als nächstes verallgemeineren wir den Begriff Homomorphismus auf Σ-Algebren.

[1] Wir schreiben ε für das *leere* Wort.

Definition 4.2.3. *Sei $\Sigma = (S, F, \mathrm{ar})$ eine Signatur und seien A und B Σ–Algebren. Ein* Σ–Homomorphismus *(oder kurz* Homomorphismus*) von A nach B ist eine Familie von Funktionen $(h_s : A_s \to B_s)_{s \in S}$, so dass für $(s_1 \dots s_n, s) \in S^* \times S$ und alle $f \in F$ mit $\mathrm{ar}(f) = (s_1 \dots s_n, s)$ gilt:*

$$h_s(f^A(x_1, \dots, x_n)) = f^B(h_{s_1}(x_1), \dots, h_{s_n}(x_n))$$

für alle $i = 1, \dots, n$ und $x_i \in A_{s_i}$.

Intuitiv bedeutet obige Bedingung, dass Algebraoperationen und (Komponenten von) h *vertauschbar* sind, der Homomorphismus also die Operationen der Algebra *respektiert*. Dies veranschaulicht folgendes Diagramm:

$$\begin{array}{ccc} A_{\vec{s}} & \xrightarrow{f^A} & A_s \\ \Big\downarrow{\scriptstyle h_{\vec{s}}} & & \Big\downarrow{\scriptstyle h_s} \\ B_{\vec{s}} & \xrightarrow[f^B]{} & B_s \end{array}$$

wobei $A_{\vec{s}} = A_{s_1} \times \cdots \times A_{s_n}$, $B_{\vec{s}} = B_{s_1} \times \cdots \times B_{s_n}$ und $h_{\vec{s}}$ definiert ist als die Funktion $h_{\vec{s}}(x_1, \dots, x_n) = (h_{s_1}(x_1), \dots, h_{s_n}(x_n))$.

Man zeigt leicht, dass Homorphismen unter Komposition abgeschlossen sind (wobei $(h' \circ h)_s = h'_s \circ h_s$).

Definition 4.2.4. *Ein Homomorphismus $h : A \to B$ von Σ-Algebren heißt* Isomorphismus*, wenn alle h_s Bijektionen sind.*

Wie üblich zeigt man, dass $h : A \to B$ genau dann ein Isomorphismus ist, wenn es einen (eindeutig bestimmten) Homomorphismus $h^{-1} : B \to A$ gibt mit $h^{-1} \circ h = \mathrm{id}_A$ und $h \circ h^{-1} = \mathrm{id}_B$.

4.3. Unteralgebren und Faktoralgebren

Analog zu Untergruppe, Unterring, Untervektorraum etc. definieren wir den Begriff einer Unteralgebra.

Definition 4.3.1. *Sei Σ eine Signatur und A eine Σ-Algebra. Eine* Unteralgebra *von A ist eine Σ-Algebra B, sodass $B_s \subseteq A_s$ für alle $s \in \Sigma_{\mathrm{sort}}$ und für alle Operationen $f : s_1 \times \cdots \times s_n \to s$ in Σ gilt, dass*

$$f^A(b_1, \dots, b_n) = f^B(b_1, \dots, b_n)$$

für alle $b_1 \in B_{s_1}$, …, $b_n \in B_{s_n}$.

Wir überlassen es als Übung zu zeigen, dass B genau dann Unteralgebra von A ist, wenn eine Homomorphismus $\iota : B \to A$ existiert mit $\iota_s(b) = b$ für alle $b \in B_s$. Solche Homomorphismen nennen wir *Inklusion(shomomorphism)en.*

Ein Homomorphismus $h : A \to B$ heißt *surjektiv*, wenn alle $h_s : A_s \to B_s$ surjektive Funktionen sind.

Satz 4.3.2. *Sei Σ eine Signatur und $h : A \to B$ ein Σ-Homomorphismus. Dann gibt es genau eine Unteralgebra $\iota : C \to B$, sodass ein surjektiver Homomorphismus $e : A \to C$ existiert mit $h = \iota \circ e$*

Beweis. Notwendigerweise muss C_s das Bild von $h_s : A_s \to B_s$ sein. Dass dann C eine Unteralgebra von B ist folgt unmittelbar aus der Homomorphieeigenschaft von h, da $f^B(h_{s_1}(a_1), \ldots, h_{s_n}(a_n)) = h_s(f^A(a_1, \ldots, a_n)) \in C_s$ für $f : s_1 \times \cdots \times s_n \to s$. Der eindeutig bestimmte surjektive Homomorphismus $e : A \to C$ ist gegeben durch $e_s(a) = h_s(a)$. □

Als nächstes definieren wir den Begriff einer Kongruenz für Σ-Algebren.

Definition 4.3.3. *Sei $\Sigma = (S, F, \mathrm{ar})$ eine Signatur und A eine Σ-Algebra. Eine* Kongruenz(relation) *auf A ist eine S-indizierte Familie $R = (R_s)_{s \in S}$ von Äquivalenzrelation R_s auf A_s, die zusammen folgende* Kongruenzbedingung *erfüllen: für alle $f : s_1 \times \cdots \times s_n \to s$ in S folgt aus $x_1 R_{s_1} y_1$, …, $x_n R_{s_n} y_n$, dass auch $f(x_1, \ldots, x_n) R_s f(y_1, \ldots, y_n)$.*

Für eine Kongruenz R auf A kann man wie folgt eine Faktoralgebra $Q = A/R$ konstruieren. Für $s \in S$ sei Q_s definiert als A_s/R_s und für $f : s_1 \times \cdots \times s_n \to s$ in F sei f^Q gegeben durch $f^Q(\widetilde{a_1}, \ldots, \widetilde{a_n}) = \widetilde{f(a_1, \ldots, a_n)}$. Beachte, dass f^Q wohldefiniert ist aufgrund der Kongruenzeigenschaft von R. Der *Quotientenhomomorphismus* $q : A \to Q$ ist gegeben durch $q_s(a) = \widetilde{a}$ für $a \in A_s$. Offenbar ist q surjektiv.

Satz 4.3.4. *Sei $h : A \to B$ ein Homorphismus von Σ-Algebren. Für jede Sorte s sei $(R_h)_s$ definiert als $\{(x, y) \in A_s \times A_s \mid h_s(x) = h_s(y)\}$. Dann ist $R_h = \big((R_h)_s\big)_{s \in S}$ eine Kongruenzrelation auf A. Dann gibt es genau einen Homomorphismus $m : A/R_h \to B$ mit $h = m \circ q$, wobei $q : A \to A/R_h$ der zu R_h gehörige Quotientenhomomorphismus ist.*

Beweis. Der gesuchte Homomorphismus m is gegeben durch $m(\widetilde{a}) = h(a)$. Er ist eindeutig, da q surjektiv ist. □

Daraus folgt, dass A/R_h zum Bild von A unter h isomorph ist (siehe Satz 4.3.2).

Übungsaufgabe 4.3.5. Für jeden surjektiven Homomorphismus $h : A \to B$ ist B isomorph zu A/R_h.

Übungsaufgabe 4.3.6. Sei A eine Σ-Algebra. Dann sind sowohl die Menge der Unteralgebren von A als auch die Menge der Kongruenzrelationen auf A unter beliebigen Durchschnitten abgeschlossen.

4.4. Terme, termerzeugte Algebren und das Prinzip der Terminduktion

Sei $\Sigma = (S, F, \mathrm{ar})$ eine Signatur.

Definition 4.4.1. *Die* freie Termalgebra $T(\Sigma)$ *über* Σ *ist* induktiv *wie folgt definiert. Wenn* $t_1 \in T(\Sigma)_{s_1}$, ..., $t_n \in T(\Sigma)_{s_n}$ *und* $f : s_1 \times \cdots \times s_n \to s$ *in* F *ist, dann ist* $f(t_1, \ldots, t_n) \in T(\Sigma)_s$. *Außerdem ist* $f^{T(\Sigma)}$ *gegeben durch* $f^{T(\Sigma)}(t_1, \ldots, t_n) = f(t_1, \ldots, t_n)$.

Die Algebra T_Σ ist also rein *syntaktischer* Natur. Man kann zeigen, dass für jede Σ-Algebra A genau ein Σ-Homomorphismus $h : T(\Sigma) \to A$ existiert, der syntaktische Objekte (Terme) in der Struktur A (*Semantik*) interpretiert. Man versteht h als *Interpretationsfunktion*.

Eine Konsequenz des ersten Teils von Übungsaufgabe 4.3.6 ist, dass jede Σ-Algebra A eine kleinste Unteralgebra $T_\Sigma(A)$ enthält, nämlich das Bild des eindeutig bestimmten Homomorphismus $h : T(\Sigma) \to A$.

Definition 4.4.2. *Eine* Σ*-Algebra heißt* termerzeugt, *wenn* $T_\Sigma(A)$ *mit* A *übereinstimmt.*

Für termerzeugte Algebren gilt folgendes Induktionsprinzip.

Satz 4.4.3. (Terminduktion)
Sei A *eine termerzeugte* Σ*-Algebra und* $(P_s)_{s \in S}$ *eine Familie von Prädikaten auf* A_s *(d.h.* $P_s \subseteq A_s$*). Dann gilt* $\forall x \in A_s : P_s(x)$ *für alle* $s \in S$ *genau dann, wenn* P *unter den Operationen* $f : s_1 \times \cdots \times s_n \to s$ *von* A *abgeschlossen ist, d.h.* $f^A(a_1, \ldots, a_n) \in P_s$, *wann immer* $a_1 \in P_{s_1}$, ..., $a_n \in P_{s_n}$.

Im Falle der 1-sortigen Algebra mit Trägermenge $\mathbb{N}$, Konstante 0 und einstelliger Operation $s(n) = n + 1$ besagt Satz 4.4.3, dass das Induktionsprinzip für $\mathbb{N}$ gilt. Satz 4.4.3 verallgemeinert dies auf beliebige termerzeugte Algebren. Da die in der Informatik betrachteten Datentypen meist termerzeugte Algebren sind, eröffet Satz 4.4.3 die Möglichkeit, allquantifizierte Aussagen über solche Datentypen mithilfe von Terminduktion zu beweisen.
Beispielsweise zeigt man leicht mit Terminduktion, dass für termerzeugte Algebren A zu jeder Algebra B höchstens ein Homomorphismus $A \to B$ existiert. Seien $h, h' : A \to B$ Homomorphismen. Für jede Sorte s sei $P_s = \{a \in A_s : h(a) = h'(a)\}$. Man zeigt leicht (Übung!), dass $(P_s)_{s \in S}$ unter den Operationen von A abgeschlossen ist. Somit folgt, dass $A_s = P_s$ für alle Sorten s, und somit $h = h'$.
Als weiteres Beispiel betrachten wir die Spezifikation

LIST_APP	$\equiv$
sorts	elem, list
funcs	c1, c2 : elem empty : $\rightarrow$ list cons : elem $\times$ list $\rightarrow$ list head : list $\rightarrow$ elem tail : list $\rightarrow$ list append : list $\times$ list $\rightarrow$ list
axioms	head(cons(a, l)) = a tail(cons(a, l)) = l tail(empty) = empty append(empty,l) = l append(cons(a, l),l') = cons(a,append(l, l'))

Sei Σ die Untersignatur, wo die Operationen head, tail und append weggelassen werden. Sei A eine Algebra, die die Spezifikation LIST_APP erfüllt und für die das *Redukt*[2] $A_{|\Sigma}$ termerzeugt ist. Dann kann man mit Terminduktion (bzgl. Σ zeigen, dass die Operation append in A assoziativ ist.

[2]d.h., man vergisst die (Interpretationen der) Operationen head, tail und append

5. Analysis – Teil I: Konvergenz und Stetigkeit

5.1. Die reellen Zahlen

Wir erinnern uns an den Begriff eines *angeordneten Körpers*:
Dies ist ein Körper K mit einer Totalordnung $\leq$, für die gilt:

- $\forall a, b, c \in K : a \leq b \Longrightarrow a + c \leq b + c$ und
- $\forall a, b, c \in K : (a \leq b \text{ und } 0 \leq c) \Longrightarrow ac \leq bc$.

Definition 5.1.1. *Die Menge der* reellen Zahlen $\mathbb{R}$ *ist der kleinste angeordnete Körper, der* $\mathbb{Z}$ *enthält, und das* Vollständigkeitsaxiom

Jede nichtleere Teilmenge, die eine obere Schranke besitzt, hat ein Supremum.

erfüllt.

Bemerkung 5.1.2. Auch die rationalen Zahlen $\mathbb{Q}$ sind ein angeordneter Körper, der $\mathbb{Z}$ enthält, aber dieser erfüllt nicht das Vollständigkeitsaxiom, denn $\{x \in \mathbb{Q} : x^2 < 2\}$ hat obere Schranken aber kein Supremum in $\mathbb{Q}$, vgl. Beispiel 1.3.9 (b).

Definition 5.1.3. *Eine Teilmenge* $M \subseteq \mathbb{R}$ *heißt*

(a) nach oben (unten) beschränkt, *wenn sie eine obere (untere) Schranke besitzt.*

(b) beschränkt, *wenn sie nach oben und unten beschränkt ist.*

Satz 5.1.4. *Jede nach unten beschränkte, nichtleere Teilmenge von* $\mathbb{R}$ *besitzt ein Infimum.*

Beweis. Es sei $M \subseteq \mathbb{R}$ eine nach unten beschränkte und nichtleere Menge. Dann gibt es eine untere Schranke C von M. Wir betrachten nun die Teilmenge $\widehat{M} := \{-x : x \in M\}$ von $\mathbb{R}$, die ebenfalls nichtleer ist. Da C eine untere Schranke von M ist, gilt $C \leq x$ für alle $x \in M$. Das bedeutet, dass $-C \geq -x$ für alle $x \in M$ ist. Anders formuliert, erhalten wir $-C \geq y$ für jedes $y \in \widehat{M}$.
Also ist $-C$ eine obere Schranke von $\widehat{M}$. Nach dem Vollständigkeitsaxiom existiert nun $s := \sup(\widehat{M})$ und wir wollen zeigen, dass $-s$ das Infimum von M ist.

Dazu zeigen wir zunächst, dass $-s$ eine untere Schranke ist: Es sei $x \in M$ beliebig. Dann gilt $-x \in \widehat{M}$ und es gilt nach der Konstruktion von s auf jeden Fall $-x \leq s$. Also ist $x \geq -s$ für jedes $x \in M$, was zeigt, dass $-s$ eine untere Schranke von M ist.
Sei nun $t \in \mathbb{R}$ eine weitere untere Schranke von M. Dann ist mit der selben Argumentation wie oben $-t$ eine obere Schranke von $\widehat{M}$. Nun ist s die kleinste obere Schranke von $\widehat{M}$, also muss $s \leq -t$ gelten. Damit ist aber auch $-s \geq t$. Also ist $-s$ die größte untere Schranke von M, d.h. das Infimum. □

Eine wichtige Rolle in der Analysis spielt die Betragsfunktion, an die wir ebenfalls noch schnell erinnern wollen.

Definition 5.1.5. *Die Funktion* $|\cdot| : \mathbb{R} \to \mathbb{R}$ *mit*

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

heißt Betragsfunktion *und* $|x|$ *heißt* Betrag *von* x.

Es gelten die folgenden Rechenregeln, vgl. Satz 2.5.12 und Beispiel 3.4.2 (a):

Satz 5.1.6. *Für alle* $x, y \in \mathbb{R}$ *gilt*

(a) $|x| \geq 0$,

(b) $|x| = |-x|$,

(c) $\pm x \leq |x|$,

(d) $|xy| = |x| \cdot |y|$,

(e) $|x| = 0$ *genau dann, wenn* $x = 0$,

(f) $|x + y| \leq |x| + |y|$ (Dreiecksungleichung),

Übungsaufgabe 5.1.7. Zeigen Sie die *umgekehrte Dreiecksungleichung*

$$\big||x| - |y|\big| \leq |x - y| \quad \text{für alle } x, y \in \mathbb{R}.$$

Wir beschließen diesen Abschnitt mit der Definition von Intervallen.

Definition 5.1.8. *Es seien zwei Zahlen* $a, b \in \mathbb{R}$ *mit* $a < b$ *gegeben. Dann heißen*

- $(a, b) := \{x \in \mathbb{R} : a < x < b\}$ offenes Intervall,
- $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ abgeschlossenes Intervall,
- $(a, b] := \{x \in \mathbb{R} : a < x \leq b\}$ *und*

- $[a, b) := \{x \in \mathbb{R} : a \leq x < b\}$ halboffene Intervalle.

Um auch die Fälle von Halbstrahlen abzudecken, definieren wir weiter:

- $[a, \infty) := \{x \in \mathbb{R} : a \leq x\}$,
- $(a, \infty) := \{x \in \mathbb{R} : a < x\}$,
- $(-\infty, \infty) := \mathbb{R}$.
- $(-\infty, a] := \{x \in \mathbb{R} : x \leq a\}$,
- $(-\infty, a) := \{x \in \mathbb{R} : x < a\}$,

Schließlich schreiben wir

- $\mathbb{R}_+ := [0, \infty)$,
- $\mathbb{R}_- := (-\infty, 0]$.

5.2. Wurzeln, Fakultäten und Binomialkoeffizienten

Eine wichtige Schlussfolgerung aus dem Vollständigkeitsaxiom ist die Existenz der n-ten Wurzeln in $\mathbb{R}_+$.
Bevor wir in diese Betrachtungen einsteigen, definieren wir der Vollständigkeit halber noch die ganzzahligen Potenzen.

Definition 5.2.1. *Für jedes $x \in \mathbb{R}$ und jedes $n \in \mathbb{N}^*$ ist*

(a) $x^n := \underbrace{x \cdot x \cdot \ldots \cdot x}_{n \text{ Faktoren}}$,

(b) $x^{-n} := \dfrac{1}{x^n}$, *falls $x \neq 0$, sowie*

(c) $x^0 := 1$.

An dieser Stelle sei noch gewarnt, dass das Potenzieren nicht assoziativ ist, d.h. im Allgemeinen ist $(x^n)^m \neq x^{(n^m)}$. Um Klammern zu sparen gilt die Konvention, dass x^{n^m} als $x^{(n^m)}$ zu lesen ist.
Der zentrale Satz für die Existenz der n-ten Wurzeln ist der folgende.

Satz 5.2.2. *Für jedes $a \in \mathbb{R}_+$ und alle $n \in \mathbb{N}^*$ gibt es genau ein $x \in \mathbb{R}_+$ mit $x^n = a$.*

Beweisidee. Man zeigt zunächst, dass $x^n \leq y^n \iff x \leq y$ für jede Wahl von $x, y \in \mathbb{R}_+$ und alle $n \in \mathbb{N}^*$ gilt.
Für den Nachweis der Eindeutigkeit seien $x, y \in \mathbb{R}$ mit $x^n = a = y^n$ gegeben. Dann gilt $x^n \leq y^n$ und $y^n \leq x^n$. Nach der Vorüberlegung ist dann $x \leq y$ und $y \leq x$, also $x = y$.
Für die Existenz betrachtet man zunächst den Fall $a = 0$. Dann ist offensichtlich $x = 0$ eine Lösung. Sei also ab jetzt $a > 0$. Wir betrachten die Menge $M := \{x \in \mathbb{R}_+ : x^n \leq a\}$. Dann ist $0 \in M$, also ist $M \neq \emptyset$. Weiterhin ist M nach oben

beschränkt, z.B. ist $1 + a$ eine obere Schranke (Achtung, a ist im Allgemeinen keine!).
Also hat die Menge M nach dem Vollständigkeitsaxiom ein Supremum x und für dieses gilt $x^n = a$. □

Definition 5.2.3. *Es seien $a \in \mathbb{R}_+$ und $n \in \mathbb{N}^*$. Die eindeutige Zahl $x \in \mathbb{R}_+$ mit $x^n = a$ heißt* n-te Wurzel *von a und man schreibt $x = \sqrt[n]{a}$. Für den wichtigsten Fall $n = 2$ gibt es die Konvention $\sqrt{a} := \sqrt[2]{a}$.*

Satz 5.2.4. *Es seien $q \in \mathbb{Q}$ und $m, p \in \mathbb{Z}$, sowie $n, r \in \mathbb{N}^*$ so, dass $q = m/n = p/r$ ist. Dann gilt für jedes $x \in \mathbb{R}_+$*

$$(\sqrt[n]{x})^m = (\sqrt[r]{x})^p.$$

Beweis. Für jedes $x \in \mathbb{R}_+$ gilt

$$\begin{aligned}\left((\sqrt[n]{x})^m\right)^r = (\sqrt[n]{x})^{mr} = (\sqrt[n]{x})^{np} = \left((\sqrt[n]{x})^n\right)^p = x^p \quad \text{und}\\ \left((\sqrt[r]{x})^p\right)^r = (\sqrt[r]{x})^{pr} = \left((\sqrt[r]{x})^r\right)^p = x^p.\end{aligned}$$

Also folgt aus der Eindeutigkeit der Wurzel die Behauptung. □

Definition 5.2.5. *Für jedes $x \in \mathbb{R}_+$ und jedes $q = n/m \in \mathbb{Q}$ mit $n \in \mathbb{Z}$ und $m \in \mathbb{N}^*$ ist die* rationale Potenz *definiert durch*

$$x^q = x^{n/m} := (\sqrt[m]{x})^n.$$

Bemerkung 5.2.6. Auch für rationale Exponenten gelten die bekannten Rechenregeln für Potenzen: Für alle $x, y \in \mathbb{R}_+ \setminus \{0\}$ und alle $p, q \in \mathbb{Q}$ gilt

- $x^p x^q = x^{p+q}$
- $\frac{x^p}{x^q} = x^{p-q}$
- $x^p y^p = (xy)^p$
- $\frac{x^p}{y^p} = \left(\frac{x}{y}\right)^p$
- $(x^p)^q = x^{pq}$

Definition 5.2.7. *(a) Es sei $n \in \mathbb{N}^*$. Dann wird die Zahl*

$$n! := 1 \cdot 2 \cdot \ldots \cdot n$$

als n Fakultät *bezeichnet.*

Weiterhin definieren wir $0! := 1$.

(b) Es seien $n, k \in \mathbb{N}$ mit $k \leq n$. Dann heißt

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

Binomialkoeffizient *„n über k“.*

Bemerkung 5.2.8. Die beiden Größen $n!$ und $\binom{n}{k}$ haben auch eine anschauliche Bedeutung:

- $n!$ ist die Anzahl der möglichen Reihenfolgen von n unterscheidbaren Dingen.
- $\binom{n}{k}$ ist die Anzahl der Möglichkeiten aus n unterscheidbaren Dingen genau k auszuwählen.

Satz 5.2.9. *Es seien $n, k \in \mathbb{N}$ mit $k \leq n$ und $a, b \in \mathbb{R}$. Dann gilt*

(a) $\binom{n}{0} = \binom{n}{n} = 1 \qquad \text{und} \qquad \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$

(b) $a^{n+1} - b^{n+1} = (a-b)\sum_{k=0}^{n} a^{n-k}b^k.$

(c) $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k.$ (Binomialformel)

Beweis. (a) Es ist

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1 \quad \text{und} \quad \binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!} = 1.$$

Außerdem ist

$$\begin{aligned}\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\ &= \frac{n!(n-k+1)}{k!(n-k+1)!} + \frac{n!k}{k!(n-k+1)!} = \frac{n!(n-k+1+k)}{k!(n-k+1)!} \\ &= \frac{n!(n+1)}{k!(n-k+1)!} = \frac{(n+1)!}{k!(n+1-k)!} = \binom{n+1}{k}.\end{aligned}$$

(b) Es gilt

$$\begin{aligned}(a-b)\sum_{k=0}^{n} a^{n-k}b^k &= a\sum_{k=0}^{n} a^{n-k}b^k - b\sum_{k=0}^{n} a^{n-k}b^k \\ &= \sum_{k=0}^{n} a^{n-k+1}b^k - \sum_{k=0}^{n} a^{n-k}b^{k+1} \\ &= a^{n+1} + \sum_{k=1}^{n} a^{n-k+1}b^k - \sum_{k=0}^{n-1} a^{n-k}b^{k+1} - b^{n+1} \\ &= a^{n+1} + \sum_{k=0}^{n-1} a^{n-k}b^{k+1} - \sum_{k=0}^{n-1} a^{n-k}b^{k+1} - b^{n+1} \\ &= a^{n+1} - b^{n+1}.\end{aligned}$$

(c) Dies ist eine gute Auffrischungsübung in vollständiger Induktion. □

Eine Summe, wie sie im Beweis von Teil (b) auftritt, bei der sich bis auf zwei alle Summanden gegenseitig wegheben, nennt man auch anschaulich *Teleskopsumme*.

Bemerkung 5.2.10. Mit der zweiten Formel aus Satz 5.2.9 (a) kann man sich die Binomialkoeffizienten gut merken. Schreibt man diese in ein dreieckiges Schema:

$$\begin{array}{ccccccc} & & & \binom{0}{0} & & & \\ & & \binom{1}{0} & & \binom{1}{1} & & \\ & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & \\ \vdots & & \vdots & & \vdots & & \vdots \end{array}$$

so sagt diese Formel gerade, dass man einen Eintrag bekommt, indem man die beiden diagonal links und rechts darüber zusammenzählt. Das ergibt das sogenannte *Pascal'sche Dreieck*

$$\begin{array}{ccccccccccc} & & & & & 1 & & & & & \\ & & & & 1 & & 1 & & & & \\ & & & 1 & & 2 & & 1 & & & \\ & & 1 & & 3 & & 3 & & 1 & & \\ & 1 & & 4 & & 6 & & 4 & & 1 & \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \end{array}$$

Aus diesem kann man nun nach Teil (c) des obigen Satzes z.B. direkt ablesen, dass

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$

gilt.

5.3. Konvergenz von Folgen

Wir wollen uns nun dem zentralen Thema der Analysis zuwenden, der mathematisch exakten Behandlung des unendlich Kleinen und unendlich Großen. Beispielsweise kann es darum gehen, unendlich viele Zahlen aufzuaddieren, wie in der unendlichen Summe

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots,$$

der wir im Folgenden einen exakten Sinn geben werden.
Hierbei ist einige Vorsicht geboten, denn beim Umgang mit dem Unendlichen können sehr unintuitive Dinge passieren, so dass anschauliche Argumentationen schnell in die Irre führen können. Unser Ziel wird also zunächst sein, eine exakte mathematische Definition für solche Grenzwertfragen zu geben. Diese Aufgabe wollen wir in diesem für alles weitere zentralen Kapitel angehen.

Wir erinnern noch einmal an den Begriff einer Folge aus Beispiel 3.1.2 (d). Eine Folge war dort eine Abbildung von $\mathbb{N}$ in einen Körper K. In Erweiterung der dortigen Definition lassen wir nun statt einem Körper allerdings eine beliebige nichtleere Menge X zu und sagen, dass eine Folge eine Abbildung $a : \mathbb{N} \to X$ ist. Um klar zu machen, was die Zielmenge dieser Abbildung ist, nennen wir a genauer eine *Folge in* X. Statt Folge in $\mathbb{R}$, bzw. $\mathbb{C}$ sagt man auch oft *reelle Folge*, bzw. *komplexe Folge*.
Wir schreiben wieder a_n statt $a(n)$ und bezeichnen die Folge a mit $(a_n)_{n\in\mathbb{N}}$ oder $(a_n)_{n\geq 0}$ oder $(a_0, a_1, a_2, \dots)$. Manchmal werden wir auch etwas verkürzt einfach (a_n) schreiben.
Zuweilen ist es praktisch mit der Zählung nicht bei Null, sondern einer anderen natürlichen Zahl zu beginnen. Wir schreiben dann beispielsweise $(a_n)_{n\geq 4}$ oder $(a_4, a_5, a_6, \dots)$.
Die meisten Betrachtungen in diesem Abschnitt gelten für die Körper $\mathbb{R}$ und $\mathbb{C}$ gleichermaßen. In diesem Abschnitt steht deshalb $\mathbb{K}$ für einen dieser beiden Körper.

5.3.1. Der Konvergenzbegriff und wichtige Beispiele

Definition 5.3.1. *(a) Es sei (a_n) eine Folge in $\mathbb{K}$ und $a \in \mathbb{K}$. Die Folge (a_n) heißt* konvergent gegen a, *falls für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert mit*

$$|a_n - a| < \varepsilon \quad \text{für alle } n \geq n_0.$$

In diesem Fall heißt a der Grenzwert *oder* Limes *von (a_n) und wir schreiben*

$$\lim_{n\to\infty} a_n = a \quad \text{oder} \quad a_n \to a \ (n \to \infty).$$

(b) Ist (a_n) eine Folge in $\mathbb{K}$, die gegen kein $a \in \mathbb{K}$ konvergiert, so heißt diese divergent.

Man kann zeigen, dass eine Folge in $\mathbb{K}$ höchstens einen Grenzwert haben kann. Wenn eine Folge konvergiert, ist der Limes also eindeutig.

Beispiel 5.3.2.

(a) Wir betrachten die Folge $(a_n) = (1/n)_{n\geq 1} = (1, 1/2, 1/3, 1/4, \dots)$.

Behauptung: (a_n) ist konvergent und es gilt $\lim_{n\to\infty} a_n = 0$.

Beweis. Sei $\varepsilon > 0$. Dann gibt es ein $n_0 \in \mathbb{N}$ mit $n_0 > 1/\varepsilon$. Also ist $1/n_0 < \varepsilon$ und wir haben für alle $n \geq n_0$

$$|a_n - a| = |a_n - 0| = |a_n| = \frac{1}{n} \leq \frac{1}{n_0} < \varepsilon. \qquad \square$$

Eine solche Folge, die gegen Null konvergiert, nennt man auch eine *Nullfolge*.

(b) Es sei $(a_n) = ((-1)^n)_{n\in\mathbb{N}} = (1, -1, 1, -1, 1, -1, \dots)$.
Behauptung: Die Folge (a_n) divergiert.

Beweis. Wir nehmen an, es gäbe ein $a \in \mathbb{K}$ mit $a_n \to a$ $(n \to \infty)$. Dann gibt es zu $\varepsilon = 1$ ein $n_0 \in \mathbb{N}$, so dass für jedes $n \geq n_0$ die Ungleichung $|a_n - a| < 1$ gilt. Für $n \geq n_0$ gilt dann aber mit Hilfe der Dreiecksungleichung

$$2 = |a_n - a_{n+1}| = |a_n - a + a - a_{n+1}| \leq |a_n - a| + |a - a_{n+1}| < 1 + 1 = 2.$$

Also folgt $2 < 2$, ein Widerspruch. □

(c) Sei $a_n = \dfrac{n^2 + 2n - 1}{n^2 + 2}$, $n \in \mathbb{N}$.
Behauptung: (a_n) konvergiert und $\lim_{n\to\infty} a_n = 1$.

Beweis. Es gilt

$$|a_n - 1| = \left|\frac{n^2 + 2n - 1 - n^2 - 2}{n^2 + 2}\right| = \frac{|2n - 3|}{n^2 + 2} \leq \frac{|2n - 3|}{n^2} \leq \frac{2n + 3}{n^2},$$

wobei wir bei der letzten Abschätzung die Dreiecksungleichung angewendet haben. Nun verwenden wir noch, dass für alle $n \geq 1$ gilt $2n+3 \leq 2n+3n = 5n$ und erhalten damit

$$|a_n - 1| \leq \frac{5n}{n^2} = \frac{5}{n}.$$

Sei nun $\varepsilon > 0$. Dann gibt es wieder ein $n_0 \in \mathbb{N}$ mit $n_0 > 5/\varepsilon$. Also haben wir nach obiger Abschätzung für alle $n \geq n_0$

$$|a_n - 1| \leq \frac{5}{n} \leq \frac{5}{n_0} < \varepsilon.$$ □

Übungsaufgabe 5.3.3. Es sei (a_n) eine Folge in $\mathbb{K}$ und $a \in \mathbb{K}$. Zeigen Sie:

(a) Gibt es eine reelle Nullfolge (α_n) mit

$$|a_n - a| \leq \alpha_n \quad \text{für alle } n \in \mathbb{N},$$

so konvergiert (a_n) gegen a.

(b) Die Folge (a_n) konvergiert genau dann gegen a, wenn die Folge $(|a_n - a|)$ gegen Null konvergiert.

Definition 5.3.4. *(a) Eine Folge (a_n) in $\mathbb{K}$ heißt* beschränkt*, wenn die Menge $\{a_n : n \in \mathbb{N}\} = \{a_0, a_1, a_2, \dots\}$ beschränkt in $\mathbb{K}$ ist.*

(b) Ist $\mathbb{K} = \mathbb{R}$, so setzen wir weiter

$$\sup_{n\in\mathbb{N}} a_n := \sup_{n=0}^{\infty} a_n := \sup\{a_n : n \in \mathbb{N}\},$$
$$\inf_{n\in\mathbb{N}} a_n := \inf_{n=0}^{\infty} a_n := \inf\{a_n : n \in \mathbb{N}\}.$$

Satz 5.3.5. *Jede konvergente Folge in $\mathbb{K}$ ist beschränkt.*

Beweis. Sei (a_n) eine konvergente Folge in $\mathbb{K}$ mit Grenzwert a. Nach der Definition der Konvergenz existiert zu $\varepsilon = 1$ ein $n_0 \in \mathbb{N}$ mit $|a_n - a| < 1$ für alle $n \geq n_0$. Wir setzen $C := \max\{|a_0|, |a_1|, |a_2|, \dots, |a_{n_0-1}|, 1 + |a|\}$. Dann gilt für alle $n < n_0$ sofort $|a_n| \leq C$ und auch für alle $n \geq n_0$ gilt diese Ungleichung, denn

$$|a_n| = |a_n - a + a| \leq |a_n - a| + |a| < 1 + |a| \leq C.$$

Zusammengenommen gilt also $|a_n| \leq C$ für alle $n \in \mathbb{N}$ und somit die Behauptung. □

Warnung 5.3.6. Die Umkehrung von Satz 5.3.5 ist *falsch*! Es gibt durchaus beschränkte Folgen, die nicht konvergieren, vgl. Beispiel 5.3.2 (b).

Die Berechnung von Grenzwerten ist allein über die Definition sehr mühsam und wird bei größeren Ausdrücken mehr oder weniger unmöglich. Eine große Hilfe ist der folgende Satz, der es erlaubt, die Berechnung komplizierter Grenzwerte auf die Betrachtung einfacherer Einzelteile zu reduzieren.

Satz 5.3.7 (Grenzwertsätze)**.** *Es seien (a_n), (b_n) und (c_n) Folgen in $\mathbb{K}$. Dann gilt:*

(a) Ist $\lim_{n\to\infty} a_n = a$, so gilt $\lim_{n\to\infty} |a_n| = |a|$.

(b) Gilt $\lim_{n\to\infty} a_n = a$ und $\lim_{n\to\infty} b_n = b$, so folgt

i) $\lim_{n\to\infty}(a_n + b_n) = a + b$.

ii) $\lim_{n\to\infty}(\alpha a_n) = \alpha a$ für alle $\alpha \in \mathbb{K}$.

iii) $\lim_{n\to\infty}(a_n \cdot b_n) = a \cdot b$.

iv) Ist zusätzlich $b_n \neq 0$ für alle $n \in \mathbb{N}$ und $b \neq 0$, so ist $\lim_{n\to\infty} a_n/b_n = a/b$.

Ist $\mathbb{K} = \mathbb{R}$, so gilt außerdem

(c) Ist $a_n \leq b_n$ für alle $n \in \mathbb{N}$ und $\lim_{n\to\infty} a_n = a$ sowie $\lim_{n\to\infty} b_n = b$, so folgt $a \leq b$.

(d) *Ist* $a_n \leq c_n \leq b_n$ *für alle* $n \in \mathbb{N}$ *und sind* (a_n) *und* (b_n) *konvergent mit* $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = a$*, so ist auch die Folge* (c_n) *konvergent und es gilt* $\lim_{n\to\infty} c_n = a$ (Sandwich-Theorem).

Beweis. (a) Sei $\varepsilon > 0$. Da (a_n) gegen a konvergiert, gibt es dann ein $n_0 \in \mathbb{N}$ mit $|a_n - a| < \varepsilon$ für alle $n \geq n_0$. Für alle diese n gilt dann nach der umgekehrten Dreiecksungleichung, s. Übungsaufgabe 5.1.7

$$\big||a_n| - |a|\big| \leq |a_n - a| < \varepsilon.$$

Also konvergiert die Folge $(|a_n|)$ gegen $|a|$.

(b) Die Teile (b)i)-(b)iii) verbleiben als Übungsaufgabe. Wir beweisen hier (b)iv).

Da (b_n) gegen b konvergiert, konvergiert nach (a) auch $(|b_n|)$ gegen $|b|$. Da weiter $b \neq 0$ ist, gilt $|b| > 0$. Zu $\varepsilon := |b|/2 > 0$ gibt es also ein $n_1 \in \mathbb{N}$ mit $||b_n| - |b|| \leq |b|/2$ für alle $n \geq n_1$. Für all diese n gilt dann mit der Dreiecksungleichung

$$|b| = \big||b|\big| = \big||b| - |b_n| + |b_n|\big| \leq \big||b| - |b_n|\big| + |b_n| \leq \frac{|b|}{2} + |b_n|.$$

Zieht man $|b|/2$ auf beiden Seiten ab, haben wir also für alle $n \geq n_1$

$$|b_n| \geq \frac{|b|}{2}, \quad \text{bzw.} \quad \frac{1}{|b_n|} \leq \frac{2}{|b|}.$$

Das liefert wiederum für alle diese n

$$\Big|\frac{1}{b_n} - \frac{1}{b}\Big| = \Big|\frac{b - b_n}{bb_n}\Big| = \frac{|b_n - b|}{|b||b_n|} \leq \frac{2|b_n - b|}{|b|^2} = \frac{2}{|b|^2}|b_n - b|.$$

Da die Folge (b_n) gegen b geht, konvergiert nach Übungsaufgabe 5.3.3 (b) die Folge $(|b_n - b|)$ gegen Null. Also konvergiert nach Teil (b)ii) dieses Satzes auch die Folge $(2|b_n - b|/|b|^2)$ gegen Null. Da dieses außerdem eine reelle Folge ist, haben wir nach Teil (a) von Übungsaufgabe 5.3.3, dass $(1/b_n)$ gegen $1/b$ konvergiert.

Die Konvergenz von (a_n/b_n) gegen a/b folgt nun aus (b)iii).

(c) Wir nehmen an, es wäre $a > b$. Dann ist $\varepsilon := (a - b)/2 > 0$ und dank der Konvergenz von (a_n) und (b_n) gibt es nun ein $n_1 \in \mathbb{N}$, so dass $b_n \in (b - \varepsilon, b + \varepsilon)$ und $a_n \in (a - \varepsilon, a + \varepsilon)$ für alle $n \geq n_1$ gilt. Da

$$b + \varepsilon = b + \frac{a - b}{2} = \frac{a}{2} + \frac{b}{2} = a - \frac{a - b}{2} = a - \varepsilon$$

gilt, haben wir also für diese n auch $b_n < a_n$ im Widerspruch zur Voraussetzung.

(d) Sei $\varepsilon > 0$. Dann gibt es ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ sowohl $|a_n - a| < \varepsilon$ als auch $|b_n - a| < \varepsilon$ gilt. Hieraus und aus der Voraussetzung folgern wir für alle diese n

$$a - \varepsilon < a_n \leq c_n \leq b_n < a + \varepsilon.$$

Also ist $a - \varepsilon < c_n < a + \varepsilon$ oder, anders ausgedrückt, $-\varepsilon < c_n - a < \varepsilon$, d.h. $|c_n - a| < \varepsilon$ für alle $n \geq n_0$ und damit konvergiert die Folge (c_n) gegen a. □

Warnung 5.3.8. Die Aussage in (c) gilt nicht mit „<“ statt „≤“! Als Beispiel können die Folgen $(a_n) = (0)_{n\geq 1}$ und $(b_n) = (1/n)_{n\geq 1}$ dienen. Dann gilt nämlich $a_n < b_n$ für alle $n \geq 1$, aber die beiden Grenzwerte sind gleich.

Wir wollen nun an zwei Beispielen zeigen, wie mit Hilfe dieses Satzes komplizierte Grenzwerte angegangen werden können.

Beispiel 5.3.9. (a) Sei $p \in \mathbb{N}^*$ fest gewählt und $a_n = 1/n^p$ für $n \in \mathbb{N}^*$. Dann gilt für alle $n \in \mathbb{N}^*$ die Ungleichung $n \leq n^p$ und damit

$$0 \leq a_n = \frac{1}{n^p} \leq \frac{1}{n}.$$

Da sowohl die Folge, die konstant Null ist, als auch die Folge $(1/n)$ gegen Null konvergiert, ist damit nach Satz 5.3.7 (d) auch die Folge (a_n) konvergent und ebenfalls eine Nullfolge.

(b) Wir untersuchen

$$a_n = \frac{n^2 + 2n + 3}{n^2 + 3}, \quad n \in \mathbb{N}.$$

Dazu kürzen wir den Bruch durch die höchste auftretende Potenz:

$$a_n = \frac{n^2 + 2n + 3}{n^2 + 3} = \frac{1 + \frac{2}{n} + \frac{3}{n^2}}{1 + \frac{3}{n^2}} \longrightarrow \frac{1 + 0 + 0}{1 + 0} = 1 \quad (n \to \infty).$$

Dabei stützen wir uns für die Berechnung des Grenzwertes von $(1/n^2)$ auf das Beispiel in (a) und zum Zusammenbau des Gesamtausdruckes auf (b)i), (b)ii) und (b)iv) aus Satz 5.3.7.

Dieses Vorgehen (Kürzen durch die höchste auftretende Potenz) ist bei allen Grenzwerten der Form „Polynom in n geteilt durch Polynom in n“ Erfolg versprechend.

Bemerkung 5.3.10. Hier finden Sie weitere wichtige Beispiele von konvergenten Folgen.

(a) Ist (a_n) eine konvergente Folge in $\mathbb{R}$ mit Grenzwert a und gilt $a_n \geq 0$ für alle $n \in \mathbb{N}$, so ist für jedes $p \in \mathbb{N}^*$ auch $\lim_{n\to\infty} \sqrt[p]{a_n} = \sqrt[p]{a}$.

(b) Die Folge $(q^n)_{n\in\mathbb{N}}$ mit $q \in \mathbb{R}$ konvergiert genau dann, wenn $q \in (-1, 1]$ ist und es gilt

$$\lim_{n\to\infty} q^n = \begin{cases} 1, & \text{falls } q = 1, \\ 0, & \text{falls } -1 < q < 1. \end{cases}$$

Ist $q \in \mathbb{C}$ mit $|q| < 1$, so gilt ebenfalls $\lim_{n\to\infty} q^n = 0$.

(c) $\lim_{n\to\infty} \sqrt[n]{c} = 1$ für jedes $c \in \mathbb{R}_+$.

(d) $\lim_{n\to\infty} \sqrt[n]{n} = 1$.

(e) Die Folge

$$a_n := \left(1 + \frac{1}{n}\right)^n, \qquad n \geq 1,$$

ist konvergent. Ihr Grenzwert

$$\mathrm{e} := \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$$

heißt *Eulersche Zahl*. Diese ist eine irrationale Zahl mit

$$\mathrm{e} \approx 2,718281828459.$$

Warnung 5.3.11. Die Folge (a_n) aus Teil (e) in obiger Bemerkung bietet eine gute Gelegenheit vor einem verbreiteten Fehler bei der Bestimmung von Grenzwerten zu warnen, der Unterteilung in „eiligere“ und „trägere“ n. Falsch ist nämlich folgende Überlegung: Die Folge $(1 + 1/n)$ geht offensichtlich gegen 1, also geht (a_n) gegen 1^n und das ist immer 1, was zu dem Ergebnis führe (a_n) würde gegen 1 streben. Das ist, vgl. oben, grob falsch. Der Grund ist folgender: Bei dieser Überlegung werden nicht alle n in der Formel gleich behandelt. Das n innerhalb der Klammer wird (quasi als Vorhut) zuerst nach ∞ geschickt, während das n im Exponenten noch warten muss, also zum „trägen“ n ernannt wird. Das geht nicht. Merke: Alle n sind gleich!
Mit der gleichen Berechtigung könnte man auch argumentieren, dass $1 + 1/n$ immer echt größer als 1 ist und da q^n für alle $q > 1$ divergiert, divergiert der Ausdruck in der Klammer, also auch die ganze Folge. Nun ist das andere n zum Warten gezwungen worden, und das Ergebnis ist genauso falsch wie das erste.
Diese Erörterung zeigt aber, was hier passiert. Das $1/n$ in der Klammer bringt den Ausdruck immer näher an 1, während es groß wird, und macht es dem n im Exponenten damit immer schwerer, die Werte von a_n zu vergrößern. Die beiden beeinflussen den Wert also in verschiedene Richtungen und die Frage ist, wer dabei erfolgreicher ist: Schafft es das n in der Klammer, die Sache nach 1 zu drücken, oder ist das n im Exponent stärker und die Folge divergiert? Daran, dass das Ergebnis irgendwo zwischen 2 und 3 liegt, sieht man, dass die beiden sich in magischer Weise im Gleichgewicht halten.

Beispiel 5.3.12. Es bleiben noch zwei wichtige Beispiele nachzutragen.

(a) Den folgenden Trick muss man mal gesehen haben. Es ist für alle $n \in \mathbb{N}$

$$0 \leq \sqrt{n+1} - \sqrt{n} = \frac{(\sqrt{n+1} - \sqrt{n})(\sqrt{n+1} + \sqrt{n})}{\sqrt{n+1} + \sqrt{n}} = \frac{(n+1) - n}{\sqrt{n+1} + \sqrt{n}}$$
$$= \frac{1}{\sqrt{n+1} + \sqrt{n}} \leq \frac{1}{2\sqrt{n}} = \frac{1}{2}\sqrt{\frac{1}{n}}.$$

Da wegen (a) aus Bemerkung 5.3.10 und Beispiel 5.3.2 (a) auch $(\sqrt{1/n})$ gegen Null geht, folgt damit aus dem Sandwich-Theorem

$$\lim_{n\to\infty} \left(\sqrt{n+1} - \sqrt{n}\right) = 0.$$

(b) Wir betrachten für jedes $q \in \mathbb{C}$ die Folge

$$a_n := \sum_{k=0}^{n} q^k = 1 + q + q^2 + q^3 + \cdots + q^n, \qquad n \in \mathbb{N}.$$

Dann gilt

$$(1-q)\sum_{k=0}^{n} q^k = \sum_{k=0}^{n} q^k - \sum_{k=0}^{n} q^{k+1} = q^0 + \sum_{k=1}^{n} q^k - \sum_{k=0}^{n-1} q^{k+1} - q^{n+1}$$
$$= 1 - q^{n+1} + \sum_{k=1}^{n} q^k - \sum_{k=1}^{n} q^k = 1 - q^{n+1}.$$

Wir erhalten so für $q \neq 1$ die *geometrische Summenformel*

$$\sum_{k=0}^{n} q^k = \frac{1 - q^{n+1}}{1-q}, \qquad q \neq 1.$$

Mit Bemerkung 5.3.10 (b) gilt also für $|q| < 1$

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1 - q^{n+1}}{1-q} = \frac{1}{1-q}, \qquad |q| < 1.$$

Definition 5.3.13. *Eine Folge (a_n) in $\mathbb{R}$* divergiert bestimmt nach ∞ *($-\infty$) und wir schreiben $\lim_{n\to\infty} a_n = \infty$ ($-\infty$), wenn es für jedes $C \geq 0$ ein $n_0 \in \mathbb{N}$ gibt, so dass $a_n \geq C$ ($a_n \leq -C$) für alle $n \geq n_0$ gilt.*

5.3.2. Konvergenzkriterien

Wir wollen uns als nächstes mit dem Monotonie-Verhalten von Folgen auseinandersetzen. Das geht naturgemäß nicht in $\mathbb{C}$, so dass wir uns auf $\mathbb{R}$ einschränken müssen.

Definition 5.3.14. *Eine reelle Folge (a_n) heißt*

(a) monoton wachsend, *wenn $a_{n+1} \geq a_n$ für alle $n \in \mathbb{N}$ gilt.*

(b) monoton fallend, *wenn $a_{n+1} \leq a_n$ für alle $n \in \mathbb{N}$ gilt.*

(c) monoton, *wenn sie monoton wachsend oder monoton fallend ist.*

Damit können wir folgendes Konvergenzkriterium beweisen.

Satz 5.3.15 (Monotonie-Kriterium)**.** *Ist die reelle Folge (a_n) nach oben (bzw. unten) beschränkt und monoton wachsend (bzw. fallend), so ist (a_n) konvergent und es gilt*

$$\lim_{n\to\infty} a_n = \sup_{n\in\mathbb{N}} a_n \quad (\textit{bzw. } \lim_{n\to\infty} a_n = \inf_{n\in\mathbb{N}} a_n).$$

Beweis. Es sei (a_n) nach oben beschränkt und monoton wachsend, sowie $a := \sup_{n\in\mathbb{N}} a_n$. Wählen wir nun ein $\varepsilon > 0$, so ist sicherlich $a - \varepsilon$ keine obere Schranke von $\{a_n : n \in \mathbb{N}\}$. Damit muss aber ein $n_0 \in \mathbb{N}$ existieren, so dass $a_{n_0} > a - \varepsilon$ ist und somit haben wir unsere Folge umzingelt, denn es gilt nun wegen der Monotonie und der Beschränktheit von (a_n) für alle $n \geq n_0$:

$$a - \varepsilon < a_{n_0} \leq a_n \leq a < a + \varepsilon$$

und hiermit $|a_n - a| < \varepsilon$ für alle $n \geq n_0$. Da $\varepsilon > 0$ beliebig war, sind wir mit der ungeklammerten Aussage fertig. Die Aussage für monoton fallende Folgen beweist man analog. □

Wir betrachten ein Beispiel für die Anwendung dieses Satzes.

Beispiel 5.3.16. Wir betrachten eine *rekursiv definierte* Folge, die gegeben ist durch

$$a_0 := \sqrt[3]{6} \quad \text{und} \quad a_{n+1} := \sqrt[3]{6 + a_n},\ n \in \mathbb{N}.$$

Also ist

$$a_1 = \sqrt[3]{6 + \sqrt[3]{6}}, \quad a_2 = \sqrt[3]{6 + \sqrt[3]{6 + \sqrt[3]{6}}}, \quad a_3 = \sqrt[3]{6 + \sqrt[3]{6 + \sqrt[3]{6 + \sqrt[3]{6}}}}, \ \ldots.$$

So abstrus dieses Beispiel auch aussieht, in dieser Weise gegebene Folgen treten sehr häufig auf, so liefert z.B. jedes iterative Näherungsverfahren eine solche Folge. Wie untersuchen wir aber ein solches Monstrum auf Konvergenz? Wir wenden unser Monotoniekriterium an, zeigen also, dass (a_n) nach oben beschränkt und monoton wachsend ist. Genauer gesagt beweisen wir

1. $a_n < 2$ und $a_{n+1} > a_n$ für alle $n \in \mathbb{N}$,

2. (a_n) konvergiert und $\lim_{n\to\infty} a_n = 2$.

Das erste ist eine Induktionsübung:

Induktionsanfang: Es gilt $a_0 = \sqrt[3]{6} < \sqrt[3]{8} = 2$ und $a_1 = \sqrt[3]{6+a_0} > \sqrt[3]{6} = a_0$, da $a_0 \geq 0$ ist. Also ist die Aussage für $n = 0$ richtig.
Induktionsvoraussetzung: Für ein $n \in \mathbb{N}$ gelte $a_n < 2$ und $a_{n+1} > a_n$.
Induktionsschritt: Es ist mit Hilfe der Induktionsvoraussetzung

$$a_{n+1} = \sqrt[3]{6+a_n} < \sqrt[3]{6+2} = \sqrt[3]{8} = 2$$

und

$$a_{n+2} = \sqrt[3]{6+a_{n+1}} > \sqrt[3]{6+a_n} = a_{n+1}.$$

Wir wenden uns also der Konvergenz in 2. zu.

Nach Satz 5.3.15 wissen wir nun, dass (a_n) konvergiert, und dass $\lim_{n\to\infty} a_n = \sup_{n\in\mathbb{N}} a_n \leq 2$ ist, denn 2 ist eine obere Schranke der Folge. Außerdem wissen wir, dass $a_{n+1}^3 = 6 + a_n$ für alle $n \in \mathbb{N}$ gilt. Nach den Rechenregeln für Grenzwertbildung aus Satz 5.3.7 konvergieren bei dieser Gleichung die Folgen auf beiden Seiten des Gleichheitszeichens. Gehen wir also in dieser Gleichung zum Limes über, so erhalten wir für $a := \lim_{n\to\infty} a_n$ die Beziehung $a^3 = 6 + a$, bzw. $a^3 - a - 6 = 0$. Eine Lösung dieser Gleichung ist $a = 2$. Dividieren wir diese ab, so erhalten wir $(a-2)(a^2+2a+3) = 0$ und $a^2 + 2a + 3 = (a+1)^2 + 2 = 0$ hat keine weiteren reellen Lösungen. Also muss $a = 2$ sein und wir haben $\lim_{n\to\infty} a_n = 2$.

Noch ein Kommentar zum Verfahren. Obwohl es nicht immer zum Ziel führt, ist dieses doch ein starkes Hilfsmittel zur Behandlung rekursiver Folgen, das man immer wieder mit Gewinn verwenden kann.

Übungsaufgabe 5.3.17. (Babylonisches Wurzelziehen)**.** Zeigen Sie, dass für jedes $x \in \mathbb{R}$ und jeden beliebigen Startwert $a_0 > 0$ die Folge (a_n) mit $a_{n+1} = \frac{a_n + x/a_n}{2}$, $n \in \mathbb{N}$, konvergiert mit Grenzwert $\sqrt{x}$.

Definition 5.3.18. *Eine Folge (a_n) in $\mathbb{K}$ heißt* Cauchy-Folge*, wenn für jedes $\varepsilon > 0$ ein Index $n_0 \in \mathbb{N}$ existiert, so dass*

$$|a_n - a_m| < \varepsilon \qquad \text{für alle } n, m \geq n_0$$

gilt.

Satz 5.3.19. *Jede konvergente Folge in $\mathbb{K}$ ist eine Cauchy-Folge.*

Beweis. Sei (a_n) eine konvergente Folge in $\mathbb{K}$ mit Grenzwert a und sei $\varepsilon > 0$. Dann gibt es ein $n_0 \in \mathbb{N}$, so dass $|a_n - a| < \varepsilon/2$ für alle $n \geq n_0$. Also ist für alle $n, m \geq n_0$ mit der Dreiecksungleichung

$$|a_n - a_m| = |a_n - a + a - a_m| \leq |a_n - a| + |a_m - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad \square$$

Tatsächlich gilt für reelle und komplexe Folgen auch die Umkehrung, die wir hier nicht beweisen wollen.

Satz 5.3.20 (Cauchy-Kriterium)**.** *Eine Folge in* $\mathbb{K}$ *konvergiert genau dann, wenn sie eine Cauchy-Folge ist.*

Bemerkung 5.3.21. (a) Bemerkenswert an den beiden behandelten Konvergenzkriterien ist, dass sie beide eine Konvergenzaussage liefern, ohne dass man eine a-priori Vermutung über den Grenzwert braucht.

(b) Während die Aussage, dass jede konvergente Folge eine Cauchy-Folge ist, allgemeingültig ist, geht in den Beweis der umgekehrten Implikation massiv das Vollständigkeitsaxiom ein. Dass die Implikation ohne dieses Axiom falsch wird, kann man sich klarmachen, wenn man sich die Folge aus Übungsaufgabe 5.3.17 mit $a = 2$ als Folge in $\mathbb{Q}$ anschaut.

Da sie in $\mathbb{R}$ konvergiert, ist sie dort (und dann auch in $\mathbb{Q}$) eine Cauchy-Folge, aber in $\mathbb{Q}$ ist sie nicht konvergent, denn ein etwaiger Grenzwert in $\mathbb{Q}$ wäre dann auch ein Grenzwert in $\mathbb{R}$ und wegen der Eindeutigkeit des Grenzwertes also $\sqrt{2}$, was nun mal nicht in $\mathbb{Q}$ liegt.

5.3.3. Teilfolgen und Häufungswerte

Definition 5.3.22. *Es sei* (a_n) *eine Folge in* $\mathbb{K}$*. Ein* $a \in \mathbb{K}$ *heißt* Häufungswert *der Folge, falls für jedes* $\varepsilon > 0$ *die Menge* $\{n \in \mathbb{N} : |a_n - a| < \varepsilon\}$ *unendlich viele Elemente hat.*

Offensichtlich ist der Grenzwert einer konvergenten Folge ein Häufungswert derselben, denn für jedes $\varepsilon > 0$ liegen ja dann sogar alle bis auf die ersten paar Folgeglieder näher als ε am Grenzwert. Aber es gibt auch divergente Folgen die Häufungswerte haben, z.B. hat $((-1)^n)_{n\in\mathbb{N}}$ die Häufungswerte 1 und -1 und die komplexe Folge $(i^n)_{n\in\mathbb{N}}$ hat deren vier, nämlich i, -1, $-i$ und 1. Es gibt sogar reelle (bzw. komplexe) Folgen, die jede reelle (bzw. komplexe) Zahl als Häufungswert haben.

Definition 5.3.23. *Es sei* (a_n) *eine Folge in* $\mathbb{K}$*. Ist* $\{n_1, n_2, n_3, \dots\} \subseteq \mathbb{N}$ *eine unendliche Menge von Indizes mit* $n_1 < n_2 < n_3 < \dots$*, so heißt die Folge* $(a_{n_k})_{k\in\mathbb{N}}$ *eine* Teilfolge *von* (a_n)*.*

Beispielsweise ist $(a_0, a_2, a_4, a_6, \dots)$ die Teilfolge der Folgeglieder mit geradem Index. Eine andere Teilfolge ist $(a_0, a_1, a_4, a_9, a_{16}, \dots)$. Keine Teilfolgen wären $(a_0, a_0, a_2, a_2, a_4, a_4, \dots)$ oder $(a_2, a_1, a_4, a_3, a_6, a_5, \dots)$.
Den engen Zusammenhang zwischen Teilfolgen, Häufungswerten und Konvergenz beschreibt der folgende Satz.

Satz 5.3.24. *Es sei* (a_n) *eine Folge in* $\mathbb{K}$*. Dann gilt*

(a) *Ein $\alpha \in \mathbb{K}$ ist genau dann ein Häufungswert von (a_n), wenn eine Teilfolge (a_{n_k}) von (a_n) existiert, die gegen α konvergiert.*

(b) *Ist (a_n) konvergent mit Grenzwert a, so konvergiert auch jede Teilfolge von (a_n) gegen a.*

(c) *Ist (a_n) konvergent, so hat (a_n) genau einen Häufungswert, nämlich den Grenzwert $\lim_{n\to\infty} a_n$.*

Beweis. Wir behandeln hier nur den Beweis von (a), die Teile (b) und (c) verbleiben als Übungsaufgabe.

„⇒“ Es sei $\alpha \in \mathbb{K}$ ein Häufungswert von (a_n). Dann existiert insbesondere für $\varepsilon = 1$ ein $n_1 \in \mathbb{N}$ mit $|a_{n_1} - \alpha| < 1$. Da es auch für $\varepsilon = 1/2$ unendlich viele Folgenglieder von (a_n) gibt, die weniger als $1/2$ von α entfernt sind, muss es auch ein $n_2 \in \mathbb{N}$ mit $n_2 > n_1$ geben, so dass $|a_{n_2} - \alpha| < 1/2$ gilt. Genauso finden wir ein $n_3 \in \mathbb{N}$ mit $n_3 > n_2$, so dass $|a_{n_3} - \alpha| < 1/3$ gilt.

Verfahren wir immer weiter so, erhalten wir schließlich eine Folge von Indizes $n_1, n_2, n_3, \ldots$ mit $n_1 < n_2 < n_3 < \ldots$, so dass

$$|a_{n_k} - \alpha| \leq \frac{1}{k} \quad \text{für alle } k \in \mathbb{N} \tag{5.1}$$

gilt. Nun ist (a_{n_k}) eine Teilfolge von (a_n), und diese konvergiert nach Übungsaufgabe 5.3.3 gegen α.

„⇐“ Sei nun (a_{n_k}) eine Teilfolge von (a_n), die gegen ein $\alpha \in \mathbb{K}$ konvergiert. Zu gegebenem $\varepsilon > 0$ existiert dann ein $k_0 \in \mathbb{N}$, so dass $|a_{n_k} - \alpha| < \varepsilon$ für alle $k \geq k_0$ gilt. Insbesondere gilt damit $|a_n - \alpha| \leq \varepsilon$ für die unendlich vielen Indizes $n_{k_0}, n_{k_0+1}, n_{k_0+2}, \ldots$, d.h. α ist ein Häufungswert von (a_n). □

Übungsaufgabe 5.3.25. Erklären Sie jemandem aus ihrem Semester, warum die Umkehrung der Aussage in Teil (c) von Satz 5.3.24 im Allgemeinen falsch ist.

5.4. Asymptotik

Ein Thema, bei dem Folgen in der Informatik prominent auftauchen, ist die Laufzeit- bzw. Aufwandsabschätzung von Algorithmen. Dabei ist a_n die Laufzeit (der Aufwand) des Algorithmus', wenn der verarbeitete Datensatz den Umfang $n \in \mathbb{N}$ hat.
Bei genauerer Betrachtung kommt es nicht auf den genauen Wert von a_n an, sondern nur auf das qualitative Verhalten, d.h. wie schnell a_n mit wachsendem n groß wird. Ist also z.B. die Laufzeit bei der Bearbeitung von n Eingabedaten $a_n = 3n^4 + 2\ \mu s$, so ist die „+2“ reichlich egal, und auch das „3·“ interessiert nur

am Rande, die wichtige Information ist, dass der Aufwand in der vierten Potenz wächst.
Wir wollen dieses „ungefähr“ rechnen nun „exakt“ formalisieren. Das ist kein Widerspruch, sondern diese Idee führt zum sehr leistungsfähigen „O-Kalkül“, dem Sie in Ihrem Studium noch häufig begegnen werden.

Definition 5.4.1. *(a) Wir bezeichnen mit*

$$F_+ := \big\{(a_n) \text{ Folge in } \mathbb{R} : a_n > 0 \text{ für alle } n \in \mathbb{N}\big\}.$$

(b) Es sei $(b_n) \in F_+$. Dann definieren wir die Landau-Symbole *durch*

$$O(b_n) := \Big\{(a_n) \in F_+ : \Big(\frac{a_n}{b_n}\Big)_{n\in\mathbb{N}} \text{ beschränkt}\Big\}, \qquad \text{(Groß-O von } b_n\text{)}$$
$$o(b_n) := \Big\{(a_n) \in F_+ : \lim_{n\to\infty} \frac{a_n}{b_n} = 0\Big\}, \qquad \text{(Klein-O von } b_n\text{)}.$$

Andere übliche Schreibweisen für $(a_n) \in O(b_n)$, gesprochen „(a_n) ist ein Groß-/Klein-O von b_n“, sind $a_n \in O(b_n)$ und $a_n = O(b_n)$, bzw. $a_n \in o(b_n)$ und $a_n = o(b_n)$.

Bemerkung 5.4.2. (a) Zur oben angeführten, sehr oft verwendeten, Schreibweise $a_n = O(b_n)$, bzw. $a_n = o(b_n)$ ist eine deutliche Warnung angebracht, denn das „=“-Zeichen wird hier nicht im mathematisch üblichen Sinne gebraucht. Z.B. ist $n = O(n)$ und $3n + 2 = O(n)$, aber $n \neq 3n + 2$.

Im Folgenden wird die Kompromiss-Notation $a_n \in O(b_n)$ verwendet werden.

(b) Es gilt immer $o(b_n) \subseteq O(b_n)$, denn jede Nullfolge ist nach Satz 5.3.5 auch beschränkt.

(c) Aus dem gleichen Grund gilt das folgende wichtige Kriterium:

$$\Big(\frac{a_n}{b_n}\Big)_{n\in\mathbb{N}} \text{ konvergent} \Rightarrow a_n \in O(b_n).$$

(d) Anschaulich bedeutet $a_n \in O(b_n)$, dass die Folge (a_n) höchstens so schnell wächst wie ein Vielfaches von (b_n), vgl. die folgende Übungsaufgabe.

Übungsaufgabe 5.4.3. Zeigen Sie für $(a_n), (b_n) \in F_+$ die folgenden Aussagen:

(a) $a_n \in O(b_n)$ genau dann, wenn es ein $C > 0$ und ein $n_0 \in \mathbb{N}$ gibt mit $a_n \leq Cb_n$ für alle $n \geq n_0$.

(b) $a_n \in o(b_n)$ genau dann, wenn es für jedes $C > 0$ ein $n_0 \in \mathbb{N}$ gibt mit $a_n \leq Cb_n$ für alle $n \geq n_0$.

Beispiel 5.4.4. Es seien $(a_n) = (2n^2 + 3n + 1)_{n\in\mathbb{N}}$, $(b_n) = (n^2)_{n\in\mathbb{N}}$ und $(c_n) = (n^7)_{n\in\mathbb{N}}$. Dann gilt $a_n \in O(b_n)$ und $a_n \in o(c_n)$, denn

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{2n^2+3n+1}{n^2} = \lim_{n\to\infty} \frac{2+\frac{3}{n}+\frac{1}{n^2}}{1} = 2$$

und

$$\lim_{n\to\infty} \frac{a_n}{c_n} = \frac{2n^2+3n+1}{n^7} = \frac{\frac{2}{n^5}+\frac{3}{n^6}+\frac{1}{n^7}}{1} = 0.$$

Allgemein gilt für jedes Polynom $p(n) = a_0 + a_1 n + \cdots + a_k n^k$ vom Grad k, dass $p(n) \in O(n^k)$ ist.

Satz 5.4.5. *Es seien $(a_n), (b_n), (c_n), (d_n) \in F_+$ und $\alpha, \beta \in \mathbb{R}_+$. Dann gilt*

(a) Sind $a_n, b_n \in O(c_n)$, so ist auch $\alpha a_n + \beta b_n \in O(c_n)$.

(b) Gilt $a_n \in O(b_n)$ und $c_n \in O(d_n)$, so ist $a_n c_n \in O(b_n d_n)$.

(c) Aus $a_n \in O(b_n)$ und $b_n \in O(c_n)$ folgt $a_n \in O(c_n)$. (Transitivität)

(d) $a_n \in O(b_n)$ genau dann, wenn $\dfrac{1}{b_n} \in O\Big(\dfrac{1}{a_n}\Big)$.

Weiterhin gelten alle diese Aussagen auch mit Klein-O anstelle von Groß-O.

Beweis. Wir führen den Beweis für große Os, die kleinen verbleiben als Übung.

(a) Nach Voraussetzung sind die Folgen (a_n/c_n) und (b_n/c_n) beschränkt, also gibt es Konstanten $C_1, C_2 > 0$ mit $a_n/c_n \le C_1$ und $b_n/c_n \le C_2$ für alle $n \in \mathbb{N}$. Dann ist auch

$$\frac{\alpha a_n + \beta b_n}{c_n} = \alpha\frac{a_n}{c_n} + \beta\frac{b_n}{c_n} \le \alpha C_1 + \beta C_2,$$

d.h. die Folge $((\alpha a_n + \beta b_n)/c_n)$ ist beschränkt, woraus $\alpha a_n + \beta b_n \in O(c_n)$ folgt.

(b) Nach Voraussetzung existieren wieder $C_1, C_2 > 0$ mit $a_n/b_n \le C_1$ und $c_n/d_n \le C_2$ für alle $n \in \mathbb{N}$. Also ist

$$\frac{a_n c_n}{b_n d_n} = \frac{a_n}{b_n} \cdot \frac{c_n}{d_n} \le C_1 C_2$$

für alle $n \in \mathbb{N}$ und damit $a_n c_n \in O(b_n d_n)$.

(c) Gilt $a_n/b_n \leq C_1$ und $b_n/c_n \leq C_2$ für alle $n \in \mathbb{N}$, so haben wir für all diese n auch

$$\frac{a_n}{c_n} = \frac{a_n}{b_n} \cdot \frac{b_n}{c_n} \leq C_1 C_2,$$

woraus die Behauptung folgt.

(d) Für die Richtung „⇒“ sei $a_n \in O(b_n)$. Dann gibt es ein $C > 0$, so dass $a_n/b_n \leq C$ für alle $n \in \mathbb{N}$ gilt. Damit ist dann ebenfalls für alle $n \in \mathbb{N}$

$$\frac{1/b_n}{1/a_n} = \frac{a_n}{b_n} \leq C,$$

also haben wir $1/b_n \in O(1/a_n)$.

Für die umgekehrte Beweisrichtung wendet man obiges Argument nochmals auf $1/b_n$ an. □

Bemerkung 5.4.6. Am häufigsten findet man die folgenden Landau-Symbole. Die entsprechende Komplexität eines Algorithmus wird auch mit passenden Namen versehen:

Landau-Symbol	Bezeichnung	Bemerkung
$O(1)$	beschränkt	
$O(\log_a(n))$	logarithmisch	$a > 1$
$O(n)$	linear	
$O(n \log_a(n))$	„n log n“	$a > 1$
$O(n^2)$	quadratisch	
$O(n^3)$	kubisch	
$O(n^k)$	polynomial	$k \in \mathbb{N}^*$
$O(a^n)$	exponentiell	$a > 1$

Die Darstellung in obiger Tabelle ist nach Größe der Menge sortiert. Es gilt also $O(1) \subseteq O(\log_a(n)) \subseteq O(n) \subseteq O(n \log_a(n)) \subseteq O(n^2) \subseteq O(n^3) \subseteq O(n^k) \subseteq O(n^\ell) \subseteq O(a^n)$ für $3 \leq k \leq \ell$.

Beispiel 5.4.7. (a) Was ist die kleinste Klasse aus Bemerkung 5.4.6 in der

$$a_n = 5n - 3\ln(n) + 9n^2 + 3n\ln(n) + n^3 + 0.1 \cdot 2^n$$

liegt? Nach Satz 5.4.5 (a) ist die Summe von mehreren Termen immer in der größten der beteiligten O-Mengen enthalten. Das größte Wachstum hat nach der Aufstellung in dieser Bemerkung hier der Term $0.1 \cdot 2^n$, also ist $a_n \in O(2^n)$.

(b) Exponentielle Algorithmen sind viel schlechter als polynomiale. Für die Laufzeit in Mikrosekunden gilt

$a_n \in$	$n = 10$	$n = 50$	$n = 100$
$O(n^2)$	100	2 500	10 000
$O(2^n)$	1 024	36 Jahre	$4 \cdot 10^{16}$ Jahre

5.5. Reihen

In diesem Abschnitt steht der Buchstabe $\mathbb{K}$ wieder für $\mathbb{R}$ oder $\mathbb{C}$.

Definition 5.5.1. *Es sei (a_n) eine Folge in $\mathbb{K}$. Dann heißt*

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + a_3 + \dots$$

die Reihe über (a_n). *Für jedes $k \in \mathbb{N}$ heißt dann $s_k := \sum_{n=0}^{k} a_n$ die k-te* Teilsumme *oder* Partialsumme *der Reihe.*
Ist die Folge $(s_k)_{k\in\mathbb{N}}$ konvergent, so nennen wir die Reihe konvergent *mit dem* Reihenwert

$$\sum_{n=0}^{\infty} a_n := \lim_{k\to\infty} s_k = \lim_{k\to\infty} \sum_{n=0}^{k} a_n,$$

ist (s_k) divergent, so nennen wir auch die Reihe divergent.

Beispiel 5.5.2. (a) In Beispiel 5.3.12 (b) haben wir schon eine Reihe betrachtet ohne sie so zu nennen, nämlich die *geometrische Reihe* $\sum_{n=0}^{\infty} q^n$. Diese ist nach dem dortigen Resultat konvergent, wenn $q \in \mathbb{C}$ mit $|q| < 1$ ist und dann gilt

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}, \qquad |q| < 1.$$

(b) Wir betrachten die Reihe

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \dots.$$

Es gilt für jedes $k \in \mathbb{N}^*$

$$\begin{aligned}\sum_{n=1}^{k} \frac{1}{n(n+1)} = \sum_{n=1}^{k} \frac{n+1-n}{n(n+1)} = \sum_{n=1}^{k} \Big(\frac{1}{n} - \frac{1}{n+1}\Big) = \sum_{n=1}^{k} \frac{1}{n} - \sum_{n=1}^{k} \frac{1}{n+1} \\ = 1 + \sum_{n=2}^{k} \frac{1}{n} - \sum_{n=2}^{k} \frac{1}{n} - \frac{1}{k+1} = 1 - \frac{1}{k+1}.\end{aligned}$$

Also ist

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{k\to\infty} \sum_{n=1}^{k} \frac{1}{n(n+1)} = \lim_{k\to\infty} \Big(1 - \frac{1}{k+1}\Big) = 1.$$

(c) Die Reihe

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

heißt *harmonische Reihe*. Für die $2k$-te Partialsumme $s_{2k} = \sum_{n=1}^{2k} 1/n$ gilt

$$s_{2k} = \sum_{n=1}^{k} \frac{1}{n} + \underbrace{\frac{1}{k+1}}_{\geq 1/(2k)} + \underbrace{\frac{1}{k+2}}_{\geq 1/(2k)} + \underbrace{\ldots}_{\ldots} + \underbrace{\frac{1}{2k}}_{\geq 1/(2k)} \geq s_k + k \cdot \frac{1}{2k} = s_k + \frac{1}{2}.$$

Nehmen wir nun an, dass die Reihe konvergent ist, so ist nach Definition der Reihenkonvergenz die Folge $(s_k)_{k\geq 1}$ konvergent. Den Grenzwert nennen wir s. Nach Satz 5.3.24 (b) konvergiert dann auch die Teilfolge $(s_{2k})_{k\geq 1}$ gegen s und wir bekommen nach Satz 5.3.7 (c) die Ungleichung

$$s = \lim_{k\to\infty} s_{2k} \geq \lim_{k\to\infty} \left(s_k + \frac{1}{2}\right) = s + \frac{1}{2}$$

und damit einen sauberen Widerspruch. Die harmonische Reihe ist also divergent.

Der folgende Satz ergibt sich direkt aus den Grenzwertsätzen für Folgen.

Satz 5.5.3. *Seien $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ zwei konvergente Reihen in $\mathbb{K}$ und $\alpha, \beta \in \mathbb{K}$. Dann ist auch $\sum_{n=0}^{\infty}(\alpha a_n + \beta b_n)$ konvergent und es gilt*

$$\sum_{n=0}^{\infty}(\alpha a_n + \beta b_n) = \alpha \sum_{n=0}^{\infty} a_n + \beta \sum_{n=0}^{\infty} b_n.$$

Im Allgemeinen ist das konkrete Ausrechnen eines Reihenwertes ein sehr schwieriges Unterfangen. Deshalb sind die wenigen Reihen, bei denen man den Wert exakt angeben kann, wertvolle Schätze. Einen besonders wichtigen solchen wollen wir nun noch ohne Beweis heben:

Satz 5.5.4. *Es gilt* $\displaystyle\sum_{n=0}^{\infty} \frac{1}{n!} = \mathrm{e}.$

Da es für kompliziertere Reihen schon schwer genug ist, überhaupt herauszubekommen, ob sie konvergieren oder nicht (vom Berechnen des Reihenwertes reden wir schon gar nicht), sind Konvergenzkriterien essentiell wichtig. Wir beginnen mit einem notwendigen Kriterium

Satz 5.5.5. *Ist $\sum_{n=0}^{\infty} a_n$ eine konvergente Reihe in $\mathbb{K}$, so ist (a_n) eine Nullfolge in $\mathbb{K}$.*

Beweis. Da die Reihe konvergiert, gilt $\lim_{k\to\infty}\sum_{n=0}^{k} a_n = \lim_{k\to\infty}\sum_{n=0}^{k-1} a_n = \sum_{n=0}^{\infty} a_n =: s$. Da außerdem für jedes $k \in \mathbb{N}^*$

$$a_k = \sum_{n=0}^{k} a_n - \sum_{n=0}^{k-1} a_n$$

gilt, ist $\lim_{k\to\infty} a_k = s - s = 0$. □

Hier sind noch zwei Kriterien, die sich direkt aus den entsprechenden Aussagen für Folgen ergeben:

Satz 5.5.6. *Es sei (a_n) eine Folge in $\mathbb{K}$ und $s_k := \sum_{n=0}^{k} a_n$, $k \in \mathbb{N}$. Dann gilt*

(a) Ist $a_n \geq 0$ für alle $n \in \mathbb{N}$ und $(s_k)_{k\in\mathbb{N}}$ nach oben beschränkt, so ist $\sum_{n=0}^{\infty} a_n$ konvergent. (Monotonie-Kriterium)

(b) Die Reihe $\sum_{n=0}^{\infty} a_n$ ist genau dann konvergent, wenn für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert mit

$$\Big|\sum_{n=\ell+1}^{k} a_n\Big| < \varepsilon \quad \text{für alle } k, \ell \in \mathbb{N} \text{ mit } k > \ell \geq n_0. \quad \text{(Cauchy-Kriterium)}$$

Beweis. (a) Die Konvergenz der Reihe ist nach Definition gleichbedeutend mit der Konvergenz der Folge $(s_k)_{k\in\mathbb{N}}$. Da alle $a_n \geq 0$ sind, gilt für diese

$$s_{k+1} = \sum_{n=0}^{k+1} a_n = \sum_{n=0}^{k} a_n + a_{k+1} \geq \sum_{n=0}^{k} a_n = s_k$$

für alle $k \in \mathbb{N}$. Also ist $(s_k)_{k\in\mathbb{N}}$ monoton wachsend und nach Voraussetzung nach oben beschränkt. Die Konvergenz folgt damit aus Satz 5.3.15.

(b) Wir zeigen, dass $(s_k)_{k\in\mathbb{N}}$ eine Cauchyfolge ist, dann folgt die Konvergenz aus dem Cauchy-Kriterium, Satz 5.3.20. Sei dazu $\varepsilon > 0$. Dann gibt es nach Voraussetzung ein $n_0 \in \mathbb{N}$, so dass für alle $k, \ell \geq n_0$ mit $k > \ell$ gilt

$$|s_k - s_\ell| = \Big|\sum_{n=0}^{k} a_n - \sum_{n=0}^{\ell} a_n\Big| = \Big|\sum_{n=\ell+1}^{k} a_n\Big| < \varepsilon.$$ □

Das folgende Konvergenzkriterium behandelt reelle Folgen, deren Summanden wechselnde Vorzeichen haben.

Satz 5.5.7 (Leibniz-Kriterium)**.** *Es sei (a_n) eine monoton fallende Folge in $\mathbb{R}$ mit $\lim_{n\to\infty} a_n = 0$. Dann ist die Reihe $\sum_{n=0}^{\infty}(-1)^n a_n$ konvergent.*

Beispiel 5.5.8. Das Leibniz-Kriterium liefert z.B. sehr schnell die Konvergenz der *alternierenden harmonischen Reihe*

$$\sum_{n=0}^{\infty}(-1)^n \frac{1}{n+1}.$$

Der Reihenwert $(\ln(2))$ ist dagegen deutlich schwerer zu bestimmen.

5.5.1. Absolute Konvergenz

In diesem Abschnitt geht es vor allem darum, weitere, alltagstauglichere Konvergenzkriterien für Reihen anzugeben. Dazu benötigen wir zunächst einen weiteren Begriff, der eine Verschärfung der Konvergenz bedeutet.

Definition 5.5.9. *Eine Reihe $\sum_{n=0}^{\infty} a_n$ in $\mathbb{K}$ heißt* absolut konvergent*, wenn die Reihe $\sum_{n=0}^{\infty} |a_n|$ in $\mathbb{K}$ konvergiert.*

Satz 5.5.10. *Jede absolut konvergente Reihe $\sum_{n=0}^{\infty} a_n$ in $\mathbb{K}$ ist auch konvergent in $\mathbb{K}$ und es gilt die* verallgemeinerte Dreiecksungleichung

$$\Big|\sum_{n=0}^{\infty} a_n\Big| \leq \sum_{n=0}^{\infty} |a_n|.$$

Beweis. Wir verwenden das Cauchy-Kriterium. Sei dazu $\varepsilon > 0$ gegeben. Dann gibt es dank der absoluten Konvergenz ein $n_0 \in \mathbb{N}$, so dass

$$\Big|\sum_{n=\ell+1}^{k} |a_n|\Big| = \sum_{n=\ell+1}^{k} |a_n| < \varepsilon$$

für alle $k > \ell \geq n_0$ gilt. Mit der Dreiecksungleichung gilt dann sofort auch

$$\Big|\sum_{n=\ell+1}^{k} a_n\Big| = |a_{\ell+1} + a_{\ell+2} + \cdots + a_k| \leq |a_{\ell+1}| + |a_{\ell+2}| + \cdots + |a_k| = \sum_{n=\ell+1}^{k} |a_n| < \varepsilon.$$

Also ist die Reihe nach dem Cauchy-Kriterium konvergent.
Mit demselben Dreicksungleichungs-Argument erhalten wir nun

$$\Big|\sum_{n=0}^{\infty} a_n\Big| = \Big|\lim_{k\to\infty} \sum_{n=0}^{k} a_n\Big| = \lim_{k\to\infty} \Big|\sum_{n=0}^{k} a_n\Big| \leq \lim_{k\to\infty} \sum_{n=0}^{k} |a_n| = \sum_{n=0}^{\infty} |a_n|. \qquad \square$$

Bemerkung 5.5.11. (a) Ein Beispiel dafür, dass die Umkehrung dieses Satzes nicht gilt, ist die alternierende harmonische Reihe aus Beispiel 5.5.8. Diese ist nach dem dortigen Ergebnis konvergent, aber nicht absolut konvergent, denn die zugehörige Reihe mit Beträgen ist die harmonische Reihe und diese divergiert, vgl. Beispiel 5.5.2 (c).

(b) Absolut konvergente Reihen haben gegenüber konvergenten einen großen Vorteil: Ist eine Reihe nur konvergent, aber nicht absolut konvergent, so kann man durch bloßes Umsortieren der Summanden den Reihenwert verändern. Dieser Effekt tritt bei absolut konvergenten Reihen nicht auf. Deren Reihenwert ist von der Summationsreihenfolge unabhängig.

Eine häufige Methode zur Konvergenzuntersuchung ist der Vergleich der zu untersuchenden Reihe mit einer bekannten Reihe. Der folgende Satz ist mit Hilfe des Cauchy-Kriteriums schnell zu beweisen, er ist aber so intuitiv, dass wir auf den Beweis hier auch verzichten können.

Satz 5.5.12. *Es seien (a_n) und (b_n) reelle Folgen und $n_0 \in \mathbb{N}$.*

(a) Ist $|a_n| \leq b_n$ für alle $n \geq n_0$ und konvergiert die Reihe $\sum_{n=0}^{\infty} b_n$, so ist $\sum_{n=0}^{\infty} a_n$ absolut konvergent. (Majorantenkriterium)

(b) Ist $a_n \geq b_n \geq 0$ für alle $n \geq n_0$ und divergiert die Reihe $\sum_{n=0}^{\infty} b_n$, so divergiert auch die Reihe $\sum_{n=0}^{\infty} a_n$. (Minorantenkriterium)

Bemerkung 5.5.13. (a) Die Vergleichsfolge (b_n) im obigen Satz heißt im Fall von Teil (a) *konvergente Majorante* und im Teil (b) *divergente Minorante*.

(b) Mit Hilfe der O-Notation kann man den Satz auch folgendermaßen formulieren:

Es sei (a_n) eine Folge in $\mathbb{R}$ und $(b_n) \in F_+$.

- Ist $\sum_{n=0}^{\infty} b_n$ konvergent, so konvergiert die Reihe $\sum_{n=0}^{\infty} a_n$ absolut, falls $|a_n| \in O(b_n)$ gilt.
- Ist $\sum_{n=0}^{\infty} b_n$ divergent, so divergiert auch die Reihe $\sum_{n=0}^{\infty} a_n$, falls $(a_n) \in F_+$ gilt und $b_n \in O(a_n)$ gilt.

Beispiel 5.5.14. (a) Da die harmonische Reihe divergiert, divergieren nach dem Minorantenkriterium die Reihen über alle langsamer fallenden Folgen. Insbesondere ist

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$$

für alle $\alpha < 1$ divergent, denn dann gilt $n^\alpha \leq n$ und damit

$$0 \leq \frac{1}{n} \leq \frac{1}{n^\alpha} \qquad \text{für alle } n \in \mathbb{N}^*.$$

(b) Die Reihe

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

ist dagegen absolut konvergent. Dazu beobachten wir zunächst, dass

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=0}^{\infty} \frac{1}{(n+1)^2}$$

ist. Weiter gilt für alle $n \in \mathbb{N}^*$ die Abschätzung $(n+1)^2 \geq n(n+1)$ und damit

$$\frac{1}{(n+1)^2} \leq \frac{1}{n(n+1)} =: b_n.$$

Die Reihe über b_n ist nach Beispiel 5.5.2 (b) konvergent, kann uns also als konvergente Majorante dienen und wir erhalten die behauptete Konvergenz aus dem Majorantenkriterium.

Bemerkung 5.5.15. Tatsächlich ist die Reihe $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ genau dann konvergent, wenn $\alpha > 1$ ist. Die harmonische Reihe ist also genau der Grenzfall.

Wir kommen nun zu zwei weiteren häufig verwendeten Konvergenzkriterien.

Satz 5.5.16. *Es sei $\sum_{n=0}^{\infty} a_n$ eine Reihe in $\mathbb{K}$.*

(a) (Wurzelkriterium) *Existiert der Grenzwert $\lim_{n\to\infty} \sqrt[n]{|a_n|}$, so ist die Reihe*

- *absolut konvergent, wenn $\lim_{n\to\infty} \sqrt[n]{|a_n|} < 1$ ist und*
- *divergent, wenn $\lim_{n\to\infty} \sqrt[n]{|a_n|} > 1$ ist.*

(b) (Quotientenkriterium) *Ist $a_n \neq 0$ für alle $n \in \mathbb{N}$ und existiert der Grenzwert $\lim_{n\to\infty} |a_{n+1}/a_n|$, so ist die Reihe*

- *absolut konvergent, wenn $\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} < 1$ ist und*
- *divergent, wenn $\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} > 1$ ist.*

Beweis. Wir beweisen nur exemplarisch das Wurzelkriterium. Es sei also (a_n) eine Folge in $\mathbb{K}$, für die $\alpha := \lim_{n\to\infty} \sqrt[n]{|a_n|}$ existiert.
Ist $\alpha < 1$, so wählen wir ein $q \in (\alpha, 1)$. Dank der Konvergenz von $(\sqrt[n]{|a_n|})$ gegen α, muss es ein $n_0 \in \mathbb{N}$ geben, so dass $|\sqrt[n]{|a_n|} - \alpha|$ für alle $n \geq n_0$ kleiner ist als der Abstand von α zu q. Insbesondere ist also

$$\sqrt[n]{|a_n|} \leq q < 1 \qquad \text{für alle } n \geq n_0.$$

Daraus folgt $|a_n| \leq q^n$ für alle $n \geq n_0$. Da $q \in (0,1)$ ist, ist die Reihe $\sum_{n=0}^{\infty} q^n$ nach Beispiel 5.5.2 (a) konvergent. Also liefert uns das Majorantenkriterium, Satz 5.5.12 (a) auch die absolute Konvergenz unserer Reihe $\sum_{n=0}^{\infty} a_n$.
Es sei nun $\alpha > 1$. Dann gibt es ein $n_0 \in \mathbb{N}$, so dass $\sqrt[n]{|a_n|} > 1$ für alle $n \geq n_0$ gilt. Damit ist auch $|a_n| > 1$ für alle $n \geq n_0$ und (a_n) ist definitiv keine Nullfolge. Nach Satz 5.5.5 kann dann die Reihe $\sum_{n=0}^{\infty} a_n$ nicht konvergent sein. □

Bemerkung 5.5.17. Liefert der Grenzwert in einem der beiden Kriterien genau Eins, so kann man daraus keine Aussage über Konvergenz oder Divergenz der Reihe ableiten, vgl. das folgende Beispiel.

Beispiel 5.5.18. (a) Wie wir gesehen haben, ist die Reihe $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ je nach der Größe von α konvergent oder divergent, vgl. Beispiel 5.5.14 und Bemerkung 5.5.15. Für jedes $\alpha \in \mathbb{Q}$ gilt allerdings

$$\lim_{n\to\infty} \sqrt[n]{\left|\frac{1}{n^\alpha}\right|} = \lim_{n\to\infty} \frac{1}{\sqrt[n]{n^\alpha}} = \left(\lim_{n\to\infty} \sqrt[n]{n}\right)^{-\alpha} = 1^{-\alpha} = 1.$$

Dies zeigt, dass im Falle, dass das Wurzelkriterium als Grenzwert 1 liefert, keine Aussage über die Konvergenz der Reihe möglich ist.

Gleiches gilt für das Quotientenkriterium.

(b) Für jedes $z \in \mathbb{C}$ betrachten wir die Reihe

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Im Fall $z = 0$ ist die Konvergenz der Reihe offensichtlich, da sie dann nur einen Summanden hat. Sei also nun $z \neq 0$. Dann gilt mit $a_n := z^n/n!$

$$\frac{|a_{n+1}|}{|a_n|} = \frac{\left|\frac{z^{n+1}}{(n+1)!}\right|}{\left|\frac{z^n}{n!}\right|} = \frac{|z|^{n+1} n!}{|z|^n (n+1)!} = \frac{|z|}{n+1}.$$

Also ist für jedes $z \neq 0$

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \frac{|z|}{n+1} = 0 < 1.$$

Nach dem Quotientenkriterium ist also die Reihe für jedes $z \in \mathbb{C}$ absolut konvergent.

Man nennt diese überaus wichtige Reihe die *Exponentialreihe*. Diese definiert uns eine Funktion

$$E : \begin{cases} \mathbb{C} & \to \mathbb{C} \\ z & \mapsto E(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}, \end{cases}$$

die *Exponentialfunktion*. Wie es zu diesem Namen kommt, wird in Kürze klar werden.

5.5.2. Das Cauchy-Produkt

In Satz 5.5.3 haben wir gesehen, dass die Summe konvergenter Reihen wieder konvergent ist. Wir wollen uns nun dem Produkt zuwenden. Naiv könnte man folgendermaßen rechnen:

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = (a_0 + a_1 + a_2 + a_3 + \dots)(b_0 + b_1 + b_2 + b_3 + \dots)$$

$$\begin{aligned}
&= a_0b_0 \\
&\quad + a_0b_1 + a_1b_0 \\
&\quad + a_0b_2 + a_1b_1 + a_2b_0 \\
&\quad + a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0 \\
&\quad + \dots \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k}.
\end{aligned}$$

Diese Rechnung ist allerdings im Allgemeinen falsch! Wir haben hier nämlich die Summanden in einer willkürlichen Reihenfolge aufsummiert und nach Bemerkung 5.5.11 (b) ist der Reihenwert nicht immer von der Summationsreihenfolge unabhängig. Bei absolut konvergenten Reihen allerdings schon und tatsächlich gilt der folgende Satz.

Satz 5.5.19. *Es seien $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ zwei absolut konvergente Reihen in $\mathbb{K}$. Dann konvergiert auch die Reihe $\sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k}$ absolut und es gilt für die Reihenwerte*

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k} = \Big(\sum_{n=0}^{\infty} a_n\Big)\Big(\sum_{n=0}^{\infty} b_n\Big).$$

Die Reihe $\sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k}$ heißt *Cauchy-Produkt* der beiden Reihen $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$.
Wir wollen diesen Satz hier nicht beweisen sondern einmal anwenden, indem wir die *Funktionalgleichung der Exponentialfunktion* zeigen.

Satz 5.5.20. *Für alle $z, w \in \mathbb{C}$ gilt $E(z+w) = E(z)E(w)$.*

Beweis. Es gilt mit Hilfe des Cauchy-Produkts, da alle beteiligten Reihen absolut konvergent sind,

$$\begin{aligned}
E(z)E(w) &= \Big(\sum_{n=0}^{\infty}\frac{z^n}{n!}\Big)\Big(\sum_{n=0}^{\infty}\frac{w^n}{n!}\Big) = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{z^k}{k!}\frac{w^{n-k}}{(n-k)!} \\
&= \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}z^k w^{n-k} = \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}z^k w^{n-k}.
\end{aligned}$$

Nun gilt nach der Binomialformel 5.2.9 (c)

$$\sum_{k=0}^{n}\binom{n}{k}z^k w^{n-k} = (z+w)^n,$$

also ist zusammengefasst

$$E(z)E(w) = \sum_{n=0}^{\infty}\frac{(z+w)^n}{n!} = E(z+w). \qquad \square$$

Das ist aber nur der erste Streich. In Satz 5.5.4 haben wir gesehen, dass

$$E(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = \mathrm{e}$$

gilt. Mit obigem Resultat ist damit für jedes $k \in \mathbb{N}^*$

$$E(k) = E(\underbrace{1 + 1 + \cdots + 1}_{k \text{ Mal}}) = E(1)^k = \mathrm{e}^k.$$

Weiter sieht man sofort, dass

$$E(0) = \sum_{n=0}^{\infty} \frac{0^n}{n!} = 1 + 0 + 0 + 0 + \cdots = 1$$

gilt. Also muss für jedes $k \in \mathbb{N}^*$ gelten

$$1 = E(0) = E(k + (-k)) = E(k)E(-k).$$

Das liefert uns, dass $E(k) \neq 0$ ist und $E(-k) = E(k)^{-1} = (\mathrm{e}^k)^{-1} = \mathrm{e}^{-k}$ für jedes $k \in \mathbb{N}$ gilt. Zusammen haben wir nun schon $E(k) = \mathrm{e}^k$ für jedes $k \in \mathbb{Z}$.
Doch hier hören wir nicht auf. Es sei nun $q = k/\ell \in \mathbb{Q}$ mit $k \in \mathbb{Z}$ und $\ell \in \mathbb{N}^*$. Zunächst beobachten wir, dass mit dem gleichen Trick wie oben

$$\mathrm{e} = E(1) = E\Big(\ell \cdot \frac{1}{\ell}\Big) = E\Big(\underbrace{\frac{1}{\ell} + \frac{1}{\ell} + \cdots + \frac{1}{\ell}}_{\ell \text{ Mal}}\Big) = E\Big(\frac{1}{\ell}\Big)^{\ell}$$

ist und damit

$$\mathrm{e}^{1/\ell} = E\Big(\frac{1}{\ell}\Big).$$

Das liefert schließlich

$$E(q) = E\Big(k \cdot \frac{1}{\ell}\Big) = E\Big(\frac{1}{\ell}\Big)^k = (\mathrm{e}^{1/\ell})^k = \mathrm{e}^{k/l} = \mathrm{e}^q \qquad \text{für alle } q \in \mathbb{Q}.$$

Es liegt also nahe e^x für alle reellen Zahlen x ebenfalls durch die Exponentialreihe zu definieren. Wir sind sogar gleich noch mutiger:

Definition 5.5.21. *Für alle $z \in \mathbb{C}$ ist*

$$\mathrm{e}^z := E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

5.6. Konvergenz in normierten Räumen

Folgen und Reihen kann man nicht nur in $\mathbb{R}$ oder $\mathbb{C}$, sondern auch in $\mathbb{R}^d$, $\mathbb{C}^d$, $d \in \mathbb{N}^*$, oder noch anderen Vektorräumen betrachten. Allerdings muss man, um den Begriff der Konvergenz einführen zu können, in irgendeiner Form Abstände messen können. Das führt uns wieder auf den Begriff des normierten Raums aus Abschnitt 3.4. Wie schon dort betrachten wir hier nur den Fall reeller Vektorräume. Im gesamten Abschnitt sei also V ein normierter $\mathbb{R}$-Vektorraum mit Norm $\|\cdot\|_V$.
Die Konvergenzdefinition ist genau identisch, wir ersetzen nur Beträge durch Normen:

Definition 5.6.1. *(a) Eine Folge $(a_n)_{n\in\mathbb{N}}$ in V heißt* konvergent *gegen ein $a \in V$, wenn für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert, so dass*

$$\|a_n - a\|_V < \varepsilon \quad \text{für alle } n \geq n_0$$

gilt.

Die Folge heißt divergent*, wenn sie nicht konvergent ist.*

(b) Eine Folge $(a_n)_{n\in\mathbb{N}}$ in V heißt Cauchy-Folge*, wenn es für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt mit*

$$\|a_n - a_m\|_V < \varepsilon \qquad \text{für alle } n, m \geq n_0.$$

(c) Eine Reihe $\sum_{n=0}^{\infty} a_n$ in V heißt konvergent *mit* Reihenwert $s \in V$*, wenn die Folge der Partialsummen $s_k := \sum_{n=0}^{k} a_n$, $k \in \mathbb{N}$, in V gegen s konvergiert.*

Konvergiert die Reihe $\sum_{n=0}^{\infty} \|a_n\|_V$ in $\mathbb{R}$, so heißt die Reihe $\sum_{n=0}^{\infty} a_n$ absolut konvergent*.*

Ist die Reihe nicht konvergent, so nennt man sie divergent*.*

Definition 5.6.2. *Eine Menge $M \subseteq V$ heißt* beschränkt*, falls es ein $C \geq 0$ gibt, so dass $\|x\|_V \leq C$ für alle $x \in M$ gilt.*

Beispiel 5.6.3. (a) In $V = \mathbb{R}^3$ mit der 1-Norm betrachten wir die Folge

$$a_n := \begin{pmatrix} 1 \\ \frac{1}{n} \\ \frac{n-1}{n} \end{pmatrix}, \qquad n \in \mathbb{N}^*.$$

Dann gilt $\lim_{n\to\infty} a_n = (1,0,1)^T$. Das sieht man so: Für jedes $n \in \mathbb{N}^*$ gilt

$$\|a_n - (1,0,1)^T\|_1 = |1-1| + \left|\frac{1}{n} - 0\right| + \left|\frac{n-1}{n} - 1\right| = \frac{1}{n} + \left|\frac{n-1-n}{n}\right| = \frac{2}{n}.$$

Ist nun $\varepsilon > 0$ vorgegeben, so gibt es ein $n_0 \in \mathbb{N}$ mit $n_0 > 2/\varepsilon$ und für alle $n \geq n_0$ gilt dann

$$\|a_n - (1,0,1)^T\|_1 = \frac{2}{n} \leq \frac{2}{n_0} < \frac{2}{2/\varepsilon} = \varepsilon.$$

(b) Die Folge $a_n = (n, 1/n)$, $n \in \mathbb{N}^*$, in $\mathbb{R}^2$ mit der 2-Norm ist wegen

$$\|a_n\|_2 = \sqrt{n^2 + 1/n^2} \geq \sqrt{n^2} = n \qquad \text{für alle } n \in \mathbb{N}^*$$

unbeschränkt. Da auch im normierten Raum jede konvergente Folge beschränkt ist, vgl. Übungsaufgabe 5.6.4 ist (a_n) damit divergent.

Übungsaufgabe 5.6.4. Übertragen Sie die Aussagen in Übungsaufgabe 5.3.3, Satz 5.3.5, Satz 5.3.7 (a) bis (b)ii), Satz 5.3.19, Satz 5.5.3 und Satz 5.5.5 in den Kontext von metrischen Räumen und beweisen Sie sie.

Satz 5.6.5. *Es sei* $(a_n)_{n\in\mathbb{N}} = \left((a_{n,1}, a_{n,2}, \ldots, a_{n,d})^T\right)_{n\in\mathbb{N}}$ *eine Folge in* $\mathbb{R}^d$ *mit der 2-Norm* $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}$. *Dann ist* (a_n) *in* $\mathbb{R}^d$ *genau dann konvergent, wenn für jedes* $j \in \{1, 2, \ldots, d\}$ *die* Koordinatenfolge $(a_{n,j})_{n\in\mathbb{N}}$ *in* $\mathbb{R}$ *konvergent ist. In diesem Fall ist*

$$\lim_{n\to\infty} \begin{pmatrix} a_{n,1} \\ a_{n,2} \\ \vdots \\ a_{n,d} \end{pmatrix} = \begin{pmatrix} \lim_{n\to\infty} a_{n,1} \\ \lim_{n\to\infty} a_{n,2} \\ \vdots \\ \lim_{n\to\infty} a_{n,d} \end{pmatrix}.$$

Beweis. „⇒“ Es sei (a_n) konvergent in $\mathbb{R}^d$ mit Grenzwert $a = (a_1, a_2, \ldots, a_d)^T \in \mathbb{R}^d$. Dann gilt für jedes $j \in \{1, 2, \ldots, d\}$ und alle $n \in \mathbb{N}$

$$|a_{n,j} - a_j| = \sqrt{(a_{n,j} - a_j)^2} \leq \sqrt{\sum_{k=1}^{d} (a_{n,k} - a_k)^2} = \|a_n - a\|_2.$$

Letzteres ist dank der Konvergenz von (a_n) eine Nullfolge in $\mathbb{R}$. Also konvergiert $(a_{n,j})$ in $\mathbb{R}$ gegen a_j.

„⇐“ Es seien nun für jedes $j \in \{1, 2, \ldots, d\}$ die Koordinatenfolgen $(a_{n,j})_{j\in\mathbb{N}}$ konvergent in $\mathbb{R}$ mit jeweiligem Grenzwert $a_j \in \mathbb{R}$. Dann gilt für jedes solche j auch $\lim_{n\to\infty} |a_{n,j} - a_j| = 0$ nach Übungsaufgabe 5.3.3 (b). Nach den Grenzwertsätzen ist dann $\sum_{j=1}^{d} (a_{n,j} - a_j)^2$ ebenfalls konvergent mit Grenzwert $\sum_{j=1}^{d} 0^2 = 0$. Schließlich folgern wir aus Bemerkung 5.3.10 (a)

$$\lim_{n\to\infty} \|a_n - a\|_2 = \lim_{n\to\infty} \sqrt{\sum_{j=1}^{d} (a_{n,j} - a_j)^2} = 0.$$

Nach Übungsaufgabe 5.6.4 folgt daraus die Konvergenz von (a_n) in $\mathbb{R}^d$ gegen $a = (a_1, a_2, \ldots, a_d)^T$. □

Beispiel 5.6.6. Die Folge in $\mathbb{R}^3$ mit

$$a_n = \begin{pmatrix} \frac{2n^2-n}{4n^2-3n+5} \\ (1+1/n)^n \\ \sqrt[n]{n} \end{pmatrix}$$

ist mit Hilfe dieses Satzes sehr schnell auf Konvergenz untersucht. Wir müssen uns nur die Koordinatenfolgen anschauen. Es ist

$$\lim_{n\to\infty} \frac{2n^2-n}{4n^2-3n+5} = \frac{1}{2}, \quad \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = \mathrm{e} \quad \text{und} \quad \lim_{n\to\infty} \sqrt[n]{n} = 1,$$

also ist (a_n) konvergent in $\mathbb{R}^3$ (mit der 2-Norm) und der Grenzwert ist $(1/2, \mathrm{e}, 1)^T$.

Bemerkung 5.6.7. Ein identischer Satz gilt für jede mögliche Norm auf $\mathbb{R}^d$. Wir werden später sehen, dass in endlichdimensionalen $\mathbb{R}$-Vektorräumen die Wahl der Norm für die Frage der Konvergenz keine Rolle spielt: Wenn eine Folge bezüglich einer Norm konvergiert, dann auch bezüglich jeder anderen und die Grenzwerte sind gleich.

In $\mathbb{R}$ haben wir offene und abgeschlossene Intervalle betrachtet. Mengen, zu denen ihr Rand gar nicht oder komplett dazugehört, spielen auch in normierten Räumen eine wichtige Rolle. Wir wollen nun die entsprechenden Begriffe definieren.

Definition 5.6.8. *(a) Es seien $x_0 \in V$ und $r \in (0,\infty)$. Dann heißt die Menge*

$$B_r(x_0) := \{x \in V : \|x - x_0\|_V < r\}$$

(offene) Kugel um x_0 *mit* Radius r.

(b) Eine Menge $M \subseteq V$ heißt offen*, falls es für jeden Punkt $x_0 \in M$ einen Radius $r > 0$ gibt, so dass $B_r(x_0) \subseteq M$ gilt.*

(c) Eine Menge $M \subseteq V$ heißt abgeschlossen*, wenn die Menge $M^c = V \setminus M$ offen ist.*

(d) Es sei $M \subseteq V$. Ein Punkt $x_0 \in M$ heißt innerer Punkt *von M, falls es ein $r > 0$ gibt, so dass $B_r(x_0) \subseteq M$ ist. Man nennt*

$$M^\circ := \{x \in M : x \text{ innerer Punkt von } M\}$$

das Innere *von M.*

Bemerkung 5.6.9. (a) Ist eine Menge abgeschlossen, so bedeutet das anschaulich, dass ihr Rand zur Menge dazugehört. Umgekehrt ist eine offene Menge eine, die keinen ihrer Randpunkte enthält. Die Begriffe „Rand“ und „Randpunkt“ definieren wir hier nicht. Trotzdem sollte dies ein richtiges intuitives Gefühl für die Begriffe offen und abgeschlossen geben.

(b) Achtung: Mengen sind keine Türen! Die meisten Mengen sind weder offen noch abgeschlossen, betrachten Sie z.B. ein halboffenes Intervall in $\mathbb{R}$. Hüten Sie sich also vor dem Fehlschluss: Ich habe festgestellt, dass meine Menge nicht offen ist, also ist sie abgeschlossen...

Beispiel 5.6.10. (a) Für jeden Mittelpunkt $x_0 \in V$ und jeden Radius $r > 0$ ist die eben definierte Kugel $B_r(x_0)$ eine offene Menge.

Um das einzusehen wählen wir ein $x \in B_r(x_0)$. Wir müssen nun zeigen, dass es einen Radius $\varrho > 0$ gibt, so dass $B_\varrho(x) \subseteq B_r(x_0)$ gilt. Da $x \in B_r(x_0)$ ist, gilt $\|x - x_0\|_V < r$. Die Zahl

$$\varrho := \frac{r - \|x - x_0\|_V}{2}$$

ist also strikt größer als Null.

Sei nun $y \in B_\varrho(x)$. Dann gilt nach der Dreiecksungleichung

$$\|y - x_0\|_V = \|y - x + x - x_0\|_V \leq \|y - x\|_V + \|x - x_0\|_V.$$

Der erste Summand ist kleiner als $\varrho = (r - \|x - x_0\|_V)/2$, denn y ist ja in der Kugel um x mit Radius ϱ. Also erhalten wir

$$\|y - x_0\|_V < \frac{r}{2} - \frac{\|x - x_0\|_V}{2} + \|x - x_0\|_V = \frac{r}{2} + \frac{\|x - x_0\|_V}{2}.$$

Weiter war $\|x - x_0\|_V < r$, also finden wir

$$\|y - x_0\|_V < \frac{r}{2} + \frac{r}{2} = r.$$

Damit haben wir $B_\varrho(x) \subseteq B_r(x_0)$ gezeigt und sind fertig.

(b) Die Kugel mit Rand $\{x \in V : \|x - x_0\|_V \leq r\}$ ist dagegen für jedes $x_0 \in V$ und alle $r > 0$ eine abgeschlossene Menge. Das sieht man am einfachsten mit Hilfe des folgenden Satzes ein.

Satz 5.6.11. *Eine Teilmenge M von V ist genau dann abgeschlossen, wenn für jede Folge in M, die in V konvergiert, der Grenzwert ein Element aus M ist.*

Beweis. „⇒“ Es sei $M \subseteq V$ abgeschlossen und (a_n) eine Folge in M, die in V konvergiert. Wir nehmen nun an, es wäre $a := \lim_{n\to\infty} a_n \notin M$, d.h. $a \in M^c$.

Da M abgeschlossen ist, ist die Menge M^c nach Definition offen. Es gibt also ein $r > 0$ mit $B_r(a) \subseteq M^c$. Weiter ist a der Limes der Folge (a_n). Also gibt es ein $n_0 \in \mathbb{N}$, so dass

$$\|a_n - a\|_V < r \quad \text{für alle } n \geq n_0$$

gilt. Das bedeutet $a_n \in B_r(a) \subseteq M^c$ für alle $n \geq n_0$ und wir haben einen Widerspruch zu der Voraussetzung, dass $a_n \in M$ für alle $n \in \mathbb{N}$ gilt.

„⇐“ Wir nehmen an, M wäre nicht abgeschlossen, d.h. M^c ist nicht offen. Dann gibt es ein $x_0 \in M^c$, so dass die Kugeln $B_r(x_0)$ für kein $r > 0$ ganz in M^c liegen. Anders gesagt, für jedes $r > 0$ gilt $B_r(x_0) \cap M \neq \emptyset$. Betrachten wir speziell $r = 1/n$ für jedes $n \in \mathbb{N}^*$, so erhalten wir für jedes $n \in \mathbb{N}^*$ ein $a_n \in B_{1/n}(x_0) \cap M$.

Die so konstruierte Folge (a_n) ist nun eine Folge in M, für die

$$\|a_n - x_0\|_V \leq \frac{1}{n} \qquad \text{für alle } n \in \mathbb{N}^*$$

gilt. Damit ist die Folge (a_n) konvergent in V mit $\lim_{n\to\infty} a_n = x_0$. Nach Voraussetzung muss also $x_0 \in M$ liegen, was ein Widerspruch zu $x_0 \in M^c$ ist. □

Damit können wir nun die Behauptung aus Beispiel 5.6.10 (b) fertig begründen. Es sei also (a_n) eine Folge in $M := \{x \in V : \|x - x_0\|_V \leq r\}$, die in V konvergiert und wir setzen $a := \lim_{n\to\infty} a_n$. Da für jede konvergente Folge (b_n) in V

$$\| \lim_{n\to\infty} b_n\|_V = \lim_{n\to\infty} \|b_n\|_V$$

gilt (vgl. Satz 5.3.7 (a) und Übungsaufgabe 5.6.4), haben wir

$$\|a - x_0\|_V = \| \lim_{n\to\infty} a_n - x_0\|_V = \| \lim_{n\to\infty} (a_n - x_0)\|_V = \lim_{n\to\infty} \|a_n - x_0\|_V.$$

Weiter liegt jedes a_n in M, also folgt nun aus der Monotonie des Grenzwertes (Satz 5.3.7 (c))

$$\|a - x_0\|_V \leq \lim_{n\to\infty} r = r.$$

Damit ist auch $a \in M$ und M somit nach Satz 5.6.11 abgeschlossen.

Übungsaufgabe 5.6.12. (a) Zeigen Sie, dass $M \subseteq V$ genau dann offen ist, wenn $M = M^\circ$ ist.

(b) Diskutieren Sie, ob $\emptyset$ und V offene und/oder abgeschlossene Mengen in V sind.

Definition 5.6.13. *Ist V ein endlichdimensionaler normierter $\mathbb{R}$-Vektorraum, so heißt eine Teilmenge $M \subseteq V$* kompakt*, wenn sie abgeschlossen und beschränkt ist.*

Warnung 5.6.14. Auch in unendlichdimensionalen Räumen gibt es den Begriff einer kompakten Teilmenge (und eigentlich wird er sogar erst dort richtig wichtig). Dann sieht allerdings die Definition völlig anders aus, und es gibt dann Mengen, die abgeschlossen und beschränkt aber nicht kompakt sind!
Im Moment kann Ihnen das egal sein. Diese Warnung soll nur vorbeugen, dass Sie gewarnt sind, wenn Sie in Ihrem Leben einmal über unendlichdimensionale normierte Räume stolpern und den Kompaktheitsbegriff brauchen.

Die Definitionen von Teilfolge und Häufungswert können wir wieder aus der eindimensionalen Situation abschreiben, indem wir Beträge durch Normen ersetzen.

Definition 5.6.15. *Es sei (a_n) eine Folge in $(V, \|\cdot\|_V)$.*

(a) Ein $a \in V$ heißt Häufungswert *von (a_n), falls für jedes $\varepsilon > 0$ die Menge*

$$\big\{n \in \mathbb{N} : \|a_n - a\|_V < \varepsilon\big\} = \big\{n \in \mathbb{N} : a_n \in B_\varepsilon(a)\big\}$$

unendlich viele Elemente hat.

(b) Ist $\{n_1, n_2, n_3, \dots\}$ eine unendliche Teilmenge von $\mathbb{N}$ mit $n_1 < n_2 < n_3 < \dots$, so heißt $(a_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge *von (a_n).*

Übungsaufgabe 5.6.16. Übertragen Sie Satz 5.3.24 mitsamt Beweis in die Situation von normierten Räumen.

Satz 5.6.17 (Satz von Bolzano-Weierstraß)**.** *Sei $(V, \|\cdot\|_V)$ ein endlichdimensionaler normierter Raum und $M \subseteq V$ kompakt. Dann besitzt jede Folge in M eine konvergente Teilfolge mit Grenzwert in M.*

Bemerkung 5.6.18. Häufig findet man die Aussage dieses Satzes auch in der folgenden Formulierung: „Ist $(V, \|\cdot\|_V)$ ein endlichdimensionaler normierter Raum, so besitzt jede beschränkte Folge in V mindestens einen Häufungswert."
Anschaulich bedeutet das: In einer beschränkten Teilmenge des $\mathbb{R}^n$ ist nicht genug Platz, als dass man mit einer Folge so wild herumspringen kann, dass sie sich nirgends häuft. Anders gesagt: Wenn man unendlich viele Punkte in einer beschränkten Menge unterbringen will, so müssen die irgendwo klumpen.

Definition 5.6.19. *Ein normierter $\mathbb{R}$-Vektorraum $(V, \|\cdot\|_V)$ heißt* vollständig*, wenn jede Cauchy-Folge in V konvergiert. Ein vollständiger normierter $\mathbb{R}$-Vektorraum wird auch* Banachraum *genannt.*
Wird die Norm $\|\cdot\|_V$ außerdem durch ein Skalarprodukt auf V induziert, so nennt man V Hilbertraum*.*

Beispiel 5.6.20. (a) Der Standardvektorraum $\mathbb{R}^d$ ist für jedes $d \in \mathbb{N}^*$ mit jeder darauf definierten Norm ein Banachraum. Wählt man speziell die durch das Standardskalarprodukt induzierte 2-Norm, so ist $(\mathbb{R}^d, \|\cdot\|_2)$ sogar ein Hilbertraum.

(b) Die Menge

$$\ell^2 := \Big\{(a_n) : (a_n) \text{ reelle Folge, so dass } \sum_{n=0}^{\infty} a_n^2 \text{ konvergent}\Big\}$$

ist ein $\mathbb{R}$-Vektorraum. Weiter ist die Abbildung $(\cdot|\cdot) : \ell^2 \times \ell^2 \to \mathbb{R}$, die für zwei Folgen $(a_n), (b_n) \in \ell^2$ durch

$$\big((a_n)\big|(b_n)\big) := \sum_{n=0}^{\infty} a_n b_n$$

definiert ist, ein Skalarprodukt. Mit der davon induzierten Norm

$$\|(a_n)\|_2 = \sqrt{((a_n)|(a_n))} = \Big(\sum_{n=0}^{\infty} a_n^2\Big)^{\frac{1}{2}}$$

wird ℓ^2 zu einem Hilbertraum.

(c) Die Menge

$$\ell^1 := \big\{(a_n) : (a_n) \text{ reelle Folge, so dass } \sum_{n=0}^{\infty} |a_n| \text{ konvergent}\big\}$$

ist mit der Norm $\|(a_n)\|_1 := \sum_{n=0}^{\infty} |a_n|$ ein Banachraum, der kein Hilbertraum ist.

Übungsaufgabe 5.6.21. Die folgenden Resultate aus den Abschnitten 5.3 und 5.5 gelten in beliebigen Banachräumen: Satz 5.3.19, Satz 5.5.10, Satz 5.5.12 (a) (Majorantenkriterium) und Satz 5.5.16 (Wurzel- und Quotientenkriterium). Übertragen Sie die Aussagen und Beweise.

Zum Abschluss dieses Abschnittes wollen wir noch einen wichtigen Satz beweisen, den Banach'schen Fixpunktsatz. Dieser gibt unter anderem ein einfaches Kriterium an eine Iterationsvorschrift an, das deren Konvergenz garantiert.

Satz 5.6.22 (Banach'scher Fixpunktsatz)**.** *Es sei $(V, \|\cdot\|_V)$ ein Banachraum, $M \subseteq V$ abgeschlossen und $f : M \to M$ eine Funktion. Weiter existiere ein $q \in (0, 1)$, so dass*

$$\|f(x) - f(y)\|_V \leq q\|x - y\|_V \qquad \textit{für alle } x, y \in M$$

gilt. Dann gelten die folgenden Aussagen:

(a) Es gibt genau ein $v \in M$ mit $f(v) = v$. (D.h. f hat genau einen Fixpunkt *in M.)*

(b) Für jedes $x_0 \in M$ konvergiert die Folge (x_n) mit $x_{n+1} = f(x_n)$, $n \in \mathbb{N}$, gegen v und es gelten die folgenden Fehlerabschätzungen für jedes $n \in \mathbb{N}^*$:

$$\|x_n - v\|_V \leq \frac{q^n}{1-q}\|x_1 - x_0\|_V \qquad \textit{(A-priori-Abschätzung),}$$

$$\|x_n - v\|_V \leq \frac{q}{1-q}\|x_n - x_{n-1}\|_V \qquad \textit{(A-posteriori-Abschätzung).}$$

Beweis. Wir wählen ein beliebiges $x_0 \in M$ und betrachten die in der Formulierung des Satzes schon erwähnte Folge mit $x_{n+1} = f(x_n)$ für jedes $n \in \mathbb{N}$. Wir zeigen zunächst

$$\|x_{n+1} - x_n\|_V \leq q^n \|x_1 - x_0\|_V \qquad \text{für alle } n \in \mathbb{N}$$

per Induktion. Für $n = 0$ lautet diese Ungleichung $\|x_1 - x_0\|_V \leq \|x_1 - x_0\|_V$, ist also wahr. Wir wenden uns dem Induktionsschritt von n nach $n+1$ zu. Tatsächlich gilt mit der Voraussetzung an f

$$\|x_{n+2} - x_{n+1}\|_V = \|f(x_{n+1}) - f(x_n)\|_V \leq q\|x_{n+1} - x_n\|_V$$

und dann der Induktionsvoraussetzung

$$\leq qq^n \|x_1 - x_0\|_V = q^{n+1}\|x_1 - x_0\|_V.$$

Nun wollen wir als nächstes zeigen, dass (x_n) eine Cauchyfolge in V ist. Seien dazu zwei Indizes $n, m \in \mathbb{N}$ mit $m > n$ gegeben. Dann gilt mit dem Ergebnis von eben

$$\begin{aligned}
\|x_m - x_n\|_V &= \|x_m - x_{m-1} + x_{m-1} - x_{m-2} + \cdots - x_{n+1} + x_{n+1} - x_n\|_V \\
&= \Big\|\sum_{k=n}^{m-1}(x_{k+1} - x_k)\Big\|_V \leq \sum_{k=n}^{m-1}\|x_{k+1} - x_k\|_V \leq \sum_{k=n}^{m-1} q^k \|x_1 - x_0\|_V \\
&= q^n \sum_{k=n}^{m-1} q^{k-n}\|x_1 - x_0\|_V = q^n \sum_{k=0}^{m-1-n} q^k \|x_1 - x_0\|_V \\
&\leq q^n \sum_{k=0}^{\infty} q^k \|x_1 - x_0\| = \frac{q^n}{1-q}\|x_1 - x_0\|_V. \qquad (5.2)
\end{aligned}$$

Wir beobachten nun zunächst, dass im Fall $x_1 = x_0$ aus dieser Ungleichung $x_m = x_n$ für alle $m > n \geq 0$ folgt. Wir haben dann also eine konstante Folge (x_n), die offensichtlich konvergent gegen x_0 ist. Es sei also in allen weiteren Überlegungen $x_1 \neq x_0$.
Sei nun $\varepsilon > 0$. Da $q \in (0, 1)$ gilt, konvergiert die Folge (q^n) gegen Null, also gibt es ein $n_0 \in \mathbb{N}$ mit $q^n < \varepsilon(1-q)/\|x_1 - x_0\|_V$ für alle $n \geq n_0$. Für alle $m > n \geq n_0$ gilt dann

$$\|x_m - x_n\|_V \leq \frac{q^n}{1-q}\|x_1 - x_0\|_V < \frac{\varepsilon(1-q)}{(1-q)\|x_1 - x_0\|_V}\|x_1 - x_0\|_V = \varepsilon.$$

Damit haben wir gezeigt, dass (x_n) eine Cauchyfolge in V ist. Da V nach Voraussetzung vollständig ist, ist dies also eine konvergente Folge in V, deren Grenzwert wir $v := \lim_{n\to\infty} x_n$ nennen.

Von diesem Grenzwert wollen wir nun natürlich zeigen, dass er ein Fixpunkt von f ist. Dazu beobachten wir, dass für jedes $n \in \mathbb{N}$ gilt

$$\begin{aligned}\|f(v) - v\|_V &= \|f(v) - x_n + x_n - v\|_V \leq \|f(v) - x_n\|_V + \|x_n - v\|_V \\ &= \|f(v) - f(x_{n-1})\|_V + \|x_n - v\|_V \leq q\|v - x_{n-1}\|_V + \|x_n - v\|_V.\end{aligned}$$

Geht man nun in dieser Ungleichung zum Grenzwert $n \to \infty$ über, so bekommt man

$$\|f(v)-v\|_V = \lim_{n\to\infty} \|f(v)-v\|_V \leq \lim_{n\to\infty} \big(q\|v-x_{n-1}\|_V + \|x_n-v\|_V\big) = q\cdot 0 + 0 = 0.$$

Die Definitheit der Norm liefert also $f(v) = v$.
Es bleiben noch die Eindeutigkeit des Fixpunktes und die beiden Fehlerabschätzungen zu zeigen. Zum Nachweis der Eindeutigkeit sei w ein weiteres Element von M mit $f(w) = w$. Dann gilt nach der Voraussetzung an f

$$\|v - w\|_V = \|f(v) - f(w)\|_V \leq q\|v - w\|_V.$$

Nehmen wir nun an, es wäre $v \neq w$, so folgt $\|v - w\|_V > 0$ und wir können durch diesen Ausdruck teilen. Das liefert den Widerspruch $1 \leq q$. Der Fixpunkt ist also eindeutig.
Die A-Priori-Abschätzung haben wir schon weiter oben fast gezeigt. Geht man nämlich in der Ungleichung (5.2) zum Grenzwert für $m \to \infty$ über, so erhält man

$$\|v - x_n\|_V = \lim_{m\to\infty} \|x_m - x_n\| \leq \lim_{m\to\infty} \Big(\frac{q^n}{1-q}\|x_1 - x_0\|_V\Big) = \frac{q^n}{1-q}\|x_1 - x_0\|_V.$$

Zum Beweis der A-posteriori-Abschätzung überlegen wir uns, dass für jedes $n \in \mathbb{N}^*$ gilt

$$\|v - x_n\|_V = \|f(v) - f(x_{n-1})\|_V \leq q\|v - x_{n-1}\|_V \leq q\big(\|v - x_n\|_V + \|x_n - x_{n-1}\|_V\big).$$

Daraus folgt

$$(1 - q)\|v - x_n\|_V \leq q\|x_n - x_{n-1}\|_V,$$

was nach Division durch $1 - q > 0$ die behauptete Abschätzung impliziert. □

Tabelle der griechischen Buchstaben

groß	klein	Name
A	α	Alpha
B	β	Beta
Γ	γ	Gamma
Δ	δ	Delta
E	ϵ, ε	Epsilon
Z	ζ	Zeta
H	η	Eta
Θ	θ, ϑ	Theta
I	ι	Iota
K	κ, $\varkappa$	Kappa
Λ	λ	Lambda
M	μ	My
N	ν	Ny
Ξ	ξ	Xi
O	o	Omikron
Π	π, ϖ	Pi
P	ρ, ϱ	Rho
Σ	σ, ς	Sigma
T	τ	Tau
Y	υ	Ypsilon
Φ	ϕ, φ	Phi
X	χ	Chi
Ψ	ψ	Psi
Ω	ω	Omega

Index